Evaluation of single nucleotide polymorphisms in virulence genes of *Mycobacterium tuberculosis* as markers of lineages and sub-lineages in Tshwane region.

By

UNARINE MATODZI

Submitted in partial fulfilment of the requirement for the degree

MAGISTER SCIENTAE

MSc (MEDICAL MICROBIOLOGY)

In the

Department of Medical Microbiology

Faculty of Health Sciences

University of Pretoria

Pretoria

South Africa

May 2020

The LORD has done this, and it is marvelous in our eyes

Psalms 118:25

DECLARATION

I, Unarine Matodzi, declare that the dissertation hereby submitted to the University of Pretoria for the MSc Degree (Medical Microbiology), and the work therein is original and neither whole or part of it has previously been submitted for another degree or at any university or tertiary education institution.

……..

Signature of candidate

ACKNOWLEDGEMENTS

First and foremost, I would like to give thanks to the Almighty God, for his guidance and divine power that enabled me to work earnestly throughout this degree.

This work has been accomplished through the contribution of my supervisors, family, and friends. Here is the list of those who helped me to made it possible.

A special gratitude to my supervisors, Dr. N.E. Maningi, Dr. A.M. Mutshembele and Dr. J Osei Sakyere for all your contribution and support in the successful completion of my research project.

I would like to thank Dr. H.M. Said for her supervision with my laboratory experiment and allowing me to carry out my laboratory work at the Center for Tuberculosis, NICD laboratory.

I would like to sincerely thank Prof R. Warren and Ms. C. Clarke for offering laboratory training (Middlebrook 7H9 culture and CTAB DNA extraction) and allowing me to carry out my work at the Faculty of Medicine and Health Sciences, Stellenbosch University.

A special appreciation to my Mom for giving me words encouragement and my family for supporting me throughout.

To my friends: O. Maluleke, J.C. Swart and T.G. Vele for your special support and encouragement.

I would like to thank NRF for funding my Degree, without them it would not have been possible.

I would like to Thank SAMRC, for their wonderful collaboration and provision of laboratory materials.

Table of contents

List of tables

List of figures

List of abbreviation

Ag	Antigen
AMK	Amikacin
ART	Antiretroviral
ATM	Amplification tagment mix
BCG	Bacille Calmette Guerin
BD	Becton Dickinson
BioEdit	Biological Sequence Alignment Editor
BSL	Biosafety level
CAS	Central Asian
CD	Cluster of differentiation
CM	Capreomycin
CS	Cycloserine
CTAB	Cetylmethylammonium bromide
dH2O	Distilled water
DNA	Deoxyribonucleotide
DR	Direct repeats
DST	Drug susceptibility test
EAI	East African-Indian
EDTA	Ethylenediaminetetraacetic acid
EMA	European Medicine Agency
EMB	Ethambutol
ETH	Ethionamide
fbp	Fibronectin binding protein
FDA	Food and Drug Administration
FQLs	Fluoroquinolones
gDNA	Genomic DNA

hbhA	Heparin-binding hemagglutinin
HCl	Hydrochloric acid
HIV	Human immunodeficiency virus
IFN- γ	Interferon gamma
IGRA	Interferon-gamma release assay
INH	Isoniazid
IS	Insertion sequence
iTOL	Interactive Tree of Life
KAN	Kanamycin
LAM	Latin American-Mediterranean
LJ	Lowenstein-Jensen
LNA	Library Normalization Additive
LNS	Library normalization storage buffer
LNW	Library normalization wash
LPA	Line probe assay
lpqH	Lipoprotein antigen
lspA	Lipoprotein signal peptidase
MAFFT	Multiple Sequence Alignment using Fast Fourier Transformation
mce	mammalian cell entry
MDR-TB	Multidrug-resistant tuberculosis
MGIT	Mycobacteria growth indicator tubes
MHC	Major histocompatibility complex
MIRU	Mycobacterial Interspersed Repetitive Units
MTBC	*Mycobacterium tuberculosis* complex
M. tuberculosis	*Mycobacterium tuberculosis*
MUSCLE	Multiple Sequence Comparison by Log-Expectation
NAAT	Nucleic acid amplification test
NaCl	Sodium chloride

NADH	Nicotinamide adenine dinucleotide dehydrogenase
NaOH	Sodium hydroxide
NICD	National Institute for Communicable Diseases
NHLS/TAD	National Health Laboratory Services/ Tshwane Academic Division
NPM	Nextera PCR master mix
NRF	National Research Funding
NTM	Nontuberculous mycobacterium
OADC	Oleic acid-albumin dextrose-catalase
OFL	Ofloxacin
PAMPs	Pathogen-associated molecular patterns
PANTA	polymyxin B, amphotericin B, nalidixic acid, trimethoprim, Azlocillin
PAS	Para-aminosalicylic acid
PATRIC	Pathosystem Resource Integration Center
PE	Pro-Glue
PZA	Pyrazinamide
RD	Region of difference
RFLP	Restriction fragments length polymorphism
RIF	Rifampicin
rip1	Regulated intermembrane proteolysis
ROS	Reactive oxygen species
RSB	Resuspension buffer
SAMRC	South African Medical Research Council
SDS	Sodium dodecyl sulfate
SLID	Injectable drugs
SM	Streptomycin
sodA	superoxide dismutase A
Spoligotyping	Spacer oligonucleotide

SSC	Standard sodium citrate
SSPE	Sodium chloride-sodium phosphate-EDTA
ST	Shared type
TA system	Toxin anti-toxin system
TB	Tuberculosis
TBE	Tris boric Ethylenediaminetetraacetic acid
TD	Tagment DNA
TE	Tris/Ethylenediaminetetraacetic acid
TLRs	Toll-like receptors
T7SS	Type VII secretion system
TST	Tuberculin skin test
U.S. A	United States of America
UV	Ultraviolet
VNTR	Variable number tandem repeats
WGS	Whole-genome sequencing
WHO	World Health Organization
XDR-TB	Extensively drug-resistant tuberculosis
zmp	Zinc-dependent metalloprotease
ZN	Ziehl-Neelsen

Summary

Tuberculosis (TB) is one of the top ten leading causes of death worldwide with millions of new TB cases reported every year. Understanding the genetic diversity of *Mycobacterium tuberculosis* (*M. tuberculosis*) is very crucial for rapid diagnosis and to reduce transmission of TB. Various diagnostic techniques, anti-tuberculosis reagents and vaccination are available, however, the disease is far from being eradicated (Brudey *et al*., 2006).

Mycobacterium tuberculosis is classified into seven major lineages that are key to the most research areas. Recently, multidrug *M. tuberculosis* have been reported as the most dangerous strains that cause a life-threatening TB. However, the *M. tuberculosis* with modified virulence and transmissibility, particularly those that are caused by mutations leading to genetic variation and increased pathogenicity are highly reported (Zaychikova *et al*., 2015). Genetic markers such as variable number tandem repeats, insertion sequence element and direct repeats have been used to identify lineages. However, the techniques (such as spoligotyping, IS6110-RLFP and MIRU-VNTR) that use these genetic markers have a lot of drawbacks and some have low discriminatory power (Mikheecheva *et al*., 2017).

Recently, single nucleotide polymorphisms (SNPs) are regarded as the most promising genetic markers for genotyping *M. tuberculosis* because they have low-level homoplasy and high discriminatory power (Zaychikova *et al*., 2015). The present study proposed that genotyping *M. tuberculosis* using polymorphisms in virulence genes may be an alternative approach to determine lineages and may help to detect the *M. tuberculosis* strains that are epidemiologically dangerous and have adapted to specific geographic regions. This study aimed to identify and evaluate a set of virulence gene SNPs as markers of *M. tuberculosis* strains circulating in the Tshwane region.

A total of 150 susceptible and resistant *M. tuberculosis* cultures stored in Mycobacteria growth indicator tubes (MGIT) tubes were collected from May to October 2018 at the National Health Laboratory Service, Tshwane Academic Division (NHLS/TAD) to conduct this study. The DNA was extracted using hexadecyltrimethylammonium bromide (CTAB) method and spoligotyping was done to screen for *M. tuberculosis* lineages. The Beijing and LAM genotypes detected by spoligotyping were sequenced using the Illumina Miseq platform. The bioinformatic analysis of

virulence genes in 56 genomes of *M. tuberculosis* belonging to Beijing and LAM genotypes was performed to detect lineage-specific SNPs markers.

Of the 150 *M. tuberculosis* collected, 57.3% were susceptible *M. tuberculosis* strains while 42.7% were drug-resistant TB. Spoligotyping of 150 isolates resulted to 86.7% previously shared type (ST) and 13.3% orphans yielding a clustering rate of 63.3%. The Beijing family was found to be the most predominant lineage by 26.7%, followed by T family (16%), LAM (13.3%), East Africa Indian (EAI) (8.7%), S (6%), Manu (4.7%), H (4.7%), CAS (4.0%) and X3 (2.7%).

The number of susceptible *M. tuberculosis* isolates per lineages was higher than drug-resistant TB with isolates detected as Beijing contributing 17.3% of all susceptible isolates, followed by isolates classified as orphans (10%), T family (9.3%), LAM family (8%) and CAS (2.67%). The association between anti-tuberculosis drug-resistant TB and lineages was found in EAI lineage (6.7%), Manu (4%) and S family (3.3%). The family with a high number of isolates which were drug-resistant TB was the EAI1-SOM sub-lineage belonging to the EAI family.

This study successfully identified 29 Beijing and 6 LAM signature SNPs that can be used to classify clinical *M. tuberculosis* isolates. Within these signature SNPs, *fadD*28 (1521 C>T), *eccCb*1 (1479 G>A), *pks*5 (6210 G>A), and *ponA*2 (372 G>T) were identified in the Beijing strains and *fadD*28 (1392 C>G) within the LAM strains that were not reported in previous studies. Furthermore, this study detected the lineage-specific SNPs: *mce*3*B* (145 T>G), *eccCb*1 (1556 G>T), *vapC*12 (95 A>G) in Beijing BO/W148 and *cyp*125 (1076 T>C), *mce*3*B* (44 T>C), *vapC*25 (221 A>C), *vapB*34 (140 C>A) F15/LAM4/KZN sub-lineages which have been reported to be virulent and associated with drug resistance.

This study showed a high genetic diversity of *M. tuberculosis* strains circulating within the Tshwane region. The Beijing lineage identified in this study was found to be more predominant than the rest of the identified genotypes. This study proposed the alternative method for genotyping *M. tuberculosis* strains using SNPs in virulence genes of *M. tuberculosis*. Observations from this study also highlight the advantage of using WGS technique over other genotyping methods such as IS6110-RFLP that has more drawbacks, as most genotypic methods discriminate *M. tuberculosis* strains using specific genes or regions in the genome of *M. tuberculosis* while WGS uses the complete genome of *M. tuberculosis* to determine different *M. tuberculosis* lineage.

Chapter 1

Introduction

1.1. Background

Globally, diseases caused by *Mycobacterium tuberculosis* (*M. tuberculosis*) remain a major problem, particularly as it is the single etiological agent implicated in most mortalities and morbidities than any pathogen presently (Trinh *et al*., 2018). Tuberculosis (TB), a major disease caused by *M. tuberculosis,* is thus the leading cause of deaths worldwide, estimated at 1.3 million in human immunodeficiency virus (HIV)-negative people, with an additional 300 000 deaths being recorded among HIV-positive people in 2018 (WHO, 2019).

Nowadays, the most dangerous TB is not only caused by drug-resistant strains of *M. tuberculosis* but also by strains that have modified virulence, transmission, and pathogenesis (Zaychikova *et al*., 2015). Virulence and pathogenicity of *M. tuberculosis* are conditioned by a variety of genes that lead to TB disease progression. These genes encode for many chemical and physical proteins, which are fundamental to *M. tuberculosis'* infection at every stage of its pathogenesis, including adhesion, colonization of mucosal membranes, inversion, evasion of host immune response, and survival under stress conditions (Mikheecheva *et al*., 2017). Mutations that occur in these genes may have an influence on the phenotypic and genotypic characteristics of pathogenic *M. tuberculosis*.

Genetic diversity of *M. tuberculosis* is caused by mutations that could affect the functionality and structural composition of a protein (Mikheecheva *et al*., 2017). The most dangerous type of mutation is the nonsynonymous mutation, which can cause a significant change in the resultant amino acid and have an impact on the phenotypic characteristics of a pathogen such as influencing the virulence and severity of infection (Bengtson *et al*., 2017). Mutations are known as the main cause of genetic variability; however, the severity of TB also depends on the host's health, genetic background, immune status, environmental status and composition of pathogen's virulence genes (virulome) (Cobat *et al*., 2013; Carding *et al*., 2015).

Mycobacterium tuberculosis is divided into seven main different lineages, which are characterized by genetic diversity that gradually occurs during evolution. These lineages represent the ancestry

and geographical origin of different strains belonging to *M. tuberculosis*. Lineage 1 comprises of East Africa-Indian (EAI) and some Manu spoligotype families, lineage 2 is the Beijing group, lineage 3 includes the Central Asia (CAS) strains and lineage 4 encompasses Ghana, Haarlem (H/T), X and Latin American-Mediterranean (LAM) strains. Lineages 5 and 6 are found frequently in Western Africa and they are related to the *M. africanum* strain, and lineage 7 has recently been reported in Ethiopia (Blouin *et al*., 2012; Yimer *et al*., 2015).

A variety of lineages and sub-lineages have been reported in different provinces in South Africa (Maguga-Phasha *et al*., 2017). Lineages such as Beijing, Haarlem, EAI, LAM, and X have been isolated in different provinces, but the Beijing lineage is the most prevalent genotype isolated across the country in almost every province (Ioerger *et al*., 2009; Maguga-Phasha *et al*., 2017). The F15/LAM4/KZN sub-lineage is a LAM variant that caused an extensively drug-resistant TB (XDR-TB) outbreak in KwaZulu-Natal province in 2006 and has been the commonest circulating TB strain in KwaZulu-Natal (Ioerger *et al*., 2009). There are speculations and suggestions that the distribution of different lineages in South Africa might be influenced by migration from different countries. For instance, a recent study done by Maguga-Phasha *et al*. (2017) detected a large number of Beijing and non-Beijing lineages in Mpumalanga and Limpopo Provinces, which share their borders with Swaziland, Zimbabwe, and Mozambique. However, TB in these provinces is also associated with poverty because they are mainly comprised of rural areas. In addition, South Africa has about 2.2 million foreigners living in different provinces as reported by Statistics South Africa in 2011, which could have an impact on the transmission of virulent *M. tuberculosis* strains from other countries (Peberdy, 2019).

Lineages are characterized by several genetic markers, such as insertion sequence (IS) element, direct repeats (DR) and variable number tandem repeats (VNTR) that have been used for lineage or sub-lineage identification (Mikheecheva *et al*., 2017). However, the techniques that use these genetic markers have low discriminatory power and some lineages have been missed by these techniques. Currently, single nucleotide polymorphisms (SNPs) are considered a promising genetic marker for genotyping *M. tuberculosis* due to their high discriminatory power and low level of homoplasy (Zaychikova *et al*., 2015). A recent study by Mikheecheva *et al*. (2017) showed the effective use of SNPs in virulence genes to identify *M. tuberculosis* lineages. This study found

a new sub-lineage, B0/N-90 inside the Beijing-B0/W148 lineage by SNPs in *irtB*, *mce3F* and *vapC46*.

In South Africa, Beijing and LAM genotypes are reported as the most prevalent genotypes circulating in different provinces and are associated with high virulence and drug resistance. Building on the work of Mikheecheva *et al.* (2017), which investigated Beijing and LAM strains from Russia, this study shall investigate the association between South African LAM or Beijing strains and SNPs of their virulence genes. This will verify the presence or absence of the same SNPs in the virulence genes of Beijing or LAM strains of South African origin as detected in the Russian strains.

1.2. Aim of the study

To identify and evaluate a set of virulence gene SNPs as markers of *M. tuberculosis* strains circulating in the Tshwane region.

1.3. Specific Objectives

- To collect susceptible and resistant *M. tuberculosis* from the National Health Laboratory Service, Tshwane Academic Division (NHLS/TAD)
- To determine lineages circulating in the Tshwane region using spoligotyping.
- To determine Beijing and LAM sub-lineages using IS6110-RFLP.
- To detect SNPs within the virulence genes using whole-genome sequencing.
- To develop a catalog of virulence gene SNPs associated with Beijing and LAM lineages and sub-lineages.
- To determine and compare the discriminatory power of SNPs in virulence genes with that of IS6110-RFLP genotyping.

Chapter 2

Literature review

2.1. Classification of *Mycobacterium tuberculosis*

*Mycobacterium tuberculosi*s is an intracellular pathogenic bacterium that belongs to the family *Mycobacteriaceae,* which is a well-known causative agent of TB, a life-threatening human disease. It was first discovered by a Nobel Prize winner, Robert Koch, in 1882 where it was previously known as "*Phthisis*" (Cambau and Drancourt, 2014). *Mycobacterium tuberculosis* belongs to the *Mycobacterium tuberculosis* complex (MTBC), a group of genetically related mycobacterium species that include *M. bovis, M. africanum, M. microti, M. canetti, M. caprae* and *M. pinnipedii*. The mycobacterium genus also includes over 50 environmental mycobacteria or nontuberculous mycobacterium (NTM), which causes a disease resembling tuberculosis (Omari, 2014).

*Mycobacterium tuberculosi*s is identified by its unique characteristics, although some of the features are shared amongst members of the MTBC. These characteristics are the ability to undergo a dormant phase (non-replicating state), causing a latent infection without developing any symptoms of infection or transmitting the infection to other individuals, ability to re-establish metabolically during chronic infection or during any state of immunosuppression, presence of a thick waxy cell wall, and its characteristically slow-growing nature as an obligate aerobe and facultative intracellular bacillus (Gordon and Parish, 2018).

Under a light microscope, *M. tuberculosis* is visualized as a long, thin, straight, or slightly curved rod measuring up to 0.5 µm by 3µm in size. It is arranged in small clumps, in pairs or as a single bacillus. The bacillus can change its morphology during any unfavorable environment to adapt and survive. It becomes very short in old cultures, filamentous inside the macrophage, ovoid during starvation and it is club-shaped or branched in extensively drug-resistant tuberculosis (XDR-TB) (Velayati and Farnia, 2012). This bacillus is metabolically flexible, which informs its classification as either prototrophic i.e. using carbon or nitrogen as a source to build its own components or as heterotrophic thus using coexisting organic compounds as a source of carbon and energy (Niederweis, 2008).

*Mycobacterium tuberculosi*s is an acid-fast bacterium; its thick waxy cell wall contains lipid-rich mycolic acids that retain carbolfuchsin dye after decolorization with acid alcohol when staining with Ziehl-Neelsen (ZN). This makes *M. tuberculosis* typical amongst other bacteria because it cannot be stained with other ordinary stains such as Gram stain since those stains cannot be absorbed or retained in its cell wall. However, the structure of its peptidoglycan layer raises suggestions that they are Gram-Positive bacteria (Gordon and Parish, 2018). The thickness, rigidity and hydrophobic condition of the cell wall makes it serve as a barrier to anti-tuberculosis drugs (Kumar *et al*., 2019).

2.2. Tuberculosis prevalence and burden

Tuberculosis is one of the top 10 leading causes of death worldwide with millions of new TB cases reported every year. Approximately 9.0 to 11.1 million people were estimated to have developed TB disease in 2018. In this estimation, 5.8 million were men, 3.2 million were women and about 1.0 million were children. The number of incidence cases in the world is estimated to have reached 87% in the 30 high TB burden countries as per the global TB report (WHO, 2019). Eight of these countries accounted for two-thirds of the global TB burden: India accounted for 27% of the total global TB burden, followed by China with 9%, Indonesia with 8%, the Philippines with 6%, Pakistan with 5%, Nigeria and Bangladesh with 4% each and South Africa with 3% (WHO, 2019).

Since 1993, TB has been declared a global health emergency and despite the major intervention programmes such as anti-TB regiments, vaccine and other preventive measures to stall the progression of the disease, it remains the topmost cause of death from an infectious disease globally (Floyd *et al*., 2018). In the Southern African regions, TB incidence was strongly associated with the HIV epidemic. According to the WHO, the HIV epidemic had the greatest effect on the health of people living in Southern Africa since 1990 (WHO, 2018). The annual national TB incidence in six Southern African countries have a huge impact on the global TB burden. Nevertheless, implementation of antiretroviral (ART) treatment played a big role in the decline of TB burden observed from 2010 to 2017 in those six countries, showing an average decline of 18% in Eswatini (formerly Swaziland), 10% in Zimbabwe, 8% in Botswana, 7% in Lesotho and South Africa, and 6% in Namibia (WHO, 2019).

Findings supporting the pivotal impact that the HIV epidemic has on the health of individuals with TB in Southern Africa was evidenced by the outbreak of XDR-TB in the Tugela Ferry Hospital,

KwaZulu-Natal province, South Africa. The high mortality rates from this outbreak were reportedly due to a TB strain that was facilitated by HIV co-infection, which elevated the transmission rate of the drug-resistant strains among the most vulnerable populations (Klopper *et al*., 2013; Ismail *et al*., 2018).

2.3. Genetic diversity of *Mycobacterium tuberculosis*

Although members of the MTBC possess a variety of genetic and phenotypic characteristics, they are thought to have been derived from a common ancestor. Studies suggest that the phylogeny of MTBC begins from *M. canettii*, which is believed to be a clonal ancestor. However, this relies on speculations and it has not been established (Gutierrez *et al*., 2005). Genetic diversity amongst *M. tuberculosis* is caused by mutations that could affect the functionality and structural composition of a protein (Mikheecheva *et al*., 2017). Some mutations are silent (synonymous), having no effect on the resultant gene product or protein. Others are nonsense mutations that code for stop codons, resulting in immature gene products, while some mutations are missense (nonsynonymous) that lead to changes in amino acid sequences (Turnpenny and Ellard, 2016). The nonsynonymous mutations can result in substantial changes in the genetic and phenotypic characteristics of a pathogen, which can influence virulence and severity of infection (Bengtson *et al*., 2017).

*M*ycobacterium *tuberculosis* strains are classified into seven main, deep-branching, different lineages (Blouin *et al*., 2012). These lineages are characterized by several genetic markers, such as IS element, VNTR, DR and SNPs. These genetic markers are results of mutations that cause genetic variations, and they are used for lineage or sub-lineage identification based on the specific technique that may be used to genotype them (Mikheecheva *et al*., 2017). Lineage 1 comprises of East Africa India (EAI) and some Manu spoligotype families, lineage 2 is the Beijing group, lineage 3 includes the Central Asian (CAS) strains and lineage 4 is the Euro-American genotype that encompasses Ghana, Haarlem, X and Latin American Mediterranean (LAM) strains. Lineages 5 and 6 are found frequently in Western Africa and correspond to the *M. africanum* strain and lineage 7 is comprised of Ethiopian strains (Blouin *et al*., 2012).

Lineages represent the origin of various strains of *M. tuberculosis* (Figures 2.1a and 2.1b). Global distribution of lineages is due to population migration from one country or continent to another (Brudey *et al*., 2006; Gagneux, 2013). It is speculated that *M. tuberculosis* might have historically originated from the horn of Africa and through migration, lineages-specific patterns were

distributed to other geographical parts of the world. Lineage 1 was originally discovered in the regions of the Indian Ocean, broadening to the Eastern part of Africa and extending to Melanesia (O'Neill *et al*., 2018). Some studies have reported that the EAI lineage is more prevalent in South-East Asia, specifically in the Philippines, Myanmar, and Malaysia, in Vietnam and Thailand (Brudey *et al*., 2006).

The Beijing lineage has been found to be the most predominant lineage associated with large outbreaks in different countries (Stavrum *et al*., 2009). This lineage was first isolated in Beijing, China, and has been found to be endemically prevalent in parts of Southern Asia, South Africa, and Northern Eurasia (Bhanu *et al*., 2002).

Corresponding to lineage 1, lineage 3 was also distributed from the Indian Ocean to Southeastern Asia. This lineage was divided into CAS-Delhi type and the CAS-Kilimanjaro. The CAS-Delhi type was predominantly recovered in Delhi, India, and in the Middle East, Central and Southern Asia; it was also isolated in Iran, Pakistan, America, Europe and Australia, where it is suggested to have been distributed through migration from Southern Asia. In contrast to CAS-Delhi, the CAS-Kilimanjaro type was historically distributed only in Tanzania (Bhanu *et al*., 2002; McHugh *et al*., 2005; Brudey *et al*., 2006).

Lineage 4 is distributed within African countries, Europe and the entire Mediterranean region and is also known as the Euro-American lineage that includes major sub-lineages such as LAM, Haarlem and the X superfamily (Duchêne *et al*., 2004). In addition, it has been reported that the X superfamily is highly restricted to North and Central America, which is linked with the history of English colonization of areas such as the United Kingdom, the United States, Australia, South Africa and in the Caribbean. The Haarlem superfamily was first discovered in European countries where it was named after the Dutch town, Haarlem. It is characterised by the absence of spacers in regions 29-30 and 33-36, forming prototype ST127 and ST777 pattern. This superfamily has only three families that have been discovered to date: i.e. Haarlem 1, 2, 3, (Duchêne *et al*., 2004; Brudey *et al*., 2006). It is prevalent in America, Australia, Finland, Georgia, Iran, and Russia. However, ST777 is likely to be found in Saudi Arabia.

The LAM superfamily is characterized by the absence of spacers in regions 21–24 and 33–36, with multiple variants that are geographically prevalent in Latin America, Africa and the Mediterranean basin (Brudey *et al*., 2006). The LAM sub-lineage has recently been reported in South Africa,

particularly the LAM4 variant, which was the commonest TB strain associated with the XDR outbreak in KwaZulu-Natal province in 2006. It was given the name F15/LAM4/KZN (Singh *et al*., 2007).

Lineages 5 and 6 are known as West African 1 and 2 respectively and they are clades of *M. africanum*. These lineages are endemic and strictly limited to Western Africa, where they are reported to be prevalent. Nevertheless, few West African strains were found in Burkina Faso, and they are rare in East Africa. The structure of these two lineages is more similar to that of *M. africanum* than that of *M. tuberculosis* (Gomgnimbo *et al*., 2012). However, Lineage 6 is currently evolving, with recent findings showing that lineage 6 *M. africanum* obtained from Djibouti has spacer S8, which is absent from lineage 6 from West Africa (Brudey *et al*., 2006).

Lineage 7 has been reported recently in Ethiopia and among the Ethiopian immigrants in Djibouti. This lineage has originally been isolated and it is still restricted in Ethiopia, particularly in the Amhara region, which was then given the name 'Aethiops vetus'. Observations show that lineage 7 strains grow slowly than any other non-lineage 7 *M. tuberculosis* and is associated with very low virulence. This observation also indicated that lineage 7 is not widespread or highly transmissible (Yimer *et al*., 2015; Nebenzahl-Guimaras *et al*., 2016).

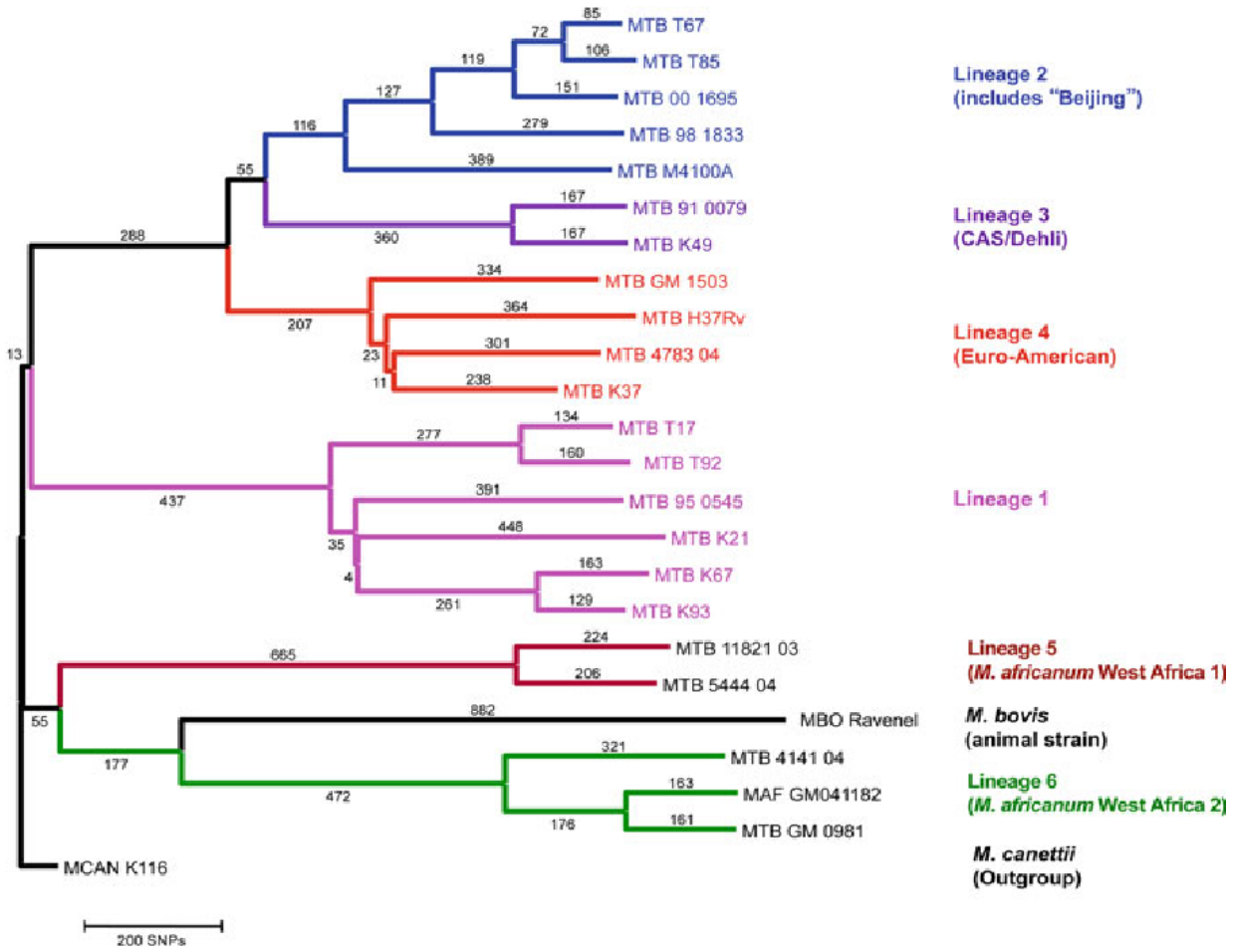

Figure 2. 1a: Illustration of the genetic diversity of *M. tuberculosis* complex showing their clonal structure and human-associated *M. tuberculosis* lineages. The *M. canetti* is used as an outgroup. Colored branches indicate human-associated *M. tuberculosis* lineages and numbers on branches indicate different SNPs (source: Gagneux, 2013).

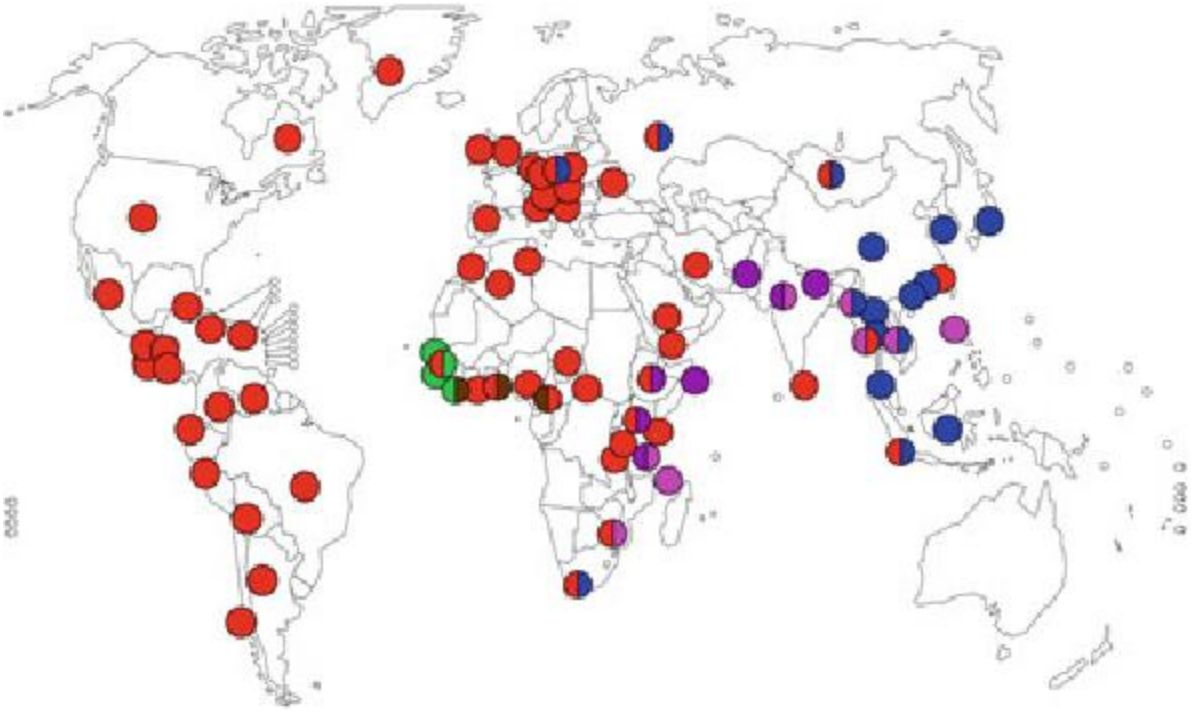

Figure 2.1b: Geographic distribution of the human-associated *M. tuberculosis* lineages. Each dot represents the country of origin, which resembles the dominant lineage(s) indicated by colors corresponding to Figure 2.1a (source: Gagneux *et al*., 2006).

2.4. Pathogenesis of *Mycobacterium tuberculosis*

Transmission of *M. tuberculosis* is through airborne pathway. About 1 to 5 µM in diameter of droplet nuclei containing *M. tuberculosis* are produced by people with active TB through sneezing, coughing and speaking, which may float through the air for longer periods until they get inhaled by other individuals (Omari, 2014). Pathogenesis begins when individuals inhale droplet nuclei containing tubercle bacilli that pass through their respiratory tract and are deposited at the alveoli in the lungs. Tubercle bacilli in the lungs are ingested by alveolar macrophages and by alveolar epithelial type II pneumocytes, which are found in greater numbers than alveolar macrophages. Dendritic cells invade the site of infection and play a key role as an antigen-presenting cell in the early stages of infection (Warner *et al*., 2015). In addition, dendritic cells are responsible for recruiting other cells of the immune system, activating T cells and disseminating the infection. During active TB, *M. tuberculosis* multiplies inside the macrophages and are released when the macrophages die (Lin *et al*., 2014). The engulfed mycobacteria are subjected to multiple bactericidal mechanisms that include phagolysosome fusion and respiratory burst to get rid of the

infection. Under normal circumstances, the ingested tubercle bacilli are contained inside the activated macrophages through the formation of a protective barrier shell called a granuloma. Nonetheless, virulent *M. tuberculosis* has its own strategy to evade or modulate the immune response in a way that favors itself and may help to activate latent or infectious active TB (Forrellad *et al*., 2013).

When *M. tuberculosis* gets into the bloodstream, it disseminates to different parts of the body and can spread to distant sites, such as peripheral lymph nodes, the kidneys, the central nervous system, and the bones, causing severe TB disease. Disseminated TB normally occurs in children and people infected with HIV, and it is known as miliary TB (Crowley, 2013).

2.5. *Mycobacterium tuberculosis* virulence genes and their role in the pathogenesis

Progression of the disease relies on the virulence factors possessed by the bacilli, which are encoded by numerous virulence genes (Lin *et al*., 2014). A variety of *M. tuberculosis* virulence genes encode enzymes responsible for the biosynthesis of diverse lipids, cell surface proteins, regulators and proteins of signal transduction systems. Some virulence genes are responsible for facilitating tolerance and survival of the bacteria inside the aggressive microenvironment of the host macrophages, thereby evading the fusion of the phagosome containing engulfed *M. tuberculosis* with the released lysosome (Forrellad *et al*., 2013). Unlike other pathogens, *M. tuberculosis* lacks distinctive toxins and most *M. tuberculosis* virulence genes are conserved in non-pathogenic mycobacteria species, suggesting that virulent species arose from non-pathogenic mycobacteria through adaptation to the intracellular environment (Smith, 2003).

Virulence in *M. tuberculosis* is coordinated by a set of genes contributing to pathogenesis. There are more than 300 virulence genes in *M. tuberculosis* that are divided into different categories depending on their function, molecular features and cellular localization. Virulence genes include proteins responsible for lipid biosynthesis, cell envelope proteins (cell wall proteins, lipoproteins and secretion systems), proteins responsible for evading the immune system and inhibiting macrophages (Forrellad *et al*., 2013).

Some of the essential virulence genes include those involved in oxidative and nitrosative stresses, phagosome arrest and inhibition of apoptosis, enzymatic (protein kinases, proteases, metalloproteases) reaction, and other proteins such as metal-transporter proteins (importers and

exporters), gene expression regulators, sigma factors, transcriptional regulators and proteins of unknown function (Forrellad *et al*., 2013; Mikheecheva *et al*., 2017). The products of these genes are fundamental to the various stages of *M. tuberculosis* infection: adhesion, colonization of mucosal membranes, invasion, evasion of host immune response, and survival under stress conditions (Mikheecheva *et al*., 2017).

2.5.1. Mycobacterium cell envelope

The cell envelope of *M. tuberculosis* is comprised of three major components: the outer membrane, the cell wall core and the plasma membrane (Figure 2.2) (Jackson, *et al*., 2013; Angala *et al*., 2014). *Mycobacterium tuberculosis*'s inner plasma membrane is typical of normal bacterial membranes. However, it also contains common components such as glycolipid, lipoprotein and lipoglycan. The outer plasma membrane is the cell wall core, which is made up of the peptidoglycan layer attached covalently by phosphoryl-*N*-acetylglucosaminosyl-rhamnosyl linkage units with the heteropolysaccharide arabinogalactan, which forms esters through the combination of non-reducing ends of α-alkyl and β-hydroxy long-chain (C_{60}-C_{90}) of mycolic acids (Barry *et al*., 2007). The cell wall mycolic acids are bound to the outer membrane, which is the layer buildup of a variety of non-covalently attached glycolipids, lipoproteins and lipoglycans. The outer membrane is covered by capsular-like structure, the outermost loose layer that consists of polysaccharides and proteins with few amounts of inner lipids. The proteins embedded in the capsule have major roles in the synthesis and maintenance of the cell wall and are vital for adhesion, transport of porins and survival of mycobacteria in the host cell during pathogenesis (Angala *et al*., 2014).

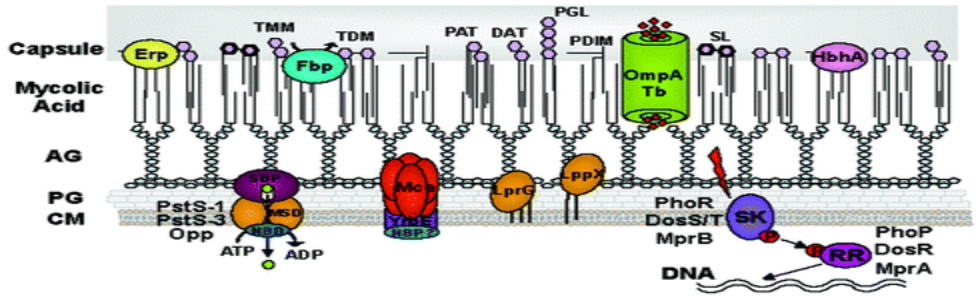

Figure 2. 2: A diagram illustrating the cell envelope proteins embedded in the outer membrane, cell wall core and the plasma membrane of the *M. tuberculosis* complex (source: Forrellad *et al*., 2013).

2.5.1.1. Cell wall proteins

The cell wall proteins consist of the outer membrane proteins that are located in the mycobacterial outer membrane bilayer. Most of these proteins are associated with the cell wall processes where more than 5% of these proteins are classified under virulence and the detoxifying category (Hoffmann *et al*., 2008). Approximately 144 proteins have been detected as outer membrane proteins that primarily participate in the uptake of hydrophobic compounds across outer membranes, efflux pump and nutrients uptake, and most importantly, they are accountable for attachment and invasion of the host cell as well as the degradation of host's tissue structures. Characterization of the cell wall proteins is fundamental to understanding bacterial survival and evasion from the host's immune response (Angala *et al*., 2014).

(i) Cell wall protein: Mammalian cell entry (*mce*1-4) protein

The *mce* is a surface-exposed protein organized in the large operon, which is comprised of six *mce* genes: A, B, C, D, E, and F. This gene encodes for adhesins or invasins, with putative signal sequences at the N-terminal of the mycobacterial cell wall, where it may facilitate the bacteria's invasion in the early stages of infection. Its name came from its function that confers to the mycobacterium the ability to enter the mammalian cell and survive inside the macrophages (Ahmad *et al*., 2004; Forrellad *et al*., 2013). There are four *mce* loci in *M. tuberculosis*, namely, *mce* 1-4. All six *mce* proteins are encoded by the *mce1* operon and *mce2A*, *mce3A* and *mce4A* are expressed during the early stages of infection. Furthermore, *mce*1*A* and *mce*1*E* are expressed and elicit the production of antibodies during natural infection with *M. tuberculosis* (Ahmad *et al*., 2004).

(ii) Cell wall protein: Fibronectin binding protein (*fbpA, B, C2*)

Fibronectin binding protein is the complex of three domains: *fbpA*, *fbpB* and *fbpC2*, which are also known as antigen 85 (Ag85) i.e. Ag85A (*fbpA)*, Ag85B (*fbpB*) and Ag85C (*fbpC2*). It mainly contributes to the virulence of *M. tuberculosis* through the synthesis of cell wall lipids and plays a role in the final assembly of the mycobacterial cell wall (Kremer *et al*., 2002). The name *fbp* comes from its ability to bind its domains to fibronectin (Forrellad *et al*., 2013). The ability to bind to fibronectin by Ag85, stimulate the complement receptor-mediated phagocytosis of *M. tuberculosis* and proliferation of their respective genes such as *fbpA*, *fbpB*, and *fbpC* (Belisle *et al*.,

1997). Moreover, adhesion of *M. tuberculosis* to the mucosal membrane is mediated by the binding of Ag85 to fibronectin, thus facilitating the bacteria's entry into the host cell (Forrellad *et al*., 2013).

(iii) Cell wall protein: Heparin-binding hemagglutinin (*hbhA*)

Heparin-binding hemagglutinin is a major adhesin, which assists with the attachment of *M. tuberculosis* to the epithelial cells and dissemination of mycobacterium from its primary site of infection. This protein influences agglutination of erythrocytes and induces aggregation of mycobacterium, thus promoting mycobacterium's capability to form biofilms (Forrellad *et al*., 2013). *Mycobacterium tuberculosis* uses *hbhA* proteins to hinder autophagy, an intracellular self-digestion whereby cytoplasmic components in a vacuole are degraded by lysosomes and amplify pathogenesis (Lamb *et al*., 2013; Zheng *et al*., 2017).

2.5.1.2. Lipoproteins

Lipoproteins make up a major component of the cell envelope in mycobacteria. They are involved in cellular functions such as protein transport mechanisms, cell wall metabolism, and adhesion to the mucosal membrane, signaling and protein degradation, which have significant roles in virulence. The *M. tuberculosis* genome encodes about 1% to 3% lipoprotein (Forrellad *et al*., 2013).

(i) Lipoprotein: 19-kDa lipoprotein antigen (*lpqH*)

Mycobacterium tuberculosis' 19-kDa lipoprotein antigen is described as a pathogen-associated molecular pattern (PAMPs), a molecular pattern from bacteria that signals through Toll-like receptors (TLRs) (Tobian *et al*., 2003). The 19-kDa lipoprotein antigen is a putative glycoprotein that hinders antigen presentation by macrophages and processing by major histocompatibility complex class two (MHC-II) through inhibition of interferon-gamma (IFN-γ) signaling by means of a TLR-2 dependent mechanism (Tobian *et al*., 2003; Forrellad *et al*., 2013). The MHC-II-restricted antigen presentation is inhibited so that *M. tuberculosis* can evade recognition by a cluster of differentiation 4^+ ($CD4^+$) T cells, causing a reduction of MHC-II synthesis. This lipoprotein induces maturation of dendritic cells, macrophages and neutrophils (causing neutrophil priming and activation), which promote apoptosis and allow the dissemination of mycobacteria (Forrellad *et al*., 2013). In addition, the 19kDa lipoprotein restricts phagosome maturation and

phagosome-antigen degradation in the myeloid differentiation factor 88-dependent manner, which inhibits the formation of peptide that binds to MHC-I and prevent recognition of *M. tuberculosis* by immune cells (Tobian *et al*., 2003).

(ii) Lipoprotein: 27-kDa lipoprotein antigen, (*lprG*)

The 27-kDa lipoprotein is also described as glycoprotein. This lipoprotein is a surface-secreted antigen that plays a role in virulence by suppressing the host‘s immune response. The immune response suppression is coordinated by a combination of operons and Rv1410c gene, a gene known as p55 that forms p27-p55 (*lprG-p55*) operon. This operon has been suggested to code for the P55 protein known as an antibiotic efflux pump wherein the conserved overexpression of *M. tuberculosis* P55 in *M. smegmatis* confers resistance to streptomycin and aminoglycosides (Forrellad *et al*., 2013).

2.5.2. Secretion system

Mycobacterium survives inside the host cell by escaping a stringent environment caused by the host cell’s immune response. Numerous genes encoding proteins secreted in that stringent environment have been identified and considered important in mycobacterium pathogenesis (Abdallah *et al*., 2007). Notably, the mycobacterium genome encodes up to five secretion systems. Nevertheless, only type II and VII secretion systems are involved in virulence.

2.5.2.1. Type II secretion system

All bacteria, including *M. tuberculosis* have a protein export system that helps them to transport their synthesized proteins which are essential in pathogenesis across the membrane (Forrellad *et al*., 2013). The proteins that need to be exported are crucial in pathogenesis and such pathways are essential to virulence because of their roles in exporting effectors that interact with the host. In mycobacterium, the type II secretion system is responsible for transporting lipoproteins which contain a lipobox motif in the C-terminal region of signal peptide across the membrane. The transported lipoproteins are processed and modified by the lipoprotein signal peptidase (*lspA)* which cleaves the lipobox motif in the C-terminal region of the signal peptide (Feltcher *et al*., 2010). Recent studies showed that *M. tuberculosis* has adopted a novel secretion system for transporting extracellular proteins across its hydrophobic and impermeable cell wall, which is an

important mechanism for effective operation in its natural environment for adaptation and survival (Feltcher *et al*., 2010; Forrellad *et al*., 2013).

2.5.2.2. Type VII secretion system

Mycobacterium tuberculosis contains a unique secretion system, which was described and named as the type VII secretion system (or T7SS) by Abdallah *et al.* (2007). The T7SS is encoded by the ESX gene clusters that have been named ESX-1 to -5 and is responsible for transporting proteins across the plasma membrane, but only ESX-1 and ESX-5 have been fully described and associated with virulence, except that combinations of these ESX genes secrete large numbers of proteins (Houben *et al*., 2014). Upon interacting with the host, these systems export toxins and signal proteins that allow the mycobacterium to cause tissue damage (Forrellad *et al*., 2013). The T7SSs mediate inflammasome activation and IL1b release during mycobacterial infections. The model of T7SS-induced translocation, necrosis, and IL-1b secretion (figure 2.3) occur when tubercle bacilli are phagocytosed and translocated to the cytosol through ESX-1–secreted effector proteins. The phagosomal translocation results in the cytosolic release of active cathepsin B and ESX-5 effector proteins. These effector proteins cause inflammasome activation, IL-1b secretion, and cell death (Abdallah *et al*., 2011).

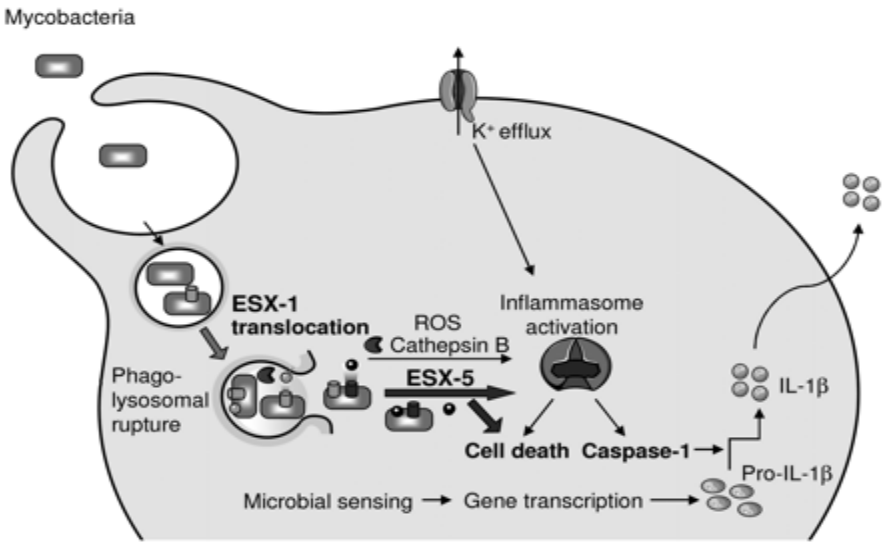

Figure 2. 3: Illustration of T7SS-induced translocation, necrosis, and IL-1β secretion. (source: Abdallah *et al*., 2011).

(i) Type VII secretion system: ESX-1

During pathogenesis, T7SS secrete ESX-1, a model of two proteins known as 6-kDa early secretory antigen target (ESAT6) and 10-kDa culture filtrate protein (CFP10) which are located in a segment called the region of difference 1 (RD1). These proteins were initially known as ESXA and ESXB respectively, and they are essential for the cell to cell migration of the pathogenic members of MTBC (Abdallah *et al*., 2007; Forrellad *et al*., 2013). It was also reported that ESAT6 and CFP10 facilitate translocation of *M. tuberculosis* phagosomes to the host cell cytoplasm at the later stage of infection (van der Wel *et al*., 2007). The ESX-1 was discovered as a major virulence gene of T7SS and it became a center of attention to researchers in vaccine development through genetic manipulation of the segment corresponding to the RD1 (Houben *et al*., 2014).

(ii) Type VII secretion system: *ESX-5*

A large number of genes that encode for *M. tuberculosis* proteins are characterized by Pro-Glue (PE) (a protein that is named after the conserved Proline and Glutamic acid) contained in their N-termini. This PE motifs contain polymorphic GC-rich sequences (PGRS), which are secreted by components of the ESX-5 genes, and they have been presumed to be vital in antigenic variation (Abdallah *et al*., 2009). Mycobacterium containing ESX-5 is speculated to be more pathogenic, and it has been shown that *ESX-5* is associated with slow-growing mycobacteria whose colonies cannot be visible within 7 days and therefore longer incubation period is required (Houben *et al*., 2014). Experiments that have been done by Abdallah *et al*. (2007) showed the contrast between ESX-1 and ESX-5, where it was revealed that ESX-5 *i*s not involved in the translocation of phagosomes to the cytosol. The experiment also gave insight that the ESX-5 plays an essential role as a transporter for PE proteins, cell wall integrity and as a virulence factor in the pathogenesis of *M. tuberculosis* (Abdallah *et al*., 2009).

2.5.3. Inhibition of immune response

Once the tubercle bacilli arrive at their initial residence in the lungs, alveolar macrophages are activated to phagocytose the tubercle bacilli (Pieters, 2008). The initial host immune defense is brought by macrophages that are localized at the site of infection. These macrophages are capable of growth inhibition of the engulfed *M. tuberculosis* and killing with the lysosome (Cooper, 2009).

However, pathogenic *M. tuberculosis* can survive killing by escaping lysosomal delivery. The *M. tuberculosis* has adopted different mechanisms that disrupt macrophages' bactericidal activity, develop resistance to host toxic components of the immune system, escape normal progression of the phagosome and avoid the induction of apoptosis. The Inhibition of the immune response is illustrated in Figure 2.4 below. Briefly, the oxidative and nitrosative stresses (A) occur when there is an imbalance between a relative shortage in antioxidant defense regarding an increased production of oxygen species (ROS) and reactive nitrogen species (RNS) with potential bactericidal activity (Forrellad *et al*., 2013). The ROS includes the superoxide anion ($O\cdot{-}2$), hydroxyl radical (·OH), nitric oxide (NO·), peroxynitrite ($ONOO^{-}$), hydrogen peroxide (H_2O_2), singlet oxygen ($^1O_2$) which react with a wide variety of molecules and virtually damage the cell domains (Ferrari *et al*., 2011). During phagocytosis, the host's immune macrophage ingests a pathogen to form a phagosome that fuses with the lysosome during their maturation process to form a phogolysosome. In phagosomal arresting (B), the *M. tuberculosis* manipulates the host's macrophage to inhibit the lysosomes from fusing with phagosomes and creating mature phagolysosomes. The proteins involved in arresting the macrophage phagosomal maturation are *ndk*, *ptpA* and PE_PGRS30 (Forrellad *et al*., 2013; Vazquez *et al*., 2017). Apoptosis or programmed death is one of the major mechanisms of the innate immune response elicited by a host cell to fight against a pathogen by killing its infected cells (Forrellad *et al*., 2013). The *M. tuberculosis* use several anti-apoptotic genes, specifically *nuoG*, *katG*, *sodA*/*secA2*, *pknE* and Rv3654c/Rv3655c to inhibit apoptosis (C) (Behar and Briken 2019; Forrellad *et al*., 2013). With these mechanisms, *M. tuberculosis* induces proteins that inhibit macrophages anti-mycobacterial response (Forrellad *et al*., 2013).

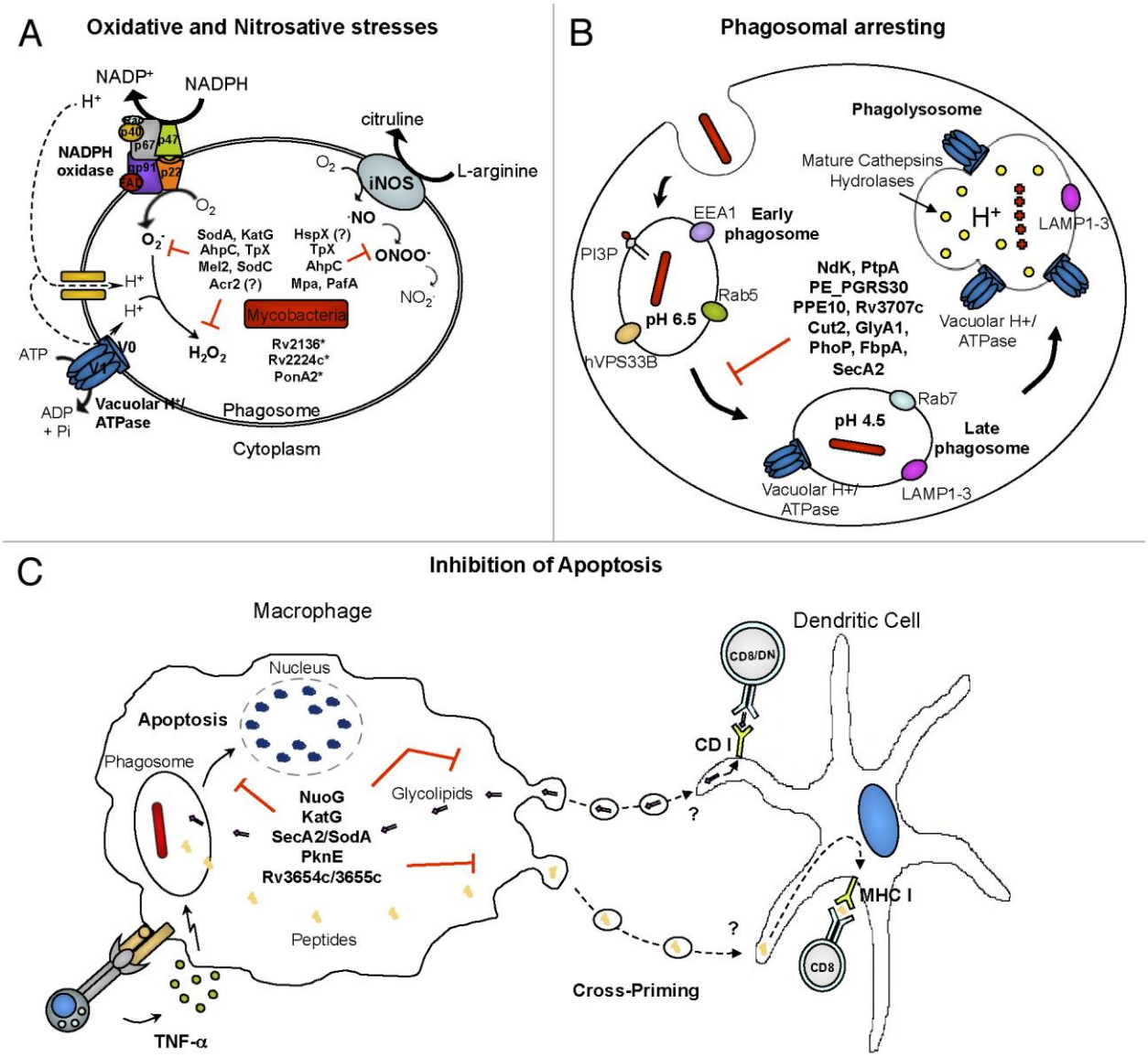

Figure 2. 4: Mechanisms of macrophage activity and proteins involved in immune response inhibition. Diagram A is Oxidative and nitrosative stresses, B is phagosome arresting and C is apoptosis and cross presentation (source: Forrellad *et al*., 2013).

2.5.3.1. Phagosome arresting

When phagosomes fuse with lysosomes, they form phagolysosomes in which the bacillus is killed and degraded into small particles that will be later presented to the adaptive immune cells. Most *M. tuberculosis* cells manipulate the host macrophages by inhibiting the formation of

phagolysosome and enabling them to grow under a host acidified intracellular compartment (Steiger *et al*., 2016). Manipulation of the host's mechanisms of killing is facilitated by virulence genes such as *ndk*, *ptpA* and *pe_pgrs30* that either suppress or interfere with the normal process of phagosome maturation (Forrellad *et al*., 2013).

(i) Phagosome arresting: *ndk*

Mycobacterium ndk is a protein that is associated with a greater survival of Mycobacterium (mostly secreted by *M. tuberculosis* and BCG strains) in the infected macrophages through the metabolism of extracellular adenosine triphosphate (ATP), which is fundamental during the inflammatory response and macrophage activation (Diaz *et al*., 2016). Mycobacterium *ndk* is described as an intracellular nucleotide pool balance mediator since it partakes in numerous enzymatic activities such as autophosphorylation, GTPase and phosphotransfer activities. A combination of ndk and ATP triggers cytotoxicity in macrophages, which drew the hypothesis that ndk might have access to the cytosolic phase of phagosomal membranes, enabling them to interact with and prevent effectors of phagosome maturation (Sun *et al*., 2010; Forrellad *et al*., 2013).

(ii) Phagosome arresting: *ptpA*

Phagosome maturation is inhibited by the low molecular weight tyrosine phosphatase, *ptpA*, which is also important in *M. tuberculosis* pathogenicity within host macrophage. The *ptpA* is involved in the regulation of membrane fusion in the endocrine pathway through dephosphorylation of host substrates' vesicle trafficking protein, Vacuolar Protein Sorting 33B (hVPS33B), leading to phagosome maturation arrest (Poirier *et al*., 2014). More interestingly, it was discovered that *ptpA* binds to subunit H of the macrophage vacuolar-H^+- adenosine triphosphatase (ATPase) machinery, which is a multi-subunit protein complex in the phagosomal membrane that initiate luminal acidification (Wong *et al*., 2011).

(iii) Phagosome arresting: *pe_pgrs*30

Mycobacterium *pe_pgrs*30 protein is a member of PE_PGRS family encoded by Rv1651c and is uniquely present in pathogenic mycobacteria (Garg *et al*., 2016). This protein mediates proinflammatory immune responses in macrophages, influences mycobacterial survival inside macrophages and modulates macrophage functionality (Chatrath *et al*., 2016). It was suggested that *pe_pgrs*30 interferes with macrophages and activates adaptive T-cell responses. This

suggestion was supported by experiments in which THP-1 macrophages were infected with recombinant *M. smegmatis* expressing deletion variants of *pe_pgrs*30 that showed survival of mycobacterium and advancement of persisters inside the macrophages (De Maio *et al*., 2018).

2.5.3.2. Inhibition of apoptosis

Apoptosis is a programmed cell death, which occurs through the innate immune response whereby infected cells are stimulated to trigger self-destruction. Two signaling pathways have been observed, namely the intrinsic and extrinsic pathways of the apoptotic signaling cascade. These signaling pathways are activated by various mitochondrial stimuli and by soluble molecules that bind to plasma-membrane receptors, allowing the host immune system to control the infection by killing the infected cells. Infected macrophages/dendritic cells also activate effector cytotoxic T cells to recognize, attack and destroy infected cells (Forrellad *et al*., 2013; Mukhopadhyay *et al*., 2014). Studies have reported different anti-apoptotic *M. tuberculosis* genes such as *nuoG, katG, sodA/secA2, pknE* and *Rv3654c/Rv3655c*, most of which were also reported to play roles in the bacterial redox homeostasis and influence full virulence of *M. tuberculosis* (Behar *et al*., 2011).

(i) Inhibition of apoptosis: *nuoG*

Mycobacterium tuberculosis nuoG is an anti-apoptotic gene that is involved in the inhibition of the extrinsic apoptotic pathway. This gene inhibits TNF-α-mediated apoptosis. That is, it encodes a subunit of the type I reduced form of nicotinamide adenine dinucleotide dehydrogenase (NADH) complex, which leads to suppression of reactive oxygen species (ROS) of the host macrophages NOX2 complex (Miller *et al*., 2010). It was also reported that lack or absence of the *nuoG* gene in *M. tuberculosis* leads to accelerated $CD4^+$ T cells priming, which suggests that inhibition of neutrophil apoptosis causes a delay in the immune response by adaptive immunity (Blomgran *et al*., 2012).

(ii) Inhibition of apoptosis: Rv3654c and Rv3655c

The Rv3654c and Rv3655c proteins that are responsible for the extrinsic pathway inhibition can accomplish Apoptosis inhibition. The role of these proteins in the anti-apoptotic mechanism was described as their ability to inhibit macrophage apoptosis of *M. tuberculosis* H37Rv transposons library (Forrellad *et al*., 2013). However, more studies must be done to confirm their roles in the

virulence of *M. tuberculosis* and determine their specific role in the manipulation of the cell apoptosis responses (Danelishvili *et al*., 2010).

(iii) Inhibition of apoptosis: *secA2*

Although inhibition of apoptosis is accomplished by either intrinsic or extrinsic apoptosis pathway arrest, some genes are capable of arresting both pathways. The *secA2* is a type of secretion system that translocate the superoxide dismutase A (*sodA*) and encodes a preprotein translocase ATPase (Braunstein *et al*., 2003). *Mycobacterium tuberculosis SecA2* has been found virulent because it facilitates the ideal secretion of *sodA* and *katG* that detoxify reactive oxygen intermediates to avoid the oxidative attack by the host cell. Recently, it has been reported that *secA2* is responsible for inhibiting phagosome maturation, thereby inhibiting both intrinsic and extrinsic apoptosis pathways induced by *M. tuberculosis*-infected macrophages during infection (Sullivan *et al*., 2012).

2.5.4. Proteins responsible for enzymes reaction

Like many other pathogenic bacteria that have their way to avoid killing by host immune cells, *M. tuberculosis* has a variety of mechanisms to protect itself from killing. These mechanisms help it to evade the host's immune responses or causes the host's tissue damage through toxins or enzymes secreted in response to defense against killing. Enzymes are proteins that play a crucial role in cell metabolisms. Their functions vary by several mechanisms such as cellular regulation, cellular homeostasis, catalytic reaction and virulence. Protein kinases and proteases have been reported as major proteins that *M. tuberculosis* use as virulence factors (Forrellad *et al*., 2013).

2.5.4.1. Protein kinases

Eukaryotic and prokaryotic cells regulate cellular metabolisms using reversible protein phosphorylation in response to external stimuli. Phosphorylation is one of the primary backbones of signal transduction pathways carried out by precise protein kinases that result in phosphorylated serine, threonine, or tyrosine residues (Av-Gay and Everett, 2000). Eleven genes, typically known as serine/threonine protein kinases viz. *pknA*-*pknL*, except *pknC*, have been identified in the *M. tuberculosis* genome and are known for modulating cellular functions such as environmental adaptation, differentiation and cell division (Koul *et al*., 2001; Prisic *et al*., 2010). Amongst these

11 genes, *pknA*, *pknB* and *pknG* have been found to be the most essential protein kinases that play significant roles in *M. tuberculosis* virulence (Forrellad *et al*., 2013).

(i) Protein kinases: *pknA* and *pknB*

Protein kinases, *pknA* and *pknB*, play vital roles in cell wall synthesis, mycolic acid biosynthesis, cell division, and they determine the morphology of the cell. These two proteins are usually described together because they are found on the same operon, playing mutual function (Forrellad *et al*., 2013). One of the essential roles of *pknA* and *pknB* is to facilitate the regulation of serine/threonine protein kinases through interdependent phosphorylation-dephosphorylation reaction. In addition, this role was demonstrated *in vitro* through protein expression in *E. coli*, which showed that *pknA* and *pknB* are targets of dephosphorylations by serine/threonine protein kinase (Sajid *et al*., 2011). The role of these two proteins has been reported in protein *pknF*, which is similarly involved in the regulation of mycolic acid biosynthesis through phosphorylation of the β-ketoacyl-acyl carrier protein synthase III. The *pknF* protein has also been shown to regulates cell growth and glucose transport (Veyron-Churlet *et al*., 2009).

(ii) Protein kinases: *pknG*

Once *M. tuberculosis* has been engulfed, a phagosome is formed and transported to the cytoplasm where it may fuse with lysosomes that degrade the bacterium into small debris (Steiger *et al*., 2016). The protein kinase G opposes the process of phagocytosis by blocking lysosomal maturation. This protein kinase is secreted directly into the cytosol to ensure that phagosome-lysosome fusion does not occur (Nguyen *et al*., 2005). The *pknG* was first observed in *M. leprae* as a soluble kinase maintained in its genome. Its presence in *M. tuberculosis* is believed to be through horizontal gene transfer and has maintained its role as a virulence factor enhancing the survival of *M. tuberculosis* within the host cell (Scherr *et al*., 2007).

2.5.4.2. Proteases

Pathogenic bacteria use extracellular proteases as virulence factors during the process of pathogenesis, leading to tissue destruction and inactivation of host immune responses by inhibiting immunoglobulins and complement pathways. Proteases are well-known enzymes important in cellular homeostasis, control of proteins responsible for transcription, cellular regulation and metabolism, virulence and activation of peptides responsible for obtaining nutrients from the host

cell by hydrolyzing host proteins and processing signaling fragments that regulate gene expression (Forrellad e*t al.*, 2013). Over 100 proteases have been reported, of which 38 were found common in *M. tuberculosis*, *M. leprae*, *M. bovis* and *M. avium paratuberculosis* species. Although there are over 100 conserved proteases, few groups such as serine proteases, ATP-depended proteases and metalloproteases have known functions (Ribeiro-Guimarães and Pessolani, 2007).

2.5.4.2.1. Serine proteases

Serine proteases are a family of proteolytic enzymes that cleave peptide bonds in other proteins, and their activities depend on the serine side-chains that act as nucleophilic amino acids at the active site (Hedstrom, 2002). Furthermore, these enzymes are prominent because they have aspartate and histidine residues, which form a catalytic triad with serine. There are two families of serine proteases that are frequently occurring with diverse functions. These families are trypsin and subtilases. Trypsins are ubiquitous in vertebrates and other phyla of higher organisms, including some prokaryotes while subtilases have been conserved in bacteria and some fungi, but they were originally identified from *Bacillus subtilis* (Hu and Leger, 2004). Bacterial subtilases are typically secreted extracellularly to cleave host peptides as a way of scavenging nutrients (Gupta *et al*., 2002).

(i) Serine protease: *mycP* genes

There are five subtilisin-like serine protease genes, also known as mycosin 1-5, which are constantly expressed in *M. tuberculosis* H37Rv and exist in high-density protein regions on the genome; they share a high degree of similarities (Brown *et al*., 2000). Catalytic residues, hydrophobic N-termini and hydrophobic regions near the C-termini, which are conserved as primary features of amino acids, characterize mycosins. Mycosin genes, *mycP1*, *mycP2* and *mycP5* have been detected in the genome of *M. leprae* while *mycP2*, *mycP3*, *mycP4* and *mycP5* were conserved in the genome of *M. avium. mycP3* was further detected in avirulent *M. smegmatis*, suggesting that diverse *mycP* genes may only occur in virulent mycobacteria (Dave *et al*., 2002). Recently, *mycP1* and *mycP3* were reported as problematic mycosin genes because they are crucial in *M. tuberculosis'* virulence, thus they regulate *ESX-1* secretion by cleaving *espB*, an *ESX-1* substrate system. The *mycP3* is also responsible for the acquisition of iron and heme (Fang *et al*., 2016).

2.5.4.3. Metalloproteases

Metalloproteases are enzymes that depend on metals, particularly zinc for their catalytic mechanisms during cell wall processes, intermediate metabolism and pathogenesis. Three zinc-dependent metalloproteases have been conserved in the *M. tuberculosis* genome. However, only two putative genes (z*mp1* and *rip1*) have been reported in *M. tuberculosis* whereas the third gene, *Rv1977* was found deleted in the genome of *M. bovis* and is therefore improbable to cause virulence (Forrellad e*t al*., 2013).

(i) Metalloprotease: zinc-dependent metalloprotease 1 (*zmp1*)

During pathogenesis, *M. tuberculosis* fights against the host for its survival. Some of the defense mechanisms include inhibition of phagolysosome formation to increase the chance of survival inside the macrophages. zinc-dependent metalloprotease 1 (*zmp1*) has an impact on phagosome maturation by hindering both caspase-1, an interleukin that proteolytically cleaves and activates precursors for processes such as inflammation, septic shock and wound healing, and pro-IL-1β activation into activated IL-1β, leading to impaired MHC class II-mediated antigen presentation (Johansen *et al*., 2011; Mori *et al*., 2014). Although *zmp1* plays a crucial role in pathogenesis, 3-(carboxymethyl) rhodanine and aminothiazole inhibitors have been developed. Furthermore, *zmp1* deletion is associated with increased protective efficacy of live attenuated BCG vaccine (Mori *et al*., 2018).

(ii) Metalloprotease: regulated intermembrane proteolysis 1 (*rip1*)

Regulated intermembrane proteolysis 1 (*rip1*) proteases are transmembrane proteins that cleave substrate proteins within the transmembrane segment, and their proteolytic activity is dependent on the transmembrane segment (Makinoshima and Glickman, 2006). Besides the fact that *rip1* proteases are metalloproteases, they are also part of three classes of regulated intermembrane proteins such as serine and aspartyl subfamilies, which are conserved at a specific active-site residue. These include rhomboid (serine), presenilin (aspartyl) and site-2 proteases (metalloprotease) (Schneider *et al*., 2013). An *in vivo* mouse model showed that the site-2 protease, *rip1,* is required for virulence by *M. tuberculosis*. A metalloprotein signature, the HExxH motif, where histidines chelate a zinc ion, glutamate is the catalytic residue, and x is any amino acid, furthermore characterizes it. The *rip1* protease is also involved in the regulation of transcription

via proteolysis of anti-SigK, anti-SigL and anti-SigM (transmembrane anti-sigma factors) (Schneider *et al*., 2014).

Table 2. 1: Summary of *M. tuberculosis* virulence genes

Groups	Genes or proteins responsible	Functions
Cell envelope proteins		
Cell wall proteins	*erp*	Not clearly known. It is crucial in cell wall biosynthesis.
	fpb	Binds fibronectin and promotes the adhesion of *M. tuberculosis* to the mucosal surface.
	mce	Synthesis of cell wall lipids. Plays a role in the final assembly of the mycobacterium cell wall.
	hbhA	They promote the attachment of *M. tuberculosis* to epithelial cells and fibroblasts.
	ompATb	It helps mycobacteria to acquire host cell molecules that help them survive during starvation.
	pstA1 and *phoT*	Involved in phosphate supply during starvation.
	caeA	It hydrolyzes ester bonds of substrates with a carbon chain length of about 3 to 7 carbon atoms.
	kefB	Interferes with phagosome maturation and elimination of the bacteria by the host cell.
	oppABC and *dppABC transporter*	They encode for permeases that are involved in the uptake of small peptides.

	ctaC	Responsible for *M. tuberculosis* growth under aerobic conditions.
Lipoprotein	*lpqH*	Hinders antigen presentation by macrophages and processing by MHC-II.
	lprG	Suppresses host immune response and plays a role in the efflux pump.
	pstS-1	Vital in phosphate metabolism that helps in survival intracellularly.
	lpqY	Transports trehalose which is needed by mycobacterium for growth.
	modA	They are a transport system for molybdenum.
Secretion system	*ESX-1*	Facilitates translocation of *M. tuberculosis* phagosome. Responsible for cell evasion.
	ESX-5	Crucial in antigenic variation. Responsible for slow-growing bacilli.
Protein inhibiting phagocytosis		
Oxidative and nitrosative stresses	*acr*1	Induce dormancy during harsh conditions.
	acr2	Induce heat shock and oxidative stress.
	136c, Rv2224c and ponA2	Display hypersensitive phenotypes in low pH or during antibiotics treatment and heat shock.
	ahpC	Involved in peroxidative homeostasis observed in INH resistance.
	sodC	Protects *M. tuberculosis* from extracellular superoxide produced by host cells.

	mel2	crucial in resistance to ROS and RNS in an activated macrophage.
	katG	Induces catalase-peroxidase activity and mycolic acid synthesis.
	tpX	Protects *M. tuberculosis* against RNS and ROS.
Phagosome arresting	*ndk*	Essential in the metabolism of extracellular ATP which is fundamental during the inflammatory response.
	ptpA	Inhibits and causes dephosphorylation of host substrates (hVPS33B) that are responsible for phagosome fusion.
	*pe_pgrs*30	Mediates proinflammatory immune response in macrophages.
Inhibition of apoptosis	*nuoG*	Encodes a subunit of type I NADH dehydrogenase complex, which leads to suppression of reactive oxygen species of the host macrophages NOX2 complex.
	Rv3654c and Rv3655c	It inhibits extrinsic and macrophage apoptosis.
	secA2	Avoids oxidative attack of the host cell. Inhibit phagosome maturation.
	pknE	Eliminates reactive nitrogen species that would be indispensable for the TNF-α-mediated induction of apoptosis.
Proteins responsible for the enzymic reaction		
Protein kinases	*pknA and pknB*	Facilitate regulation of serine/threonine protein kinases through interdependent phosphorylation-dephosphorylation reaction.

	pknF	Regulates mycolic acid biosynthesis by phosphorylation of the β-ketoacyl-acyl carrier protein synthase III. It regulates cell growth as well as glucose transport.
	pknG	Blocks lysosomal maturation.
Proteases	*mycP1*	Regulates *ESX-1* secretion by cleaving the *espB* protein.
	mycP3	Responsible for iron and heme acquisition.
	clpP1 and clpP2	Required for growth *in vitro* and in a mouse model of infection.
Other virulence proteins		
Fatty acids and lipids biosynthesis	*fadD33*	Supports tissue specific for replication during *M. tuberculosis* pathogenesis.
	Icl1 and Icl2	Involved in the Krebs cycle to help *M. tuberculosis* rely on carbon sources obtained from the host's cells to survive during phagocytosis by macrophages.
Metal transporter protein	*mgtC*	Involved in the uptake of magnesium, which supports the survival of *M. tuberculosis.*
	mbtB, irtAB and ideR	Involved in iron acquisition used in the heme of cytochromes and hemeproteins.
	ctpC	It is a zinc efflux P-type ATPase that works as an exporter.
	ctpV	It is a copper efflux transporter P-type ATPase that is essential for *M. tuberculosis* to maintain resistance to copper toxicity.
Regulator proteins	*phoP-phoR*	Encodes a histidine protein kinase (p*hoR*), and a transcriptional regulator (*phoP*) that receives a phosphate from *phoR.*

	dosR	Activates hypoxia and nitric oxide responsible for dormancy.
	sigA	Activates housekeeping genes and their expression during stress.
	mosR	Regulates the transcription of operons and regulons, by repression or activation of proteins involved in mammalian cell entry (*mce1*), hypoxia (*tgs1*) and starvation.
	pckA	Encodes a phosphoenolpyruvate carboxykinase that is responsible for reversible decarboxylation and phosphorylation of oxaloacetate to form phosphoenolpyruvate.

2.6. Treatment of tuberculosis

Treatment of TB requires a combination of multiple drugs that have different modes of action to avoid resistance in *M. tuberculosis* populations that could occur with monotherapy. The main goal for treatment is to clear or reduce bacillary load and reduce the risk of transmission of *M. tuberculosis* to other individuals. Effective drug treatment recommended for cases of susceptible TB require 6-months therapy with four first-line anti-tuberculosis drugs: rifampicin (RIF), isoniazid (INH), ethambutol (EMB) and pyrazinamide (PZA) (WHO, 2019). The 6-month TB treatment is divided into two phases, i.e. intensive and the continuous phase. The intensive phase is a 2-months phase that includes all four first-line anti-tubercular drugs. This phase is followed by a 4-months continuous phase, which includes only RIF and INH that are administered to clear the remaining bacilli that have entered the dormant phase or that are slowly replicating (Hoagland *et al*., 2016).

During the 2-months first-line TB effective therapy, the bacillary population decline. Most of the actively replicating bacilli are killed, except the slow-growing or non-replicating bacteria which are classified as persisters (figure 2.5). At the end of this phase, most patients produce sputum that has low bacteria load and may be culture negative. However, patients are expected to continue

with the 4-months phase to ensure that tuberculosis is completely treated (Horsburgh *et al*., 2015).

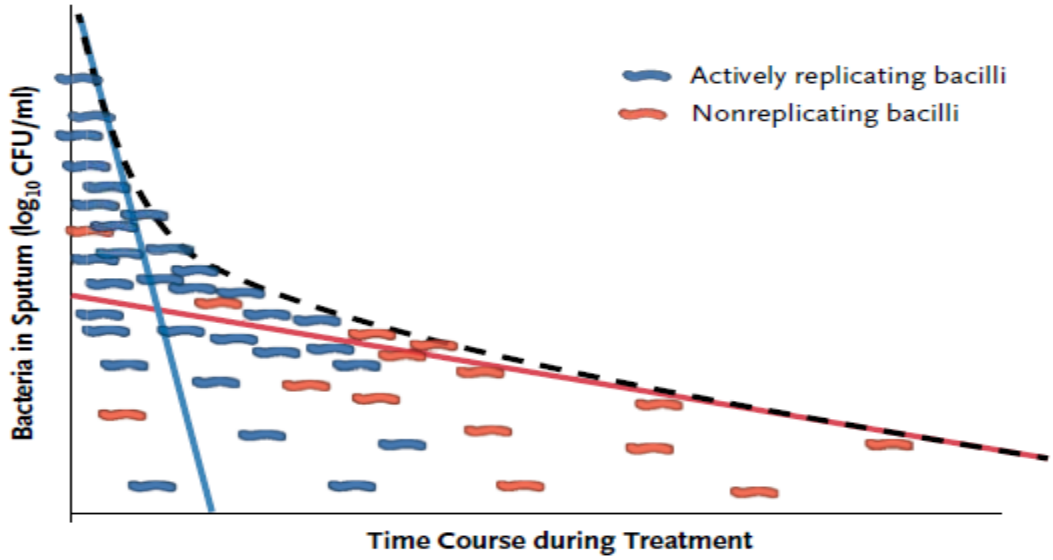

Figure 2. 5: Decline in viable bacteria during TB treatment in a sputum sample (log_{10}CFU/ml) from a TB patient during the course of effective treatment (Horsburgh *et al*., 2015). The number of tubercle bacilli declines rapidly during the early phase of first-line TB therapy (blue line), while a gradual decline of bacilli is observed during the continuous phase of anti-TB therapy (red line). The biphasic pattern (black dotted curve) denotes that the decline of the bacterial population is dependent on either actively or non-replicating bacilli.

There are instances that lead to treatment failure, which may be caused by drug toxicity, hypersensitivity, gastrointestinal disorders and drug-induced hepatitis that may lead to patients not completing their treatments. Some patients believe that they are cured because the symptoms are alleviated during the course of their treatment and leave treatment at the early phase of treatment (Menzies *et al*., 2018). Successively, some patients acquire resistance (i.e. mono, multi-drug resistance TB) throughout the treatment as a result of spontaneous mutations in *M. tuberculosis* and due to noncompliance to treatment prescriptions (Palomino and Martin, 2014). If *M. tuberculosis* acquire resistance to first-line anti-tuberculosis agents, particularly with RIF resistance (RR) or MDR-TB, patients are transferred to start the second-line anti-TB drugs. Second-line anti-TB drugs include fluoroquinolones (FQLs) (levofloxacin, moxifloxacin, gatifloxacin), and injectable drugs (amikacin (AMK), capreomycin (CM), kanamycin (KAN), streptomycin (SM)). However, over 42 countries have started using bedaquiline and delamanid drugs for treatment of MDR-TB (WHO, 2018). The WHO has recommended that delamanid may

be used together with longer MDR-TB regimens in patients aged 6–17 years and bedaquiline is only recommended to some selected adults over the age of 18 but less than 75 years who do not have liver or heart complications (WHO, 2018; Schnippel *et al*., 2018). Resistance to the second-line drugs leads to XDR-TB, a life-threatening drug-resistant *M. tuberculosis*, which is defined as MDR-TB in addition to resistance to FQLs and at least one injectable second-line drug. For cases of XDR-TB, there are no specific anti-TB drugs that are used for the treatment of XDR-TB, and therefore, some second-line anti-TB drugs that were not included in the treatment of MDR-TB may be used as an alternative (WHO, 2018).

2.7. Prevention and control

Although TB remains a global concern, the prevention of both new infections and the development of the disease is critical to reducing morbidity and mortality. The main aspect of TB prevention is to reduce the transmission of *M. tuberculosis*, prevent the progression of latent TB into active TB and reduce the risk of TB to individuals in close contact with people with TB (WHO, 2018).

Vaccination is the major strategy of prevention and control of TB worldwide. It has been reported that vaccination with Bacille Calmette Guerin (BCG) prevents about 120 000-childhood deaths in a year. Nevertheless, BCG is less effective, and it is given only to infants and young children (Voss *et al*., 2018). The low efficacy of BCG gives clear indications that there is a need for vaccines that have enhanced efficacy to reduce the risk of both infections with *M. tuberculosis* and progression of latent to active TB in adults. There are several vaccine candidates in the pipeline, but to date, there is no hope that a new TB vaccine will be available in the immediate future (WHO, 2018).

Other prevention and control measures include the use of preventive therapy, particularly in children aged 5 years or more with household TB contacts or to health care workers. These reduce the chances of contracting active TB as well as activation of latent TB. Isolation of individuals with TB, use of N95 respiratory mask, avoiding travelling to high burden TB countries, and proper ventilation may help reduce the risk of being infected with *M. tuberculosis*. (Marais, 2017; WHO, 2019).

2.8 Laboratory diagnosis of *Mycobacterium tuberculosis*

Detection of *M. tuberculosis* is performed to control the transmission, morbidity and mortality of tuberculosis. Clinically, laboratory examinations are conducted immediately for patients who

show symptoms of TB. However, not all TB patients are symptomatic, and therefore people who are suspected to have TB but not showing symptoms (including latent TB) may be diagnosed using chest radiography or either positive tuberculin skin test (TST), or interferon-gamma release assay (IGRA) should be provided to further proceed with laboratory identification (Al-Zamel, 2009). Common specimens that are used for identifications are respiratory specimens (for pulmonary TB), including sputum, bronchial aspirates and pulmonary biopsy; and non-pulmonary specimens (for extra-pulmonary TB), which includes pleural fluid, gastric aspirates, cerebrospinal fluid, urine, stool, ascitic fluid, lymph node aspirate, skin biopsy and mammary abscess (Moure *et al*., 2011).

2.8.1. Conventional method for diagnosis of *Mycobacterium tuberculosis*

Traditionally, either Auramine or Ziehl-Neelsen (ZN) stain is used to detect *M. tuberculosis* based on physiological structure or morphology by means of microscopic detection (Agrawal *et al*., 2016). Microscopy is the cheapest, simplest and one of the fastest methods for the baseline diagnosis of TB. However, its sensitivity is reduced if the bacterial load is less than 10 000 and when detecting *M. tuberculosis* in patients infected with HIV, paediatric TB, or with extrapulmonary TB (Desikan, 2013). The "gold standard" method for the diagnosis of *M. tuberculosis* is culture on Lowenstein-Jensen (LJ) solid agar medium that supports the growth of both slow- and fast-growing bacteria (Hines *et al*., 2009; Moure *et al*., 2011; Asmar, 2015). A major limitation of culture is its longer turnaround time due to the slow growth of *M. tuberculosis*, resulting in longer incubation time of about 4 weeks or more to get tangible results (Dimareli-Malli and Sarris, 2001). The BACTEC Mycobacteria Indicator Growth Tube (MIGT) 960 (Becton Dickson and Company, Sparks, USA) has been used as an alternative growth media due to its shorter turnaround time, although contamination from other microorganisms such as Nocardia species is high as MGIT allows non-specific growth (Hines *et al*., 2006). Therefore, the MPT64 antigen identification test or microscopic confirmation by ZN is required to confirm positive TB culture (Ryu, 2015).

In 2008, the WHO endorsed the Hain Line Probe Assay (LPA) (Hain Lifescience, Germany) known as Hain GenoType MTBDRplus (Hain Lifescience, Germany) for rapid diagnosis of *M. tuberculosis* and resistant to RIF and INH anti-TB drugs. The RIF and INH resistance are detected by identifying mutations in the *rpoB* and *katG*, plus *inhA* genes respectively (WHO, 2018). Studies revealed that more than 95% of RIF resistance can be obtained by targeting mutations in the 81-

base pair (also called “core region”) of the *rpoB* gene (Cavusoglu *et al*., 2002). Due to the high occurrence of resistance and shortage of diagnostic techniques for drug-resistant TB (MDR-TB and XDR-TB), Hain Lifescience has improved the Hain GenoType MTBDRplus (Hain Lifescience, Germany) by including a second-line drug resistance in their assay that led to the implementation of Hain GenoType MTBDE*sl* (Hain Lifescience, Germany). Detection of second-line drug resistance is based on the mutations in the *gyrA* gene for FQLs resistance, *rrs* gene for KAN, AMK, and CM resistance, and *embB* gene for resistance to EMB (Barnard *et al*., 2012).

The LPA is a strip-based method that uses reverse hybridization technology for the detection of both first- and second-line drug resistances associated with *M. tuberculosis*. This technology has been beneficially used for treatment follow-ups. Thus, LPA can be used for treatment assessment while waiting for conventional culture-based drug susceptibility testing (DST). The sensitivity and specificity of GenoType MTBDR*plus* (Hain Lifescience, Germany) assay is 97.7% and 66.7% for detection of RIF resistance, 69.9% and 69.2% for INH resistance, and 69.8% and 76.8% for multidrug-resistance TB (MDR-TB), respectively. The GenoType MTBDR*sl* (Hain Lifescience, Germany) assay has sensitivities and specificities of approximately 100% and 98.9% for detection of FQLs resistance, 86.2% and 99.5% for detection of second-line injectable drugs (SLID) resistance, 63.7% and 86.4% for detection of EMB resistance, and 46.2% and 100.0% for detection of XDR-TB resistance, respectively (Gardee *et al*., 2016; Liu *et al*., 2017). Hain LPA has a turnaround time of 2 days, although it can take more days depending on the volume of specimens received by the laboratory (Sypabekova *et al*., 2017). One of the main disadvantages of Hain’s LPAs is that they only identify common mutations and rely on the reverse hybridization results from the wildtype-specific probe as markers of additional mutation (Mäkinen *et al*., 2006).

In 2011, the World Health Organization (WHO) recommended another rapid diagnostic test, the GeneXpert (Cepheid Sunnyvale, CA, USA), which is an automated cartridge-based nucleic acid amplification test (NAAT) used for simultaneous identification of MTBC and RIF drug resistance (Agrawal *et al*., 2016; WHO, 2017). The WHO endorsed GeneXpert (Cepheid Sunnyvale, CA, USA) to be used as a primary diagnostic test in suspected new TB cases (WHO, 2018). Detection of *M. tuberculosis* by GeneXpert assay (Cepheid Sunnyvale, CA, USA) is based on the amplification of *M. tuberculosis*-specific sequence of *rpoB* gene using ultra-sensitive hemi-nested real-time PCR. The *rpoB* gene probe is mobilized with the molecular beacons within each cartridge

that bind to the *M. tuberculosis* deoxyribonucleic acid (DNA) within the RIF-resistance region (81 base pair) that helps to determine RIF-resistant *M. tuberculosis* (Helb *et al*., 2010).

GeneXpert (Cepheid Sunnyvale, CA, USA) is user-friendly. DNA extraction, amplification and detection occur within the cartridge in a period of 2 hours without requiring a highly trained laboratory personnel or any additional biosafety measures. The sensitivity of this assay ranges from 72.5 to 98.2% in smear-negative, culture-positive and all smear- and culture-positive patients respectively, while the specificity is approximately 99.2% (Hillemann *et al*., 2011). The main limitation of this assay is that detection of the *M. tuberculosis* relies on DNA Amplification, which may give positive results to a smear-negative culture (due to non-viable bacilli). Due to this limitation, this technique is only used for diagnosis of new TB cases and therefore, it cannot be used for treatment assessment. Cepheid GeneXpert (Cepheid Sunnyvale, CA, USA) cannot detect INH mono-resistance and XDR-TB (Held *et al*., 2014).

A new GeneXpert known as GeneXpert Ultra assay has been developed to overcome the limitations of the old GeneXpert with an improved sensitivity in detecting the *M. tuberculosis* and the RIF resistance (Osei Sekyere *et al*., 2019). The improvement includes the improved assay chemistry and cartridge design that incorporates two different multicopy amplification targets (IS6110, IS1081) and RIF resistance-determining region of the rpoB gene (Perez-Risco *et al*., 2018; Osei Sekyere *et al*., 2019). Furthermore, laboratory data showed an improved differentiation of certain silent mutations, detection of mixed infections and false-positive results in detecting RIF resistance in a paucibacillary specimen (Osei Sekyere *et al*., 2019).

2.8.2. Molecular methods for typing *Mycobacterium tuberculosis*

Genetic diversity of *M. tuberculosis* has successively increased than previously expected, and it has been suggested that genetic variance is brought by point mutations, which have a huge impact on the pathobiological phenotype of bacilli (Jagielski *et al*., 2014). Some strains that were geographically restricted are now globally distributed due to human migration, bringing some pathogens of various virulence characteristics into contact, leading to unique disease manifestations or the development of new sub-lineages (Reed *et al*., 2007). New strains or lineages are continuously identified because of advances in phylogenetic analysis, and several typing methods have been developed to classify isolates into different lineages based on various polymorphisms (Mikheecheva *et al*., 2017). Spacer oligonucleotide typing (spoligotyping),

Mycobacterial Interspersed Repetitive Units-Variable Number Tandem Repeats (MIRU-VNTR), Insertion Sequence 6110 restriction fragment length polymorphism (IS6110-RFLP) and SNPs are applied in molecular epidemiology studies to determine *M. tuberculosis* genotypes and lineages (Zaychikova *et al*., 2015).

2.8.2.1. Insertion sequence 6110-Restriction Fragment Length Polymorphism (IS6110-RFLP)

The IS6110-RFLP typing method is based on the digestion of genomic DNA at a specific site using restriction enzymes. Fragments produced after digestion are separated by gel electrophoresis and detected by Southern blotting to determine the sizes of localized restriction fragments (Zaychikova *et al*., 2015; Mikheecheva *et al*., 2017). The IS6110 is recognized as a genetic marker for molecular typing methods and is found in abundance at multiple sites in the *M. tuberculosis* genome with their copies varying between 1 to 25 in different strains. This element is known as mobile element, or "jumping gene", which is easily copied in different sites and has been used as a popular tool for strain typing by RFLP (Heersma *et al*., 1998; Roychowdhury *et al*., 2015). The IS6110-RFLP method is currently used as a gold standard for genotypic diagnosis or typing because of its high discriminatory power (Said *et al*., 2016). The drawback of this method is that it is time-consuming, labor-intensive and the resulting DNA fingerprinting image makes inter-laboratory interpretation and comparisons difficult (Weniger *et al*., 2010; Said *et al*., 2016).

2.8.2.2. MIRU-VNTR

A PCR-based genotyping method that targets VNTRs has been developed to overcome the RFLP drawbacks. The MIRU-VNTR typing is a specific *M. tuberculosis* typing method that utilizes conserved tandem repeats in the mycobacterium genomic loci as a genetic marker. The MIRU-VNTR typing is based on the amplification of multiple loci (either 12, 15, or 24) using primers specific to each VNTR region, and the determination of the PCR product size, which reflects the number of the targeted MIRU-VNTR copies (Said *et al*., 2012). Tandem repeats differ in various strains and that gives the opportunity for studying transmission and epidemiology of TB infection (Zaychikova *et al*., 2015; Suzana *et al*., 2017). This method is very important in constructing phylogenetic trees, which shows the genetic relatedness of *M. tuberculosis* strains that can be found in a certain region; it also helps distinguish different lineages (Weniger *et al*., 2010). Several studies have reported that higher lineage discriminatory power is obtained upon combining MIRU-

VNTR and spoligotyping methods than using either spoligotyping or MIRU-VNTR methods alone (Sola *et al*., 2003; Said *et al*., 2012).

2.8.2.3. Spoligotyping

Spoligotyping is based on the *in vitro* PCR amplification of highly polymorphic direct repeat loci in the *M. tuberculosis* genome that contain variable numbers of 36 base pair nucleotides separated by spacers of 34-41bp (Zaychikova *et al*., 2015). The *in vitro* DNA amplification targets direct repeats and amplified DNA is hybridized by binding them to various synthetic spacer oligonucleotides bound covalently on a membrane, which results in different patterns that are seen on the x-ray film as black bars (Kamerbeek *et al*., 1997). Spoligotyping is advantageous in that it can be used to identify both pulmonary and extrapulmonary TB lineages; it is simple, highly reproducible and rapid. Results are obtainable within a day as simple digital black squares patterns visualized on an X-ray film (Suzana *et al*., 2017). The main disadvantage of spoligotyping is that all genetic polymorphisms are restricted to a single direct repeat cluster that provides an insufficient discriminatory power and is thus often used in combination with another genotypic method that further discriminates *M. tuberculosis* lineages and sub-lineages (Ramos *et al*., 2014).

2.8.2.4. Whole-genome-based SNP typing

The above-stated genotyping techniques are widely used for detecting and typing *M. tuberculosis*. However, these methods have a lower resolution. The whole-genome-based SNP typing has been reported as a method that is more promising due to its higher resolution or discriminatory power and low level of homoplasy (Zaychikova *et al*., 2015). Single nucleotide polymorphisms are phylogenetically informative in population analysis because of the paucity of sequence heterogeneity caused by either synonymous or nonsynonymous substitutions within *M. tuberculosis* genomes that ensure that the probability of encountering repeated synonymous SNPs within the individual lineage is extremely low (Reed *et al*., 2009). It has been reported that SNPs in virulence genes might be crucial in the transmission of *M. tuberculosis* lineages and drug resistance. Recently, studies showed that comprehensive analysis of SNPs by whole-genome sequencing (WGS) can be used to detect unique mutations in the genome of particular strains/lineages, which may be associated with specific *M. tuberculosis* phenotypic characteristics (Coscolla and Gagneux, 2014).

2.9. Conclusion

Mycobacterium tuberculosis lineages are globally distributed and are associated with the massive spread of MDR- and XDR-TB (Mikheecheva *et al*., 2017). Today the *M. tuberculosis* with modified virulence and transmissibility, particularly those that are caused by mutations leading to genetic variation and increased pathogenicity are highly reported (Zaychikova *et al*., 2015).

Various molecular methods for genotyping *M. tuberculosis* using different markers are available, however, these techniques have number of disadvantages and some have low discriminatory power (Brudey *et al*., 2006; Mikheecheva *et al*., 2017). The single nucleotide polymorphisms (SNPs) are regarded as the most promising genetic markers for genotyping *M. tuberculosis* because they have low-level homoplasy and high discriminatory power (Zaychikova *et al*., 2015). Recently, studies showed that comprehensive analysis of SNPs WGS can be used to detect unique mutations in the genome of particular strains/lineages, which may be associated with specific *M. tuberculosis* phenotypic characteristics Coscolla and Gagneux, 2014). The present study proposed that genotyping *M. tuberculosis* using polymorphisms in virulence genes may be an alternative approach to determine lineages and may help to detect the *M. tuberculosis* strains that are epidemiologically dangerous and have adapted to specific geographic regions.

Chapter 3

Materials and Methods

3.1. Study design and sample collections

This was a descriptive study to identify and evaluate virulence genes' signature SNPs as markers of *M. tuberculosis* lineages and sub-lineages. A total of 150 susceptible (n= 86 isolates) and resistant (n= 64 isolates) *M. tuberculosis* cultures stored in Mycobacteria growth indicator tubes (MGIT) tubes (Becton Dickson and Company, Sparks, USA) were collected from May to October 2018 to conduct this study. These cultures were obtained from sputum specimens collected at various clinics and hospitals in the Tshwane district that were submitted to the National Health Laboratory Service/Tshwane Academic division (NHLS/TAD) for diagnostic purposes.

3.2. Study site

This study was conducted at the University of Pretoria; all samples were collected and sub-cultured at the NHLS/TAD situated at the University of Pretoria's Department of Microbiology. Samples were sent to Stellenbosch University where cultivation of *M. tuberculosis* in Middlebrook 7H9 broth base (Sigma-Aldrich, St Louis, Missouri, USA) and DNA extraction by hexadecyltrimethylammonium bromide (CTAB) (Sigma-Aldrich, St Louis, Missouri, USA) were done in a biosafety level 3 (BSL3) laboratory. The extracted DNA, which was used to perform spoligotyping at the University of Pretoria, and DNA aliquots were sent to the sequencing core facility, National Institute for Communicable Disease (NICD), Johannesburg, where whole-genome sequencing (WGS) was performed and at the Center for Tuberculosis, NICD, Johannesburg, where IS6110-RFLP was done.

3.3. Cultivation of *Mycobacterium tuberculosis*

3.3.1. Cultivation of *Mycobacterium tuberculosis* in the Mycobacteria growth indicator tubes

Stored *Mycobacterium tuberculosis* cultures were sub-cultured in Becton Dickinson (BD) MGIT tubes (Becton Dickson and Company, Sparks, USA) until the tubes flagged positive for the growth of *M. tuberculosis*. Briefly, polymyxin B, amphotericin B, nalidixic acid, trimethoprim, and Azlocillin (PANTA) (Becton Dickson and Company, Sparks, USA) antibiotic mixtures were reconstituted with 15 mL of the oleic acid-albumin dextrose-catalase (OADC) (Becton Dickson

and Company, Sparks, USA) growth enrichment medium (see appendix A, for detailed preparation of reagents). From each clean MGIT tube (Becton Dickson and Company, Sparks, USA) containing 7 mL of modified Middlebrook 7H9 broth base (Becton Dickson and Company, Sparks, USA), 0.8 mL of growth supplement/PANTA (Becton Dickson and Company, Sparks, USA) mixture was added aseptically. About 0.5 mL of the *M. tuberculosis* cultures were inoculated into the MGIT tubes (Becton Dickson and Company, Sparks, USA) containing the growth supplements mentioned above. The suspension and the inoculum were mixed vigorously and loaded into an automated BACTEC MGIT 960 instrument (Becton Dickson and Company, Sparks, USA). All positive *M. tuberculosis* cultures were removed from the BACTEC MGIT instrument (Becton Dickson and Company, Sparks, USA) immediately after they flagged positive and incubated at 37ºC for 10 days to further support growth and increase the yield of bacilli.

3.3.2. Cultivation of *Mycobacterium tuberculosis* in the blood culture agar plates

All positive MGIT cultures were assessed for purity using blood agar (DMP, Johannesburg, South Africa) to check if no contaminants were growing in the MGIT tubes (Becton Dickson and Company, Sparks, USA). Few drops of MGIT cultures were transferred onto blood agar plates (DMP, Johannesburg, South Africa) and incubated at 37ºC for 48 hours. The blood agar plates (DMP, Johannesburg, South Africa) were read after 48 hours and the plates that showed growth were recorded as contaminated MGIT cultures and they were discontinued from this study.

3.3.3. Growth enrichment with Middlebrook 7H9 broth base medium

All culture-positive non-contaminated *M. tuberculosis,* which were confirmed by ZN stain, were sent to the BSL3 laboratory at the Stellenbosch University to sub-cultured isolates in Middlebrook 7H9 broth base (Sigma-Aldrich, St Louis, Missouri, USA) medium to increase the yield of bacilli. Briefly, 2.35 g of Middlebrook 7H9 broth base (Sigma-Aldrich, St Louis, Missouri, USA) powder was weighed and dissolved with distilled water in an autoclavable bottle (see appendix A, for detailed preparation of reagents). Once the Middlebrook broth base (Sigma-Aldrich, St Louis, Missouri, USA) had dissolved, distilled water was added up to 450 mL. Approximately 1 mL glycerol was added and then autoclaved at 121ºC for 15 minutes. The media was removed from the autoclave and placed on a water bath (Labnet, New Jersey, USA) to cool to at least 45ºC. Aseptically, 50 mL of OADC (Becton Dickson and Company, Sparks, USA) was added to the media. Aliquots of 5 mL Middlebrook 7H9 broth (Sigma-Aldrich, St Louis, Missouri, USA) were

dispensed to the cell culture Flasks (Inqaba Biotechnical Industries, South Africa) followed by addition of 1 mL MGIT cultures and incubated at 37°C for 10 up to 15 days.

3.4. Identification by Ziehl-Neelsen stain

Positive *M. tuberculosis* cultures were further confirmed by microscopic examination with a Ziehl-Neelsen (ZN) stain. A thin smear was prepared by mixing a drop of smear fixative (DMP, Johannesburg, South Africa) with a drop of growth sediments of MGIT (Becton Dickson and Company, Sparks, USA) cultures on the microscope slide (Merck & Co, Inc, Kenilworth, Germany) and allowed to air dry. The smear was heat fixed and placed on the staining rack. Carbolfuchsin (Diagnostic Media Product (DMP), Johannesburg, South Africa) was dispensed on the smear and heated gently from the underside of the slide by passing a flame under the rack until fumes appeared. After 5 minutes, carbolfuchsin (DMP, Johannesburg, South Africa) stain was washed off with clean water and the smear was then decolorized with 3% acid alcohol (3% hydrochloric acid and 70% ethanol) (DMP, Johannesburg, South Africa) for 30 seconds. The smear was well rinsed with water and flooded with the methylene blue stain (DMP, Johannesburg, South Africa) for 30 seconds. The stain was washed off with clean water and the back of the slide was blotted dry to remove access water. The slides (Merck & Co, Inc, Kenilworth, Germany) were placed on the draining rack for the smear to air dry. The smears were examined microscopically using the 100X oil immersion objective (Labnet, New Jersey, USA). The smear was declared ZN positive for *M. tuberculosis* if there were pink rod-shaped bacteria against the blue background.

3.5. DNA extraction

Five milliliters of liquid cultures were transferred into 15 mL falcon tubes (Inqaba Biotechnical Industries, South Africa) and heat-killed at 80ºC for 1 hour before being removed from the BSL3. The heat-killed cultures were centrifuged (Labnet, Edison, New Jersey, USA) at 4000 x g for 30 minutes at room temperature. The supernatant was discarded, and the pellets were resuspended with 400 µL Tris/Ethylenediaminetetraacetic acid (EDTA) (TE) (Sigma-Aldrich, St Louis, Missouri, USA) buffer. Fifty microliters of lysozyme (10 mg/mL) (Sigma-Aldrich, St Louis, Missouri, USA) was added and incubated at 37°C overnight (approximately 18 to 24 hours). Roughly 70 µL of 10% sodium dodecyl sulfate (SDS) (Sigma-Aldrich, St Louis, Missouri, USA) and 5 µL proteinase K (10 mg/mL) (Sigma-Aldrich, St Louis, Missouri, USA) were added, mixed and incubated at 65°C for 10 minutes. A hundred microliters of sodium chloride (NaCl) (Sigma-

Aldrich, St Louis, Missouri, USA) and hexadecyltrimethylammonium bromide (CTAB) (Sigma-Aldrich, St Louis, Missouri, USA Sigma-Aldrich, St Louis, Missouri, USA) solutions were added and vortexed (Labnet, Edison, New Jersey, USA) until they became milky. Approximately 750 µL chloroform/isoamyl alcohol (Sigma-Aldrich, St Louis, Missouri, USA) was added to each falcon tube (Inqaba Biotechnical Industry, South Africa) and vortexed (Labnet, Edison, New Jersey, USA) for 10 seconds. The mixtures were transferred to 2 mL Eppendorf tubes (Inqaba Biotechnical Industries, South Africa) and centrifuged (Labnet, Edison, New Jersey, USA) at 12000 x g for 5 minutes, resulting in the formation of 3 phases: top (aqueous) phase, middle white-like solid phase and bottom liquid phase. The top aqueous solution was aspirated into new clean 1.5 mL Eppendorf tubes (Inqaba Biotechnical Industry, South Africa) and 0.6 x volume ice-cold isopropanol (Sigma-Aldrich, St Louis, Missouri, USA) was added and incubated at -20°C for 1 hour. The solutions were centrifuged (Labnet, Edison, New Jersey, USA) at room temperature at 12000 x g for 30 minutes. Carefully, the supernatant was aspirated without touching the pellet. The pellets were resuspended with 1 mL of 70% ethanol (Sigma-Aldrich, St Louis, Missouri, USA) and centrifuged (Labnet, Edison, New Jersey, USA) at 12000 x g for 15 minutes. Ethanol (Sigma-Aldrich, St Louis, Missouri, USA) was aspirated and excess ethanol (Sigma-Aldrich, St Louis, Missouri, USA) was allowed to evaporate at room temperature until pellets were dry. About 50 µL of TE (Sigma-Aldrich, St Louis, Missouri, USA) buffer was added to the dried pellets and they resuspend overnight. The tubes were vortexed (Labnet, Edison, New Jersey, USA) to mix well the DNA. The DNA was quantified by nanodrop 200 (Thermofischer, Waltham, USA) and stored at -80°C.

3.6. Screening and discrimination of *Mycobacterium tuberculosis* lineages

Screening for circulating lineages was done using PCR-based spoligotyping methods according to the protocol described by Kamerbeek *et al*. (1997). Lineages were further determined by IS6110-RFLP to confirm and identify sub-lineages within Beijing and LAM lineages.

3.6.1. Spoligotyping

Spoligotyping was done using DNA extracted from all 150 *M. tuberculosis* cultures to screen and detect the presence or absence of 43 spacer sequences located between the direct repeat sequences using protocols described by Kamerbeek *et al*. (1997). The method is based on the amplification of the direct repeat sequences in *M. tuberculosis* genomes, followed by hybridization on a

membrane containing covalently linked oligonucleotides that correspond to various spacer sequences.

3.6.1.1. in Vitro amplification of spacer DNA PCR

The direct repeat regions of the *M. tuberculosis* genome were amplified using direct repeat forward (DRa) 5' biotinylated primers (Ocimum Biosolutions, India) and direct repeat reverse (DRb) primers (Ocimum Biosolutions, India).

The PCR reaction was performed in a total volume of 50 µL reaction mixture consisting of 25 µL Kapa Taq ready mix (KapaBiosystems, South Africa), 2 µL of primers DRa and DRb (Ocimum Biosolutions, India) each, 17 µL of nuclease-free water (Thermofischer, Waltham, USA) and 4 µL of the template DNA. *Mycobacterium tuberculosis* H37Rv and BCG DNA were used as a positive control whilst de-ionized water was used as a negative control. The PCR amplification conditions consisted of initial denaturation step at 96^{o}C for 3 minutes, followed by 30 cycles of denaturation at 95^{o}C for 1 minute, annealing (Tm-5°C) at 55^{o}C for 1 minute and extension at 72^{o}C for 30 and a final extension at 72^{o}C for 2 minutes. The PCR products were removed from the PCR thermal cycler (Labnet, Edison, New Jersey, USA) and stored at 4°C.

3.6.1.2. Hybridization of PRC products and detection of spoligotype patterns

Hybridization of the biotin-labelled PCR products to the spacer-oligonucleotides immobilized on the nylon membrane (Ocimum Biosolutions, India) was done. All buffers required were prepared and pre-warmed before used (see appendix A, for detailed preparation of reagents). Briefly, 250 mL 2X sodium chloride-sodium phosphate-EDTA (SSPE)/0.1% SDS (Sigma-Aldrich, St Louis, Missouri, USA), 250 mL 2X SSPE/0.5% SDS (Sigma-Aldrich, St Louis, Missouri, USA) were prepared from the stock concentrated buffers by diluting with the de-mineralized water and pre-warmed at 60°C. Approximately, 250 mL of 2X SSPE/0.5% SDS (Sigma-Aldrich, St Louis, Missouri, USA) was prepared and pre-warmed at 42°C and 250 mL of 2X SSPE (Sigma-Aldrich, St Louis, Missouri, USA) was prepared and placed at room temperature. The 20 µL of PCR product was added to PCR tubes (Inqaba Biotechnical Industries, South Africa) containing 150 µL of 2X SSPE/0.1% SDS (Sigma-Aldrich, St Louis, Missouri, USA) and heat-denatured at 99°C for 5 minutes. The membrane (Ocimum Biosolutions, India) was washed for 5 minutes at 60°C in 250 µL 2X SSPE/0.1% SDS (Sigma-Aldrich, St Louis, Missouri, USA) and placed on top of support

cushion (Sigma-Aldrich, St Louis, Missouri, USA) between the Miniblotter 45 (Ocimum Biosolutions, India) in a way that the slots were perpendicular to the line pattern of the applied oligonucleotide. The residual fluids from the Miniblotter 45 (Ocimum Biosolutions, India) slots were removed by aspiration. The slots were filled with the diluted denatured PCR products and hybridized for 60 minutes at 60°C. Samples were removed from the Miniblotter 45 (Ocimum Biosolutions, India) by aspiration. The membrane (Ocimum Biosolutions, India) was removed from the miniblotter 45 (Ocimum Biosolutions, India) and washed twice in 250 mL 2X SSPE/0.5% SDS (Sigma-Aldrich, St Louis, Missouri, USA) for 10 minutes at 60°C. The membrane was placed in a rolling bottle and allowed to cool down and streptavidin-peroxidase conjugate (Sigma-Aldrich, St Louis, Missouri, USA) was added and incubated at 42°C for 60 minutes. The membrane (Ocimum Biosolutions, India) was washed twice with 250mL of 2X SSPE/0.5% SDS (Sigma-Aldrich, St Louis, Missouri, USA) for 10 minutes at 40°C. The membrane (Ocimum Biosolutions, India) was then rinsed in 250mL 2X SSPE (Sigma-Aldrich, St Louis, Missouri, USA) for 5 minutes at room temperature and incubated for 1 minute in 20 µL ECL (GE Healthcare, Amersham, United Kingdom) detection reagents. The membrane (Ocimum Biosolutions, India) was covered with a transparent plastic sheet and exposed to a light-sensitive X-ray film (GE Healthcare, Amersham, United Kingdom) for 20 minutes inside the X-ray cassette. The X-ray film (GE Healthcare, Amersham, United Kingdom) was exposed to the developer (Guangzhou MeCan Medical, China) for 2 minutes to develop dark spots on the film (GE Healthcare, Amersham, United Kingdom). It was then immediately fixed with the fixer for 1 minute (Guangzhou MeCan Medical, China). The resultant spoligotypes were entered in an Excel sheet as a binary code presenting either positive or negative (1 or 0 respectively) hybridization results. The spreadsheet was then deposited to the MIRU-VNTRplus database (http://www.miru-vntrplus.org) to get the lineage name assigned to each isolate.

3.6.2. Confirmation of lineages and detection of Beijing and LAM sub-lineages

The IS6110-RLFP genotypic method was used to confirm and detect the sub-lineages within Beijing and LAM lineages detected by spoligotyping. This method included DNA digestion by the *PvuII* restriction endonuclease (Inqaba Biotechnical Industries, South Africa) and Southern transfer of the digested DNA fragments within the fingerprinting gel.

3.6.2.1. DNA digestion by *PvuII* restriction endonuclease

To perform the restriction digest of the DNA, the DNA concentrations obtained were re-constituted to 6 µg to get the volume (in microliters) of DNA that was used and to calculate the amount of water that was added to make 100 µL, the final volume of the entire restriction enzyme mixture (see appendix A, for detailed preparation of reagents). About 10 µL of the restriction buffer (Inqaba Biotechnical Industries, South Africa) was added into a clean 1.5 mL tubes (Inqaba Biotechnical Industries, South Africa), followed by the amount (the amount of water added was depended on the volume of the DNA added to make up 100 µL of restriction endonuclease reaction mixture, hence the volumes of 10 µL restriction buffer and 3 µL *PvuII* restriction endonuclease were constant to every reaction mixture) of nuclease-free water (Thermofischer, Waltham, USA). Approximately 6 µg of DNA was added and 3 µL of the *PvuII* restriction endonuclease (Inqaba Biotechnical Industries, South Africa) was then added as the last solution to ensure that it does not start excising activity before DNA is added. The mixture was mixed well and incubated at 37°C for 16 hours. A test gel for testing DNA digestion by *PvuII* endonuclease (Inqaba Biotechnical Industries, South Africa) was made by adding 3 g agarose (Sigma-Aldrich, St Louis, Missouri, USA) and 300 mL 1X tris boric EDTA (TBE) buffer (Sigma-Aldrich, St Louis, Missouri, USA). About 8 µL of digested DNA was removed from the digested samples and mixed with 4 µL of 6X loading dye (Inqaba Biotechnical Industries, South Africa). About 12 µL of DNA with loading dye (Inqaba Biotechnical Industries, South Africa) was loaded into the wells of the gel, which has been placed in a buffer-filled box. The gel was run for 4 hours at 100 volts. The gel was placed in a flat container with 500 mL of 1X TBE buffer (Sigma-Aldrich, St Louis, Missouri, USA) mixed with 50 µL of ethidium bromide (Inqaba Biotechnical Industries, South Africa) and placed on a shaker (Labnet, Edison, New Jersey, USA) for 30 minutes at room temperature, which allowed the ethidium bromide to (Inqaba Biotechnical Industries, South Africa) penetrate the gel. The gel was viewed using Syngene UV transilluminator (Ingenius, United Kingdom).

3.6.2.2. Southern transfer of the fingerprinting gel

The remaining *PvuII*-digested DNA (92 µL) was precipitated by adding 9 µL of 3 M sodium acetate (Sigma-Aldrich, St Louis, Missouri, USA) and 300 µL ice-cold 100% ethanol (Sigma-Aldrich, St Louis, Missouri, USA) and incubated for 16 hours at 20°C. The DNA fragments were obtained in a pellet form by centrifuging pellets at 10000 x g for 30 minutes at room temperature.

The supernatant was aspirated, and the pellets were washed with 500 µL ice-cold 70% ethanol (Sigma-Aldrich, St Louis, Missouri, USA) and centrifuged (Labnet, Edison, New Jersey, USA) at 10000 x g for 30 minutes at 4°C. The supernatant was aspirated and DNA in pelleted form was dried at room temperature overnight. The DNA was re-dissolved in a 1X loading buffer (Inqaba Biotechnical Industries, South Africa) with an internal molecular weight marker and incubated at 4°C for 16 hours. Eight percent agarose gel was prepared by dissolving 2.4 g agarose (Sigma-Aldrich, St Louis, Missouri, USA) with 300 mL 1X TBE (Sigma-Aldrich, St Louis, Missouri, USA). About 10 µL *PvuII*-digested DNA, with external marker MTB14323 (Inqaba Biotechnical Industries, South Africa), was loaded into each well and the gel was run overnight at 60 volts. After the gel run, the gel was stained with ethidium bromide (Inqaba Biotechnical Industries, South Africa) to visualize the bands by adding 50 µL ethidium bromide (Inqaba Biotechnical Industries, South Africa) into a flat container with 500 mL 1X TBE buffer (Sigma-Aldrich, St Louis, Missouri, USA). The gel was placed on the shaker (Labnet, Edison, New Jersey, USA) for 30 minutes at room temperature to allow the ethidium bromide to penetrate the gel. The gel was visualized by placing the gel on the Syngene UV transilluminator (Ingenius, United Kingdom). The gel was inverted and placed into a flat container. The DNA in the gel was denatured by adding 500 mL of denaturation solution (Sigma-Aldrich, St Louis, Missouri, USA) for 30 minutes. The denaturation solution (Sigma-Aldrich, St Louis, Missouri, USA) was discarded and the gel was neutralized by incubating the gel in 500 mL of neutralizing solution (Sigma-Aldrich, St Louis, Missouri, USA) for 30 minutes. The nylon membrane hybond N^+ (GE Healthcare, Amersham, United Kingdom) was labelled with a black ballpoint and spotted with 0.2 µL aliquots of the orientation marker (Roche, USA) and internal markers (Roche, USA). The membrane (GE Healthcare, Amersham, United Kingdom) was hydrated by soaking with distilled water and equilibrated in 20X SSPE solution (Sigma-Aldrich, St Louis, Missouri, USA). A Whatman 3 MM paper (GE Healthcare, Amersham, United Kingdom) was soaked in a 20X SSPE (Sigma-Aldrich, St Louis, Missouri, USA) and placed in a flat blotting tray. The inverted agarose gel was placed on top of the Whatman 3 MM paper (GE Healthcare, Amersham, United Kingdom). The air bubbles were removed by gently rolling a 10 mL pipette (Inqaba Biotechnical Industries, South Africa) over the gel. The strips of parafilm (Sigma-Aldrich, St Louis, Missouri, USA) were placed around the gel to ensure the fluid flowed through the gel during Southern transfer (Sigma-Aldrich, St Louis, Missouri, USA). The nylon membrane (GE Healthcare, Amersham, United Kingdom) was placed on the

agarose gel with the orientation makers (Roche, USA) facing the agarose gel and the air bubbles were removed by rolling with 10 mL pipette (Inqaba Biotechnical Industries, South Africa). Two sheets of Whatman 3 MM paper (GE Healthcare, Amersham, United Kingdom) were soaked in a 20X SSPE (Sigma-Aldrich, St Louis, Missouri, USA) and placed onto the nylon membrane (GE Healthcare, Amersham, United Kingdom). Air bubbles were removed, and the stack of folded paper towels were placed onto the Whatman 3 MM paper (GE Healthcare, Amersham, United Kingdom). The glass plate was put on top of the stack paper towel and a bottle filled with 1 L of water was placed on top to ensure that the fluid flowed through the agarose gel. The blotting tray was filled with 20X SSPE (Sigma-Aldrich, St Louis, Missouri, USA) and the Southern transfer was allowed to proceed for 16 hours. The nylon membrane (GE Healthcare, Amersham, United Kingdom) was removed after Southern transfer and washed with 2X SSPE (Sigma-Aldrich, St Louis, Missouri, USA) for 10 minutes. The nylon membrane (GE Healthcare, Amersham, United Kingdom) was then placed between two sheets of the Whatman 3 MM paper (GE Healthcare, Amersham, United Kingdom) and baked at 80°C for 2 hours to covalently bond the DNA to the membrane (GE Healthcare, Amersham, United Kingdom). The membrane (GE Healthcare, Amersham, United Kingdom) was sealed with the plastic sleeve containing 48 mL of ECL gold buffer (GE Healthcare, Amersham, United Kingdom). The membrane (GE Healthcare, Amersham, United Kingdom) was rolled with 10 mL pipette (Inqaba Biotechnical Industries, South Africa) to remove air bubbles and to distribute the buffer to the membrane (GE Healthcare, Amersham, United Kingdom) and incubated at 42°C for 60 minutes. Approximately 0.9 µL of the IS6110 probe DNA (200 ng amplified IS6110) was added to a 14.1 µL deionized water. The probe DNA was denatured by incubating for 5 minutes at 100°C followed by cooling on ice for 5 minutes to keep the DNA single-stranded. The probe was labeled with a 15 µL labelling mix (GE Healthcare, Amersham, United Kingdom) and 15 µL glutaraldehyde solution and incubated for 10 minutes at 37°C. After 60 minutes of membrane (GE Healthcare, Amersham, United Kingdom) incubation, the labelled probe was added to the membrane (GE Healthcare, Amersham, United Kingdom) and sealed. The African wonder bag was prepared by adding approximately 1 L water in a plastic bag and sealed. The membrane (GE Healthcare, Amersham, United Kingdom) was placed in a flat container and the African wonder bag was then placed on top of the membrane (GE Healthcare, Amersham, United Kingdom) and incubated in a shaking water bath (VELP Scientifica, Monza and Brianza, Italy) for 16 hours at 42°C. After hybridization had occurred, the membrane (GE

Healthcare, Amersham, United Kingdom) was removed from the plastic bag and washed with 400 mL primary wash (Sigma-Aldrich, St Louis, Missouri, USA) twice for 20 minutes at 42°C. The membrane (GE Healthcare, Amersham, United Kingdom) was then washed twice in 400 mL 2X standard sodium citrate (SSC) (Sigma-Aldrich, St Louis, Missouri, USA) for 5 minutes at room temperature on a shaker (Labnet, Edison, New Jersey, USA). The membrane (GE Healthcare, Amersham, United Kingdom) was exposed to X-ray film (GE Healthcare, Amersham, United Kingdom) for 3 hours and washed with a developer (Guangzhou MeCan Medical, China) and a fixer (Guangzhou MeCan Medical, China).

3.7. Whole-genome sequencing by Illumina Miseq platform

Isolates detected as Beijing and LAM genotypes by spoligotyping (n= 56 isolates) were further genotyped by whole-genome sequencing (WGS) which was carried out using the Illumina Miseq platform (Illumina, San Diego, USA) to detect lineage-specific mutations in the virulence genes of *M. tuberculosis*. This procedure included genomic DNA (gDNA) tagmentation, library PCR amplification, library quantification and normalization, and library pooling and sequencing

3.7.1. Genomic DNA tagmentation

Genomic DNA was tagmented with Nextera XT transposome (Illumina, San Diego, USA). Ten microliters of tagment DNA (TD) buffer, 5 µL of normalized gDNA and 5 µL of amplification tagment mix (ATM) were transferred to a 96-well PCR plate (Bio-rad, Hercules, California, USA) and centrifuged (Labnet, New Jersey, USA) at 280 x g for 1 minute. The PCR plate was placed on a thermal cycler and run at 55^{o}C for 5 minutes. Five microliters of neutralizing tagmentation buffer (Illumina, San Diego, USA) was added to each well, centrifuged (Labnet, Edison, New Jersey, USA) at 280 x g for 1 minute at 20^{o}C and then incubated at room temperature for 5 minutes.

3.7.2. Library PCR amplification

Index 1 and index 2 adapters were arranged in a TruSeq Index Plate Fixture (Illumina, San Diego, USA) wherein, index 1 adapters were arranged in columns 1-6 and index 2 adapters were arranged in rows A-H. The PCR reaction was made by adding 5 µL Index 1 and 2 primers (Illumina, San Diego, USA) and 15 µL of Nextera PCR master mix (NPM) (Illumina, San Diego, USA) into the wells containing tagmented DNA. The plate was sealed with Microseal 'A' (Bio-rad, Hercules, California, USA) centrifuged (Labnet, Edison, New Jersey, USA) at 280 x g for 1 minute at 20^{o}C.

Polymerase chain reaction amplification was performed under the following parameters: 72°C for 3 minutes, 95°C for 30 seconds followed by 12 cycles (95°C for 10 seconds, 55°C for 30 seconds and 72°C for 30 seconds) and 72° C for 5 minutes.

AMPure XP beads (Beckman Coulter Inc, California, US) were used to purify DNA libraries. Briefly, 50µL amplicons were transferred to corresponding wells on the midi plate (Thermofischer, Waltham, USA) and 30 µL of AMPure beads (Beckman Coulter Inc, California, US) were added to each well. The mixture was incubated at room temperature for 5 minutes and then placed on the magnetic stand until the liquid became clear. Supernatant from each well was discarded and DNA libraries were washed 2 times by adding 200 µL of freshly prepared 80% ethanol (Merck & Co, Inc, Kenilworth, Germany) to each well and incubated on the magnetic stand (Ambion, Forster City, US) for 30 seconds. The supernatant was discarded from each well and the plates were air-dried on the magnetic stand for 15 minutes. A volume of 52.5 µL of resuspension buffer (RSB) (Illumina, San Diego, USA) was added to each well and shaken for 2 minutes. The plate was then incubated at room temperature for 2 minutes and placed on the magnetic stand until the liquid became clear. About 50 µL of the supernatant was transferred into a new PCR plate (Bio-rad, Hercules, California, USA).

3.7.3. Library quantification and normalization

The DNA library was quantified using cubit assay (Thermofischer, Waltham, USA) and library normalization was done by transferring 20 µL of the supernatant from the PCR plate (Bio-rad, Hercules, USA, California) into a new PCR plate. In a conical tube (Sigma-Aldrich, St Louis, Missouri, USA) 4.4 mL library Normalization Additive 1 (LNA1) (Illumina, San Diego, USA) was mixed with 800 µL library normalization storage buffer 1 (LNS1). The bead mixture was added to the conical tube and 45 µL LNA1/LNS1 mixture was added to each well containing the libraries. The plates were placed on the magnetic stand until the liquid became clear. The supernatant was discarded, and libraries were washed twice with 45 µL of library normalization wash (LNW1). The supernatant was discarded and 30 µL of 0.1 sodium hydroxide (NaOH) (Merck & Co, Inc, Kenilworth, Germany) was added to each well. The 30 µL LNS1 was added to a new 96-well PCR plate (Bio-rad, Hercules, USA, California). The plate with the libraries was placed on the magnetic stand (Ambion, Forster City, US) until the liquid was clear and the supernatant

was transferred into the PCR plate containing the LNS1. The plate was centrifuged (Labnet, Edison, New Jersey, USA) at 1000 x g for 1 min.

3.7.4. Library pooling and sequencing

About 5 µL library was transferred into a new PCR 8-tube strip (Inqaba Biotechnical Industries, South Africa). The DNA library was denatured by mixing 5 µL of 4 nM DNA library with 5 µL of 0.2 N NaOH (Merck & Co, Inc, Kenilworth, Germany) in a microcentrifuge tube (Sigma-Aldrich, St Louis, Missouri, USA). The mixture was well mixed by vortexing (VELP Scientifica, Monza and Brianza, Italy) for 10 seconds, centrifuged (Labnet, Edison, New Jersey, USA) at 280 x g for 1 minute and then incubated at room temperature for 5 minutes. A volume of 990 µL pre-chilled HT1 (Illumina, San Diego, USA) was added to the Eppendorf tube (Inqaba Biotechnical Industries, South Africa) containing a denatured library. The 20 pM libraries were diluted and kept on ice until the next step. The final library mixture contained at least 5% PhiX (Illumina, San Diego, USA). A volume of 2 µL 10 nM PhiX library (Illumina, San Diego, USA) and 3 µL 10 mM Tris pH 8.5 (Sigma-Aldrich, St Louis, Missouri, USA) were mixed to dilute the PhiX library (Illumina, San Diego, USA). Five microlitres of 4 nM PhiX library (Illumina, San Diego, USA) and 5 µL of 0.2 N NaOH (Merck & Co, Inc, Kenilworth, Germany) were mixed in the Eppendorf tube (Inqaba Biotechnical Industries, South Africa) to make 2 nM PhiX library (Illumina, San Diego, USA) solution. The mixture was centrifuged (Labnet, Edison, New Jersey, USA) at 280 x g for 1 minute, followed by incubation for 5 minutes at room temperature to denature the PhiX library (Illumina, San Diego, USA) into single strands. A volume of 10 µL of PhiX library (Illumina, San Diego, USA) and 990 µL pre-chilled HT1 (Illumina, San Diego, USA) was added to the tube, which resulted in 20 pM PhiX library (Illumina, San Diego, USA). The denatured 20 pM PhiX library (Illumina, San Diego, USA) was diluted to the same loading concentration as the amplicon library. The DNA solution was inverted to mix and then pulse-centrifuged (Labnet, Edison, New Jersey, USA). The denatured and diluted PhiX library (Illumina, San Diego, USA) was then placed on ice. The amplicon library and PhiX control (Illumina, San Diego, USA) were mixed. Thirty microliters of the denatured and diluted PhiX control (Illumina, San Diego, USA) and 570 µL of the denatured and diluted amplicon library were combined in an Eppendorf tube (Inqaba Biotechnical Industries, South Africa). The combined sample library and PhiX control (Illumina, San Diego, USA) were set aside on ice until heat denaturation. The mixture of library

and PhiX control (Illumina, San Diego, USA) were incubated at 96°C for 2 minutes in a heat block (Labnet, New Jersey, USA). After the incubation, the Eppendorf tube (Inqaba Biotechnical Industries, South Africa) were inverted two times to mix and immediately placed in the prepared ice-water bath. The tubes were kept in the ice-water bath for 5 minutes and loaded into the Illumina MiSeq v3 reagent cartridge (Illumina, San Diego, USA) and sequenced. The paired-end read sequencing was performed with reading lengths of about 150 bp and then transferred into 2 mL Eppendorf tube (Inqaba Biotechnical Industries, South Africa).

3.8. Bioinformatic analysis

3.8.1 Detection of *M. tuberculosis* lineages and drug susceptibility

The resistance sniffer online software tool (http://resistance-sniffer.bi.up.ac.za) was used to detect the clades-specific polymorphism and the genetic determinants of drug resistance. The program mapped the 56 raw sequences to the embedded reference genome (*M. tuberculosis* H37Rv, NC_000962.3) and discriminated the clades-specific polymorphism and genetic determinants of drug resistance of each genome using the diagnostic keys consisting of bifurcating split for each decision points. The bifurcating splits were root, followed by node and clade name. At each intermediate node, the program detected the pattern of polymorphisms and calculated normalized counts of power depending on the states of sites in the genome. The bifurcating branch scores accepted for every single clade was 0.75 to 1.0. For each clade below than 0.75, the program discriminated the clades using all alternative branches to determine the similarities of polymorphisms that may be shared with sister clades and detect the clades at the highest score level.

3.8.2 Detection of lineage-specific SNPs

For the detection of lineage-specific SNPs and the construction of the phylogenetic tree, the WGS outputs were analyzed using the Pathosystem Resource Integration Center (PATRIC) version 3.6.2 (http://www.patricbrc.org). The genome sequence paired reads in fastq file format were assembled and aligned to the *M. tuberculosis* H37Rv reference genome (NC_000962.3). Alignments with an identity >98% and coverage >95% were retained for SNP-calling. Multiple sequence alignment of virulence genes was accomplished by Multiple Sequence Alignment using Fast Fourier

Transformation (MAFFT) version 7.452 (https://mafft.cbrc.jp) and the alignment results were analyzed and edited using Biological Sequence Alignment Editor (BioEdit) version 7.2.5.

3.8.3 Detection of lineage-specific SNPs and construction of phylogenetic tree using *M. tuberculosis* isolated from different countries.

Publicly available *M. tuberculosis* genomes from different countries were used to determine lineage-specific SNPs in virulence genes within Beijing and LAM lineage. The genomes were searched using the accession numbers and downloaded from PATRIC version 3.6.2 (http://www.patricbrc.org). Over 100 *M. tuberculosis* were downloaded and selected according to their genotypes. The Beijing and LAM genomes from different countries that have been deposited on PATRIC version 3.6.2 (http://www.patricbrc.org) were used to confirm and compare the signature SNPs detected in this study

The global phylogenetic tree of various published *M. tuberculosis* genotypes and phylogenetic tree of the Beijing and LAM genotypes from this study were constructed using the PATRIC version 3.6.2. One hundred genomes (86 genomes from PATRIC and 14 genomes from this study) were used to construct the phylogenetic tree. The genome sequences were aligned using Multiple Sequence Comparison by Log-Expectation (MUSCLE). The joint set of the amino acids and nucleotides alignment were concatenated into a data matrix, and RaxML was used to analyze this matrix, with fast bootstrapping to generate the support value in the tree. The tree was visualized using Interactive Tree Of Life (iTOL) online version 5 (https://itol.embl.de).

3.9. Quality control

The DNA from *M. bovis* BCG and *M. tuberculosis* H37Rv strains were used as a positive control and de-ionized water was used as a negative control for every spoligotyping and IS-6110 RLFP runs. The *M. tuberculosis* H37Rv reference genome was used for sequence alignment.

3.10. Statistical analysis

The sample size was calculated with the help of Mr. Charl Janse Van Rensburg, who is a biostatistician at SAMRC, Biostatistics Unit, Pretoria. Previously published articles were used to determine the sample size of Beijing and LAM isolates that were used to genotype *M. tuberculosis* by IS6110-RFLP and WGS. The clustering rate for *M. tuberculosis* genotypes was calculated using the formula defined as $(n_c\text{-}c)/n$, where n_c is the total number of clustered isolates, c is the number

of isolates' clusters and *n* is the total number of isolates genotyped by spoligotyping. A cluster was defined as two or more *M. tuberculosis* isolates with an identical genetic pattern determined by spoligotyping and all isolates with unmatched genetic profile were considered non-clustered strains.

3.11. Ethical approval

Permission was granted from the manager of the NHLS/TAD, Medical Microbiology laboratory, to collect isolates and perform the experiments in their laboratory. Ethical approval (610/2018) (see attached document on page 103) was obtained from the Research Ethics Committee, Faculty of Health Sciences, University of Pretoria.

Chapter 4

Results

In this study, a total number of 150 *M. tuberculosis* cultures stored in MGIT tubes were collected. Spoligotyping was done directly from the original cultures and for IS6110-RLFP and whole-genome sequencing (WGS), the isolates were sub-cultured in a Middlebrook 7H9 to increase the yield of *M. tuberculosis* cells and to assess the purity of the cultures. Only four Isolates were contaminated and were discontinued from further genotypic analysis.

4.1. Screening of *M. tuberculosis* lineages using spoligotyping

Spoligotyping of the 150 *M. tuberculosis* isolates resulted in 86.7% (n = 130 isolates) previously shared type (ST) and 13.3% (n = 20 isolates) orphans, which were defined as isolates with unique patterns that could not be found on the SpolDB4 SITVIT database. The overall spoligotyping resulted in 16 clusters and 39 unique cases. In this study, 19 genotypes were identified, which included Beijing (SIT 1) and non-Beijing lineages and sub-linages: CAS1_Delhi (SIT 36), CAS1_Kili (SIT 21), EAI1_SOM (SIT 48 and 1251), EAI5 (SIT 236), H1 (SIT 47, 185 and 833), H3 (SIT 268) H3-T3 (SIT 36), LAM1 (SIT 271), LAM3 (SIT 33 and 111), LAM4 (SIT 60), LAM9 (SIT, 42 and 385), LAM11_ZWE (SIT 1607), Manu 1 (SIT 100) Manu 2 (SIT 54 and 1088), S (SIT 34 and 789), T1 (SIT 53, 245, 613 and 1800) T2 (SIT 52 and 118), T2-T3 (SIT 73) and X3 (SIT 70 and 93) as shown in Table 4.1 below. The Beijing family was found to be the most predominant lineage with a 26.7% prevalence, followed by T family with 16%, Latin American and Mediterranean (LAM) with 13.3%, East Africa Indian (EAI) with 8.7%, S with 6%, Manu with 4.7%, H with 4.7%, CAS with 4.0% and X3 as the least lineage with 2.7% prevalence (Figure 4.1). Spoligotyping identified 111 isolates that were clustered into 16 groups or clusters, giving a clustering rate of 63.3%.

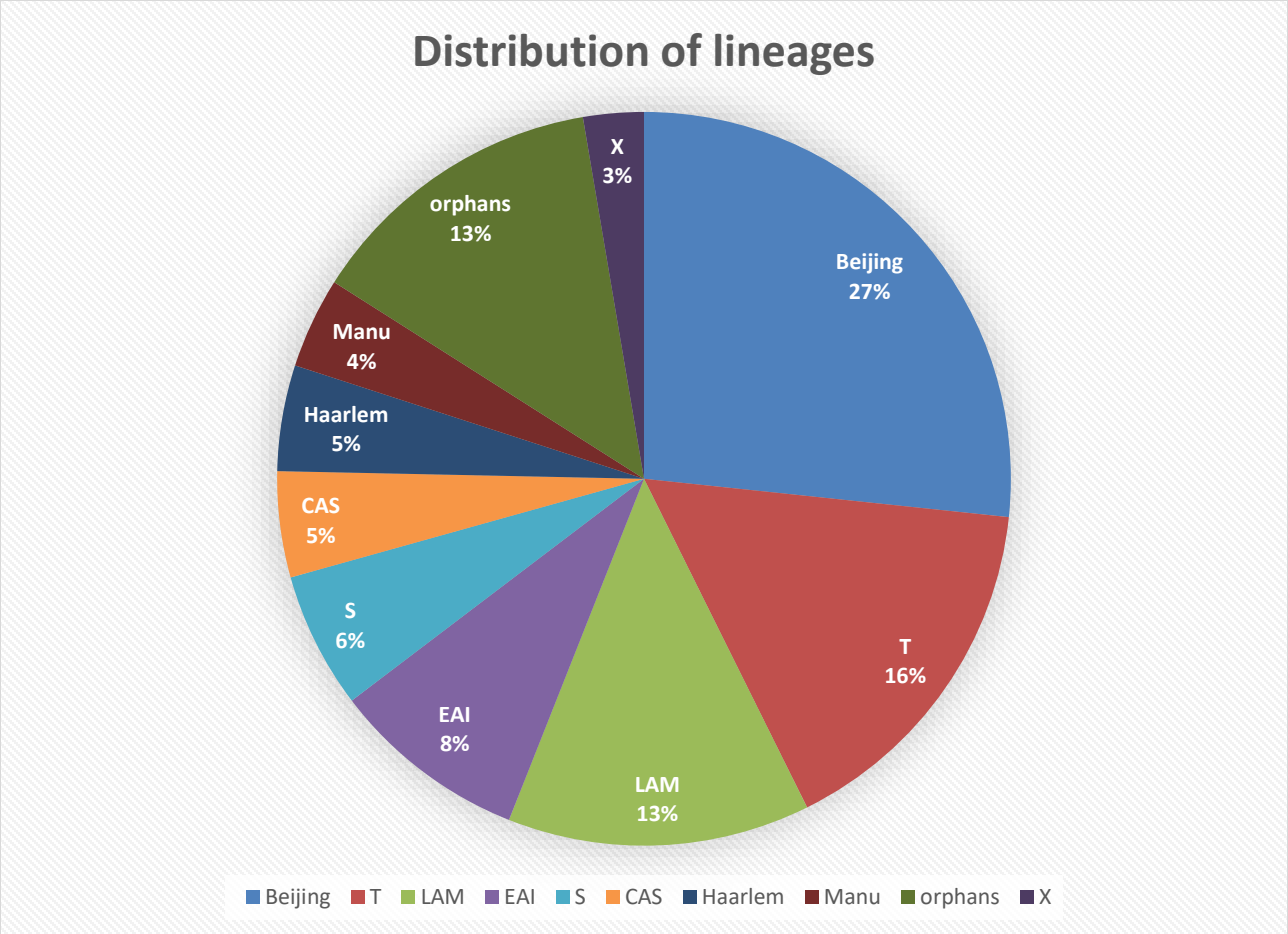

Figure 4. 1: Distribution of lineages detected by spoligotyping

Of the 150 *M. tuberculosis* culture samples collected from the National Health Laboratory Service/Tshwane Academic Division (NHLS/TAD), 57.3% (n = 86) of the *M. tuberculosis* isolates were susceptible to the first-line anti-tuberculosis regimen viz, rifampicin (RIF), isoniazid (INH), pyrazinamide (PZA), and ethambutol (EMB). The overall drug-resistant isolates were 42.7% (n = 64 isolates); of these isolates, 4.7% (n = 7 isolates) were RIF (3.3%, n = 5) and INH (1.3%, n = 2) mono-resistant. Further, 22.7% (n = 34 isolates) were multidrug-resistant (MDR)-TB and 15.3% (n = 23 isolates) were extensively drug-resistant (XDR)-TB as shown on Table 4.1 below.

Table 4. 1: Association between *M. tuberculosis* lineages and drug susceptibility profiles

Lineages	Number of isolates per lineage	Antibiotic Susceptibility Results				
		Sensitive	MDR-TB	XDR-TB	RIF (R)	INH (R)
Beijing family	**40**	26	6	7	1	0
T family	**24**					
T1	19	11	6	1	1	0
T2	4	2	2	0	0	0
T2-T3	1	1	0	0	0	0
LAM family	**20**					
LAM1	1	1	0	0	0	0
LAM3	11	7	2	0	2	0
LAM4	3	2	0	1	0	0
LAM9	2	1	1	0	0	0
LAM11_ZWE	3	1	1	1	0	0
EAI family	**13**					
EAI1_SOM	11	3	3	5	0	0
EAI5	2	0	1	1	0	0
S family	**9**	4	1	3	0	1
CAS family	**6**					
CAS1_Delhi	3	2	1	0	0	0
CAS1-Kili	3	3	0	0	0	0
Haarlem family	**7**					
H1	4	3	1	0	0	0
H3	1	0	1	0	0	0
H3-T3	2	1	1	0	0	0
Manu family	**7**					
Manu1	1	0	0	1	0	0
Manu2	6	1	2	2	0	1
X3 family	**4**	2	1	1	0	0
Orphan	20	15	4	0	1	0
Total	150	86	34	23	5	2

4.2. Association between lineages and drug-resistant TB

The number of susceptible *M. tuberculosis* isolates per lineages was higher than drug-resistant TB (Figure 4.2). The genotypes with a greater number of isolates that were sensitive were: Beijing (17.3%), T family (9.3%), LAM family (8%) CAS (2.67%) and isolates classified as orphans (10%). The Beijing lineage was the predominant lineage with more isolates that were sensitive to the first-line anti-tuberculosis regimen. The association between anti-tuberculosis drug-resistant TB and lineages was found in EAI lineage (6.7%), Manu (4%) and S family (3.3%). The family with a high number of isolates which were drug-resistant TB was the EAI1-SOM sub-lineage belonging to the EAI family. The association between the lineages and their susceptibility was represented by the phylogenetic tree (Figure 4.3) which was used to group the isolates within their lineages according to their susceptibility profile. The tree was contracted using the MIRU-VNTRplus database (http://www.miru-vntrplus.org).

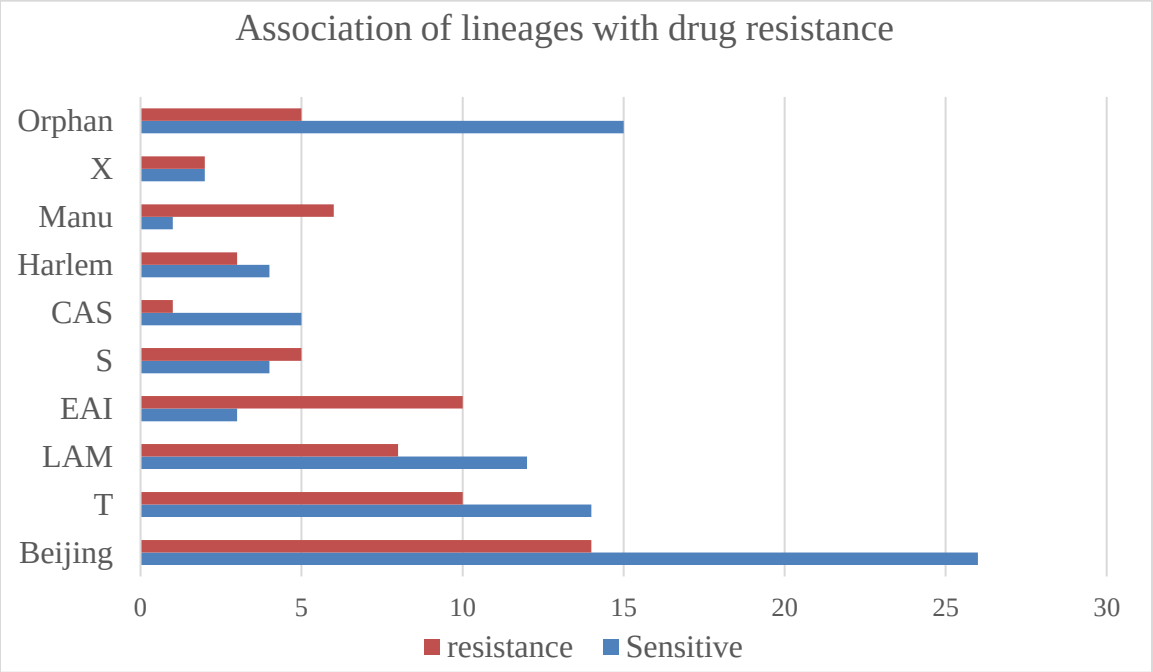

Figure 4. 2: Association of TB lineages with anti-tuberculosis drug resistance. The column shaded in blue color are isolates that are sensitive to anti-TB regimen and the red color denotes isolates that are drug-resistant: MDR-TB, XDR-TB, RIF mono resistant TB and INH mono resistant TB.

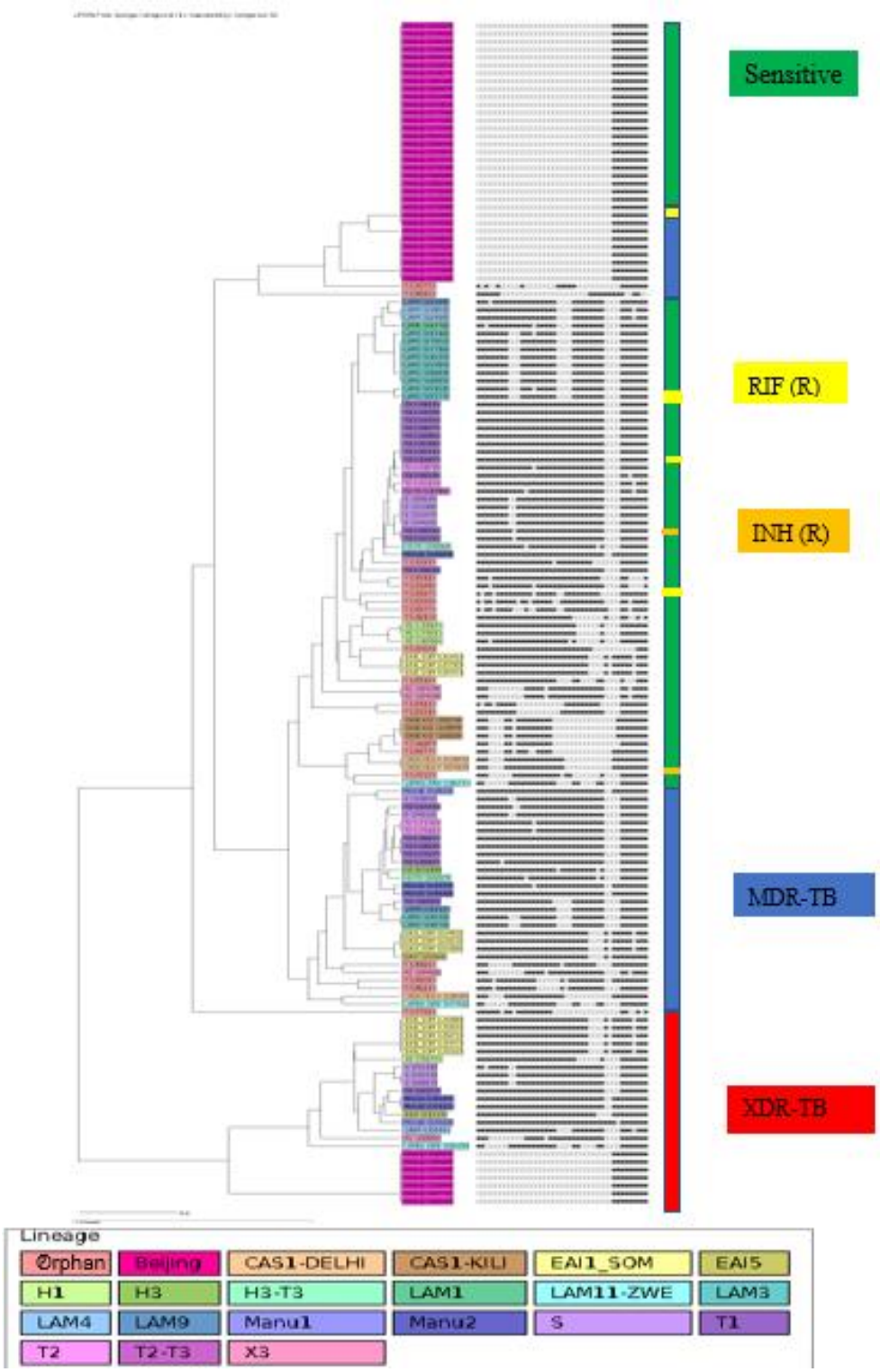

Figure 4. 3: A dendrogram (UPGMA) of the *M. tuberculosis* genotypes detected by spoligotyping. The isolates within the genotype were grouped according to their susceptibility profile.

4.3. The IS6110-RLFP results

IS6110-RLFP did not give good results that could be interpreted. For all 60 isolate which were identified as Beijing (40 isolates) and LAM (20 isolates) lineages using spoligotyping, 4 isolates showed contamination from the blood culture plate and were excluded from this study. The results obtained from the 56 genomic DNA samples used were showing faint bands (Figure 4.4) on the X-ray film, while some digested DNA fragments were not transferred to the membrane during western blotting. The experiment was repeated but resulted in faint bands that could not be interpreted and due to lack of reagents, this technique could not be done, and the results were not included in this study.

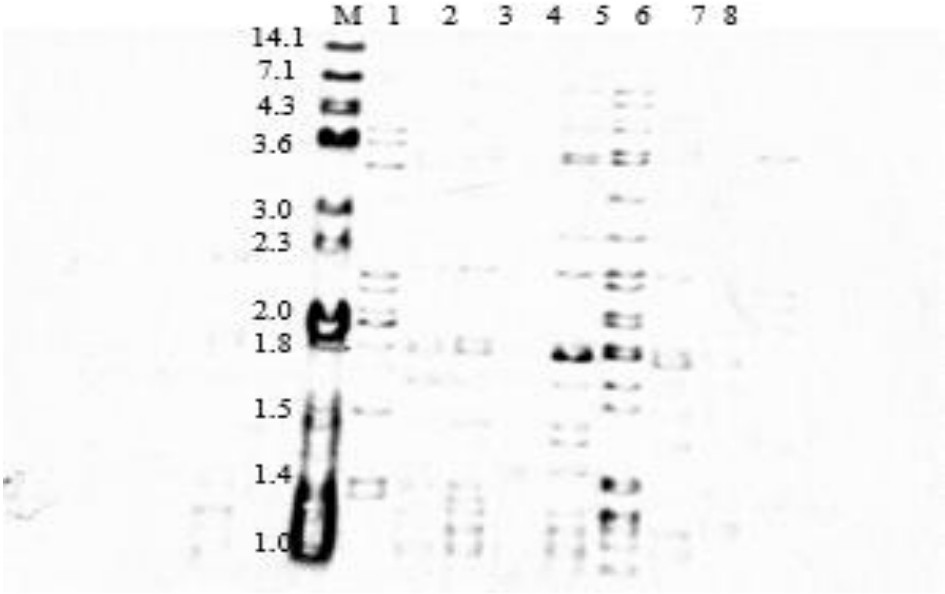

Figure 4. 4: The IS6110-RLFP results for Beijing and LAM isolates

4.4. Whole-genome sequencing and bioinformatic analysis

The WGS fastq paired-read files of the 56 genomic DNA sequenced by the Illumina Miseq platform were obtained; 24 sequence runs (14 from Beijing isolates and 10 from LAM isolates) failed the prediction of lineages and WGS drug sensitivity due to poor sequence quality, and 32 remaining sequenced genomic DNA were analyzed to identify lineages and drug susceptibility using resistance sniffer bioinformatic online software tool (http://resistance-sniffer.bi.up.ac.za).

4.4.1. Detection of drug susceptibility using resistant sniffer online software tool

The resistance sniffer bioinformatic online tool program was used to determine whether the isolates were resistant or susceptible to the first- and second-line anti-TB regimens. The anti-TB drugs included were amikacin (AMK), capreomycin (CM), cycloserine (CS), EMB, ethionamide (ETH), INH, fluoroquinolones (FLQ), kanamycin (KAN), ofloxacin (OFL), para-aminosalicylic acid (PAS), PZA, RIF and streptomycin (SM). For each clade, the associated sets of polymorphic sites, which distinguished between the antibiotic-resistant and antibiotic sensitive variants were calculated.

The drug susceptibility test (DST) results detected by BACTEC MGIT 960 susceptibility testing and LPA (done at the NHLS/TAD) were compared to the WGS results detected by resistance sniffer bioinformatic online software tool (http://resistance-sniffer.bi.up.ac.za). The analysis showed that 16 isolates were similarly identified by both phenotypic DST and WGS. Of these isolates, 13 were susceptible *M. tuberculosis*, and three isolates were drug-resistant TB, i.e. two isolates were MDR-TB and one isolate was XDR-TB. There was one isolate detected as RIF mono resistant by LPA which the WGS showed that the isolate was also resistant to INH, PZA, and EMB. Six isolates were sensitive by phenotypic DST but were detected by WGS as isolates that were most likely to be resistant to both RIF and INH (MDR-TB). Four isolates were sensitive by phenotypic DST while WGS revealed that they were sensitive but with a likelihood of this isolates to become MDR-TB (have intermediate (shown in yellow columns in Figure 4.5) results to RIF and IHN) and may have reduced sensitivity. Three isolates were MDR-TB (1 isolate) and XDR-TB (2 isolates) phenotypically, whereas the WGS results showed that RIF was still susceptible (shown in white columns, Figure 4.5) with no evidence of mutations. One of these isolates was sensitive to RIF and INH, while two isolates were susceptible to RIF and PZA. Three isolates were detected as susceptible-TB (2 isolates) and MDR-TB (1 isolate) phenotypically, but the WGS failed to detect their sensitivity. These were isolates that were predicted as Beijing by spoligotyping while WGS predicted them as X lineage.

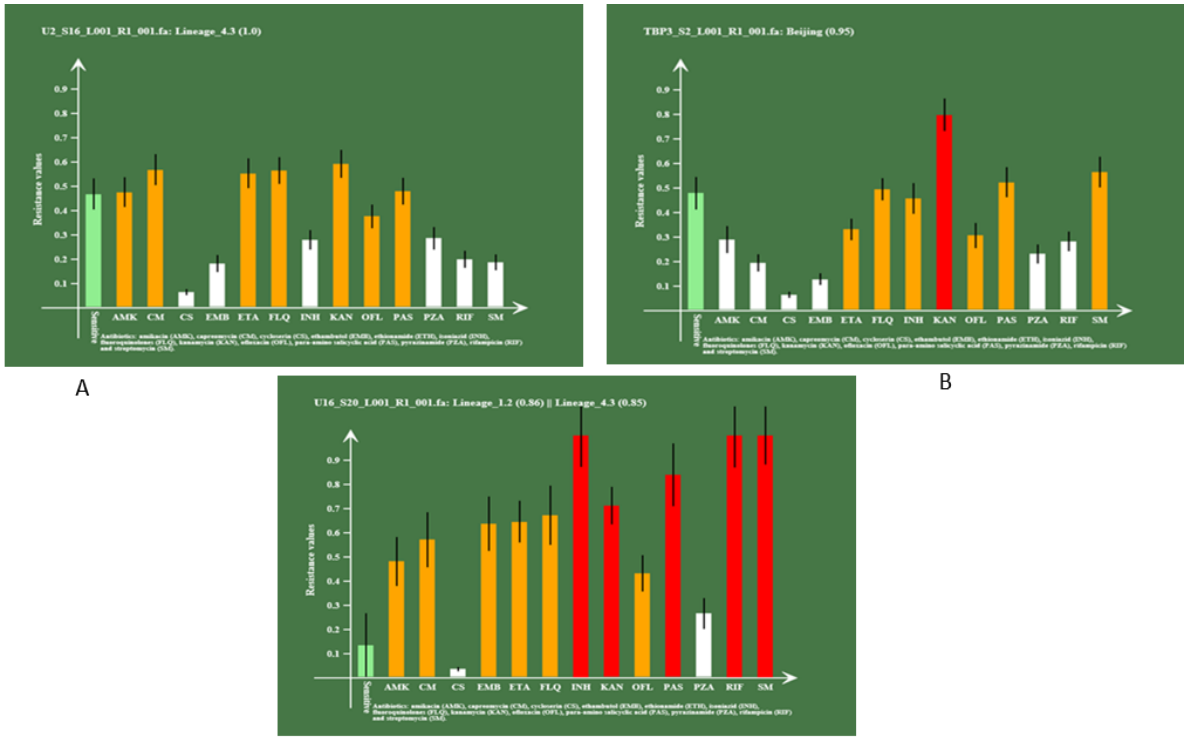

Figure 4. 5: Drug resistance predictions by Resistance Sniffer. (A) shows strain (U6369) is sensitive to first-line (RIF, INH, PZA, EMB) TB regimen with an emergence of drug resistance (intermediate) to second-line (FLQ, AMK, KAN, CM, SM) TB regimen, (B) shows strain (U6136) which is also sensitive with similar results to (A), however, KAN is likely to be resistant, (C) shows strain (U9560) is more likely to be XDR-TB. White columns show sensitivity to antibiotics with the confidence above 55%, red columns predict the resistance with the confidence above 55%, and orange columns show intermediate results. The green most right column depicts the likelihood for this strain to be sensitive to all 13 antibiotics.

4.3.2. Detection of lineages by WGS

The detection of lineages was done using the resistance sniffer online software tool (http://resistance-sniffer.bi.up.ac.za), which discriminated the polymorphic clade from its root, followed by node and clade name. At each intermediate node, the program detected the pattern of polymorphisms and calculated normalized counts of power depending on the states of sites in the genome. The bifurcating branch scores accepted for every single clade was above 0.75. Of the 32 sequenced genomic DNA, WGS agreed with spoligotyping detection of lineages in 20 isolates, and there were discordant results in 12 isolates. Thus, three isolates detected as Beijing by spoligotyping were detected as Asian strain (node) by the resistance sniffer online software tool (http://resistance-sniffer.bi.up.ac.za) and were denoted as X lineage, and the remaining 9 isolates could not be completely discriminated because the resultant genome sequences had poor quality.

Therefore, the lineages were discriminated by their root, node, but the program could not decide the clade name to which the isolates belonged to.

4.3.3. Lineage-specific SNP in virulence genes

The whole-genome sequences were aligned using the Pathosystem Resource Integration Center (PATRIC) version 3.6.2 (https://www.patricbrc.org). Alignments with an identity >98% and coverage >95% were retained for SNP-calling. Multiple sequence alignment against *M. tuberculosis* H37Rv for all virulence genes was accomplished by Multiple Sequence Alignment using Fast Fourier Transformation (MAFFT) version 7.452 online version (https://mafft.cbrc.jp) and the alignment results were visualized using Biological Sequence Alignment Editor (BioEdit) software version 7.2.5 to detect the SNPs within the virulence genes.

As guided by previous studies (Dou *et al*., 2017; Naidoo and Pillay, 2017; Mikheecheva *et al*., 2017; Zaychikova *et al*., 2015), a total of 41 *M. tuberculosis* virulence genes were used to detect SNPs specific to Beijing (26 genes) and LAM (15 genes) lineages. From the 41 virulence genes, 27 virulence genes (Table 4.2) were shown to have SNPs that are specific for Beijing and LAM lineages. For SNPs specific to lineages, only genome sequences with unambiguously defined genotypes and a reasonable number of mutations were used to discriminate the lineages and added to the lineage-specific SNPs catalog. A total of 14 (10 Beijing and 4 LAM) lineages were used to identify the lineage and sub-lineage signature SNPs and to construct the phylogenic tree (Figure 4.7). Of the 10 Beijing strains, 22 genes were used; thus, this study found 29 overall SNPs specific for Beijing lineage. From these 22 genes, 10 SNPs were detected and found to be common in the Beijing family, and 19 SNPs were signature SNPs for the Beijing-modern sub-lineage. From the four LAM strains used, six SNPs specific for LAM sub-lineages were found: two isolates were LAM3, the other two were LAM9 and LAM11-ZWE as detected by spoligotyping and WGS. Mutations found in *fadE*28 (874 A>C), *mce*3*F* (992 C>G), *mce*1*D* (794 C>T) and *hsp*22.5 (183 C>A) were found to be common in all LAM sub-lineages, while SNP (1392 C>G) in *fadD*28 was found to be lineage signature for LAM3, and SNP (494 A>C) in mmaA4 gene was found only on LAM9 and LAM11-ZWE. This study detected SNPs in *fadD*28 (1521 C>T), *eccCb*1 (1479 G>A), *pks*5 (6210 G>A), and *ponA*2 (372 G>T) among Beijing strains and *fadD*28 (1392 C>G) within LAM strains which were not reported in previous studies.

Table 4. 2: Lineage-specific SNPs within the virulence genes of Beijing and LAM *M. tuberculosis* lineages

Virulence genes	Groups	Lineage	SNPs
glcB	Lipid and fatty acid metabolism	Beijing-modern	G310A
kefB	Cell wall protein	Beijing	A304G
mce4C	Cell wall protein	Beijing-modern	C571A
mce1D	Cell wall proteins	LAM	C794T
mce3F	Cell wall proteins	LAM	C992G
fadE28	Cholesterol catabolism	LAM	A874C
cyp125	Cholesterol catabolism	Beijing-modern	C1125T
ltp2	Cholesterol catabolism	Beijing	G670C
fadD28	Lipid and fatty acid metabolism	Beijing-modern LAM3	A1306G C1521T C1392G
lipF	Lipid and fatty acid metabolism	Beijing-modern	C697T
mas	Lipid and fatty acid metabolism	Beijing-modern	A6013C
pks5	Lipid and fatty acid metabolism	Beijing Beijing-modern	T6182G G6210A
pks7	Lipid and fatty acid metabolism	Beijing	A2441C
plcA	Lipid and fatty acid metabolism	Beijing Beijing-modern	A1336G G705A
plcC	Lipid and fatty acid metabolism	Beijing-modern	T753C G1081T
mbtB	Metal transporter proteins	Beijing	G2020C
mmaA4	Mycolic acid synthesis	LAM	A494G
ponA2	Proteins inhibiting antimicrobial responses of the macrophage	Beijing-modern	G372T G1855A
sigG	Sigma factor	Beijing	G860A G994T
mazF8	TA systems	Beijing	G122T
mazF3	TA systems	Beijing	C194T
vapB47	TA systems	Beijing-modern	C250T
vapC37	TA systems	Beijing-modern	A46G
vapC38	TA systems	Beijing-modern	T143C
eccCb1	Type VII secretion system	Beijing-modern	G1479A
secA2	Type VII secretion system	Beijing modern	G1228A C1830T
hsp22.5	Heat-shock protein	LAM	C183A

4.3.4 Detection of lineage-specific SNPs using *M. tuberculosis* isolated from various countries.

The lineage-specific SNPs detected in the virulence genes of *M. tuberculosis* Beijing and LAM genome from different countries which were obtained from PATRIC version 3.6.2

(https://www.patricbrc.org) (see Appendix B) were found to be similar to the lineage-specific SNPs detected in this study (Table 3). Within the Beijing-modern sub-lineages, this study detected signature SNPs for Beijing-BO/W148 which was isolated in Russia. The present study also confirmed the signature SNP, *fadD*28 (1392 C>G) detected in LAM3 isolated in this study with LAM3 which were isolated from Western Cape, South Africa. Within the South African strains, this study also found lineage-specific SNPs for F15/LAM4/KZN strain which was first isolated in KwaZulu Natal.

Table 4. 3: A set of lineage-specific SNPs catalog for genotyping *M. tuberculosis* Beijing and LAM lineages.

Lineages	Country of isolation	Lineage-specific SNPs
Beijing-modern	South Africa Swaziland Uganda Mozambique Thailand Russia Papua New Guinea	*glcB* (G310A), *mce*4*C* (C571A), *cyp*125 (C1125T), *lipF* (C697T), *mas* (A6013C), *pks*5 (G6210A), *plcA* (G705A), *plcC* (T753C), *plcC* (G1081T), *fadD*28 (A1306G), *fadD*28 (C1521T), *ponA*2 (G372T), *ponA*2 (G1855A), *vapB*47 (C250T), *vapC*37 (A46G), *vapC*38 (T143C), *eccCb*1 (G1479A), *secA*2 (G1228A), *secA*2 (C1830T)
Beijing-BO/W148	Russia	*mce*3*B* (T145G), *eccCb*1 (G1556T), *vapC*12 (A95G)
LAM3	South Africa	*FadD*28 (C1392G)
F15/LAM4/KZN	South Africa	*cyp*125 (T1076C), *mce*3*B* (T44C), *vapC*25 (A221C), *vapB*34 (C140A)

4.4 Phylogenetic Analysis

The global phylogenetic tree of various published *M. tuberculosis* genotypes and phylogenetic tree showing the relationship within Beijing and LAM genotypes from this study were constructed using the PATRIC version 3.6.2 (http://www.patricbrc.org) and visualized using Interactive Tree Of Life (iTOL) online version 5 (https://itol.embl.de). The global phylogenetic tree (Figure 4.6)

was constructed using 5 main lineages which were lineage 1 (EAI), lineage 2 (Beijing), Lineage3 (CAS), lineage 4 (Euro-American), and lineage 7 (Aethiops vetus). The lineages detected from this study are encircled on the tree. Genotypes detected in this study were grouped with other isolates of the same genotype with the bootstrap support value of >50%.

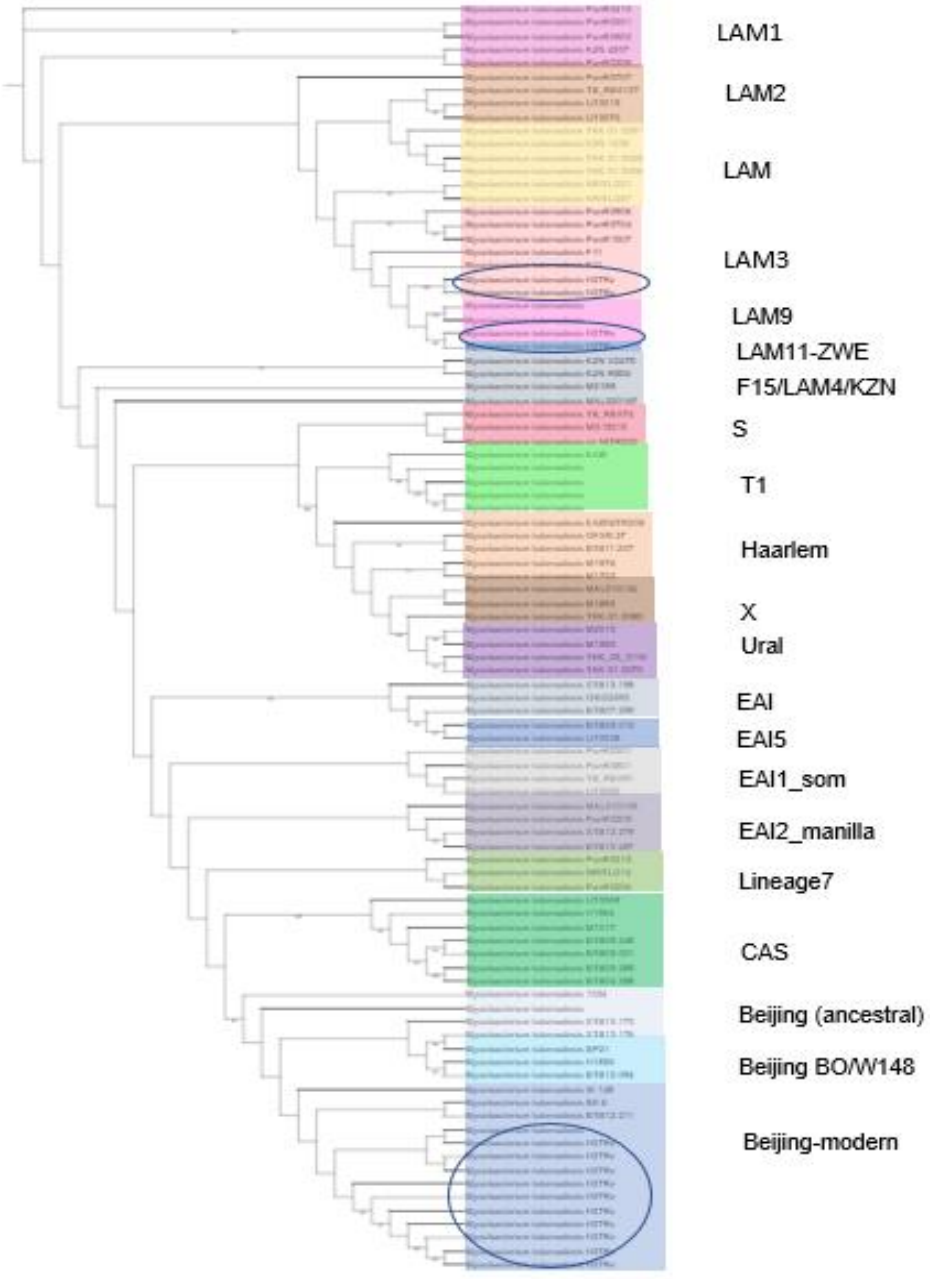

Figure 4. 6: Phylogenetic relationship between 5 main *M. tuberculosis* lineages. The tree was constructed using PATRIC version 3.6.2 and visualized using iTOL version 5.

The phylogenetic tree of 14 isolates that were used for the detection of lineage-specific SNPs in this study was constructed. The tree was showing the division of the strains into the main Beijing or LAM lineages. Within the Beijing branch, was the Beijing-modern genotype to which all Beijing isolates were branched showing that all Beijing isolates detected in this study were found within the Beijing-modern sub-lineage. Within the LAM sub-lineage, the branch containing LAM 9 was located nearer to LAM 11-ZWE sub-lineage, with a bootstrap support value of 99%. Two isolates were identified as LAM3 with a bootstrap support value of 100%. The tree showed high bootstrap support value, which was over 60% as indicated in Figure 4.7.

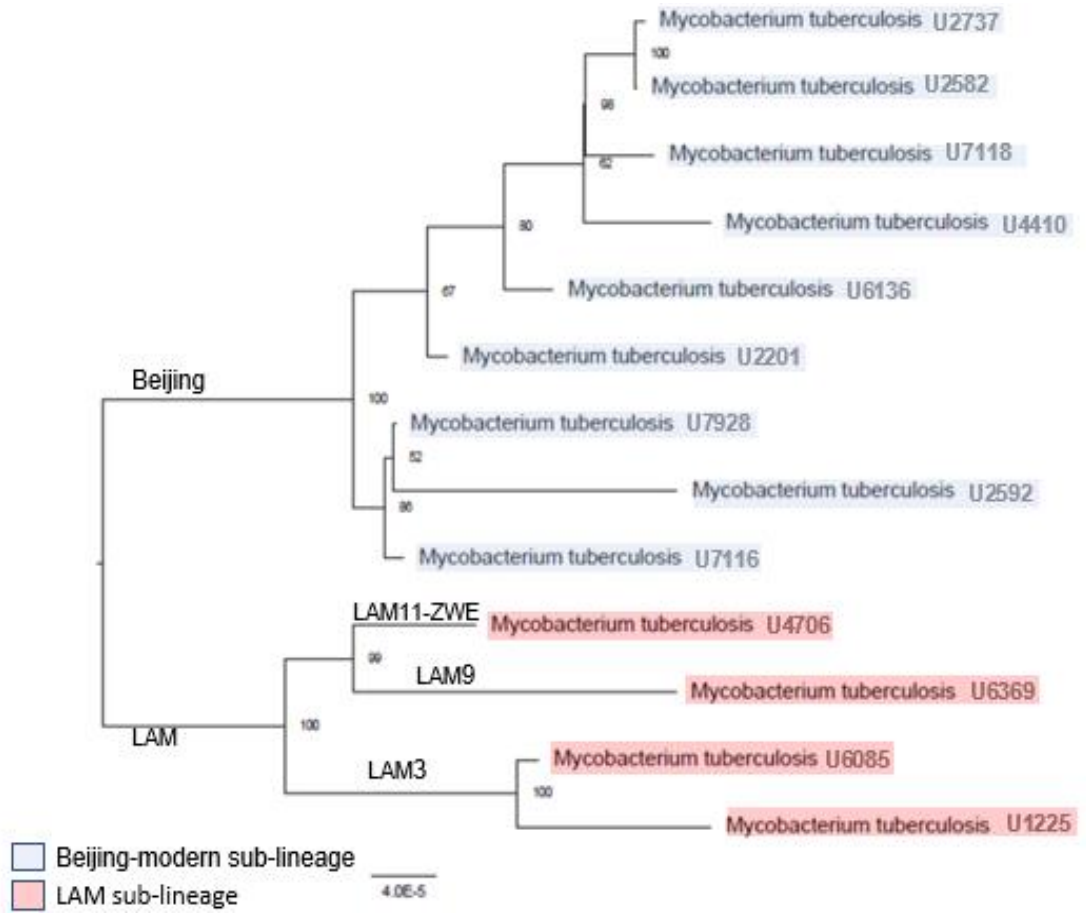

Figure 4. 7: A dendrogram showing genetic relatedness of 14 isolates belonging to Beijing (10 isolates) and LAM (4 isolates) family. These isolates were used for lineage-specific SNPs identification. The tree was constructed using PATRIC bioinformatics analysis software. The tree showed a high bootstrap support value, which was over 60%.

Chapter 5

Discussion and conclusion

5.1. Discussion

South Africa is one of the top 30 countries with a remarkably high TB burden (WHO, 2018). Therefore, understanding the genetic diversity of *M. tuberculosis* is very crucial for rapid diagnosis and reduction of TB transmission. *M. tuberculosis* evolve to adapt inside host cells in response to the host's immune reactions (Madacki *et al*., 2019). The leading factors that cause genetic diversity in *M. tuberculosis* are anti-tuberculosis drugs, suppression of the immune system caused by underlying diseases, environmental conditions, and microbiota dysbiosis. These factors may also contribute to the significant change in the genome of *M. tuberculosis* and may cause the emergence of new lineages/sub-lineages with higher virulence and drug resistance (Mikheecheva *et al*, 2017).

Lineages identification is of great importance for proper treatment assignment, and to eradicate the global transmission of different strains of *M. tuberculosis (*Zaychikova *et al*., 2015*)*. The diversity of *M. tuberculosis* lineages has been described in various provinces of South Africa including the Western Cape, Gauteng, Kwazulu-Natal, North West, and Limpopo where Beijing was found to be the commonest lineage *(*Maguga-Phasha *et al*., 2017). This study found substantial genetic diversity of *M. tuberculosis* in the Tshwane region with the Beijing lineage (26.7%) being the most predominant strain. Furthermore, Beijing lineage was found to be over-represented in South Africa as compared to other African countries (Osei Sekyere *et al*.,2019).

All studied lineages belonged to lineage 1 to 4 of *M. tuberculosis* and lineage 5, 6, and 7 could not be identified in this study. That could be because lineage 5, 6 and 7 are believed to be highly restricted to their localized region. Lineages 5 and 6 are found across West Africa, lineage 7 is found in Ethiopia, and they are rarely found in other countries (Mikheecheva *et al*, 2017).

Our study agrees with reports found by Chihota *et al*., (2018), which reported that the LAM sub-lineage that is highly isolated in large proportion in South Africa is the LAM3. Of all LAM sub-lineages isolated in this study, LAM3 was the most isolated genotype. Although the LAM4/F15/KZN sub-lineage was firstly isolated in South Africa, after the outbreak that occurred in the Tugela hospital, Kwazulu-Natal, this lineage has not been found to be highly prevalent in

the Gauteng Province (Maguga-Phasha *et al*., 2017). Spoligotyping results detected only 2% (3/150) of LAM4/F15/KZN sub-lineages in this study, however, this lineage was not further detected by WGS because isolates were contaminated and excluded from WGS method.

The IS6110-RFLP could not give enough results that could be used in this study even after being repeated. Some samples gave results that were faint and could not be interpreted while some failed to be transferred from the agarose gel to the membrane. This could be because we got insufficient genomic DNA for genotyping since IS6110-RFLP requires 6 µg genomic DNA. On the other hand, the probe used could have been degraded and we could not get a newly prepared probe due to a lack of funds to buy all reagent needed. For these reasons, the Beijing and LAM lineages detected by spoligotyping were not confirmed using IS6110-RFLP and the discriminatory power of SNPs in virulence genes could not be compared to that of IS6110-RFLP.

All the 150 *M. tuberculosis* cultures were already tested for antimicrobial susceptibility for first-line antituberculosis drugs with BACTEC MGIT 960 instrument and with line probe assay to further confirm drug-resistant TB at the NHLS/TAD. The findings showed that many of the strains were still susceptible to first-line TB regimen viz., 57.3 %. However, *M. tuberculosis* resistance of 42.7% was also reported. South Africa has been rated among the top 20 countries with high TB drug resistance worldwide (WHO, 2019). Furthermore, our country is also rated among the top 20 countries with a high burden of HIV-TB co-infection, which is one of the predisposing factors of drug-resistant TB. This study also identified mono resistance, RIF monoresistance and INH resistance.

The present study has further detected anti-tuberculosis drug susceptibility for all first-line anti-tuberculosis drugs viz., RIF, INH, EMB, and PZA, as well as selected second-line drugs such as FLQ, capreomycin (CM), AMK, OFL, KAN para-aminosalicylic acid (PAS), streptomycin (SM), cycloserine (CS), and ethionamide (ETH) using the resistance sniffer online tool (http://resistance-sniffer.bi.up.ac.za). This software has helped to identify the probability of every single drug to become resistant, by detecting all known and un-known mutations in the genes that are associated with drug resistance. From our results, this study identified 31.3% (10/32) of isolates that were reportedly sensitive by BACTEC MGIT 960 while the resistance sniffer online tool (http://resistance-sniffer.bi.up.ac.za) showed the likelihood of these isolates being drug-resistant

M. tuberculosis. A study by Omar *et al*. (2019) suggested that the application of WGS in detection of *M. tuberculosis* drug resistance is shown to be valuable.

The association between drug-resistant TB and lineages was found in EAI lineage (6.7%), Manu (4%) and S family (3.3%), where more isolates were found to be resistant than those which were susceptible. In most studies, isolates belonging to the Beijing lineage were most likely found to be drug-resistant TB, specifically MDR-TB (Gupta *et al*., 2014). However, this study found a high number of Beijing isolates to be drug-susceptible and were not fully associated with drug resistance. Similarly, a study by Lui *et al*. (2017) found that the drug-resistant rate of Beijing family stains was not higher than that of non-Beijing strains. The results are inconsistent with several publications in different regions worldwide. Some studies reported the association of drug-resistant TB with Beijing lineage while other studies clearly reported no association between drug resistance and specific clades/lineages (Liu *et al*., 2017).

In agreement with results obtained using spoligotyping and DST (MGIT and LPA), genotyping using WGS also revealed that there was no association between MDR-TB or XDR-TB and Beijing or LAM lineages seen in this study. However, the resistance sniffer bioinformatics tool showed that all susceptible Beijing strains were resistant to the KAN. Köser *et al*. (2013) showed that the Beijing-modern isolates harbored a frameshift mutation in *tap* (at codon 194 out of 419 codons), which likely lead to low-level kanamycin mono-resistance in this group of Beijing isolates.

This study established a range of virulence genes that have mutations specific to Beijing and LAM lineages. Virulence is defined as the ability of the pathogen to cause disease through the process of invasion and adhesion to the host cells, and adaptation to the hostile environment, escaping the host immune response, resulting in tissue damage (Mikheecheva *et al*, 2017). The virulence genes included were regulatory and structural genes that are essential for various stages of pathogenesis.

This study used virulence genes to determine the relationship between the mutations within virulence genes and a particular lineage, suggesting that higher virulence poses a threat to patients by worsening the infection and may also increases the risk of emergence of drug resistance (Mikheecheva *et al*., 2017). The results obtained in this study are in agreement with the finding obtained by Mikheecheva *et al*. (2017); Naidoo and Pillay (2017) and Zaychikova *et al*. (2015) that showed that genotyping *M. tuberculosis* using polymorphisms in virulence genes may be an alternative approach to determine the *M. tuberculosis* genotypes. The present study has used *M.*

tuberculosis genomes from different countries, which were downloaded from PATRIC to compare and confirm the lineage-specific SNPs detected in this study. Our study found similar mutations in the genomes from different countries, suggesting that genotyping using polymorphisms in virulence genes of *M. tuberculosis* approach can be globally applicable.

Lineages represent the ancestry and geographical origin of different strains of *M. tuberculosis* and are globally distributed with some families still restricted to their country of origin (Gagneux, 2013; O'Neill *et al*., 2018). As these strains are distributed geographically to different countries, they adapt to the new environment and mutate, and may be different from the ancestral strain. This study detected the modern Beijing strain, Beijing BO/W148, isolated from Russia and LAM4 sub-lineage, F15/LAM4/KZN, isolated in South Africa, which has been reported to be more virulent. The F15/LAM4/KZN strain caused the largest XDR-TB outbreak in HIV-positive patients in the Tugela Ferry hospital, KwaZulu-Natal Province, South Africa, while the Beijing BO/W148 was reported to have an increased virulence, associated with MDR-TB and increased transmissibility when compared to other Beijing strains (Bespyatykh *et al*., 2019; Gandhi *et al*., 2006; Naidoo *et al*., 2017). The lineage-specific SNPs detected in Beijing BO/W148 were *mce*3*B* (145 T>G), *eccCb*1 (1556 G>T), and *vapC*12 (95 A>G) whilst F15/LAM4/KZN specific-SNPs detected were *cyp*125 (1076 T>C), *mce*3*B* (44 T>C), *vapC*25 (221 A>C), and *vapB*34 (140 C>A). Similar lineage-specific SNPs from these genotypes were reported by Mikheecheva *et al*. (2017).

Although detection of lineages by spoligotyping method is simpler and more straightforward, its discriminatory power is low as compared to IS6110-RFLP technique (Jeon *et al*., 2018), even though genotyping of *M. tuberculosis* by spoligotyping may give correct phylogenetic lineage in about 90% of cases (Dou *et al*., 2017). This study found that some strains may not be classified at all, which may be because the spoligotyping database lacks the signature spacers used to classify lineages or it might be a newly developed lineage or sub-lineage, and some isolates may be misclassified. The present study found that WGS can be used to determine lineages that may be misclassified or that may not be classified by spoligotyping, however, the results from this study may not be used to conclude these findings, since some genome were affected by poor sequence quality and were excluded.

5.2. Conclusion

This study showed a high genetic diversity of *M. tuberculosis* strains circulating within the Tshwane region. The Beijing lineage identified in this study was found to be more predominant than the rest of the identified genotypes. The prevalence of the Beijing lineage should be of great concern in the Tshwane region, because this lineage may be associated with the severity of TB, rapid transmission, drug resistance and poor treatment outcome. The spoligotyping technique has lower discriminatory power than any other genotyping method, but it is cost-effective and rapid. Thus, this technique could be used for primary screening and selection of *M. tuberculosis* lineages. The study found that the Illumina Miseq WGS could be considered a promising alternative method to discriminate lineages and sub-lineages. This method may also be used to identify some lineages that have been misdiagnosed using spoligotyping due to human errors or for those lineages that are assigned as orphans by spoligotyping as those lineages might be newly developed. This study established that WGS is a promising technique for genotyping *M. tuberculosis* during the event of outbreaks and to detects transmission events missed during epidemiological investigations.

This study proposed the alternative method for genotyping *M. tuberculosis* strains using SNPs in virulence genes of *M. tuberculosis*. Through this approach, the study successfully identified 29 Beijing and six LAM signature SNPs that can be used to classify clinical *M. tuberculosis* isolates. This study also identified lineages specific for F15/LAM4/KZN (4 SNPs) and Beijing BO/W148 (3 SNPs) sub-lineages, which were reported to be more virulent and associated with drug resistance. The lineage-specific SNPs detected in this study were shown to be globally applicable for genotyping the *M. tuberculosis* Beijing and LAM lineage. There were lineage-specific SNPs detected in this study that have not been reported from the previous literatures. Observations from this study also highlight the advantage of using WGS technique over other genotyping methods such as IS6110-RFLP that has more drawbacks, as most genotypic methods discriminate *M. tuberculosis* strains using specific genes or regions in the genome of *M. tuberculosis* while WGS uses the complete genome of *M. tuberculosis* to determine different *M. tuberculosis* lineages. Furthermore, this study suggests that WGS can be used to resolve discordant results and to determine mutations that may lead to the emergence of drug resistance.

Chapter 6

Concluding remarks

6.1. Concluding remarks

Tuberculosis remains a major global concern. Despite the global increase in TB cases reported every year, there are still challenges with the diagnosis and effective treatment of *M. tuberculosis* (WHO, 2019). Classification of *M. tuberculosis* to various lineages is considered as one of the good approaches in diagnosing *M. tuberculosis* to stop global transmission of different strains of *M. tuberculosis*, to identify strains that are more virulent than others and to associate a certain lineage with treatment outcomes. In addition, identifying genetic diversity may help to determine the genetic relatedness of the *M. tuberculosis* strains as a way of resolving epidemiological challenges such as determining sources of infection, monitoring the global distribution of a particular genotype, differentiating of recent transmission from reactivation and tracing the cause of relapse or treatment failure (Dou *et al*., 2017*)*.

Objective 1: To collect susceptible and resistant *M. tuberculosis* from the National Health Laboratory Service, Tshwane Academic Division (NHLS/TAD).

A total of 150 susceptible and resistant *M. tuberculosis* cultures stored in Mycobacteria growth indicator tubes (MGIT) tubes (Becton Dickson and Company, Sparks, USA) were used to conduct this study. These cultures were obtained from sputum specimens collected at various clinics and hospitals in the Tshwane district and were submitted to the NHLS/TAD for diagnostic purposes. The samples were already identified for drug resistance testing (DST) using the BACTEC MGIT 960 system (Becton Dickson and Company, Sparks, USA) and line probe assay (LPA) (Hain Life science, USA) at the NHLS/TAD.

This study observed that the number of susceptible *M. tuberculosis* isolates per lineages was higher than the drug-resistant TB and the Beijing lineage was the predominant genotype with more susceptible isolates than other genotypes. This study also found that the EAI lineage (6.7%), Manu (4%) and S family (3.3%), had more resistant isolates than susceptible ones.

Objective 2: To determine lineages circulating in the Tshwane region using spoligotyping.

Spoligotyping results showed that the Beijing lineage was the most prevalent lineage amongst all *M. tuberculosis* cultures collected at the NHLS/TAD. Similar to most studies done to genotype *M. tuberculosis* in South Africa (Chihota *et al*., 2018; Said *et al*., 2016), the Beijing genotype has been found to be the most predominant lineage. The present study highlighted that despite the low discriminatory power produced by the spoligotyping technique, this method can be used for screening and selecting *M. tuberculosis* lineages, to give insight into the lineages that may be included in the study before proceeding with the WGS technique. Although few lineages may be misclassified while using spoligotyping, over 90% of the lineages are discriminated correctly and that may help to reduce the cost since WGS is considered the most expensive technique.

Objective 3: To determine Beijing and LAM sub-lineages using *IS*6110-RFLP.

The IS6110-RFLP could not give enough results that could be used in this study even after repetition. Some samples gave results that were faint and could not be interpreted while some failed to be transferred from the agarose gel to the membrane. This could have been due to the following limitation:

1. Insufficient DNA used for genotyping, as IS6110-RFLP requires 6 µg genomic DNA
2. Some isolates might have had a low IS6110 copy number of less than six, which is one of the drawbacks of genotyping with IS6110-RFLP.
3. The genomic DNA used could have degraded during the DNA extraction technique.

For these reasons, the Beijing and LAM lineages detected by spoligotyping were not confirmed using IS6110-RFLP and the discriminatory power of SNPs in virulence genes could not be compared to that of IS6110-RFLP.

Objective 4: To detect SNPs within the virulence genes using whole-genome sequencing.

This study established a range of virulence genes that have mutations specific to Beijing and LAM lineages. The virulence genes used to determine the SNPs markers for Beijing and LAM lineages were regulatory and structural genes that are essential for various stages of pathogenesis. The protein groups and virulence genes used were carbohydrate metabolism (*glcB*), cell wall protein (*kefB, mce*4*c*, *mce*1*D*, *mce*3*F*), Cholesterol catabolism (*fadE*28, *ltp*2, and *cyp*125) Lipid and fatty

acid metabolism *(fadD*28, *lipF, mas, pks*5, *pks*7 *plcA* and *plcC*), Metal transporter proteins (*mbtB*), Mycolic acid synthesis (*mmaA*4), Proteins inhibiting antimicrobial responses of the macrophage (*ponA*2), Sigma factor (*sigG*), TA systems (*mazF*3, *mazF*8, *vapB*34, *vapB*47, *vapC*12, *vapC*37, *vapC*38), Type VII secretion system (*eccCb*1, *secA*2) and heat shock proteins (*hsp*22.5).

Objective 5: To develop a catalog of virulence gene SNPs associated with Beijing and LAM lineages and sub-lineages.

This study used virulence genes to evaluate the signature SNP markers that could be used to detect *M. tuberculosis* lineages and to determine if the lineages are associated with the virulence of strains that causes a life-threatening TB disease. Through this approach, the study successfully identified 29 Beijing and 6 LAM signature SNPs that can be used to classify clinical *M. tuberculosis* isolates. This study also identified lineages specific for F15/LAM4/KZN (four SNPs) and Beijing BO/W148 (three SNPs) sub-lineages from the genome that were downloaded from PATRIC. These strains were reported to be more virulent and are associated with drug resistance. Possibly, the mutations detected in this study may also contribute to the rapid transmission of *M. tuberculosis* between immunosuppressed patients (including healthy individuals).

Objective 6: To determine and compare the discriminatory power of SNPs in virulence genes with that of IS6110-RFLP genotyping.

The study could not compare the discriminatory power of the SNP and IS6110-RFLP because there were not enough results obtained when genotyping *M. tuberculosis* using IS6110-RFLP.

6.2. Limitations of this study

The limitations of this study include the following:

- Some isolates were contaminated and were excluded from this study. Contamination could have been increased by using a liquid medium; this study used MGIT and Middlebrook 7H9 broth media.
- results. The Beijing and LAM lineages and sub-lineages obtained were not confirmed using this technique and discordant results between spoligotyping and WGS were not confirmed. Moreover, the detection of SNP signature per lineage was only done using all genomes

with high-quality sequences; only genome sequences with unambiguously defined genotypes and a reasonable number of mutations were used to discriminate the lineages.

- One of our objectives was to find signature SNPs for all LAM sub-lineages. However, due to a limited number of LAM sub-lineages obtained, there were not enough lineage-specific SNPs to discriminate LAM sub-lineages.
- The sample size used for this study was small due to the lack of funds needed to perform WGS. This also contributed to a limited number of SNPs being obtained to discriminate our lineages.

6.3. Future research

Future studies will be done to determine the polymorphisms within new functional subgroups inside other lineages. Lineage-specific SNPs within drug resistance genes of *M. tuberculosis* that are distributed in different regions of South Africa also need to be studied. Finally, a bioinformatics program that will simultaneously determine the *M. tuberculosis* lineage, lineage-specific SNPs and drug resistance mutations should be created to advance *M. tuberculosis* typing, resistance detection and epidemiology analyses.

6.4. Acknowledgement

The authors would like to thank the National Health Laboratory Service, Tshwane Academic Division (NHLS/TAD) for providing the samples used in this study. The National Institute for Communicable Disease (NICD) for providing IS6110-RFLP training and the University of Stellenbosch for providing CTAB training. The SAMRC for funding the project. Ms. Unarine Matodzi would like to thank National Research Funding (NRF) for the financial support received for her studies.

6.5. Conflict of interest

No conflict of interest was declared by this study

References

Abdallah AM, Bestebroer J, Savage ND, De Punder K, Van Zon M, Wilson L, Korbee CJ, Van Der Sar AM, Ottenhoff TH & Van Der Wel NN (2011) Mycobacterial secretion systems ESX-1 and ESX-5 play distinct roles in host cell death and inflammasome activation. *The Journal of Immunology* **187**:4744-4753.

Abdallah AM, Van Pittius NCG, Champion PAD, Cox J, Luirink J, Vandenbroucke-Grauls CM, Appelmelk BJ & Bitter W (2007) Type VII secretion—mycobacteria show the way. *Nature Reviews Microbiology* **5**:883.

Abdallah AM, Verboom T, Weerdenburg EM, Gey Van Pittius NC, Mahasha PW, Jiménez C, Parra M, Cadieux N, Brennan MJ & Appelmelk BJ (2009) PPE and PE_PGRS proteins of *Mycobacterium marinum* are transported via the type VII secretion system ESX-5. *Molecular Microbiology* **73**:329-340.

Agrawal M, Bajaj A, Bhatia V & Dutt S (2016) Comparative study of GeneXpert with ZN stain and culture in samples of suspected pulmonary tuberculosis. *Journal of Clinical and Diagnostic Research* **10**: DC09.

Ahmad S, El-Shazly S, Mustafa A & Al-Attiyah R (2004) Mammalian cell-entry proteins encoded by the mce3 operon of *Mycobacterium tuberculosis* are expressed during natural infection in humans. *Scandinavian Journal of Immunology* **60**:382-391.

Al-Zamel FA (2009) Detection and diagnosis of *Mycobacterium tuberculosis*. *Expert Review of Anti-Infective Therapy* **7(9)**:1099-1108.

Angala SK, Belardinelli JM, Huc-Claustre E, Wheat WH & Jackson M (2014) The cell envelope glycoconjugates of *Mycobacterium tuberculosis*. *Critical Reviews In Biochemistry And Molecular Biology* **49**:361-399.

Asmar S, Chatellier S, Mirande C, Van Belkum A, Canard I, Raoult D & Drancourt M (2015) A novel solid medium for culturing *Mycobacterium tuberculosis* from clinical specimens. *Journal of Clinical Microbiology* **1**: 149-15.

Av-Gay Y & Everett M (2000) The eukaryotic-like Ser/Thr protein kinases of *Mycobacterium tuberculosis*. *Trends In Microbiology* **8**:238-244.

Barnard M, Warren R, Van Pittius NG, Van Helden P, Bosman M, Streicher E, Coetzee G & O'brien R (2012) Genotype MTBDR sl line probe assay shortens time to diagnosis of extensively drug-resistant tuberculosis in a high-throughput diagnostic laboratory. *American Journal Of Respiratory And Critical Care Medicine* **186**:1298-1305.

Barry CE, Crick DC & Mcneil MR (2007) Targeting the formation of the cell wall core of *M. tuberculosis*. *Infectious Disorders-Drug Targets (Formerly Current Drug Targets-Infectious Disorders)* **7**:182-202.

Behar SM & Briken V (2019) Apoptosis inhibition by intracellular bacteria and its consequence on host immunity. *Current Opinion in Immunology* **60**:103-110.

Behar S, Martin C, Booty M, Nishimura T, Zhao X, Gan H, Divangahi M & Remold H (2011) Apoptosis is an innate defense function of macrophages against *Mycobacterium tuberculosis*. *Mucosal Immunology* **4**:279.

Belisle JT, Vissa VD, Sievert T, Takayama K, Brennan PJ & Besra GS (1997) Role of the major antigen *of Mycobacterium tuberculosis* in cell wall biogenesis. *Science* **276**:1420-1422.

Bengtson HN, Homolka S, Niemann S, Reis AJ, Da Silva PE, Gerasimova YV, Kolpashchikov DM & Rohde KH (2017) Multiplex detection of extensively drug-resistant tuberculosis using binary deoxyribozyme sensors. *Biosensors and Bioelectronics* **94**: 176-183.

Bespyatykh J, Shitikov E, Guliaev A, Smolyakov A, Klimina K, Veselovsky V, Malakhova M, Arapidi G, Dogonadze M & Manicheva O (2019) System OMICs analysis of *Mycobacterium tuberculosis* Beijing B0/W148 cluster. *Scientific Reports* **9**:1-11.

Bhanu NV, Van Soolingen D, Van Embden J, Dar L, Pandey R & Seth P (2002) Predominance of a novel *Mycobacterium tuberculosis* genotype in the Delhi region of India. *Tuberculosis* **82**:105-112.

Bigi F, Alito A, Romano M, Zumarraga M, Caimi K & Cataldi A (2000) The gene encoding P27 lipoprotein and a putative antibiotic-resistance gene form an operon in *Mycobacterium tuberculosis* and *Mycobacterium bovis*. *Microbiology* **146**:1011-1018.

Blomgran R, Desvignes L, Briken V & Ernst JD (2012) *Mycobacterium tuberculosis* inhibits neutrophil apoptosis, leading to delayed activation of naive CD4 T cells. Cell host & microbe **11**:81-90.

Blouin Y, Hauck Y, Soler C, Fabre M, Vong R, Dehan C, Cazajous G, Massoure P-L, Kraemer P & Jenkins A (2012) Significance of the identification in the Horn of Africa of an exceptionally deep branching *Mycobacterium tuberculosis* clade. *PloS One* **7**:e52841.

Braunstein M, Espinosa BJ, Chan J, Belisle JT & R. Jacobs Jr W (2003) SecA2 functions in the secretion of superoxide dismutase A and in the virulence of *Mycobacterium tuberculosis*. *Molecular Microbiology* **48**:453-464.

Brown GD, Dave JA, Van Pittius NCG, Stevens L, Ehlers MR & Beyers AD (2000) The mycosins of *Mycobacterium tuberculosis* H37Rv: a family of subtilisin-like serine proteases. *Gene* **254**:147-155.

Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, Al-Hajoj SA, Allix C, Aristimuño L, Arora J & Baumanis V (2006) *Mycobacterium tuberculosis* complex genetic diversity: mining the

fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiology* **6**:23.

Cambau E & Drancourt M (2014) Steps towards the discovery of *Mycobacterium tuberculosis* by Robert Koch, 1882. *Clinical Microbiology and Infection* **20**:196-201.

Carding S, Verbeke K, Vipond DT, Corfe BM & Owen LJ (2015) Dysbiosis of the gut microbiota in disease. *Microbial Ecology in Health and Disease* **26**:26191.

Cavusoglu C, Hilmioglu S, Guneri S & Bilgic A (2002) Characterization of rpoB mutations in rifampin-resistant clinical isolates of *Mycobacterium tuberculosis* from Turkey by DNA sequencing and line probe assay. *Journal Of Clinical Microbiology* **40**:4435-4438.

Chatrath S, Gupta VK, Dixit A & Garg LC (2016) PE_PGRS30 of *Mycobacterium tuberculosis* mediates suppression of proinflammatory immune response in macrophages through its PGRS and PE domains. *Microbes and Infection* **18**:536-542.

Chen Y, Yuan Z, Shen X, Wu J, Wu Z & Xu B (2016) Resistance to second-line antituberculosis drugs and delay in drug susceptibility testing among multidrug-resistant tuberculosis patients in Shanghai. *BioMed Research International*.

Chihota VN, Niehaus A, Streicher EM, Wang X, Sampson SL, Mason P, Källenius G, Mfinanga SG, Pillay M & Klopper M (2018) Geospatial distribution of *Mycobacterium tuberculosis* genotypes in Africa. *PloS One* 13:

Cobat A, Orlova M, Barrera L & Schurr E (2013) Host genomics and control of tuberculosis infection. *Public Health Genomics* **16**:44-49.

Cooper AM (2009) Cell-mediated immune responses in tuberculosis. *Annual Review Of Immunology* **27**:393-422.

Coscolla M & Gagneux S (2014) Consequences of genomic diversity in *Mycobacterium tuberculosis*. Seminars in Immunology. *Elsevier* **15**: 431-444.

Crowley LV 2013. An introduction to human disease: Pathology and pathophysiology correlations. *Jones & Bartlett Publishers.*

Danelishvili L, Yamazaki Y, Selker J & Bermudez LE (2010) Secreted *Mycobacterium tuberculosis* Rv3654c and Rv3655c proteins participate in the suppression of macrophage apoptosis. *PloS One* **5**:e10474.

Dave JA, Van Pittius NCG, Beyers AD, Ehlers MR & Brown GD (2002) Mycosin-1, a subtilisin-like serine protease of *Mycobacterium tuberculosis*, is cell wall-associated and expressed during infection of macrophages. *BMC Microbiology* **2**:30.

De Maio F, Battah B, Palmieri V, Petrone L, Corrente F, Salustri A, Palucci I, Bellesi S, Papi M & Rubino S (2018) PE_PGRS3 of *Mycobacterium tuberculosis* is specifically expressed at low phosphate concentration, and its arginine-rich C-terminal domain mediates adhesion and persistence in host tissues when expressed in *Mycobacterium smegmatis*. *Cellular Microbiology* **20**:e12952.

Desikan P (2013) Sputum smear microscopy in tuberculosis: is it still relevant? *The Indian Journal of Medical Research* 137:442

Diaz G, Wolfe LM, Kruh-Garcia NA & Dobos KM (2016) Changes in the membrane-associated proteins of exosomes released from human macrophages after *Mycobacterium tuberculosis* infection. *Scientific reports* **6**:37975.

Dimareli-Malli Z & Sarris K (2001) Comparison of DNA probe test and cultivation methods for detection of *Mycobacterium avium* subsp paratuberculosis in caprine and ovine faeces. *Australian Veterinary Journal* **79**: 47-50.

Dou H-Y, Lin C-H, Chen Y-Y, Yang S-J, Chang J-R, Wu K-M, Chen Y-T, Chin P-J, Liu Y-M & Su I-J (2017) Lineage-specific SNPs for genotyping of *Mycobacterium tuberculosis* clinical isolates. *Scientific Reports* **7**:1-9

Duchêne V, Ferdinand S, Filliol I, Guégan JF, Rastogi N & Sola C (2004) Phylogenetic reconstruction of *Mycobacterium tuberculosis* within four settings of the Caribbean region: tree comparative analyse and first appraisal on their phylogeography. *Infection, Genetics and Evolution* **4**:5-14.

Fang Z, Schubert W-D & Van Pittius NCG (2016) Expression and production of soluble *Mycobacterium tuberculosis* H37Rv mycosin-3. *Biochemistry and Biophysics Reports* **5**:448-452.
Feltcher ME, Sullivan JT & Braunstein M (2010) Protein export systems of *Mycobacterium tuberculosis*: novel targets for drug development? *Future Microbiology* **5**:1581-1597.

Ferrari CK, Souto PC, França EL & Honorio-França AC (2011) Oxidative and nitrosative stress on phagocytes' function: from effective defense to immunity evasion mechanisms. *Archivum Immunologiae et Therapiae Experimentalis* **59**:441-448.

Floyd K, Glaziou P, Zumla A & Raviglione M (2018) The global tuberculosis epidemic and progress in care, prevention, and research: an overview in year 3 of the End TB era. *The Lancet Respiratory Medicine* **6**:299-314.

Forrellad MA, Klepp LI, Gioffré A, Sabio Y Garcia J, Morbidoni HR, Santangelo MdlP, Cataldi AA & Bigi F (2013) Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence* **4**:3-66.

Gagneux S 2013. Genetic diversity in *Mycobacterium tuberculosis*. Pathogenesis of *Mycobacterium tuberculosis* and its Interaction with the Host Organism. *Springer*.

Gagneux S, Deriemer K, Van T, Kato-Maeda M, De Jong BC, Narayanan S, Nicol M, Niemann S, Kremer K & Gutierrez MC (2006) Variable host–pathogen compatibility in *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences* **103**:2869-2873.

Gandhi NR, Moll A, Sturm AW, Pawinski R, Govender T, Lalloo U, Zeller K, Andrews J & Friedland G (2006) Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. The Lancet 368:1575-1580.

Gardee Y, Dreyer A, Koornhof H, Omar S, Da Silva P, Bhyat Z & Ismail NA (2017) Evaluation of the GenoType MTBDRsl version 2.0 assay for second-line drug resistance detection of *Mycobacterium tuberculosis* isolates in South Africa. *Journal of Clinical Microbiology* **55**:791-800.

Garg LC, Chatrath S, Gupta VK & Dixit A (2016) PE_PGRS30 of *Mycobacterium tuberculosis* mediates suppression of proinflammatory immune response in macrophages through its PGRS and PE domains.

Gaude GS, Hattiholli J & Kumar P (2014) Risk factors and drug-resistance patterns among pulmonary tuberculosis patients in northern Karnataka region, India. Nigerian medical journal: *Journal of the Nigeria Medical Association* **55**:327.

Gordon SV & Parish T (2018) Microbe profile: *Mycobacterium tuberculosis*: humanity's deadly microbial foe. *Microbiology* **164**:437-439

Gupta R, Beg Q & Lorenz P (2002) Bacterial alkaline proteases: molecular approaches and industrial applications. *Applied Microbiology and Biotechnology* **59**:15-32.

Gutierrez MC, Brisse S, Brosch R, Fabre M, Omaïs B, Marmiesse M, Supply P & Vincent V (2005) Ancient origin and gene mosaicism of the progenitor *of Mycobacterium tuberculosis*. *PLoS Pathogens* **1**:e5.

Heersma HF, Kremer K & Van Embden JD 1998. Computer analysis of IS6110 RFLP patterns of *Mycobacterium tuberculosis*. Mycobacteria protocols. *Springer*.

Hedstrom L (2002) Serine protease mechanism and specificity. *Chemical reviews*. **102**:4501-4524.

Helb D, Jones M, Story E, Boehme C, Wallace E, Ho K, Kop J, Owens MR, Rodgers R & Banada P (2010) Rapid detection of *Mycobacterium tuberculosis* and rifampin resistance by use of on-demand, near-patient technology. *Journal of Clinical Microbiology* **48**:229-237.

Held M, Laubscher M, Zar H & Dunn R (2014) GeneXpert polymerase chain reaction for spinal tuberculosis: an accurate and rapid diagnostic test. *The Bone & Joint Journal* **96**:1366-1369.

Hillemann D, Rüsch-Gerdes S, Boehme C & Richter E (2011) Rapid molecular detection of extrapulmonary tuberculosis by the automated GeneXpert MTB/RIF system. *Journal of Clinical Microbiology* **49**:1202-1205.

Hines N, Payeur JB & Hoffman LJ (2006) Comparison of the recovery of *Mycobacterium bovis* isolates using the BACTEC MGIT 960 system, BACTEC 460 system, and Middlebrook 7H10 and 7H11 solid media. *Journal of Veterinary Diagnostic Investigation* **18**: 243-250.

Hoagland DT, Liu J, Lee RB & Lee RE (2016) New agents for the treatment of drug-resistant *Mycobacterium tuberculosis*. *Advanced Drug Delivery Reviews* **102**: 55-72.

Hoffmann C, Leis A, Niederweis M, Plitzko JM & Engelhardt H (2008) Disclosure of the mycobacterial outer membrane: cryo-electron tomography and vitreous sections reveal the lipid bilayer structure. *Proceedings of the National Academy of Sciences* **105**:3963-3967.

Horsburgh Jr CR, Barry Iii CE & Lange C (2015) Treatment of tuberculosis. *New England Journal of Medicine* **373**:2149-2160.

Houben EN, Korotkov KV & Bitter W (2014) Take five—Type VII secretion systems of Mycobacteria. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1843**:1707-1716.

Hu G & Leger RS (2004) A phylogenomic approach to reconstructing the diversification of serine proteases in fungi. *Journal of evolutionary biology* **17**:1204-1214.

Ioerger TR, Koo S, No E-G, Chen X, Larsen MH, Jacobs Jr WR, Pillay M, Sturm AW & Sacchettini JC (2009) Genome analysis of multi-and extensively-drug-resistant tuberculosis from KwaZulu-Natal, South Africa. *PloS One* **4**: e7778.

Ismail NA, Mvusi L, Nanoo A, Dreyer A, Omar SV, Babatunde S, Molebatsi T, Van Der Walt M, Adelekan A & Deyde V (2018) Prevalence of drug-resistant tuberculosis and imputed burden in South Africa: a national and sub-national cross-sectional survey. *The Lancet Infectious Diseases* **18**:779-787.

Jackson M, Mcneil MR & Brennan PJ (2013) Progress in targeting cell envelope biogenesis in *Mycobacterium tuberculosis*. *Future Microbiology* **8**:855-875.

Jagielski T, Van Ingen J, Rastogi N, Dziadek J, Mazur PK & Bielecki J (2014) Current methods in the molecular typing of *Mycobacterium tuberculosis* and other mycobacteria. *BioMed Research International* **2014**.

Jeon S, Lim N, Park S, Park M & Kim S (2018) Comparison of PFGE, IS6110-RFLP, and 24-Locus MIRU-VNTR for molecular epidemiologic typing of *Mycobacterium tuberculosis* isolates with known epidemic connections. *Journal of Microbiology and Biotechnology* **28**:338-346.

Johansen P, Fettelschoss A, Amstutz B, Selchow P, Waeckerle-Men Y, Keller P, Deretic V, Held L, Kündig TM & Böttger EC (2011) Relief from Zmp1-mediated arrest of phagosome maturation is associated with facilitated presentation and enhanced immunogenicity of mycobacterial antigens. *Clin. Vaccine Immunol.* **18**:907-913.

Kamerbeek J, Schouls L, Kolk A, Van Agterveld M, Van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R & Goyal M (1997) Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *Journal of Clinical Microbiology* **35**:907-914.

Klopper M, Warren RM, Hayes C, Van Pittius NCG, Streicher EM, Müller B, Sirgel FA, Chabula-Nxiweni M, Hoosain E & Coetzee G (2013) Emergence and spread of extensively and totally drug-resistant tuberculosis, South Africa. *Emerging Infectious Diseases* **19**:449.

Köser CU, Bryant JM, Parkhill J & Peacock SJ (2013) Consequences of whiB7 (Rv3197A) mutations in Beijing genotype isolates of the *Mycobacterium tuberculosis* complex. *Antimicrobial agents and chemotherapy* **57**:3461-3461.

Koul A, Choidas A, Tyagi AK, Drlica K, Singh Y & Ullrich A (2001) Serine/threonine protein kinases PknF and PknG of *Mycobacterium tuberculosis:* characterization and localization. *Microbiology* **147**:2307-2314.

Kremer L, Maughan W, Wilson R, Dover L & Besra G (2002) The *M. tuberculosis* antigen 85 complex and mycolyltransferase activity. *Letters in Applied Microbiology* **34**:233-237.

Kumar K, Sagar BC, Giribhattanavar P & Patil SA (2019) Ultrastructural analysis of cell wall of drug resistant and sensitive *Mycobacterium tuberculosis* isolated from cerebrospinal fluid by transmission electron microscope. *Microscopy Research and Technique* **82**:122-127.

Lamb CA, Yoshimori T & Tooze SA (2013) The autophagosome: origins unknown, biogenesis complex. *Nature Reviews Molecular Cell Biology* **14**:759.

Lin PL, Ford CB, Coleman MT, Myers AJ, Gawande R, Ioerger T, Sacchettini J, Fortune SM & Flynn JL (2014) Sterilization of granulomas is common in active and latent tuberculosis despite within-host variability in bacterial killing. *Nature Medicine* **20**:75.

Liu Q, Li G-L, Chen C, Wang J-M, Martinez L, Lu W & Zhu L-M (2017) Diagnostic performance of the GenoType MTBDRplus and MTBDRsl assays to identify tuberculosis drug resistance in Eastern China. *Chinese Medical Journal* **130**:1521.

Madacki J, Fiol GM & Brosch R (2019) Update on the virulence factors of the obligate pathogen *Mycobacterium* tuberculosis and related tuberculosis-causing mycobacteria. *Infection, Genetics and Evolution* **72**:67-77.

Maguga-Phasha N, Munyai N, Mashinya F, Makgatho M & Mbajiorgu E (2017) Genetic diversity and distribution of *Mycobacterium tuberculosis* genotypes in Limpopo, South Africa. *BMC Infectious Diseases* **17**: 764.

Makinoshima H & Glickman MS (2006) Site-2 proteases in prokaryotes: regulated intramembrane proteolysis expands to microbial pathogenesis. *Microbes and Infection* **8**:1882-1888.

Mäkinen J, Marttila HJ, Marjamäki M, Viljanen MK & Soini H (2006) Comparison of two commercially available DNA line probe assays for detection of multidrug-resistant *Mycobacterium tuberculosis*. *Journal of Clinical Microbiology* **44**:350-352.

Marais BJ (2017) Improving access to tuberculosis preventive therapy and treatment for children. *International Journal of Infectious Diseases* **56**:122-125.

McHugh TD, Batt SL, Shorten RJ, Gosling RD, Uiso L & Gillespie SH (2005) *Mycobacterium tuberculosis* lineage: a naming of the parts. Tuberculosis. **85**:127-136.

Menzies D, Adjobimey M, Ruslami R, Trajman A, Sow O, Kim H, Obeng Baah J, Marks GB, Long R & Hoeppner V (2018) Four months of rifampin or nine months of isoniazid for latent tuberculosis in adults. *New England Journal of Medicine* 3**79**:440-453.

Mikheecheva NE, Zaychikova MV, Melerzanov AV & Danilenko VN (2017) A nonsynonymous SNP catalog of *Mycobacterium tuberculosis* virulence genes and its use for detecting new potentially virulent sublineages. *Genome Biology and Evolution* **9**: 887-899.

Miller JL, Velmurugan K, Cowan MJ & Briken V (2010) The type I NADH dehydrogenase of *Mycobacterium tuberculosis* counters phagosomal NOX2 activity to inhibit TNF-α-mediated host cell apoptosis. *PLoS Pathogens* **6**:e1000864.

Mori M, Deodato D, Kasula M, Ferraris DM, Sanna A, De Logu A, Rizzi M & Botta M (2018) Design, synthesis, SAR and biological investigation of 3-(carboxymethyl) rhodanine and aminothiazole inhibitors of *Mycobacterium tuberculosis* Zmp1. *Bioorganic & Medicinal Chemistry Letters* **28**:637-641.

Mori M, Moraca F, Deodato D, Ferraris DM, Selchow P, Sander P, Rizzi M & Botta M (2014) Discovery of the first potent and selective *Mycobacterium tuberculosis* Zmp1 inhibitor. *Bioorganic & Medicinal Chemistry Letters* **24**:2508-2511.

Moure R, Muñoz L, Torres M, Santin M, Martín R & Alcaide F (2011) Rapid detection of *Mycobacterium tuberculosis* complex and rifampin resistance in smear-negative clinical samples by use of an integrated real-time PCR method. *Journal of Clinical Microbiology* **49**:1137-1139.

Mukhopadhyay S, Panda PK, Sinha N, Das DN & Bhutia SK (2014) Autophagy and apoptosis: where do they meet? Apoptosis **19**:555-566.

Mustazzolu A, Piersimoni C, Iacobino A, Giannoni F, Chirullo B & Fattorini L (2019) Revisiting problems and solutions to decrease *Mycobacterium tuberculosis* pyrazinamide false resistance when using the Bactec MGIT 960 system. *Annali dell'Istituto superiore di sanita* **55**:51-54.

Nguyen L, Walburger A, Houben E, Koul A, Muller S, Morbitzer M, Klebl B, Ferrari G & Pieters J (2005) Role of protein kinase G in growth and glutamine metabolism of *Mycobacterium bovis* BCG. Journal of bacteriology **187**:5852-585.

Niederweis M (2008) Nutrient acquisition by mycobacteria. *Microbiology* **154**:679-692.

Omar SV, Joseph L, Said HM, Ismail F, Ismail N, Gwala TL & Ismail NA (2019) Whole genome sequencing for drug resistance determination in *Mycobacterium tuberculosis*. *African journal of laboratory medicine* **8**:1-5.

Omari A (2014) Characteristics of Isolates of *Mycobacterium Tuberculosis* in Extrapulmonary Tuberculosis in Korle-Bu Teaching Hospital. Doctoral Dissertation University of Ghana.

O'neill MB, Mortimer TD & Pepperell CS (2015) Diversity of *Mycobacterium tuberculosis* across evolutionary scales. *PLoS Pathogens* **11**: e1005257.

Osei Sekyere J, Reta MA, Maningi NE & Fourie PB (2019) Antibiotic resistance of Mycobacterium tuberculosis complex in Africa: A systematic review of current reports of molecular epidemiology, mechanisms and diagnostics. *Journal of Infection* **79**:550-571

Osei Sekyere J, Maphalala N, Malinga LA, Mbelle NM & Maningi NE (2019) A Comparative Evaluation of the New Genexpert MTB/RIF Ultra and other Rapid Diagnostic Assays for Detecting Tuberculosis in Pulmonary and Extra Pulmonary Specimens. *Scientific Reports* **9:**1-9.

Palomino J & Martin A (2014) Drug resistance mechanisms in *Mycobacterium tuberculosis*. *Antibiotics* **3**:317-340.

Peberdy SA (2019) Changing Geographies of Immigration to South Africa. *The Geography of South Africa.* Springer.

Perez-Risco D, Rodriguez-Temporal D, Valledor-Sanchez I & Alcaide F (2018) Evaluation of the Xpert MTB/RIF Ultra assay for direct detection of *Mycobacterium tuberculosis* complex in smear-negative extrapulmonary samples. *Journal of clinical microbiology* **56**:e00659-18.

Pieters J (2008) *Mycobacterium tuberculosis* and the macrophage: maintaining a balance. *Cell Host & Microbe* **3**:399-407.

Poirier V, Bach H & Av-Gay Y (2014) *Mycobacterium tuberculosis* promotes anti-apoptotic activity of the macrophage by PtpA protein-dependent dephosphorylation of host GSK3α. *Journal of Biological Chemistry* **289**:29376-29385.

Prisic S, Dankwa S, Schwartz D, Chou MF, Locasale JW, Kang C-M, Bemis G, Church GM, Steen H & Husson RN (2010) Extensive phosphorylation with overlapping specificity by *Mycobacterium tuberculosis* serine/threonine protein kinases. *Proceedings of the National Academy of Sciences* **107**:7521-7526.

Ramos DF, Tavares L, Silva PEAd & Dellagostin OA (2014) Molecular typing of *Mycobacterium bovis* isolates: a review. *Brazilian Journal of Microbiology* **45**: 365-372.

Reed MB, Gagneux S, Deriemer K, Small PM & Barry CE (2007) The W-Beijing lineage of *Mycobacterium tuberculosis* overproduces triglycerides and has the DosR dormancy regulon constitutively upregulated. *Journal of bacteriology* **189**:2583-2589.

Reed MB, Pichler VK, Mcintosh F, Mattia A, Fallow A, Masala S, Domenech P, Zwerling A, Thibert L & Menzies D (2009) Major *Mycobacterium tuberculosis* lineages associated with patient country of origin. *Journal of Clinical Microbiology* **47**: 1119-1128.

Ribeiro-Guimarães ML & Pessolani MCV (2007) Comparative genomics of mycobacterial proteases. *Microbial pathogenesis* **43**:173-178.

Roychowdhury T, Mandal S & Bhattacharya A (2015) Analysis of IS6110 insertion sites provide a glimpse into genome evolution of *Mycobacterium tuberculosis*. *Scientific Reports* **5**:12567.

Ryu YJ (2015) Diagnosis of pulmonary tuberculosis: recent advances and diagnostic algorithms. *Tuberculosis and Respiratory Diseases* **78**: 64-71.

Said HM, Kock MM, Ismail NA, Mphahlele M, Baba K, Omar SV, Osman AG, Hoosen AA & Ehlers MM (2012) Molecular characterization and second-line antituberculosis drug resistance patterns of multidrug-resistant *Mycobacterium tuberculosis* isolates from the northern region of South Africa. *Journal of Clinical Microbiology* **50**:2857-2862.

Said HM, Kushner N, Omar SV, Dreyer AW, Koornhof H, Erasmus L, Gardee Y, Rukasha I, Shashkina E & Beylis N (2016) A novel molecular strategy for surveillance of multidrug-resistant tuberculosis in high burden settings. *PloS One* **11**: e0146106.

Sajid A, Arora G, Gupta M, Upadhyay S, Nandicoori VK & Singh Y (2011) Phosphorylation of *Mycobacterium tuberculosis* ser/thr phosphatase by PknA and PknB. *PloS One* **6**:e17871.

Scherr N, Honnappa S, Kunz G, Mueller P, Jayachandran R, Winkler F, Pieters J & Steinmetz MO (2007) Structural basis for the specific inhibition of protein kinase G, a virulence factor of *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences* **104**:12151-12156.

Schneider JS, Reddy SP, Evans HW & Glickman MS (2013) Site-2 protease substrate specificity and coupling in trans by a PDZ-substrate adapter protein. *Proceedings of the National Academy of Sciences* **110**:19543-19548.

Schneider JS, Sklar JG & Glickman MS (2014) The Rip1 protease of *Mycobacterium tuberculosis* controls the SigD regulon. *Journal of Bacteriology* **196**:2638-2645.

Schnippel K, Ndjeka N, Maartens G, Meintjes G, Master I, Ismail N, Hughes J, Ferreira H, Padanilam X & Romero R (2018) Effect of bedaquiline on mortality in South African patients with drug-resistant tuberculosis: a retrospective cohort study. *The Lancet Respiratory Medicine* **6**:699-706.

Shitikov EA, Bespyatykh JA, Ischenko DS, Alexeev DG, Karpova IY, Kostryukova ES, Isaeva YD, Nosova EY, Mokrousov IV & Vyazovaya AA (2014) Unusual large-scale chromosomal rearrangements in *Mycobacterium tuberculosis* Beijing B0/W148 cluster isolates. *PLoS On*e **9**:

Singh J.A, Upshur R. and Padayatchi N (2007) XDR-TB in South Africa: no time for denial or complacency. *PLoS Medicine* **4**: e50.

Smith I (2003) *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. *Clinical Microbiology Reviews* **16**: 463-496.

Sola C, Filliol I, Legrand E, Lesjean S, Locht C, Supply P & Rastogi N (2003) Genotyping of the *Mycobacterium tuberculosis* complex using MIRUs: association with VNTR and spoligotyping for molecular epidemiology and evolutionary genetics. *Infection, Genetics and Evolution* **3**: 125-133.

Stavrum R, Mphahlele M, Øvreås K, Muthivhi T, Fourie PB, Weyer K & Grewal HM (2009) High diversity of *Mycobacterium tuberculosis* genotypes in South Africa and preponderance of mixed infections among ST53 isolates. *Journal of Clinical Microbiology* **47**: 1848-1856.

Steiger J, Stephan A, Inkeles MS, Realegeno S, Bruns H, Kröll P, De Castro Kroner J, Sommer A, Batinica M & Pitzler L (2016) Imatinib triggers phagolysosome acidification and antimicrobial activity against *Mycobacterium bovis* bacille Calmette–Guérin in glucocorticoid-treated human macrophages. *The Journal of Immunology* **197**:222-232.

Sullivan JT, Young EF, Mccann JR & Braunstein M (2012) The *Mycobacterium tuberculosis* SecA2 system subverts phagosome maturation to promote growth in macrophages. *Infection and Immunity* **80**:996-1006.

Sun J, Wang X, Lau A, Liao T-YA, Bucci C & Hmama Z (2010) Mycobacterial nucleoside diphosphate kinase blocks phagosome maturation in murine RAW 264.7 macrophages. *PloS One* **5:**e8769.

Suzana S, Shanmugam S, Uma Devi K, Swarna Latha P & Michael JS (2017) Spoligotyping of *Mycobacterium tuberculosis* isolates at a tertiary care hospital in India. *Tropical Medicine & International Health* **22**:703-707.

Sypabekova M, Bekmurzayeva A, Wang R, Li Y, Nogues C & Kanayeva D (2017) Selection, characterization, and application of DNA aptamers for detection of *Mycobacterium tuberculosis* secreted protein MPT64. *Tuberculosis* **104**: 70-78.

Tobian AA, Potter NS, Ramachandra L, Pai RK, Convery M, Boom WH & Harding CV (2003) Alternate class I MHC antigen processing is inhibited by Toll-like receptor signaling pathogen-associated molecular patterns: *Mycobacterium tuberculosis* 19-kDa lipoprotein, CpG DNA, and lipopolysaccharide. **171**:1413-1422.

Trinh T, Hoang NN, Srinivasan V, Hoang H, Do T, Vu H, Tran D, Ashton P, Nguyen H & Nguyen P (2018) Virulence of *Mycobacterium tuberculosis* clinical isolates is associated with sputum pre-treatment bacterial load, lineage, survival in macrophages, and cytokine response. *Frontiers in Cellular and Infection Microbiology* **8**:417.

Turnpenny PD & Ellard S (2016) Emery's Elements of Medical Genetics E-Book. *Elsevier Health Sciences* **43**: 153-170

Van Der Wel N, Hava D, Houben D, Fluitsma D, Van Zon M, Pierson J, Brenner M & Peters PJ (2007) *M. tuberculosis and M. leprae* translocate from the phagolysosome to the cytosol in myeloid cells. *Cell* **129**:1287-1298.

Velayati AA and Farnia P (2012) Morphological characterization of *Mycobacterium tuberculosis*. Understanding tuberculosis-Deciphering the secret life of the bacilli. *IntechOpen*.

Veyron-Churlet R, Molle V, Taylor RC, Brown AK, Besra GS, Zanella-Cléon I, Fütterer K & Kremer L (2009) The *Mycobacterium tuberculosis* β-ketoacyl-acyl carrier protein synthase III

activity is inhibited by phosphorylation on a single threonine residue. *Journal of Biological Chemistry* **284**:6414-6424.

Voss G, Casimiro D, Neyrolles O, Williams A, Kaufmann SH, Mcshane H, Hatherill M & Fletcher HA (2018) Progress and challenges in TB vaccine development. *F1000Research* **7**: 152-198.
Warner D.F, Koch, A. and Mizrahi, V (2015) Diversity and disease pathogenesis in *Mycobacterium tuberculosis*. *Trends in microbiology* **23**:14-21.

Weniger T, Krawczyk J, Supply P, Niemann S & Harmsen D (2010) MIRU-VNTR plus: a web tool for polyphasic genotyping of *Mycobacterium tuberculosis* complex bacteria. *Nucleic Acids Research* **38**: 326-331.

World Health Organization (2019) Global Tuberculosis Report. Available at http://apps.who.int/iris.

Wong D, Bach H, Sun J, Hmama Z & Av-Gay Y (2011) *Mycobacterium tuberculosis* protein tyrosine phosphatase (PtpA) excludes host vacuolar-H+–ATPase to inhibit phagosome acidification. *Proceedings of the National Academy of Sciences* **108**:19371-19376.

Yimer SA, Norheim G, Namouchi A, Zegeye ED, Kinander W, Tønjum T, Bekele S, Mannsåker T, Bjune G & Aseffa A (2015) *Mycobacterium tuberculosis* lineage 7 strains are associated with prolonged patient delay in seeking treatment for pulmonary tuberculosis in Amhara Region, Ethiopia. *Journal of Clinical Microbiology* **53**:1301-1309.

Zaychikova MV, Zakharevich NV, Sagaidak MO, Bogolubova NA, Smirnova TG, Andreevskaya SN, Larionova EE, Alekseeva MG, Chernousova LN & Danilenko VN (2015) *Mycobacterium tuberculosis* type II toxin-antitoxin systems: genetic polymorphisms and functional properties and the possibility of their use for genotyping. *PloS One* **10**: e0143682.

Vazquez CL, Bianco MV, Blanco FC, Forrellad MA, Gutierrez MG & Bigi F (2017) *Mycobacterium bovis* requires P27 (LprG) to arrest phagosome maturation and replicate within bovine macrophages. *Infection and Immunity* **85**:e00720-16.

Zheng Q, Li Z, Zhou S, Zhang Q, Zhou L, Fu X, Yang L, Ma Y & Hao X (2017) Heparin-binding hemagglutinin of *Mycobacterium tuberculosis* is an inhibitor of autophagy. *Frontiers in Cellular and Infection Microbiology* **7**:.33.

Appendix A

Preparation of reagents and Buffers used in experiment procedures

100 mL 10 mg/mL lysozyme

Add 10 mg lysozyme into 1 mL to reconstitute the lyophilized lysozyme

1 L of 0.5 M (pH 8.0) EDTA stock solution

0.5 M EDTA = C × V × Mr

= 0.5 M ×1 L × 372.24 g/mol

= 186.1 g

∴ To make 0.5 M EDTA add:

EDTA disodium salt (Sigma-Aldrich, St. Louis, USA)	186.1 g
Sodium hydroxide (NaOH) pellets (Sigma-Aldrich, St. Louis, USA)	~20g
Distilled water (dH2O)	800mL

1 L of 5 × Tris-EDTA (TE) buffer

Prepare: 1 L of 1 M Tris-HCl buffer (pH ~7.5) stock solution

Dissolve 121 g of tris base (Sigma-Aldrich, St. Louis, USA) in 800 mL. Adjust the pH using HCl (Sigma-Aldrich, St. Louis, USA) to pH 7.5. Then fill to 1 L with dH2O.

∴ To make TE buffer Add:

1 M Tris-HCl	10 mL
0.5 M EDTA	2 mL
dH2O	988 mL

100 mL sodium chloride (NaCl)

NaCl (Sigma-Aldrich, St. Louis, USA)	29.2 g
dH2O	100 mL

100 mL Hexadecyltrimethylammonium bromide (CTAB)/NaCl

NaCl (Sigma-Aldrich, St. Louis, USA)	4.1 g
dH2O	80 mL
CTAB (Sigma-Aldrich, St. Louis, USA)	10 g

NB: dissolve 4.1 g NaCl in 80 mL of water and then slowly add 10 g of CTAB while stirring. Then, head the solution in a 65ºC incubator

1 L 70% Ethanol

Use ratio 7:3 to dilute absolute ethanol with distilled water

100% Ethanol (Sigma-Aldrich, St. Louis, USA)	700 mL
dH2O	300 mL

100 mL of 3 M sodium acetate (pH 5.2)

Sodium acetate (Sigma-Aldrich, St. Louis, USA) $= C \times V \times Mr$

$= 3mol/\ dm3 \times 0.1dm3 \times 82.0337$

$= 24.61\ g$

NB. Dissolve 24.61 g Sodium acetate (Sigma-Aldrich, St. Louis, USA) into 40 mL dH2O. Adjust pH to 5.2 using acetic acid and fill to 100 mL with dH2O.

1 L of 2×saline-sodium phosphate-EDTA (SSPE)

20×SSPE (Sigma-Aldrich, St. Louis, USA)	100 mL
De-mineralized water	900 mL

1 L of 2×SSPE/0.5%SDS

20×SSPE (Sigma-Aldrich, St. Louis, USA)	100 mL
10%SDS (Sigma-Aldrich, St. Louis, USA)	50 mL
De-mineralized water	850 mL

1 L of 2×SSPE/0.1%SDS

20×SSPE (Sigma-Aldrich, St. Louis, USA)	100 mL
10%SDS (Sigma-Aldrich, St. Louis, USA)	10 mL
De-mineralized water	900 mL

Middlebrook 7H9 broth base medium

Middlebrook 7H9 powder (Sigma-Aldrich, St. Louis, USA)	2.35 g
dH2O	450 mL
Glycerol (Sigma-Aldrich, St. Louis, USA)	1 mL

Sterilize by autoclaving at 15 lbs pressure (121ºC) for 10 minutes and cool to room temperature and add 50 mL OADC growth supplement.

Appendix B

Data and detailed results

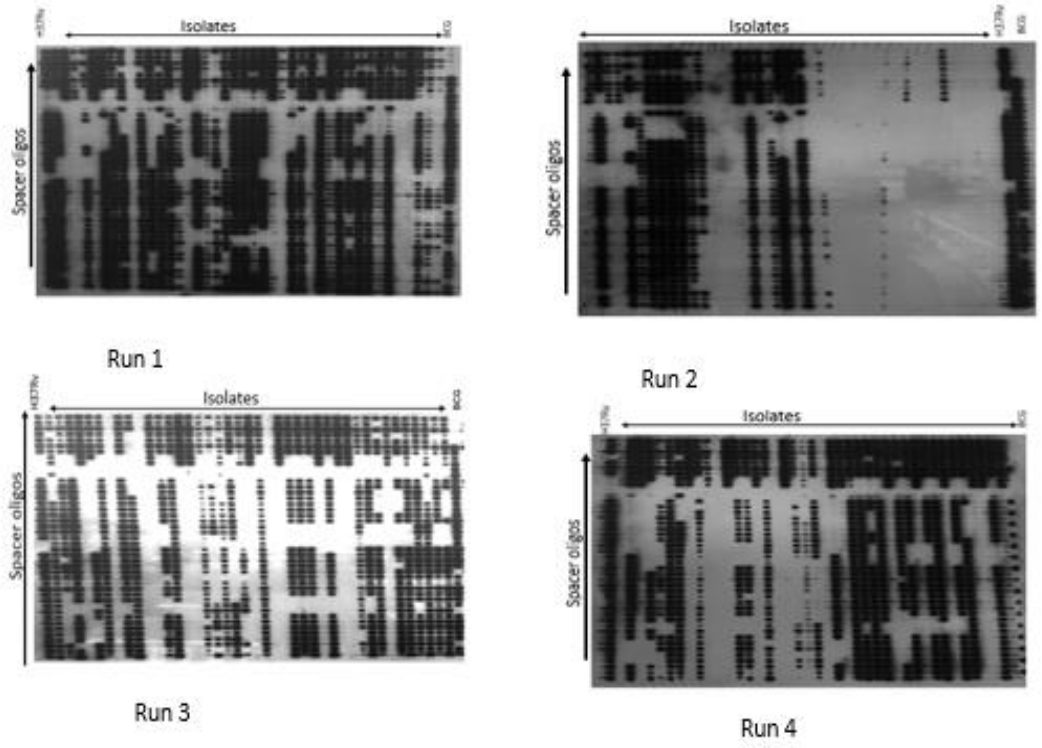

Figure 1. Resultant spoligotyping patterns in a form of black spots on X-ray film. Four spoligotyping runs were conducted wherein 42 samples and 3 controls (H37Rv, BCG as positive control and 2xSSPE/0.1% SDS as negative control) were run on the spoligo-membrane with 43 spacer oligotypes hybridized on it.

Table 1: A catalog of virulence genes used to detect *M. tuberculosis* Beijing and LAM lineages in the Tshwane region.

Protein groups	**Genes**
Carbohydrate metabolism	*glcB* (Rv1837c)
Cell wall protein	*kefB* (Rv3236c), *mce*4*C* (Rv3497c), *mce*1*D* (Rv0172), *mce*3*F* (Rv1971)
Cholesterol catabolism	f*adE*28 (Rv3544c), *ltp*2 (Rv3540c), *cyp*125 (Rv3545c)
Lipid and fatty acid metabolism	*lipF* (Rv3487c), *mas* (Rv2940c), *pks*5 (Rv1527c), *pks*7 (Rv1661), *plcA* (Rv2351c), *plcC* (Rv2349c), *fadD*28 (Rv2941)
Metal transporter proteins	*mbtB (Rv2383c*)
Mycolic acid synthesis	*mmaA*4 (Rv0642c)
Proteins inhibiting antimicrobial responses of the macrophage	*ponA*2 (Rv3682)
Sigma factor	*sigG* (Rv0182c)
TA systems	*mazF*8 (Rv2274c), *vapB*47 (Rv3407), *vapC*37 (Rv2103c), *vapC*38 (Rv2494) *mazF3* (Rv1102c), *vapB34* (Rv1740), *vapC12* (Rv1720c)
Type VII secretion system	*eccCb*1 (Rv3871), *secA*2 (Rv1821)
Heat shock proteins	*hsp*22.5 (Rv0990c)

Table 2: A catalog of Beijing and LAM SNPs detected in the virulence genes of *M. tuberculosis*.

Lineages	**Lineage-specific SNPs**
Beijing ancestral	*mazF*8 (G122T), *mazF3* (C194T), *sigG* (G860A) *sigG* (G994T), *mbtB* (G2020C), *plcA* (A1336G), *pks*7 (A2441C), *pks*5 (T6182G), *ltp*2 (G670C), *kefB* (A304G)
Beijing-modern	*glcB* (G310A), *mce*4*C* (C571A), *cyp*125 (C1125T), *lipF* (C697T), *mas* (A6013C), *pks*5 (G6210A). *plcA* (G705A), *plcC* (T753C), *plcC* (G1081T), *fadD*28 (A1306G), *fadD*28 (C1521T), *ponA*2 (G372T), *ponA*2 (G1855A), *vapB*47 (C250T), *vapC*37 (A46G), *vapC*38 (T143C), *eccCb*1 (G1479A), *secA*2 (G1228A), *secA*2 (C1830T)
Beijing-BO/W148	*mce*3*B* (T145G), *eccCb*1 (G1556T), *vapC*12 (A95G)

LAM	*mce*1*D* (C794T), *mce*3*F* (C992G), *fadE*28 (A874C), *hsp*22.5 (C183A)
LAM3	*FadD*28 (C1392G)
F15/LAM4/KZN	*cyp*125 (T1076C), *mce3B* (T44 C), *vapC*25 (A221C), *vapB*34 (C140A)

Table 3: *Mycobacterium tuberculosis* genomes from different countries that have been used to construct phylogenetic tree and detection of signature SNPs in Beijing and LAM lineages

Lineage	Country of origin	Accession number
Beijing	Swaziland	PRJEB7281
	Nigeria	PRJEB7281
	Kenya	PRJNA218286
	China	PRJNA273899
Beijing-modern	Swaziland	PRJEB6273
	Mozambique	PRJNA273400
	Papua New Guinea	PRJNA386696
	South Africa	PRJEB25997
	Uganda	PRJNA233338
	Thailand	PRJEB2777
	Russia	PRJNA509547
Beijing BO/W148	Russia	PRJNA181180
	Russia	PRJEB7281
	Russia	PRJNA476697
	Germany	PRJEB6273
EAI	Vietnam	PRJDB5608
	Mali	PRJNA211652
	South Africa	PRJNA235455
EAI5	India	PRJNA210717
	India	PRJNA193387
EAI1_SOM	India	PRJNA235733
	India	PRJNA235715
	India	PRJNA235757
EAI2_Manila	India	PRJNA235720
	Sweden	PRJNA229212
	Thailand	PRJNA209239
	Sweden	PRJNA229323
LAM	Moldova	PRJNA234579

	Kazakhstan	PRJNA229943
	Iran	PRJNA237427
	Sweden	PRJNA229217
	Peru	RJNA232701
	Uruguay	PRJNA230410
LAM1	Panama	PRJNA208575
	Panama	PRJNA210069
	Panama	PRJNA182109
	Panama	PRJNA208576
	Uganda	PRJNA223585
LAM2	Uganda	PRJNA223572
	Panama	PRJNA208603
	Panama	PRJNA210062
	Panama	PRJNA208576
LAM3	South Africa	PRJNA15642
	South Africa	PRJNA218190
	Panama	PRJNA181126
	Panama	PRJNA210076
	Panama	PRJNA210071
F15/LAM4/KZN	South Africa	PRJNA39765
	South Africa	PRJNA198186
	South Africa	PRJNA21055
	South Africa	PRJNA198122
Lineage 7	Ethiopia	PRJEB8432
	Ethiopia	PRJEB8432
	Ethiopia	PRJEB8432
Haarlem	India	PRJNA192916
	Netherlands	PRJNA218188
	Argentina	PRJNA317008
	South Africa	PRJNA198115
CAS	Sweden	PRJNA229222
	India	PRJNA235634
	India	PRJNA445659
	Iran	PRJNA237429
	Uganda	PRJNA233395
	Uganda	PRJNA233398
	South Africa	PRJNA228084
T1	Mali	PRJNA211698
	Mali	PRJNA211683
	Mali	PRJNA211680
	Iran	PRJNA237410
	India	PRJNA235783
S	South Africa	PRJNA228096
	Romania	PRJNA233357
	China	PRJNA177783

X	Belarus	PRJNA229725
	South Africa	PRJNA235537
	Uganda	PRJNA223563
Ural	Moldova	PRJNA234626
	India	PRJNA213484
	Uganda	PRJNA223578
	Uganda	PRJNA223554

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Health Sciences

The Research Ethics Committee, Faculty Health Sciences, University of Pretoria complies with ICH-GCP guidelines and has US Federal wide Assurance.
- FWA 00002567, Approved dd 22 May 2002 and Expires 03/20/2022.
- IRB 0000 2235 IORG0001762 Approved dd 22/04/2014 and Expires 03/14/2020.

10/10/2018

Approval Certificate
New Application

Ethics Reference No: 610/2018

Title: Evaluation of single nucleotide polymorphisms in virulence genes of Mycobacterium tuberculosis as markers of lineages and sub-lineages in the Tshwane region

Dear Miss U Matodzi

The **Amendment** as described in your documents specified in your cover letter dated 27/09/2018 received on 27/09/2018 was approved by the Faculty of Health Sciences Research Ethics Committee on its quorate meeting of 10/10/2018.

Please note the following about your ethics approval:
- Ethics Approval is valid for 1 year
- Please remember to use your protocol number (**610/2018**) on any documents or correspondence with the Research Ethics Committee regarding your research.
- Please note that the Research Ethics Committee may ask further questions, seek additional information, require further modification, or monitor the conduct of your research.

Ethics approval is subject to the following:
- The ethics approval is conditional on the research being conducted as stipulated by the details of all documents submitted to the Committee. In the event that a further need arises to change who the investigators are, the methods or any other aspect, such changes must be submitted as an Amendment for approval by the Committee.

We wish you the best with your research.

Yours sincerely

Dr R Soma[illegible], MBChB, MMed (Int); MPharMed,PhD
Deputy Chairperson of the Faculty of Health Sciences Research Ethics Committee, University of Pretoria

The Faculty of Health Sciences Research Ethics Committee complies with the SA National Act 61 of 2003 as it pertains to health research and the United States Code of Federal Regulations Title 45 and 46. This committee abides by the ethical norms and principles for research, established by the Declaration of Helsinki, the South African Medical Research Council Guidelines as well as the Guidelines for Ethical Research: Principles Structures and Processes, Second Edition 2015 (Department of Health).

Research Ethics Committee
Room 4-60, Level 4, Tswelopele Building
University of Pretoria, Private Bag X323
Arcadia 0007, South Africa
Tel +27 (0)12 356 3084
E-mail deepeka.behari@up.ac.za
www.up.ac.za

Fakulteit Gesondheidswetenskappe
Lefapha la Disaense tša Maphelo