HISTORIC ORIGIN OF *MYCOBACTERIUM TUBERCULOSIS* COMPLEX STRAINS IN THE FREE STATE PROVINCE, SOUTH AFRICA

By

PAKISO MOSES MAKHOAHLE

DISSERTATION

submitted in fulfilment of the requirements for the degree

Magister Scientiae in Medical Science
M.Med.Sc (Medical Microbiology)

in the

DIVISION OF MEDICINE
FACULTY OF HEALTH SCIENCES
UNIVERSITY OF THE FREE STATE

Promotor: Mrs Anneke van der Spoel van Dijk
Co-Promotor: DR Leen Rigouts

November 2009

Declaration

I declare that the dissertation hereby submitted by me for the M.Med.Sc degree at the University of the Free State, Bloemfontein, is my own independent work and has not previously been submitted by me at another institution/faculty. I further more cede copyright in favour of the University of the Free State.

Pakiso Moses Makhoahle

30/11/2009

Dedication

This work is dedicated to my family; your unwavering love is the wind beneath my wings

Acknowledgement

MY SINCERE GRATITUDE GOES TO THE FOLLOWING PERSONS AND INSTITUTIONS FOR THEIR ASSISTANCE:

MS ANNEKE VAN DER SPOEL VAN DIJK, MY SUPERVISOR AND COLLEAGUE, FOR HER PERSISTENT SUPPORT, GUIDANCE AND WORDS OF ENCOURAGEMENT THROUGHOUT THIS STUDY AND BELIEVING IN ME, YOU REALLY MADE MY DREAM COME TRUE.

DR LEEN RIGOUTS, MY CO-SUPERVISOR, FOR HER GUIDANCE AND WORDS OF ENCOURAGEMENT THROUGHOUT THIS STUDY.

MR SEHLOHO Z MOKHETHI, MR J MAQALA KHUMALO AND MS NELIA VAN HEERDEN FOR PROCESSING SOME SPECIMEN IN THEIR PREVIOUS PROJECTS

DR J RAUBENHEIMER OF BIOSTATISTICS FOR TECHNICAL ASSISTANCE AND STATISTICAL ANALYSIS OF SOME DATA.

DR LOEKIE BADENHORST (HEAD OF DEPARTMENT OF MEDICAL MICROBIOLOGY) FOR HIS SUPPORT, ENCOURAGEMENT IN ENSURING THAT THE RESEARCH IS EXPANDING AND SUPPORTED IN OUR DEPARTMENT.

PROF. NOLAN JANSE VAN RENSBURG AND PROF ANNE-MARIE PRETORIUS (FORMER HEADS OF DEPARTMENT OF MEDICAL MICROBIOLOGY) FOR THEIR SUPPORT AND ENCOURAGEMENT.

THE STAFF OF MEDICAL MICROBIOLOGY AND NATIONAL HEALTH LABORATORY SERVICES (REGISTRAS, TECHNOLOGISTS, TECHNICIANS AND LABORATORY ASSISTANTS) FOR YOU THEIR UNDIVIDED SUPPORT DURING MY STUDY

ARBO VIRUSES GROUP (DEPARTMENT OF MEDICAL MICROBIOLOGY AND VIROLOGY, UFS) FOR THEIR RESEARCH BASED IDEAS AND SUPPORT

MRC, NHLS AND FACULTY OF HEALTH SCIENCE FOR FUNDING THE PROJECT

TABLES OF CONTENTS

List of figures vi
List of tables vii
Abbreviations viii
Abstract xi
Chapter 1 Introduction, Aim and objectives 1
AIM and OBJECTIVES 5
Chapter 2 Literature Review 6
2.1 TUBERCULOSIS 7
2.1.1 Historic background 7
2.1.2 Prevalence of TB 8
2.1.3 Transmission of TB 10
2.1.4 Reservoirs of infection 10
2.1.5 Pathogenesis 11
2.1.6 Tuberculosis and HIV/AIDS 12
2.1.7 Chemotherapy and drug resistance 13
2.1.8 Laboratory diagnosis of TB 15
2.1.8.1 Specimen: 15
2.1.8.2 Microscopy: 16
2.1.8.3 Culture of mycobacteria: 16
2.1.8.4 Nucleic acid amplification techniques 17
2.1.8.5 Sensitivity testing: 17
2.2 *M. TUBERCULOSIS* 19
2.2.1 Genome 19
2.2.2 Genetic heterogeneity of the MTB complex 20
2.3 MOLECULAR EPIDEMIOLOGY OF MTBC STRAINS 25
2.3.1 Molecular markers and techniques 25
2.3.1.1 IS6110-based restriction fragment length polymorphism (RFLP) typing 25
2.3.1.2 Spoligotyping 25
2.3.1.3 MIRU-VNTR typing 27
2.3.2 Principal lineages 32
2.3.2.1 The Beijing lineage 34
2.3.2.2 The East African-Indian (EAI) lineage 35
2.3.2.3 The Manila lineage 35
2.3.2.4 The Central-Asian (CAS) or Delhi lineage 37
2.3.2.5 The Haarlem family 38
2.3.2.6 The Latin American and Mediterranean (LAM) family 38
2.3.2.7 The X family: the European IS6110 low banders 39

2.3.2.8 The T family and others 40

2.3.2.9 Prevalent genotypes in South Africa 41

2.4 DEMOGRAPHIC SETTING AND HISTORY OF THE FS AND STUDY AREA 42

Chapter 3 Materials and methods 45

3.1 SAMPLE 46

3.2 ISOLATION OF MYCOBACTERIA 48

3.3 IDENTIFICATION OF THE ISOLATES 48

3.3.1 Ziehl-Neelsen (ZN) staining 48

3.3.2 Nitrate test 49

3.3.3 Catalase test 49

3.4 DNA EXTRACTION 50

3.5 RESTRICTION FRAGMENT LENGTH POLIMORPHISM (RFLP) 51

3.6 SPOLIGOTYPING 51

3.6.1 DNA amplification 51

3.6.2 Hybridization with PCR product and detection 52

3.6.3 Analysis of results 53

3.7 MULTIPLEX PCR ANALYSIS 54

3.7.1 DNA amplification 54

3.7.2 Gel electrophoresis 56

3.8 MYCOBACTERIAL INTERSPERSED REPETITIVE UNITS OR VARIABLE NUMBER OF TANDEM REPEATS (MIRU-VNTR) TYPING 56

3.8.1 DNA Amplification 56

3.8.2 Gel Electrophoresis 57

3.8.3 Analysis of data 57

3.9 GENETIC RELATIONSHIP AND PHYLOGENETIC ANALYSIS 62

Chapter 4 Results 63

4.1 STRAINS ANALYSED 64

4.2 GENETIC DIVERSITY AND FAMILY ASSIGNMENT 64

4.2.1 Spacer-oligotyping "Spoligotyping" 69

4.2.2 Mycobacterial interspersed repetitive units or variable number of tandem repeats (MIRU-VNTR) typing 73

4.3 PHYLOGENETIC STUDIES 74

4.4 *MULTIPLEX PCR TO DIFFERENTIATE BEIJING STRAINS* 77

Chapter 5 Discussion 79

Chapter 6 Conclusion 87

Chapter 7 References 90

Chapter 8 Summary 115

Chapter 9 0psomming 120

LIST OF FIGURES

Figure 2-1: Circular diagram representing the chromosome of *M. tuberculosis* H37Rv. 22
Figure 2-2: Evolutionary relationships of the MTBC as proposed by Wirth *et al*. 2008. 23
Figure 2-3: The evolution of MTB out of Mesopotamia, a scenario proposed by Wirth *et al*. 2008. 24
Figure 2-4: Position *of* the MIRU loci on the MTB H37Rv chromosome. 29
Figure 2-5: Proposed evolutionary groups of MTB. 33
Figure 2-6: Circular diagram representing genomic deletions in 16 clinical MTB isolates tested by Kato-Maeda et al. 2001. 36
Figure 2-7: Location of the three areas where M. tuberculosis isolates were collected. 44
Figure 3-1: Structural representation of IS6110 direct repeat with the intervening NTF-1 sequence. . 55
Figure 4-1: UPGMA-similarity tree of the spoligopatterns of the 142 isolates plus a control strain (H37Rv) of MTB compared to the reference strains of the MIRU-VNTRplus DB. 70
Figure 4-2: A combined MIRU-VNTR [12] Categorical (1) Spoligo: Categorical (1) UPGMA-similarity tree obtained with MIRU-VNTRplus DB. 75
Figure 4-3: Amplification products from a multiplex PCR assay for W/Beijing strain identification. ... 78

LIST OF TABLES

Table 3.1. Treatment history, gender and age distribution of patients. 47

Table 3.2: Sequences of primers for multiplex PCR 55

Table 3.3: Pairs of primers for amplification of different MIRUs that was developed from the flanking region of the different MIRUs from MTB H37Rv as indicated on Figure 2-4 58

Table 3.4: Specific MIRU-VNTR alleles in *M. tuberculosis* H37Rv control strain. 59

Table 3.5: MIRU-VNTR allele scoring table. 60

Table 3.6: Specific MIRU-VNTR alleles in locus 4 61

Table 4.1: Results of spoligo-MIRU-VNTR typing arranged according to obtained 66

Table 4.2: Results of spoligo-MIRU-VNTR typing arranged according to obtained MIRU-VNTR clones. 67

Table 4.3 Best-match-based analyses performed on the 142 isolates from 68

Table 4.4a: Demographic distribution of lineages over three districts in FS province 72

Ala	**Alanine**
Arg	**Arginine**
AIDS	**Acquired Immune deficiency syndrome**
BCG	**Bacillus Calmette Guérin**
bp	**Base pair**
CDC	**Centre for Disease Control**
CI	**Confidence interval**
CLSI	**Clinical and Laboratory Standards Institute**
CAS	**Central –Asian**
DC	**District**
DR	**Direct repeat**
DOTS	**Directly Observed Treatment Shortcourse**
DB	**Database**
DNA	**Deoxyribonucleic acid**
ETR-A	**Exact Tandem Repeat Allele**
EAI	**East African India**
EDTA	**Ethylenediaminetetraacetic acid**
FS	**Free State**
FSDOH	**Free State Department of Health**
Gly	**Glycine**
h	**hour**
HCl	**Hydrochloric acid**
H_2O_2	**Hydrogen peroxide**
HIV	**Human immunodeficiency virus**
INH	**Isoniazid**
IS	**Insertion sequence**
KZN	**KwaZulu-Natal**
LJ	**Lowenstein –Jensen**
LSP	**Long Sequence Polymorphism**
LAM	**Latin American and Mediterranean**
LM	**Low Melting**

ABBREVIATIONS CONTINUE

$MgCl_2$	**Magnesium chloride**
MDR	**Multidrug resistant**
MIRU-VNTR	**Mycobacterial interspersed repetitive units-variable number of tandem repetitive**
MLVA	**Multiple Loci VNTR Analysis**
MTBC	**Mycobacterium tuberculosis complex**
MPTR	**Major Polymorphic Tandem Repeat**
MTB	**Mycobacterium tuberculosis**
min	**minute (time)**
NGO's	**Non-Government Organisations**
NTF	**Intervening region**
PCR	**Polymerase Chain Reaction**
PGRS	**Polymorphic GC-rich Tandem Repeat Sequence**
PPE	**Proline-Proline-Glutamine**
PE	**Proline-Glutamine**
PTB	**Pulmonary Tuberculosis**
Pro	**Proline**
RFLP	**Restriction fragment length polymorphism**
RIF	**Rifampicin**
SSR	**Short Sequences Repeats**
Spacer-oligos	**Spoligotyping**
SNPs	**Single nucleotide polymorphisms**
SADC	**Southern African Democratic countries**
ST	**Shared Type**
SIT	**Shared Type Information**
sec	**second**
SSPE	**Saline-sodium phosphate EDTA**
SDS	**sodium dodecyl sulphate**
TB	**Tuberculosis**
TBE	**Tris Borate EDTA buffer**
TE	**Tris EDTA**
Trp	**Trypsin**
Tris	**2-Amino-2-(hydromethly)-1, 3-propanediol**
US	**United States**

UPGMA	**Unweighed Pair Group Method with Arithmetic averages**
VNTR	**Variable Number Tandem Repeats**
V	**Volts**
WHO	**World Health Organisation**
XDR	**Extensive drug resistant**
ZN	**Ziehl-Neelsen**
Units/µl	**units per microlitre**
v/v	**volume per volume**
mm	**millimetre**
ml	**millilitre**
mg/ml	**milligram per millilitre**
M	**Molar**
mM	**millimolar**
ng/µl	**nanogram per microlitre**
pmol/µl	**picomole per microlitre**
U/µl	**unit per microlitre**
U/ml	**unit per millilitre**
µg/ml	**microgram per milliliter**

ABSTRACT

With TB still a major threat worldwide and South Africa (SA) being one of the countries with the highest TB incidence, the development in the field of epidemiology gained importance to elucidate the history and dynamics of the disease.

The aim of the study was to elucidate the diversity and historic origin of *Mycobacterium tuberculosis* (MTB) strains in the FS by studying the molecular epidemiology of isolates in three high burden areas.

During 2001-2003 MTB isolates were collected during two studies in the FS from consenting patients and DNA from these isolates was further investigated in this study. It involved the use of molecular methods such as spoligotyping and 12MIRU-VNTR typing that provides unique patterns for different highly transmissible genotypes, permitting historic relationships and the origin of circulating strains to be deciphered and to make predictions concerning their spread. Multiplex PCR was used to further characterise the Beijing lineage types and complex software packages including the SpolDB3 and 4, Excell 2007 and MIRU-VNTRplus (to draw phylogenetic trees) aided the analysis of the results.

Strains of the FS were found to be extremely diverse with spoligo analysis resulting in an overall diversity of 73% and 12 MIRU-VNTR analyses a diversity of 50%.

Nine lineages, namely Beijing, Haarlem, LAM, S, T, X, CAS, *M. bovis* and U were represented by the FS isolates with the three main families (T, LAM and X) respresenting 67.6%. Spoligopattern diversity within each of these three families varied substantially from 76%, 56% and 80% for T, LAM and X respectively. The T family was the most prominent, but phylogenetic tree analysis comparing the results of different typing methods, except for one

closely related T1 (SIT53) cluster, showed little relationship between them as expected from reports globally. SIT 53 strains were the most prominent family found amongst the T isolates FS stains.

The LAM3/SIT33 was the largest clonal group of FS strains and showed little diversity even by MIRU-VNTR code analyses. Only 8 Beijing type isolates were present, a family that is prominent in the Western Cape and many other countries. Comparison of all strains to international strains on the MIRU-VNTRplus database shows that FS strains can be linked to other countries such as Uganda, Ghana, Cameroon, China, Delhi counties, Zimbabwe and Tanzania.

Nevertheless, it remains hard to determine the detailed history of MTBC strains in the FS province. It was not possible to determine unique pattern for the MTB strain using the 12-loci based MIRU-VNTR typing methods. Given the fact that both the X3 and Beijing strains - dominant in the Western Cape – are less important in the FS, it is more likely that MTB in the FS migrated down from the Gauteng province and Northern countries including Uganda, Ghana, Zimbabwe and Cameroon where similar lineages are present.

This study served as a pilot as it contains isolates collected in the mid 2001. The incidence of legal and illegal immigrants entering the country plus global economic partnership allowing people free movement in and out of the country might have changed the dynamics of the disease.

A need for follow up on current strains is urgently required to detect changes that influence TB TB management in the FS.

Chapter 1

Introduction,

Aim and objectives

Tuberculosis (TB) has become a world threat fuelled by the modern world of inter dependence and globalization, where people can travel to all parts of the world freely and easily, global trading, and changing socio-cultural patterns.

South Africa, a country that is extremely diverse in populations and topography, has high rates of poverty in sparsely populated relatively large provinces with vast open spaces as the Free State (FS), which leads to limited travel except for work or study related purposes. With little resources for research, policy makers in such provinces often have to rely on molecular epidemiological surveillance and resistance monitoring data for TB done in a few areas of the country to make decisions about treatment programs without the certainty that the data is applicable to the country as a whole, e.g. molecular epidemiological TB studies of a high incidence area in the Western Cape, sporadic studies in mining populations and a prison outbreak in KwaZulu Natal.

In the early 1990's the newly developed molecular techniques for DNA fingerprinting of TB using restriction fragment length polymorphism (RFLP) based on the number and location of the insertion sequence (IS)*6110* has globally enhanced our understanding of transmission, spread and emergence of multidrug resistant (MDR) strains and TB infections (Warren *et al.,* 1996; Bifani *et al.,* 1999; 2002).

DNA fingerprinting has revealed extensive genetic polymorphism among *Mycobacterium tuberculosis* (MTB) isolates and facilitated the recognition of various MTB groups and genotype families. IS*6110*-based RFLP typing represents the gold standard in molecular epidemiology of MTB and reveals extensive diversity amongst TB strains. Since then several polymerase chain reaction (PCR) based tools for molecular typing of bacteria including variable-numbers of tandem repeat (VNTR) sequences and polymorphism of

the chromosomal direct repeat (DR) locus that contains a variable number of short repeats, interspersed with non-repetitive spacers that provide unique, easy identifiable patterns for different lineages (spoligotyping), multiplex PCR to distinguish between the different Beijing strains and sequencing lead to the current field of molecular epidemiology (Bifani *et al.,* 1999; Kamerbeek *et al*., 1997; Lari *et al.,* 2005). The latter two methods, although showing less diversity, allow for TB strains to be divided into distinct lineages and to determine the historic origin of strains present in different populations.

Molecular typing of MTB has thus become a high priority for researchers globally to gain more information about circulation of specific strains, determining genetic changes that may result in antibiotic resistance, highly transmissible strains and area specific lineages, and changes in the dynamics of the disease in populations (Spugiesz *et al.,* 2003). Both molecular epidemiological surveillance and characterisation of resistant strains are valuable new disciplines in understanding the disease dynamics of TB and ultimately assist in keeping abreast of the development of highly transmissible MDR/ extensively drug resistant (XDR) strains and outbreaks to facilitate intervention. Molecular studies have discovered multiple markers including the IS, DR locus, deletion regions, mini-satellites and single nucleotide polymorphisms (SNPs) to track similarities and differences in the molecular evolution of the MTB complex (Allix-Béguec *et al*., 2008).

A highly homogeneous MTB genotype, designated the Beijing genotype family, has been found worldwide, but predominantly in South-East Asia (van Soolingen *et al*., 1995; Kremer *et al*., 2004a; Glynn *et al*., 2002). It then spread across the world and has been reported to predominate in both high and low burden countries and the Western Cape in South Africa (van Soolingen *et al*., 1995; Soini *et al.,* 2000; Narvskaya *et al*., 2005; Borgdorff *et al*., 2003). Spoligo- and MIRU-VNTR typing provides unique patterns for the Beijing and other highly transmissible genotypes permitting historic

relationships, and the origin of currently circulating strains to be deciphered and to follow spreading of outbreaks (Supply *et al.,* 2001; Kremer *et al.*, 2005a; Sampson *et al.*, 1999).

Molecular typing of MTB strains in the Western Cape of South Africa has shown that over half of MDR-TB patients share DNA profiles with other patients. Some of the highly prevalent multidrug resistant strains are Beijing types, an area specific strain DRF 150 and the KZN strain from Kwazulu-Natal (Victor *et al.*, 2004; Ndimande *et al.* 2005). Although the Beijing genotype is well known and associated with drug resistance in some parts of South Africa, the contribution of the Beijing genotype to the TB problem in the FS is unknown. A recent study in the FS hypothesized that the FS province has a very diverse population of TB strains and a relatively low prevalence of MDR-TB (van der Spoel van Dijk *et al.,* 1996, 2005). It is therefore important for the TB program in the FS to determine the genotype and historic origin of these strains and their importance to the local pool of strains.

Determining the MTB genotypes circulating in the Free State and the possible historic relationship of the Free State strains to that of other provinces and countries will be of importance to the national TB program adding knowledge about the contribution of known genotypes to the local pool of strains and the relevance of countrywide studies to the FS. Secondly, although the strains are not a totally representative sample of the Free State strains, analysis of these strains will serve as a pilot study building capacity in the use of several molecular techniques with the potential to be used as rapid epidemiological typing tools in the Free State.

AIM and OBJECTIVES

The aim of the study was to investigate the diversity and the possible historic relationship of MTB strains from three areas in the Free State, South Africa previously analysed by IS6110-based RFLP typing.

The objectives of this study were:

- To assign previously IS*6110*-based RFLP fingerprinted MTB strains to different genotype families of interest using spoligotyping
- To utilise multiplex PCR analysis for the IS*6110* direct repeat with an intervening NTF-1 sequence to distinguish W and U type Beijing strains
- To use 12-loci based MIRU-VNTR typing to determine a unique pattern for the MTB strains and possibly providing a faster and more cost effective method to study strain diversity
- To compare the obtained MTB spoligo- and MIRU-VNTR profiles with the international databases SpolDB3.0, SpolDB4.0, MIRU-VNTRplus and literature about the history of MTB to determine the possible historic origin of the strains.

Chapter 2

Literature Review

2.1 Tuberculosis

2.1.1 Historic background

TB is an infectious disease of humans, which dates back to ancient generations. Evidence of TB in fossils dates back 6 000 to 8000 years indicating the disease being as old as human history (Salo *et al.,* 1994). Robert Koch discovered the causative agent of TB in 1882 (Zumla *et al.*, 1999). Edward Livingston Tredau was the first to look for MTB in clinical specimens to confirm clinical diagnosis (Zumla *et al.*, 2000).

During Hippocrates times (460-377 BC) and later, TB was known as "phthisis" and became well known in Europe in the 17th century during a massive epidemic from where it was spread through colony sites including South Africa (Edginton, 2000). Different populations who were not exposed to TB before were rapidly infected because they had no immunity against TB (Edginton, 2000). Infection rates increased rapidly following the discovery of diamonds in 1867 and gold in 1886 as a result of poor living conditions and poverty (Metcalf, 1991, Cronje *et al.,* 2006). This created a demand for labour, drawing migrant workers to mines from all over Southern African Democratic countries (SADC). Conditions at mines favoured the spread of infection: overcrowded mine compounds; long shifts; poor ventilations; stress; inadequate diets and prevailing diseases such as malaria and pneumonia initiated an ideal breeding-ground for TB and increased the spread of the disease (Collins, 1982, Metcalf, 1991). Miners who were infected were sent home immediately and that facilitated the spread of TB among the families mostly in rural areas. In the 20th century, urbanisation, housing shortages, economic recession and hot climates contributed to the TB epidemic in South Africa (Metcalf, 1991).

2.1.2 Prevalence of TB

It was estimated that more than 60% of the black population of South Africa was infected in 1930 (Collins, 1982). The incidence rate of TB in 1953 was estimated to be 780/100 000 of the population in the northern and eastern parts of South Africa. Incidence rates of TB increased rapidly at the early 20th century with a peak in the 60's and it started to decline in the early 70's when certain parts of the country were declared independent states and were excluded in reported statistics (van Rensburg *et al.,* 1982). Apartheid policies were the major contributing factors to the spread of TB during this era (van Rensburg *et al.,* 1982). In East London, access to urban areas was forbidden for the black population without a pass as early as 1848. By 1903 Cape Town joined this strategy, facilitating the permanent isolation and concentration of blacks in one area, creating favourable conditions to spread the disease (Dubovsky, 1987, Collins, 1982). Rural poverty and rapid urbanisation created living conditions that favoured continuation of the epidemic (Collins, 1982). Health services were inadequate for the majority of the population, causing delay in diagnosis and facilitating the development of resistance to drugs due to start-stop treatment (Bradshaw *et al.,* 1987, van Rensburg *et al.,* 2005).

In the beginning of the 21st century, TB was predicted as virtually eradicated (Blumberg, 1995), but has emerged again as one of the world's most serious afflictions. Deteriorating socio-economic conditions with increasing poverty and homelessness, a breakdown in TB control programmes, the human immunodeficiency virus (HIV) epidemic and the emergence of MDR and XDR strains have instead created an unfavourable situation for TB control (Department of Health, 1998; Pillay *et al.,* 2007).

TB is the most common cause of infectious disease mortality worldwide and was declared a global emergency by the World Health Organisation (WHO)

in 1993 (WHO, 1994). In 2003, the Centre for Disease Control (CDC) reported that despite all the TB control programmes, the incidence of TB is still on the increase (CDC, 2003). It was estimated that by the year 2020 nearly one billion people will be newly infected, 200 million will get sick and 35 million will die from TB (CDC, 2003). The WHO estimates that one-third of the world population has latent TB. In 2006, 9.2 million new cases were reported, while another 1.7 million people died worldwide of TB of which 0.2 million deaths were HIV-positive people. TB is responsible for a quarter of preventable deaths globally. Ninety percent of TB-related deaths occur in developing countries (WHO, 2003).

Approximately two thirds of the world TB cases occur in Asian countries, whereas Africa has the highest incidence of TB in the world, with about 1.5 million infected individuals developing TB and more than half a million deaths reported yearly (WHO, 2008). More than 75 percent of the newly infected cases occur in 22 high-burden countries with South Africa ranking position four on the list (WHO, 2008).

The incidence of all forms of TB in South Africa was 526/100 000 in 2001 and 940/100 000 population in 2006, an incidence rate classified as a serious epidemic by the WHO (WHO, 2008; The South African TB control programme practical guidelines, 2000). In South Africa, the cure rate is low for all diagnosed smear-positive cases, with death and default the most frequent negative outcomes (WHO, 2008). Case notification rates continue to increase; a reassessment of the incidence estimates, based on registered deaths, suggested that the 70% case detection rate target was reached for the first time in 2007 (WHO, 2008).

Western Cape has the highest incidence of TB in South Africa, and the Free State with an incidence of 352/100 000 ranking 4th. From 144 910 new cases of pulmonary TB in 2001, the Free State contributed 9 978 cases (Nicol *et al.*, 2005, FSDOH, 2002). In the South African mining industry,

the incidence has doubled since the advent of the HIV epidemic, with the TB incidence estimated to be 2000 per 100 000 population annually (http://www.weforum.org/pdf/Initiatives/GHI_TB_Anglogold_AppendixA.pdf).

2.1.3 Transmission of TB

M. tuberculosis is spread by airborne particles, known as droplet nuclei or aerosols, which can be generated when persons with pulmonary or laryngeal TB sneeze, cough, speak, or sing.

Persons who share the same airspace with persons with open pulmonary TB are at greatest risk for infection by inhalation of droplet nuclei containing tubercle bacilli. These bacilli penetrate macrophages in the alveoli of the lungs. From thereon they can later spread in the lungs or throughout the body.

2.1.4 Reservoirs of infection

Humans are the main reservoir of infection, with persons suffering from active pulmonary disease posing the biggest threat for infecting other persons. The incidence of TB varies in the developed and developing countries. The distribution of TB disease is also uneven within countries. Certain groups within the societies bear a disproportionately high burden, such as AIDS patients, close contacts of TB patients, immigrants, medically under-serviced populations, alcoholics and intravenous drug users, people residing in long-term care facilities (e.g. nursing homes) or correctional institutions, mine workers and homeless people (CDC, 1990).

2.1.5 Pathogenesis

M. tuberculosis can survive within macrophages. It rises to slowly developing, chronic disease in which much of pathology is attributed to host immune responsiveness rather than direct bacterial toxicity. Clinically, TB can present in three stages: primary, latent and post-primary.

Primary infection

The majority of cases are asymptomatic with a primary local lung lesion (Ghon focus), with marked enlargement of the lymph nodes. A primary focus can alternatively include others sites (Blumberg, 1995).

Primary infection may progress to tuberculous bronchopneumonia, caseation, miliary TB, tuberculous meningitis, bone and joint TB or genitourinary TB.

Latent infection

At this phase the tubercle bacilli remain dormant before initiating active disease up to many years after the primary infection.

Post-primary infection

This phase generally involves the lungs, with lesions in the apices: if untreated, the tubercle bacillus is activated and progressive chronic disease develops, with areas of local exudation and caseation surrounded by dense fibrosis. Caseous lesions enlarge with fluid contained in the cavities, can then be seen on radiography.

Presenting symptoms include non-specific illness, with fever, night sweats, weight loss, and respiratory symptoms such as chronic cough, haemoptysis

and a pneumonic illness that fails to respond to conventional antibiotics.

Factors that induce post-primary infection can be:

Endogenous: reactivation of latent foci formed at the time of primary infection because of a weakened immune system.

Exogenous: reinfection, by inhalation of the infected respiratory secretions from an infected individual with tubercle bacilli in the sputum. Approximately between 30 to 40% of cases that have been treated before are the result of reinfections, but this rate can be higher in cases of reinfections with MDR strains (Metchock *et al.*, 1999).

2.1.6 Tuberculosis and HIV/AIDS

The advent of the HIV/AIDS epidemic has fuelled the spread of TB globally, overwhelming the TB control programmes (WHO, 2008). TB infection in HIV-positive patients is more likely to progress to active disease than in HIV-negative patients. HIV-infected persons have a 10% per year risk of reactivation of a dormant TB focus, as compared to a 10% lifetime risk for HIV-negative persons. The clinical picture parallels the degree of immune suppression (Blumberg, 1995).

The HIV/AIDS epidemic poses a serious challenge to the success of TB control programmes globally. The United States of America have reported a significant increase in case notifications, with approximately 30% increase at the beginning of 1986. This trend has certainly been associated with the spread of AIDS. Sub-Saharan Africa is the most devastated with up to 60% of children with active TB and above 70% of TB adults being co-infected with HIV (Haries, 1998). Mortality in dually infected patients is approximately 50 % (Blumberg, 1995; Mwinga, 1999), although the introduction of anti-retroviral treatment has improved TB cure rates in these patients (Schluger, 1999).

In South Africa, more than 60% of TB patients are also HIV-positive and this proportion is still increasing. In September 2006, in the rural area of Kwazulu-Natal Tugela Ferry, it was reported that almost 50% of 544 TB patients were MDR cases, and 53 of the MDR patients were XDR positive (Gandhi *et al*., 2006). A disturbing issue is that about 80% of the XDR patients were HIV positive. In 2001, 410 patients diagnosed with TB were tested for HIV in the Free State province, and 288 (70%) were HIV positive (Kironde *et al.,* 2002). It has been shown that HIV/AIDS increases the chance of reactivating dormant TB from 10% to 50% (Kironde *et al.,* 2002).

2.1.7 Chemotherapy and drug resistance

The breakthrough in TB treatment was in the mid 20th century when para-aminosalicylic acid (PAS) and streptomycin were discovered (1944) followed by isoniazid (INH) in 1951 (Iseman, 2002). During that period of anti-TB agent's discovery, combination therapy of PAS and streptomycin was found to be more effective by the British Medical Research Council than the single agents in killing drug resistant strains of MTB (Iseman, 2002). Since that time chemotherapy has been the most potent tool available to fight TB and consists of 6 to 8 months combined treatment, including an intensive phase (4 to 5 drugs for 2 to 3 months) and a continuation phase (2 to 3 drugs for 4 to 5 months). When used properly, available anti-TB drugs are able to reach cure rates above the 85% target recommended by the WHO (British Thoracic Association, 1982). However, emerged resistance to treatment has become a point of concern worldwide, mainly associated with increased treatment failures. Resistance is defined as single-drug, multi-drug, or poly-drug resistance depending on the number of drugs and /or the specific drugs involved (Rieder, 2002). Of particular concern is the increasing prevalence of MDR-TB organisms, i.e. resistant to isoniazid (INH) and rifampicin (RIF), the two backbones of modern short-course therapy, and the emergence of the

deadly XDR-TB strains. Resistance to RIF is been used as an indicator of MDR-TB, as to date only few RIF-mono-resistance strains have been reported. In most cases RIF resistance occurs in conjunction with INH resistance (over 92% of cases) (Traore *et al*., 2000, 2006, Watterson *et al*., 1998, Nikolayevskyy *et al.,* 2009).

Drug resistance in MTB occurs mainly as a result of random spontaneous chromosomal mutations during natural cell division. These mutations are not drug induced or linked. The probability of a drug-resistant mutant to occur is directly proportional to the size of the bacterial population, and the frequency of primary resistant organisms varies for each drug: spontaneous resistance to INH is estimated to occur once in every 10^6 organisms and to RIF once in every 10^8 organisms. The probability of spontaneous mutants being simultaneously resistant to two or more drugs is the product of the individual mutants. Drug-resistant mutants can be selected if patients are treated inappropriately. The current (M) DR-TB epidemic is a man-made amplification of a naturally occurring phenomenon. Previous treatment to TB predisposes to selection of MDR or even XDR organisms. Non-adherence to treatment and HIV status are the major factors allowing resistant organisms to survive (Portaels *et al*., 1999, Pillay *et al*., 2007). Combined drug therapy is used to prevent the selection of drug-resistant mutants (American Thoracic Society, Centers for Disease Control and Prevention, Infectious Diseases Society of America (2003)).

Literature reports show an unequal global distribution of drug resistant TB. Countries with a high prevalence of MDR-TB include Latvia (1998: 9.0%) Estonia (1998: 14.0%), the Dominican Republic (1994-1995: 6.6%), Ivory Coast (1995-1996:5.3%), Argentina (1994:4.6%), Russia (Ivanovo Oblast) (1998: 9.0%), Iran (1998: 5.0%) and the Henan province in China (1996: 10.8%). South Africa's neighbours Botswana (1995-96), Lesotho (1994-95) and Swaziland (1994-95) have reported encouraging results of 0.2%, 0.9% and 0.9% respectively. Acquired MDR rates of over 20% were reported in

Guinea (1998:28%), Latvia (1996:54.4%), Mexico (1997:22.4%), Italy (1999:33.9%), Russia (Ivanovo Oblast) (1998: 25.9%), Tomsk Oblast (1999: 26.7%), Estonia (1998: 37.8%), Iran (1998 48.2%), Sierra Leone (1997: 23.1%), Argentina (1994:22.2%) and Spain (Barcelona) (1995-96:20.5%). Again acquired MDR-TB was low in Botswana (1998:9.0%), Mozambique (1999:3.3%), Lesotho (1994-95:5.7%) and Swaziland (1994-95:9.1%) (Cohn, 1997, Espinal *et al.*, 2001).

In high burden countries, the WHO has recommended the implementation of a Directly Observed Treatment System (DOTS) plus for the management of MDR-TB, which involves the use of specific treatment regimes together with isolation of TB bacilli from sputum and subsequent drug-susceptibility testing. Epidemiological surveillance of resistance and mutation monitoring are still not fully employed in South Africa and should be the pillar of the governmental continuous programmes as well as DOTS-plus programmes, including the help of non-governmental organizations (NGO's) (Consensus statement, 2003).

An XDR-TB strain was identified and reported in KZN in 2006 as being resistant to the most effective anti-TB drugs (Pillay *et al.,* 2007).

2.1.8 Laboratory diagnosis of TB

2.1.8.1 Specimen:

Early morning sputum is the most important specimen for the diagnosis of pulmonary TB. For a patient who cannot produce sputum (in the case of pulmonary TB in many HIV patients), a gastric aspiration method can be used. Cerebrospinal fluid is required for the diagnosis of meningitis TB. Early morning urine can be used for the diagnosis of renal TB, although routine urine cultures may be positive in only 7%-10% of patients. Blood culture has become more important in the diagnosis of generalized mycobacterial

disease in HIV-positive patients (Blumberg, 1995).

2.1.8.2 Microscopy:

The direct microscopic examination of sputum smears is central in the diagnosis of pulmonary TB, especially for the detection of infectious cases. Slides can be stained in two ways; by a modified Ziehl-Nielsen method, in which carbol-fuchsine is used to stain the bacilli; or by the phenol auramine technique, which uses a fluorescent dye easily visible under ultraviolet illumination. Auramine straining enables a large number of specimens to be screened quickly and is particularly suited for laboratories with a large throughput. In microscopy diagnostic methods the detection limit is between 10^4 and 10^5 bacilli per ml of specimen, meaning that patients with fewer bacilli than the limit will be most probably classified as smear negative and thus less infectious (Blumberg, 1995).

Microscopy is rapid, cheap and relatively easy to perform. Sensitivity of auramine microscopy to detect pulmonary TB approaches about 60-70%. The sensitivity of carbolfuchsine-based smear microscopy of sputum is even lower (i.e., 50% in adults) (Blumberg, 1995). Organisms other than TB may also demonstrate various degrees of acid fastness (*Nocardia asteroides)* leading to false smear-positive results (Blumberg, 1995).

2.1.8.3 Culture of mycobacteria:

Most species of mycobacteria are slowly growing, with a doubling time of 20-22 hours. Culture of sputum is more sensitive than microscopy and detects 100-1000 organisms per ml. Because of the slow growth, cultures of clinical specimens should be held for two months before they can be recorded as negative. Suitable media for growing mycobacteria include, among others, Lowenstein-Jensen (LJ) and the BACTEC liquid medium. Higher levels of

false-positive results have been associated with the use of broth-based systems, but since classical cultures are very slow, methods that allow rapid growth of mycobacterium using liquid media have become the preferred method worldwide. These methods include several auto- and semi-automated systems such as the BACTEC TB-460, BACTEC (Mycobacterial Growth Indicator Tube) MGIT^{960} which is, currently the preferred method in South Africa, VersaTREK and BacT/Alert 3D (Blumberg, 1995; Badak *et al*., 1996, Brunello *et al*., 1999; Alcaide *et al.,* 2000; Williams-Bouyer *et al.,* 2000).

2.1.8.4 Nucleic acid amplification techniques

Also molecular methods have been proposed for use in the rapid diagnosis of TB (Shamputa *et al*. 2004). Several mycobacterial target genes have been investigated in assay systems, which allow the identification of a single or multiple mycobacterium species. However the sensitivity for these assays so far have not yet reached those of culture (Shamputa *et al.,* 2004).

2.1.8.5 Sensitivity testing:

The method that is recommended by the Clinical and Laboratory Standards Institutes (CLSI) for susceptibility testing of MTB is the modified agar proportion method. This method involves the growth of mycobacteria on solid media (either LJ or Middlebrook-Cohn7H10) that contain various antituberculous drugs. This method is inexpensive and relatively simple providing the results in four to six weeks from a culture isolate. However this system has been standardized only for the first-line drugs (i.e. INH, RIF, streptomycin, ethambutol and pyrazinamide). The BACTEC system (Becton Dickson) is also used globally. It provides results in up to five or 10 days from a culture isolate, but requires expensive reagents, equipment and technical expertise. E-test and luciferase-based reporter mycobacteriophage

can also be used as a surrogate marker for drug susceptibility (Blumberg, 1995). Molecular tools, especially line probe assays such as GenoType® MTBDR assay (Hain LifeScience GmbH, Nehren, Germany) and Inno-LiPa®Rif.TB test (Innogenetics, Ghent, Belgium), have become more important with the advent of MDR and XDR-TB to provide resistance data in time to ensure effective treatment (Nikolayevskyy *et al.,* 2009, Traore *et al.,* 2006). A meta-analysis by Ling *et al.,* 2008 found that the specificity for the GenoType® MTBDR assay were high for both rifampicin and isoniazid (98.7%, 95% at confidence interval (CI) 97.3-99.4 and 99.5%, 95% CI 97.5-99.9 pooled specificity respectively). The pooled sensitivity (98.1%, 95% CI 95.9-99.1) of rifampicin was good in all studies, but for isoniazid the sensitivity was lower and variable (84.3%, 95% CI 76.6-89.8). An improved assay, the GenoType® MTBDR plus assay with new target genes for the isoniazid regulatory region and improved primer and probe designs increased concordance results for isoniazid in smear positive samples from 72.3% to 91.5% compared to culture-based drug susceptibility testing in one study and from 88% to 92% in another. A study in a high burden area in South Africa also reported excellent results and the assay is now used in many laboratories across the globe (Ling *et al.,* 2008, Miotto *et al.,* 2006, Hillemann *et al.,* 2007, Barnard *et al.,* 2008).

2.2 *M. tuberculosis*

TB is caused by a group of closely-related bacterial species belonging to the *M. tuberculosis* complex (MTBC). Today the principal cause of human TB is *M. tuberculosis.* Other members of the MTBC include the human pathogens *M. africanum* which is responsible for about 60% of cases in certain regions in Africa (Zumla *et al.*, 1999; Brosch *et al.*, 2002; Gagneux *et al.*, 2006, 2007; Gutacker *et al.*, 2002, 2006; Brudey *et al.*, 2006b)*, M. canetti* and species usually associated with animal infections, such as *M. bovis, M. microti* and *M. pinnipedii* (Gutierrez *et al.,* 2005)*.*

The exact host-association of *M. africanum* subtype I strains has not been studied so far. There is some evidence that *M. africanum*, which is less virulent than other MTBC genotypes, is currently extinct in settings where it was the most prevalent strain only three decades ago. Instead, it is being replaced by imported, more virulent genotypes (Homolka *et al.,* 2008). The genetic susceptibility of the indigenous African population to TB during World War I is a well-known fact. This supports the idea that TB caused by a more virulent genotype evokes a different, acute and even fatal disease, very different from that produced by *M. africanum*.

2.2.1 Genome

The genome of MTB, made up of a single chromosome, consists of approximately 4.4 million base pairs, and contains around 4000 genes. The DNA material is rich in repetitive DNA, namely insertion sequences (IS) and short repetitive DNA, new multi gene families and duplicated housekeeping genes. Most of the IS in MTB appear to have inserted in intergenic or non-coding regions. Many are clustered, suggesting the existence of insertional hot spots that prevent genes from being inactivated. Genetic DNA elements

called short repetitive DNA, associated with some degree of diversity have been identified in the MTBC. Three of these, the polymorphic GC-rich tandem repeat sequence (PGRS), a repeat of the triplet GTG, and the major polymorphic tandem repeat (MPTR), are present at multiple chromosomal loci. The genome is also rich in G+C nucleotides. The presence of a high proportion of the G+C rich codons results in an increased number of the amino acids alanine (Ala), glycine (Gly), proline (Pro), arginine (Arg) and trypsin (Trp) (Cole *et al.*, 1998).

After its isolation in 1905, the H37Rv strain of MTB has found extensive global use as a control strain in medical research, because it has retained full virulence in animal models and is the first strain to be sequenced completely, as shown in Figure 2.1 (Phillip *et al.*, 1996, Brosch *et al.*, 1998, Cole *et al.*, 1998).

2.2.2 Genetic heterogenity of the MTB complex

There is ongoing debate to consider *"M. prototuberculosis"* to be the common ancestor to all MTBC members. Some researchers disagree and argue that the computation time frame of more than 3 centuries is less likely to be true and cannot be reliable. Furthermore, there is no evidence to prove that *"M. prototuberculosis"* is a more likely ancestor to the MTBC than any animal pathogen still to be characterized (Smith, 2006a, Cohan, 2002, Smith *et al.*, 2006, Godreuil *et al.,* 2007). Even though Gutierrez *et al* have shown that the gene mosaicism found in *"M. prototuberculosis" is* real, more studies on the genetic diversity of *"M. prototuberculosis"* are needed to increase our understanding on lateral genetic transfer and homologous recombination events in the MTBC (Gutierrez *et al.,* 2005).

It has been believed for a long time that MTB emerged from *M. bovis* after adaptation to humans. This idea was initially supported by molecular findings. As Brosch *et al.* identified deletions in *M. bovis* by comparing it with

the only MTB chromosome sequence available at that time (Fig. 2.1), it was inevitable to conclude that *M. bovis* was the terminal group in the lineage (Smith 2006a). The speculation that the region of difference 9 (RD9; deleted in the ancient lineages and in *M. bovis*) descended from a MTB like ancestor, also implies that the most recent common ancestor of these strains was adapted to humans.

The molecular evolution of *M. bovis* provides an interesting framework for comparison with that of MTB (Smith *et al.,* 2006). In particular, Smith and collaborators, using the genetic diversity of *M. bovis* in the United Kingdom as a model, demonstrated that all *M. bovis* genotypes derive from a single clonal complex that is likely to have emerged as a result of the actions of bovine TB control programs, which have been in force for the last 100 years (Smith *et* al., 2006).

Wirth *et al* demonstrated the origin, spread and demography of MTBC in Fig. 2.2 and 2.3 (Wirth *et al.*, 2008). In Fig. 2.2 Wirth *et al* evidently showed *M. prototuberculosis* as the ancestor and proposed evolution of MTBC (Wirth *et al.*, 2008). In Fig. 2.3, with numbers, they indicated where different lineages originated and spread according to human migration patterns to Africa (Wirth *et al.*, 2008). This evolutionary figure unveils the dynamic dimension of the association between human and MTBC pathogens.

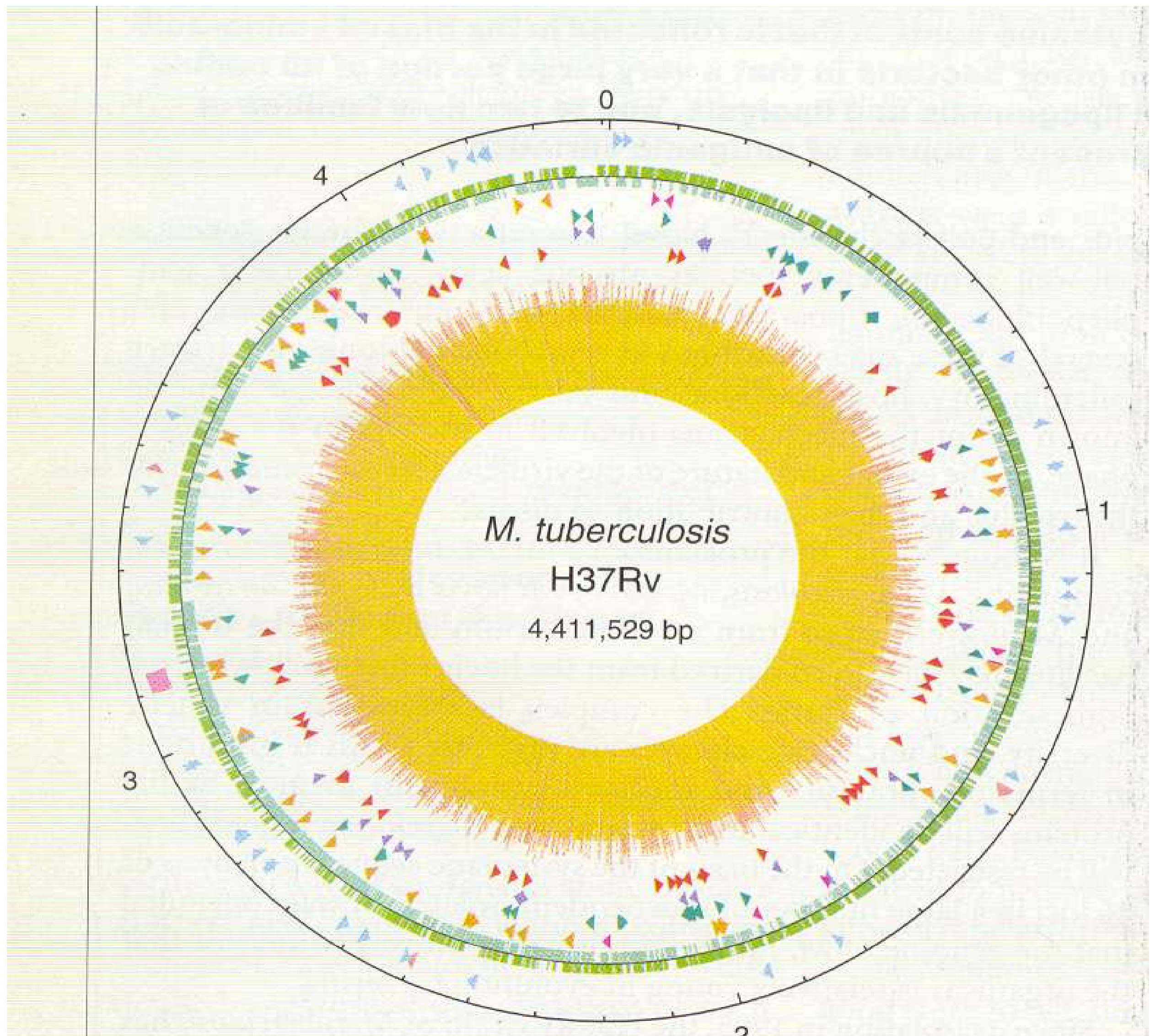

Figure 2-1: Circular diagram representing the chromosome of *M. tuberculosis* H37Rv.

This diagram was generated by Cole *et al* using software from DNASTAR. "The outer numbers represent the scale in Mb, 0 represent the origin of replication. The first ring from the exterior denotes the position of stable tRNAs (blue, or pink) and the direct repeat region (pink cubic); the second ring shows the coding sequence by strand (clockwise, dark green, anticlockwise, light green); the third ring depicts repetitive DNA (Insertion sequence, orange ; 13E12 REP family ,dark pink, prophage, blue; the fourth ring shows the position of PPE family members (green); the fifth ring shows the PE family members (purple excluding PGRS); and the sixth ring shows the position of the PGRS sequences (dark red). The histogram (center) represent G+C content, with <65% G+C content in the yellow, and >65% G+C content in red" (Cole *et al.*, 1998).

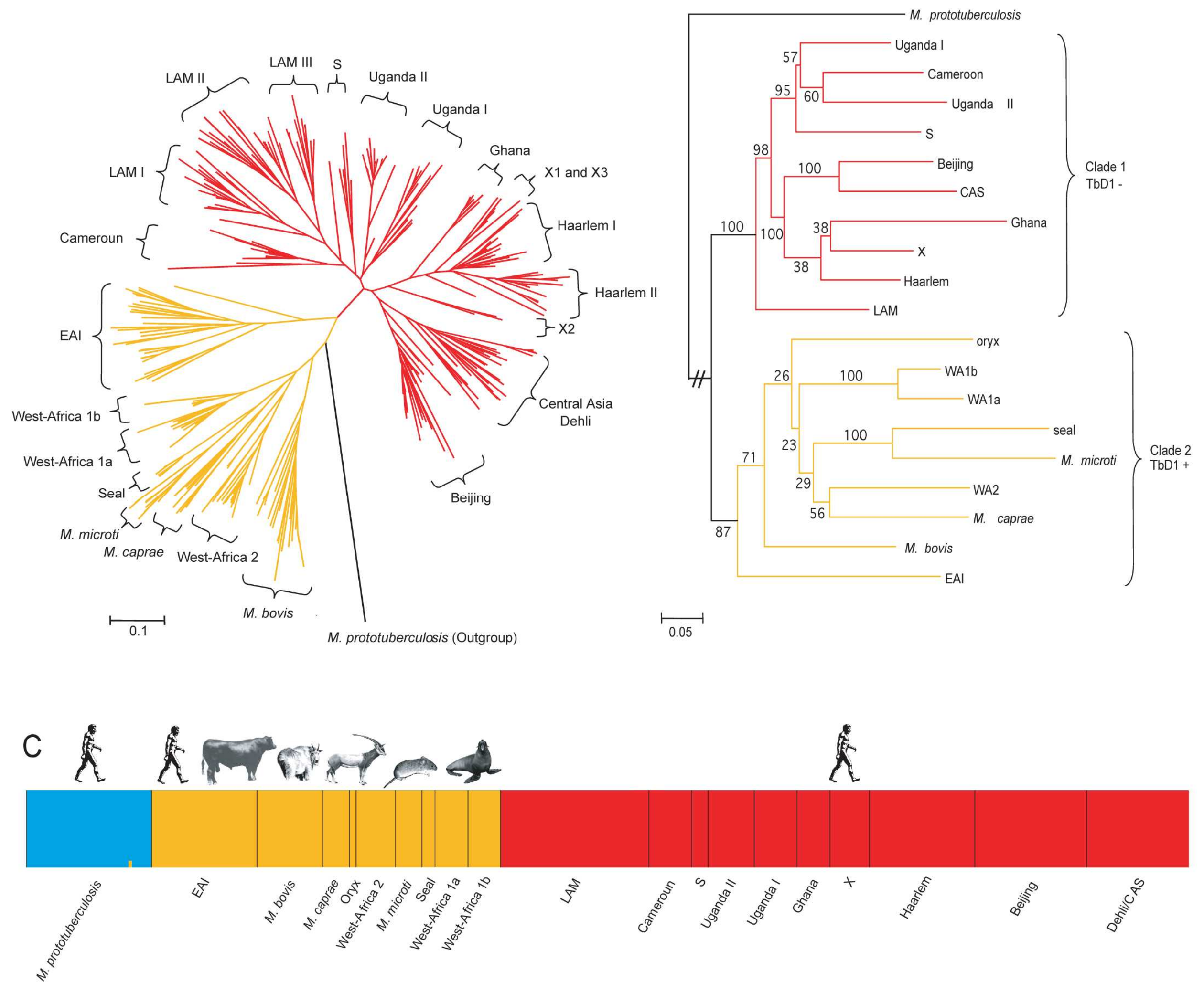

Figure 2-2: Evolutionary relationships of the MTBC as proposed by Wirth *et al.* 2008.

"(A) Unrooted MIRU Neighbour-joining phenogram depicting genetic distance relationships among tubercle bacilli isolates based on Nei et al.'s D_A distances. (B) Rooted MIRU population Neighbour-joining tree based on genetic distance. *M. prototuberculosis* was used as an out-group. Values on the nodes represent the percentage of bootstrap replicates over individuals (N = 1000) showing the particular nodes. Branch lengths are proportional to the genetic distance between the tubercle lineages. It is noteworthy that low bootstrap values within clade 2 prevent us from drawing further inferences on the branching order in this clade. WA, West-Africa. (C) Population structure of 20 MTBC clonal lineages using the no-admixture model, where $K = 3$. Each colour represents one cluster, and the length of the color segment shows the strains' estimated proportion of membership in that cluster. Results shown are averages over 10 STRUCTURE runs. For clarity, strains codes are also given according to Gagneux et al. (2006)" (Wirth *et al.*, 2008).

Figure 2-3: The evolution of MTB out of Mesopotamia, a scenario proposed by Wirth *et al*. 2008.

"The main migrations events are numbered and correspond to: 1, *M. prototuberculosis*, the ancestor of the MTBC, this bacterium reached the Fertile Crescent some 40,000 years ago by sea or land; 2 and 3, two distinct basal lineages arose, EAI and LAM and spread out of Mesopotamia some 10, 000 years ago; 4, 5 and 6, later on (8–5000 years ago) derived populations from clade 1 followed main human migration patterns to Africa, Asia and Europe, giving rise to locally adapted tubercle strains and further diversifications. Note that the depicted borders are "artificial" and are used for the demonstration. Global movements and intercontinental exchanges tend to blur this phylogenetic signal though strong enough to be detected nowadays" (Wirth *et al.*, 2008).

2.3 Molecular epidemiology of MTBC strains

Molecular epidemiological studies over the last decade described several typing techniques that gave rise to the development of a new discipline that has a huge impact on research and maybe future TB control programmes. A number of TB lineages were discovered of which the most prominent ones have been characterized in great detail. Several polymorphic genetic markers have been useful to discriminate or sub-speciate clinical isolates of MTBC*.*

2.3.1 Molecular markers and techniques

2.3.1.1 IS6110-based restriction fragment length polymorphism (RFLP) typing

DNA fingerprinting of TB using RFLP analysis based on the number and location of the IS*6110* insertion sequence has worldwide become the most useful and important tool to monitor the geographical diversity and spread of MTB strains, to determine transmission versus reactivation and to track outbreaks of MDR strains (Warren *et al.,* 1996, Bifani *et al.,* 1999, 2002).

2.3.1.2 Spoligotyping

Spacer-oligotyping ‘’spoligotyping” depends on the polymorphism of the Direct Repeat (DR) locus of MTBC strains containing multiple, well-conserved short repetitive DNA sequences (DRs) of 36 base pairs (bp) separated by non-repetitive spacer sequences of 34-41 bp (Hermans *et al*., 1991). Ninety four spacers have been identified of which 43 are commonly used for MTBC strain differentiation (van Embden *et al.,* 2000).

Spoligotyping has proven to be extremely valuable in molecular epidemiological studies for more than a decade. It is easy to perform and can be used to detect and type MTBC simultaneously and even directly from the clinical specimen (Kamerbeek *et al.,* 1997). A drawback of the current spoligotyping method is its limited discriminatory power for isolates with higher IS*6110* RFLP copies compared to that of IS*6110* RFLP typing (Kamerbeek *et al*., 1997, Kremer *et al*., 1999, van Soolingen *et al*., 1998). Fifty-one extra spacers representing spacers 44-95 described by van Embden *et al* and 10 oligonucleotides 95-104 described in the study of Caimi *et al* were included to the existing 43 spacers of the traditional method to develop a new spoligotying method evaluated by van der Zanden *et al* (van der Zanden *et al*., 2002). These new DRs showed to have a high differentiation and interpretability for MTBC (Caimi *et al*., 2001, van Embden *et al*., 2000, van der Zanden *et al*., 2002). Continuation of the traditional method was supported by van der Zanden *et al,* because this new method has a significant advantage only when used in areas populated with isolates with five or less IS*6110* copies and when extended discrimination is required (van der Zanden *et al*., 2002).

Spoligotyping has been found to be highly reproducible and reliable in discriminating MTB and *M. bovis* clinical isolates into different lineages (Kamerbeek *et al.,* 1997, Kremer *et al.,* 1999). Roring *et al* found spoligotyping 97% sensitive and 100% specific for simultaneous detection and typing of *M. bovis* from bovine tissue specimens (Roring *et al.,* 2000).

SPOTCLUST (Spoldb3.0), the first publicly available database on MTBC consisting of 535 spoligotyping patterns was developed in 2002 and provided a method to identify the similarity to MTBC isolates. Spoligotyping patterns can be entered as octal or binary format and results shown as probability as they are compared to 535 entries obtained between 1996-2004 from TB patients from New York (http://cgi2.cs.rpi.edu/~bennek/Run.html). By means of this, TB strains could be divided into families, sub-

families and or variants.

SpolDB4 is one of the largest publicly available databases on MTBC genetic polymorphism with a universal nomenclature system of spoligotyping. Spoligotyping patterns can be entered as octal or binary format and compared to 46286 entries obtain from 119 isolation countries in a global epidemiological information system (Brudey *et al.*, 2006a, http://www.pasteur-guadelope.fr:8081/SITVITDemo/?).

2.3.1.3 MIRU-VNTR typing

Most homologous pathogenic bacteria species such as *Bacillus anthracis*, *Yersinia pestis* and MTB contain thousands of tandem repeated sequences in their genome (Cox *et al.*, 1997, Cole *et al.*, 1998). Micro and minisatellites are distinguished based on the number of short sequence repeats (SSR) being between 1-13 bp and 10-100bp respectively (Tautz *et al.,* 1984).

Typing techniques looking at the variation in number of repeats of these SSR are referred to as VNTR (Variable number of tandem repeats) typing systems. For TB typing Multiple Loci VNTR Analysis (MLVA) looks for variation in elements known as mycobacterial interspersed repetitive units (MIRUs) consisting of homologous 40-100bp DNA sequences scattered in 41 locations throughout the genome of *Mycobacterium tuberculosis* H37Rv as shown on Fig. 2.4 (Supply *et al.,* 1997, Le Fleche *et al.,* 2002, Oelemann *et al.,* 2007). These MIRUs are inserted in the intergenic or non coding region of the MTBC chromosome (Supply *et al.,* 1997). Supply *et al* classified MIRUs into three categories, type I containing repeats of 77 bp, type II and type III having a gap of 24 bp and 15 bp at the 3' and 5' ends of type I sequences respectively. Locus 20 and locus 8 have shown to contain repeats of both type II/III MIRUs (Supply *et al.,* 1997). MIRUs contain an open reading frame which overlaps the stop and start codons of their flaking regions (Supply *et al.,* 2000).

Initially, 12 of the 41 MIRU loci displaying variation in tandem repeat copies were studied and found to have high discriminatory power to distinguish between MTBC clinical isolates and provide data that can easily be compared on-line (Supply *et al.,* 2000, Le Fleche *et al.,* 2002.).

MIRU copy number variability was evaluated among three strains namely *M. tuberculosis* H37Rv, *M. tuberculosis* CDC 151 and *M. bovis* AF212/97 (Supply *et al.* 2000). Nearly all MIRUs were present in all 3 species, except for the *M. bovis* strain which has lost loci 21 and the MIRU flanking regions. This region corresponds to the region of difference (RD) 9 or RD12, a region known to be deleted in *M. bovis* Bacillus Calmette Guérin (BCG) (Gordon *et al.*, 1999, Behr *et al.*, 1999). All other loci contained identical MIRU sequences except for six loci which contain variable MIRU copy numbers among the three strains. Loci 2 and 24 contain an additional type III and type I MIRU, respectively (Supply *et al.,* 2000). Locus 4 contains an additional 53-bp type II MIRU in 3' of 77-bp VNTR units in nearly all clinical isolates (Supply *et al.,* 2000).
Micro-minisatellites were used successfully as a powerful genetic element for evolutionary and also in population genetic studies in higher eukaryotes (Jeffreys *et al.*, 1991, Sutherland *et al.*, 1994). In a study by Mazars *et al*, VNTR approach based on the minisatellites was used and perfectly clustered all related isolates, indicating the stability of MIRU-VNTR (Mazars *et al.*, 2001). When compared with IS*6110* RFLP, MIRU-VNTR showed a high resolution, and the advantage of being faster. It also showed a high resolution power for strains with few IS*6110* copies or that are devoid of IS*6110*. Furthermore, results proved to be reproducible between independent laboratories (Supply *et al.,* 2000, 2001, Mazars *et al.*, 2001, Cowan *et al.*, 2002, Sola *et al.,* 2003). More studies have indicated that this method can efficiently be used to discriminate Beijing strains sharing the same spoligotyping patterns (Supply *et al.*, 2001).

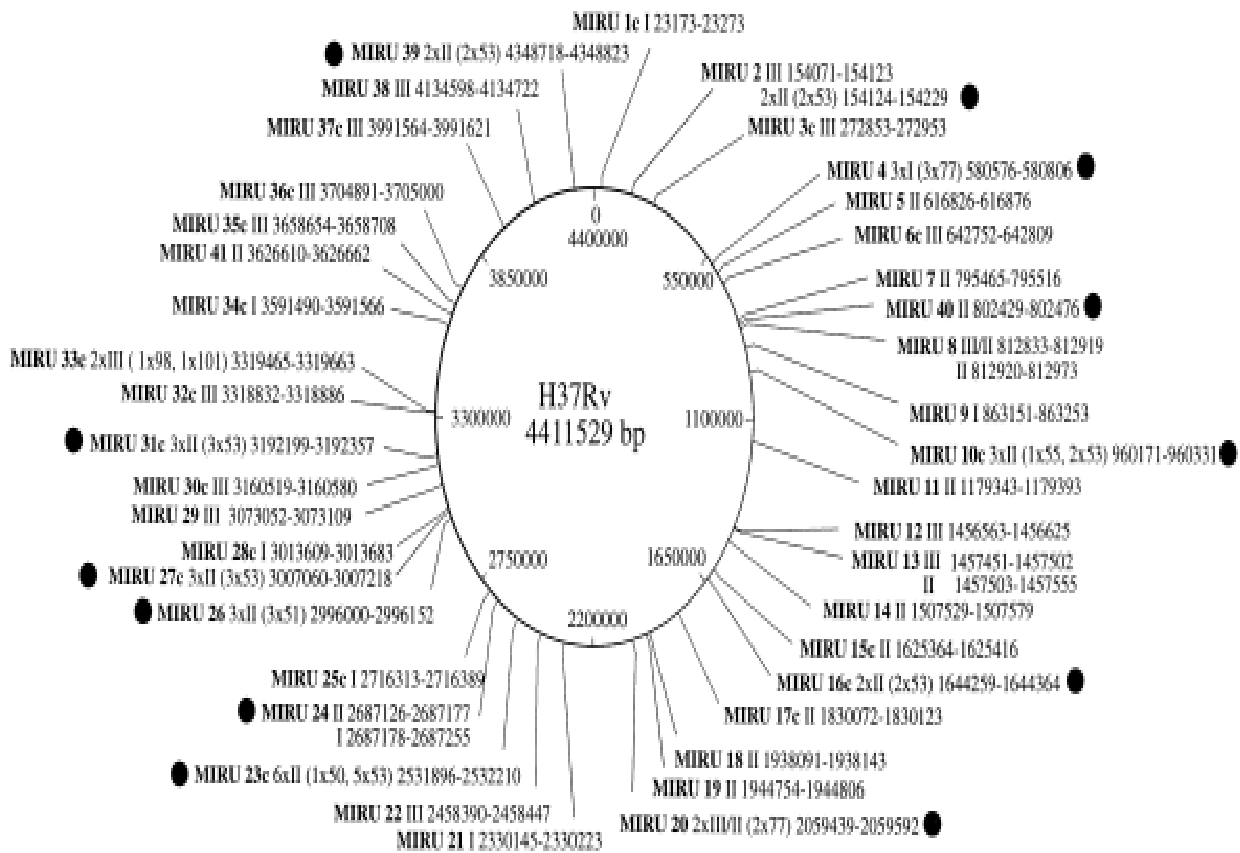

Figure 2-4: Position of the MIRU loci on the MTB H37Rv chromosome.
“Numbers in bold specify the respective MIRU locus numbers. The symbols 'c' designates that the corresponding MIRUs are reversely orientated to that defined by Cole et al. 1998. Roman figures give the type of MIRU (type I, II or III). The exact positions of the MIRU loci are given in numbers after the type numbers. The 12 loci containing variable numbers of MIRUs are indicated by black dots” (Supply *et al*., 2000).

In Singapore, MIRU-VNTR typing was able to differentiate between ancestral and modern MTB as described by deletion TbD1 analysis (Sun *et al*, 2004, Brosch *et al*., 2002). Brosch *et al* defined ancestral strains to contain TbD1 which is deleted in modern MTB (Brosch *et al*., 2002).

Among different sets of MIRU-VNTR loci described and suggested for typing MTB isolates (Roring *et al*., 2002, 2004, Skuce *et al.,* 2002; Le Fleche *et al*., 2002; Magdalene *et al*., 1998; Kremer *et al.,* 2005a; Frothingham *et al.,* 1998), the 12 MIRU-VNTR system is currently used worldwide and was even integrated in the national TB control system program of the United States of America (Allix *et al*., 2004; Cowan *et al*., 2005; Mazars *et al.,* 2001; Supply *et al.,* 2000). The discriminatory power of 12 MIRU-VNTR approached that of *IS6110*-RFLP typing in differentiating between epidemiologically unrelated isolates, while the genotyping based on this set is stable between epidemiologically linked isolates (Mazars *et al.,* 2001, Supply *et al.,* 2001, Blackwood *et al*., 2004, Hawkey *et al.,* 2003, Kwara *et al.,* 2004, Savine *et al.,* 2002). A recent population-based study has indicated that the use of 12 MIRU-VNTR as a first line method in combination with spoligotyping provides adequate discrimination in most large-scale populations, but a significant proportion of unrelated isolates remained falsely clustered (Cowan *et al,* 2005; Scott *et al,* 2005). Therefore, IS*6110*-RFLP typing is still needed especially where the contact investigation or demographic or epidemiological data do not provide independent clues for the existence or the absence of links between the patients (Cowan *et al.,* 2005; Blackwood *et al*., 2004)

To overcome this problem, alternative sets of MIRU-VNTR were investigated to further improve the discrimination of unrelated cases compared to that of 12-loci MIRU-VNTR typing (Kamerbeek *et al.,* 2006, Kremer *et al.,* 2005b, Le Fleche *et al.,* 2002, Roring *et al*., 2004, Supply *et al.,* 1997, 2006, Warren *et al.,* 2004). The resolution and stability of 29 loci were analyzed (Supply *et*

al., 2006). Five loci (3232, 3336, 2163a, QUB-1895, and QUB-18) were excluded from final selection due to lack of robustness and/or stability (Supply *et al.,* 2006). The clustering rate of the combined 24 loci set system and spoligotyping was decreased by fourfold and by threefold under the same condition (Oelemann *et al.,* 2007, Supply *et al.,* 2006). Finally, a discriminatory subset of 15 loci with the highest evolutionary rate was defined showing a concentration of 96% of the total resolution obtained by 24 loci and a predictive value equal to that of IS*6110*-based RFLP typing (Oelemann *et al.,* 2007, Supply *et al.,* 2006). Spoligotyping combined with MIRU-VNTR typing will be of great benefit, especially as a quick and convenient independent control (Supply *et al.,* 2006). Fifteen–loci MIRU-VNTR typing was proposed for epidemiological studies and the use of 24 loci mostly for phylogenetic studies (Supply *et al.,* 2006).

In 2007-2008 an international free accessible MIRU-VNTR database was established by Allix-Beguec and colleagues, to compare strains based on spoligotyping, MIRU-VNTR's, region of difference (RD), single nucleotide polymorphism, susceptibility testing or by a combination of different data types (www.miruvntrplus.org, Allix-Beguec *et al.,* 2008). Data can also be compared to 122 MTB reference strains belonging to different lineages including the W/Beijing, Cameroon, Delhi/Central Asia, East African-Indian, Ghana, Latin American-Mediterranean, Turkish, S, Uganda I and II, Ural, and X lineages. Reference strains for *M. africanum, M. canettii* (*M. prototuberculosis*), *M. bovis, M. caprae, M. microti,* and *M. pinnepedii* are also included adding up to 186 strains in total (Allix-Beguec *et al.,* 2008).

2.3.2 Principal lineages

Apart from DNA-fingerprinting methods based on repeated DNA sequences, TB isolates can also be typed or grouped based on single nucleotide polymorphisms (SNPs) in specific genes or in the whole genome. In a study of 842 MTBC isolates recovered from diverse geographical locations 3 evolutionary groups were proposed based on single nucleotide polymorphism in the *katG* gene codon 463 CTG Leucine (Leu) and *gyrA* gene codon 95 ACC Threonine (Thr) (Sreevatsan *et al*., 1997).

Genotype group1 isolates are ancestors to the other two groups because they link predominantly animal pathogens *M. microti* and *M. bovis* to human pathogens *M. africanum and* MTB. The allelic *katG* and *gyrA* gene codon combinations are shown in Fig. 2.5 (Sreevatsan *et al*., 1997). They also showed that the IS*6110* copy number frequency distribution among the 3 groups was different and there was little sharing of IS*6110* profiles among them (Sreevatsan *et al.,* 1997). This evolutionary grouping will help in determining the history of lineages based on the proposed groups by Sreevatsan and colleagues (Sreevatsan *et al.,* 1997).

Based on the combined use of the above mentioned DNA-fingerprinting techniques and genome sequence analyses various major lineages amongst the MTB complex have been identified.

Kato-Maeda *et al.* have designed Fig. 2.6 to show the non-randomness of deletions in 16 clinical isolates. These isolates were tested by micro-array against the H37Rv genome (Kato-Maeda *et al*., 2001). Different isolates contained unique deletions whereas other deletions were shared by many isolates. The study of Kato-Maeda was extended to 100 different and unique IS*6110*-based RFLP types representing the global genetic diversity of the MTBC observed in San Francisco over 12 years (Tsolaki *et al*., 2004). Long sequence polymorphism (LSP) size varied between 105 and 11,985 bp, with

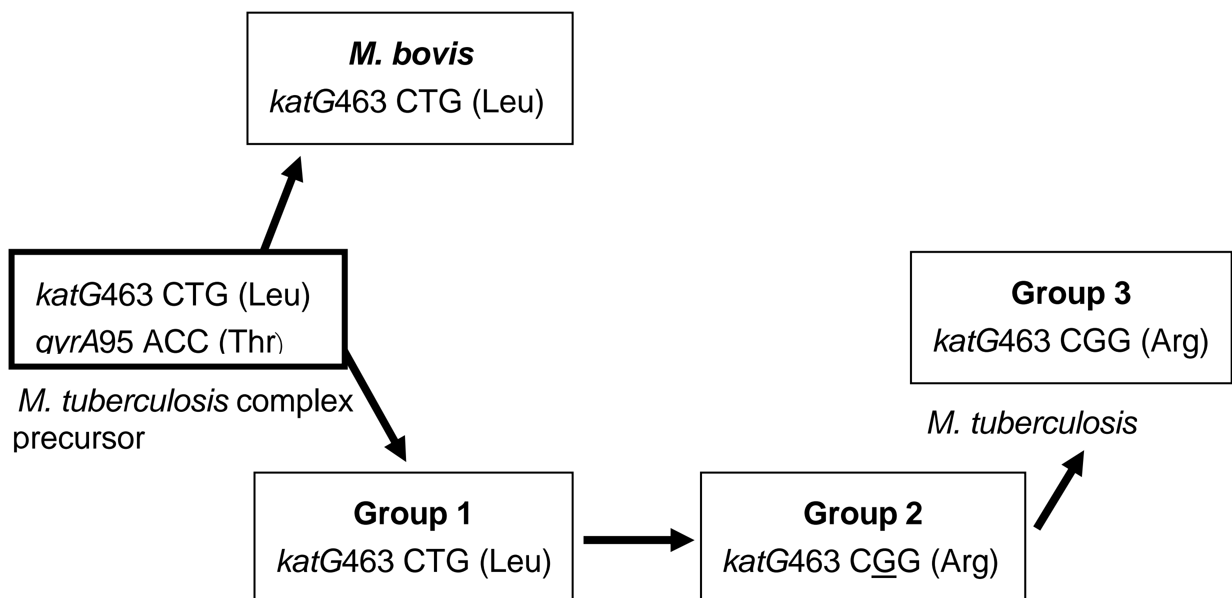

Figure 2-5: Proposed evolutionary groups of MTB.

The precursor of *M. tuberculosis* complex was characterised by *kat*G gene codon 463 (Leu) and *gyr*A gene codon 95 (Thr). Strains corresponding to this combination were defined as group 1. Group 2 depict a point mutation in the *kat*G gene codon 463 changing Leu to arginine (Arg) and Group 3 the same mutation in the *kat*G codon plus a point mutation in the *gyr*A95 gene changing Thr to serine (Ser).

eight deleted sequences larger than 5,000 bp. LSPs tend to occur in genomic regions that are prone to repeated insertion-deletion events and may be responsible for a high degree of genomic variation in the MTBC (Alland *et al*., 2007).

2.3.2.1 The Beijing lineage

The Beijing genotype belongs to the principal genetic group 1 as characterized by Sreevatsan *et al*, and its specific spoligotype signature (absence of spacer 1-33, presence of spacer 35-43) was discovered in 1995 (van Soolingen *et al.,* 1995). A notorious outbreak in New York at the beginning of '90s was due to a MDR clone of one of its offspring (W strain) that had been characterized before (Plikaytis *et al*., 1994, Bifani *et al*., 2002). It was hypothesized that this genotype emerged successfully in East Asia due to mass BCG vaccination during the 20th century (van Soolingen *et al*., 1995, Abebe *et al*., 2006). The Beijing lineage was also considered as a group of variant clones evolving from a common ancestor, maybe during the Genghis Khan reign or before (Mokrousov *et al*., 2005).

Beijing strains are further characterized by the presence of an inverted IS*6110*-based RFLP copy within the direct repeat (DR) region, an IS*6110* element at a particular insertion site and one or two IS*6110* copies in a DNA region called the intervening region (NTF) (Plikaytis *et al*, 1994, Kurepina *et al*., 1998). A characteristic Beijing lineage-defining SNP (G81A in Rv3815c) has been reported, and according to the SNP analysis, the Beijing cluster was designated as SCG 2 or single SNP-II (sSNP) (Filliol *et al*., 2006, Gutacker *et al*., 2006).

Other characteristic SNPs of the Beijing lineage were described in putative DNA repair genes (Rad *et al*., 2003). More recently, new phylogenetically-informative specific LSP markers were found, such as RD105, which is present in all Beijing/W or RD142, RD150 and RD181. It allows a further

division of the Beijing lineage into four monophyletic subgroups (Tsolaki *et al*., 2005). Some researchers named Beijing lineage the East Asian Lineage (Gagneux *et al*., 2006). Its most frequent variable number of tandem repeats (VNTR) signature is 42435 (Kremer *et al*., 1999).

2.3.2.2 The East African-Indian (EAI) lineage

This lineage was first described in Guinea-Bissau (Källenius *et al*., 1999) and was shown to be frequent in South-East Asia, India, and East Africa (Kremer *et al*., 1999). EAI strains are characterized by a small number (less than 6) of IS*6110*-based RFLP copies. A subgroup of these strains gaining a single copy of IS*6110* was shown to be widespread in areas like in Malaysia, Tanzania, and Oman (Fomukong *et al*., 1994). This lineage demonstrated congruence between spoligotypes (absence of spacers 29-32 and 34, presence of spacer 33), VNTR [exact tandem repeat A (ETR-A) allele ≥4], *kat*G*gyr*A grouping (Group 1), and later the presence of the TbD1 sequence (Filliol *et al*., 2006, Gutacker *et al*., 2006, Soini *et al*., 2000, Sola *et al*., 2001). Some believe that this lineage, which is endemic in South-East Asia, South-India, and East-Africa, may have originated in Asia, where TB could have historically found favorable spreading conditions (Soini *et al*., 2000, Filliol *et al.,* 2002).

2.3.2.3 The Manila lineage

The Manila family was first identified by Douglas in 1997, and was later thoroughly characterized by the same group (Douglas *et al*., 2003). The Manila family was identified based on the prevalence of clustered strains isolated from Philippino immigrants in the United States (US) and was later shown to be prevalent in the Philippines. The Manila family is identical to

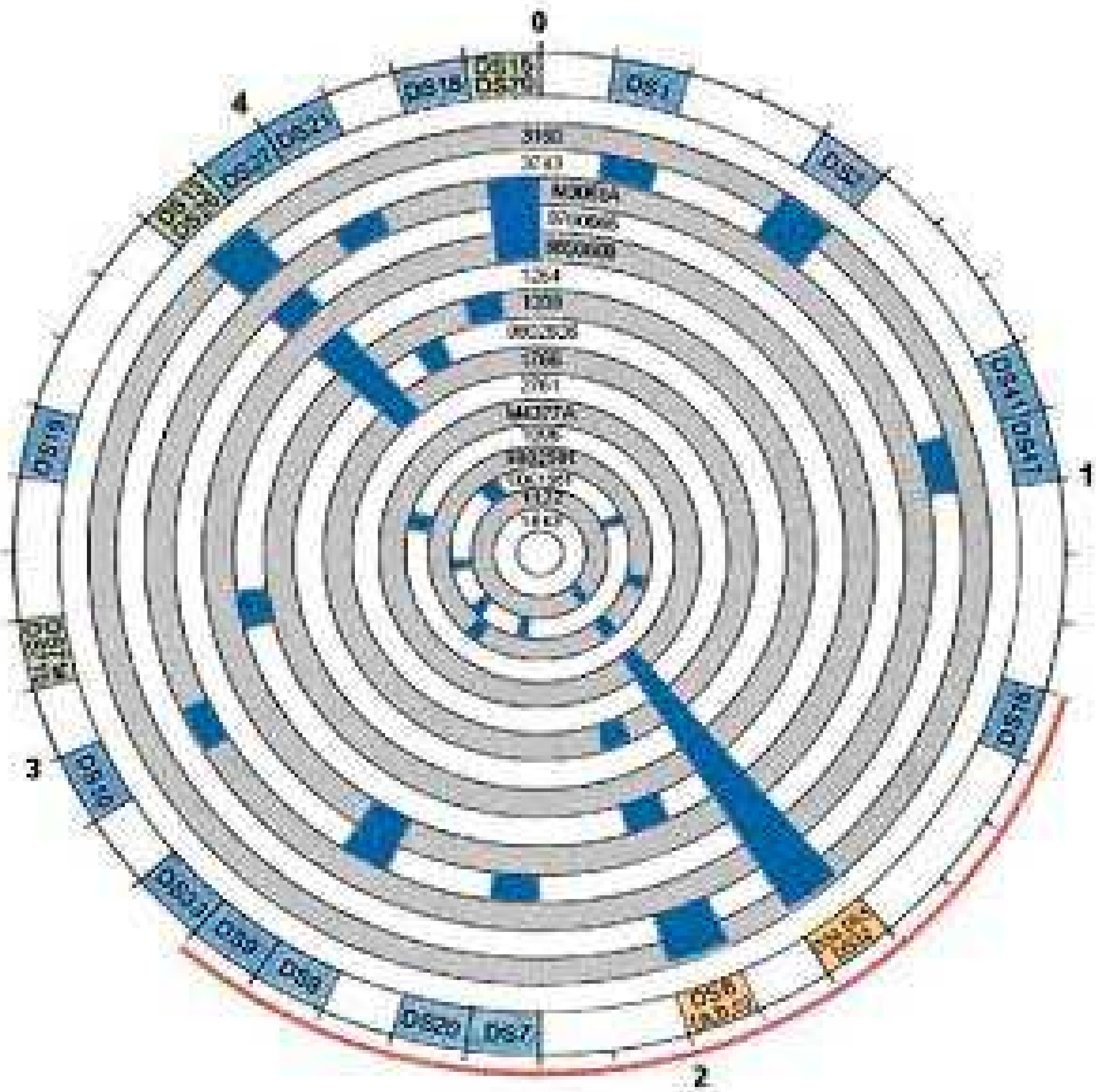

Figure 2-6: Circular diagram representing genomic deletions in 16 clinical MTB isolates tested by Kato-Maeda *et al*. 2001.

"The diagram show that deletions differ among *M. tuberculosis* clones and is not spatially random. The outer numbers represent the scale in mega base pairs, 0 represent the origin of replication. In blue: shows genomic area of deleted sequences. The outer circle represents combination of all detected deletions. The following colors blue, orange and green are linked to number of deletions (1, 2, and 3) respectively. The thin red line spans the genomic region of the genome where the number of deletions detected is greater than expected by chance alone. *M. tuberculosis* CDC 1551 appears as the third ring on the above picture" (Kato-Maeda et al., 2001).

subtype EAI-2 and gives a shared international type (SIT) spoligo-signature (SIT19) (Filliol *et al.*, 2002). SIT89, which defines the Nonthaburi (Thailand) group of strains, is a derived clone of the Manila family (Namwat *et al.*, 1998). In this family, specific variants have been described for Madagascar (SIT109, EAI-6), Bangladesh (SIT591, SIT1898 or EAI-6 and 7) and Vietnam (SIT139 or EAI-4) (Brudey *et al.*, 2006a).

2.3.2.4 The Central-Asian (CAS) or Delhi lineage

The presence of a specific lineage of the MTBC complex in India was independently reported by two different groups of researchers using IS*6110*-based RFLP and spoligotyping, respectively (Bhanu *et al.*, 2002, Filliol *et al.*, 2003). This lineage was later shown to be endemic in Sudan, Pakistan and other sub-Saharan countries (Brudey *et al.*, 2006a). Using IS*6110*-based RFLP typing, the Delhi lineage shows a characteristic band pair in the high molecular weight region, whereas its specific spoligotype signature is characterized by the absence of spacers 4-7 and 23-34. This spoligo-signature shows numerous variants and several subgroups such as CAS1-Kili (for Kilimanjaro) and CAS1-Dar (with absence of spacer 2, for Dar-es-Salaam) have been defined on the basis of new spoligotype-signatures that are specific for each new clonal group (Mc Hugh *et al.*, 2005, Eldholm *et al.*, 2006). VNTR signatures of clinical isolates from South-Asian immigrants in London and native patients in Rawalpindi, Pakistan, were identical (allele combination 42235) and correlated with the CAS spoligotype (Gascoyne-Binzi *et al.*, 2002). The CAS genotype could be the ancestor of the Beijing family since they cluster closely when analyzed by a combination of MIRU-VNTR and spoligotyping (Sola *et al.*, 2003). In India, its distribution varies from one region to another; it is more prevalent in the North than in the South, where the EAI family predominates (Suresh *et al.*, 2006).

2.3.2.5 The Haarlem family

The Haarlem family was described first in the Netherlands in the late 90's (Kremer *et al*., 1999). On IS*6110*-based RFLP typing, these strains result with a double band at 1.4 kb. Their spoligotyping pattern is characterized by the absence of the spacer 31, which is due to the presence of a second IS*6110* copy in the DR region (Groenen *et al*., 1993). Due to an asymmetric insertion within the DR locus, this second IS*6110* copy hinders the detection of spacer 31 (Filliol *et al*., 2000, Legrand *et al*., 2001). Three main spoligotype-signatures define the variants H1 to H3 (Filliol *et al*., 2002). Another characteristic of the Haarlem lineage is the frequent VNTR 5 loci pattern 33233 (Kremer 1999). A SNP in the *mgt* gene of the Haarlem genotype was discovered recently (Alix *et al*., 2006).

The Haarlem family is highly prevalent in Northern Europe, but is also present in the Caribbean, although to a lesser extent, and in Central Africa, where it is believed to have been introduced during the European colonization (Filliol *et al*., 2003).

2.3.2.6 The Latin American and Mediterranean (LAM) family

The LAM family is one of the most prevalent and widely distributed lineages in the world (Lazzarini *et al*., 2007, Brudey *et al*., 2006a, 2006b).

The LAM family is defined by the linkage disequilibrium between the absence of spoligotype spacers 21-24 and the presence of an MIRU-VNTR ETR-A allele equal to 2 (Sola *et al*., 2001). However, this genotype family with subtypes LAM1 to LAM12, characterized according to the latest international spoligotypes database project SpolDB4 (Brudey *et al*., 2006a) is more complicated than initially thought. The phylogenetic significance of the common absence of spacers 23-24 has also not been demonstrated in this lineage. Some genotypes that show strong geographical specificity (for example the LAM10-Cameroon or the LAM7-Turkey) were initially classified

as LAM, although there is no evidence of their phylogenetical relation to other LAM spoligo-signatures (Niobe-Eyangoh *et al*., 2003, Zozio *et al*., 2005). Strains belonging to LAM3/F11 and S/F28 was reported individually, but harbor identical spoligotypes (ST4), indicating that they might be the same (Warren *et al.,* 2002).

The LAM clade is frequent in Mediterranean countries and its presence in Latin America is supposed to be linked to the Lusitanian-Hispanian colonization of the New World. Conversely, it may have been endemic in Africa and/or in South America, spreading to Europe later (Arriaza *et al*., 1995, Salo *et al.,* 2001).

In the Ouest province of Cameroon it was reported that 47% of their MTB isolates belong to the Cameroon family (LAM10-Cam) (Niobe-Eyangoh *et al*., 2004) and in Dar es Salaam CAS (missing spacers 4-7, 10 and 20-35), LAM and EAI were the abundant families respectively (Eldholm *et al*., 2006). In Harare Zimbabwe, 32% of their MTB isolates are LAM-ZWE variants (Easterbrook *et al*., 2004).

Paleopathological and ancient DNA data support the existence of TB in Latin America before the arrival of Spanish settlers to Latin America in the 15th century (Arriaza *et al*., 1995, Salo 2001).

Recently, a specific deletion designated as RD^{Rio} was shown to be linked to certain LAM spoligo-signatures present in Rio de Janeiro, Brazil (Lazzarini *et al*., 2007). It was found that all RD^{Rio} strains from Rio de Janeiro, Brazil belong to the LAM-9 family characterized by absence of spacers 21-24, and 33-36, including sub-families originating from Latino-American Mediterranean countries (Lazzarini *et al*., 2007).

2.3.2.7 The X family: the European IS6110 low banders

The X family of strains is defined by two correlated features: a low number of IS*6110* copies and the absence of spacer 18 in the spoligo-signature

(Sebban *et al.*, 2002). The latter is indeed an important characteristic common to at least three SIT spoligo-signatures: SIT119, SIT137, and SIT92. Both characteristics are present in the CDC1551 strain, which was once suggested to be highly virulent. The X family was the first group identified in Guadeloupe (Sola *et al.*, 1997) and the French Polynesia (Torrea *et al.*, 1995). Specific epidemic variants of the X family were described in the Western Cape Province of South Africa (Streicher *et al.*, 2004). The absence of spacer 18 bears phylogenetical significance because it is improbable that this spacer was deleted more than once in the evolution of MTB. The distribution of the X family appears to be linked to Anglo-Saxon countries (Dale *et al.*, 2003). It is also highly prevalent in South Africa (e.g. KZN) and to a lesser extent in the Caribbean (Soini *et al.,* 2000, Brudey *et al.,* 2006a).

2.3.2.8 The T family and others

The T group (the modern MTB strains) is characterized by default. It includes strains that miss spacers 33-36 and can hardly be classified in other groups. The T family has general characteristics of strains belonging to the principal genetic groups 2 and 3, together with the absence of an intact *pks* 15/1 gene (Marmiesse *et al.*, 2004). The presence of intact polyketide synthase genes, active in the synthesis of the specific lipid complex of the MTBC is now known to be linked to virulence (Constant 2002).

Alternatively the 7 bp frameshift deletion in *pks*15/1 may be considered as a phylogenetical marker specific for the modern MTB strains (Gagneux *et al.*, 2006) and may define the recently designated Euro-American lineage (Gagneux *et al.*, 2006).

Gagneux *et al.* suggest that West African 2 diverged from an ancestral branch of *M. bovis,* whereas West African 1, characterized by a deletion of RD711, did not (Gagneux *et al.*, 2006).

2.3.2.9 Prevalent genotypes in South Africa

On a three year study conducted in the Western Cape, various strains of TB that are new in South Africa were identified including DRF 150, which has never been identified globally before. DRF 150 was found to be resistant to four of five drugs used to combat TB. Isolates of this new strain have also been identified in the Northern Province, Mpumalanga Province and Nairobi, Kenya. DRF 150 was found to be different from the deadly W-Beijing strain. The prevalence of DRF150 is not known (Tuberculosis, New multidrug-resistant strain-South Africa 2003, Promed-mail).

Using the spoligo databases, it was possible to successfully classify the emerged XDR-TB strain in Kwazulu-Natal province of South Africa as the KZN strain (Pillay *et al.,* 2007). After classification of the KZN MTB strain they found that it belongs to the F15 family and LAM4 which form part of the (LAM) family as they posses similar spoligo patterns characterised by the absence of spacers 21-24, 33-36, and 40. This variant is recorded in the SPOLCLUST3 database as SIT 60 (Pillay *et al.,* 2007, Streicher *et al.,* 2007), and has been reported in 12 European countries and North and South America (Filliol *et al., 2003*).

Spoligotyping was also used to show the importance of age shift in Western Cape South Africa, where scientists found that the prevalence of the Beijing genotype in children seems to differ from that described in adults from the same geographic area (Nicol *et al.,* 2005, Marais *et al*., 2006). Beijing and LAM3/F11 predominates in children (Marais *et al*., 2006) while Haarlem and W/Beijing strains predominates in adults (Victor *et al*., 2004). Mostly children tend to progress to disease within 12 months of primary infection and it has been reported that children provide valuable epidemiologic perspective, as the burden of childhood TB reflects the level of the epidemiologic control achieved within a particular area of that specific community (Marais *et al.,* 2004*, 2005,* 2006b).

2.4 Demographic setting and history of the FS and study area

The FS province is found in the heart of South Africa, with Lesotho nestling in the hollow bean-like shape (Fig. 2-7). The province is the third largest of nine provinces, taking up 10.6% of South Africa's land with a total area of 129 825 square km. The province is divided into five districts (Fezile Dabi, Xhariep, Thabo Mofutsanyane, Motheo and Lejweleputswa) (Fig. 2-7) and consists of 20 towns. Bloemfontein is the capital city, and other important towns include Welkom, Odendaalsrus, Sasolburg, and Phuthadithjaba to mention a few. Its population was estimated 2.9 million people in 2006. The majority of people speak Sesotho, followed by Afrikaans and a sprinkling of isiXhosa. Mining is the province major employer, with 12 gold mines, producing 30% of South African's output. (www.nationsonline.org/oneworld/south_africa_provinces.htm#Free%20State).

Gold was discovered in South Africa in 1892 but because of financial constraints little mining could be done. The Second World War also played a part in delaying this exploration, and it was only in 1946 when more reefs were discovered that the Goldfields mine was established (Fig. 2-7). Welkom was only founded in 1948. In 1996 the population size was 450 176 which was 15% of the overall population of the FS. Eighty three percent of the FS population consists of black people, 5% of white people and 2% of both Asians and coloureds. Fifty six percent of the overall population were males and 44% females. This was mainly because of the high number of men working as migrant labour from different parts of the country as well as other neighbouring states to provide a better life for their families and leaving their wives behind. People from farms moved to townships for the same reason and that put pressure on the infrastructure and resources available in small towns such as Allanridge, Odendaalsrus and Virginia (www.

nationsonline.org/oneworld/south_africa_provinces.htm#Free%20State).

Qwaqwa is part of the Thabo Mofutsanyane district municipality (DC19). The study area comprised of the local municipality Maluti a phofung (FS194) (Fig. 2-7). In 1974 Qwaqwa became a self-governing homeland, within a period of 30 years a large influx of people came in to settle here, as the area was declared an industrial growth point and also because of forced resettlements by the then nationalist government. As a result this area became overpopulated straining on the natural resources that were already limited. The population size was 303 781 in 1996 and this contributed to 10% of the overall FS population. Ninety nine percent of the population was black and 1% whites, Asians and coloureds. This area has more women (54%) than men based on the fact that the men moved away from the area to seek employment elsewhere (Cronje *et al.,* 2006). Seventy four percent of the entire population is unemployed (Cronje *et* al., 2006).

Thaba Nchu (Fig. 2-7) is a former homeland to the Basothos, but joined by the Batswanas in 1833 before the arrival of Vootrekker families, in 1936. It is situated 60 km from Bloemfontein, the industrial capital of the Free State. People only started to settle in Thaba Nchu when Selosesha and Botshabelo, the two neighbouring areas, were acknowledged as industrial growth points. People were also forced to move from the former Bophuthatswana (a different ethnic homeland) to their own ethnic homeland areas in Thaba Nchu. In 1996 the population size was 77 208 people which was only 3% of the overall Free State population (www.nationsonline.org/oneworld/south_africa_provinces htm#Free%20 State).

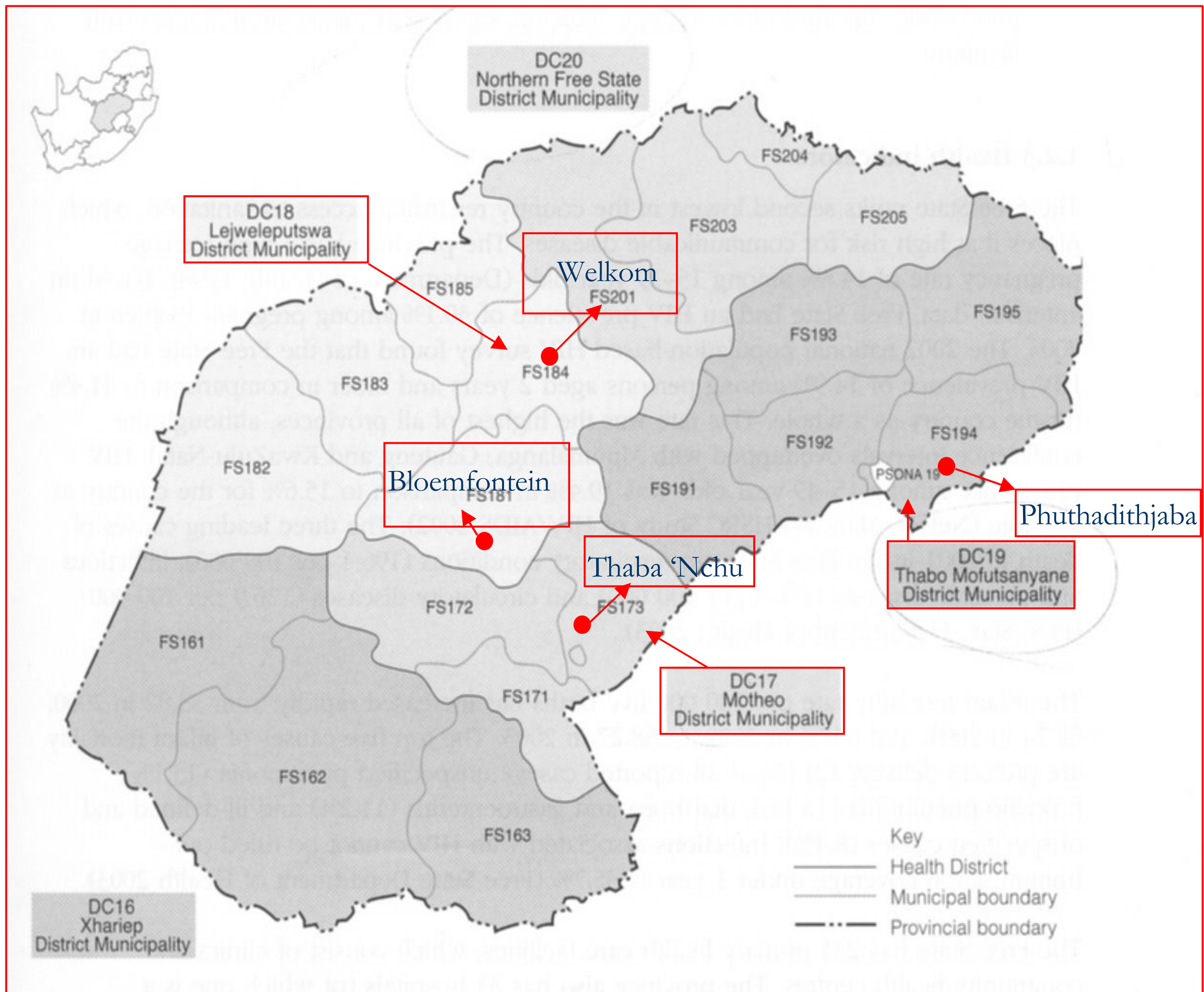

Figure 2-7: Location of the three areas where *M. tuberculosis* isolates were collected.

Chapter 3

Materials and methods

3.1 Sample

The study was performed on 126/174 and 15/27 MTB isolates that were typed using RFLP fingerprinting during two previous studies in Bloemfontein (2001-2004 respectively). The 126 isolates were cultured during a collaborative study with Belgium (ETOVS no: 01/74) and 15 INH resistant FS isolates that were cultured and kindly provided by the National Tuberculosis Lead Research Programme, Medical Research Council (MRC), Pretoria for another former study (van der Spoel van Dijk *et al.*, 2005, Mokhethi *et al.,* 2005). The 126 isolates were collected in the FS over a period from June 2001 to June 2003 from smear-positive pulmonary TB (PTB) patients in 3 high-burden clinics in each of the districts: Motheo (DC17) in the town Thaba 'Nchu (60), Thabo Mofutsanyane (DC19) in the town Phuthadithjaba (the former homeland Qwaqwa) (31) and from the Thabong clinic in the town Welkom (35) in the Lejweleputswa (DC18), also known as the Goldfields (Fig. 2-7). The 15 INH-resistant isolates used had been collected during May 2001-April 2002 from pulmonary tuberculosis suspects from mostly the same clinics during a nation-wide TB drug resistance monitoring study of the MRC.

All enrolled patients were subjected to written consent. During collection of sputum for the first study, additional sputum was requested from the patients after TB had been confirmed, but before treatment was initiated, because culturing of mycobacteria in new cases is not routinely done for the diagnosis of TB in the FS. In the study from where the 15 INH resistant strains originated, extra sputum was collected from the patients with the sputa collected for diagnosis.

The characteristics of the 141 isolates used in this study are shown in Table 3.1. Ninety six of the 141 (68%) isolates were from patients infected with TB for the first time (new cases) and 46 were from patients treated for at least one month at a previous occasion (retreatment cases). Of these

Table 3.1. Treatment history, gender and age distribution of patients.

	Male	**%**	**Female**	**%**	**Total**
New cases	56	58.9	39	41.1	95
Retreatment	33	71.7	13	28.3	46
Total	**89**		**52**		**141**
			Age(years)		
18-30	21		18		39
31-40	31		20		51
41-50	23		10		33
>50	14		3		17
Unknown			1		1
Total					**141**

patients 87% were within the age group 20 – 50 and 75 (53.2 %) of the 141 patients were men. The case, gender and age distribution correlates well with the statistical analysis of the former study where these parameters were used to prove that the isolates analysed were representative samples of TB cases of the Free State.

DNA of the 141 isolates was extracted and subjected to RFLP typing during the previous studies resulting in good RFLP patterns and their samples still had sufficient DNA left for the current study.

3.2 Isolation of mycobacteria

Strains used during this study were previously isolated using the following methods. The decontaminated specimen (10 µl) was inoculated onto a 1.2% (v/v) glycerol containing LJ slant medium. The LJ slopes were incubated in a slanted position for at least 24 hours at 37°C. All cultures were examined after 5-7 days of incubation and weekly thereafter for 4-6 weeks. Caps were opened once a week for a short interval to aerate and to examine bottles for positive growth. ZN stain, nitrate and catalase tests were performed for MTBC identification. All MTB isolates *were* stored at -20°C.

3.3 Identification of the isolates

3.3.1 Ziehl-Neelsen (ZN) staining

ZN staining was performed on the decontaminated sputum specimen for case detection or from growth on a solid LJ slant to confirm the acid-fast character of the observed growth. Carbol Fuchsin stained the

M. tuberculosis red, while 5% acid alcohol was used to decolorize and Lofflër methylene blue was used to counter stain. The slides were examined under the 100x oil immersion objective of a microscope for the presence of acid fast organisms.

3.3.2 Nitrate test

Two millilitres of nitrate buffer in screw-capped tubes were inoculated with 2 loops full of bacilli, whereas a negative control tube did not contain bacilli. The contents of the tubes were mixed and incubated at 37°C for 4 hours. After incubation, a drop of HCl, 2 drops of 0.2% sulphanilamide, and 2 drops 0.1% *N*-naphthylethylene-diamine were added. The solution was examined for the development of a reddish/pink colour contrasting with the negative control.

3.3.3 Catalase test

Two loops full of bacilli were suspended in 0.5 ml Sorenson's buffer in screw-cap tubes (16 mm by 125 mm), whereas no bacilli were added to the negative control. The tubes were placed in a 68°C water bath for 20 min and then cooled to room temperature. Half a millilitre each of 10% Tween 80 and 30% H_2O_2 were added and the production of bubbles indicated a positive catalase test. Tubes were allowed to stand for 20 min before being regarded as catalase negative. Negative results shows that the isolate is not MTBC. The blank also have to be negative (Metchock *et al.*, 1995).

3.4 DNA Extraction

For chromosomal DNA preparation, the chloroform/phenol-extraction method described by Warren *et al.* 1996 was used (Warren *et al.*, 1996).

After heat inactivation of the growth on LJ slants at 80°C for an hour, DNA was extracted in a P2 safety cabinet. The growth from a slant culture was suspended in 6 ml extraction buffer (5% monosodium glutamate, 50 mM Tris-HCl, pH 7.0 and 25 mM EDTA) in a 50ml polypropylene tube containing approximately thirty 5 mm glass balls. The bacterial clumps (50 mg/ml) were disrupted by vigorous shaking and vortex mixing. Five hundred microlitres of lysozyme (50 mg/ml) (Amersham Biosciences, Greece) and 10 μl of RNAse A (10 mg/ml) (Amersham Biosciences, Greece) were added in a tube. The contents of the tube were mixed by inverting, and then incubated at 37°C for 2 hours. After incubation, 600 μl of 10x Proteinase K buffer and 150 μl of Proteinase K (10 mg/ml) (Amersham Biosciences, Greece) were added. The tubes were gently mixed by inverting several times and then incubated overnight at 45°C. Proteins were removed by phenol/chloroform (5 ml) and the chloroform/isoamyl-alcohol (5 ml) extraction. DNA was then precipitated by addition of 600 μl 3 M sodium-acetate ph 5.5 and 7 ml of cold (-20°C) isopropanol. The precipitated DNA was collected on a glass loop and washed with 1 ml of 70% ethanol for approximately 1 min. The DNA was air dried and dissolved in TE buffer (10 mM Tris, 1 mM EDTA, ph 8.0) and stored at -20°C until further use.

The DNA concentrations of the extracts were determined by spectrophotometry using a NanoDrop® ND-100 Spectrophotometer v3.01 (NanoDrop Technologies Inc).

3.5 Restriction fragment length polimorphism (RFLP)

Restriction fragment length polymorphism (RFLP) analyses were performed based on the standardised IS*6110* technique involving *Pvu* II restriction, gel electrophoresis, Southern blotting and chemiluminescent detection of IS*6110* fragments (van der Spoel van Dijk *et al.*, 2005, Van Embden *et al.,* 1993).

Bands of the scanned autoradiographs (Hewlett Packard Scanjet II CX/T) were normalised and analysed using the Dice unweighted pair group method with arithmetic averages (UPGMA) (GelCompar, version 4.1; Applied Maths, Kortrijk, Belgium) (van der Spoel van Dijk *et al.*, 2005).

3.6 Spoligotyping

All MTB isolates were spoligotyped to detect the 43 known spacer sequences in the DR section using a commercially available kit (Isogen Bioscience BV, Maarsen, The Nederlands) according to the manufacturer's instructions (Kamerbeek *et al.,* 1997) as briefly described below.

3.6.1 DNA amplification

Two microlitres (20-30 ng/µl) of extracted DNA of the isolates was added to 50 µl of PCR mixture consisting of a final concentration of 1.5 mM $MgCl_2$ solution (AB gene, USA), primers DRa and DRb (20 pmol/µl)(Isogen life Science, Maarsen, The Netherlands), 1x Tth buffer (New England Bio-labs inc, UK), 0.5 U/µl Tth polymerase (New England Bio-labs inc, UK), 2.5 mM deoxynucleotide triphosphate (dNTP's) (AB gene, USA) and the volume

was filled up with commercial Sabax water.

The reaction was carried out using a Perkin-Elmer 9600 cycler starting with a denaturation step of 3 min at 96°C. After denaturing, the PCR was performed for 30 cycles of denaturation for 1 min at 96°C, annealing at 55°C for 1 min and extension at 72°C for 30 sec. The final extension step was done by incubating for 5 min at 72°C. Purified genomic DNA of *M. tuberculosis* H37Rv and *M. bovis* supplied with the kit were used as positive controls. The negative controls consisted of PCR mixtures without mycobacterial DNA that were handled the same way as the isolates.

3.6.2 Hybridization with PCR product and detection

Biotin-labeled PCR products were hybridized to the immobilized spacer-oligos using a miniblotter as follows: All buffers were pre-warmed to the required temperatures according to the protocol. Twenty microlitres (20 µl) of the PCR product was added to 150 µl 2x SSPE/0.1% SDS, followed by heat denaturation of the diluted PCR product for 10 min at 99°C and immediate cooling on ice.

The membrane and support cushion were placed into the miniblotter, in such a way that the slots are perpendicular to the line pattern of the applied oligonucleotides. Residual fluid (the membrane must always be kept wet) was removed by pipetting from the slots. The slots where filled with diluted PCR product (avoiding air bubbles) and hybridized for 60 min at 60°C on a horizontal surface without shaking. Masking tape was used to cover the slot ends to avoid contamination. If less than 45 samples were used, the neighbouring slots were filled with 2x SSPE/0.1% SDS to prevent cross flow.

After the elapse of 60 min, samples were removed by pipetting from the slots and the membrane was removed from the miniblotter using forceps and placed in a square plastic dish containing 250 ml 2xSSPE/0.5% SDS.

The membrane was washed twice with 250 ml 2xSSPE/0.5% SDS for 10 min at 60°C. The membrane was then placed in a rolling bottle in a hybridization oven (Techne hybridiser HB-1, Techne (Cambridge) Ltd, Duxford Cambridge, USA) and allowed to cool to prevent inactivation of the peroxidase enzyme that is involved in the next step. To make detection possible, 2.5 µl streptavidin-peroxidase conjugate (500 U/ml) (Amersham International, Amersham, Buckinghamshire, UK) was added to 10 ml of 2x SSPE/0.5% SDS; this solution was poured in the rolling tube with the membrane and incubated for 60 min at 42°C.

The membrane was rinsed twice with 250 ml of 2x SSPE for 5 min at room temperature. For chemiluminescent detection, the blot was covered with 20 ml ECL detection liquid (10 ml of each of the supplied solutions (A & B); Amersham biosciences, UK) for 1 min. The membrane was covered with saran-wrap, and inside a dark room a light sensitive hyperfilm was exposed to the membrane in an X-ray cassette for 20 min.

After development of the film, the presence of the spacers was visualized as black squares. If the signalling was too weak or too strong the membrane was exposed again to another film for a shorter or longer period.

The used membrane was treated for re-use after rinsing in 1% SDS which was heated at 80°C for 30 min, washed for 15 min in 20 mM EDTA at pH 8 at room temperature, and further storage at 4°C wrapped in saran-wrap to avoid dehydration until next use. Each membrane was used up to 6 times.

3.6.3 Analysis of results

The results were recorded in a binary format representing the 43 spacers (Brudey *et al.*, 2006a). We performed spoligotyping blinded to the previously reported RFLP typing results (van der Spoel van Dijk *et al.*, 2005). The spoligotyping patterns were compared with an updated SpolDB4 database

of Pasteur Institute of Guadeloupe (http://www.pasteur-guadelope.fr:8081/SITVITDemo/?) that provides information on SIT distribution of MTBC spoligotypes and SpolDB3 (SPOTCLUST) (http://cgi2.cs.rpi.edu/~bennek/Run.html). The updated "in-house" version of the Pasteur Institute of Guadeloupe database, at the time of comparison (May 2008) contained a total of 39609 entries from 121 isolation countries and SPOTCLUST identified spoligotyping families based on the 535 patterns.

3.7 Multiplex PCR analysis

Eight isolates identified as Beijing and Beijing-like strains by the spoligotyping method were further evaluated by multiplex PCR which targets a direct repeat of IS*6110* within a 556 base pair (bp) intervening sequence (NTF-1) (Fig. 3.1) using primers (Table 3.2) and the method described by Plikaytis *et al.,* 1994 and the numbering system of Thierry *et al.,* 1990.

3.7.1 DNA amplification

The amplification mixture contained 5 ul of the template DNA and 45 µl of a reaction mixture consisting of 200 uM dNTP's, 1.0 µM (each) primers (Table 3.2), 1.25 U of Supertherm Taq polymerase (Southern Cross Biotechnology, Pty, Ltd), 10 mM Tris hydrochloride (pH8.3), 50 mM KCl, 1.5 mM $MgCl_2$, and 0.01% gelatin. The reaction was carried out using a Perkin-Elmer 9600 cycler starting with a denaturation step of 12 min at 96 °C. After denaturing, the PCR was performed for 30 cycles of denaturation for 1.5 min at 94°C, annealing at 60°C for 2.15 min and extension at 72°C for 2.5 min. The final extension step was done by incubating for 10 min at 72°C. Purified genomic DNA of *M. tuberculosis* H37Rv was used as positive controls. The negative controls consisted of PCR mixtures without mycobacterial DNA.

Table 3.2: Sequences of primers for multiplex PCR

Primer	Target	Sequences(5'-3')	Location[a]
IS54	IS*6110*	TCGACTGGTTCAACCATCGCCG	1210-1231
IS56	IS*6110*	GCGACCTCACTGATCGCTGC	170-151c
IS59	IS*6110*	GCGCCAGGCGCAGGTCGATGC	216-236
IS60	IS*6110*	GATCAGCGATCGTGGTCCTGC	738-718c
IS61	IS*6110*	GACCGCGGATCTCTGCGACC	133-114
IS62	IS*6110*	ACCAGTACTGCGGCGACGTC	1237-1256
MDR-6	NTF-1	CCAGATATCGGGTGTGTCGAC	473-494
MDR-7	NTF-1	CGCGAGATCTCATCGACAACC	52-32c

[a] denotes the residues of IS*6110* numbered according to the Thierry *et al* system. The first base of the 556bp *NTF-1* sequence is defined as the base adjacent to the 3' end of the flanking IS*6110* element of strain W of *Mycobacteriun tuberculosis.* The letter c denotes that the listed sequence is the complement of the published sequence (Thierry *et al.,* 1990).

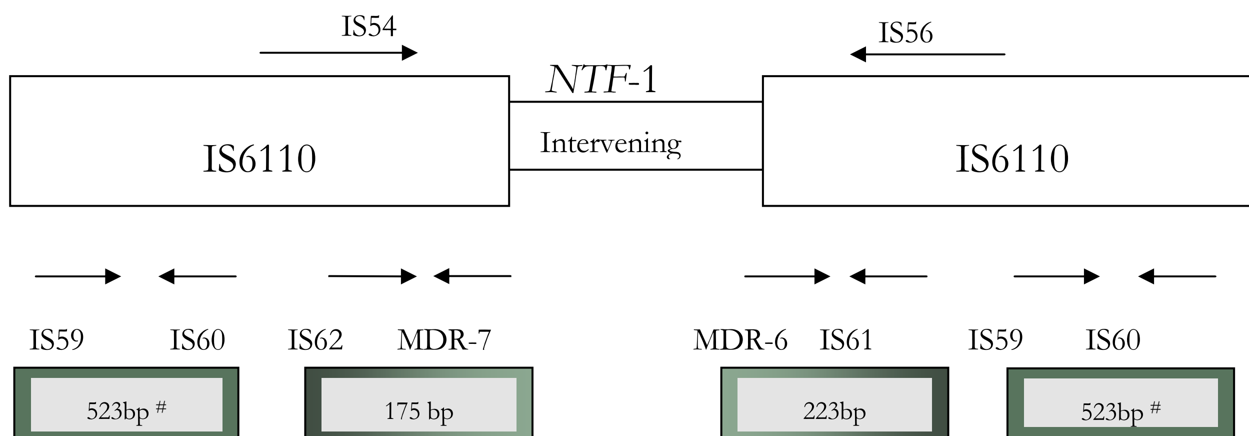

Figure 3-1: Structural representation of IS*6110* direct repeat with the intervening NTF-1 sequence.
The position of the primers used in the PCR, and the expected results from strain W. #, positive internal PCR control product are indicated. U strains will show internal control and atypical bands of 223 and various other sizes.

3.7.2 Gel electrophoresis

Amplicons were analysed by electrophoresis on 2% Low Melting (LM) Sieve agarose (BioWhittaker molecular applications, USA) at 100 V for 2 h using a 100 bp ladder (New England Bio-labs inc, UK) as the size marker. The gel was stained with 0.5 µg/ml ethidium bromide, observed under ultraviolet light and photographed using an uvipro (WhiteHead Scientific, Pty, Ltd) system.

3.8 Mycobacterial interspersed repetitive units or variable number of tandem repeats (MIRU-VNTR) typing

The original 12 MIRU-VNTR loci were individually amplified as described previously (Supply *et al.,* 2001), and results from each of the 12 loci were combined to form a 12 numerical allelic profile.

3.8.1 DNA Amplification

Simplex PCR was carried out for amplification of the loci using the HotStart Taq DNA polymerase kit (Qiagen, Hilden, Germany). The twelve pairs of primers used are listed in Table 3.3. One microliter of DNA solution was added to a final volume of 25 µl. The $MgCl_2$ concentrations varied depending on the locus. A 1.5 mM $MgCl_2$ solution was used in reaction mixtures for MIRUs 20, 24, 26, and 27; 2 mM $MgCl_2$ for MIRUs 2, 4, 10,16, 31 and 40; and 2.5 mM $MgCl_2$ for MIRUs 23 and 39.

The reactions were carried out using a Perkin-Elmer 9600 cycler starting with a denaturation step of 15 min at 95°C. After denaturation PCR was performed for 40 cycles of 1 min at 94°C, 1 min (varied depending on annealing temperatures) at 59°C and extension at 72°C for 1 min 30 sec. Annealing temperatures used were 5°C lower than the melting temperature obtained using Biomath-TM calculation for oligos

(www.promega.com/biomath/calc11.htm#mat_results). A final extension step of 10 min at 72°C was included at the end of the 40 cycles. The *M. tuberculosis* H37RV strain was used as a positive control. The negative controls consisted of PCR performed on PCR mixtures without mycobacterial DNA.

3.8.2 Gel Electrophoresis

The amplicons were fractionated electrophoretically in 3% LM Sieve (Whitehead Scientific Pty Ltd) agarose gels prepared in 1 × TBE buffer. Electrophoresis was performed at 120 V for 4 h using a 100 DNA bp ladder (New England Bio-labs Inc, UK) as the size marker. The gel was stained with 0.5 µg/ml ethidium bromide, observed under ultraviolet light and photographed using an uvipro (Whitehead Scientific Pty Ltd) system.

3.8.3 Analysis of data

The data was valid only if the *M. tuberculosis* H37Rv PCR results gave a band size corresponding to the standardised allele as indicated in Table 3.4 as been suggested by Supply *et al.* and Mazars *et al.* (Supply *et al.*, 2000, 2001, Mazars *et al.*, 2001). Obtained band sizes were assigned to corresponding numerical allelic codes as shown on Table 3.5. The oligonucleotides against the regions flanking locus 23 and 26 (ea. 23^1 and 26^1) have been changed to shorten the PCR products as shown in Table 3.3 (Supply *et al.*, 2000, 2001, Mazars *et al.*, 2001). Loci 2 and 24 contain an additional invariable type III and type I MIRU, respectively. Locus 4 contains an additional invariable 53 bp type II MIRU in 3’ of 77 bp VNTR units in nearly all clinical isolates (Table 3.6).

Table 3.3: Pairs of primers for amplification of different MIRUs that was developed from the flanking region of the different MIRUs from MTB H37Rv as indicated on Figure 2-4

Locus	**Primer (5' – 3')**
MIRU 2	2F - TGGACTTGCAGCAATGGACCAACT 2R - TACTCGGACGCCGGCTCAAAAT
MIRU 4	4F – GCGCGAGAGCCCGAACTGC 4R - GCGCAGCAGAAACGTCAGC
MIRU 10	10F – GTTCTTGACCAACTGCAGTCGTCC 10R - GCCACCTTGGTGATCAGCTACCT
MIRU 16	16F – TCGGTGATCGGGTCCAGTCCAAGTA 16R - CCCGTCGTGCAGCCCTGGTAC
MIRU 20	20F – TCGGAGAGATGCCCTTCGAGTTAG 20R - GGAGACCGCGACCAGGTACTTGTA
MIRU 23	23F – CTGTCGATGGCCGCAACAAAACG 23R - AGCTCAACGGGTTCGCCCTTTTGTC
MIRU 24	24F – CGACCAAGATGTGCAGGAATACAT 24R - GGCGAGTTGAGCTCACAGAA
MIRU 26	26F – TAGGTCTACCGTCGAAATCTGTGAC 26R - CATAGGCGACCAGGCGAATAG
MIRU 27	27F – TCGAAAGCCTCTGCGTGCCAGTAA 27R - GCGATGTGAGCGTGCCACTCAA
MIRU 31	31F – ACTGATTGGCTTCATACGGCTTTA 31R - GTGCCGACGTGGTCTTGAT
MIRU 39	39F – CGCATCGACAAACTGGAGCCAAAC 39R - CGGAAACGTCTACGCCCCACACAT
MIRU 40	40F – GGGTTGCTGGATGACAACGTGT 40R - GGGTGATCTCGGCGAAATCAGATA

Supply *et al.*, 2000, 2001 and Mazars *et al.*, 2001

Table 3.4: Specific MIRU-VNTR alleles in *M. tuberculosis* H37Rv control strain.

Strain ID	*M. tuberculosis* H37Rv											
MIRU'S	2	4	10	16	20	23	24	26	27	31	39	40
Allele	2	3'	3	2	2	6	1	3	3	3	2	1

Supply *et al.*, 2000, 2001, Mazars *et al.*, 2001

Table 3.5: MIRU-VNTR allele scoring table.

Allele	MIRU	MIRU 04	MIRU 10	MIRU 16	MIRU 20	MIRU	MIRU	MIRU	MIRU	MIRU	MIRU	MIRU	MIRU	MIRU
0	402	175	482	565	437	558	150	395	461	285	498	492	540	354
1	455	252	537	618	514	608	200	447	512	336	551	545	593	408
2	508	329	590	671	591	661	253	501	563	387	604	598	646	462
3	561	406	643	724	668	714	306	555	614	438	657	651	699	516
4	614	483	696	777	745	767	359	609	665	489	710	704	752	570
5	667	560	749	830	822	820	412	663	716	540	763	757	805	624
6	720	637	802	883	899	873	465	717	767	591	816	810	858	678
7	773	714	855	936	976	926	518	771	818	642	869	863	911	732
8	826	791	908	989	1053	979	571	825	869	693	922	916	964	786
9	879	868	961	1042	1130	1032	624	879	920	744	975	969	1017	840
10	932	945	1014	1095	1207	1085	677	933	971	795	1028	1022	1070	894
11	985	1022	1067	1148	1284	1138	730	987	1022	846	1081	1075	1123	948
12	1038	1099	1120	1201	1361	1191	783	1041	1073	897	1134	1128	1176	1002
13	1091	1176	1173	1254	1438	1244	836	1095	1124	948	1187	1181	1229	1056
14	1144	1253	1226	1307	1515	1297	889	1149	1175	999	1240	1234	1282	1110
15	1197	1330	1279	1360	1592	1350	942	1203	1226	1050	1293	1287	1335	1164

[1]represent alleles obtained with oligonucleotides used in Supply *et al.*, 2000, Mazars *et al.*, 2001.

[2]represent alleles obtained with oligonucleotides used in Supply *et al.*, 2001.

Table 3.6: Specific MIRU-VNTR alleles in locus 4

Allele	MIRU 04
0’	122
1’	199
2’	276
3’	353

Supply *et al*., 2000, Mazars *et al*., 2001.

3.9 Genetic relationship and Phylogenetic analysis

The genetic relationship of the 142 isolates to strains in the MIRU-VNTR DB were determined using both the spoligo and MIRU-VNTR data with equal weights and the categorical distance with a cutoff value of 0.17 or if no matching strains were found the tolerance were increased to 0.3. The categorical distance merely scores the number of markers with a different allele or spacer divided by the number of markers used and in our measurement with i representing loci and j spacers it will be $[0.5 \times (i/12)] + [0.5 \times (j/43)] \leq 0.17$.

Phylogenetic trees were drawn using the unweighted pair group method with arithmatic averages (UPGMA), MIRU-VNTR [12] with categorical coefficient rooting from a *M. canettii* (*M. prototuberculosis*) strain.

Chapter 4

Results

4.1 Strains analysed

The 141 isolates and one control strain H37Rv were analysed with both spoligotyping and MIRU-VNTR. Three of the 141 analyzed strains (GF32, GF56a and ZT71) were unintentionally repeated and gave the same results for MIRU-VNTR typing and spoligotyping both times indicating that laboratory practises were of acceptable standard (results not included). Furthermore, for 9 of the 141 strains an additional isolate was available numbered as GF53a, GF56a, Q01a, QO2a, Q28a, ZT16a, ZT53a, ZT70a and ZT80a by a previous student, and also tested (results not included in the study). It is however not clear if these isolates were from sputum that was taken two months after treatment started or why they were numbered differently. All these additional isolates were also tested unintentionally and gave different results compared to the initial specimens. The original RFLP data could not be found and both MIRU-VNTR and spoligo patterns were different for the second isolate, these results were therefor deemed correct and the samples as faulty and they were discarded from the study before analyses and thus not included as part of the 141 isolates. They might have been contaminated or mixed up during the previous study and therefor does not prove contamination of the spoligotyping or MIRU-VNTR typing results.

4.2 Genetic diversity and family assignment

Results of spoligo- and MIRU-VNTR typing with lineage assignment using SpolDB3, SpolDB4, MIRU-VNTRplus DB and clones identified with both methods are summarised in Tables 4.1 and 4.2

Strains of the FS were extremely diverse with the 141 selected isolates analysed giving 104 different spoligo patterns resulting in an overall diversity

(no of spoligo patterns of a family/ total number of isolates of a family) of 73%. MIRU-VNTR analysis gave 72 different 12-digit codes resulting in an overall diversity of 50% (no of patterns/ total number of isolates). The diversity of MIRU-VNTR within spoligo lineages however are more diverse as discussed in section 4.2.2.

SpolDB4 assigned names were used whenever a matching spoligo pattern was found in the database. This resulted in 69 isolates represented in 9 lineages, namely Beijing, Haarlem, LAM, S, T, X, CAS, *M. bovis* and U, with the last three lineages mentioned each represented by one isolate only.

Seventy three (51%) spoligo patterns were orphans as they were not reported before. These isolates were assigned to probable families or subfamilies using SpolDB3 “SPOTCLUST” whenever a probability of more than 0.77 was obtained. Except for two isolates (ZT60 and GF57) which gave a probability of 0.77 and one isolate ZT63, with a probability of 0.65.

Using the MIRU-VNTRplus DB best-match-based analyses with “spoligo: categorical” at a similarity level of 99.7% 3 miss matches was found and with “MIRU-VNTR [12]: categorical (1), spoligo: categorical (1)” gave different results at similarity levels of 99.87%, 99.7% and 99.5% respectively (Table 4.3). At the 99.5% similarity level all but 6 orphan strains could be assigned to SpolDB4 lineages (Table 4.3).

Table 4.1: Results of spoligo-MIRU-VNTR typing arranged according to obtained *SPOTCLUST (SpolDB3)* lineages.

Table 4.1: Results of spoligo-MIRU-VNTR typing arranged according to obtained SPOTCLUST (SpolDB3) lineages

Spoligo typing Pattern	UFS	Spoligo	MIRU-VNTR	12 MIRU-VNTR patterns												SpolDB4	RD	VNTRplus	VNTRplus	Distance
Spacers 1-43	Strain no	Clones	Clones	M 2	M 4	M10	M16	M20	M23	M24	M26	M27	M31	M39	M40	SIT no	Lineages	SIT no	Lineages	
	Q06	1	15	2	2	3	3	2	5	1	7	3	5	3	3	1	Beijing	1	Beijing	0
	Q08	1	15	2	2	3	3	2	5	1	7	3	5	3	3	1	Beijing	1	Beijing	0
	ZT69	1	15	2	2	3	3	2	5	1	7	3	5	3	3	1	Beijing	1	Beijing	0
	ZT71	1	15	2	2	3	3	2	5	1	7	3	5	3	3	1	Beijing	1	Beijing	0
	Q07	1	38	2	2	3	3	2	5	1	5	3	5	4	3	1	Beijing	1	Beijing	0
	ZT65	1	69	2	2	4	3	3	5	1	7	3	5	3	3	1	Beijing	1	Beijing	0
	ZT67	1	70	2	2	3	3	2	4	1	5	3	5	3	3	1	Beijing	1	Beijing	0
	ZT68		15	2	2	3	3	2	5	1	7	3	5	3	3	796	Beijing		Beijing	
	Q12		40	2	3	5	4	2	4	1	5	3	2	3	3	26	CAS	26	Dehli/CAS	0.0233
	GF02		16	2	5	4	4	3	6	1	4	3	5	2	3	Orphan	EAI1	539	vole	0.2326
	GF11		16	2	5	4	4	3	6	1	4	3	5	2	3	Orphan	EAI1	539	vole	0.1395
	Q51		16	2	5	4	4	3	6	1	4	3	5	2	3	Orphan	EAI1	818	Caprae	0.1163
	GF23		17	2	5	4	4	2	6	2	4	3	5	3	3	Orphan	EAI1	539	vole	0.2093
	ZT19		17	2	5	4	4	2	6	2	4	3	5	3	3	Orphan	EAI1		West African	0.2558
	Q10		39	2	5	4	3	3	6	1	5	3	2	2	3	Orphan	EAI5	1092	Dehli/CAS	0.1628
	Q11		5	2	3	3	2	3	6	1	5	3	2	2	3	Orphan	Family33		Ugandal	0.2326
	ZT30		2	2	3	3	2	3	5	1	5	3	2	2	3	Orphan	Family33	236	EAI	0.1628
	Q04		7	2	2	3	3	3	6	1	5	3	2	2	3	Orphan	Family35	539	vole	0.2093
	Q25		43	2	3	3	2	3	4	1	5	3	2	2	3	Orphan	Family35	127	NEW-1	0.0698
	ZT10	2	55	2	3	3	3	3	3	1	5	3	2	2	3	Orphan	Family36	125	Ugandal	0.093
	ZT13	2	56	5	3	3	3	3	5	1	5	3	2	2	3	Orphan	Family36	125	Ugandal	0.093
	BFN01		3	2	3	3	3	1	6	1	5	3	2	2	3	Orphan	Family36	539	vole	0.1163
	Q71		11	2	3	3	3	3	6	1	5	3	2	2	3	Orphan	Family36	539	vole	0.1163
	ZT46		11	2	3	3	3	3	6	1	5	3	2	2	3	Orphan	Family36	125	Ugandal	0.0698
	ZT09		1	2	2	4	3	2	5	1	5	3	3	2	2	62	Haarlem1	47	Haarlem	0.0233
	Q47		51	5	5	4	3	3	6	1	5	3	2	2	3	148	Haarlem1	236	EAI	0.0698
	ZT33		1	2	2	4	3	2	5	1	5	3	3	2	2	Orphan	Haarlem1	47	Haarlem	0.1628
	ZT31		60	2	5	4	2	3	6	1	4	3	5	2	3	Orphan	Haarlem1	47	Haarlem	0.1395
	GF01		14	5	2	3	2	3	6	1	5	3	2	2	3	Orphan	LAM1	20	LAM	0.0465
	GF08	3	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3	33	LAM	0
	GF13	3	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3	33	LAM	0
	GF22	3	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3	33	LAM	0
	Q18	3	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3	33	LAM	0
	Q33	3	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3	33	LAM	0
	Q36	3	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3	33	LAM	0
	Q37	3	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3	33	LAM	0
	ZT17	3	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3	33	LAM	0
	ZT18	3	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3	33	LAM	0
	ZT20	3	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3	33	LAM	0
	ZT42	3	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3	33	LAM	0
	ZT44	3	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3	33	LAM	0
	GF38	3	22	2	2	4	3	2	7	1	5	3	3	2	3	33	LAM3	33	LAM	0
	Q17		41	2	2	4	3	3	6	1	5	3	3	2	3	33	LAM3	33	LAM	0.093
	GF05		2	2	3	3	2	3	5	1	5	3	2	2	3	Orphan	LAM3	33	LAM	0.0698
	GF12		4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	LAM3			
	ZT24		12	2	2	4	3	2	6	1	5	3	3	2	3	Orphan	LAM3	376	LAM	0.2093
	ZT50		12	2	2	4	3	2	6	1	5	3	3	2	3	Orphan	LAM3	33	LAM	0.1628
	Q35		46	2	2	3	0	3	6	1	5	3	2	2	3	Orphan	LAM3	33	LAM	0.093
	ZT40		62	2	2	3	2	3	4	1	5	3	2	2	3	Orphan	LAM3	33	LAM	0.2791

BFN02		4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	LAM7	539	vole	0.186
ZT03		4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	LAM7	539	vole	0.093
ZT15		4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	LAM7	539	vole	0.1628
ZT25		59	3	3	3	3	3	7	1	5	3	3	2	3	Orphan	LAM7			
Q02	4	36	2	5	4	3	3	6	1	5	3	2	2	1	811	LAM8	41	TUR	0.1163
ZT32	4	4	2	2	3	2	3	6	1	5	3	2	2	3	811	LAM8	41	TUR	0.1163
Q20		9	2	2	4	2	3	6	1	5	3	5	2	1	813	LAM8	42	LAM	0.1395
ZT08		9	2	2	4	2	3	6	1	5	3	5	2	1	815	LAM8	42	LAM	0.1395
GF59		28	2	2	4	0	3	6	1	5	3	5	2	1	Orphan	LAM8		Dehli/CAS	0.2326
ZT48		13	2	2	3	2	3	6	1	5	3	4	2	3	162	LAM9	42	LAM	0.0233
Q41		48	2	2	3	3	3	5	1	4	3	2	2	3	Orphan	LAM9	33	LAM	0.186
ZT52		63	2	2	3	2	3	6	2	5	3	4	2	3	Orphan	LAM9	254	LAM	0.093
GF60		29	2	2	4	4	2	4	1	4	3	2	2	3	Orphan	*M. africanum*	125	Ugandal	0.2093
ZT59		67	2	2	2	3	2	4	2	5	3	2	2	3	482	*M. bovis*-BCG	482	Bovis	0.0233
ZT16	5	7	2	2	3	3	3	6	1	5	3	2	2	3	71	S	71	S	0
ZT27	5	7	2	2	3	3	3	6	1	5	3	2	2	3	71	S	71	S	0
GF71	6	13	2	2	3	2	3	6	1	5	3	4	2	3	71	S	71	S	0.0233
ZT39	6	8	2	3	3	3	2	5	1	5	3	3	2	3	34	S	71	S	0.0233
ZT64	7	8	2	3	3	3	2	5	1	5	3	3	2	3	34	S	34	S	0
ZT60	7	8	2	3	3	3	2	5	1	5	3	3	2	3	34	S/T1	34	S	0
ZT61	7	8	2	3	3	3	2	5	1	5	3	3	2	3	34	S	34	S	0
GF26		8	2	2	3	3	2	5	1	5	3	3	2	3	34	S	34	S	0.0465
GF07		4	2	2	3	2	3	6	1	5	3	2	2	3	34	S	71	S	0.1628
ZT49	8	5	2	3	3	2	3	6	1	5	3	2	2	3	Orphan	S	34	S	0.0233
Q32	8	45	2	3	3	2	11	6	1	5	3	2	2	3	Orphan	S	34	S	0.0233
ZT14		57	2	2	3	2	1	6	1	1	3	2	2	3	Orphan	S	262	URAL	0.1395
ZT56		66	5	3	3	2	3	6	1	5	3	2	2	1	Orphan	S	71	S	0.1163
ZT63		68	2	3	3	0	3	5	1	5	3	2	2	3	Orphan	S/T1	34	S	0.0465
GF44	9	24	2	2	3	2	3	5	1	4	3	2	2	3	53	T1	53	Ghana	0
GF53	9	5	2	3	3	2	3	6	1	5	3	2	2	3	53	T1	53	Ghana	0
ZT54	9	5	2	3	3	2	3	6	1	5	3	2	2	3	53	T1	53	Ghana	0
Q24	9	3	2	3	3	3	1	6	1	5	3	2	2	3	53	T1	53	Ghana	0
Q40	9	11	2	3	3	3	3	6	1	5	3	2	2	3	53	T1	53	Ghana	0
ZT04	9	11	2	3	3	3	3	6	1	5	3	2	2	3	53	T1	53	Ghana	0
ZT38	9	11	2	3	3	3	3	6	1	5	3	2	2	3	53	T1	53	Ghana	0
ZT80	10	11	2	3	3	3	3	6	1	5	3	2	2	3	53	T1	53	Ghana	0.0233
GF62	10	31	2	3	3	2	3	6	1	5	3	2	2	4	53	T1	53	Ghana	0.0233
ZT53	10	64	2	5	4	4	3	0	1	4	3	5	2	3	53	T1	53	Ghana	0.0233
GF56		26	2	5	4	3	3	7	2	5	3	2	3	4	53	T1	236	EAI	0.0465
Q28		44	2	3	4	3	11	6	1	5	3	2	2	3	53	T1	52	Ugandal	0
GF51		5	2	3	3	2	3	6	1	5	3	2	2	3	44	T1	53	Ghana	0.0233
ZT55		65	2	3	3	3	3	4	1	5	3	2	2	3	118	T1	53	Ghana	0.0465
Q26		4	2	2	3	2	3	6	1	5	3	2	2	3	245	T1	53	Ghana	0.0465
ZT28		4	2	2	3	2	3	6	1	5	3	2	2	3	1122	T1	53	Ghana	0.0233
ZT29		4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	T1	135	Ugandall	0.0698
GF18		7	2	2	3	3	3	6	1	5	3	2	2	3	Orphan	T1	53	Ghana	0.186
ZT66		7	2	2	3	3	3	6	1	5	3	2	2	3	Orphan	T1	53	Ghana	0.0233
ZT35		18	2	3	3	2	3	6	1	5	3	3	2	3	Orphan	T1	53	Ghana	0.0698
ZT57		18	2	3	3	2	3	6	1	5	3	3	2	3	Orphan	T1	53	Ghana	0.093
GF33		21	2	3	3	2	3	7	1	4	3	2	2	3	Orphan	T1		Ugandall	0.2326
GF40		23	2	2	0	0	3	6	1	5	3	2	2	3	Orphan	T1	53	Ghana	0.0698
Q21		42	2	3	3	0	3	4	1	5	3	2	2	3	Orphan	T1		Ugandal	0.0465
Q39		47	2	2	3	3	2	5	1	5	3	2	2	3	Orphan	T1		Ghana	0.0698
ZT23		58	2	3	3	3	3	6	1	4	3	3	2	3	Orphan	T1	33	LAM	0.1163
ZT36		61	5	3	3	3	3	6	2	5	3	3	2	3	Orphan	T1	53	Ghana	0.0465
GF31		8	2	3	3	3	2	5	1	5	3	3	2	3	Orphan	T2			
BFN03		14	5	2	3	2	3	6	1	5	3	2	2	3	Orphan	T2		Ugandall	0.2558
GF64		32	2	3	2	5	2	3	5	3	5	1	3	2	Orphan	T2	78	Ugandal	0.1395
Q01		35	5	2	4	3	2	5	1	5	3	3	2	2	Orphan	T2	647	Caprae	0.2791
GF32		10	5	2	3	2	3	5	1	5	3	2	2	3	Orphan	T2		Dehli/CAS	0.1628

Q72		10	5	2	3	2	3	5	1	5	3	2	2	3	Orphan	T3	818	Caprae	0.186
GF61		30	2	2	4	0	3	7	1	5	3	2	2	3	Orphan	T3		West African	0.2093
ZT76		72	2	2	3	2	2	6	1	5	3	2	2	3	Orphan	T3	818	Caprae	0.2791
ZT22		4	2	2	3	2	3	6	1	5	3	2	2	3	1737	T4	159	LAM	0.0233
GF49		4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	T4	592	Canetti	0.1395
GF45		6	2	2	4	2	3	6	1	5	3	2	2	3	254	T5-RUS	254	LAM	0
GF30		6	2	2	4	2	3	6	1	5	3	2	2	3	834	U	53	Ghana	0.093
BFN04	11	19	5	2	3	2	2	3	5	1	5	2	2	3	119	X1	119	X	0
ZT05	11	53	2	3	3	3	3	6	1	5	3	2	2	4	119	X1	119	X	0
ZT73		4	2	2	3	2	3	6	1	5	3	2	2	3	336	X1/T1	52	Ugandal	0.0233
GF21		13	2	2	3	2	3	6	1	5	3	4	2	3	1470	X1-LAM9	42	LAM	0.0233
GF57		27	2	7	4	3	2	6	2	2	3	4	1	4	Orphan	X1/T1	119	X	0.093
ZT06		54	2	3	3	2	3	6	1	5	3	2	2	2	Orphan	X1/T1		Ugandal	0.1163
ZT62		4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	X1		Ugandal	0.1628
ZT77		4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	X1	34	S	0.1163
GF20		7	2	2	3	3	3	6	1	5	3	2	2	3	Orphan	X2	642	llama	0.2791
Q43		49	2	3	3	3	3	5	1	5	3	2	2	4	12	X3	34	S	0.2558
GF03	12	14	5	2	3	2	3	6	1	5	3	2	2	3	92	X3	34	S	0.186
GF54	12	4	2	2	3	2	3	6	1	5	3	2	2	3	92	X3	34	S	0.186
Q03	12	37	2	2	3	2	3	5	1	5	3	2	2	3	92	X3	34	S	0.186
ZT74	12	71	2	2	3	2	3	6	1	5	4	2	2	3	92	X3	34	S	0.186
BFN05		20	2	2	4	3	2	5	1	4	3	2	2	2	Orphan	X3		Ugandal	0.186
GF52		25	2	3	3	3	3	5	1	5	3	2	2	3	Orphan	X3	647	Caprae	0.2093
GF70		33	2	2	4	3	2	5	1	5	3	3	2	4	Orphan	X3	34	S	0.2093
GF72		34	2	2	4	3	3	5	1	5	3	3	2	3	Orphan	X3	818	Caprae	0.2326
Q05		3	2	3	3	3	1	6	1	5	3	2	2	3	Orphan	X3	34	S	0.2558
ZT70		3	2	3	3	3	1	6	1	5	3	2	2	3	Orphan	X3		Ugandal	0.2558
Q44	13	50	2	2	4	3	2	5	1	5	3	3	5	4	Orphan	X3		Ugandal	0.2326
Q45	13	11	2	3	3	3	3	6	1	5	3	2	2	3	Orphan	X3		Ugandal	0.2326
Q70		4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	X3	592	Canetti	0.1395
ZT43		4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	X3	539	vole	0.1628
ZT01		52	2	2	3	1	2	6	1	5	3	2	2	2	Orphan	X3	125	Ugandal	0.2558
H37RV		Control	2'	3	2	5	2	3	2	3	6	1	3	2	451	H37Rv	451	H37Rv	0.0465

Table 4.2: **Results of spoligo-MIRU-VNTR typing arranged according to MIRU-VNTR clones using excel 2007.**

Table 4.2 Results of spoligo-MIRU-VNTR typing arranged according to MIRU-VNTR clones using excel 2007.

Spoligo typing Pattern	*M. tuberculosis*	MIRU-VNTR	12 MIRU-VNTR patterns												SpolDB4	SpolDB3
Spacers 1-43	Strain no	Clones	M 2	M 4	M10	M16	M20	M23	M24	M26	M27	M31	M39	M40	SIT no	Lineages
	ZT09	1	2	2	4	3	2	5	1	5	3	3	2	2	62	Haarlem1
	ZT33	1	2	2	4	3	2	5	1	5	3	3	2	2	Orphan	Haarlem1
	GF05	2	2	3	3	2	3	5	1	5	3	2	2	3	Orphan	LAM3
	ZT30	2	2	3	3	2	3	5	1	5	3	2	2	3	Orphan	Family33
	BFN01	3	2	3	3	3	1	6	1	5	3	2	2	3	Orphan	Family36
	Q05	3	2	3	3	3	1	6	1	5	3	2	2	3	Orphan	X3
	Q24	3	2	3	3	3	1	6	1	5	3	2	2	3	53	T1
	ZT70	3	2	3	3	3	1	6	1	5	3	2	2	3	Orphan	X3
	GF12	4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	LAM3
	BFN02	4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	LAM7
	ZT03	4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	LAM7
	ZT15	4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	LAM7
	ZT32	4	2	2	3	2	3	6	1	5	3	2	2	3	811	LAM8
	GF07	4	2	2	3	2	3	6	1	5	3	2	2	3	34	S
	ZT28	4	2	2	3	2	3	6	1	5	3	2	2	3	1122	T1
	ZT29	4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	T1
	Q26	4	2	2	3	2	3	6	1	5	3	2	2	3	245	T1
	ZT22	4	2	2	3	2	3	6	1	5	3	2	2	3	1737	T4
	GF49	4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	T4
	ZT62	4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	X1
	ZT77	4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	X1
	ZT73	4	2	2	3	2	3	6	1	5	3	2	2	3	336	X1/T1
	GF54	4	2	2	3	2	3	6	1	5	3	2	2	3	92	X3
	ZT43	4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	X3
	Q70	4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	X3
	GF51	5	2	3	3	2	3	6	1	5	3	2	2	3	44	T1
	GF53	5	2	3	3	2	3	6	1	5	3	2	2	3	53	T1
	Q11	5	2	3	3	2	3	6	1	5	3	2	2	3	Orphan	Family33
	ZT49	5	2	3	3	2	3	6	1	5	3	2	2	3	Orphan	S
	ZT54	5	2	3	3	2	3	6	1	5	3	2	2	3	53	T1
	GF30	6	2	2	4	2	3	6	1	5	3	2	2	3	834	U
	GF45	6	2	2	4	2	3	6	1	5	3	2	2	3	254	T5-RUS
	GF18	7	2	2	3	3	3	6	1	5	3	2	2	3	Orphan	T1
	GF20	7	2	2	3	3	3	6	1	5	3	2	2	3	Orphan	X2
	Q04	7	2	2	3	3	3	6	1	5	3	2	2	3	Orphan	Family35
	ZT16	7	2	2	3	3	3	6	1	5	3	2	2	3	71	S
	ZT27	7	2	2	3	3	3	6	1	5	3	2	2	3	71	S
	ZT66	7	2	2	3	3	3	6	1	5	3	2	2	3	Orphan	T1
	GF26	8	2	2	3	3	2	5	1	5	3	3	2	3	34	S
	GF31	8	2	3	3	3	2	5	1	5	3	3	2	3	Orphan	T2
	ZT39	8	2	3	3	3	2	5	1	5	3	3	2	3	34	S
	ZT60	8	2	3	3	3	2	5	1	5	3	3	2	3	34	S/T1
	ZT61	8	2	3	3	3	2	5	1	5	3	3	2	3	34	S
	ZT64	8	2	3	3	3	2	5	1	5	3	3	2	3	34	S

Q20	9	2	2	4	2	3	6	1	5	3	5	2	1	813	LAM8
ZT08	9	2	2	4	2	3	6	1	5	3	5	2	1	815	LAM8
GF31	10	5	2	3	2	3	5	1	5	3	2	2	3	Orphan	T2
GF32	10	5	2	3	2	3	5	1	5	3	2	2	3	Orphan	T2
GF32	10	5	2	3	2	3	5	1	5	3	2	2	3	Orphan	T2
Q72	10	5	2	3	2	3	5	1	5	3	2	2	3	Orphan	T3
Q40	11	2	3	3	3	3	6	1	5	3	2	2	3	53	T1
Q45	11	2	3	3	3	3	6	1	5	3	2	2	3	Orphan	X3
ZT04	11	2	3	3	3	3	6	1	5	3	2	2	3	53	T1
ZT38	11	2	3	3	3	3	6	1	5	3	2	2	3	53	T1
ZT46	11	2	3	3	3	3	6	1	5	3	2	2	3	Orphan	Family36
ZT80	11	2	3	3	3	3	6	1	5	3	2	2	3	53	T1
Q71	11	2	3	3	3	3	6	1	5	3	2	2	3	Orphan	Family36
GF08	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3
GF13	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3
GF22	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3
Q18	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3
Q33	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3
Q36	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3
Q37	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3
ZT17	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3
ZT18	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3
ZT20	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3
ZT24	12	2	2	4	3	2	6	1	5	3	3	2	3	Orphan	LAM3
ZT42	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3
ZT44	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3
ZT50	12	2	2	4	3	2	6	1	5	3	3	2	3	Orphan	LAM3
GF21	13	2	2	3	2	3	6	1	5	3	4	2	3	1470	X1-LAM9
GF27	13	2	2	3	2	3	6	1	5	3	4	2	3	1154	LAM9
GF71	13	2	2	3	2	3	6	1	5	3	4	2	3	71	S
ZT48	13	2	2	3	2	3	6	1	5	3	4	2	3	162	LAM9
GF01	14	5	2	3	2	3	6	1	5	3	2	2	3	Orphan	LAM1
GF03	14	5	2	3	2	3	6	1	5	3	2	2	3	92	X3
BFN03	14	5	2	3	2	3	6	1	5	3	2	2	3	Orphan	T2
Q06	15	2	2	3	3	2	5	1	7	3	5	3	3	1	Beijing
Q08	15	2	2	3	3	2	5	1	7	3	5	3	3	1	Beijing
ZT68	15	2	2	3	3	2	5	1	7	3	5	3	3	796	Beijing
ZT69	15	2	2	3	3	2	5	1	7	3	5	3	3	1	Beijing
ZT71	15	2	2	3	3	2	5	1	7	3	5	3	3	1	Beijing
GF02	16	2	5	4	4	3	6	1	4	3	5	2	3	Orphan	EAI1
GF11	16	2	5	4	4	3	6	1	4	3	5	2	3	Orphan	EAI1
Q51	16	2	5	4	4	3	6	1	4	3	5	2	3	Orphan	EAI1
GF23	17	2	5	4	4	2	6	2	4	3	5	3	3	Orphan	EAI1
ZT19	17	2	5	4	4	2	6	2	4	3	5	3	3	Orphan	EAI1

ZT35	18	2	3	3	2	3	6	1	5	3	3	2	3	Orphan	T1
ZT57	18	2	3	3	2	3	6	1	5	3	3	2	3	Orphan	T1
BFN04	19	5	2	3	2	2	3	5	1	5	2	2	3	119	X1
BFN05	20	2	2	4	3	2	5	1	4	3	2	2	2	Orphan	X3
GF33	21	2	3	3	2	3	7	1	4	3	2	2	3	Orphan	T1
GF38	22	2	2	4	3	2	7	1	5	3	3	2	3	33	LAM3
GF40	23	2	2	0	0	3	6	1	5	3	2	2	3	Orphan	T1
GF44	24	2	2	3	2	3	5	1	4	3	2	2	3	53	T1
GF52	25	2	3	3	3	3	5	1	5	3	2	2	3	Orphan	X3
GF56	26	2	5	4	3	3	7	2	5	3	2	3	4	53	T1
GF57	27	2	7	4	3	2	6	2	2	3	4	1	4	Orphan	X1/T1
GF59	28	2	2	4	0	3	6	1	5	3	5	2	1	Orphan	LAM8
GF60	29	2	2	4	4	2	4	1	4	3	2	2	3	Orphan	*M. africanum*
GF61	30	2	2	4	0	3	7	1	5	3	2	2	3	Orphan	T3
GF62	31	2	3	3	2	3	6	1	5	3	2	2	4	53	T1
GF64	32	2	3	2	5	2	3	5	3	5	1	3	2	Orphan	T2
GF70	33	2	2	4	3	2	5	1	5	3	3	2	4	Orphan	X3
GF72	34	2	2	4	3	3	5	1	5	3	3	2	3	Orphan	X3
Q01	35	5	2	4	3	2	5	1	5	3	3	2	2	Orphan	T2
Q02	36	2	5	4	3	3	6	1	5	3	2	2	1	811	LAM8
Q03	37	2	2	3	2	3	5	1	5	3	2	2	3	92	X3
Q07	38	2	2	3	3	2	5	1	5	3	5	4	3	1	Beijing
Q10	39	2	5	4	3	3	6	1	5	3	2	2	3	Orphan	EAI5
Q12	40	2	3	5	4	2	4	1	5	3	2	3	3	26	CAS
Q17	41	2	2	4	3	3	6	1	5	3	3	2	3	33	LAM3
Q21	42	2	3	3	0	3	4	1	5	3	2	2	3	Orphan	T1
Q25	43	2	3	3	2	3	4	1	5	3	2	2	3	Orphan	Family35
Q28	44	2	3	4	3	11	6	1	5	3	2	2	3	53	T1
Q32	45	2	3	3	2	11	6	1	5	3	2	2	3	Orphan	S
Q35	46	2	2	3	0	3	6	1	5	3	2	2	3	Orphan	LAM3
Q39	47	2	2	3	3	2	5	1	5	3	2	2	3	Orphan	T1
Q41	48	2	2	3	3	3	5	1	4	3	2	2	3	Orphan	LAM9
Q43	49	2	3	3	3	3	5	1	5	3	2	2	4	12	X3
Q44	50	2	2	4	3	2	5	1	5	3	3	5	4	Orphan	X3
Q47	51	5	5	4	3	3	6	1	5	3	2	2	3	148	Haarlem1
ZT01	52	2	2	3	1	2	6	1	5	3	2	2	2	Orphan	X3
ZT05	53	2	3	3	3	3	6	1	5	3	2	2	4	119	X1
ZT06	54	2	3	3	2	3	6	1	5	3	2	2	2	Orphan	X1/T1
ZT10	55	2	3	3	3	3	3	1	5	3	2	2	3	Orphan	Family36
ZT13	56	5	3	3	3	3	5	1	5	3	2	2	3	Orphan	Family36
ZT14	57	2	2	3	2	1	6	1	1	3	2	2	3	Orphan	S
ZT23	58	2	3	3	3	3	6	1	4	3	3	2	3	Orphan	T1
ZT25	59	3	3	3	3	3	7	1	5	3	3	2	3	Orphan	LAM7
ZT31	60	2	5	4	2	3	6	1	4	3	5	2	3	Orphan	Haarlem 1
ZT36	61	5	3	3	3	3	6	2	5	3	3	2	3	Orphan	T1
ZT40	62	2	2	3	2	3	4	1	5	3	2	2	3	Orphan	LAM3
ZT52	63	2	2	3	2	3	6	2	5	3	4	2	3	Orphan	LAM9
ZT53	64	2	5	4	4	3	0	1	4	3	5	2	3	53	T1
ZT55	65	2	3	3	3	3	4	1	5	3	2	2	3	118	T1
ZT56	66	5	3	3	2	3	6	1	5	3	2	2	1	Orphan	S
ZT59	67	2	2	2	3	2	4	2	5	3	2	2	3	482	*M. bovis*-BCG
ZT63	68	2	3	3	0	3	5	1	5	3	2	2	3	Orphan	S/T1
ZT65	69	2	2	4	3	3	5	1	7	3	5	3	3	1	Beijing
ZT67	70	2	2	3	3	2	4	1	5	3	5	3	3	1	Beijing
ZT74	71	2	2	3	2	3	6	1	5	4	2	2	3	92	X3
ZT76	72	2	2	3	2	2	6	1	5	3	2	2	3	Orphan	T3
H37RV	Control	2'	3	2	5	2	3	2	3	6	1	3	2	451	H37Rv

Table 4.3 **Best-match-based analyses performed on the 141 isolates from the FS.**

Cutoff results	No(%) of isolates with the indicated results by:	
	Spoligo: categorical	MIRU-VNTR[12], categorical (1) Spoligo: categorical (1)
Cutoff of 0.17		
Best Match		0 (0)
No match		141 (100)
conflict		0 (0)
Cutoff of 0.3		
Best Match	101 (71.1)	91 (64.1)
No match	3 (2.1)	21 (14.8)
conflict	38 (26.8)	30 (21.1)
Cutoff of 0.5		
Best Match		70 (49.3)
No match		6 (4.2)
conflict		66 (0)

4.2.1 Spacer-oligotyping "Spoligotyping"

All our strains except for 8 miss spacers 33 to 36.

Using spoligotyping alone, the biggest group (n=38, 26.8 %) of isolates belonged to the T family which could be divided into five subfamilies T1, T2, T3, T4 and T5-Rus1 with 27 (71.0%), 5 (13.2%), 3 (7.9%), 2 (5.3%) and 1 (2.6%) strains respectively. Eighteen (47%) of these isolates could be identified with SpolDB4 with 12 (32%) being SIT53, and 53% were orphans.

The 38 T strains were further characterized and divided into 26 different MIRU-VNTR patterns. Loci 4, 10, 16, 20, 23, 26 and 31 showed a high discrimination for T family isolates as indicated by red rectangular boxes (Table 4.2) within the MIRU-VNTR patterns. Most T strains have identical alleles for loci 2, 24, 27, 39 and 40.

The LAM family, consisting of five subfamilies (1, -3, -7, -8 and-9), was identified to be the second largest group comprising 33 (23.2%) of the isolates and including two clones (3 and 4). The largest clonal group consists of 13 isolates identified as SIT33. A 14th isolate that was assigned a SIT33 strains using the probabilities in SpolDB3 had a slightly different spoligopattern and is thus SIT33-like. The second clone comprises 2 isolates as shown in Fig 4.1. Of the 33 LAM strains 19 (57.6%) could be identified with SpolDB4 and 14 (41.2%) were orphans.

The 33 LAM strains were further characterised and divided into 13 different MIRU-VNTR patterns. Loci 16, 20, 23 and 31 showed high diversity for the LAM family isolates while other loci have almost the same allelic values. The 13/14 LAM3/SIT33 strains all had identical MIRU-VNTR patterns except for isolate GF38 that has a different allelic value for the MIRU-VNTR loci 23. The 14th isolate with a slightly different spoligo pattern has a different value for MIRU-VNTR loci 20.

Figure 4-1: UPGMA-similarity tree of the spoligopatterns of the 141 isolates plus a control strain (H37Rv) of MTB compared to the reference strains of the MIRU-VNTRplus DB.

Figure 4-1: UPGMA-similarity tree of the spoligopatterns of the 142 isolates plus a control strain (H37Rv) of *M. tuberculosis* compared to the reference strains of the MIRU-VNTRplus DB. The isolates from the study are represented by the coloured blocks with isolates that form clones (identical spoligopatterns) indicated by other colours than yellow. Isolates encased by a red block illustrate Free State strain identical to strains from other countries of origin. The SIT[a] numbers indicated in the dendogram and the lineages[b] indicate data determined from the MIRU-VNTRplus DB during analysis.

The third largest group comprising of 25 (17.6%) isolates, was the X family consisting of subfamilies X-1, -2, and -3 and including three clones (11, 12 and 13). Nine (36%) of the isolates could be identified in the SpolDB4 and 16 (64%) were orphans. Sixteen (64%) isolates belonged to the X3 subfamily with 11 (68.8%) of them being orphans.

The X family was split into 19 groups by MIRU-VNTR with high diversity in all loci. Even strains that form spoligo clones all have a certain amount of difference in their allelic constitution.

These three main families (T, LAM and X) caused 67.6% of the individual cases of MTB and were responsible for 7 spoligo clones in the FS as shown in Table 4.2. The rate of spoligo pattern diversity within each main family varies substantially and was 76%, 56% and 80% for T, LAM and X respectively.

Other identified families such as the S family (14 isolates) comprised 4 clones with 3 of them consisting of 2 isolates each and 3 isolates for the fourth one. *M. bovis, M. africanum*, CAS, the Haarlem1, U, F33-, 35-, 36 families (1 clone) and the EAI family accounted together for 15.5% of the strain pool (Tables 4.2 and 4.4a). Seven (4.9%) isolates were identified as belonging to the Beijing family and all of them formed a clone by spoligotyping. One isolate belonged to the Beijing-like group. Beijing isolates were divided into four groups by MIRU-VNTR with a few variable allelic values observed (Table 4.2).

The demographic distribution of the spoligotypes is summarised in Tables 4.4a and b. The Beijing, S, T1 and X strains were more prominent in the Motheo district, but the difference was not significant (Tables 4.4a and b).

Table 4.4a: Demographic distribution of lineages over three districts in FS province

Lineage(s)	U	LAM1	LAM3	LAM7	LAM8	LAM9	S	T1	T2	T3	T4	T5-Rus1	X1	X2	X3
Total isolate(s)	n=1	n=33					n=14	n=38					n=25		
District (Strain abbreviation)															
Motheo (ZT,BFN)	0	0	8	4	2	2	10	13	1	1	1	0	6	0	5
Thabo Mofutsanyane (Q)	0	0	6	0	2	1	1	6	1	1	0	0	0	0	6
Lejweleputswa (GF)	1	1	6	0	1	0	3	8	3	1	1	1	2	1	5

ZT-Thaba Nchu, BFN-Boemfontein, Q-Qwaqwa, GF-Goldfields

Table 4.4b: Demographic distribution of lineages over three districts in FS province

Lineage(s)	Beijing	CAS	EAI1	EAI 5	F33	F35	F36	Haarlem1	*M. bovis*	*M. africanum*
Total isolate(s)	n=8	n=1	n=6		n=2	n=2	n=5	n=4	n=1	n=1
District (Strain abbreviation)										
Motheo (ZT,BFN)	5	0	1	0	1	0	4	3	1	0
Thabo Mofutsanyane (Q)	3	1	1	1	1	2	1	1	0	0
Lejweleputswa (GF)	0	0	3	0	0	0	0	0	0	1

4.2.2 Mycobacterial interspersed repetitive units or variable number of tandem repeats (MIRU-VNTR) typing

Fifty four (38%) MIRU-VNTR profiles occurred only once and 18 clones comprised 88 isolates (61.9%) with each clone consisting of 2 or more isolates. The two largest clones (no 4 and 12) comprised 17 (11.9 %) and 12 (8.5%) isolates respectively (Table 4.2).

Clone 4 consists of the MIRU-VNTR profile 223236153223 and includes 5 LAM strains, 5 T strains, 6 X strains and one isolate from the S family. Ten of these isolates have a spoligo pattern that was not reported before and six of them have less than 10 spacers. Clone 12, with the MIRU-VNTR profile 224326153223, consists of LAM3 (SIT33) spoligotypes.

Twenty-six different MIRU-VNTR patterns were found among 38 T strains analysed (diversity of 58%). Nine of the MIRU-VNTR clones contain isolates of the T lineage: 8 MIRU-VNTR clones (no 3, 4, 5, 6, 7, 8, 11 and 14) containing T strains as well as other spoligo lineages and 1 clone (no 10) with T strains only. Twenty T strains with unique spoligo and MIRU-VNTR patterns were found.

Nine-teen different MIRU-VNTR patterns were obtained for 25 X strains analysed (diversity of 76%), with 2 clones (no 3 and 4) comprising of 8 isolates. Seventeen of the 25 strains have unique MIRU-VNTR patterns, of which 4 of form part of other clones (no 7, 11, 13 and 14).

Cluster 15 consists of 5 Beijing strains*.* Three other Beijing strains analysed gave unique MIRU-VNTR patterns.

Of 4 Haarlem strains analysed, we identified one clone (no 1) with two isolates, and 2 strains with totally different MIRU-VNTR patterns. Of 14 S

family strains analysed, 9 different MIRU-VNTR patterns were found, which formed part of clones no 7 (2 strains) and 8 (5 strains) and 7 strains had unique MIRU-VNTR patterns. Three of these unique strains form part of other clusters (no 4, 5, and 13).

The EAI family was split into two clones (no 16 and 17) comprising of 3 and 2 strains respectively and one unique MIRU-VNTR pattern. The strains in the two clones differed with one allele in three different loci.

Other families that were analysed, which gave unique MIRU-VNTR patterns but were found to be part of other clones are Family 33 - 35- and 36, and CAS. The CAS family was not found to be part of any clone.

4.3 Phylogenetic studies

Phylogenetic comparison of spoligo patterns UPGMA-Tree, Spoligo: Categorical (MIRU-VNTR*plus* DB) (Fig. 4.1) divided the strains into 2 major groups. The smaller group could be divided into 2 again with the smaller of the 2 consisting of 10 strains assigned probable lineages by SpolDB3, but with less than 14 spacers. When compared to a combined UPGMA-Tree, MIRU-VNTR [12]: Categorical (1), Spoligo: Categorical (1) (Fig. 4.1), 6 of these were grouped once again in a group with mixed strains. The larger group contained the Beijing strains and further unrecognisable strains with some distance relation to *M. canetti* and *M. microti* strains.

Overall, 91 spoligo patterns occurred only once (64%) and 13 clones comprised 51 isolates (36%) with this tree analysis (Fig 4.1). Two of the clones represented the dominant T family with the largest one consisting of the previously mentioned 7 SIT53 isolates and the other one consisting of 3 strains assigned SIT7. The 7 SIT53 isolates cloned together with ten Ghana

Figure 4-2: A combined MIRU-VNTR [12] Categorical (1) Spoligo: Categorical (1) UPGMA-similarity tree obtained with MIRU-VNTRplus DB.

Table 4.2 Results of spoligo-MIRU-VNTR typing arranged according to MIRU-VNTR clones using excel 2007.

M. tuberculosis	MIRU-VNTR	12 MIRU-VNTR patterns												SpolDB4	SpolDB3
Strain no	Clones	M 2	M 4	M10	M16	M20	M23	M24	M26	M27	M31	M39	M40	SIT no	Lineages
ZT09	1	2	2	4	3	2	5	1	5	3	3	2	2	62	Haarlem1
ZT33	1	2	2	4	3	2	5	1	5	3	3	2	2	Orphan	Haarlem1
GF05	2	2	3	3	2	3	5	1	5	3	2	2	3	Orphan	LAM3
ZT30	2	2	3	3	2	3	5	1	5	3	2	2	3	Orphan	Family33
BFN01	3	2	3	3	3	1	6	1	5	3	2	2	3	Orphan	Family36
Q05	3	2	3	3	3	1	6	1	5	3	2	2	3	Orphan	X3
Q24	3	2	3	3	3	1	6	1	5	3	2	2	3	53	T1
ZT70	3	2	3	3	3	1	6	1	5	3	2	2	3	Orphan	X3
GF12	4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	LAM3
BFN02	4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	LAM7
ZT03	4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	LAM7
ZT15	4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	LAM7
ZT32	4	2	2	3	2	3	6	1	5	3	2	2	3	811	LAM8
GF07	4	2	2	3	2	3	6	1	5	3	2	2	3	34	S
ZT28	4	2	2	3	2	3	6	1	5	3	2	2	3	1122	T1
ZT29	4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	T1
Q26	4	2	2	3	2	3	6	1	5	3	2	2	3	245	T1
ZT22	4	2	2	3	2	3	6	1	5	3	2	2	3	1737	T4
GF49	4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	T4
ZT62	4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	X1
ZT77	4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	X1
ZT73	4	2	2	3	2	3	6	1	5	3	2	2	3	336	X1/T1
GF54	4	2	2	3	2	3	6	1	5	3	2	2	3	92	X3
ZT43	4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	X3
Q70	4	2	2	3	2	3	6	1	5	3	2	2	3	Orphan	X3
GF51	5	2	3	3	2	3	6	1	5	3	2	2	3	44	T1
GF53	5	2	3	3	2	3	6	1	5	3	2	2	3	53	T1
Q11	5	2	3	3	2	3	6	1	5	3	2	2	3	Orphan	Family33
ZT49	5	2	3	3	2	3	6	1	5	3	2	2	3	Orphan	S
ZT54	5	2	3	3	2	3	6	1	5	3	2	2	3	53	T1
GF30	6	2	2	4	2	3	6	1	5	3	2	2	3	834	U
GF45	6	2	2	4	2	3	6	1	5	3	2	2	3	254	T5-RUS
GF18	7	2	2	3	3	3	6	1	5	3	2	2	3	Orphan	T1
GF20	7	2	2	3	3	3	6	1	5	3	2	2	3	Orphan	X2
Q04	7	2	2	3	3	3	6	1	5	3	2	2	3	Orphan	Family35
ZT16	7	2	2	3	3	3	6	1	5	3	2	2	3	71	S
ZT27	7	2	2	3	3	3	6	1	5	3	2	2	3	71	S
ZT66	7	2	2	3	3	3	6	1	5	3	2	2	3	Orphan	T1
GF26	8	2	2	3	3	2	5	1	5	3	3	2	3	34	S
GF31	8	2	3	3	3	2	5	1	5	3	3	2	3	Orphan	T2
ZT39	8	2	3	3	3	2	5	1	5	3	3	2	3	34	S
ZT60	8	2	3	3	3	2	5	1	5	3	3	2	3	34	S/T1
ZT61	8	2	3	3	3	2	5	1	5	3	3	2	3	34	S
ZT64	8	2	3	3	3	2	5	1	5	3	3	2	3	34	S

Isolate	Cluster	1	2	3	4	5	6	7	8	9	10	11	12	SIT	Family
Q20	9	2	2	4	2	3	6	1	5	3	5	2	1	813	LAM8
ZT08	9	2	2	4	2	3	6	1	5	3	5	2	1	815	LAM8
GF31	10	5	2	3	2	3	5	1	5	3	2	2	3	Orphan	T2
GF32	10	5	2	3	2	3	5	1	5	3	2	2	3	Orphan	T2
GF32	10	5	2	3	2	3	5	1	5	3	2	2	3	Orphan	T2
Q72	10	5	2	3	2	3	5	1	5	3	2	2	3	Orphan	T3
Q40	11	2	3	3	3	3	6	1	5	3	2	2	3	53	T1
Q45	11	2	3	3	3	3	6	1	5	3	2	2	3	Orphan	X3
ZT04	11	2	3	3	3	3	6	1	5	3	2	2	3	53	T1
ZT38	11	2	3	3	3	3	6	1	5	3	2	2	3	53	T1
ZT46	11	2	3	3	3	3	6	1	5	3	2	2	3	Orphan	Family36
ZT80	11	2	3	3	3	3	6	1	5	3	2	2	3	53	T1
Q71	11	2	3	3	3	3	6	1	5	3	2	2	3	Orphan	Family36
GF08	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3
GF13	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3
GF22	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3
Q18	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3
Q33	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3
Q36	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3
Q37	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3
ZT17	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3
ZT18	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3
ZT20	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3
ZT24	12	2	2	4	3	2	6	1	5	3	3	2	3	Orphan	LAM3
ZT42	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3
ZT44	12	2	2	4	3	2	6	1	5	3	3	2	3	33	LAM3
ZT50	12	2	2	4	3	2	6	1	5	3	3	2	3	Orphan	LAM3
GF21	13	2	2	3	2	3	6	1	5	3	4	2	3	1470	X1-LAM9
GF27	13	2	2	3	2	3	6	1	5	3	4	2	3	1154	LAM9
GF71	13	2	2	3	2	3	6	1	5	3	4	2	3	71	S
ZT48	13	2	2	3	2	3	6	1	5	3	4	2	3	162	LAM9
GF01	14	5	2	3	2	3	6	1	5	3	2	2	3	Orphan	LAM1
GF03	14	5	2	3	2	3	6	1	5	3	2	2	3	92	X3
BFN03	14	5	2	3	2	3	6	1	5	3	2	2	3	Orphan	T2
Q06	15	2	2	3	3	2	5	1	7	3	5	3	3	1	Beijing
Q08	15	2	2	3	3	2	5	1	7	3	5	3	3	1	Beijing
ZT68	15	2	2	3	3	2	5	1	7	3	5	3	3	796	Beijing
ZT69	15	2	2	3	3	2	5	1	7	3	5	3	3	1	Beijing
ZT71	15	2	2	3	3	2	5	1	7	3	5	3	3	1	Beijing
GF02	16	2	5	4	4	3	6	1	4	3	5	2	3	Orphan	EAI1
GF11	16	2	5	4	4	3	6	1	4	3	5	2	3	Orphan	EAI1
Q51	16	2	5	4	4	3	6	1	4	3	5	2	3	Orphan	EAI1
GF23	17	2	5	4	4	2	6	2	4	3	5	3	3	Orphan	EAI1
ZT19	17	2	5	4	4	2	6	2	4	3	5	3	3	Orphan	EAI1

ZT35	18	2	3	3	2	3	6	1	5	3	3	2	3	Orphan	T1
ZT57	18	2	3	3	2	3	6	1	5	3	3	2	3	Orphan	T1
BFN04	19	5	2	3	2	2	3	5	1	5	2	2	3	119	X1
BFN05	20	2	2	4	3	2	5	1	4	3	2	2	2	Orphan	X3
GF33	21	2	3	3	2	3	7	1	4	3	2	2	3	Orphan	T1
GF38	22	2	2	4	3	2	7	1	5	3	3	2	3	33	LAM3
GF40	23	2	2	0	0	3	6	1	5	3	2	2	3	Orphan	T1
GF44	24	2	2	3	2	3	5	1	4	3	2	2	3	53	T1
GF52	25	2	3	3	3	3	5	1	5	3	2	2	3	Orphan	X3
GF56	26	2	5	4	3	3	7	2	5	3	2	3	4	53	T1
GF57	27	2	7	4	3	2	6	2	2	3	4	1	4	Orphan	X1/T1
GF59	28	2	2	4	0	3	6	1	5	3	5	2	1	Orphan	LAM8
GF60	29	2	2	4	4	2	4	1	4	3	2	2	3	Orphan	*M. africanum*
GF61	30	2	2	4	0	3	7	1	5	3	2	2	3	Orphan	T3
GF62	31	2	3	3	2	3	6	1	5	3	2	2	4	53	T1
GF64	32	2	3	2	5	2	3	5	3	5	1	3	2	Orphan	T2
GF70	33	2	2	4	3	2	5	1	5	3	3	2	4	Orphan	X3
GF72	34	2	2	4	3	3	5	1	5	3	3	2	3	Orphan	X3
Q01	35	5	2	4	3	2	5	1	5	3	3	2	2	Orphan	T2
Q02	36	2	5	4	3	3	6	1	5	3	2	2	1	811	LAM8
Q03	37	2	2	3	2	3	5	1	5	3	2	2	3	92	X3
Q07	38	2	2	3	3	2	5	1	5	3	5	4	3	1	Beijing
Q10	39	2	5	4	3	3	6	1	5	3	2	2	3	Orphan	EAI5
Q12	40	2	3	5	4	2	4	1	5	3	2	3	3	26	CAS
Q17	41	2	2	4	3	3	6	1	5	3	3	2	3	33	LAM3
Q21	42	2	3	3	0	3	4	1	5	3	2	2	3	Orphan	T1
Q25	43	2	3	3	2	3	4	1	5	3	2	2	3	Orphan	Family35
Q28	44	2	3	4	3	11	6	1	5	3	2	2	3	53	T1
Q32	45	2	3	3	2	11	6	1	5	3	2	2	3	Orphan	S
Q35	46	2	2	3	0	3	6	1	5	3	2	2	3	Orphan	LAM3
Q39	47	2	2	3	3	2	5	1	5	3	2	2	3	Orphan	T1
Q41	48	2	2	3	3	3	5	1	4	3	2	2	3	Orphan	LAM9
Q43	49	2	3	3	3	3	5	1	5	3	2	2	4	12	X3
Q44	50	2	2	4	3	2	5	1	5	3	3	5	4	Orphan	X3
Q47	51	5	5	4	3	3	6	1	5	3	2	2	3	148	Haarlem1
ZT01	52	2	2	3	1	2	6	1	5	3	2	2	2	Orphan	X3
ZT05	53	2	3	3	3	3	6	1	5	3	2	2	4	119	X1
ZT06	54	2	3	3	2	3	6	1	5	3	2	2	2	Orphan	X1/T1
ZT10	55	2	3	3	3	3	3	1	5	3	2	2	3	Orphan	Family36
ZT13	56	5	3	3	3	3	5	1	5	3	2	2	3	Orphan	Family36
ZT14	57	2	2	3	2	1	6	1	1	3	2	2	3	Orphan	S
ZT23	58	2	3	3	3	3	6	1	4	3	3	2	3	Orphan	T1
ZT25	59	3	3	3	3	3	7	1	5	3	3	2	3	Orphan	LAM7
ZT31	60	2	5	4	2	3	6	1	4	3	5	2	3	Orphan	Haarlem 1
ZT36	61	5	3	3	3	3	6	2	5	3	3	2	3	Orphan	T1
ZT40	62	2	2	3	2	3	4	1	5	3	2	2	3	Orphan	LAM3
ZT52	63	2	2	3	2	3	6	2	5	3	4	2	3	Orphan	LAM9
ZT53	64	2	5	4	4	3	0	1	4	3	5	2	3	53	T1
ZT55	65	2	3	3	3	3	4	1	5	3	2	2	3	118	T1
ZT56	66	5	3	3	2	3	6	1	5	3	2	2	1	Orphan	S
ZT59	67	2	2	2	3	2	4	2	5	3	2	2	3	482	*M. bovis*-BCG
ZT63	68	2	3	3	0	3	5	1	5	3	2	2	3	Orphan	S/T1
ZT65	69	2	2	4	3	3	5	1	7	3	5	3	3	1	Beijing
ZT67	70	2	2	3	3	2	4	1	5	3	5	3	3	1	Beijing
ZT74	71	2	2	3	2	3	6	1	5	4	2	2	3	92	X3
ZT76	72	2	2	3	2	2	6	1	5	3	2	2	3	Orphan	T3
H37RV	Control	2'	3	2	5	2	3	2	3	6	1	3	2	451	H37Rv

type reference strains from the MIRU-VNTRplus DB that originated from Ghana and 1 T1 isolate cloned with 3 Ugandal I reference strains. The above mentioned T isolates in our study all grouped in a cluster of 21 isolates in the combined tree analyses in Fig. 4.2 that included three clones with 6 of the 10 SIT53, SIT7 isolates and 2 S isolates. The rest of the T strains were distributed over the complete dendrogram showing no relationship between them as expected.

The LAM cluster (17 isolates) included a LAM strain of the MIRU-VNTRplus DB, 11 X3 and 3 Family36 isolates diverged from the same inner node with an Ugandal strain. The 8 Beijing isolates correlated 100% with the Beijing strains and 6 S family isolates to the S family. Two LAM9, a LAM1 and X1-LAM9 correlated to LAM strain and 2 X1 isolates correlated with X strains. An EAI5 and a Haarlem isolate correlated to EAI1 strains, while a CAS isolate was found to correlate with Delhi/CAS strains.

The second most dominant LAM family is represented by 2 clones with the LAM3/SIT33 (13 isolates) being the biggest clonal group of FS strains. The X family was identified as the third most dominant with 3 clones each consisting of at least two isolates. Minor families which were found to form clones were the S family (4 clones), Family 36 (1 clone) and Beijing (1 clones) as seen in Fig. 4.1.

With the UPGMA-Tree Fig. 4.1 (comparing the MIRU-VNTR and spoligotyping patterns) only 6 clones were found having both identical MIRU-VNTR and spoligo patterns with the most significant one containing 12 of the 13 LAM3/SIT33 isolates. Five Beijing isolates (SIT1) was the second largest clone and the other four consisted of the two T1 (SIT53) and two S (SIT34 and SIT71) groups mentioned before. These clonal groups represent 18.3% of the strains analysed. Isolates showing 100% similarity (combined allocation of the MIRU-VNTRplus DB) from the inner nodes,

representing a cluster with almost identical MIRU-VNTR patterns were colour coded different from yellow. Combination of the two typing method resulted in 20 clusters comprising of 78 isolates. The T isolates as shown in Fig. 4.1 is represented in 4 (20%) different clusters comprising of 13 isolates, the LAM family in 4 (20%) of 24 isolates and the X family in 4 (20%) clusters comprising of 14 isolates. These three families T, LAM and X represent 60% of the clusters obtained by the combined methods. One cluster was found for each of the following families Beijing (8), EAI (5), Haarlem (2), 2 clusters for S (5) and 2 clusters for Family 36 (5) respectively. Sixty-four unique patterns were found consisting of T (24) (16.9%), LAM (10) (7%), X (11) (7.8%), EAI (2) (1.4%), Haarlem (2) (1.4%), Family 33 (2) (1.4%), Family 35 (2) (1.4%), S (7) (4.9%), U (1) (0.7%), *M. africanum* (1) (0.7%)*, M. bovis* (1) (0.7%), and CAS (1) (0.7%).

4.4 *Multiplex PCR to differentiate Beijing strains*

The results of 8 isolates found to belong to the Beijing family by spoligotyping and further characterised to establish the presence of the W/Beijing type in the FS province, using a multiplex PCR assay are shown in Fig. 4.3.

For 5 strains (Q08, Q07, Q06, ZT68 and ZT71) three-fragments were amplified with molecular weights of 523, 223 and 175 bp indicative of the W Beijing type. Strains ZT67, ZT69 and H37Rv (control) contained two fragments (523 and 150 bp) each. The 523 band serves as an internal positive control and is present in each lane. ZT65 (lane 9) did not amplify, but could not be repeated due to insufficient DNA (Fig 4-3).

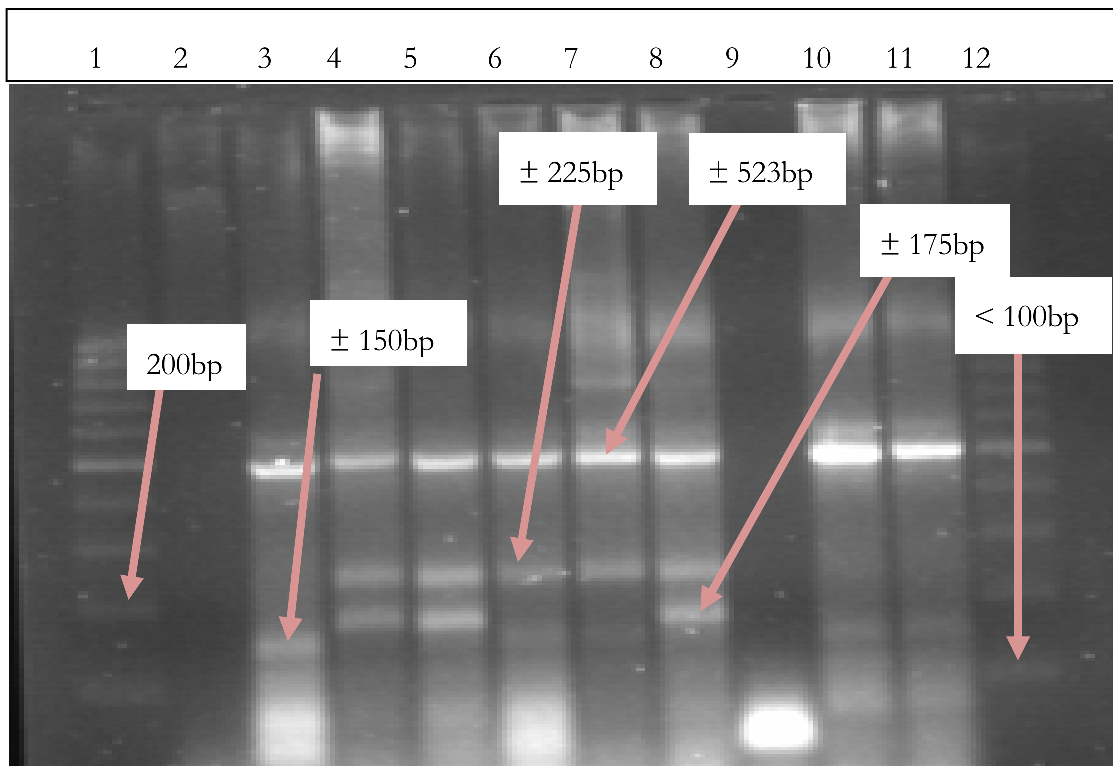

Figure 4-3: Amplification products from a multiplex PCR assay for W/Beijing strain identification.

DNA templates were MTB of the Beijing type as follows. From lane 1 (100bp DNA ladder), 2 (Negative control), 3 (ZT67), 4 (Q08), 5 (Q07), 6 (Q06), 7 (ZT68), 8 (ZT71), 9 (ZT65), 10 (ZT69), 11 (Positive control, H37Rv), and 12 (100bp DNA ladder).

Chapter 5

Discussion

In our study the diversity of the spoligotyping results were extremely high, but the reason for this is not clear. At one stage during the experimental phase of the study our fridge broke down and the temperature rose unacceptable high, but clear spoligotypes were still found and results from isolates that were spoligotyped before and after the event did not give obviously worse results. MIRU-VNTR typing was done after the event and good results were obtained through-out. The DNA of the isolates were extracted in 2001-2003 and normal degrading of DNA over time might have been responsible for especially the many orphan strains found with spoligotyping since this technique amplify a larger stretch of DNA than when MIRU-VNTR loci are amplified. The results obtained with repeated testing of some isolates however did not support this idea of DNA degradation. Differences observed between initial and follow-up isolates of the same patient could reflect the phenomenon of reinfection, as seen in various studies by indicating that reinfection in high burden communities are common, especially among HIV-positive subjects (De Boer *et al.,* 2003, Van Rie *et al.,* 1999., Shamputa *et al.,* 2004). The specific reasons for the higher variability among spoligotypes as compared to MIRU-VNTR profiles thus remain unclear. The MIRU-VNTR data show similar variability as the RFLP data from a previous study with the same strains (van der Spoel van Dijk *et al.,* 2005).

The diversity and history of MTB in the Western Cape and KZN have been described before with the Beijing lineage as one of the most prominent (Streicher *et al.* 2004). Until 2009, in the FS province, IS*6110* RFLP typing data was the only information available for a representative sample of *M. tuberculosis* strains. In this study, 142 isolates from the FS were genotyped by spoligotyping and MIRU-VNTR typing, revealing the presence of a very diverse population of strains that included 3 predominant lineages: the ill defined T, LAM and X families. This data is in concordance with a recent study by Stavrum *et al* that was the first to report typing data from 8

provinces in SA. They reported that the FS had the second most diverse population of strains, which could mean that the population is conducive to mutation. They reported that T lineage was the most prominent amongst 25 FS isolates and in the Gautang province. The LAM and X lineages occurred in all the provinces with the LAM being the most prominent in the Eastern Cape and KZN, while the highest frequencies of the X lineage was seen in the Western Cape and Northern province (Stavrum *et al*. 2009).

The rate of diversity among the T family in our study is high (76 %) compared to the LAM family and could indicate that the T family is adapted best to the FS. This could also mean that there is selective pressure for the T family and other lineages with the loss of spacers 33-36. T1 (ST53) isolates were the most prominent amongst the FS stains as was reported for the 8 provinces by Stavrum *et al.* (Stavrum *et al*. 2009). This is in concordance with the ranking of this sublineage in the SpolDB4 (Brudey *at al.* 2006a). The FS T strains all miss spacers 33-36, but this is the only conserved region in this family. This finding is in agreement with global published literature (Frothingham *et* al., 1998, Soini *et* al., 2000, Streicher *et al*., 2007, Gutierrez *et al*., 2006, Eldholm *et al*., 2006, Dou *et al*., 2008). These strains can hardly be classified into one cluster as this family is distributed among clusters (Fig 4.1 and Fig 4.2). Sixteen isolates (40%) of the T family in our study were found in Motheo and Lejweleputswa districts respectively and 20% (16) in Thabo Mofutsanyana district as shown in Table 4.4a and b. This agrees with previous RFLP typing data of these strains from the FS that has been published by our department as they are not grouped as one cluster (van der Spoel van Dijk *et. al.,* 2005).

Comparison of all strains to international strains on the MIRU-VNTRplus database shows that T family can be linked to Ugandan (1 strain), Ghanan (8 of the T1/SIT53 strains) and X1 (2 X1/SIT34 strains) as shown on Fig 4.2. On Fig. 4.2 2 T strains GF56 and Q39 were found to correlate with EAI1 and

Ugandan strains respectively. While combination of MIRU-VNTR-spoligo typing on Fig 5.4 showed that Q38 and GF44 T strains correlated with Cameroon and Ugandan strains respectively. The T family is shown to be distributed globally with less information of its origin.

The rate of diversity of the LAM family is low (56%) compared to the T and X families that were also found in high numbers in the province. The LAM family is the second largest group in the FS accounting for 23.2% of all strains; this clearly correlate to previous studies indicating that LAM family is one of the most prevalent and widely distributed lineages in the world (Brudey *et al.*, 2006a; 2006b). The LAM family was distributed among the three districts as follows 48.5% (16) in Motheo, 27.3% (9) in Thabo Mofutsanyana and 24.2% (8) in Lejweleputswa (Table 4.2). However, this genotype family is more diverse as it consists of 12 different sub-families and its study is more complicated than initially thought (Table 4.2 and 4.3, Brudey *et al.*, 2006b).

The LAM family in our study was defined by equilibrium absence of spacers, 21-24 and 33-36 (table 4.1) by spoligotyping and this correlates with what was published in the literature (Sola *et al.*, 2001) except for one LAM7 strain (ZT25) that showed the presence of spacer 21.

The LAM3 subfamily in our study also missed spacers 9-11 and was found to be the largest clone of strains present in the FS as demonstrated both by spoligotyping (Fig 4.1 and 4.2) and MIRU-VNTR typing (clone 12). This agrees with RFLP typing data in the literature where the strains (GF05, GF08, ZT03, ZT44, ZT40, Q35, Q36, Q37, GF12, GF13, ZT43 and Q33) were present in a cluster (van der Spoel van Dijk *et al.*, 2005). This subfamily was reported as the F11/ST33 strain in the Western Cape where it is highly successful and most probable diverged to the FS (Victor *et al.,* 2004, Warren *et al.,* 2002, 2004).

The LAM7 subfamily in our study was found to miss spacers 7-12, 26-29, 31, 39-40 and 42-43. The LAM8 subfamily missed spacers 28-31 and 40. Our LAM strain correlates with previous reports that there is no common presence of spacer 21-22 in these lineages (Niobe-Eyangoh *et al.*, 2003, Zozio *et al.*, 2005). There was disequilibrium absence of spacers 15-17, 19-20 and 40-43 for the LAM9 subfamily.

FS LAM9 strains differ from LAM9 RD^{Rio} strains from Rio de Janeiro, Brazil belonging to the LAM9 family characterized by the absence of spacers 21-24, and 33-36 only (Lazzarini *et al.*, 2007). The two LAM9 strains (GF27 and ZT48) differ from the deadly KZN/F15/LAM4 strain by one spacer as they miss spacer 39 and 41 respectively instead of missing spacer 40 identified as the KZN XDR-TB strain (Pillay *et al.*, 2007).

Comparison of FS LAM strains to reference strains in MIRU-VNTRplus database showed that there is agreement between MIRU-VNTR patterns (Fig. 4.2) and spoligopatterns (Fig 4.2) as they are closely related to Cameroon LAM strains. Combined methods (MIRU-VNTR-spoligotyping) analysis to reference strains (fig 5.4) showed that the LAM8 strain Q70 in our study was identical to LAM11_ZWE SIT833. A spoligotyping study in the Ouest province of Cameroon reported that 47% of their *M. tuberculosis* isolates belong to the Cameroon family (LAM10-Cam) (Niobe-Eyangoh *et al.*, 2004).

The genetic diversity of MTB in Dar es Salaam using spoligotyping, reported high prevalence of LAM family and in Zimbabwe, 32% of their MTB isolates are LAM-ZWE variants (Eldholm *et al.*, 2006, Easterbrook *et al.*, 2004). Correlation to African countries thus shows that this family is circulating globally and has spread throughout Africa. The LAM families originates from Latin American countries and were brought to Africa and South Africa by settlers and Africans migrating to international countries in search of better jobs as high flow of immigrants as was seen after 1994. The rate of diversity

may be due to the movement of migrants to the city, Bloemfontein and then mixed with South Africans who also converged in the center city on their way to other places. One has to go through Bloemfontein when traveling by road from the North and South using public transport, which is the case with most of the migrant workers. This could then be the route of spread to Thabo Mofutsanyana and Lejweleputswa.

The rate of diversity of the X family is very high (80%) compared to the T and LAM families which represents the highest number of isolates in the province. X family of strains is defined by two correlated features, a low number of IS*6110* copies and the absence of spacer 18 in the spoligotyping (Sebban *et al*., 2002). This agrees with our current results and that of van der Spoel van Dijk in previous RFLP typing data where 24 (16.7%) isolates had low (n<6) IS*6110* RFLP copies (van der Spoel van Dijk *et al.,* 2005). Moreover there is a common absence of spacers 5-12 especially to X3 subfamily as shown in table (Table 4.1). ST119, SIT137, and SIT92 have been mentioned widely in the literature and of these SIT119 and SIT92 were found to be the two common types in the FS. Both characteristics (missing spacer 18 and low copy numbers of IS*6110* are present in the CDC1551 strain, which was once suggested to be highly virulent, (Bishai *et al*., 1999, Cowan *et al*., 2002). These support the assumption that in the history of the FS this virulent strain was present. Three Lejweleputswa X3, 4 Motheo and 5 Thabo Mofutsanyana X3 strains including 2 X1 strains from the Motheo districts were found to correlate with Ugandal strain in Fig 4.2. The X family was also reported in Guadeloupe (Sola *et al*., 1997), the French Polynesia (Torrea *et al*., 1995), Anglo-Saxon countries (Dale *et al*., 2003), the Western Cape Province of South Africa (Streicher *et al*., 2004) and is highly prevalent in Kwazulu Natal (Pillay *et al.,* 2006). Combination of MIRU-VNTR–spoligo typing does not show much correlation of our X family strains to other countries. The X strains showed to have originated from South Africa and other parts of the world such as French Guiana and Guadelope.

The 7 isolates identified as Beijing genotypes with a specific spoligotype signature (absence of spacer 1-33, presence of spacer 34-43) and one Beijing-like strain that missed an extra spacer (34) is well known and have been characterized extensively (van Soolingen *et al.,*1995, Kamerbeek *et al.*, 1997, Qian *et al.*, 1999, Kremer *et al.*, 2004b, 2005a). None of the low IS*6110* RFLP copy isolates were identified as Beijing genotypes. Five of our Beijing strains belonged to the W type, which caused a notorious outbreak in New York at the beginning of '90s due to a multidrug resistant mutant strain (Plikaytis *et al.*, 1994, Bifani *et al.*, 2002). The same strain is also present in great numbers in the Western Cape, and was reported by Stavrum *et al* as the second most prevalent lineage in SA (10.3%), but probable due to the limited travel between these two provinces in the apartheid era, it still contributes only 5.6% of the FS isolates (Hanekom *et al.,* 2007, Stavrum *et al.,* 2009). The Beijing strains originate from China where most of the MTB population consists of Beijing types (Glynn *et al.,* 2002). Van Sooligen and colleagues hypothesized that this genotype emerged successfully in East Asia due to mass BCG vaccination during the 20th century (van Soolingen *et al.*, 1995, Abebe *et al.*, 2006).

Six EAI strains were identified belonging to two subfamilies EAI1 (5) and EAI5 (1). Two (EAI1) strains were characterized by low *IS6110* RFLP copies as been reported on the previous studies (Källenius *et al.*, 1999 Kremer *et al.*, 1999), while the remaining 3 (EAI1) strains and 1 (EAI5) strain comprise of high *IS6110* RFLP copies (n> 6). The EAI5 strain demonstrated congruence between spoligotypes as been reported before (absence of spacers 29-32 and 34, presence of spacer 33) (Soini *et al.*, 2000), moreover missing extra spacers 4-9, 15, and 23-24. This lineage was first reported in Guinea-Bissau (Källenius *et al.*, 1999) and later was shown to be frequent in South-East Asia, India, and East Africa (Kremer *et al.*, 1999). Subgroup of these strains gaining a single copy of IS*6110* was shown to be widespread in areas like in Malaysia, Tanzania, and Oman (Fomukong *et al.*, 1994). It is

believed that this lineage, which is endemic in South-East Asia, South-India, and East-Africa, may have originated in Asia, where TB could have historically found favorable spreading conditions (Källenius *et al*., 1999, Kremer *et al.,* 1999).

Chapter 6

Conclusion

In conclusion, there is a high strain diversity of MTBC strains in the FS province which correlates nicely to the history of South Africa immigrants, missionaries and settlers who came already infected in the country. On the other hand, apartheid regime policies of not allowing other tribes free access to other areas of the country facilitated favourable breeding environment and spreading of the infection within the community especially if one or more family members were infected.

Nevertheless, it remains hard to determine the detailed history of MTBC strains in the FS province. Given the fact that both the X3 and Beijing strains - predominant in the Western Cape – are less important in the FS, it is more likely that MTB in the FS migrated down from the Gautang province and Northern countries including Uganda, Ghana, Zimbabwe and Cameroon where similar lineages are present.

Three dominant families in the Free State (the “ill define” T, LAM and X) account for more than two-thirds of the MTBC strains identified. Identification of this T family correlates with the history of TB in the country as the presence of the European dates back to the beginning of the 20th century; they are linked to the X family. LAM family is found worldwide. Some of our T-family strains are linked to other African strains such as the Cameroon strains, sharing the same spoligo pattern with Tanzanian strains and strains from the Western Cape Province in South Africa. This indicates that the T family is best adapted within these communities. Presence of other identified strains could be linked to migration in the country.

The presence of only few Beijing-family strains within Motheo and Thabo Mofutsanyana districts could be linked to the fact that these two areas comprise factories with a history of Chinese ownership.

The data obtained by 12MIRU-VNTR typing was far too diverse to conclude on any possible pattern uniquely to the FS and much more strains will have to be analysed and 15MIRU-VNTR typing performed before MIRU-VNTR typing can be considered for use as a tool for specific patterns as been suggested in the objectives number 3.

This study served as a pilot as it contains isolates collected almost a decade ago. There still remains a need to type current strains. Because there have been reports about high legal and illegal immigrants in the country, high global economic partnership allowing people coming and going out of the country.

Chapter 7

References

Abebe, F., and Bjune, G., 2006.The emergence of Beijing family genotypes of *Mycobacterium tuberculosis* and low-level protection by bacille Calmette-Guerin (BCG) vaccines: is there a link? Clinical and Experimental Immunology.145:389-397.

Alcaide, F., Benítez, M.A., Escribà, J.M., and Martín, R., 2000. Evaluation of the BACTEC MGIT 960 and the MB/BacT Systems for Recovery of Mycobacteria from Clinical Specimens and for Species Identification by DNA AccuProbe. Journal of Clinical Microbiology. 38:398-401.

Alix, E., Godreuil, I.S., and Blanc-Potard, A.B., 2006. Identification of a Haarlem genotype-specific single nucleotide polymorphism in the mgtC virulence gene of *Mycobacterium tuberculosis.* Journal of Clinical Microbiology. 44:2093-2098.

Alland, D., Lacher, D.W., Hazbon, M.H., Motiwala, A.S., Qi, W., Fleischmann, R.D., and Whittam, T.S., 2007. Role of large sequence polymorphisms (LSPs) in generating genomic diversity among clinical isolates of *Mycobacterium tuberculosis* and the utility of LSPs in phylogenetic analysis. Journal of Clinical Microbiology. 45:39-46.

Allix, C., Supply, P., and Fauville-Dufaux, M., 2004. Utility of fast mycobacterial interspersed analysis. Clinical Infectious Diseases. 39:783-789.

Allix-Beguec, C., Harmsen, D., Weniger, T., Supply, P., and Niemann, S., 2008. Evaluation and user-strategy of MIRU-VNTRplus, multifunctional database for online analysis of genotyping data and phylogenetic identification of *Mycobacterium tuberculosis* complex isolates. Journal of Clinical Microbiology. 46:2692-2699.

American Thoracic Society / Centers for Disease Control and Prevention / Infectious Diseases Society of America: Treatment of tuberculosis. American Journal of Respiratory and Critical Care Medicine. 2003. 167:604-661.

Arriaza, B.T., Salo, W., Aufderheide, A.C., and Holcomb, T.A., 1995. Pre-Columbian tuberculosis in northern Chile: molecular and skeletal evidence. American Journal of Physical Anthropology. 98:37-45.

Badak, F.Z., Kiska, D.L., Setterquist, S., Hartley, C., O'Connell, M.A., and Hopfer, R.L., 1996. Comparison of mycobacteria growth indicator tube with BACTEC 460 for detection and recovery of mycobacteria from clinical specimens. Journal of Clinical Microbiology. 34:2236-2239.

Barnard, M., Albert, H., Coetzee, G., O'Brien, R., and Bosman, M.E., 2008. Rapid molecular screening for multidrug-resistant tuberculosis in a high volume public health laboratory in South Africa. American Journal of Respiratory and Critical Care Medicine. 177:787-792

Behr, M.A., and Small, P.M., 1999. Molecular fingerprinting of *Mycobacterium tuberculosis*. How can it help the clinician? Clinical Infectious Diseases. 25:806-810.

Bhanu, N.V., van Soolingen, D., van Embden, J.D., Dar, L., Pandey, R.M., and Seth, P., 2002. Predominance of a novel *Mycobacterium tuberculosis* genotype in the Delhi region of India. Tuberculosis. 82:105-112.

Bifani, P.J., Mathema, B., Kurepina, N.E., and Kreiswirth, B.N., 2002. Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. Trends Microbiology. 10:45-52.

Bifani, P.J., Mathema, B., Lui, Z., Moghazeh, S.L., Shopsin, B., Tempalski, B., Driscoll, J., Frothingham, R., Musser, J.M., Alcabes, P., and Kreiswirth, B.N., 1999. Identification of a W variant outbreak of *Mycobacterium tuberculosis* via population-based molecular epidemiology. Journal of American Medical Association. 282:2321-2327.

Bishai, W.R., Dannenberg, A.M., Parrish, N., Ruiz, R., Chen, P., Zook, B.C., Johnson, W., Boles, J.W., and Pitt, M.L., 1999. Virulence of *Mycobacterium tuberculosis* CDC1551 and H37Rv in rabbits evaluated by Lurie's pulmonary tubercle count method. Infection and Immunity. 67:4931-4934.

Blackwood, K.S., Wolfe, J.N., and Kabani, A.M., 2004. Application of Mycobacterial Interspersed Repetitive Unit Typing to Manitoba Tuberculosis Cases: Can Restriction Fragment Length Polymorphism Be Forgotten? Journal of Clinical Microbiology. 42:5001-5006.

Blumberg, L., 1995. Developments in tuberculosis. The Southern African Journal of Epidemiology and Infection. 10:118-122.

Borgdorff, M.W., de Haas, P., Kremer, K., and van Soolingen, D., 2003. *Mycobacterium tuberculosis* Beijing genotype, the Netherlands. Emerging Infectious Disease. 9:1310-1313.

Bradshaw, D., Botha, H., Joubert, G., Pretorious, J.P., Van Wyk, R., and Yach, D., 1987. Review of South African mortality (1984). Parow: South African Medical Reasearch Council.

Brosch, R., Gordon, S.V., Billault, A., Ganier, T., Eiglmeier, K., Soravito, C., Barrell, B.G., and Cole, S.T., 1998. Use of a *Mycobacterium tuberculosis* H37RV bacterial artificial chromosome library for the genome mapping, sequencing, and comparative genomics. Infection and Immunity. 66:2221-2229.

Brosch, R., Gordon, S.V., Marmiesse, M., Brodin, P., Buchrieser, C., Eigmeier, K., Ganier, T., Gutierrez, C., Hewison, G., Kremer, K., Parsons, L.M., Pym, A.S., Samper, S., van Soolingen, D., and Cole, S.T., 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proceeding of the National Academy of Sciences of U S A. 99:3684-3689.

Brudey, K., Driscoll, J.R., Rigouts, L., Prodinger, W.M., Gori A., Al-Hajoj, S.A., Allix, C., Aristimuño, L., Jyoti, A., Baumanis, V., Binder, L., Cafrune, P., Cataldi, A., Cheong, S., Diel, R., Ellermeier, C., Evans, J.T., Fauville-Dufaux, M., Ferdinand, S., de Viedma, D.G., Garzelli, C., Gazzola, L., Gomes, H.M., Guttierez, M.C., Hawkey, P.M., van Helden, P.D., Kadival, G.V., Kreiswirth, B.N., Kremer, K., Kubin, M., Kulkarni, S.P., Liens, B., Lillebaek, T., Ly, H.M., Martin, C., Martin, C., Mokrousov, I., Narvskaïa, O., Ngeow, Y.F., Naumann, L., Niemann, S., Parwati, I., Rahim, Z., Rasolofo-Razanamparany, V., Rasolonavalona, T., Rossetti, M.L., Rüsch-Gerdes, S., Sajduda, A., Samper, S., Shemyakin, I.G., Singh, U.B., Somoskovi, A., Skuce, R.A., van Soolingen, D., Streicher, E.M., Suffys, P.N., Tortoli, E., Tracevska, T., Vincent, V., Victor, T.C., Warren, R.M., Yap, S.F., Zaman, K., Portaels, F., Rastogi, N., Sola, C., Prodinger, W.M., Aristimuño, L., Arora, J., Diel, R., Ellermeier, C., Kubin, M., Kulkarni, S.P., Liens, B., Lillebaek, T., Parwati, I., and Zeaur, R., 2006a. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BioMedCentral Microbiology.6:23 doi:10.1186/1471-2180-6-

23.

Brudey, K., Filliol, I., Ferdinand, S., Guernier, V., Duval, P., Maubert, B., Sola, C., and Rastogi, N., 2006b. Long-term population-based genotyping study of *Mycobacterium tuberculosis* complex isolates in the French departments of the Americans. Journal of Clinical Microbiology. 44:183-191.

Brunello, F., Favari, F., and Fontana, R., 1999. Comparison of the MB/BacT and BACTEC 460 TB Systems for Recovery of Mycobacteria from Various Clinical Specimens. Journal of Clinical Microbiology. 37:1206-1209.

Caimi, K., Romano, M.I., Alito, A., Zumarraga, M., Bigi, F., and Cataldi, A., 2001. Sequence Analysis of the Direct Repeat Region in *Mycobacterium bovis*. Journal of Clinical Microbiology. 39:1067-1072.

Centre for Disease Control and Prevention, 1990. Screening for tuberculosis and tuberculosis infection in high risk populations. Recommendations of the Advisory committee for elimination of tuberculosis. MMWR. 39:1-7.

Center for Disease Control (CDC), 2003.

Cohan, F.M., 2002. What are bacterial species? Annual Review Microbiology. 56:457-487.

Cohn, D.L., 1997. Use of the Bacille Calmette-Guerin Vaccination for the Prevention of Tuberculosis: Renewed Interest in an Old Vaccine. American Journal of the Medical Sciences. Mycobacterial Diseases. 313:372-376.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, MA., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S., and Barrell, B.G., 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature. 393:537-544.

Collins, T.F.B., 1982. The history of southern Africa's first tuberculosis epidemic. South African Medical Journal. 62:780-788.

Consensus Statement., 2003. A prioritized research agenda for DOTS-Plus for multidrug- resistance tuberculosis MDRTB. Stop TB working group on DOTS-Plus for MDRTB. International Journal of Tuberculosis and Lung Disease. 7:410-414.

Constant, P., Perez, E., Malaga, W., Laneelle, M.A., Saurel, O., Daffe, M., and Gilhot, C., 2002. Role of the *pks*15/1 gene in the biosynthesis of phenolglycolipids in the *Mycobacterium tuberculosis* complex. Evidence that all strains synthesize glycosylated p-hydroxybenzoic methly esters and those strains devoid of phenolglycolipids harbor a frameshift mutation in the *pks*15/1 gene. Journal of Biological Chemistry. 277:38148-38158.

Cowan, L. S., Diem, L., Monson T., Wand, P., Temporado, D., Oemig, T. V., and Crawford, J. T., 2005. Evaluationof a Two-Step Approach for Large-Scale, Prospective Genotyping of *Mycobacterium tuberculosis* Isolates in the United States. Journal of Clinical Microbiology. 43:688-695.

Cowan, L.S., and Crawford, J.T., 2002. Genotype analysis of the *Mycobacterium tuberculosis* isolates from a sentinel surveillance population. Emerging Infectious Diseases. 8:1294-1302.

Cox, R., and Mirkin, S.M., 1997. Characteristic enrichment of DNA repeats in different genomes. Proceedings of the National Academy Science of the United States of America. 94:5237-5242.

Cronje, L., and Barker, C.H., 2006. Tuberculosis in the Free State province: Present trends, Future prognosis. South African Geographical Journal. 88:39-47.

Dale, J.W., Al-Ghusein, H., Al-Hashmi, S., Butcher, P., Dickens, A.L., Drobniewski, F., Forbes, K.J., Gillespie, S.H., Lamprecht, D., McHugh, T.D., Pitman, R., Rastogi, N., Smith, A.T., Sola, C., and Yesilkaya, H., 2003. Evolutionary relationships among strains of *Mycobacterium tuberculosis* with few copies of IS*6110*. Journal of Bacteriology. 185:2555-2562.

De Boer, A.S., Borgdoff, M.W., Vynnycky, E., Sebek, M.M., and Soolingen, D., 2003. Exogenous re infection as a cause of recurrent tuberculosis in

alow incidence area. International Journal of Tuberculosis and Lung Disease. 7:145-152.

Department of Health (DOH), RSA., 1998. *Tuberculosis: A training manual for health workers.* Cape Town: Cape and Transvaal book Printers.

Dou, H-Y., Tseng, F-C., Lin, C-W., Chang, J-R., Sun, J-R., Tsai, W-S., Lee, S-Y., Su, I-J., and Lu, J-J., 2008. Molecular epidemiology and evolutionary genetics of *Mycobacterium tuberculosis* in Tapei. BioMedCentral Infectious Diseases. doi:10.1186/1471-2334-8-170.

Douglas, J.T., Qian, L., Montoya, J.C., Musser, J.M., van Embden, J.D.A., van Soolingen, D., and Kremer, K., 2003. Characterization of the Manila family of *Mycobacterium tuberculosis*. Journal of Clinical Microbiology. 41:2723-2726.

Dubovsky, H., 1987. Rural tuberculosis in the Republic of South Africa. Public Health Journal. 101:85-91.

Easterbrook, P.J., Gibson, A., Murad, S., Lamprecht, D., Ives, N., Ferguson, A., Lowe, O., Mason, P., Ndudzo, A., Taziwa, A., Makombe, R., Mbengeranwa, L., Sola, C., Rostogi, N., and Drobniewski, F., 2004. High Rates of Clustering of Strains Causing Tuberculosis in Harare, Zimbabwe: a Molecular Epidemiological Study. Journal Clinical Microbiology. 42:4536-4544.

Edginton, M., 2000.Tuberculosis in South Africa: TB-past, present and future. Health Sytem Trust Update. 56:3-7.

Eldholm, V., Matee, M., Sayoki, G.M., Heun, M.M., and Dahle, U.R., 2006. A first insight into the genetic diversity of *Mycobacterium tuberculosis* in Dar es Salaam, Tanzania, assessed by spoligotyping. BioMedCentral Microbiology. 6:76. doi: 10.1186/1471-2180-6-76.

Espinal, M. A., Adalbert Laszlo, P.H., Simonsen, L., Boulahbal, F., Kim, S.J., Reniero, A., Hoffner, S., Rieder, H.L., Binkin, N., Dye, C., Williams, R., and Raviglione, M.C., 2001. Global Trends in Resistance to Antituberculosis Drugs for the World Health Organization–International Union against Tuberculosis and Lung Disease Working Group on Anti-Tuberculosis Drug Resistance Surveillance. The New England Journal of Medicine. 344:129-133.

Filliol, I., Motiwala, A.S., Cavatore, M., Qi, W., Hazbón, M.H., del Valle, M.B., Fyfe, J., García-García, L., Rastogi, N., Sola, C., Zozio, T., Guerrero, M.I., León, C.I., Crabtree, J., Angiuoli, S., Eisenach, K.D., Durmaz, R., Joloba, M.L., Rendón, A., Sifuentes-Osornio, J., de León, A.P., Cave, M.D., Fleischmann, R., Whittam, T.S., and Alland, D., 2006. Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. Journal of Bacteriology. 188:759-772.

Filliol, I., Driscoll, J.R., van Soolingen, D., Kreiswirth, B.N., Kremer, K., Valétudie, G., Anh, D.D., Barlow, R., Banerjee, D., Bifani, P.J., Brudey, K., Cataldi, A., Cooksey, R.C., Cousins, D.V., Dale, J.W., Dellagostin, O.A., Drobniewski, F., Engelmann, G., Ferdinand, S., Gascoyne-Binzi, D., Gordon, M., Gutierrez, M.C., Haas, W.H., Heersma, H., Källenius, G., Kassa-Kelembho, E., Koivula,T., Ly, H.M., Makristathis, A., Mammina, C., Martin, G., Moström, P., Mokrousov, I., Narbonne, V., Narvskaya, O., Nastasi, A., Niobe-Eyangoh, S.N., Pape, J.W., Rasolofo-Razanamparany, V., Ridell, M.M., Rossetti, L., Stauffer, F., Suffys, P.N., Takiff, H., Texier-Maugein, J., Vincent, V., de Waard, J.H., Sola, C., and Rastogi, N., 2002. Global distribution of *Mycobacterium tuberculosis* spoligotypes. Emerging Infectious Diseases. 8:1347-1349.

Filliol, I., Driscoll, J.R., van Soolingen, D., Kreiswirth, B.N., Kremer, K., Valétudie, G., Anh, D.D., Barlow, R., Banerjee, D., Bifani, P.J., Brudey, K., Cataldi, A., Cooksey, R.C., Cousins, D.V., , Dale, J.W., Dellagostin, O.A., Drobniewski, F., Engelmann, G., Ferdinand, S., Gascoyne-Binzi, D., Gordon, M., Gutierrez, M.C., Haas, W.H., Heersma, H., Källenius, G., Kassa-Kelembho, E., Koivula, T., Ly, H.M., Makristathis, A., Mammina, C., Martin, G., Moström, P., Mokrousov, I., Narbonne, V., Narvskaya, O., Nastasi, A., Niobe-Eyangoh, S.N., Pape, J.W., Rasolofo-Razanamparany, V., Ridell, M.M., Rossetti, L., Stauffer, F., Suffys, P.N., Takiff, H., Texier-Maugein, J., Vincent, V., de Waard, J.H., Sola, C., and Rastogi, N., 2003. Snapshot of moving and expanding clones of *Mycobacterium tuberculosis* and their global distribution assessed by spoligotyping in an international study. Journal of Clinical Microbiology. 41:1963-1970.

Filliol, I., Sola, C., and Rastogi, N., 2000. Detection of a previously unamplified spacer within the DR locus of *Mycobacterium tuberculosis*:

epidemiological implications. Journal of Clinical Microbiology. 38:1231-1234.

Fomukong, N.G., Tang, T.H., Al-Maamary, S., Ibrahim, W.A., Ramayah, S., Yates, M., Zainuddin, Z.F., Dale, J.W., 1994. Insertion sequence typing of *Mycobacterium tuberculosis*: characterization of a widespread subtype with a single copy of IS6110. Tubercle Lung Diseases. 75:435-440.

Free State Department of Health (FSDOH), 2002. Free State Government tuberculosis database summary. Cape Town: CTP Book Printers. <http://www.fshealth.gov.za/portal>

Frothingham, R., and Meeker-O´Connell, W.A., 1998. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeat. Microbiology. 144:1189-1196.

Gagneux, S., DeRiemer, K., Van, T., Kato-Maeda, M., de Jong, B.C., Narayanan, S., Nicol, M., Niemann, S., Kremer, K., Gutierrez, M.C., Hilty, M., Hopewell, P.C., Small, P.M., 2006. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. Proceedings of the National Academy of Sciences U S A. 103:2869-2873.

Gandhi, N.R., Moll, A., Sturm, A .W., Pawinski, R., Govender, T., Lalloo, U., Zeller, K.., Andrews, J., and 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. Lancet. 368:1575-1580.

Gascoyne-Binzi, D.M., Barlow, R.E., Essex, A., Gellitlie, R., Khan, M.A., Hafiz, S., Collyn, T.A., Frizzel, R., Hawkey, P.M., 2002. Predominant VNTR family of strains of *Mycobacterium tuberculosis* isolated from South Asian patients. International Journal of Tuberculosis and Lung Disease. 6:492-496.

Glynn, J.R., Whitely, J., Bifani, P.J., Kremer, K., and van Soolingen, D., 2002. Worldwide occurrence of Beijing/W Strains of *Mycobacterium tuberculosis*: A systematic review. Emerging Infectious Disease. 8:843-849.

Godreuil, S., Torrea, G., Terru, D., Chevenet, F., Diagbouga, S., Supply, P., van de Perre, P., Carriere, C., and Banuls, A.L., 2007. First Molecule epidemiology study of *Mycobacterium tuberculosis* in Burkina Faso. Journal

of Clinical Microbiology. 45:921-927.

Gordon, S.V., Brosch, R., Billault, A., Garnier, T., Eiglmeier, K., Cole, .ST., 1999. Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. Molecular Microbiology. 32:643-655.

Groenen, P.M., Bunschoten, A.E., van Soolingen, D., van Embden, J.D., 1993. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*: application for strain differentiation by a novel typing method. Molecular Microbiology .10:1057-1065.

Gutacker, M.M., Mathema, B., Soini, H., Shashkina, E., Kreiswirth, B.N., Graviss, E.A., and Musser, J.M., 2006. Single-nucleotide polymorphism-based population genetic analysis of *Mycobacterium tuberculosis* strains from 4 geographic sites. Journal of Infectious Diseases. 193:121-128.

Gutacker, M.M., Smoot, J.C., Migliaccio, C.A., Ricklefs, S.M., Hua, S., Cousins, D.V., Graviss, E.A., Shashkina, E., Kreiswirth, B.N., and Musser, J.M., 2002. Genome-wide analysis of synonymous single nucleotide polymorphisms in *Mycobacterium tuberculosis* complex organisms: resolution of genetic relationships among closely related microbial strains. Genetics. 162:1533-1543.

Gutierrez, M.C., Brisse, S., Brosch, R., Fabre, M., Omaïs, B., Marmiesse, M., Supply, P., and Vincent, V., 2005. Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. Public Library of Science Pathogens. 1:5 doi:10.1371/journal.ppat.0010005.g002.

Hanekom, M., van der Spuy, G.D., Streicher, E., Ndabambi, S.L., McEvoy, C.R., Kidd, M., Beyers, N., Victor, T.C., van Helden, P.D., and Warren, R.M., 2007. A recently evolved sublineage of the *Mycobacterium tuberculosis* Beijing strain family is associated with an increased ability to spread and cause disease. Journal of Clinical Microbiology. 45:1483-1490.

Harries, A.D., 1998. Tuberculosis in HIV-infected persons with special emphasis on Sub-Saharan Africa. Journal of Infection. 37:205-209.

Hawkey, P.M., Smith, E.G., Evans, J.T., Monk, P., Bryan, G., Mohamed, H.H., Bardhan, M., Pugh, R.N., 2003. Mycobacterial Interspersed

Repetitive Unit Typing of Mycobacterium tuberculosis Compared to IS6110-Based Restriction Fragment Length Polymorphism Analysis for Investigation of Apparently Clustered Cases of Tuberculosis. Journal of Clinical Microbiology. 41:3514-3520.

Hermans, P.W., van Soolingen, D., Bik, E.M., de Haas, P.E., Dale, J.W., and van Embden, J.D., 1991. Insertion element IS987 from *Mycobacterium bovis* BCG is located in a hot-spot integration region for insertion elements in Mycobacterium tuberculosis complex strain. Infection and Immunity. 59:2695-2705.

Hilleman, D., Rusch-Gerdes, S., Richter, E., 2007. Evaluation of the Genotype MTBDRplus assay for rifampicin and isoniazid susceptibility testing of *Mycobacterium tuberculosis* strains in clinical specimens. Journal of Clinical Microbiology. 45:2635-2640

Homolka, S., Post, E., Oberhauser, B., George, A.G., Westman, L., Dafae, F., Rüsch-Gerdes, S., and Niemann, S., 2008. High genetic diversity among *Mycobacterium tuberculosis* complex strains from Sierra Leone. BioMedCentral Microbiology 2008, 8:103 doi:10.1186/1471-2180-8-103

http://cgi2.cs.rpi.edu/~bennek/Run.html

http://www.pasteur-guadeloupe.fr:8081/SITVITDemo/index.jsp

http://www.weforum.org/pdf/Initiatives/GHI_TB_Anglogold_AppendixA.pdf

Iseman, M.D., 2002. Tuberculosis therapy: past, present and future. European Respiratory Journal. 20: 87s-94s

Jeffreys, A.J., MacLeod, A., Tamaki, K., Neil, D.L., and Monckton, D.G., 1991. Minisatellite repeats coding as a digital approach to DNA typing. Nature. 354:204-209.

Joint Tuberculosis Committee of the British Thoracic Association, 1982. Notification of tuberculosis: a code of practice for England and Wales. British Medical Journal. 284:1454-1456.

Kallenius, G., Koivula, T., Ghebremichael, S., Hoffner, S.E., Norberg, R., Svensson, E., Dias, F., Marklund, B.I., and Svenson, S.B., 1999. Evolution and clonal traits of *Mycobacterium tuberculosis* complex in Guinea-Bissau.

Journal of Clinical Microbiology. 37:3872-3878.

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

Kapur, V., Whittam, T.S, Musser, and J.M., 1994. Is *Mycobacterium tuberculosis* 15,000 years old? Journal of Infectious Disease. 170:1348-1349.

Kato-Maeda, M., Rhee, J.T., Gingeras, T.R., Salamon, H., Drenkow, J., Smittipat, N., and Small, P.M., 2001. Comparing genomes within the species *Mycobacterium tuberculosis*. Genome Research. 11:547-554.

Kironde, S., and Kahirimbanyi, M., 2002. Community participation in primary health care (PHC) programmes: Lessons from tuberculosis treatment delivery in South Africa. African Health Science. 2:16-23.

Kremer, K., van Soolingen, D.S., Frothingham, R., Haas, W.H., Hermans, P.W. M., Martin, C., Palittapongarnpim, P., Plikaytis, B.B., Riley, L.W., Yakrus, M. A., Musser, J. M., and van Helden, J.D.A. 1999. Comparison of Methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: Interlaboratory study of discriminatory power and reproducibility. Journal of Clinical Microbiology. 37:2607-2618.

Kremer, K., Glynn, J.R., Lillebaek, T., Niemann, S., Kurepina, N.E., Kreiswirth, B.N., Bifani, P.J., and van Soolingen, D., 2004a. Definition of the Beijing/w lineage of *Mycobacterium tuberculosis* on the basis of Genetic markers. Journal of Clinical Microbiology. 42:4040-4049.

Kremer, K., van der Werf, M.J., Au, B.K.Y., Ahn, D.D., Kam, K.M., van Doorn, H.R., and Borgdorff M.W., 2004b. Typical and atypical strains of the *Mycobacterium tuberculosis* Beijing genotype have different levels of association with BCG vaccination and drug resistance. Emerging Infectious Diseases. 15:335-339.

Kremer, K., Au, B.K.Y., Skuce, R., Supply, P., Kam, K.M., and van Soolingen, D., 2005a. Use of Variable-Number Tandem-Repeat Typing To

differentiate *Mycobacterium tuberculosis* Beijing family isolates from Hong Kong and Comparison with IS6110 Restriction Fragment Length Polymorphism typing and Spoligotyping. Journal of Clinical Microbiology. 43:314-320.

Kremer, K., Arnold, C., Cataldi, A., Gutierrez, M.C., Haas, W.H., Panaiotov, S., Skuce, R.A., Supply, P., van der Zanden, A.G.M., and van Soolingen, D., 2005b. Discriminatory Power and Reproducibility of Novel DNA Typing Methods for *Mycobacterium tuberculosis* Complex Strains. Journal of Clinical Microbiology. 43:5628-5638.

Kurepina, N.E., Sreevatsan, S., Plikaytis, B.B., Bifani, P.J., Connell, N.D., Donnelly, R.J., van Sooligen, D. Musser, J.M., and Kreiswirth, B.N., 1998 Characterization of the phylogenetic distribution and chromosomal insertion sites of five IS*6110* elements in *Mycobacterium tuberculosis*: non-random integration in the dnaA-dnaN region. Tubercle Lung Diseases. 79: 31-42.

Kwara, A., Carter, E.J., Rich, J.D., and Flanigan, T.P., 2004. Development of Opportunistic Infections after Diagnosis of Active Tuberculosis in HIV-Infected Patients. AIDS Patient Care and STDs. June:341-347.

Lari, N., Rindi, L., Sola, C., Bonanni, D., Rastogi, N., Tortoli, E., and Garzelli, C., 2005. Genetic diversity, determined on the basis of *katG463* and *gyrA95* polymorphisms, spoligotyping and IS6110 typing of *Mycobacterium tuberculosis* complex isolates from Italy. Journal of Clinical Microbiology. 43:1617-1624.

Lazzarini, L.C.O., Huard, R.C., Boechat, N.L., Gomes, H.M., Oelemann, M.C., Kurepina, N., Shashkina, E., Mello, F.C.Q., Gibson, A.L., Virginio, M.J., Marsico, A.G., Butler, W.R., Kreiswirth, B.N., Suffys, P.N., Lapa e Silva, J.R., and Ho, J.L., 2007. Discovery of a Novel *Mycobacterium tuberculosis* Lineage That Is a Major Cause of Tuberculosis in Rio de Janeiro, Brazil. Journal of Clinical Microbiology. 45:3891-3902.

Le Fleche, P., Fabre, M., Denoeud, F., Koeck, J.L., Vergnaud, G., 2002. High resolution on line identification of strains from the *Mycobacterium tuberculosis* complex based on tandem repeat typing. BioMedCentral Microbiology. 2:37 doi:10.1186/1471-2180-2-37.

Legrand, E., Filliol, I., Sola, C., and Rastogi, N., 2001. Use of spoligotyping

to study the evolution of the direct repeat locus by IS*6110* transposition in *Mycobacterium tuberculosis*. Journal of Clinical Microbiology. 39:1595-1599.

Ling, D.I., Zwerling, A.A., Pai, M., 2008. GenoType MTBDR assays for the diagnosis of multidrug-resistant tuberculosis: a meta-analysis. European Respiratory Journal. 32:1165-1174.

Magdalene, J., Vachee, A., Supply, P., and Locht, C., 1998. Identification of a new DNA region for members of *Mycobacterium tuberculosis* complex. Journal of Clinical Microbiology. 36:937-943.

Marais, B.J., Gie, R.P., Schaaf, H.S., Hesseling, A.C., Obihara, C.C., Starke, J.J., Enarson, D.A., Donald, P.R., and Beyers, N., 2004. The natural history of disease of childhood intra-thoracic tuberculosis–a critical review of the pre-chemotherapy literature. International Journal of Tuberculosis and Lung Diseases. 8:392-402.

Marais, B.J., Victor, T.C., Hesseling, A.C., Barnard, M., Jordaan, A., Brittle, W., Reuter, H., Beyers, N., van Helden, P.D., Warren, R.M., and Schaaf, H. S., 2006. Beijing and Haarlem Genotypes Are Overrepresented among Children with Drug-Resistant Tuberculosis in the Western Cape Province of South Africa, Journal of Clinical Microbiology. 44:3539-3543.

Marais, B. J., Gie, R. P., Schaaf, H. S., Beyers, N., Donald, P. R., and Starke, J., 2006b. Childhood pulmonary tuberculosis–old wisdom and new challenges. American Journal of Respiratory and Critical Care Medicine. 173:1078-1090.

Marais, B.J., Obihara, C.C., Warren, R.W., Schaaf, H.S., Gie, R.P., and Donald, P.R., 2005. The burden of childhood tuberculosis: a public health perspective. International Journal of Tuberculosis and Lung Diseases. 9:1305-1313.

Marmiesse, M., Brodin, P., and Buchrieser, C., 2004. Macro-array and bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex. Microbiology. 150:483-496.

Mazars, E., Lesjean, S., Banuls, A-L., Gilbert, M., Vincent, V., Gicquel, B., Tiayrenc, M., Locht, C., and Supply, P., 2001. High-resolution minisatellite-

based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. Proceedings of the National Academy of Sciences, USA. 98:1901-1906.

McHugh, T.D., Batt, S.L., Shorten, R.J., Gosling, R.D., Uiso, L., and Gillespie, S.H., 2005. *Mycobacterium tuberculosis* lineage: a naming of the parts. Tuberculosis (Edinb). 85:127-136.

Metcalf, C., Coonavia, H.M., and Benetar, S.R., 1991. A history of tuberculosis. A century of tuberculosis: South African perspectives. Oxford, Cape Town.1-31

Metchock, B.G., Noltes, F.S., and Wallace, Jr R.J. Mycobacterium. In: Murray, P.R., Baron, E.J., Pfaller, M.A., Tenover, F.C., Yolken, R.H., eds., 1995. Manual of Clinical Microbiology. 6th ed. American Society for Microbiology, Washington DC:399-437

Miotto, P., Piana, F., Penati, V., Canducci, F., Migliori, G.B., Cirillo, D.M., 2006. Use of genotype MTBDR assay for molecular detection of rifampicin and isoniazid resistance *in Mycobacterium tuberculosis* in clinical strains isolated in Italy. Journal of Clinical Microbiology. 44:2485-2491.

Mokhethi, S.Z., van der Spoel van Dijk, A., and van der Zander, A.G.M., 2005. KatG and rpoB mutations in clinical isolates of *Mycobacterium tuberculosis* with isoniazid and multidrug resistant phenotypes from the Free State and Northern Cape, South Africa. First Joint Congress: Federation of Infectious Diseases Societies of Southern Africa.

Mokrousov, I., Ly, H.M., Otten, T., Lan, N.N., Vyshevskyi, B., Hoffner, S., and Narvskaya, O., 2005. Origin and primary dispersal of the *Mycobacterium tuberculosis* Beijing genotype: clues from human phylogeography. Genome Research. 15:1357-1364.

Mwinga, A., Hosp, M., and Godfrey-Fausset, P., 1998. Twice weekly tuberculosis preventive therapy in HIV infection in Zambia. AIDS. 12:2247-2257.

Namwat, W., Luangsuk, P., Palittapongarnpim, P., 1998.The genetic diversity of *Mycobacterium tuberculosis* strains in Thailand studied by amplification of DNA segments containing direct repetitive sequences. International Journal of Tuberculosis and Lung Diseases. 2:153-159.

Narvskaya, O., Mokrousov, I., Limeschenko, E., Otten, T., Steklova, L., Graschenkova, O., and Vishnevsky, B., 2005. Molecular characterization of *Mycobacterium tuberculosis* strains from the northwest region of Russia. EpiNorth. 1:22-24.

Ndimande, B., Pillay, M., and Sturn A.W., 2005. Spoligotyping of *Mycobacterium tuberculosis* KZN (Kwazulu-Natal). First Joint Congress: Federation of Infectious Diseases Societies of Southern Africa.

Nicol, M.P., Sola, C., February, B., Rastogi, N., Steyn, L., Wilkinson and R.J., 2005. Distribution of Strain Families of *Mycobacterium tuberculosis* Causing Pulmonary and Extrapulmonary Disease in Hospitalized Children in Cape Town, South Africa. Journal of Clinical Microbiology. 43:5779-5781.

Nikolayevskyy, V., Balabanova, Y., Simak, T., Malomanova, N., Fedorin, I., and Drobniewski, F., 2009. Performance of the Genotype®MTBDRPlus assay in the diagnosis of tuberculosis and drug resistance in Samara, Russia Federation. BioMedCentral Clinical Pathology. doi:10.1186/1472-6890-9-2.

Niobe-Eyangoh, S.N., Kuaban, C., Sorlin, P., Cunin, P., Thonnon, J., Sola, C., Rastogi, N., Vincent, V., and Gutierrez, M.C., 2003. Genetic biodiversity of *Mycobacterium tuberculosis* complex strains from patients with pulmonary tuberculosis in Cameroon. Journal of Clinical Microbiology. 41:2547-2553.

Niobe-Eyangoh, S.N., Kuaban, C., Sorlin, P., Thonnon, J., Vincent, V., Gutierrez, M.C., 2004. Molecular Characteristics of Strains of the Cameroon Family, the Major Group of *Mycobacterium tuberculosis* in a Country with a High Prevalence of Tuberculosis. Journal of Clinical Microbiology. 42:5029-5035.

Oelemann, M.C., Diel, R., Vatin, V., Haas, W., Rüsch-Gerdes, S., Locht, C., Niemann, S., and Supply, P., 2007. Assessment of an optimized Mycobacterial interspersed repetitive unit –variable –number tandem –repeat typing system combined with spoligotyping for population –based molecular epidemiology studies of tuberculosis. Journal of Clinical Microbiology. 45:691-697.

Phillip, W.J., Poulet, S., Eiglmeier, K., Pascopella, L., Balasubramian, V., Heym, B., Bergh, S., Bloom, B.R., Jacobs, W.R., and Cole, S.T., 1996**.** An integrated map of the genome of the tubercle bacillus, *Mycobacterium tuberculosis* H37Rv, and comparison with *Mycobacterium leprae*. Proceedings of the National Academy of Science, USA. 93:3132-3137.

Pillay, M., and Sturm, A.W., 2007. Evolution of the Extensively Drug-Resistant F15/LAM4/KZN Strain of *Mycobacterium tuberculosis* in KwaZulu-Natal, South Africa. Clinical Infectious Diseases. 45:1409-1414

Plikaytis, B.B., Marden, J.L., Crawford, J.T., Woodley, C.L., Butler, and W.R., Shinnick, T.M., 1994. Multiplex PCR assay specific for the multidrug-resistant strain W of *Mycobacterium tuberculosis*. Journal of Clinical Microbiology. 32:1542-1546.

Portaels, F., Rigouts, L., and Bastian, I., 1999. Addressing multidrug-resistant tuberculosis in penitentiary hospitals and in the general population of the former Soviet Union. International journal of tuberculosis and lung disease. 3:582-588.

Qian, L., van Embden, J.D.A., van der Zander, A.G.M., Weltevreden, E.F., Duanmu, H., and Douglas, J.T., 1999. Retrospective analysis of the Beijing family of *Mycobacterium tuberculosis* in preserved lung tissues. Journal of Clinical Microbiology. 37:471-474.

Rad, M. E., Bifani, P., Martin, C., Kremer, K., Samper, S., Rauzier, J., Kreiswirth, B., Blazquez, J., Jouan, M., van Soolingen, D., and Gicquel, B., 2003. Mutations in putative mutator genes of *Mycobacterium tuberculosis* strains of the W-Beijing family. Emerging Infectious Disease. 9:838-845.

Rieder, H.L., 2002. Interventions for Tuberculosis Control and Elimination. International Union against Tuberculosis and Lung Disease. 5-169.

Roring, S., Hughes, M.S., Skuce, R.A., and Neill, S.D., 2000. Simultaneous detection and strain differentiation of *Mycobacterium bovis* directly from bovine tissue specimens by spoligotyping. Vetenary Microbiology. 74:227-236.

Roring, S., Scott, A., Brittain, D., Walker, I., Hewinson, G., Neill, S., Skuce, R. 2002. Development of Variable-Number Tandem Repeat Typing of *Mycobacterium bovis*: Comparison of Results with Those Obtained by

Using Existing Exact Tandem Repeats and Spoligotyping. Journal of Clinical Microbiology. 40:2126-2133.

Roring, S., Scott, A.N., Glyn-Hewinson, R., Neill, S.D., and Skuce, R.A., 2004. Evaluation of variable number tandem repeat (VNTR) loci in molecular typing of *Mycobacterium bovis* isolates from Ireland. Vetenary Microbiology. 101:65-73.

Salo, W.L., Aufderheide, A.C., Buikstra, J., and Holcomb, T.A., 1994. Identification of *Mycobacterium tuberculosis* DNA in a pre-Columbian Peruvian mummy. Proceedings of the National Academy of Science, U S A. 91:2091-2094.

Sampson, S.L., Warren, R.M., Richardson, M., van der Spuy, G.D., and van Helden, P.D., 1999. Disruption of coding regions by IS6110 insertion in *Mycobacterium tuberculosis.* Tubercle and Lung Disease. 79:349-359.

Savine, E., Warren, R. M., van der Spuy, G. D., Beyers, N., van Helden, P. D., Locht, C., Supply, P., 2002. Stability of Variable-Number Tandem Repeats of Mycobacterial Interspersed Repetitive Units from 12 Loci in Serial Isolates of *Mycobacterium tuberculosis.* Journal of Clinical Microbiology. 40: 4561-4566.

Schluger, N.W., 1999. Issues in the treatment of active tuberculosis in Human Immmunodeficiency virus- infected patients. Tuberculosis Commentary. 130-135.

Scott, A.N., Menzies, D., Tannenbaum, T-N., Thibert, L., Kozak, R., Joseph, L., Schwartzman, K., and Behr, M.A., 2005. Sensitivities and Specificities of Spoligotyping and Mycobacterial Interspersed Repetitive Unit-Variable-Number Tandem Repeat Typing Methods for Studying Molecular Epidemiology of Tuberculosis. Journal of Clinical Microbiology. 43:89-94.

Sebban, M., Mokrousov, I., Rastogi, N., and Sola, C., 2002. A data-mining approach to spacer oligonucleotide typing of *Mycobacterium tuberculosis*. Bioinformatics. 18:235-243.

Shamputa, I.C., Rigouts, L., and Portaels, F., 2004. Molecular genetic methods for diagnosis and antibiotic resistance detection of mycobacteria from clinical specimens. Acta Pathologica, Microbiologica et Immunologica

Scandinavica. 112:728-752.

Shamputa, I.C., Rigouts,L., Eyongeta, L.A., El Aila, N.A., van Deun, A., Salim, A.H., Willery, E., Locht, C., Supply, P., and Portaels, P., 2004. Genotypic and phenotypic heterogeneity among *Mycobacterium tuberculosis* isolates form pulmonary tuberculosis patients.Journal of Clinical Microbiology. 42:5528-5536.

Skuce, R.A., McCorry, T.P., McCarroll, J.F., Roring, S.M.M., Scott, A. N., Brittain, D., Hughes, S.L., Hewinson, R.G., and Neill, S.D., 2002. Discrimination of *Mycobacterium tuberculosis* complex bacteria using novel VNTR-PCR targets. Microbiology. 148:519-528.

Smith, N.H., 2006. A re-evaluation of M. prototuberculosis. Public Library of Science Pathogens 2:e98. doi: 10.1371/journal.ppat.0020098

Smith, N.H., Gordon, S.V., Rua-Domenech, R., Clifton-Hadley, R.S., and Hewinson, R.G., 2006a. Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis*. Nature Review Microbiology. 4:670-681.

Soini, H., Pan, X., Amin, A., Gravis, W. A., Siddiqui, A., and Musser, J.M., 2000. Characterization of *Mycobacterium tuberculosis* isolates from patients in Houston, Texas, by spoligotyping. Journal of Clinical Microbiology. 38:669-676.

Sola, C., Horgen, L., Goh, K.S., and Rastogi, N., 1997. Molecular fingerprinting of *Mycobacterium tuberculosis* on a Caribbean island with IS*6110* and DRr probes. Journal of Clinical Microbiology. 35:843-846.

Sola, C., Filliol, I., Gutierrez, M.C., Mokrousov, I., Vincent, V., and Rastogi, N., 2001. Spoligotype database of *Mycobacterium tuberculosis*: biogeographic distribution of shared types and epidemiologic and phylogenetic perspectives. Emerging Infectious Diseases. 7:390-396.

Sola, C., Filliol, I., Legrand, E., Lesjean, S., Locht, C., Supply, P., and 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.

Spurgiesz, R.S., Quitugua, T.N., Smith, T.L., Schupp, J., Palmer, E.G.,

Cox, R.A., and Keim, P., 2003. Molecular typing of *Mycobacterium tuberculosis* by using nine novel variable number tandem repeats across Beijing family and low copy numbers IS*6110* isolates. Journal of Clinical Microbiology. 41:4224-4230.

Sreevatsan, S., Pan, X., Stockbauer, K.E., Connell, N.D., Kreiswirth, B.N., Whittam, T.S., and Musser, J.M., 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proceedings of the National Academy of Science, U S A. 94:9869-9874.

Stavrum, R., Mphahlele, M., Ovreas, K., Muthivhi, T., Fourie, P.B., Weyer, K., and Grewal, H.M.S., 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.

Streicher, E.M., Victor, T.C., van der Spuy, G., Sola, C., Rastogi, N., van Helden, P.D., and Warren, R.M., 2007. Spoligotype Signatures in the *Mycobacterium tuberculosis* Complex. Journal of Clinical Microbiology. 45:237–240.

Streicher, E.M., Warren, R.M., Kewley, C., Simpson, J., Rastogi, N., Sola, C., van der Spuy, G.D., van Helden, P.D., and Victor, T.C., 2004. Genotypic and phenotypic characterization of drug-resistant *Mycobacterium tuberculosis* isolates from rural districts of the Western Cape Province of South Africa. Journal of Clinical Microbiology. 42:891-894.

Sun, Y-J., Bellamy, R., Lee, A.S.G., Ng, S.T., Ravindran, S., Wong, S-Y., Locht, C., Supply, P., and Paton, N.I., 2004. Use of Mycobacterial Interspersed Repetitive Unit-Variable-Number Tandem Repeat Typing to Examine Genetic Diversity of *Mycobacterium tuberculosis* in Singapore. Journal of Clinical Microbiology. 42:1986-1993.

Supply, P., Mazars, E., Lesjean, S., Vincent, V., Gicquel, B., and Locht, C., 2000. Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. Molecular Microbiology. 36:762- 771.

Supply, P., Allix, C., Lesjean, S., Cardoso-Oelemann, M., Rüsch-Gerdes, S., Willery, E., Savine, E., de Haas, P., van Deutekom, H., Roring, S.,

Bifani, P., Kurepina, N., Kreiswirth, B., Sola, C., Rastogi, N., Vatin, V., Gutierrez, M.C., Fauville, M., Niemann, S., Skuce, R., Kremer, K., Locht, C., and van Soolingen, D., 2006. Proposal for standardization of optimized Mycobacterial Interspersed Repetitive Unit-Variable Number Tandem Repeat typing of *Mycobacterium tuberculosis*. Journal of Clinical Microbiology. 44:4498-4510.

Supply, P., Himpens, M.J., and Locht, C., 1997. Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. Molecular Microbiology. 26:991-1003.

Supply, P., Lesjean, S., Savine, E., Kremer, K., van Soolingen, D., and Locht, C., 2001. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. Journal of Clinical Microbiology. 39:3563-3571.

Suresh, N., Singh, U.B., Arora, J., Pant, H., Seth, P., Sola, C., Rastogi, N., Samantaray, J.C., and Pande, J.N., 2006. *rpoB* gene sequencing and spoligotyping of multidrug-resistant *Mycobacterium tuberculosis* isolates from India. Infection, Genetics and Evolution. 6:474-483.

Sutherland, G.R., and Richards, R.J., 1994. DNA - a treasury of human variation. New England Journal of Medicine. 331:191-193.

Tautz, D., and Renz, M., 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes.

The South African tuberculosis programme practical guidelines, 2000.

Thierry, D., Cave, M.D., Eisenach, K.D., Crawford, J.T., Bates, J.H., Gicquel, B., and Guesdon, J.L., 1990. IS6110, and IS-element of *Mycobacterium tuberculosis* complex. Nucleic Acids Research. 18: 188

Torrea, G., Levee, G., Grimont, P., Martin, C., Chanteau, S., and Gicquel, B., 1995. Chromosomal DNA fingerprinting analysis using the insertion sequence IS*6110* and the repetitive element DR as strain-specific markers for epidemiological study of tuberculosis in French Polynesia. Journal of Clinical Microbiology. 33:1899-18904.

Traore, H., Eisette, K., Bastian, I., Devleeschouwer, M., and Protaels, F.,

2000. Detection of rifampicin resistance in *Mycobacterium tuberculosis* isolates from diverse countries by a commercial line probe assay as an initial indicator of multidrug resistance. International Journal of Tubercle and Lung Disease. 4: 481-484

Traore, H., van Deun, A., Shamputa, I.C., Rigouts, L., and Portaels, F., 2006. Direct detection of *Mycobacterium tuberculosis* Complex DNA and rifampin resistance in clinical specimen from Tuberculosis patients by line probe assay. Journal of Clinical Microbiology. 44:4384-4388.

Treatment of tuberculosis, 2003. American Thoracic Society, Centre of Disease Control, and Infectious Diseases Society of America .MMWR 52 (RR11): 1-77.

Tsolaki, A.G., Gagneux, S., Pym, A.S., Goguet de la Salmoniere, Y-O.L., Kreiswirth, B.N., Van Soolingen, D., and Small P.M., 2005. Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis*. Journal of Clinical Microbiology.43:3185-3191.

Tsolaki, A.G., Hirsh, A.E., DeRiemer, K., Enciso, J.A., Wong, M.Z., Hannan, M., Goguet de la Salmoniere, Y-O.L., Aman, K., Kaeto-Maeda, M., and Small, P.M., 2004. Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. Proceedings of the National Academy of Science, U S A. 101:4865-4870.

van der Spoel van Dijk, A., Botha, P.L., and Chalkley, L.J., 1996. DNA fingerprint analysis of *Mycobacterium tuberculosis* strains from Bloemfontein: Is Bloemfontein following national trends? Bushpath '96 36th Congress of the Federation of South African Sociaties of Pathology (FSASP). Mpumalanga, South Africa. Abstract MV15.

van der Spoel van Dijk, A., Mokhethi, S.Z., Khumalo, M.J., Shamputa, I.C., Matebesi, S.Z., van Rensburg, H.C.J., Portaels, F., and Rigouts, L., 2005. DNA fingerprinting analyses of *Mycobacterium tuberculosis*-complex isolates from the Free State, South Africa, as part of a multidisciplinary study. *Acta Academica Supplementum.* 1:82-109.

van der Zanden, A.G.M., Kremer, K., Schouls, L.M., Caimi, K., Cataldi, A.,

Hulleman, A., Nagekerke, N.J.D., and van Soolingen, D., 2002. Improvement of differentiation and interpretability of spoligotyping for *Mycobacterium tuberculosis* Complex isolates by introduction of new spacer oligonucleotides. Journal of Clinical Microbiology. 40:4628-4639.

van Embden, J.D., Cave, M.D., Crawford, J.T., Dale, J.W., Eisenach, K.D., Gicquel, B., Hermans, P., Martin, C., McAdam, R., and Shinnick, T.M., 1993. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. Journal of Clinical Microbiology. 31:406-409.

van Embden, J.D., van Gorkom, T., Kremer, K., Jansen, R., Der-Zeijst, B.A., and Schouls, L.M., 2000. Genetic variation and evolutionary origin of the direct repeat locus of the *Mycobacterium tuberculosis* complex bacteria. Journal of Bacteriology. 182:2393-2401.

van Rensburg, D., Janse van Resnburg-Bonthuyzen, E., Heunis, C., Meulemans, H., 2005. Tuberculosis control in South Africa: reasons for persistent failure. Acta academica Supplementum. 1:1-55.

van Rensburg, H.C.J and Mans, A., 1982. Profile of disease and health care in South Africa. Pretoria: Academia.

van Soolingen, D., Qian, L., de Haas, P.E.W., Douglas, J.T., Traoke, H., Portaels, F., Zi Qing, H., Enkhsaikan, D., Nymadawa, P., and van Embden, J.D.A., 1995. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of East Asia. Journal of Clinical Microbiology. 33:3234-3238.

van Soolingen, D., van der Zanden, A.G., de Haas, P.E., Noordhoek, G.T., Kiers, A., Foudrame, N.A., Portaels, F., Kolk, A.H., Kremer, K., and van Embden, J.D., 1998. Diagnosis of *Mycobacterium tuberculosis* infections among humans by using novel genetic markers. Journal of Clinical Microbiology. 36:1840- 1845.

Van Rie, A., Warren, R., Richardson, M., Victor, T.C., Gie, R.P., Enarson, D.A., Beyers, N., and van Helden, P.D., 1999. Exogenous reinfection as a cause of recurrent tuberculosis after curative treatment. New England Journal of Medicine. 341:1174-1179.

Victor, T.C., de Haas, P.E.W., Jordaan, A.M., van der Spuy, G.D.,

Richardson, M., van Soolingen, D., van Helden, P.D., and Warren, R., 2004. Molecular characteristics and global spread of *Mycobacterium tuberculosis* with a Western Cape F11 genotype. Journal of Clinical Microbiology. 42:769-772.

Warren, R., Hauman, J., Beyers, N., Richardson, M., Schaaf, H.S., Donald, P., and van Helden, P., 1996. Unexpectedly high strain diversity of *Mycobacterium tuberculosis* in a high incidence community. Southern African Medical Journal. 86:45-49.

Warren, R.M., Streicher, E.M., Sampson, S.L., van der Spuy, G.D., Richardson, M., Nguyen, D., Behr, M. A., Victor, T.C., and van Helden, P.D., 2002. Microevolution of the direct repeat region of *Mycobacterium tuberculosis*: implications for interpretation of spoligotyping data. Journal of Clinical Microbiology. 40:4457-4465.

Warren, R.M., Victor, T.C., Streicher, E.M., Richardson, M., van der Spuy, M., Johnson, R., Chihota, V.N., Locht, C., Supply, P., and van Helden, P.D., 2004. Clonal expansion of a globally disseminated lineage of *Mycobacterium tuberculosis* with low IS*6110* copy numbers. Journal of Clinical Microbiology. 42:5774-5782.

Watterson, S.A., Wilson, S.M., Yates, M.D., and Drobniewski, F.A., 1998. Comparison of three molecular assays for rapid detection of rifampin resistance in *Mycobacterium tuberculosis*. Journal of Clinical Microbiology. 36:1969-1973.

William-Bouyer, N., Yorke, R., Lee, H.I., and Woods, G.I., 2000. Comparison of the BACTEC MGIT 960 and ESP Culture System II for growth and detection of mycobacteria. Journal of Clinical Microbiology. 38:4167-4170.

Wirth, T., Hildebrand, F., Allix-Beguec, C., Wolbeling, F., Kubica, T., Kremer, K., van Soolingen, D., Ruscj-Gerdes, S., Locht, C., Brisse, S., Meyer, A., Supply, P., and Niemann, S., 2008. Origin, spread and demography of the Mycobacterium Complex. Public Library Of Science. 4:e1000160.doi:10.131371/journal.ppat.1000160.

World Health Organization, report 2008. http://www.who.int/tb/publications/globalreport/en/index.html. Global Tuberculosis Control-Surveillance, Planning, Financing.17-304.

World Health Organization. WHO Tuberculosis Programme framework for effective tuberculosis control. World Health Organization, Geneva, 1994. World Health Organization Document WHO/TB/94.1:-*7.*

www.nationsonline.org/oneworld/south_africa_provinces.htm#Free%20State

www.promega.com/biomath/calc11.htm#mat_results

Zozio, T., Allix, C., Gunal, S., Saribas, Z., Alp, A., Durmaz, R., Fauville-Dufaux, M., Rastogi, N., and Sola, C., 2005. Genotyping of *Mycobacterium tuberculosis* clinical isolates in two cities of Turkey: description of a new family of genotypes that is phylogeographically specific for Asia Minor. BioMedCentral Microbiology. 5:44 doi:10.1186/1471-2180-5-44.

Zumla, A., Maoln, P., Henherson, J., and Grange, J.M., 2000.The impact of HIV infections on tuberculosis. Post Graduate Medical Journal. 76:259-268.

Zumla, A., Mwaba, P., Squire, S.B., Grange, J.M., 1999. The tuberculosis pandemic-which way now. Journal of Infection. 38: 74-79.

Chapter 8

Summary

The aim of the study was to elucidate the diversity and historic origin of *Mycobacterium tuberculosis* (MTB) strains in the Free State by studying the molecular epidemiology of isolates in three high burden areas.

During 2001-2003 MTB isolates were collected during two studies in the FS from consenting patients and DNA from these isolates was further investigated in this study. It involved the use of molecular methods such as spoligotyping and MIRU-VNTR typing that provides unique patterns for the different highly transmissible genotypes (the most prominent ones have been characterized in great detail), permitting historic relationships and the origin of currently circulating strains to be deciphered and to make predictions concerning their spread. Multiplex PCR was used to further characterise the Beijing lineage types and complex software packages including the SpolDB4 and SpolDB3 database (DB), Excell 2007 and MIRU-VNTRplus (to draw phylogenetic trees) aided the analysis of the results.

Strains of the FS were found to be extremely diverse with spoligo analysis resulting in an overall diversity of 73% and MIRU-VNTR analysis a diversity of 50%.

Nine lineages, namely Beijing, Haarlem, LAM, S, T, X, CAS, *M. bovis* and U were represented by the FS isolates. Spoligotyping identified three main families (T, LAM and X) that caused 67.6% of the individual cases of MTB with the rate of spoligopattern diversity within each of these families varying substantially from 76%, 56% and 80% for T, LAM and X respectively. The rate of diversity of the T family is high compared to the LAM family and this could indicate that the T family is especially well adapted to the FS.

The T family was the most prominent, but phylogenetic tree analysis showed that it was distributed over the complete dendrogram with the exception of one closely related T1 (SIT53) cluster, no relationship between them as

expected from reports globally. SIT 53 strains were the most prominent family found amongst the T isolates FS stains.

The demographic distribution of the isolates showed that Beijing, S, T1 and X strains were more prominent in the Motheo district, but no significant difference in the demographic spread of the strains were found in the three areas. Five of 8 Beijing type isolates on further characterisation were found to belong to the W/Beijing subtype, which originated from East Asia and caused several drug resistant outbreaks globally. It is found frequently in the Western Cape province of South Africa.

The LAM family was the second most common lineage with the LAM3/SIT33 being the biggest clonal group of FS strains. This group show little diversity even by MIRU-VNTR code analyses probable due to its ancient history. It might have been brought to South Africa via Europe since the LAM3/F11 is abundant in the Western Cape or through migrant workers from Western Africa countries where it could have been endemic.

The X family was identified as the third most common with 3 clones each consisting of at least two isolates. Minor families which were found to form clones were the S family (4 clones), Family 36 (1 clone) and Beijing (1 clones).

High correlation was observed when FS strains were compared to international reference strains in MIRU-VNTRpuls DB. Some isolates from the LAM, X3 and Family36 grouped together in a cluster of similar strains with an Uganda strain. So do all Beijing isolates with the Beijing strains and also some S family isolates to the S family. LAM9, LAM1 and X1-LAM9 correlated with a LAM strain and X1 isolates correlated with X strains. An EAI5 and a Haarlem isolate correlated to EAI1 strains, while a CAS isolate was found to correlate with Delhi/CAS strains.

UPGMA-Tree analysis (comparing the MIRU-VNTR and spoligotyping patterns) found 6 clones having both identical MIRU-VNTR and spoligo patterns with the most significant ones containing LAM3/SIT33 isolates followed by Beijing, two T1 (SIT53) and two S (SIT34 and SIT71). These clonal groups represent 18.3% of the strains analysed.

Combination of the two typing method resulted in 20 clusters (100% similarity) with the three main families represented by 20% each.

Comparison of all strains to international strains on the MIRU-VNTRplus database shows that FS strains can be linked to other countries such as Uganda, Ghana, Cameroon, China, Delhi counties, Zimbabwe and Tanzania.

The high strain diversity is in agreement with the history of South African TB when it spread from Europe in the 17th century during a massive epidemic as colonies were established in the Northern Cape on route to the FS. More recently the influx of migrant workers to the FS mines from northern areas such as Gauteng and Ghana is seen. The presence of the T family correlates with the history of TB in the country as the presence of the Europeans dates back to the beginning of the 20th century; they are related to the X family. LAM family is found worldwide. Some of our T-family strains are linked to other African strains such as the Cameroon strains and sharing the same spoligo pattern with Tanzanian strains and strains from the Western Cape Province in South Africa. This indicates that the T family is best adapted within these communities.

Nevertheless, it remains hard to determine the detailed history of MTBC strains in the FS province. Given the fact that both the X3 and Beijing strains - dominant in the Western Cape – are less important in the FS, it is more likely that MTB in the FS migrated down from the Gauteng province

and Northern countries including Uganda, Ghana, Zimbabwe and Cameroon where similar lineages are present.

The presence of only few Beijing-family strains within Motheo and Thabo Mofutsanyana districts could be linked to the fact that these two areas comprise factories with a history of Chinese owners.

The diversity and history of MTB in the Western Cape and KZN have been described before with the Beijing lineage as one of the most prominent. Until 2009, in the FS province, IS*6110* RFLP typing data was the only information available for a representative sample of *M. tuberculosis* strains. In this study, the presence of 3 dominant lineages: the ill defined T, LAM and X families is in concordance with a recent study by Stavrum *et al* that was the first to report typing data from 8 provinces in SA. They reported that the T lineage was the most prominent amongst 25 FS isolates and in the Gauteng province. The LAM and X lineages occurred in all the provinces with the LAM being more prominent in the Eastern Cape and KZN, while the highest frequencies of the X lineage was seen in the Western Cape and Northern province.

This study served as a pilot as it contains isolates collected in the mid 2001. There remains a need to type current strains as there has been an influx of both reports about high legal and illegal immigrants into the country plus new global economic partnerships allowing people free movement into and out of the country.

Chapter 9

Opsomming

Die doel van die studie was om lig te werp op die diversiteit en geskiedenis van *Mycobacterium tuberculosis* (MTB) stamme in die Vrystaat (VS) deur 'n studie van die molekulêre epidemiologie van isolate in drie hoë insidensie gebiede.

Gedurende 2001-2003 is MTB isolate versamel tydens twee studies in die VS van pasiënte wat hul toestemming gegee het en DNS van hierdie isolate is verder ondersoek in hierdie studie. Dit het die gebruik van molekulêre metodes soos spoligotipering en MIRU-VNTR tipering behels. Hierdie tegnieke verskaf unieke patrone vir verskillende hoogs oordraagbare genotipes (waarvan die prominentste al in diepte bestudeer is) en maak dit moontlik om geskiedkundige verwantskappe en die oorsprong van huidig sirkulerende stamme te ontrafel en om voorspellings te maak betreffende die verspreiding van verwante groepe.

Multipleks PKR is gebruik om die Beijing familie tipe isolate verder te karakteriseer terwyl komplekse sagteware pakkette insluitend die SpolDB4 en SpolDB3 databasisse (DB), Excell 2007 en die MIRU-VNTRplus (om filogenetiese stambome te trek) gebruik is om resultate te analiseer.

Dit is bevind dat die stamme van die VS buitegewoon divers is deur die gebruik van spoligo analise met 'n diversiteit van 73% en met MIRU-VNTR analise 'n diversiteit van 50%.

Die VS isolate verteenwoordig nege familielyne, naamlik Beijing, Haarlem, LAM, S, T, X, CAS, *M. bovis* en U. Spoligotipering het drie hoof families in die VS geïdentifiseer (T, LAM en X) wat gesamentlik 67.6% van die individuele gevalle van MTB verteenwoordig. Die diversiteitssyfer vir spoligotipering binne elkeen van hierdie families is baie uiteenlopend, naamlik 76%, 56% en 80% onderskeidelik vir die T, LAM en X families. Die

diversiteitssyfer van die T familie was hoog in vergelykig met die LAM familie en dit kon moonlik aandui dat die T familie veral aangepas is in die VS.

Die T familie was die mees prominente groep, maar filogenetiese stamboom analise het getoon dat hierdie familie oor die hele dendrogram verspreid was met geen verwantskap tussen hulle nie, behalwe vir die nabyverwante T1 (SIT53) groep. Dit stem ooreen met wat wêreldwyd gerapporteer word. SIT 53 stamme was die prominentste groep onder die T isolate van die VS stamme.

Die demografiese verspreiding van die isolate het aangetoon dat die Beijing, S, T1 en X stamme meer prominent in die Motheo distrik was, maar geen noemenswaardige verskille in die demografiese verspreiding van die stamme is in die drie gebiede gevind nie. Met verdere karakterisering is gevind dat 5 van die 8 Beijing familie isolate aan die W/Beijing subtipe behoort. W/Beijing subtipe het verskeie middelweerstandige uitbrake wêreldwyd veroorsaak en word gevind in groot getalle in die WesKaap provinsie van Suid Afrika en is oorspronklik van Oos Asië afkomstig.

Die LAM familie is die tweede algemeenste familielyn met die LAM3/SIT33 die grootste klonale groep van VS stamme. Hierdie groep toon min verskille selfs met MIRU-VNTR kode analise en dit is waarskynlik weens die ou geskiedenis. Hierdie stamme kon moontlik na Suid-Afrika versprei het via Europa, aangesien die LAM/F11 in groot getalle in die WesKaap teenwoordig is, of deur migrasie van werkers vanaf Wes-Afrika lande waar dit moontlik endemies kon wees.

Die X familie is die derde algemeenste familie wat ïdentifiseer is, met 3 klone wat elkeen ten minste twee isolate elk bevat het. Klein families wat klone gevorm het was die S familie (4 klone), Familie 36 (1 kloon) en Beijing (1 kloon).

‘n Hoë korrelasie is opgemerk toe VS stamme vergelyk is met internasionale verwysingsstamme in die MIRU-VNTRplus DB. Sommige isolate van die LAM, X3 en Familie 36 het saamgegroepeer in ‘n groep van soortgelyke stamme met ‘n Uganda stam. Alle Beijing isolate het saam met Beijing stamme gegroepeer asook sommige S familie isolate met ander S familie isolate. LAM9, LAM1 en X1-LAM9 het gekorreleer met ‘n LAM stam en X1 isolate met X stamme. ‘n EAI5 en ‘n Haarlem isolaat het gekorreleer met EAI1 stamme terwyl ‘n CAS isolaat gevind is wat gekorreleer het met Delhi/CAS stamme.

UPGMA-stamboom analise (vergelyking van die MIRU-VNTR en spoligotiperingspatrone) het 6 klone geïdentifiseer wat beide identiese MIRU-VNTR klone en spoligopattrone het met die mees betekenisvolle klone bevattende LAM3/SIT33 isolate gevolg deur Beijing, twee T1 (SIT53) en twee S (SIT34 en SIT71) isolate. Hierdie klonale groepe verteenwoordig 18.3% van die stamme wat analiseer is.

Kombinasie van die twee tiperingsmetodes het 20 groepe (100% ooreenstemming) getoon met die drie hoof families verteenwoordig deur 20% elk.

Vergelyking van alle stamme met internasionale stamme in die MIRU-VNTRplus databasis toon dat die VS stamme gekoppel kan word aan stamme van ander lande soos Uganda, Ghana, Cameroon, China, Delhi lande, Zimbabwe en Tanzania.

Die hoë stam diversiteit is in ooreenstemming met die geskiedenis van Suid-Afrikaanse TB deurdat dit versprei het van Europa in die 17th eeu gedurende ’n massiewe epidemie waarna kolonies gevorm is in die Noord-Kaap onderweg na die VS. Die invloei van trekarbeiders na die VS myne vanaf ander gebiede verder noord soos Gauteng en Ghana het plaasgevind. Die

teenwoordigheid van die T familie korreleer verder met die geskiedenis van TB in die land aangesien die teenwoordigheid van die Europeërs terugdateer na die begin van die 20ste eeu; hulle is verbind met die X familie. Die LAM familie word ook wêreldwyd gevind. Sommige van ons T-familie stamme is verbind met ander Afrika stamme soos die Cameroon stamme. Hulle deel dieselfde spoligopattroon met Tanzanise stamme en stamme van die Wes-Kaap provinsie van Suid Afrika. Hierdie dui waarskynlik aan dat die T familie die beste aangepas is binne hierdie gemeenskappe.

Daarbenewens bly dit steeds moeilik om die volledige geskiedenis van MTBC stamme in die VS provinsie te bepaal. As die feit in ag geneem word dat beide die X3 en Beijing stamme – prominent in die Wes-Kaap – minder belangrik is in die VS, dan is dit meer waarskynlik dat MTB in die VS migreer het vanuit die noordelike Gauteng provinsie en Noordelike lande insluitend Uganda, Ghana, Zimbabwe en Cameroon waar soortgelyke familielyne teenwoordig is.

Die teenwoordigheid van slegs elkele Beijing-familie stamme in die Motheo en Thabo Mofutsanyana distrikte kan gekoppel word aan die feit dat hierdie twee gebiede bestaan uit fabrieke met 'n geskiedenis van Chinese eienaars.

Die diversiteit en geskiedenis van MTB in die Wes-Kaap en KZN is voorheen beskryf met die Beijing familie as een van die prominentste familielyne. Tot en met 2009 was IS*6110* RFLP tipering die enigste data beskikbaar vir 'n verteenwoordigende monster van *M. tuberculosis* stamme in die VS provinsie. In hierdie studie is die bevinding van 3 dominante familielyne, die swak gedefinieerde T, LAM en X families in ooreenstemming met 'n onlangse studie van Stavrum *et al* wat die eerste was om tiperings data van 8 provinsies in SA te publiseer. Hulle raporteer dat die T-familielyn

die prominentste is tussen die 25 VS isolate asook in die Gauteng provinsie. Die LAM en X familielyn is in al die provinsies gevind, die LAM was die prominentste in die Oos Kaap en KZN; terwyl die hoogste voorkoms van die X familielyn in die Weskaap en Noordelike provinsie gevind is.

Die studie het gedien as ‘n lootsstudie, want dit bevat isolate wat al in 2001 versamel is. Daar is nogsteeds ‘n groot behoefte om huidige stamme te tipeer omdat daar regmatige en onregmatige immigrante in die land is as gevolg van veelvuldige internasionale ekonomiese venootskappe wat veroorsaak dat mense deesdae vrylike toegang in en uit die land het.