# HOST INDUCED MICROEVOLUTION OF ESX SECRETION SYSTEMS OF *M. TUBERCULOSIS*

By

## MELISHA SUKKHU

Submitted for complete fulfilment of the academic requirements for the degree of Doctor of Philosophy in Medical Microbiology (PhDMD)

Supervisor: Dr. Alexander S. Pym

**15 November 2013**

*College of Health Sciences, School of Laboratory Medicine & Medical Sciences, Doris Duke Medical Research Institute, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa*

*"An experiment is a question which science poses to Nature and a measurement is the recording of Nature's answer."*
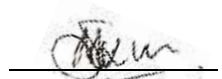
*Max Planck (1858-1947)*

# PLAGIARISM DECLARATION

I, **Melisha Sukkhu**, declare that

(i)  The research proposed in this thesis, except where otherwise indicated, is my original work.

(ii)  This thesis has not been submitted for any degree or examination at any other university.

(iii) This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

(iv) This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other persons. Where other written sources have been quoted, then:

> a) their words have been re-written but the general information attributed to them has been referenced;
>
> b) where their exact words have been used, their writing has been referenced and acknowledged.

(v)  Where I have reproduced a publication of which I am an author, co-author or editor, I have indicated in detail which part of the publication was actually written by myself alone and have fully referenced such publications.

(vi) This thesis does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.
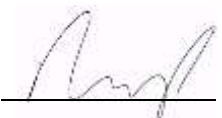
Signed: _____          Date:  **15 November 2013**
**Melisha Sukkhu**

Signed: _____          Date:  **15 November 2013**
**Dr Alexander S. Pym (Supervisor)**

# DEDICATION

*In loving memory of my late grandfather, Mr. Ramchandar Sukkhu (Snr).*

*Dhada, I have always looked toward you for strength, protection and guidance throughout my studies. You are my first God whom I have always turned to when I felt lost. Even though you are not here physically, I know in my heart you have heard all my conversations with you during my daily commute from Pietermaritzburg to Durban during this PhD. Thank you for keeping me safe and always being with me in some form or another. I hope I have made you proud!*

*Love you and miss you everyday.*

# PRESENTATIONS

**Sukkhu, M.**

*esx* Gene Expression across *BCG* sub-strains.

NOVSEC-TB Consortium (Dublin, Ireland), $15^{th}$-$16^{th}$ May 2010. (Oral Presentation).

**Sukkhu, M.**

Expression of *esx* genes in *Mtb* Clinical and Laboratory Isolates.

NOVSEC-TB Consortium (Rome, Italy), $23^{rd}$-$24^{th}$ May 2011. (Oral Presentation).

**Sukkhu, M., Pym, A. S.**

Variation in sequence and expression profiles of ESX genes in *Mtb* Clinical and Laboratory isolates.

$4^{th}$ FIDSSA Congress (Durban, South Africa), $8^{th}$-$11^{th}$ September 2011. (Oral Presentation).

**Sukkhu, M.**

Variation in sequence and expression profiles of ESX genes in *Mtb* Clinical and Laboratory isolates.

UKZN College of Health Sciences Research Symposium (Durban, South Africa), $14^{th}$ September 2011. (Poster Presentation).

**Sukkhu, M.**

Variation in sequence and expression profiles of ESX genes in *Mtb* Clinical and Laboratory isolates.

Tuberculosis 2012 Conference (Institute Pasteur, Paris, France), $15^{th}$-$20^{th}$ September 2012. (Poster Presentation).

**Sukkhu, M.**

Variation in sequence and expression profiles of ESX genes in *Mtb* Clinical and Laboratory isolates.

$5^{th}$ FIDSSA Congress (Drakensberg, South Africa), $10^{th}$-$12^{th}$ October 2013. (Oral Presentation).

# LIST OF PUBLICATIONS

**Sukkhu, M. and Pym, A. S.** A Review of ESX secretion systems within *M. tuberculosis. Manuscript in preparation.*

**Sukkhu, M. and Pym, A. S.** ELISPOT responses in healthy and TB positive donors for QILSS and MTB9.9 peptides. *Manuscript in preparation.*

# ACKNOWLEDGMENTS

# ABSTRACT

The ESX family of genes (*esxA-W*) in *Mycobacterium tuberculosis (Mtb)* encodes 23 effector molecules influencing immunogenicity and pathogenicity. This study was aimed at identifying and evaluating variations in ESX sequence and protein expression profiles in clinical isolates and examining how diversity might influence immune responses. 23 ESX genes from 55 clinical isolates (20 Beijing, 25 KZN and 10 Other) and 3 Laboratory strains (H37Rv, H37Ra and BCG) were sequenced. 482 single nucleotide polymorphisms (SNPs) were identified in 12 ESX genes relative to H37Rv. Majority of the identified 363 nsSNPs occured in Beijing isolates. No mutations were observed in *esxA, B, C, E, G, H, J, R, S* and *T*. Six unique nsSNPs were identified in the Beijing isolates: *esxI* (Q20L), *esxO* (E52G), 2 in *esxP* (T3S; N83D), *esxU* (P63S) and *esxW* (T2A). Three unique nsSNPs were identified in the KZN isolates: *esxK* (A58T), *esxL* (R33S). The *esxL* polymorphism resulted from a dinucleotide change.

ESX gene transcription levels were evaluated using RT-qPCR. Varying expression levels were observed for *esxA, B, C, F, M* and *Q* across all clinical isolates with lowest levels seen amongst the Beijing isolates. This correlated with immunoblots with confirmed decreased *esxAB* protein expression relative to the other strains. The Matrix-Assisted Laser Desorption Ionization Time of Flight (MALDI-TOF) spectral protein profiles were quantitatively compared within and between *Mtb* clinical and laboratory isolates. Protein spectral profiles within the mass range of the CFP-10 protein with variations in peak intensities were observed across all isolates.

QILSS and Mtb9.9 peptides were tested individually for immune responses in TB infected patients. Healthy patients displayed no responses to QILSS and Mtb9.9, strong but variable immune responses were detected for specific regions of QILSS and Mtb9.9 in TB infected patients. These findings demonstrate that differences in sequence, transcriptional profiles and protein expression patterns in ESX secreted proteins exist between clinical isolates, and may translate into differences in human immune responses. Further research is needed to correlate human host immune responses to the phenotype and genotype of the infecting strain of *Mtb* to determine the consequences of specific variations of the other ESX members. These studies are important for the development of improved immune diagnostics and vaccines.

# ETHICS

The research described in this thesis was approved by the Faculty of Medicine Biomedical Research Ethics Committee (BREC REF: BCA27409) of the University of KwaZulu-Natal.

# TABLE OF CONTENTS

## CHAPTER TWO:

## Evaluation of ESX Sequence Variations within *M. tuberculosis* Clinical

## CHAPTER THREE:

## CHAPTER FOUR:

## ESX Protein and Expression Profiles for *M. tuberculosis* Clinical

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| *C. diphtheria* | *Corynebacterium diphtheria* |
| *E. coli* | *Escherichia coli* |
| *M. africanum* | *Mycobacterium africanum* |
| *M. bovis* | *Mycobacterium bovis* |
| *M. canettii* | *Mycobacterium canettii* |
| *M. marinum (Mma)* | *Mycobacterium marinum* |
| *M. microti* | *Mycobacterium microti* |
| *M. smegmatis* | *Mycobacterium smegmatis* |
| *Mtb* | *Mycobacterium tuberculosis* |
| *M. vaccae* | *Mycobacterium vaccae* |

| | |
|---|---|
| ADC | Albumin Dextrose Catalase |
| Asn | Asparagine |
| ATPase | Adenosine triphophatase |
| ATP | Adenosine triphosphate |
| BCA | Bicinchoninic Acid |
| BCG | Bacillus Calmette-Guérin |
| BCIP | 5-Bromo-4-Chloro-3-Indolyphosphate Ptoluidine salt |
| BLAST | Basic Local Alignment Search Tool |
| BREC | Biomedical Research Ethics Committee |
| BSA | Bovine Serum Albumin |

| | |
|---|---|
| BSL3 | Biosafety Level 3 |
| C- | Carboxy- |
| CD4 | Cluster of Differentiation 4 (Glycoprotein) |
| cDNA | Complementary DNA |
| CFP-10 | Culture Filtrate Protein 10 |
| $C_T$ | Cycle Threshold |
| CTAB | *N*-cetyl-*N,N,N,*-trimethyl-ammonium bromide |
| CUBS | Collection of Urine Blood and Sputum |
| DEPC | Diethylpyrocarbonate |
| DMSO | Dimethyl Sulfoxide |
| DNA | Deoxyribose Nucleic Acid |
| DOH | Department of Health |
| DPBS | Dulbecco's Phosphate Buffered Saline |
| DTT | Dithiothreitol |
| ESAT-6 | Early Secretory Antigenic Target of 6kDa |
| *ecc* | *esx* Conserved Component |
| *esp* | *esx-1* Secreted-associated Protein |
| *esx*/ESX | ESAT-6 Secretion System |
| ExoI | Exonuclease I |
| EDTA | Ethylenediaminetetraacetic acid |
| ELISA | Enzyme-Linked Immunosorbent Assay |
| ELISPOT | Enzyme-Linked Immunosorbent Spot |

| | |
|---|---|
| FRET | Fluorescence Resonance Energy Transfer |
| Fur | Ferric Uptake Regulator |
| GC | Guanine and Cytosine |
| Gly | Glycine |
| HIV | Human Immunodeficiency Viurs |
| HLA | Human Leukocyte Antigen |
| $H_2O$ | Water |
| IFN-$\gamma$ | Interferon-gamma |
| IGRA | Interferon Gamma Release Assay |
| IL-1$\beta$ | Interleukin-1 beta |
| IS1 | Ion Source Voltage 1 |
| kDa | Kilo Dalton |
| K-RITH | KwaZulu-Natal Research Institute for Tuberculosis and HIV/AIDS |
| KZN | KwaZulu-Natal |
| LAM | Lipoarabinomannan |
| MAb | Monoclonal Antibody |
| MALDI-TOF | Matrix-Assisted Laser Desorption/Ionisation-Time of Flight |
| MDR | Multiple Drug Resistant |
| MIGIT | Mycobacteria Growth Indicator Tube |
| MIQE | Minimum Informaion of Quantitative Real-Time PCR Experiments |
| MPTR | Major Polymorphic Tandem Repeat |

| | |
|---|---|
| mRNA | Messenger RNA |
| MS | Mass Spectrometry |
| MTC | *Mycobacterium tuberculosis* complex |
| *mycP* | Mycosin Protease |
| N- | Amino- |
| NaCl | Sodium Chloride |
| NaOAc | Sodium Acetate |
| NBT | Nitro-Blue Tetrazolium |
| NC | Negative Control |
| nsSNP | Non-synonymous SNP |
| OD | Optical Density |
| PBMC | Peripheral Blood Mononuclear Cell |
| PBS | Phosphate Buffer Saline |
| PCR | Polymerase Chain Reaction |
| PE | Proline-glutamic acid |
| PGRS | Polymorphic GC-Rich Repetitive Sequence |
| PHA | Phytohemagglutinin |
| PIE | Pulsed Ion Extraction |
| PPE | Proline-proline-glutamic acid |
| PVDF | Polyvinylidene Difluoride |
| RD | Region of Difference |
| RFLP | Restriction Fragment Length Polymorphism |
| RNA | Ribonucleic Acid |

| | |
|---|---|
| RQI | RNA Quality Indicator |
| RT-qPCR | Real-Time Quantitative PCR |
| SDS-PAGE | Sodium Dodecyl Sulfate Polyacrylamide |
| SANBS | South African National Blood Bank |
| SCID | Severe Combine Immune-Deficient |
| SFC | Spot Forming Counts |
| SNP | Single Nucleotide Polymorphisms |
| sRNA | Small RNA |
| sSNP | Synonymous SNP |
| TB | Tuberculosis |
| T-cell | T lymphocytes |
| TNF-$\alpha$ | Tumor Necrosis Factor-alpha |
| TraSH | Transposon Site Hybridisation |
| T7SS | Type VII Secretion System |
| U | Unit |
| UKZN | University of KwaZulu-Natal |
| WXG | Tryptophan-X-Glycine |
| WHO | Wolrd Health Organization |
| X | variable amino acid |
| x g | Times Gravity |
| Zur | Zinc Uptake Regulator |
| $\alpha$ | Alpha |
| $\beta$ | Beta |
| $\gamma$ | Gamma |

| | |
|---|---|
| μl | Micro Litre |
| °C | Degrees Celcius |
| 2-D | 2-Dimensional |
| 3-D | 3- Dimensional |
| A | Amps |
| aa | Amino Acid |
| bp | Base Pair |
| Da | Dalton |
| dN | number of nonsynonymous mutations per nonsynonymous site |
| dS | number of synonymous mutations per synonymous site |
| F | Forward |
| Hz | Hertz |
| kV | Kilo Volt |
| M | Molar |
| mA | Milliamps |
| ml | Milli litre |
| mM | Milli Molar |
| MW | Molecular Weight |
| ng | Nanogram |
| nm | Nano Metre |
| ns | Nano Second |
| pH | Potential of Hydrogen |
| pM | Pico Molar |

| | |
|---|---|
| R | Reverse |
| U | Unit |
| V | Volts |
| v/v | Volume per Volume |
| w/v | Weight per Volume |
| x | Times |
| 7H9 | Middlebrook Broth Medium |
| 7H11 | Middlebrook Agar Medium |

# GENERAL INTRODUCTION

Globally, tuberculosis (TB), the disease caused by the *Mycobacterium tuberculosis* (*Mtb*) remains a major health problem, with an estimated 8.7 million incident cases of TB (13% co-infected with HIV) reported in 2011 (WHO, 2012). According to the 2012 WHO Global Tuberculosis report, the majority of the estimated TB cases occurred in Asia (59%) and Africa (29%). Of the 22 high burden countries, South Africa is the 3rd country with regard to the largest number of reported incident cases (0.4 million-0.6 million).

It is believed that the *Mycobacterium* genus dates back to more than 150 million years and the current members of the *Mycobacterium tuberculosis* complex (MTC) originated from a common progenitor approximately 15,000-35,000 years ago (Gutierrez *et. al.*, 2005; Daniel, 2006).

Compatible with a relatively recent emergence, the MTC is a tightly-knit slow-growing complex, comprising of *M. tuberculosis* (causative agent of human tuberculosis), *M. bovis* (infects a variety of mammals and humans), *M. africanum* (causes human tuberculosis in sub-Saharan Africa), *M. canettii* (smooth variant) and *M. microti* (causative agent of tuberculosis in voles). Based on the DNA homology analyses, it was revealed that all the members share a significant level of genetic similarity at the nucleotide level (99.9%) and identical 16SrRNA, but phenotypic and host preferences differ (Sreevatsan *et. al.*, 1997; Brosch *et. al.*, 2000; Gutacker *et. al.*, 2002; Huard *et. al.*, 2006). In 1997, Sreevatsan and colleagues reported that this complex lacks inter-strain genetic diversity and nucleotide changes are infrequent. Based on sequence analyses of 26 loci, it was concluded that the *Mtb* genome is unusually inert and in evolutionary terms, this organism is fairly young (Sreevatsan *et. al.*, 1997). To gain additional insight into the possible factors influencing the MTC diversity, several research groups implemented phylogenetic approaches in the analysis of clinical *Mtb* isolates from diverse geographical locations. The main aim was to associate unique Single Nucleotide Polymorphisms (SNPs) as genetic markers in different lineages. Those studies revealed definitive evidence for a clonal

population structure for this complex as well as a lack of on going horizontal gene transfer (Filliol *et. al*., 2006; Gagneux *et. al.,* 2006; Gutacker *et. al*., 2006; Hershberg *et. al.,* 2008; Comas *et. al.,* 2010; Uplekar *et. al.,* 2011). The chronic nature of this disease has also been influenced by selective pressures, namely, the host immune response, changes in human demography, HIV co-infection and the usage of anti-TB drugs (Malik and Godfrey-Fausset*,* 2005; Gagneux *et. al.,* 2006; Nicol and Wilkinson*,* 2008; Kato-Maeda *et. al.,* 2001; Parwati *et. al.,* 2010; Comas and Gagneux, 2011; Brites and Gagneux, 2012).

The G+C rich genome sequence of *Mtb* H37Rv, has stimulated much focus on the development of vaccines to curb the incidence and prevalence of the disease. Within this genome is the ESX gene family comprising of five gene clusters (*esx-1* to *esx-5*) which are part of the Type VII protein secretion system (T7SS) in *Mtb*. Likewise, similar secretion systems have been identified in other *Mtb* species and in a few Gram positive bacteria (Tekaia *et. al*., 1999; Gey van Pittius *et. al*., 2001; Pallen, 2002; Finn *et. al*., 2006). Cole and colleagues speculated possible immunological functions for the ESX gene family (Cole *et. al.,* 1998). This proposal propelled research in the direction of antigenic variation and a closer examination of this gene family and associated proteins. The ESAT-6 antigen was recognised to be of importance, due to its absence in all BCG sub-strains (Maheiras *et. al.,* 1996). This is the genomic region that has been implicated in the attenuation of all vaccine strains as the region of difference (RD1), which encodes *esx-1* (Sørensen *et. al.,* 1995) is absent from all BCG strains. As a result, this has created interest in exploring the use of this antigen for potential downstream vaccine and diagnostic applications (Andersen, *et. al.,* 2000; Doherty, *et. al.,* 2002; Pym *et. al.,* 2003; Brodin *et. al.,* 2004).

The availability of the sequenced mycobacterial genome has led to a better understanding into the biology and genomics of the organism. Since then, public databases of several mycobacterial sequences have been established and it is now possible to compare genes and establish whether the genes are conserved to specific species. They have also been successful in identifying *Mtb* outbreaks, where polymorphic genetic markers have been utilized in the discrimination and subtyping of *Mtb* strains (Brosch, *et. al.,* 2001). Two independent studies were published recently (during the work of this thesis), assessing the genetic diversity of the ESX

family members (Uplekar *et. al.,* 2011) and their potential for antigenic variation (Comas *et. al.,* 2010). Uplekar and colleagues showed that some of the identified mutations did affect known ESX epitopes. Both studies also noted a high genetic variability in the Mtb9.9 and QILSS subfamilies. It therefore became imperative and vital that further studies are executed in identifying and characterizing these proteins and their secretion systems to identify variability from a wider selection of strains.

Despite extensive research efforts and availability of anti-tuberculosis drugs, there is very little understanding of the virulence mechanisms of the organism. An additional contributing factor to the prevalence of this disease is the lack of an effective vaccine. The only vaccine currently in use is the Bacille Calmette-Guérin (BCG) vaccine that was developed in 1921. This vaccine is partially effective as it provides protection against the severe forms of pediatric TB (TB meningitis and military TB) but has variable efficacy in preventing pulmonary TB in adults. Despite the extensive usage of this vaccine, the global disease burden is still on the rise. For this reason, there is an emphasis on developing new and more effective vaccine designs are being developed and aimed at the identification and selection of immunodominant antigens of *Mtb* that are capable of inducing protective immune responses. The final outcome is to improve and transform the current therapy and diagnosis of this disease (WHO, 2013). According to the 2012 WHO Gobal Tuberculosis Report, 12 TB vaccine candidates were submitted into clinical trial programs, but presently there are two adjuvanted protein subunit vaccines (Aeras 402 and MVA-85A) in Phase II trails. Both these candidates are virus vectored vaccines that express one or more TB antigens. Aeras 402 is based on the adenovirus 35 virus vector and expresses a few of the TB antigens. Whereas MVA-85A is based on a modified vaccinia Ankara vector which expresses the TB antigen 85A. However, recent efficacy data from an MVA-85A clinical trial was not encouraging, and hope of the availability of a new vaccine in 2018 is optimistic (WHO, 2012; 2013).

In 2009, Bitter and colleagues proposed a general systematic gene nomenclature for the T7SS gene cluster components (Bitter *et. al*., 2009), however, the old terminology will be used throughout this thesis. The old terminology together with the proposed terminology is represented in Figure 1.1 and Table 1.1.

# SCOPE AND AIMS OF THESIS

The hypothesis to be explored in this study is that the ESX systems, through their wide range of effector molecules, that influence immunogenicity and pathogenicity, could also be important in generating phenotypic diversity in clinical isolates of *Mtb*. Characterizing such diversity also has important implications for diagnostics and vaccine design. The primary interest lies in establishing whether differences in sequence, transcriptional profiles and protein expression patterns related to ESX secreted proteins exist between different clinical isolates.

This study is aimed at investigating the ESX system of *Mtb* to:

(i)     Evaluate the sequence diversity by reference to the completed genome sequences and comparison with other publicly available *Mtb* sequences.

(ii)    Analyze and identify potential genetic variations within and between the ESX systems in clinical isolates using amplicon sequencing.

(iii)   Determine, using Real-Time qPCR, whether differences exist in ESX transcription.

(iv)    Compare quantitatively, using ESAT-6 and CFP-10 antibodies, the secretion profiles within and between clinical isolates.

(v)     Utilize the MALDI-TOF proteomics approach to compare ESX protein spectra within and between clinical isolates.

(vi)    Design and incorporate peptides in an ELISPOT assay to determine whether an immune response is elucidated from regions of diversity.

This thesis is divided into six chapters. The first chapter comprises the literature review. Chapter two is the evaluation of ESX sequence variation for the clinical and laboratory isolates used in this study. In chapter three, the transcriptomics approach in the analysis of ESX gene expression is explored. Chapter four is the proteomics approach to assess the ESX protein profiles for the clinical and laboratory isolates used in this study. Included in chapter five are the ELISPOT responses in healthy and TB positive donors to peptides designed from the QILSS and Mtb9.9 subfamilies. Chapter six presents the conclusions and thoughts on future research.

# CHAPTER ONE



# Literature Review

## 1.1    ESX Protein Family and Secretion System

Within the *Mtb* genome, is the ESX family of genes encoding 23 proteins of approximately 100 amino acids (aa). Twenty-two of the genes occur in couples, and one gene (*esxQ)* occurs in isolation. The 23 genes are labelled as *esxA* to *esxW* with *esxA* and *esxB* encoding two proteins ESAT-6, and CFP-10 that were the first members of the family to be identified. Five of the gene couples occur in larger homologous gene clusters or regions named ESX-1 (Rv3866-Rv3883c), ESX-2 (Rv3884c-Rv3895c), ESX-3 (Rv0282-Rv0292), ESX-4 (Rv3444c-Rv3450c) and ESX-5 (Rv1782-1798) (Table 1.2)*.* Phylogenetic analyses indicated that they were probably duplicated from the ancestral region, Region 4, in the order 3, 1, 2 and then 5 and the remaining ESX genes arose from duplication events and the integration of additional genes (Figure 1.1) (Gey van Pittius *et. al.,* 2001).
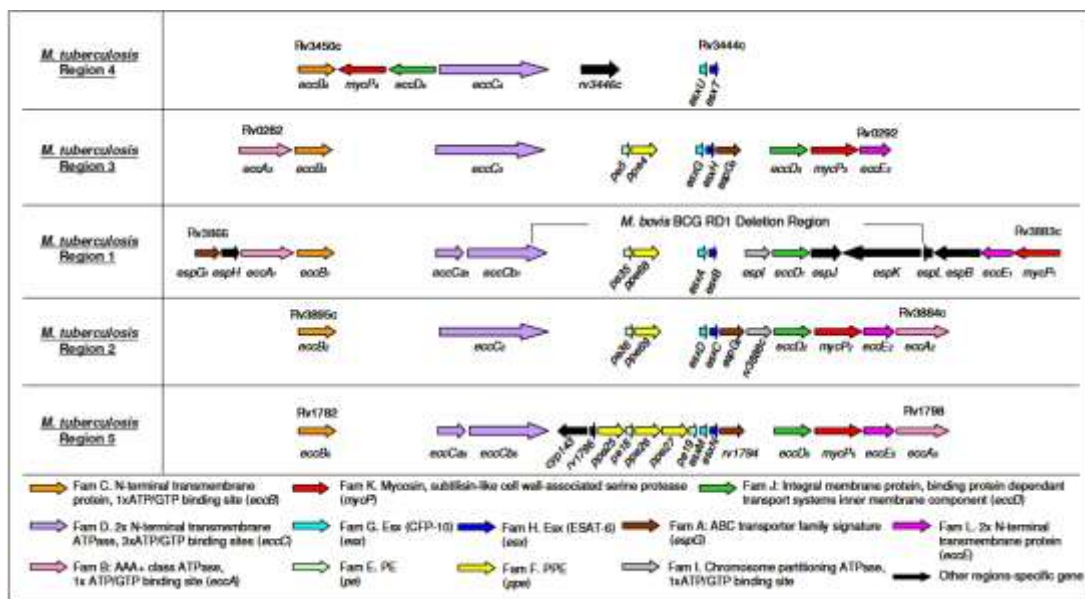


**Figure 1.1**    The ESAT-6 gene clusters of *Mtb* evolved in the order 4, 3, 1, 2 and then 5. Genes of the ESAT-6 gene cluster region 4 are maintained through all the duplications and the PE/PPE genes, incorporated into Region 3, in the subsequent duplications (Gey van Pittius *et. al.,* 2001).

Due to the high sequence identity between members, the ESX family can be classified into subfamilies, as represented in Table 1.2 (adapted from Uplekar *et. al.,* 2011). The TB10.4 subfamily includes *esxH* (from the ESX-3 cluster), while the Mtb9.9 and QILSS subfamilies contain 5 members together with *esxM* and *esxN* (from the ESX-5 cluster) (Tekaia *et. al.*, 1999; Louise *et. al.*, 2001; Skjøt *et. al.,* 2002).

### 1.1.1   ESX Genetic Locus

The five ESX gene clusters are comprised of two ESX homologs (Family G and H), a transmembrane ATPase (Family D), a transmembrane ATP-binding protein, a subtilisin-like membrane-anchored cell wall associated serine protease (mycosin) and a putative integral membrane pore protein (Family J) (Figure 1.2) (Tekaia *et. al.*, 1999; Gey van Pittius *et. al.*, 2001; Abdallah *et. al.*, 2007). Furthermore, ESX-1, 2, 3 and 5 include genes encoding the PE and PPE proteins. These regions encode elements for the T7SS secretion apparatus (Brodin *et. al.,* 2004, Fortune *et. al.*, 2005; MacGurn *et. al.,* 2005; Abdallah *et. al.*, 2007). The positions and presence of family A to L genes in 5 of the ESAT-6 gene clusters are specified in Table 1.1 together with the encoded proteins. It has been proposed that the T7SS machinery is comprised of the ESX proteins (Family G and H) of each ESAT-6 gene cluster forming a complex that interacts with the FtsK/SpoIIE transmembrane ATPase (Family D) providing the energy protein translocation through the membrane protein pore (Family J) (Abdallah *et. al.*, 2007). Recently, the purpose of some of the remaining ESAT-6 gene cluster components where identified for the first time. Using computational analyses involving phylogenetic profiles, transmembrane helices, 3-D folds, signal peptides and protein-protein associated prediction, Das *et. al.* (2011) assigned possible functions to additional ESX-1 components (Rv3866, Rv3876, Rv3869 and Rv3882c). They proposed that Rv3866 is involved in the transcriptional activation of ESAT-6, while Rv3876 was predicted to be the negative regulator for this machinery. In addition, Rv3869 and Rv3882c where proposed to be the components required for the formation of the mycomembrane translocation apparatus (Das *et. al.,* 2011). It is yet to be established exactly how the ESX proteins cross the mycobacterial outer membrane.
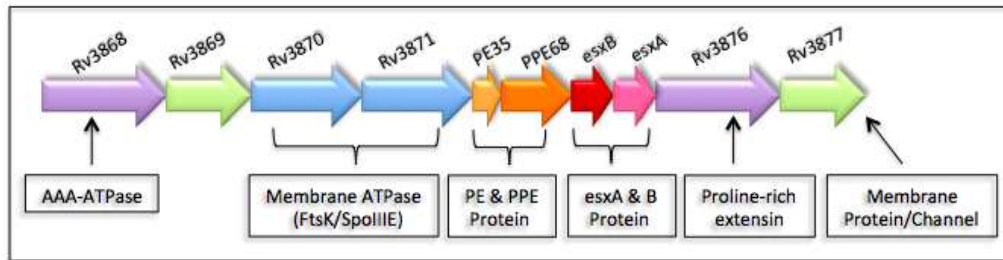
**Figure 1.2** Diagrammatic representation of the genetic neighbourhood for the ESX-1 system (Young, 2003).

**Table 1.1**      Components of the 5 ESAT-6 gene clusters within *Mtb*.

| Gene family | Description | Presence of genes in the ESAT-6 gene cluster regions | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| A | ABC transporter family signature | Rv3866 (*espG₁*) | Rv3889c (*espG₂*) | Rv0289 (*espG₃*) | | Rv1794 |
| B | AAA+ class ATPases, CBXX/CFQX family, SpoVK, 1x ATP/GTP-binding site | Rv3868 (*eccA₁*) | Rv3884c (*eccA₂*) | Rv0282 (*eccA₃*) | | Rv1798 (*eccA₅*) |
| C | Amino terminal transmembrane protein, possible ATP/GTP-binding motif | Rv3869 (*eccB₁*) | Rv3895c (*eccB₂*) | Rv0283 (*eccB₃*) | Rv3450c (*eccB₄*) | Rv1782 (*eccB₅*) |
| D | DNA segregation ATPase, ftsK chromosome partitioning protein, SpoIIIE, YukA, 3x ATP/GTP-binding sites, 2x amino-terminal transmembrane protein | Rv3870 (*eccCa₁*) -Rv3871 (*eccCb₁*) | Rv3894c (*eccC₂*) | Rv0284 (*eccC₃*) | Rv3447c (*eccC₄*) | Rv1783 (*eccCa₅*) -Rv1784 (*eccCb₅*) |
| E | PE | Rv3872 (*pe35*) | Rv3893c (*pe36*) | Rv0285 (*pe5*) | | Rv1788 (*pe18*) Rv1791 (*pe19*) |
| F | PPE | Rv3873 (*ppe68*) | Rv3892c (*ppe69*) | Rv0286 (*ppe4*) | | Rv1787 (*ppe25*) Rv1789 (*ppe26*) Rv1790 (*ppe27*) |
| G | CFP-10, Esx family protein | Rv3874 (*esxB*) | Rv3891c (*esxD*) | Rv0287 (*esxG*) | Rv3445c (*esxU*) | Rv1792 (*esxM*) |
| H | ESAT-6, Esx family protein | Rv3875 (*esxA*) | Rv3890c (*esxC*) | Rv0288 (*esxH*) | Rv3444c (*esxT*) | Rv1793 (*esxN*) |
| I | ATPases involved in chromosome partitioning, 1x ATP/GTP-binding motif | Rv3876 (*espI*) | Rv3888c | | | |
| J | Integral inner membrane protein, binding-protein-dependent transport systems inner membrane component signature, putative transporter protein | Rv3877 (*eccD₁*) | Rv3887c (*eccD₂*) | Rv0290 (*eccD₃*) | Rv3448 (*eccD₄*) | Rv1795 (*eccD₅*) |
| K | Mycosin, subtilisin-like cell wall-associated serine protease | Rv3883c (*mycP₁*) | Rv3886c (*mycP₂*) | Rv0291 (*mycP₃*) | Rv3449 (*mycP₄*) | Rv1796 (*mycP₅*) |
| L | 2x amino-terminal transmembrane protein | Rv3882c (*eccE₁*) | Rv3885c (*eccE₂*) | Rv0292 (*eccE₃*) | | Rv1797 (*eccE₅*) |

Systematic genetic nomenclature for T7SS within brackets.      ecc: ESX conserved component; esp: ESX-1 secretion-associated protein; mycP: mycosin protease (Bitter *et. al*., 2009)

**Table 1.2**     Summary of ESX genes for each cluster and subfamily (Adapted from Uplekar *et. al.,* 2011).

| ESX Locus | EsxA Paralog | | EsxB Paralog | |
|---|---|---|---|---|
| | Within ESX Locus | Outside ESX Locus | Within ESX Locus | Outside ESX Locus |
| ESX-1 | *esxA* (Rv3875) | | *esxB* (Rv3874) | |
| ESX-2 | *esxC* (Rv3890c) | | *esxD* (Rv3819c) | |
| ESX-3 | *esxH* (Rv0288) | *esxR* (Rv3019c), *esxQ* (Rv3017c) | *esxG* (Rv0287) | *esxS* (Rv3020c) |
| ESX-4 | *esxT* (Rv3444c) | | *esxU* (Rv3445c) | |
| ESX-5 | *esxN* (Rv1793) | *esxI* (Rv1037c), *esxL* (Rv1198), *esxO* (Rv2346c), *esxV* (Rv3619c) | *esxM* (Rv1792) | *esxJ* (Rv1038c), *esxK* (Rv1197), *esxP* (Rv2347c), *esxW* (Rv3620c) |
| None | | *esxE* (Rv3904c) | | *esxF* (Rv3905c) |

■ TB10.4 subfamily;    ■ Mtb9.9 subfamily;    ■ QILSS subfamily

10

### 1.1.2 ESX-1 System

The ESX-1 system is restricted to the mycobacterium genus, and is located near the origin of replication (Gey Van Pittius *et. al.,* 2001). This system is responsible for the secretion of the 6 kDa early secreted antigenic target (ESAT-6) (Rv3875) and the 10 kDa culture filtrate protein (CFP-10) (Rv3874). These two small proteins contain a WXG amino acid motif that forms the hairpin bend in the helix-turn-helix structure (Figure 1.3) (Cole, *et. al.*, 1998; Pallen *et. al.*, 2002; Renshaw *et. al.*, 2002).

Both the genes are co-transcribed, as a consequence of being located directly adjacent to each other, and are secreted from the cell as a heterodimer despite the absence of known secretion signals (Sørensen *et. al.*, 1995; van Pinxteren *et. al.*, 2000; Pym *et. al.*, 2003). In a study published in 2010 by Callahan and colleagues, they reported that the WXG proteins from *Mtb* (*esxG* and *esxH*), *M. smegmatis* (*esxA* and *esxB*) and *C. diphtheria* (*esxA* and *esxB*) are heterodimers that fold into the α–helix structure. By fusing *esxA* and *esxB* into a single polypeptide (MtbEsxAB), they created a biomimetic bait, which interacted with several ESX-1 encoded proteins, as well as ESX proteins encoded by ESX-2. Therefore, their study suggested that WXG proteins are similar in structure regardless of the host species (Callahan *et. al.*, 2010).

Additional studies have shown that several ESX pairs form tight complexes including *esxR* and *esxS* (Rv3019c and Rv3020c) and *esxO* and *esxP* (Rv2346c and Rv2347c) (Renshaw *et. al.,* 2002; Meher *et. al.,* 2006; Lightbody *et. al.*, 2008; Arbing *et. al.*, 2010).

Despite being the most characterized system, seven additional proteins that are part of this system, have been also reported to be secreted by ESX-1. This includes EspA (Rv3616c), EspB (Rv3881c), EspC (Rv3615c), EspE (Rv3864), EspF (Rv3865), PE35 (Rv3872), and EspR (Rv3849) (Fortune *et. al.,* 2005; Abdallah *et. al.,* 2007; McLaughlin *et. al.,* 2007; Xu *et. al.,* 2007; Raghavan *et. al.,* 2008; Bitter *et. al.,* 2009).
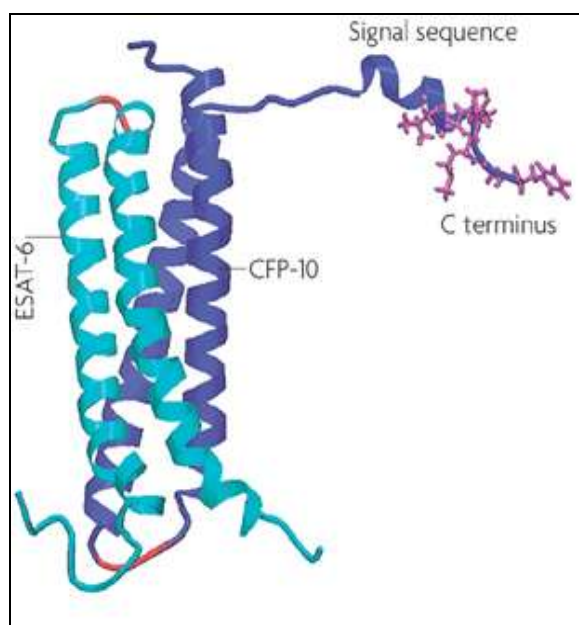
**Figure 1.3** ESAT-6 and CFP-10 interaction results in a helix-turn-helix motif and an unstructured C-terminal region (Abdallah *et. al.*, 2007).

### 1.1.2.1 ESX-1 Regulation and Virulence

As described in the introduction, a genomic deletion results in loss of the ESX-1 system from *M. bovis* BCG (Harboe *et. al.,* 1996; Behr *et. al.*, 1999; Gordon *et. al.,* 1999; Colangeli *et. al.,* 2000; Wards *et. al.,* 2000). However, the secretion of the ESAT-6 and CFP-10 antigens in the context of a recombinant BCG vaccine in the background of a reconstituted RD1 has been shown to restore attenutation and have a beneficial effect on protection against challenge with *Mtb* (Pym *et. al.,* 2003). ESAT-6 expression has been identified in the early stage of infection (Andersen *et. al.*, 2000) and is crucial for the survival and spread of the bacteria *in vivo* (Chapman *et. al.*, 2002). Subsequent *in silico* investigations and genetic screens have indicated that the genes surrounding the *esxAB* operon are vital in the secretion of ESAT-6 and CFP-10. This evidence was provided in the disruption of the *Rv3870, Rv3871* and *Rv3877* genes that prevented the secretion of ESAT-6 and CFP-10 (Gey van Pittius *et. al.*, 2001; Pallen *et. al.*, 2002; Hsu *et. al.*, 2003; Sassetti *et. al.*, 2003a; Sassetti *et. al.*, 2003b; Stanely *et. al.*, 2003; Guinn *et. al.*, 2004).

The ESX-1 system is controlled by multiple regulators including the DNA binding transcription factor EspR (Rv3849), the PhoP two component regulator system and the serine protease MycP1 (Frigui *et. al.*, 2008; Gonzalo-Asensio *et. al.,* 2008; Raghavan *et. al.*, 2008; Ohol *et. al.,* 2010).

An earlier study of ESX-1 virulence and regulation in mice by Raghavan and colleagues showed that EspR (Rv3849) is a key regulator of this ESX system. Their study revealed that EspR is responsible in activating *espACD* (Rv3616c-Rv3614c)*,* components of ESX-1, and these expressed components promoted the secretion of ESX-1 substrates upon macrophage infection. EspR also binds and activates the Rv3616c-Rv3614c promoter (Raghavan *et. al.,* 2008). Not only were the authors able to prove that EspR ia a key regulator of *Mtb* virulence, but it is controlled by a unique mechanism, whereby it activates ESX-1 secretion after phagocytosis, but during macrophage infection, its activity is titrated away from the cell, thus lowering its concentration within the cell. By using ChIP-Seq experiments (chromatin immunoprecipitation followed by ultra-high throughput DNA sequencing), Blasco and colleagues recently showed that EspR binds to 165 loci on the *Mtb* mostly associated with cell wall function in the genome and its expression varied with bacterial growth. This implied that EspR acts as a nucleoid-associated protein that impacts cell wall functioning and pathogenesis through multiple genes. This suggests a more global regulatory function rather than specific control of ESX-1 (Blasco *et. al.,* 2012).

The study conducted by Ohol and colleagues in 2010, uncovered the role of the MycP1 protease (a component of the ESX-1 secretion system) in the regulation and virulence of the ESX-1 system. Using homologous recombination, a MycP1 mutant was created in *Mtb*, and the culture supernatant for both the wild type and mutant strains were probed with ESAT-6 antibodies in southern blot experiments. These experiments confirmed that MycP1 is required for ESX-1 secretion as ESAT-6 was detected in the wild type strain but detection failed in the mutant strains. In macrophage infection experiments, the investigators reported on a hyperactivation of the ESX-1 stimulated innate signalling pathways in an inactive MycP1 protease mutant. However, in a mouse model, MycP1 was proven to be essential for growth during the acute stage of infection, while the loss of protease activity lead to

attenuated virulence (Ohol *et. al.,* 2010). Thus, the dual role of MycP1 is vital in the regulation and virulence of the ESX-1 system.

As of 2012, a novel proteomics approach to evaluate and monitor the ESX-1 protein secretion system was successfully developed by Champion and colleagues (Champion *et. al.,* 2012). They were able to directly detect on agar, ESX-1 protein secretion and the presence and absence of ESAT-6 and CFP-10 from the surface of wild-type and mutant *M. marinum* colonies using whole colony matrix-assisted laser desorption/ionisation-time of flight (MALDI-TOF) mass spectrometry. The authors had generated a library of *M. marinum* mutant strains containing independent transposon insertions. Using the MALDI-TOF method, they were able to analyse their library to successfully identify mutant strains that failed in secreting ESAT-6 and CFP-10. In addition, they also reported on a transposon insertion in *eccB1,* which is known to be essential in ESAT-6 and CFP-10 secretion in *M. smegmatis* and *M. microti* complemented models (Converse and Cox, 2005; Brodin *et. al.,* 2006). This is the first report of a proteomics confirmation demonstrating that these two genes are an essential requirement for an intact ESX-1 system in pathogenic bacteria.

Thus *esxA* and *esxB* are very important proteins of *Mtb* involved in host-pathogen interaction and in addition to the induction of strong T-cell mediated immune responses (Berthet *et. al*., 1998).

## 1.1.2.2 ESX Virulence in Cell Models

Several possibilities exist for the role that the ESX-1 system plays in pathogenesis. Stanely and colleagues showed that ESX-1 is responsible for an interaction with the cytokine interferon beta resulting in the induction of interferon responsive genes (Stanely *et. al*., 2007). On the other hand, de Jonge and colleagues noted a particular interest in the role of ESAT-6 and CFP-10 in biomembrane interaction once the bacterium was engulfed by phagocytic cells. They reported that under acidic conditions, ESAT-6 dissociates from CFP-10 and exhibits membrane-lysing activity (de Jonge *et. al.,* 2007). This finding suggested that the ESX-1 system does in fact play a role in phagosomal escape.

Using an ingenious single cell fluorescence resonance energy transfer (FRET) based method that detected the presence of cytoplasmic *Mtb*, Simeone and colleagues where able to show that THP-1 macrophages infected with wild or mutant *Mtb* strains provoked a change in the FRET signal after 3 to 4 days of infection. No cytoplasmic *Mtb* was detected with mutant strains not processing the ESX-1 *esx* factors. In addition, the full complement of the ESX-1 system restores the ability of these mutants in causing phagosomal ruptures. They also observed in infected macrophages, for both *Mtb* and *M. marinum,* phagosomal rupture followed by necrotic cell death (Simeone *et. al.,* 2012).

The ESX-1 secretion system was credited for the activation of the DNA-dependant cytosolic surveillance pathway within macrophages by the exposure of extracellular mycobacterial DNA to the cytosol of the macrophages (Manzanillo *et. al.,* 2012; Watson, Manzanillo and Cox, 2012). The ESX-1 system was reported as the vehicle that allowed for the extracellular *Mtb* DNA access to cytoplasmic DNA receptors through limited permeabilization of the phagosomal membrane. This response is generated by the IFN production during macrophage infection, where the IFI204 DNA receptor stimulates the STING/TBK1/IRF3 signaling axis (Manzanillo *et. al.,* 2012). In a later study by Watson, Manzanillo and Cox (2012), the authors identified critical steps in the pathway within macrophages used in *Mtb* recognition. Upon activation of the STING/TBK1/IRF3 signaling axis, a population of the engulfed bacteria are surrounded and marked by host ubiquitin chains resulting in the delivery to lysosomes. This selective autophagy is a mechanism of host control employed during *in vivo Mtb* infection (Watson, Manzanillo and Cox, 2012). This finding has aided in the speculation that the activation of the cytosolic surveillance pathway is a survival strategy used by *Mtb* to remain persistent within the host (Novikov *et. al.,* 2011). In addition, recent studies in *Salmonella* have also shown that the bacterium initiates an inflammatory immune response that facilitates the persistence within the host (Winter *et. al.,* 2010; Arpaia *et. al.,* 2011).

### 1.1.2.3 ESX Virulence in Animal Models

In 2010, Carlsson and colleagues investigated the role of the ESX-1 mediated inflammasome activation in *M. marinum* infection of mice tails. Tails of the mice infected with ESX-1 deficient bacteria (wild-type) showed that ESX-1 increased the generation of proinflammatory IL-1β cytokines *in vivo* (Carlsson *et. al.,* 2010). Their findings implicate a distinct role for ESX-1 in disease promotion and inflammation due to its capacity to activate the inflammasome. On the other hand, Hsu and colleagues proposed that ESAT-6 has a toxic role as it lyses the cellular membranes (Hsu *et. al.*, 2003). This finding was consistent with a study conducted by Junqueira-Kipnis and collagues, whereby they showed that the RD1 is essential for tissue necrosis in the lungs of infected mice (Junqueira-Kipnis *et. al.*, 2006).

The *M. marinum* zebrafish model has been a useful model in the study of *Mtb* infections. Besides being a close genetic relative of *Mtb*, in ectotherms, it causes a similar tuberculosis-like disease with the typical granuloma formation. The zebrafish is a natural host for *M. marinum* within which the pathogen produces granulomas that bear a resemblance to human granulomas (Ramakrishnan *et. al.*, 1997; Talaat *et. al.*, 1998; Davis *et. al.*, 2002; Prouty *et. al.*, 2003; van der Sar *et. al.*, 2003; Broussard and Ennis, 2007). An advantage of the zebrafish model, is the ease of investigating the innate immune response during microbial infection as the embryos initially lack an adaptive immune system (Trede *et. al.*, 2004). A consequence of the transparent zebrafish embryos is that it is an easier vessel to monitor the early stages of infection in real-time (Davis *et. al.*, 2002). Hence, the zebrafish model is a good model in studying and understanding *Mtb* pathogenesis. Evidence of pore formation from the ESX-1 system was shown in experiments in *M. marinum*, where a genetic deletion of the ESAT-6 gene eliminated pore formation (Smith *et. al.,* 2008).

In addition to the ESX-1 system, for both *Mtb* and *M. marinum,* being a major virulent determinant, studies have found that the substrates secreted by this system play a role in granuloma formation (Stoop *et. al.*, 2011). This probably occurs through intracellular bacterial spread between macrophages and interference in the phagosomal membrane integrity, leading to "leaky membranes" and bacterial entrance into the cytosol (Stanely *et. al.,* 2003; Lewis *et. al.,* 2003; Gao *et. al.*, 2004; Guinn *et.*

*al.,* 2004; Swaim *et. al.,* 2006; Volkman *et. al.,* 2004; Stanely *et. al.,* 2007; Brodin *et. al.*, 2010; Smith *et. al.,* 2008; Manzanillo *et. al.,* 2012; Simeone *et. al.,* 2012; Watson, Manzanillo and Cox, 2012).

By exploiting the zebrafish embryo model in *M. marimun*, Stoop and colleagues developed an *in vivo* screen to identify genes involved in granuloma formation. They reported 3 mutants from an initial screen of 200 mutants that initiated granuloma formation (Stoop *et. al.*, 2011). Of the 3 reported mutants, *espL*, was the ESX-1 related mutant that effected ESAT-6 secretion and was unable to induce granuloma formation. Their findings were also in agreement with Brodin and colleagues whom demonstrated that phagosomal maturation arrest in macrophages was affected in a *Mtb espL* mutant (Brodin *et. al.*, 2010). The findings from this study show that *espL* is an important component in the ESX-1 system, playing a vital role in virulence and granuloma formation.

### 1.1.3 ESX-2 System

This system is located directly adjacent to the ESX-1 system and bears similarities to ESX-1 structurally (Gey van Pittius *et. al.,* 2001). The *esxC* (*Rv3890c*) and *esxD* (Rv3891c) genes are found in this system, however, their functional role is still unknown (Sassetti *et. al.*, 2003a; Sassetti *et. al.*, 2003b). Due to a partial deletion of the membrane-bound component, MMAR_5460, it is believed that the ESX-2 system may be defective in *M. marinum* (Gey van Pittius *et. al*., 2001; Abdallah *et. al.,* 2006). To date, this secretion system has not been investigated and therefore its function remains unknown.

### 1.1.4 ESX-3 System

The ESX-3 system is structurally similiar to the ESX-1 system, but encodes *esxG* and *esxH*. For *esxH*, expression studies by Rindi and colleagues in *Mtb* H37Ra, revealed it to be down-regulated in this strain (Rindi, Lari and Garzelli*,* 1999; Rindi, Lari and Garzelli, 2001). A later study in 2002, by Skjot and colleagues, reported that the *esxH* protein was a potent T-cell antigen that was strongly recognized in *Mtb* infected patients (Skjot *et. al., 2002)*. In 2010, Davila and colleagues conducted a study investigating the genetic diversity of *esxA, esxH* and *fbpB* genes amongst clinical isolates and their possible implications in future TB subunit vaccines. 88 strains were included in the investigation, but no nucleotide polymorphisms were reported for *esxA* and *esxH* (Davila *et. al.,* 2010). Apart from the above citations, studies on *esxG* and *esxH* have been limited.

With insertional mutagenesis and transposon site hybridisation (TraSH) analyses, the ESX-3 system was shown to be essential for *in vitro* growth of *Mtb* (Sassetti *et. al.,* 2003a; Sassetti *et. al.,* 2003b) but not for pathogenicity *in vivo* (Siegrist *et. al.,* 2009). Maciag and colleagues later revealed ESX-3 expression is controlled by the zinc uptake regulator Zur/FurB (Maciag *et. al.,* 2007). The role of ESX-3 in iron and zinc homeostasis was confirmed with the construction of ESX-3 conditional mutants, where the transcription could be down-regulated (Serafini *et. al.*, 2009). It was shown that the wild-type was vital for *Mtb* survival, however, the mutant phenotype can be seen in high iron and zinc concentrations. As a consequence, this lead to the suggestion that ESX-3 expression is regulated by diavalent cation levels (Serafini *et.*

*al.*, 2009). In 2009, Maciag and colleagues reported that ESX-3 is regulated by iron concentration and not by zinc concentration in *M. smegmatis*. This difference in regulation could be due to the fact that despite the presence of irons limiting in the soil and human host environments, *M. smegmatis* in its natural environment is unlikely to experience zinc deficiency (Maciag *et. al.,* 2009).

Further investigations of the *esxGH* crystal structure by Ilghari and colleagues in 2011 revealed that this complex contained a specific $Zn^{2+}$ binding site from a cluster of histidine residues on *esxH*. Thus concluding that perhaps this site might be instrumental in the scavenging of zinc ions or the regulation of $Zn^{2+}$ (Ilghari *et. al.*, 2011). The authors also reported that despite the similarities to the *esxAB* folds of the protein structure, *esxGH* does infact differ from it. They noted that the helices of *esxGH* are much shorter than the *esxAB* complex and the C-terminal region of *esxH* has the tendency to adapt a helical confirmation whereas in *esxAB*, the C-terminal of *esxB* is responsible for this. The final difference reported was that a cleft exists on the *esxGH* structure, hinting at the existence of a functional binding site for $Zn^{2+}$ that was noted to be absent in the *esxAB* complex (Ilghari *et. al.*, 2011).

Orthologous systems of ESX-3 are found in all mycobacterial species, whose genomes were analysed so far and ESX-3 is the most conserved, which is compatible with the essential character of the system that appears to play an elementary role in the mycobacterial lifecycle. In preliminary work it was recently suggested that ESX-3 prevents innate immune killing of mycobacteria (Sassetti *et. al.,* 2003a; Sassetti *et. al.*, 2003b). Hence, ESX-3 also represents an interesting potential new drug target that needs to be explored more comprehensively.

### 1.1.5 ESX-4 System

Phylogentic investigations have implied that the ESX-4 system is the progenitor system, based on its presence in other actinobacteria. The remaining systems arose with gene duplications, proceeded by divergence events as well as the recruitment and accumulation of other genes (Tekaia *et. al.,* 1999; Gey van Pittius *et. al*., 2001). ESX-4 is the smallest gene cluster, and does not encode the PE or PPE genes. In addition, the ESX-4 system has loci that are well conserved within corynebacteria (*Corynebacterium diptheriae, Corynebacteruim efficiens* and *Corynebacterium glutamicum*) (Cerdeño-Tárraga *et. al.,* 2003) (Figure 1.4), *Nocardia farcinica, Gordonia bronchialis* and various Rhodococcus species. These are high GC Gram-positive bacteria, and suggest that the ESX-4 secretion system may be conserved amongst these organisms (Gey van Pittius *et. al*., 2001). In other low GC Gram-positive bacteria, *esx*-like proteins have been identified, as well as in *Bacillus anthracis* and *Staphylococcus aureus* (Pallen, 2002; Burts *et. al*., 2008; Garufi *et. al*., 2008). Whole genome mutagenesis studies have not predicted a requirement of the ESX-4 system for virulence or *in vitro* growth (Sassetti *et. al.*, 2003a; Sassetti *et. al.*, 2003b).
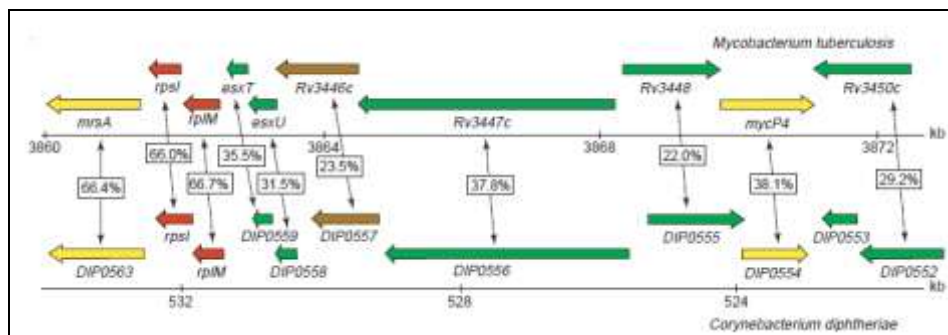


**Figure 1.4**     The ESX-4 system gene map of *Mtb* and *C. diptheriae* showing the conserved genes and pair-wise percentage amino acid identity (Cerdeño-Tárraga *et. al.,* 2003; http://www.sanger.ac.uk/Projects/C_diphtheriae).

### 1.1.6 ESX-5 System

This is a poorly understood group of proteins due to the lack of reported studies in *Mtb*. The ESX-5 locus encodes the *esxM* and *esxN* proteins that induce strong CD4+ T-cell responses in human and animal models (Alderson *et. al.,* 2000; Jones *et. al.,* 2010b). Extensive studies have been conducted with *Mycobacterium marinum (Mma)* where the ESX-5 system was shown to be involved in the secretion of proteins from the PE/PPE family and the PPE-MPTR and PE-PGRS subgroups (Abdallah *et. al.,* 2008; 2009; 2011; Cascioferro *et. al.,* 2011; Bottai *et. al.,* 2012). This system is the more recent duplication and the region occurs only in slow growing mycobacteria (Gey van Pittius *et. al.,* 2001; Abdallah *et. al.,* 2006).

There are an additional gene pairs (*esxKL, esxIJ, esxOP* and *esxVW*) that belong to the ESX-5 system and encode a variant QILSS and Mtb9.9 motif, which are required for normal microbial growth, and likely to have a role in bacterial multiplication during active infection. In 2011, Bukka and colleagues investigated factors that may contribute to the survival of active *Mtb* during infection in the lung (Bukka *et. al.,* 2011). By analyzing the RNA expression patterns in sputum, they reported that *esxIJ* and *esxKL* transcripts were differentially expressed under varying growth conditions. They further reported that mutants for both these genes altered the growth of laboratory broth cultures (Bukka *et. al.,* 2011).

Abdallah and colleagues reported the initial link to this system and PPE protein secretion in 2006. Their study showed that a transposon mutant of ESX-5 in *M. marinum* was incapable of secreting the PPE41 protein (Rv2430c) from *Mtb*. The authors of this study complemented *M. smegmatis* with various portions of the ESX-5 cluster of *M. marinum* and were able to confirm the involvement of ESX-5 in PPE secretion (Abdallah *et. al.,* 2006). Subsequent studies with ESX-5, *M. marinum* mutants showed the mutants were unable to modulate macrophage cytokine responses and were not involved in the translocation to the cytosol (Abdallah *et. al.,* 2008; 2009; 2011). It has been recently shown that ESX-5 secreted substrates are involved in IL-1β conversion and cell death in *Mtb*, but in *M. marinum*, the ESX-5 mutant does not induce IL-1β activation and inflammasome activities. This system was also shown to induce caspase independent cell death following translocation. With the aid of small interfering RNA experiments, Abdallah and colleagues demonstrated that cathepsin B

is also involved in inflammasome activation and cell death. With microarray and real time PCR analysis, the roles of PE and PPE in *Mtb* virulence were explored. The analyses revealed that the deletion of the Rv0485 gene can decrease PE13 and PPE18 expressions that lead to the attenuation of *Mtb* virulence and in infected murine macrophages the secretion of pro-inflammatory cytokines was reduced (Goldstone *et. al.,* 2009).

In an attempt to characterize the role of ESX-5 in *Mtb*, Bottai and colleagues constructed 5 knockout/deletion ESX-5 mutants in *Mtb* H37Rv, namely *eccA5, eccD5,* Rv1794*, esxM* genes and PPE25-PE19. The authors reported no phenotypic change with the Rv1794 knockout, however, the remaining four mutants displayed defective secretions of *esxN* and PPE41. Upon evaluation of the virulence of the *eccD5* knockout mutant and PPE25-PE19 deletion, they were found attenuated in in macrophages and in the severe combined immune-deficient (SCID) mouse infection model (Bottai *et. al.,* 2012). Based on the reported results, it has been proposed that these proteins of the ESX-5 system may play a role in maintaining a fully functional cell envelope and the virulence of *Mtb.*

**Table 1.3**      Literature searches of members of ESX gene Family adapted from Mycobrowser (Lew *et. al.,* 2011).

| Gene Name | Rv Number | Cluster | Function | Functional Category | Proteomics | Transcriptome | Essential Gene |
|---|---|---|---|---|---|---|---|
| *esxA* | Rv 3875 | ESX-1 | Unknown | Cell Wall & Cell Processes | Identified in H37Rv culture filtrates (Malen *et. al.,* 2007) | mRNA identified by RT-PCR (Amoudy *et. al.,* 2006) | Essential (Elicits immune response) |
| *esxB* | Rv 3874 | ESX-1 | Unknown | Cell Wall & Cell Processes | Identified in H37Rv culture filtrates (Malen *et. al.,* 2008). | mRNA identified by RT-PCR (Amoudy *et. al.,* 2006). Possibly down-regulated by hrcA/Rv 2374c (Steward *et. al.,* 2002) | Essential (Exported protein co-transcribed with *esxA*) |
| *esxC* | Rv 3890c | ESX-2 | Unknown | Cell Wall & Cell Processes | - | mRNA identified by microarray analysis & up-regulated after 4hr, 24hr and 96hr of starvation (Betts *et. al.,* 2002) | Non-Essential by Himar1-based transposon mutagenesis in H37Rv (Sassetti *et. al.,* 2003). Deleted (partially/completely) in 1 or more clinical isolates (Tsolaki *et. al.,* 2004) |
| *esxD* | Rv 3891c | ESX-2 | Unknown | Cell Wall & Cell Processes | - | Microarray analysis detected expression in H37Rv *in vivo* (in BALB/c & SCID mice) but not *in vitro* (7H9 medium) (Talaat *et. al.,* 2004). Possibly down-regulated by hrcA/Rv 2374c (Steward *et. al.,* 2002). Up-regulated after 4hr, 24hr and 96hr of starvation (Betts *et. al.,* 2002). | Non-Essential by Himar1-based transposon mutagenesis in H37Rv (Sassetti *et. al.,* 2003). |
| *esxE* | Rv 3904c | - | Unknown | Cell Wall & Cell Processes | - | - | Non-Essential by Himar1-based transposon mutagenesis in H37Rv (Sassetti *et. al.,* 2003). |
| *esxF* | Rv 3905c | _ | Unknown | Cell Wall & Cell Processes | - | - | Non-Essential by Himar1-based transposon mutagenesis in H37Rv (Sassetti *et. al.,* 2003). |

| esxG | Rv 0287 | ESX-3 | Unknown | Cell Wall & Cell Processes | Identified in H37Rv culture filtrates (Malen *et. al.,* 2008) | Down-regulated after 4hr, 24hr and 96hr of starvation (Betts *et. al.,* 2002). Repression by iron & IdeR/Rv 2711 in H37Rv (Rodriguez *et. al.,* 2002). Predicted to be in Zur/Rv 2359 regulon (Maciag *et. al.,* 2007). | Essential |
|---|---|---|---|---|---|---|---|
| esxH | Rv 0288 | ESX-3 | Unknown | Cell Wall & Cell Processes | Spots identified on gel in short term culture filtrate at Stateus Serum Institute (Denmark) | Down-regulated after 24hr starvation (Betts *et. al.,* 2002). Repression by iron & IdeR/Rv 2711 in H37Rv (Rodriguez *et. al.,* 2002). | Non-Essential by Himar1-based transposon mutagenesis in H37Rv (Sassetti *et. al.,* 2003). |
| esxI | Rv 1037c | - | Unknown | Cell Wall & Cell Processes | Identified in H37Rv culture filtrates (Malen *et. al.,* 2007) | - | Non-Essential by Himar1-based transposon mutagenesis in H37Rv (Sassetti *et. al.,* 2003). |
| esxJ | Rv 1038c | - | Unknown | Cell Wall & Cell Processes | Identified in H37Rv culture filtrates (Malen *et. al.,* 2007) | - | Non-Essential by Himar1-based transposon mutagenesis in H37Rv (Sassetti *et. al.,* 2003). |
| esxK | Rv 1197 | - | Unknown | Cell Wall & Cell Processes | Identified in H37Rv culture filtrates (Malen *et. al.,* 2007) | - | Non-Essential by Himar1-based transposon mutagenesis in H37Rv (Sassetti *et. al.,* 2003). |
| esxL | Rv 1198 | - | Unknown | Cell Wall & Cell Processes | Identified in H37Rv culture filtrates (Malen *et. al.,* 2007) | - | Non-Essential by Himar1-based transposon mutagenesis in H37Rv (Sassetti *et. al.,* 2003). |
| esxM | Rv 1792 | ESX-5 | Unknown | Cell Wall & Cell Processes | Spots identified on gel in short term culture filtrate at Statens Serum Institute (Denmark) | - | Non-Essential by Himar1-based transposon mutagenesis in H37Rv (Sassetti *et. al.,* 2003). |
| esxN | Rv 1793 | ESX-5 | Unknown | Cell Wall & Cell Processes | Spots identified on gel in short term culture filtrate at Statens Serum Institute (Denmark) | - | Non-Essential by Himar1-based transposon mutagenesis in H37Rv & CDC1551 (Sassetti *et. al.,* 2003; Lamichhane *et. al.,* 2003). |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *esxO* | Rv 2346c | - | Unknown | Cell Wall & Cell Processes | Spots identified on gel in short term culture filtrate at Statens Serum Institute (Denmark) | - | Non-Essential by Himar1-based transposon mutagenesis in CDC 1551 (Lamichhane *et. al.,* 2003). |
| *esxP* | Rv 2347c | - | Unknown | Cell Wall & Cell Processes | - | - | Non-Essential by Himar1-based transposon mutagenesis in CDC 1551 (Lamichhane *et. al.,* 2003). |
| *esxQ* | Rv 3017c | - | Unknown | Cell Wall & Cell Processes | Identified by proteomics (Skjot *et. al.,* 2002) | - | Non-Essential by Himar1-based transposon mutagenesis in H37Rv (Sassetti *et. al.,* 2003). |
| *esxR* | Rv 3019c | - | Unknown | Cell Wall & Cell Processes | Identified by proteomics (Skjot *et. al.,* 2002) | Down-regulated after 24hr starvation (Betts *et. al.,* 2002). | Non-Essential by Himar1-based transposon mutagenesis in H37Rv (Sassetti *et. al.,* 2003). |
| *esxS* | Rv 3020c | - | Unknown | Cell Wall & Cell Processes | - | Down-regulated after 24hr starvation (Betts *et. al.,* 2002). Predicted to be in Zur/Rv 2359 regulon (Maciag *et. al.,* 2007). | Non-Essential by Himar1-based transposon mutagenesis in H37Rv (Sassetti *et. al.,* 2003). |
| *esxT* | Rv 3444c | ESX-4 | Unknown | Cell Wall & Cell Processes | - | - | Non-Essential by Himar1-based transposon mutagenesis in H37Rv (Sassetti *et. al.,* 2003). |
| *esxU* | Rv 3445c | ESX-4 | Unknown | Cell Wall & Cell Processes | - | Transcription repressed at low pH *in vitro* conditions (Fisher *et. al.,* 2002). | Non-Essential by Himar1-based transposon mutagenesis in H37Rv & CDC1551 (Sassetti *et. al.,* 2003; Lamichhane *et. al.,* 2003). |
| *esxV* | Rv 3619c | - | Unknown | Cell Wall & Cell Processes | - | - | - |
| *esxW* | Rv 3620c | - | Unknown | Cell Wall & Cell Processes | Identified in culture supernatant of H37Rv using mass spec (Mattow *et. al.,* 2003) | - | - |

### 1.1.7 PE and PPE Protein Families

The genomic organization of the ESX members is associated with two *Mtb* protein families. The PE and PPE protein families, which in *Mtb* H37Rv consist of 100 proline-glutamic acid (PE) and 67 proline-proline-glutamic acid (PPE) members, show conserved N-terminal domains bearing characteristic PE or PPE motifs and occupy more than 8% of the coding capacity of the genome. They also lack detectable secretion signals (Cole *et. al.,* 1998). Despite being highly conserved, no homology exists between the N-terminal domains of both families (Gey van Pittius *et. al.,* 2006) (Figure 1.5). Both the gene families are further divided into sub-families, based on homology and presence of characteristic motifs in the C-terminal domains. Both these protein families include members that have variable stretches of repetitive motifs in the C-terminal domains.

The PE family is subdivided into 2 subfamilies, the polymorphic GC-rich-repetitive sequence (PGRS) subfamily and the other consisting of proteins with low C-terminal homology. The PGRS subfamily consists of 65 protein of multiple tandem repeats in the C-terminal domain of either glycine-glycine-asparagine or glycine-glycine-alanine motifs (Figure 1.5). The PPE family is subdivided into 4 subfamilies, namely the PPW, SVP, MPTR and low C-terminal homology proteins. The SVP subfamily contains 24 PPE proteins with a Gly-X-X-Ser-Val-Pro-X-X-Trp motif. The MPTR subfamily comprises of 23 proteins with tandem repeats of a Asp-X-Gly-X-Gly-Asn-X-Gly motif. The PPW subfamily is a 44 amino acid region that harbors the Phe-X-Gly-Thr and Pro-X-X-Pro-X-X-Trp motifs (Figure 1.6) (Gey van Pittius *et. al.,* 2006). Given the genetic variability it was originally suggested that gene family could be a source of antigenic variation but there is no convincing functional evidence to date that has established this (Cole *et. al.,* 1998).

Through phylogenetic and sequence analyses, Gey van Pittius and colleagues reported that the PE and PPE gene families are linked to duplications of the ESAT-6 gene cluster, thus implying that they are the ancestral copies of the two gene families. In addition, the PGRS and MPTR subfamilies recently evolved from defined branching points in mycobacteria and are only found in members of the *Mtb* complex and close relatives

(Gey van Pittius *et. al.,* 2006). Several PE and PPE proteins were identified on the cell envelope of *Mtb*, and this feature further emphasises the potential importance of these proteins for their interaction with host cells. Their abundance in the genome of *Mtb* is intriguing and suggests that these proteins have an important biological role yet to be discovered (Abdallah *et. al.,* 2006).

To date, very little is known regarding the function of the varying domains of the PE and PPE proteins. The linkage and location of the PE and PPE genes relative to ESX genes, and most of the ESX clusters, implies a functional link between these different gene families. PPE68 (encoded from the ESAT-6 gene cluster region 1) has been shown to interact with the ESX proteins *esxA, B* and *H* (Okkels and Andersen, 2004). One PE protein LipY is the only PE protein that has been functionally characterized (Deb *et. al.,* 2006). The investigators established under nutrient deprived conditions, that this protein was involved in the degradation of triacyl-glycerols (TAGs). Since this protein is a major active lipase, it was hypothesised that LipY is crucial in fatty acid metabolism during the dormant and reactivated stage of *Mtb* infection (Deb *et. al.,* 2006). This hypothesis was in agreement with data published later by Mishra and colleagues in 2008, with the LipY protein in *M. marinum* that contains a PPE domain instead of the PE domain (Mishra *et. al.,* 2008). In a study by Daleke and colleagues in 2011, they demonstrated that the LipY homologes in *Mtb* and *M. marinum* are secreted by ESX-5 and the LipY homologues and rely on the respective PE and PPE domains for secretion (Daleke *et. al.,* 2011).
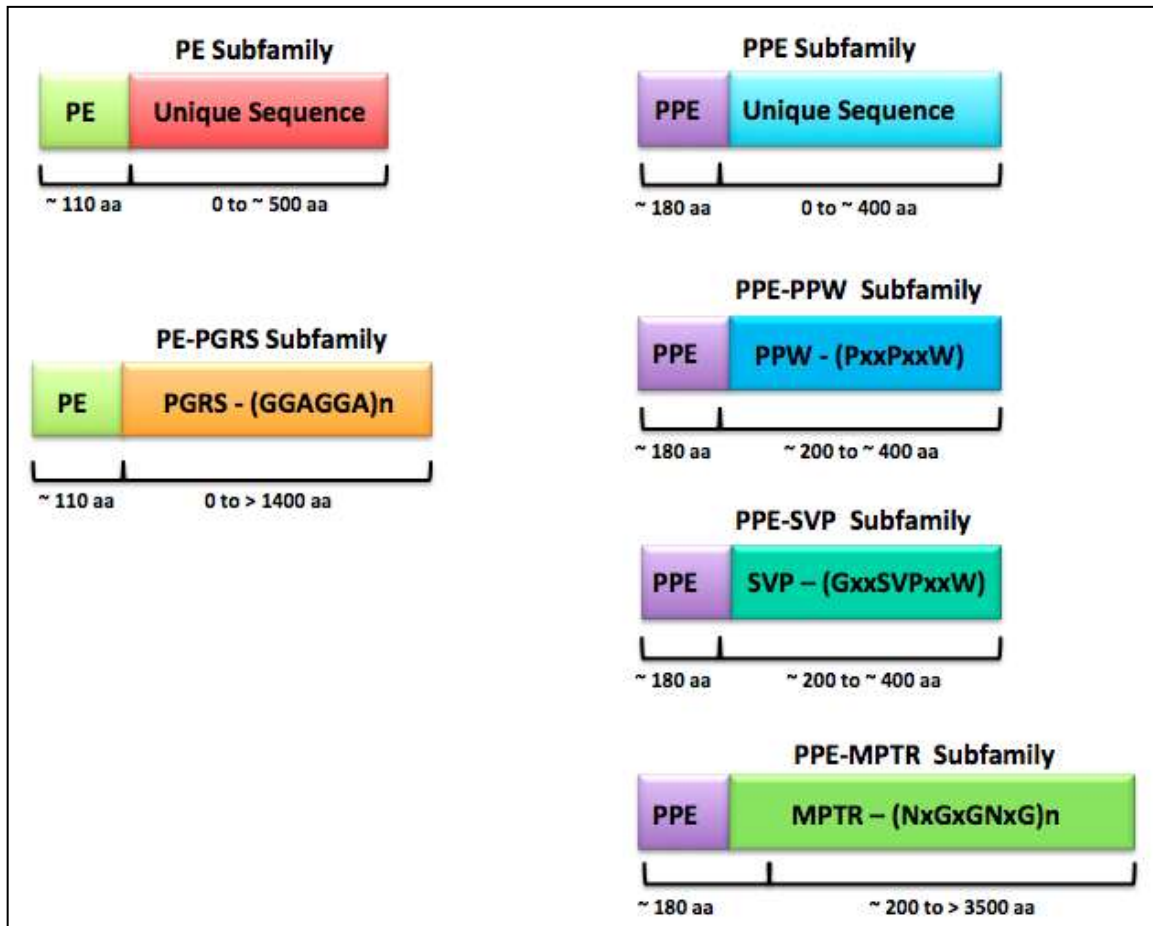
**Figure 1.5** Illustration of conserved N-terminal region in the PE and PPE gene family, and the varying C-terminal region between the sub-families (adapted from Gey van Pittius *et. al.,* 2006).

## 1.2    Immunogenicity of ESX Proteins

TB was one of the first diseases for which an attenuated vaccine was developed (Calmette and Plotz, 1929). It is widely believed that antigenic proteins induce T-cell mediated immune responses. Studies have demonstrated that upon *Mtb* infection, the inflammatory immune response is activated by elevated expression levels of TNF-α (Ogawa *et. al.,* 1991; Flynn *et. al.,* 1995; Mohan *et. al.,* 2001; Abebe *et. al.,* 2010) and IFN-γ (Kawamura*,* 2006; Ma *et. al.,* 2003; Hussain *et. al.,* 2007), in response to the *Mtb* antigens (Sullivan *et. al.,* 2005; Goter-Robinson *et. al.,* 2006; Hervas-Stubbs *et. al.,* 2006; Fletcher, 2007).

With the discovery of the ESX members within *Mtb,* research has been nudged in the direction of their usage as potential immunological markers for *Mtb* infections. Consequently, the implementation of enzyme-linked immunospot (ELISPOT) assays with specific *Mtb* antigens has been significantly instrumental in the diagnosis of TB and latent TB infections in some settings (Pai *et. al.,* 2004). This was demonstrated in TB outbreaks occurring in the United Kingdom where the reported number of ELISPOT responses among TB contacts showed a better correlation with the degree of exposure to the infection than the tuberculin skin tests (TSTs) (Lalvani *et. al.,* 2001a; Lalvani *et. al.,* 2001b; Ewer *et. al.,* 2003).

### 1.2.1 Immunogenicity of Other ESX members

There has been an increase in the identification of T-cell antigens in *Mtb*, and interestingly within the genome, these antigens are localised in clusters and in the neighbourhood of ESX operons (Okkels and Andersen, 2004). The clusters have genes encoding for membrane proteins and ATP-binding proteins. Based on this arrangement, it promotes the idea that the T-cells antigens form part of an immunogenicity island. Examples of other T-cell antigens encoded by a single region include Rv1196 (PPE18) (Dillon *et. al*., 1999) and Rv1198 *(esxL)* (Alderson *et. al*., 2000), Rv0286 (PPE4) (Okkels *et. al*., 2003) and Rv0288 *(esxH)* (Skjøt *et. al*., 2000), Rv3021c (PPE47) (Okkels *et. al*., 2003) and Rv3019c *(esxR)* (Skjøt *et. al*., 2002), Rv3018c (PPE46) (Okkels *et. al*., 2003) and Rv3017c *(esxQ)* (Skjøt *et. al*., 2002). Choudhary and colleagues have shown the Gly-X-Gly-Asn-X-Gly repeat motif of PPE41 can stimulate humoral and cell mediated immunity (Choudhary *et. al.*, 2003). Using cattle models of *M. bovis* infection and BCG vaccination, Cockle and colleagues analysed highly immunogenic antigens from the RD1, RD2 and RD14 of BCG Pasteur. They identified 8 highly immunogenic antigens, of which 4 belonged to the PE and PPE families (Rv3872, Rv3873, Rv1983 and Rv1768) that had responder frequencies between 45 and 82% (Cockle *et. al.,* 2002). Using ELISA and T-cell proliferation assays, PPE42 displayed a positive reaction to patient serum samples, therefore suggesting that PPE proteins play a role in T-cell response (Chakhaiyar *et. al.,* 2004).

Proteins from the ESX-5 system may also be potential biomarkers for tuberculosis serodiagnosis (Romano *et. al.*, 2008). The PPE17 (Rv1168c) protein has been shown to display strong immunoreactivity against serum from patients with active TB thus being used to distinguish TB infected patients from *M. bovis* BCG vaccinated patients as well as detect smear-negative pulmonary TB (Khan *et. al.*, 2008).

### 1.3 TB Vaccine Developments

Currently, new approaches, strategies and technological advances are providing significant insights in TB vaccine development. The strategies for a new vaccine include: (1) the development of a "prime-boost" strategy where BCG is given to newborn babies and the new vaccine is implemented as a "booster" dose later on, (2) an "early booster" vaccine that is implemented alongside childhood vaccines to help amplify immune responses, (3) administering of "late-booster" vaccines to schoolchildren, adolescents and adults with the intension of the reduction in progression from latent TB to active TB and (4) implement a new vaccine with existing TB drug therapy to shorten and simplify the treatment regimen (Kaufmann, Hussey and Lambert, 2010; WHO; 2012; WHO, 2013).

As of 2012, 12 vaccine candidates were in the clinical trial stage of development, where 11 are aimed at the prevention of the disease while the remaining candidate is an immunotherapeutic vaccine (WHO, 2012; WHO, 2013).

### 1.3.1 ESAT-6 and CFP-10 in Vaccine Development

The ESX systems represent a new area to localize and target vaccine. In 1996, a study by Brandt and colleagues reported that ESAT-6 had a strong T-cell immune response in 5 of the 6 genetically varied strains of inbred mice, which was initiated by the recognition of the two major epitopes in ESAT-6 N-terminus (named P1) and C-terminus (named P6). Furthermore, approximately 25-35% of these ESAT-6 specific T-cells belonged to the mycobacterium-reactive T-cell repertoire that is activated in the initial phases of the disease (Brandt *et. al.,* 1996). A similar trend where ESAT-6 was a target for T-cell responses was observed within the first phases of infection in cattle, guinea pigs and humans (Pollock and Andersen, 1997; Elhay *et. al*., 1998; Ravn *et. al*., 1999). Moreover, in murine models, ESAT-6 has induced protective immune responses as well as being recognised by the sera from TB patients (Andersen *et. al*., 1995; Kanaujia *et. al*., 2003; Lopez-Vidal *et. al*., 2004). In addition, ESAT-6 has been shown to display a discriminatory role, in differentiating tuberculosis patients from healthy individuals (Ulrichs *et. al.,* 1998; Cardoso *et. al.,* 2002). Macedo and colleagues conducted a study

amongst Brazilian patients to discriminate infected from non-infected patients with potential *Mtb* antigens (Macedo *et. al.,* 2011). Their findings indicated that the ESAT-6 and AG85B antigens were able to differentiate tuberculosis patients from healthy patients by IFN-γ and TNF-α production, thus behaving as biomarkers for the clinical status of the patients.

The usage of ESAT-6 as a fusion protein was explored by Tebianian and colleagues (Tebinanian *et. al.*, 2011). They developed the E6H70C fusion protein, by the fusion of ESAT-6 to the C-terminus of *Mtb* HSP70 (HSP70$_{359-610}$). Immunogenicity of this fusion protein was tested in the mouse model. Immunisation with the E6H70C protein resulted in the induction of a specific immune response in comparison to ESAT-6 immunised mice. This was indicative of increased production levels of the ESAT-6 specific antibody, IFN-γ and proliferation of splenocytes. Hence, this protein has been proposed as a potential vaccine candidate in the fight against TB.

Likewise, CFP-10 also plays a role in eliciting the immune response, as it is a target for human B and T-cell responses. In Guinea pigs, the CFP-10 antigen induces high levels of IFN-γ and delayed-type hypersensitivity (Colangeli *et. al.*, 2000; Dillon *et. al.*, 2000; van Pinxteren *et. al.*, 2000; Skjøt *et. al.*, 2000).

### 1.3.2 Other ESX Members in Vaccine Development

The ESAT-6 family contains sub-families of nearly identical genes. The Rv3019c *(esxR)* and Rv3017c *(esxQ)* genes are two such members of the ESAT-6 family exhibiting 84% and 74% amino acid identity to the Rv0288 *(esxH)* gene. However, epitope mapping of the three proteins lead to the discovery of unique epitope patterns in Rv3019c *(esxR)* and Rv3017c *(esxQ)* (Skjøt *et. al*., 2002). Similarly, based on a 90% amino acid sequence identity, a further 5 members of the ESAT-6 family displayed similar epitope patterns: Rv1793 (*esxN*), Rv3619c *(esxV),* Rv1198 *(esxL),* Rv1037c *(esxI)* and Rv2346c *(esxO)* (Alderson *et. al*., 2000). In a study conducted by Majlessi and colleagues (Majlessi *et. al.*, 2003), they reported that two ESAT-6 members, Rv0288 *(esxH)* and Rv3019c *(esxR),* present as targets for CD8 T-cells. This finding insinuates that these proteins are capable, in some way, of accessing the cytoplasm and the class-I processing machinery housed within the proteasome.

Other studies have also revealed the powerful diagnostic potential of incorporating the ESAT-6 and CFP-10 and other ESX antigens in the detection of *Mtb* in cattle which suggest they may also be useful for vaccine development (Buddle *et. al.,* 2001; Lalvani *et. al.,* 2001a; Lalvani *et. al.,* 2001b; Vordermeier *et. al.,* 2001; Aagaard *et. al.,* 2003; Aagaard *et. al.,* 2006; Jones, Hewinson and Vordermeier, 2010a; Jones *et. al.,* 2010b). Vordermeier and colleagues in 2001 reported that these antigens are also differentially recognized by *M. bovis* infected cattle compared to the BCG vaccinated cattle (Vordermeier *et. al.,* 2001). In 2010, two studies conducted by Jones and colleagues, involved blood based assays to screen potential *M. bovis* secreted antigens to identify immunogenic targets with the aid of overlapping peptides representing the ESAT-6 family. The aim was to use these peptide cocktails to distinguish between *M. bovis* infected and BCG vaccinated animals. In the earlier study conducted in 2010, Jones and colleagues reported on 3 peptide pools, from 379 pools of overlapping peptides, that was frequently recognised in infected cattle but did not induce IFN-γ responses in the vaccinated animals. Of the three reported peptide pools, two comprised of Rv2346c (*esxO*) and Rv3020c (*esxS*) overlapping peptides and the other, Sec 2, consisted of a cocktail of nine peptides from multiple antigens. The peptide pools of Rv2346c (*esxO*)

and Rv3020c (*esxS*) antigens induced a 61% and 57% response frequency in infected cattle respectively, while antigens of the Sec 2 cocktail elicited a 54% response in infected cattle (Jones, Hewinson and Vordermeier, 2010a; Jones *et. al.,* 2010b).

In a later study conducted in 2010, Jones and colleagues proposed that the analysis of secreted proteins could reveal further immunogenic antigens, which may be potential vaccine candidates for *Mtb* infections or diagnostic components in tests. They screened 119 secreted proteins, from a pool of 382 overlapping peptides, of which 70 (59%) induced positive responses in infected cattle. The positive responder frequencies ranged from 4 to 65%. Interestingly, one of the 15 ESAT-6 proteins tested was recognised by approximately 30% of the infected cattle. Curiously, the authors noted an IFN-γ response for Rv2346c (*esxO*) in more than half of the animals tested, despite the deletion of the Rv2347c (*esxP*) binding partner in *M. bovis*. In this study, peptide-mapping experiments on the recurrently recognised peptides revealed those particular peptides were members of the Mtb9.9 and the QILSS subfamilies and were located in specific regions of sequence diversity amongst this group of proteins. Thus implying that their specific locations within the regions of amino acid diversity could be linked to antigenicity (Jones *et. al.,* 2010b).

Further *Mtb* ESX members were identified by Betts and colleagues (Betts *et. al.,* 2002) using microarray and proteome analyses to examine the bacterium response to nutrient starvation. The proteome experiments revealed that several proteins, including *esxC, esxD, esxR, esxS, esxG* and *esxH,* were induced as a result of starvation. Microarray analyses revealed the reduction of the transcription apparatus and other cell mechanisms including the induction of genes involved in the mainatinance of the bacterium within the host. Up-regulation was observed for *esxC, esxD, esxR* and *esxS* gene expression after 4 hours, 24 hours and 96 hours of starvation. Down-regulation of *esxG* and *esxH* gene expression was observed after a 24 hour starvation period.

The members of the ESX-5 system are additional vaccine candidates where some PPE proteins have shown potential to be incorporated as vaccine components. These include PPE42 (Rv2608), PPE44 (Rv2770c), PPE57 (Rv3425) and PPE68 (Rv3878). PPE42 (Rv2608) together with the TLR-9 agonist CpG conferred partial protection in mice. Also the ID83 fusion protein (combination of Rv1813, Rv3620 and PPE42) stimulated immunity in mice (Bertholet *et. al.,* 2008).

## 1.4 Impact of Genetic Variability in *Mtb*

With the availability of whole-genome sequencing, this technology has allowed for the comparative analysis of the H3Rv genome sequence with other laboratory and clinical strains, consequently providing new insights into the characteristics of genetic diversity between strains (Brosch *et. al.,* 2001). This has also enabled the identification of SNPs that can be used as potential genetic markers in strain typing as well as to gain insight into the evolution history of these organisms and aid in the genotyping of worldwide strain collections (Gutacker *et. al.,* 2002; Filliol *et. al.,* 2006; Hershberg *et. al.,* 2008; Comas *et. al.,* 2010, Mestre *et. al.,* 2011). RDs, or indels represent unique events in the genealogy of an *Mtb* strain and provide additional sources of genetic variability (Schürch *et. al.,* 2011).

SNP analysis is beneficial in addressing various biological questions. Non-synonymous SNPs (nsSNPs) that confer drug resistance could be favourable in understanding the nature and spread of drug resistance between and within populations. Synonymous SNPs (sSNPs) are generally considered functionally neutral (although this is not always the case), therefore their inclusion in analysis may confer additional insight into genetic drift and evolutionary relationships of *Mtb* strains (Filliol *et. al.,* 2006). Furthermore, the presence of sSNPs and RDs are symbolic of an ancestral state of that particular strain and can be used as molecular markers for clonal complexes (Schürch *et. al.,* 2011).

Comas and colleagues reported on the hyperconservation of human T cell epitopes among a representation of 21 strains of 6 geographical *Mtb* lineages. Using whole genome sequencing, the uncovered more than 9 000 SNPs with the Illumina genome analyzer sequencing technology (Comas *et. al.,* 2010). On the other hand, Uplekar and colleagues identified 109 unique SNPs across the 23 ESX genes from 108 clinical samples using amplicon sequencing. They reported a high prevalence of 3 SNPs in *esxV* that occurred in 74 isolates from different lineages, thus implying these positions as lineage markers (Uplekar *et. al.,* 2011). In a recent study by Mestre and colleagues, sequence data from 58 Beijing clinical isolates, from various geographic locations, lead to the identification of polymorphisms in the DNA replication, recombination and repair

36

(3R) genes (Mestre *et. al.,* 2011). The SNPs identified in this study allowed for the construction of a phylogenetic network that facilitates the implementation of these unique SNPs as potential genetic markers for these isolates. However, despite these recent studies indicating that there is diversity in antigens, albeit limited compared to other organisms, there have been few studies actually linking genotype of *Mtb* strains to an immune-phenotype.

# CHAPTER TWO

# Evaluation of ESX Sequence Variations within *M. tuberculosis* Clinical and Laboratory Isolates

## 2.1 Chapter Summary

In this chapter, Clinical and Laboratory isolates are evaluated for potential ESX sequence variations using amplicon sequencing. Using spoligotyping and a strain specific test, the genotypes for all isolates were confirmed. Following amplicon sequencing, all 23 ESX genes were evaluated for variations between the Beijing, KwaZulu-Natal (KZN) and Other genetic isolate groupings as well as within isolates. These results identify new mutations in the ESX family, some of which were not detected by genome sequencing.

## 2.2 Materials and Method

### 2.2.1 Bacterial Culture and Growth Conditions

In this study, 55 Mycobacterial clinical isolates and laboratory strains (Table 2.1a and b) were kindly provided by Dr. M. Pillay (Department of Medical Microbiology, UKZN). These isolates were obtained from archived collections in the Department of Medical Microbiology within the last 15 years that were isolated as part of routine Department of Health diagnostics. The strains were used from 3 studies that have ethical approval from BREC to store and molecularly analyze. These studies were: Rapid Detection Study (Ethic Number: E157/04); Department of Health (DOH) surveillance; *M. vaccae* clinical trial and Westville Prison Molecular Epidemiology Study (Ethic Number: H084/00).

Mycobacteria were maintained in either liquid Middlebrook 7H9 medium (Difco) or solid Middlebrook 7H11 agar enriched with Albumin Dextrose Catalase (ADC) (Difco) and 0.05% Tween 80. Cultures were grown at 37˚C, with agitation in the biosafety level three (BSL3) facility within the Department of Medical Microbiology. Glycerol stocks of the cultures were maintained at -70˚C in 30% (v/v) glycerol.

## 2.2.2 Genotypic Confirmation of Clinical Isolates

The Research Laboratory team in the Department of Medical Microbiology at the University of KwaZulu-Natal, Nelson Mandela Medical School under the leadership of Dr. M. Pillay performed confirmations of the isolates genotypes. Both the *IS6110* RFLP method (adapted from van Soolingen *et. al.,* 1994) and the spoligotyping method (Isogen BioSciences, Maarssen, Netherlands) were used.

A strain specific molecular test, developed and performed by Miss N. Pillay in the Department of Medical Microbiology at the University of KwaZulu-Natal as part of her Masters dissertation, was additionally performed to confirm the isolates genotypes. This test was developed to assess the prevalence and distribution of the F15/LAM4/KZN within the province of KZN and used as a screening tool for a province wide drug susceptibility survey (Pillay, 2010).

**Table 2.1a**     Bacterial strains utilized in this study.

| Isolate Number: | Isolates: | RFLP / Strains: | Isolated From: | Isolation Date: |
|---|---|---|---|---|
| 1 | R 26 | KZN variant 1 | Sputum | 1995-96 |
| 2 | R 62 | Neither | Sputum | 1995-96 |
| 3 | R 224 | Neither | Sputum | 1995-96 |
| 4 | R 226 | Neither | Sputum | 1995-96 |
| 5 | R 252 | Neither | Sputum | 1995-96 |
| 6 | R 253 | KZN variant 1 | Sputum | 1995-96 |
| 7 | R 257 | KZN variant 1 | Sputum | 1995-96 |
| 8 | R 295 | KZN variant 1 | Sputum | 1995-96 |
| 9 | R 299 | Neither | Sputum | 1995-96 |
| 10 | R 300 | Neither | Sputum | 1995-96 |
| 11 | R 339 | KZN variant 1 | Sputum | 1995-96 |
| 12 | R 351 | KZN variant 1 | Sputum | 1995-96 |
| 13 | R 375 | KZN variant 1 | Sputum | 1995-96 |
| 14 | R 389 | KZN variant 1 | Sputum | 1995-96 |
| 15 | R 402 | Neither | Sputum | 1995-96 |
| 16 | R 413 | KZN variant 1 | Sputum | 1995-96 |
| 17 | R 426 | KZN Archetype | Sputum | 1995-96 |
| 18 | R 434 | KZN variant 1 | Sputum | 1995-96 |
| 19 | R 443 | F28 | Sputum | 1995-96 |
| 20 | R 467 | KZN variant 1 | Sputum | 1995-96 |
| 21 | R 492 | F28 | Sputum | 1995-96 |
| 22 | R 502 | KZN variant 1 | Sputum | 1995-96 |
| 23 | R 503 | KZN variant 1 | Sputum | 1995-96 |
| 24 | R 504 | KZN variant 1 | Sputum | 1995-96 |
| 25 | R 506 | KZN variant 1 | Sputum | 1995-96 |
| 26 | R 525 | KZN variant 1 | Sputum | 1995-96 |
| 27 | R 576 | Neither | Sputum | 1995-96 |
| 28 | R 623 | KZN variant 1 | Sputum | 1995-96 |
| 29 | H37Rv | Reference Strain | | |

**Table 2.1b**     Bacterial strains utilized in this study.

| Isolate Number: | Isolates: | RFLP/Strains: | Isolated From: | Isolation Date: |
|---|---|---|---|---|
| 30 | H37Ra | Avirulent Strain | | |
| 31 | 910 P4 | Beijing | Sputum | 1996-01 |
| 32 | 1784 P5 | Beijing | Sputum | 1996-01 |
| 33 | 1528 P8 | Beijing | Sputum | 1996-01 |
| 34 | Vac 1435 | KZN variant 1 | Sputum | 1996-01 |
| 35 | Vac 4207 | KZN variant 1 | Sputum | 1996-01 |
| 36 | Vac 4258 | KZN variant 1 | Sputum | 1996-01 |
| 37 | Vac 2475 | KZN variant 1 | Sputum | 1996-01 |
| 38 | KZN 605 | KZN variant 1 | Sputum | 1996-01 |
| 39 | Vac 666 | KZN variant 1 | Sputum | 1996-01 |
| 40 | Vac 8426 | KZN variant 1 | Sputum | 1996-01 |
| 41 | 43178 | Beijing | Sputum | 2002 |
| 42 | WPO6 25210 | Beijing | Sputum | 2002 |
| 43 | WPO6 10263 | Beijing | Sputum | 2002 |
| 44 | WPO6 2705 | Beijing | Sputum | 2002 |
| 45 | 39321 | Beijing | Sputum | 2002 |
| 46 | 43117 | Beijing | Sputum | 2002 |
| 47 | 48246 | Beijing | Sputum | 2002 |
| 48 | WPO6 21386 | Beijing | Sputum | 2002 |
| 49 | WPO7 7021344 | Beijing | Sputum | 2002 |
| 50 | WPO7 7027955 | Beijing | Sputum | 2002 |
| 51 | WPO7 7037345 | Beijing | Sputum | 2002 |
| 52 | WPO7 7075078 | Beijing | Sputum | 2002 |
| 53 | WPO7 7084874 | Beijing | Sputum | 2002 |
| 54 | WPO7 7023338 | Beijing | Sputum | 2002 |
| 55 | WPO7 7041516 | Beijing | Sputum | 2002 |
| 56 | WPO6 2731 | Beijing | Sputum | 2002 |
| 57 | WPO7 7115992 | Beijing | Sputum | 2002 |
| 58 | BCG | Reference Strain | | |

### 2.2.3 *In Silico* Genome Comparisons of Publically Available Databases:

*In silico* comparisons of the ESX sequences was conducted using the DNAStar MegAlign software (LaserGene 7.2, Madison, Wisconsin, USA). Genome comparisons were employed using partial and completed genome sequences that were downloaded from public websites and databases (TBDB, TubercuList and GenoList) (Table 2.2). Comparisons were done with the corresponding ESX sequences from the *Mtb* H37Rv reference strain. This was done at the initiation of the studies and does not reflect the rapid expansion of genome sequences that have become available in the last year.

**Table 2.2**    List of Public Available Websites and Databases for ESX sequences.

| Website/Database | Website Adress | Reference |
|---|---|---|
| TBDB | http://www.tbdb.org | Reddy, *et. al.,* 2009 |
| TubercuList | http://genolist.pasteur.fr/TubercuList | Lew, *et. al.,* 2011 |
| GenoList | http://genodb.pasteur.fr/cgi-bin/WebObjects/GenoList | Lechat, *et. al.,* 2008 |

### 2.2.4 Genomic DNA Extraction using CTAB/NaCl Method:

The genomic DNA extraction was achieved using the CTAB/NaCl method as previously described, (van Soolingen, *et. al*., 1994). Scraped bacterial colonies were heat inactivated for 30 minutes at 80˚C in 500μl TE buffer (100mM Tris/HCL; 10mM EDTA, pH 8.0), and in turn treated with 50μl lysozyme (10mg/ml). This mixture was incubated for 1 hour at 37˚C, and subsequently treated with 75μl of the Proteinase K-SDS mixture (10mg/ml; 10% v/v). Following brief vortexing and a further 10 minutes incubation at 65˚C, 100μl of pre-warmed CTAB/NaCl (10% *N*-cetyl-*N,N,N,*-trimethyl-ammonium bromide, 0.73M NaCl) was added. The liquid contents were vortexed until the mixture appeared milky and was incubated for 10 minutes at 65˚C. This was followed by an addition of 750μl of chloroform/isoamyl alcohol (24:1, v/v), vortexing for 10 seconds following a 20 minute centrifugation at maximum speed. The nucleic acid was precipitated with the addition of 500μl isopropanol to the supernatant. Upon DNA thread formation, centrifugation was

performed at maximum speed for 30 minutes. The DNA pellet was washed twice with 70% cold ethanol to remove residual CTAB/NaCl, and air-dried for 5 minutes at room temperature. The resultant pellet was re-dissolved in the appropriate amount of TE Buffer (100mM Tris/HCL; 10mM EDTA, pH 8.0) and stored at 4˚C. DNA concentration and purity was measured by optical density at 260nm.

### 2.2.5 PCR of ESX genes:

Primers were designed for the 11 ESX gene pairs of interest using the Primer3 program (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) (Table 2.3a). In some instances it was necessary to amplify single genes in order to retrieve suffucient single PCR products. Individual ESX gene primers were also designed for *esxA, esxB, esxR, esxS, esxT and esxU* (Table 2.3b). The PCR reactions for the ESX gene pairs were performed using EconoTaq Buffer and 5U/μl of EconTaq polymerase (Invitrogen, Saint Aubin, France), 25mM nucleotide mix, 2pM of each primer, 1-10ng of template DNA and nuclease-free water to a final volume of 25μl. For the PCR of the individual ESX genes, the GoTaq Colorless Mastermix (Promega Corp., USA) was applied as stipulated by protocol dictated by the manufacturer. The thermal cycling was performed in a Bio-Rad T100 Thermal Cycler machine and the GeneAmp PCR system 9700 PCR machine (Applied Biosystems, Foster City, CA, USA) respectively. Primer specificity was confirmed through gel electrophoresis of the amplicons with Sybr safe (Invitrogen, Saint Aubin, France).

**Table 2.3a**      Primers for Sequencing ESX Gene Pairs.

| Genes | Primer Sequence (5' – 3') |
|---|---|
| *esxCD* | F:  ATCGACAGGTCCGCAGAG |
|  | R:  TAGCAGCAAGCAGAAGGTG |
| *esxEF* | F:  GTGCTGTGTGCTGGTGA |
|  | R:  GAGATCACCGCACCCAAAC |
| *esxGH* | F:  GACCGCAACCAAAGAAC |
|  | R:  CCAGCACCCACGGAAAG |
| *esxIJ* | F:  AGTCATAACCTGTCCGCCAC |
|  | R:  TCCCAGTTCAGCACCATCC |
| *esxKL* | F:  GGCGCAGACTGTCGTTATTT |
|  | R:  AACACCCCAGCACTGACCAC |
| *esxMN* | F:  AAGGAGAGGGGGAACATCC |
|  | R:  ATCCATCGCTACCTCA |
| *esxOP* | F:  GGGCGCAGACTGTCATTATT |
|  | R:  CTTAGCGGAGGCACCAGAG |
| *esxQ* | F:  TTCGATCGAAAGAGTGTCTA |
|  | R:  ACGAACATCGCCGCCAAC |
| *esxVW* | F:  TTTAACAACTTCGCTGC |
|  | R:  AGTGTTCCCAACGACGAC |

F: Forward Primer; R: Reverse Primer

45

**Table 2.3b**    Primers for Sequencing Individual ESX genes.

| Genes | Primer Sequence (5' – 3') |
| --- | --- |
| *esxA* | F: AGGCCGGCGTCCAATACT |
| | R: TCAGAGTGCGCTCAAACGTA |
| *esxB* | F: GGTGAGCTCCCGTAATGACA |
| | R: GTGACATTTCCCTGGATTGC |
| *esxR* | F: CAAGCCAATTTGGGTGAGG |
| | R: CTACCGGATCCACCAACAG |
| *esxS* | F: GGGGCCGGATTTGGTCG |
| | R: GCACGCTGCAGAGCTTG |
| *esxT* | F: GCTCTACCACGTCCTGCAC |
| | R: AGCTTACCGACGAGCATCC |
| *esxU* | F: GTGGACGATTCTGCTCGAC |
| | R: CGGTGGTGTGGATCTCCT |

F: Forward Primer; R: Reverse Primer

### 2.2.6 Sequencing of ESX Genes:

Upon amplification of PCR products, primers were eliminated by incubating 10μl of the PCR product with 1U of Shrimp Alkaline Phosphatase (SAP) (Fermentas Life Sciences, USA) and 10U of Exonuclease I (ExoI) (Fermentas Life Sciences, USA) for 25 minutes at 37℃ followed by 15 minutes at 80℃. To 1μl of this treated reaction mixture, 0.4μl of Big Dye sequencing mix (Applied Biosystems, Foster City, CA, USA), 1.6μl of the 2pM primer and 2μl of the 5x Buffer (5mM $MgCl_2$/200 mM Tris-HCl, pH 8.8) and nuclease-free water were added to a final volume of 10μl. Thermal cycling was performed on this mixture with an initial denaturation step of 1 minute at 96℃, followed by 35 cycles of 10sec at 96℃, 5 seconds at 50℃ and 4 minutes at 60℃. The DNA precipitation reactions were performed in 96-well plates to a final volume of 10μl. To each well, 1μl of EDTA

(125mM, pH 8.0) was added, followed by 26µl of a combined mixture of NaOAc (3M, pH 5.2) and 100% ethanol. Following centrifugation at 3000xg for 10 minutes at 18ºC, pre-chilled 35µl of 70% ethanol (v/v) was added to each well and centrifuged at 3000xg for 5 minutes. The plate was dried in a thermal cycler at 50°C for 5 minutes. The reactions were dissolved in 10µl of Hi-Di Formamide (Applied Biosystems, Foster City, CA, USA) denatured in a thermal cycler (95°C for 3 minutes, 4°C for 3 minutes) and subjected to automated sequencing on an ABI Prism 310 genetic sequence analyzer (Applied Biosystems, Foster City, CA, USA). Resultant gene sequences were subjected to comparison and alignment with the DNAStar SeqMan sequence assembler and MegaAlign software (LaserGene 7.2, Madison, Wisconsin, USA).

### 2.2.7 SNP Detection:

The 23 ESX sequences for the H37Rv reference strain were downloaded from the TubercuList database (Lew *et. al.,* 2011) (http://tuberculist.epfl.ch). Using the DNAStar MegaAlign software (LaserGene 7.2, Madison, Wisconsin, USA), these sequences were aligned and a phylogenetic tree was complied, using the Geneious software (Drummond *et. al.,* 2012), to survey the clustering patterns of the sequences.

The ESX sequences from the 58 clinical isolates were compared to the corresponding sequences of *Mtb* H37Rv reference strain. Using the BLAST function available on the Tuberculist website (http://tuberculist.epfl.ch), positions of the variant nucleotides were documented as SNPs. The SNPs were further characterised by comparing the amino acid resulting from the substitution with the reference amino acid from H37Rv, into the sSNPs, no change in amino acid, and non-synonymous nsSNPs, resulting in a change in the amino acid mutations.

### 2.2.8 Detection of Selection:

The dN/dS ratio allows for the measurement of the type of selection occurring on codon alignments.  However, the number of SNPs occurring in the individual ESX genes was too low for inclusion in this ratio. Therefore, for each isolate, the 23 ESX genes were concatenated to generate single sequences, and used in the subsequent analyses. dS and dN are the numbers of synonymous and nonsynonymous substitutions per site, respectively. The variance of the difference was computed using the bootstrap method (500 replicates). Analyses were conducted using the Nei-Gojobori method (Nei and Gojobori, 1986). The analysis involved the 58 nucleotide concatenated sequences. All positions containing gaps and missing data were eliminated. There were a total of 880 positions in the final dataset. Evolutionary codon-based analyses were conducted in MEGA5 (Tamura, *et. al.,* 2011). The program estimates the number of synonymous mutations per synonymous site (dS) and the number of nonsynonymous mutations per nonsynonymous site (dN) as well as the variances of the estimates. This estimate was then used in testing the null hypothesis that the genes are undergoing neutral ($H_o$: dN=dS), purifying ($H_o$: dN<dS) or positive ($H_o$: dN>dS) selection.

## 2.3  Results and Discussion

### 2.3.1 Survey of ESX Diversity of publically available genomes

An initial review of available sequenced genomes on public databases revealed the ESX sequences are highly conserved amongst the genomes reviewed (Table 2.4). This exercise was conducted at a time when a sizeable number of genome sequences were uncompleted in this survey, due to the lack of closure at the time of data retrieval. As a result, this is not a complete assessment of ESX gene diversity in the public available databases relative to the H37Rv ESX sequences. However, this preliminary survey indicated that not all annotated members of the ESX gene family were potentially transcriptionally active due to deletion, frameshift or truncation. This is consistent with an earlier study cataloguing indels (Marmiesse *et. al.,* 2004) and confirms that some ESX members are not essential for human infection.

**Table 2.4**     Survey of ESX genes in sequenced genomes available on public databases representing presence (red), deletion (yellow), stop codon (green), frame shift or truncation (purple) and unavailable sequence (white).

| esx Genes | esx Cluster | H37Rv | Bovis AF2122/97 | BCG Pasteur 1173P2 | M. canetti | CDC 1551 | TBC | Haarlem | F11 | 98-R604 | I435 | 4207 | 605 | V:148 | H37Ra | 82-1387 | EAS054 | GM.1503 | T17 | T85 | T92 | 94-M4241A | T46 | CPHL-A | K85 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | Esx-1 | | | | | | | | | | | | | | | | | | | | | | | |
| B | | | | | | | | | | | | | | | | | | | | | | | | |
| C | Esx-2 | | | | | | | | | | | | | | | | | | | | | | | |
| D | | | | | | | | | | | | | | | | | | | | | | | | |
| E | | | | | | | | | | | | | | | | | | | | | | | | |
| F | Esx-3 | | | | | | | | | | | | | | | | | | | | | | | |
| G | | | | | | | | | | | | | | | | | | | | | | | | |
| H | | | | | | | | | | | | | | | | | | | | | | | | |
| I | | | | | | | | | | | | | | | | | | | | | | | | |
| J | | | | | | | | | | | | | | | | | | | | | | | | |
| K | | | | | | | | | | | | | | | | | | | | | | | | |
| L | | | | | | | | | | | | | | | | | | | | | | | | |
| M | Esx-5 | | | | | | | | | | | | | | | | | | | | | | | |
| N | | | | | | | | | | | | | | | | | | | | | | | | |
| O | | | | | | | | | | | | | | | | | | | | | | | | |
| P | | | | | | | | | | | | | | | | | | | | | | | | |
| Q | | | | | | | | | | | | | | | | | | | | | | | | |
| R | | | | | | | | | | | | | | | | | | | | | | | | |
| S | | | | | | | | | | | | | | | | | | | | | | | | |
| T | Esx-4 | | | | | | | | | | | | | | | | | | | | | | | |
| U | | | | | | | | | | | | | | | | | | | | | | | | |
| V | | | | | | | | | | | | | | | | | | | | | | | | |
| V | | | | | | | | | | | | | | | | | | | | | | | | |

50

## 2.3.2 Genotypic confirmation of Clinical Isolates

In this study, a total of 55 clinical isolates and 3 laboratory strains were included to investigate the ESX genetic diversity. I carefully selected these strains to represent two important groups within KwaZulu-Natal. The first was the Beijing lineage that represents approximately 20 percent of all circulating strains in KwaZulu-Natal and the second a group of closely related strains known as the F15/LAM4/KZN lineage that was associated with a large outbreak of drug resistant tuberculosis (Pillay and Sturm, 2007). I reasoned that successful strains, implied by their ecological abundance, would be involved in ongoing cycles of transmission in the local population and would therefore be most likely to have been under the influence of immune selection.

The clinical isolates originated from Tugela Ferry; Rapid Detection Study *M. vaccae* clinical trial and Westville Prison Molecular Epidemiology Study, conducted at the department of Infection, Prevention and Control (UKZN) to follow the transmission of *Mtb* strains in host populations. The IS6110 RFLP typing method, the gold standard of typing methods for *Mtb* (van Embden *et. al*., 1993), was used to confirm the genotypes of the clinical isolates and laboratory strains and dendograms were generated on the band placement for the isolates (Figures 2.1-2.3). Hawkey and colleagues reported on the difficulty of this method in sizing up the DNA fragments, since one cannot distinguish different sized fragments appropriately. In addition, software used in the analysis may not necessarily be accurate in the placement of the bands at the correct positions for the strain comparisons (Hawkey *et. al.,* 2003). With this typing method, all the Bejing isolates displayed the expected Beijing RFLP pattern (Figure 2.1). Gagneux and colleagues screened unique large sequence polymorphisms (LSPs) in Beijing strains and reported those strains to be a monophyletic clade based on the RD105 deletion (Gagneux *et. al.,* 2006), and these strains were confirmed to have RD105 deletions.

The isolates grouped as KZN (Figure 2.2) displayed the ST60 spoligotype signature listed on the online spoligotype database (Pillay and Sturm, 2007), except for BCG that had two unique bands that was not in agreement with the KZN patterns. In addition, the remaining 8 clinical isolates (Figure 2.3) lacked any signatures similar to a Beijing or KZN pattern,

and were denoted as the "Other" grouping. The pattern observed in Figure 2.2 is a typical pattern unique to the F15 family that forms part of the Latino-American and Mediterranean family (LAM) and corresponds to the LAM4 subgroup, thus being named the F15/LAM4/KZN strain (Pillay and Sturm, 2007).
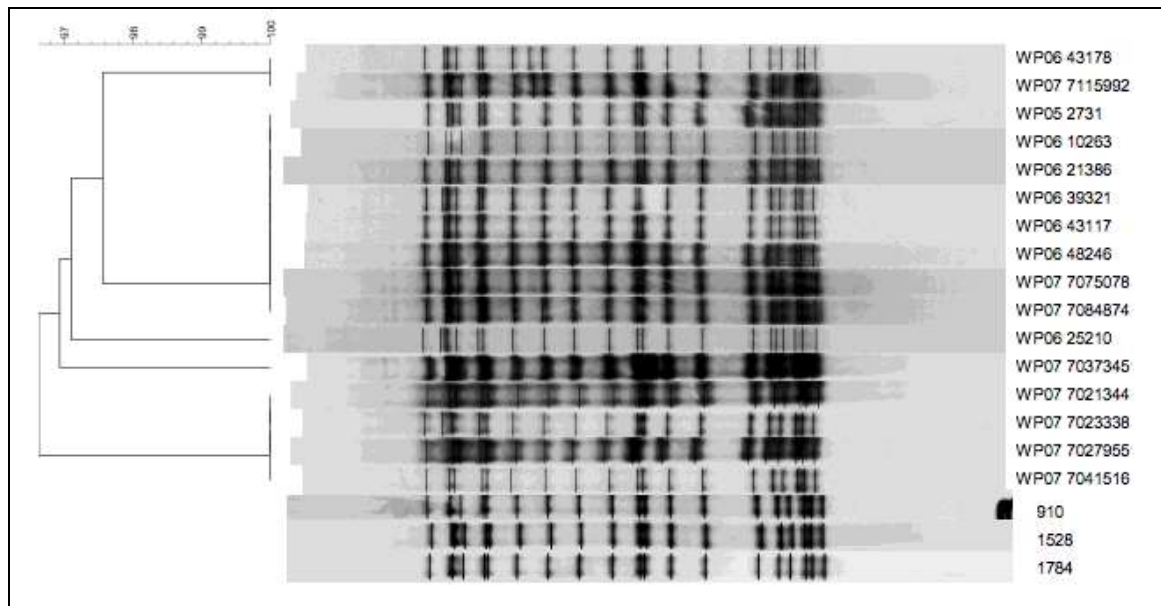


**Figure 2.1** IS6110 RFLP pattern generated for clinical isolates corresponding to the Beijing genotype.
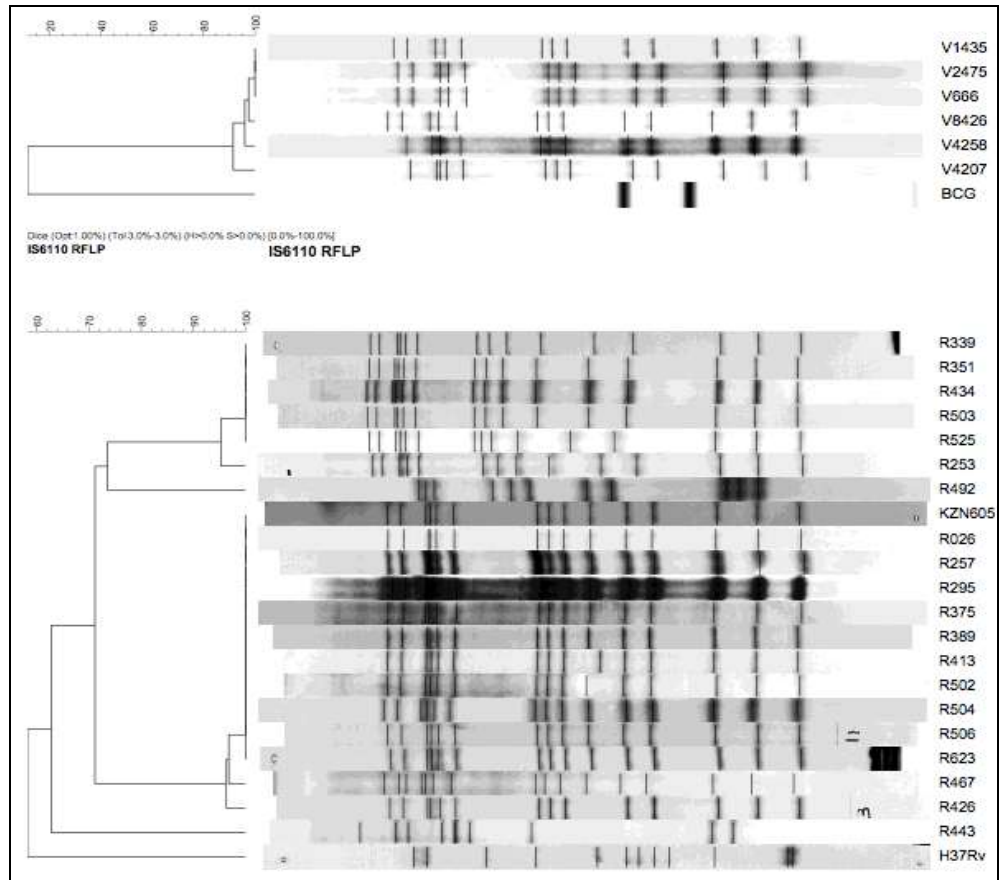
**Figure 2.2** IS6110 RFLP pattern generated for clinical isolates corresponding to the F15/LAM4/KZN strain genotype.
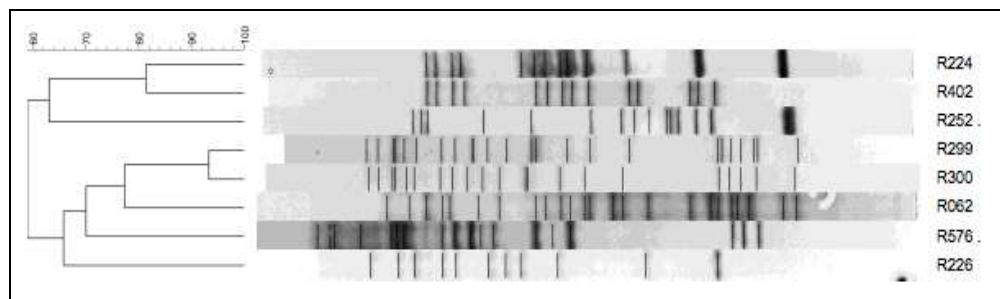


**Figure 2.3** IS6110 RFLP pattern generated for clinical isolates corresponding to neither Beijing or F15/LAM4/KZN strain genotypes.

An additional confirmation of the isolates genotypes was performed. In Figure 2.4a, a representation of the isolates were selected and tested positive for the *helZ* and *fadE22* deletions, indicated by a smaller amplicon size for a KZN strain relative to the non-KZN H37Rv strain. These two deletions are unique to the KZN strains (Pillay, 2010). The isolates were then subjected to a multiplex spoligotype PCR assay to verify whether those positive results were in fact true positives for the deletions detected. With this test (Figure 2.4b), the presence of two bands confirms the positive test for KZN strains, whereas the Beijing isolates displayed a single band that was larger than 1031 bp. However, isolate 8426 displayed two bands in Figure 2.4a, and was not concordant to the spoligotype pattern (Figure 2.2) for a KZN strain. However, after being subjected to the multiplex spoligotype PCR assay (Figure 2.4b), it was confirmed to be a KZN genotype. The characteristic pattern for the KZN strain shows that the strain has the spacer 20 but is missing the spacers 21-24 and spacer 40. Any deviation from this pattern indicated that the isolate was not of the KZN strain (Pillay, 2010).

The remaining isolates were tested using both assays and their genotypes were successfully confirmed and correlated to the spoligotype patterns (results not shown).
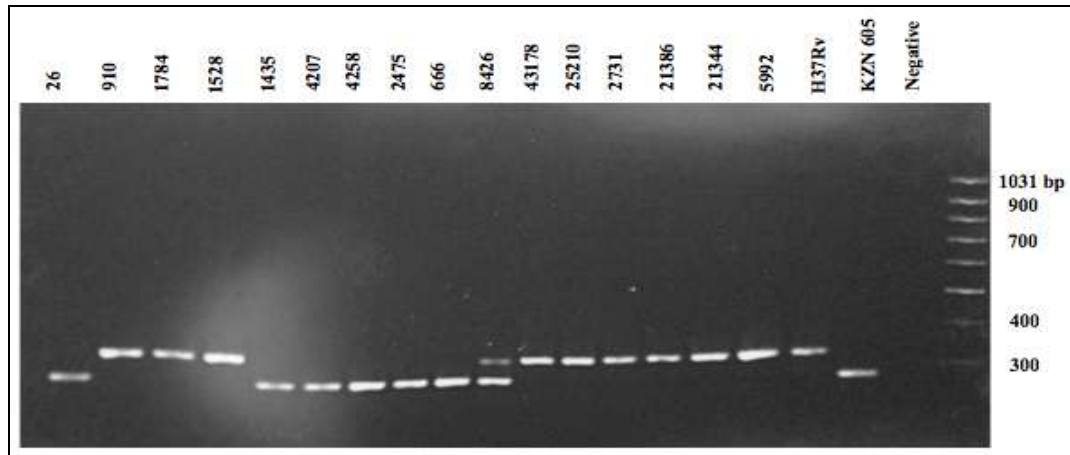
**(a)**
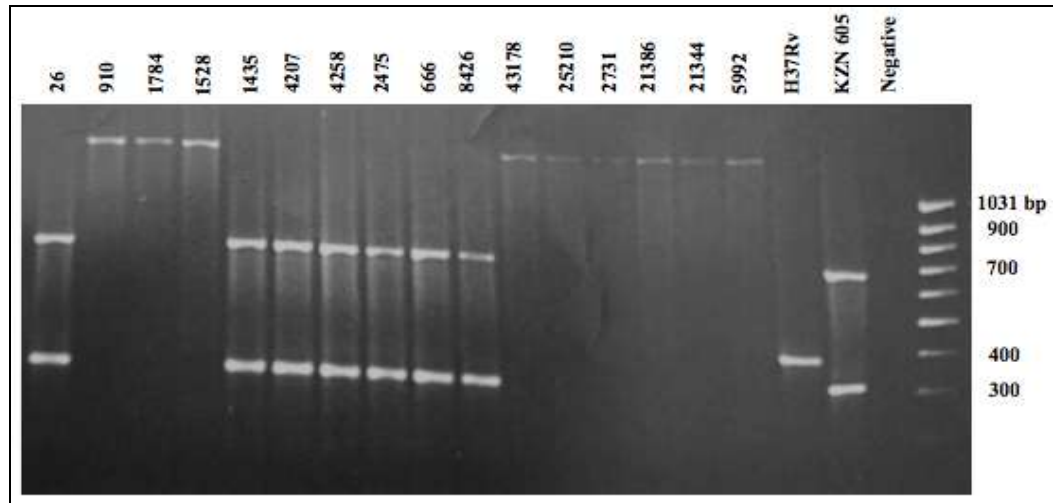


**(b)**

**Figure 2.4** Agarose gel of amplified DNA for clinical isolates using a strain specific PCR that selectively amplifies specific DR region spacer sequences. Agarose gel **(a)** shows the amplified DNA for KZN and non-KZN strains. KZN strains have a band close to 266 bp that is similar to the KZN 605 positive control. Agarose gel **(b)** is the verification assay for false positive strains. KZN strains have a similar amplification pattern to the KZN 605 positive control.

### 2.3.4 PCR Amplification of ESX genes

The PCR primers for the ESX genes were initially optimised with the H37Rv and BCG genomic DNA (Fig 2.5). PCR amplification was successful as each gene pair was amplified at the expected band size (Table 2.5). The lack of a PCR product for *esxA, esxB, esxOP* and *esxVW* in BCG was expected, as it is well documented that these genes are absent in the organism (Mahairas *et. al.,* 1996). This result is indicative of the specificity in the primer design for the amplification of the genes of interest. As a result, these primer sets were applied to amplify the ESX genes in genomic DNA extracted from the 55 clinical isolates (results not shown) and used in subsequent sequencing reactions.
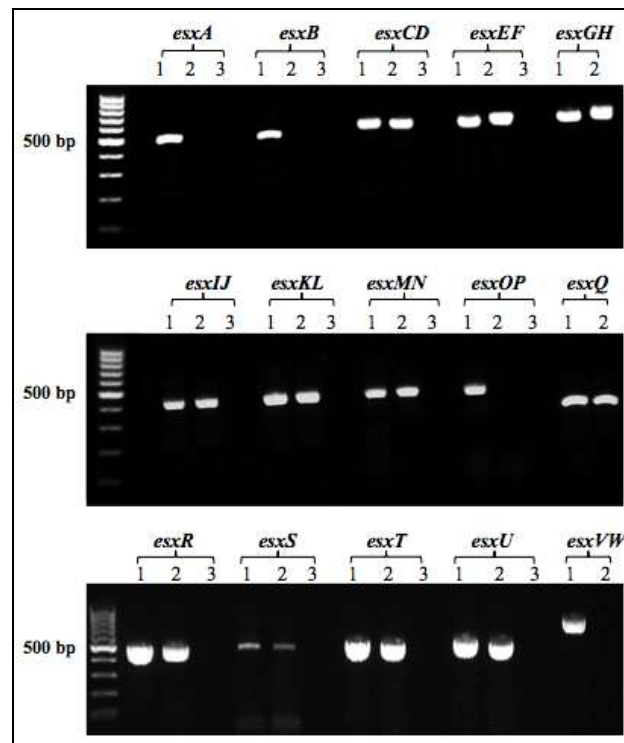


**Figure 2.5**    PCR amplification of ESX genes and gene pairs on a 1.5% (w/v) agarose gel with H37Rv (1) and BCG (2) genomic DNA and no DNA control (3). A 100bp molecular weight marker was run on the gel to verify amplified products.

**Table 2.5**     Expected PCR product sizes for ESX gene pairs.

| ESX Genes | Expected Molecular Weight in bp (base pairs) |
|-----------|:---:|
| *esxA* | 510 bp |
| *esxB* | 525 bp |
| *esxCD* | 610 bp |
| *esxEF* | 776 bp |
| *esxGH* | 850 bp |
| *esxIJ* | 644 bp |
| *esxKL* | 670 bp |
| *esxMN* | 725 bp |
| *esxOP* | 764 bp |
| *esxQ* | 551 bp |
| *esxR* | 500 bp |
| *esxS* | 584 bp |
| *esxT* | 510 bp |
| *esxU* | 520 bp |
| *esxVW* | 625 bp |

**2.3.5 Clustering of the 23 ESX Sequences from H37Rv Reference Strain**

A phylogenetic tree was constructed using the Geneious software (Drummond *et. al.*, 2012) from the H37Rv ESX sequences (Figure 2.6). The three ESX subfamilies are represented within the tree and cluster into distinctive clades as a consequence of their high sequence identity between members. However, within the Mtb9.9 subfamily, the *esxI* and *esxV* sequences lack any SNPs thus rendering those sequences to be identical. This was further confirmed upon alignment of both the nucleotide and protein sequences (Figure 2.7). However, these genes could be amplified independently because of unique flanking sequences.
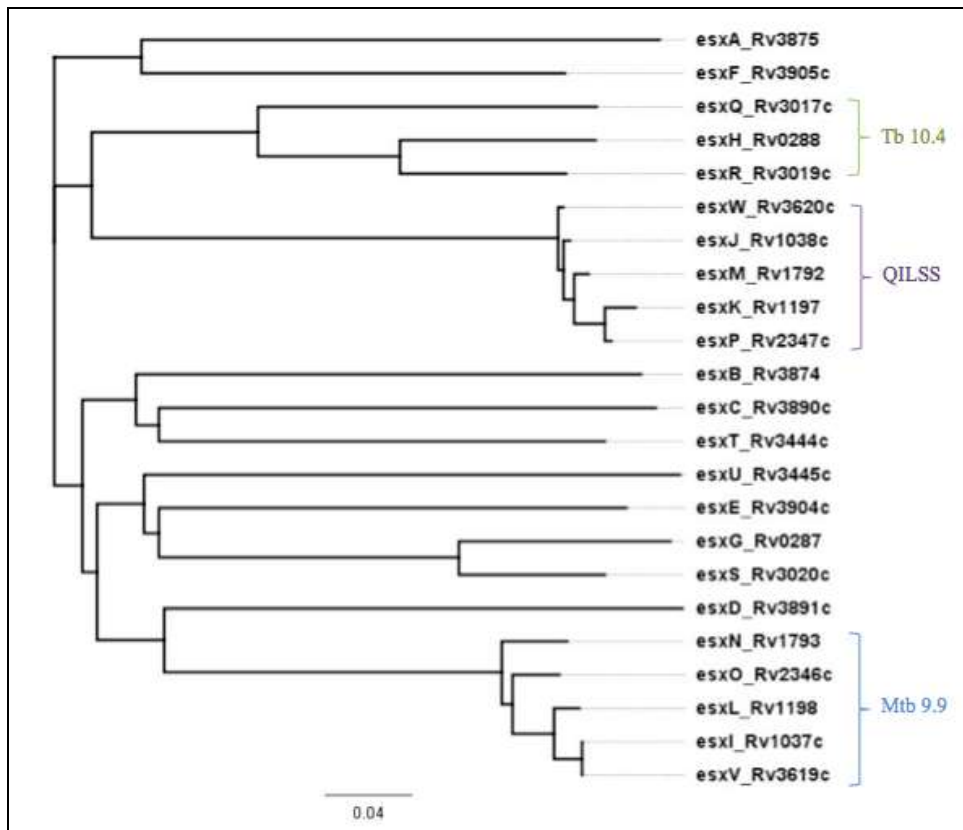


**Figure 2.6**   Phylogenetic tree showing the clustering of all 23 ESX members and subfamilies in H7Rv using Geneious software (Drummond *et. al.,* 2012).

**Figure 2.7** Nucleotide and protein sequence alignment of *esxI* and *esxV* sequences from H37Rv using the DNAStar MegAlign software.

### 2.3.6 ESX Sequence Diversity within Clinical Isolates

Comparative sequence analysis was conducted by screening all the ESX sequences for variation using the MegAlign software, with the H37Rv ESX sequences as a reference. All 23 ESX genes were successfully sequenced in the 55 clinical isolates as well as the 3 laboratory strains (H37Rv, H37Ra and *M. bovis BCG*). Polymorphisms relative to H37Rv were located in 12 of the 23 ESX genes (Figure 2.8, Tables 2.6 and 2.7) with the majority of the polymorphisms occurring in the Beijing isolates (Figure 2.9). The number of nsSNPs was three and a half times more than the number of sSNPs for the Beijing isolates (Figure 2.10). Mestre and colleagues analysed polymorphisms in the DNA repair, replication and recombination (3R) genes from a collection of 305 Bejing isolates. Interestingly, they reported that the number of nsSNPs was twice the number of the sSNPs from their Beijing dataset (Mestre *et. al.,* 2011).
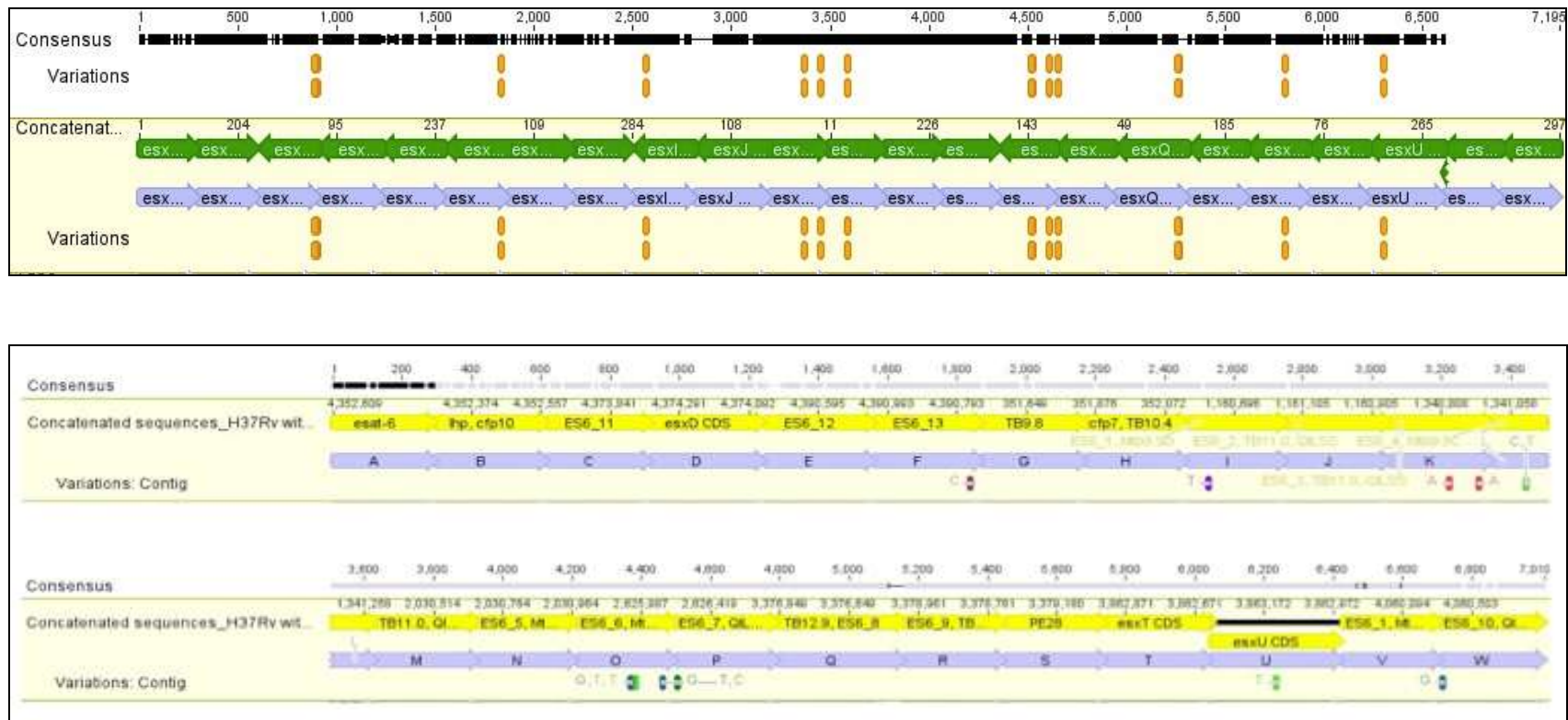
**Figure 2.8** Snapshot of clustal alignment of concatenated sequences using Geneious software of all the SNPs identified in the 23 ESX genes sequences for the 55 clinical isolates and 3 laboratory strains (blue arrows). The sequences were aligned to the concatenated H37Rv reference sequence (yellow and green arrows with coding sequence annotations) and aligned using the ClustalW software available on the Geneious package (Drummond *et. al.,* 2012).
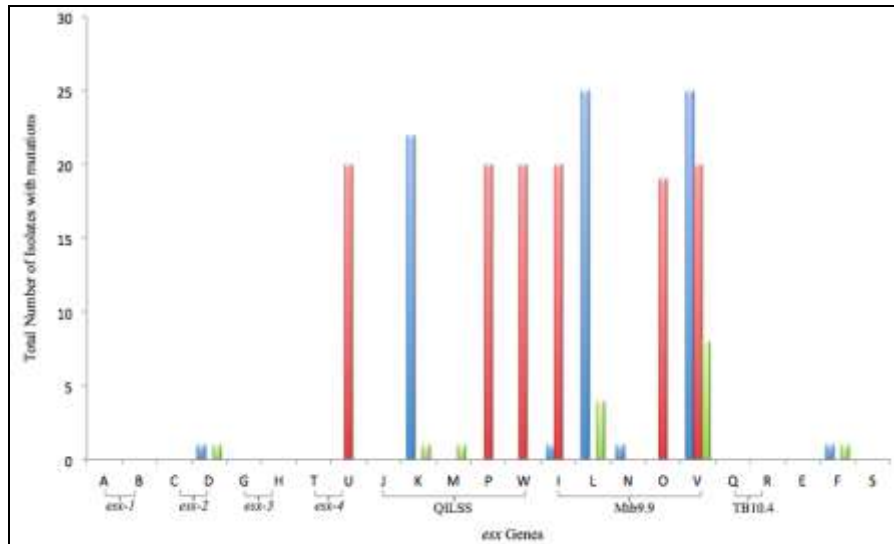
**Figure 2.9**    Distribution of SNPs relative to H37Rv for the 23 ESX genes and subfamilies seen across the three isolate groupings. Blue bars: KZN Isolates, Red bars: Beijing Isolates and Green bars: Other Isolate groupings.
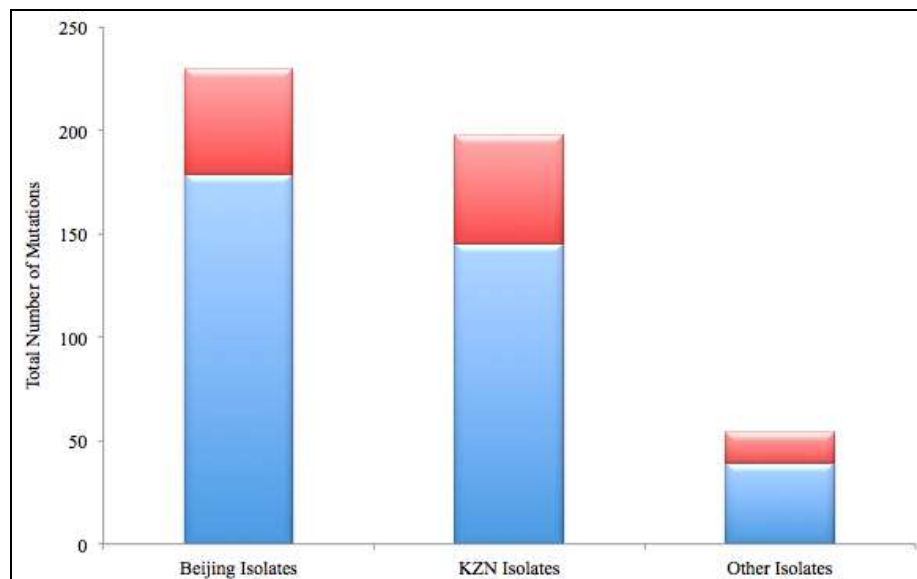


**Figure 2.10**    Total numbers of Non-Synonymous (blue) and Synonymous (red) SNPs across the three isolate groupings: 20 Beijing Isolates; 25 KZN Isolates and 13 Other Isolates.

A total of 482 SNPs were identified, of which 363 were nsSNPs and 119 were sSNPS. A closer inspection of the nsSNPs distribution amongst the isolate groupings revealed 179 occurred in the Beijing isolates with 145 in the KZN isolates and 39 in the Other isolate grouping (Figure 2.10). Similarly, the distribution of the sSNPs amongst the isolate groupings revealed 51 occurred in the Beijing isolates with 53 in the KZN isolates and 15 in the Other isolate grouping (Figure 3.10). No mutations occurred within 10 ESX genes (*esxA, B, C, E, G, H, J, R, S* and *T*). Similarly a study conducted by Musser and colleagues involving the sequencing of 24 *Mtb* antigens from 16 clinical isolates resulted in no variation for *esxA* and *esxB* (Musser *et. al*., 2000). These findings are also in agreement with the 2010 study conducted by Davila and colleagues where no sequence variation was observed for the *esxA* and *esxH* genes sequenced from 88 clinical isolates (Davila *et. al.,* 2010). Despite the report of SNPs in *esxH* (Comas *et. al.,* 2010 and Uplekar *et. al.,* 2011), this was not the case in my data. Based on my data, the lack of SNPs in *esxA, B, C, E, G, H, J, R, S* and *T* genes suggests that these gene regions are conserved among different genetic groups of *Mtb* analyzed in this study.

Overall, the clinical isolates harboured more than one non-synonymous mutation in *esxI, K, L, O, P, V* and *W* relative to H37Rv (Table 2.8). Four SNPs associated with the KZN strain grouping, with 3 being nsSNPs and the other a sSNP, respectively. Of the 3 nsSNPs (Table 2.8), one occurred in *esxK* (A58T) and 2 in *esxL* (R33S) in 22 KZN isolates, whereas the sSNP occurred in *esxK* (E86) for 17 KZN isolates (Table 2.9). Similarly, 9 unique SNPs were identified only in the Beijing isolates, of which 6 were nsSNPs (Table 2.8) and 3 were sSNPs (Table 2.9). The nsSNPs included *esxI* (Q20L), *esxO* (E52G), 2 in *esxP* (T3S; N83D), *exU* (P63S) and *esxW* (T2A). Of the 3 remaining sSNPs, 2 occurred in *esxO* (I54; L57) and 1 in *esxP* (A2), respectively.

5 SNPs occurred in isolates belonging to the 3 different lineages of the data set. Four SNPs occurred in *esxV*, which included 3 nsSNPs (Q20L, S23L and A57V) and 1 sSNP (H13), while the remaining SNP, being synonymous, occurred in *esxF* (S93) (Tables 2.8 and 2.9). Interestingly, the SNPs identified for *esxV,* occurred in almost all the isolates for

the dataset. The Q20L, S23L nsSNPs identified in my dataset also confirmed a similar finding in the dataset of Uplekar and colleagues (Uplekar *et. al*., 2011).

Interestingly, two nsSNPs, one in *esxF* (W58stop) and one in *esxD* (T49A), occurred only in a *M. bovis BCG* laboratory strain and the KZN 8426 isolate. Upon comparison to the genome sequences of *M. bovis* BCG Pasteur and *M. bovis* AF2122/97, these particular mutations were confirmed to be *M. bovis* specific. Nonetheless, isolate 8426 was confirmed to be a KZN strain by spoligotype (Figure 2.2) and the genotypic test (Figure 2.5b) and no further mutations within the ESX family for this isolate was recognised to be similar to *M. bovis* BCG. Although the PCR and sequencing was replicated and repeated independently, the mutation for this isolate remained unchanged.

Additionally, 5 other SNPs in the BCG laboratory strain were validated as *M. bovis* specific. These included 2 nsSNPs in *esxM* (M48Q) and 3 sSNPs, of which 2 occurred in *esxK* (S3 and N39) and 1 in *esxM* (G47), respectively.

The *esxI* gene had 2 nsSNPs (L55F and A88G) for isolate 426, which was not observed in the other isolates. Similarly, isolate 389 showed a single sSNP in *esxN* (V90) that was not present in the other isolates. Both isolates 389 and 429 were confirmed to represent the KZN grouping in my dataset.

SNPs encoding stop codons have substantial impacts on the structure and functionality of proteins. *esxM* is the only other gene that possesses a stop codon (stop59Q) which occurred in the *BCG* laboratory strain and was not present in any of the other strains. This result was contrary to the findings by Uplekar and colleagues in which this mutation was accounted for in 9 clinical isolates (Uplekar *et. al*., 2011).
Overall the 55 clinical isolates used in my study were approxiamtely half the 108 isolates used in the previous study conducted by Uplekar and colleagues. Taking into account this difference the absolute number of SNPs identified in this study is compatible with the numbers identified in the study of Uplekar and colleagues (Uplekar *et. al*., 2011).

### 2.3.7 SNP Diversity across ESX gene sub-families

In Figure 2.5, isolates harbouring SNPs were grouped according to the ESX gene subfamilies to establish the SNP distribution and diversity across them. Overall, genes encoded within ESX-1 to ESX-4 presented with low levels of SNP variation except for *esxD* and *esxU*. Majority of the SNPs occur in genes constituting the ESX-5, Mtb9.9 and QILSS subfamilies (Figure 2.9). Intriguingly, the co-occurrence of 2 nsSNPs was noticed in the neighbouring positions 97 and 98 of *esxL* (Tables 2.6 and 2.8). This particular trait was also observed in *esxM* of the BCG laboratory strain containing a single sSNP in codon 47 followed by 2 nsSNPs in codon 48 (Tables 2.6 to 2.9). In my dataset, the SNPs reported for *esxV* are highly prevalent in over 90% of the clinical isolates. The remainder of nucleotide sequences for the ESX-5, Mtb9.9 and QILSS subfamilies were identical upon comparison to the *Mtb* H37Rv sequences.

### 2.3.8 ESX Phylogenetic and Evolutionary Analysis of the Clinical Isolates

The information from the spoligotype patterns and the genotypic confirmation assay were inspected closely to correlate to lineage specificity based on the SNPs identified in the sequence dataset (Tables 2.6-2.9). As expected from the topology of the phylogenetic tree, two distinct clades were observed in Figure 2.11, based on the concatenated sequences containing all the SNPs identified in this study.

Most mutations can decrease the function of the protein and can eventually be eliminated from the population through a negative or purifying selection. Nevertheless, in other instances, strains with a higher mutation rate may have a selective advantage under certain conditions (Mestre *et. al.,* 2011). In extremely unusual circumstances, a mutation can be beneficial and is as a consequence fixed into the population by positive or diversifying selection (Filliol *et. al.,* 2006; Gutacker *et. al*., 2006). To infer the type of evolution occurring in this data set, the dN/dS ratio was used to measure the type of selection in operation on the ESX genes. In order to test the hypothesis that positive selection was in operation as these sequences diverged, the codon-based Z test was used on all 58 concatenated codon alignments. This yielded a value of 0.328. It was noted that my dataset harboured majority of the SNPs in the Mtb9.9 and QILSS genes (Figure 2.9), with sSNPs occurring in a larger proportion in the clinical isolates compared to the

nsSNP (Figure 2.10), thus influencing the low dN/dS value. The analysis was repeated, but excluded the Mtb9.9 and QILSS subfamily of genes resulting in a value of 1.307 (dN/dS >1) indicating positive selection for these genes. My findings replicate the results reported by Uplekar and colleagues. (Uplekar *et. al.,* 2011). A recent publication by Comas and colleagues investigated the antigenic variation and diversity of human T-cell epitopes amongst 21 strains representative of the 6 major lineages of the MTBC. Using the Illumina sequencer, whole genome sequences were generated from the 21 strains from which more than 9000 unique SNPs were identified. Their analysis of the sequences revealed that the T-cell epitopes showed little sequence variation and estimated low dN/dS ratio values for changes observed in essential genes to that of the non-essential genes (Comas *et. al.,* 2010).
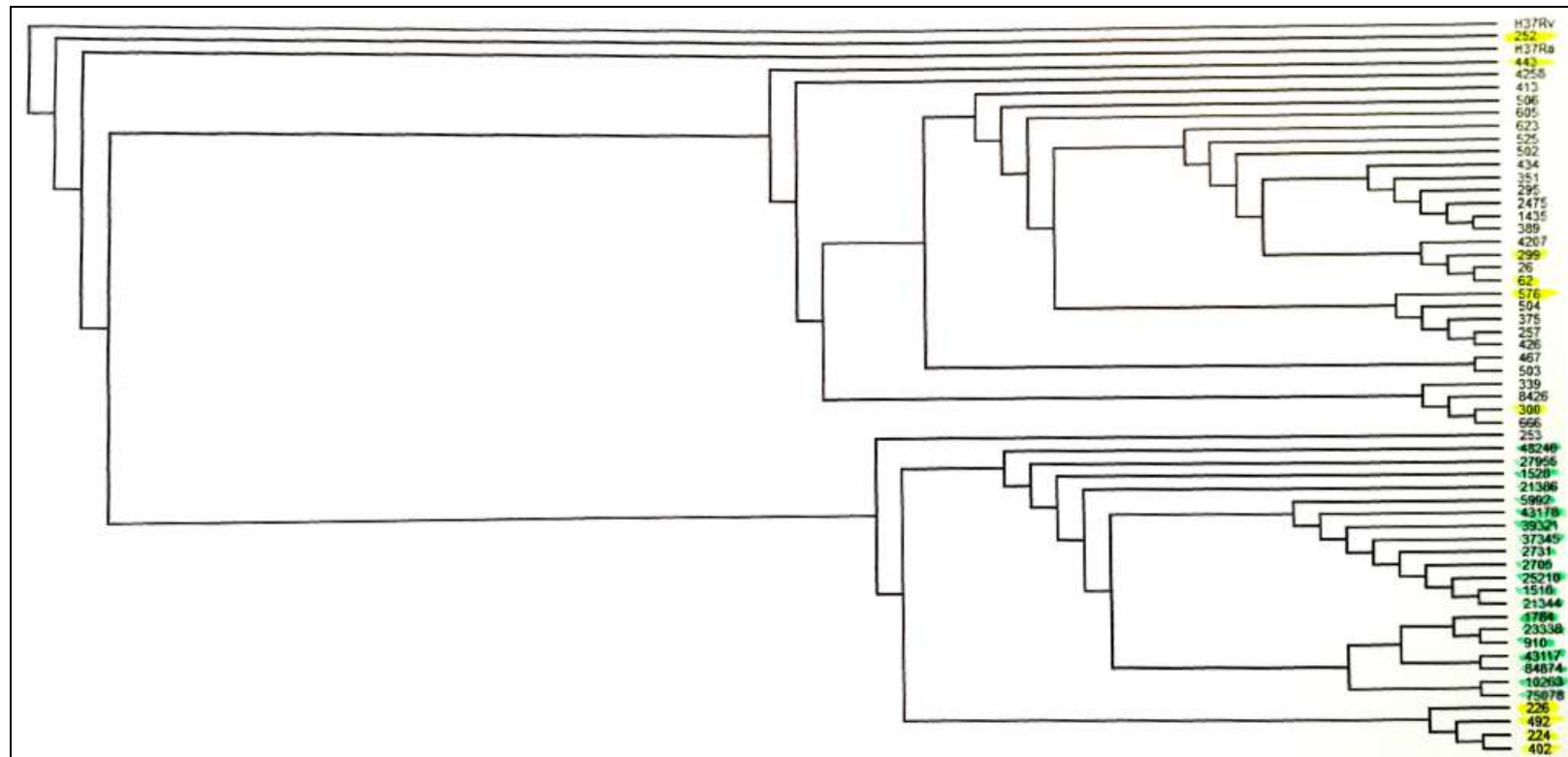
**Figure 2.11** Phylogenetic tree of concatenated sequences using the DNAStar TreeView software rooted to H37Rv. Highlighted in Green: the Beijing isolates, Yellow: the Other Isolate grouping, and unhighlighted sequences: the KZN grouping of isolates. The concatenated BCG sequences was not included as part of the phylogenetic tree.

**Table 2.6**     Table of Non-Synonymous Mutations Identified from the 58 sequenced Clinical Isolates.

| esx Cluster | esx Genes | Number of Non-Synonymous Mutations | | | Non-Synonymous Mutations | | |
|---|---|---|---|---|---|---|---|
| | | Total Clinical Isolates Sequenced | Ref. Strain (H37Rv) | Clinical Isolates | Isolates | Mutation | Number of Isolates Mutation Occurs in: |
| esx-1 | A | 58 | - | - | - | - | - |
| | B | 58 | - | - | - | - | - |
| esx-2 | C | 58 | - | - | - | - | - |
| | D | 58 | - | 1 | 40; 58 | T49A (a145g) | 2 |
| | E | 58 | - | - | - | - | - |
| | F | 58 | - | 1 | 40; 58 | W58Stop (g173a) | 2 |
| esx-3 | G | 58 | - | - | - | - | - |
| | H | 58 | - | - | - | - | - |
| | I | 58 | - | 1 | 31-33; 41-58 | Q20L (a59t) | 21 |
| | J | 58 | - | - | - | - | - |
| | K | 58 | - | 1 | 32 | A58P (g172c) | 1 |
| | | | | 1 | 2; 4; 8-14; 16; 20; 22-28; 34; 36; 38-40 | A58T (g172a) | 25 |
| | L | 58 | - | 1 | 4 | L23S (t68c) | 1 |
| | | | | 2 | 2; 8-14; 16; 18-20; 22-28; 30; 34; 38-40 | R33S (c97t; g98c) | 24 |
| esx-5 | M | 58 | - | - | - | - | - |
| | N | 58 | - | - | - | - | - |
| | O | 58 | - | 1 | 31-33; 41-57 | E52G (a155g) | 20 |
| | P | 58 | - | 1 | 31-33; 41-57 | T3S (a7t) | 20 |
| | Q | 58 | - | - | - | - | - |
| | R | 58 | - | - | - | - | - |
| | S | 58 | - | - | - | - | - |
| esx-4 | T | 58 | - | - | - | - | - |
| | U | 58 | - | 1 | 31-33; 41-57 | P63S (c187t) | 20 |
| | V | 58 | - | 1 | 1-28; 31-57 | Q20L (a59t) | 56 |
| | | | | 1 | 1-28; 31-57 | S23L (c68t) | 56 |
| | | | | 1 | 1-28; 31-57 | A57V(c170t) | 56 |
| | W | 58 | - | 1 | 31-33; 41-57 | T2A (a4g) | 20 |

**Table 2.7**     Table of Synonymous Mutations Identified from the 58 sequenced Clinical Isolates.

| | | | Number of Synonymous Mutations | | Synonymous Mutations | | |
|---|---|---|---|---|---|---|---|
| *esx* Cluster: | *esx* Genes | Total Clinical Isolates Sequenced | Ref. Strain (H37Rv) | Clinical Isolates | Isolates: | Mutation: | Number of Isolates Mutation Occurs in: |
| esx-1 | A | 58 | - | - | - | - | - |
| | B | 58 | - | - | - | - | - |
| esx-2 | C | 58 | - | - | - | - | - |
| | D | 58 | - | - | - | - | - |
| | E | 58 | - | - | - | - | - |
| | F | 58 | - | 1 | 1-2; 8-14; 16-19; 22; 26; 34-39 | S93S (g279c) | 21 |
| esx-3 | G | 58 | - | - | - | - | - |
| | H | 58 | - | - | - | - | - |
| | I | 58 | - | - | - | - | - |
| | J | 58 | - | - | - | - | - |
| | K | 58 | - | 1 | 3; 21; 58 | S3S (a9g) | 3 |
| | | | | 1 | 21 | H10H (c30t) | 1 |
| | | | | 1 | 58 | N39N (c117t) | 1 |
| | | | | 3 | 4; 21 | I40I (t120c); S41S (c123g); G42G (t126c) | 2 |
| | L | 58 | - | - | - | - | - |
| esx-5 | M | 58 | - | - | - | - | - |
| | N | 58 | - | - | - | - | - |
| | O | 58 | - | 3 | 31-33; 41-57 | I54I (c162t); L57L (c169t); A83A (a249g) | 20 |
| | P | 58 | - | 1 | 31- 33; 41-57 | A2A (a6c) | 20 |
| | Q | 58 | - | - | - | - | - |
| | R | 58 | - | - | - | - | - |
| | S | 58 | - | - | - | - | - |
| esx-4 | T | 58 | - | - | - | - | - |
| | U | 58 | - | - | - | - | - |
| | V | 58 | - | 1 | 1-28; 31-57 | H13H (c39t) | 56 |
| | W | 58 | - | - | - | - | - |

**Table 2.8**     Binary representation for Non-synonymous SNPs identified in the 23 ESX genes across 55 clinical isolates and the 3 laboratory strains. Presence of the mutation is represented by 1 and no mutation represented by 0. The dataset is a color coded representation of isolate groupings. Green: Beijing Isolates, Blue: KZN Isolates and Yellow: Other Isolate grouping.

| Nucleotide | | | | | | a145g | | g173a | | | a59t | a163c | c261g | | g172a | c97t | g98t | a142c | t143a | t175c | | a155g | a7t | a247g | | | | | c187t | a59t | c68t | c170t | a4g |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Protein | | | | | | T49A | | W58Stop | | | Q20L | L55F | A88G | | A58T | R33S | R33S | M48Q | M48Q | Stop59Q | | E52G | T3S | N83D | | | | | P63S | Q20L | S23L | A57V | T2A |
| RFLP / Strains: | Isolates: | Isolate Number: | esxA | esxB | esxC | esxD | esxE | esxF | esxG | esxH | esxI | esxJ | esxI | esxJ | esxK | esxL | esxL | esxM | esxM | esxN | esxN | esxO | esxP | esxP | esxQ | esxR | esx5 | esxT | esxU | esxV | esxV | esxV | esxW |
| Beijing | 910 | 31 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| Beijing | 1784 | 32 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| Beijing | 1528 | 33 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| Beijing | 43178 | 41 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| Beijing | 25210 | 42 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| Beijing | 10263 | 43 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| Beijing | 2705 | 44 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| Beijing | 39321 | 45 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| Beijing | 43117 | 46 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| Beijing | 48246 | 47 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| Beijing | 21386 | 48 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| Beijing | 7021344 | 49 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| Beijing | 7027955 | 50 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| Beijing | 7037345 | 51 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| Beijing | 7075078 | 52 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| Beijing | 7084874 | 53 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| Beijing | 7023338 | 54 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| Beijing | 7041516 | 55 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| Beijing | 2731 | 56 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| Beijing | 7115992 | 57 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| KZN | 26 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| KZN | 253 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| KZN | 257 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| KZN | 295 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| KZN | 339 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| KZN | 351 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| KZN | 375 | 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| KZN | 389 | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| KZN | 413 | 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| KZN | 426 | 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| KZN | 434 | 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| KZN | 467 | 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| KZN | 502 | 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| KZN | 503 | 23 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| KZN | 504 | 24 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| KZN | 506 | 25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| KZN | 525 | 26 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| KZN | 623 | 28 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| KZN | 1435 | 34 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| KZN | 4207 | 35 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| KZN | 4258 | 36 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| KZN | 2475 | 37 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| KZN | 605 | 38 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| KZN | 666 | 39 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| KZN | 8426 | 40 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| Other | 62 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| Other | 224 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| Other | 226 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| Other | 252 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Other | 299 | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| Other | 300 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| Other | 402 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 |
| Other | 443 | 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| Other | 492 | 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 |
| Other | 576 | 27 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| Reference Strain | H37Rv | 29 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Avirulent Strain | H37Ra | 30 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Reference Strain | BCG | 58 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Table 2.9**      Binary representation for Synonymous SNPs identified in the 23 ESX genes across 55 clinical isolates and the 3 laboratory strains. Presence of the mutation is represented by 1 and no mutation represented by 0. The dataset is color coded according to isolate grouping. Green represents Beijing Isolates, blue representing KZN Isolates and yellow for the Other Isolate grouping.

| Nucleotide | | | | | | | | g279c | | | | | a9g | c117t | g258a | | c141t | c267g | c162t | c169t | a6c | c30t | c168t | | | | | c39t | |
| Protein | | | | | | | | S93 | | | | | S3 | N39 | E86 | | G47 | V90 | I54 | L57 | A2 | H10 | A57 | | | | | H13 | |
| RFLP / Strains | Isolates | Isolate Number | esxA | esxB | esxC | esxD | esxE | esxF | esxG | esxH | esxI | esxJ | esxK | esxK | esxK | esxL | esxM | esxN | esxO | esxO | esxP | esxP | esxQ | esxR | esxS | esxT | esxU | esxV | esxW |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Beijing | 910 P4 | 31 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Beijing | 1784 P5 | 32 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Beijing | 1528 P8 | 33 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Beijing | 43178 | 41 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Beijing | WPO6 25210 | 42 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Beijing | WPO6 10263 | 43 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Beijing | WPO6 2705 | 44 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Beijing | 39321 | 45 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Beijing | 43117 | 46 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Beijing | 48246 | 47 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Beijing | WPO6 21386 | 48 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Beijing | WPO7 7021344 | 49 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Beijing | WPO7 7027955 | 50 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Beijing | WPO7 7037345 | 51 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Beijing | WPO7 7075078 | 52 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Beijing | WPO7 7084874 | 53 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Beijing | WPO7 7023338 | 54 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Beijing | WPO7 7041516 | 55 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Beijing | WPO6 2731 | 56 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Beijing | WPO7 7115992 | 57 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| KZN variant 1 | R 26 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| KZN variant 1 | R 253 | 6 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| KZN variant 1 | R 257 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| KZN variant 1 | R 295 | 8 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| KZN variant 1 | R 339 | 11 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| KZN variant 1 | R 351 | 12 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| KZN variant 1 | R 375 | 13 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| KZN variant 1 | R 389 | 14 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| KZN variant 1 | R 413 | 16 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| KZN Archetype | R 426 | 17 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| KZN variant 1 | R 434 | 18 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| KZN variant 1 | R 467 | 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| KZN variant 1 | R 502 | 22 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| KZN variant 1 | R 503 | 23 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| KZN variant 1 | R 504 | 24 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| KZN variant 1 | R 506 | 25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| KZN variant 1 | R 525 | 26 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| KZN variant 1 | R 623 | 28 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| KZN variant 1 | Vac 1435 | 34 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| KZN variant 1 | Vac 4207 | 35 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| KZN variant 1 | Vac 4258 | 36 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| KZN variant 1 | Vac 2475 | 37 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| KZN variant 1 | KZN 605 | 38 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| KZN variant 1 | Vac 666 | 39 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| KZN variant 1 | Vac 8426 | 40 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Neither | R 62 | 2 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Neither | R 224 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Neither | R 226 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Neither | R 252 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Neither | R 299 | 9 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Neither | R 300 | 10 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Neither | R 402 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| F28 | R 443 | 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| F28 | R 492 | 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Neither | R 576 | 27 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Reference Strain | H37Rv | 29 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Avirulent Strain | H37Ra | 30 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Reference Strain | BCG | 58 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**2.3.9 ESX Sequence Diversity of Clinical Isolates Using Full Genome Sequencing**

With more recent advances in genome sequencing it was possible to augment my dataset with ESX sequence data from 130 full genome sequences of clinical isolates from KZN. These were sequenced as part of the ongoing collaboration between KwaZulu-Natal Research Institute for Tuberculosis and HIV/AIDS (K-RITH) and The Broad Institute. These isolates were from two sources. Ninety strains were collected from 20 sites across the province of KZN as part of a drug susceptibility study, and 80 of these isolates were at least resistant to one first-line anti-tuberculosis drug. The second study collected isolates from a single site in Durban, King Dinizulu hospital, which is the centralized referral centre for all drug-resistant tuberculosis in KZN. The majority of these isolates were resistant to at least one drug. Full genome sequencing using the Illumina platform was carried out on two libraries, a short fragment library and a jumping library and all SNPs were identified using BROAD developed variant calling software. A total of 10,781 SNP positions, synonymous and non-synonymous, were identified and used to generate a phylogeny shown in Figure 2.12. nsSNP Mutations in the 23 ESX genes were identified and have been plotted against each strain in the peacock graph shown in Figure 2.12. The comparative analysis was performed relative to the H37Rv sequence.

The phylogeny was able to resolve the Beijing and KZN families. In Figure 2.12 the KZN strains are labeled as LAM4. In terms of the distribution of nsSNPs there were differences between the amplicon and genome sequenced clinical isolates. There was no overlap between these two collections of strains in that no strains were sequenced by both methods. In the amplicon sequenced group of strains no nSNPs were found in 12 genes (*esx A, B, C, E, G, H, J, N, Q, R, S, T*) (Table 2.6 and 2.8). In the genome sequenced group 5 *esx* genes (*esxA, G, H, J, Q, S* and *T*) had no nsSNPs (Figure 2.12). The differences can be accounted for by 7 nsSNPs in 5 genes (Figure 2.12) that occurred in strains other than the KZN or Beijing and reflect a broader sample size in the full genome sequenced group.

Similiarly an excellent correlation was observed for the nsSNPs in the Beijing and KZN strains sequenced by both methods. In Figure 2.12, the largest number of polymorphisms

71

occurred in the Beijing clade, probably in part due to the phylogenic distance from the reference H37Rv used in this analysis. Unique nsSNPs, for the Beijing and KZN isolates that were identified earlier from my dataset, also occur in Figure 2.12.

The unique nsSNPs for the Beijing isolates that were found by both sequencing methods were: *esxI* (Q20L); e*sxO* (E52G); e*sxP* (T3S); *esxU* (P43S) and *esxW* (T2A). In the previous amplicon sequencing description above I annotated the *esxU* mutation at codon 63, but subsequent to this analysis there was a change in the annotated start codon (http://genolist.pasteur.fr/TubercuList) by -20 AA. However, in the amplicon sequencing an additional mutation was found in e*sxP* corresponding to N83D (Tables 2.6 and 2.8). Interestingly this portion of the gene is conserved amongst the QILSS family of genes and is likely that the stringency of the SNP calling software used by the BROAD bioinformatics systems would exclude a SNP in 1 out of the 5 QILSS genes because of its low occurrence. Similarly when I analysed the clade specific nsSNPs for the KZN isolates both methodologies revealed: *esxK* (A58T) and *esxL* (R33S). The *esxL* mutation corresponds to two nucleotide changes (g97t and g98t) (Table 2.6 and 2.8). In Figure 2.12 these have been annotated as individual codon changes (R33C and R33P) but taken together they result in an R33S mutation. Two mutations in *esxV* that occurred in nearly all the isolates (Q20L and S23L) were also found in both datasets, including in the dataset of Uplekar and colleagues (Uplekar *et. al*., 2011), but an additional *esxV* mutation (A57V) was found across the same broad range of strains only by amplicon sequencing (Table 2.6 and 2.8). Similiarly e*sxV* is an Mtb9.9 gene and the residue 57 is in a highly conserved portion of the protein.

**Figure 2.12** Phylogenetic tree based on all SNPs identified in 130 full genome sequenced isolates rooted to H37Rv. The grid to the right indicates the presence (bold) or absence (blank) of nSNPs identified in 23 esx genes relative to H37Rv. The unlabelled column on the left signifies if there are no mutations present.

**2.4    Conclusion**

73

The key finding of my study was that the sequence variation in the ESX genes is phylogenetic, and no mutations were identified that appeared to have arisen independently in distinct clades. Although this points against immune selection, previous work that has identified a strong phylo-geographic relationship amongst global strains of *Mtb* and raised the hypothesis that host adaptation may have occurred. The Beijing strains for example evolved in Asia and only recently spread. Genetic diversity might therefore have been selected for in the context of an Asiatic host immune system and sufficient time has not yet elapsed for the other human populations to exert a contrary immune selection.

ESX genes SNP variation appeared to be concentrated in Beijing and KZN isolates as well as other isolates in the Mtb9.9 and QILSS subfamilies (Figure 2.10). *esxI, K, L, M, N, O, P, U, V* and *W* all had multiple SNPs identified in them when both amplicon and genome sequencing sets were combined.

It is unclear yet how SNPs might influence the phenotype of clinical isolates. Mestre and colleagues, for example studied variation in DNA repair genes and were able to subdivide the Bejing clade but it is not known whether there are distinct phenotypes in these sublineages (Mestre *et. al.,* 2011). An alternative approach such as a study conducted in 2010 by Ashiru and colleagues has been to attempt to discover phenotypes for specific clades. They looked for the mechanism that may confer a more virulent phenotype that would account for strains being prominently clustered and associated with drug resistance. They found that there was an increase in the Beijing and KZN strains ability to adhere and invade A549 alveolar epithelial cells. They also found that the XDR KZN strain appeared to adhere more effectively than all the other strains to the A549 cells (Ashiru *et. al.,* 2010).

Clearly further studies are required to examine the consequences of phenotypic variation and my approach of investigating how microbial diversity impinges on the hosts' immune responses is a more direct and feasible approach than looking for genetic associations with imprecise epidemiological phenotypes.

My study involved the use of traditional sequencing of the ESX genes with PCR primers designed specifically to each gene generating long sequence reads from the ABI sequencer. The sequence assembly and SNP identification was done traditionally with the DNAStar software package. On the contrary, Comas and colleagues used the Illumina technology for sequence assembly and SNP identification as a result of the short sequence reads generated (Comas *et. al.,* 2010). This has been the major pitfall in their assessment of SNP variation, as my data suggests higher genetic variability in the Mtb9.9 and QILSS subfamilies (Figure 2.9), inferring that they could have underestimated the amount of variation in these ESX families. Since the Illumina technology generates short and random reads during sequencing, SNPs occurring in the highly homologous ESX subfamilies may have been incorrectly assembled resulting in Comas and co-workers overlooking polymorphisms identified in these data. They reported a high frequency of SNPs in *esxH,* which was conserved in my sequencing results. For this reason, my results are in disagreement with the findings of Comas and colleagues and in agreement with Uplekar and colleagues (Uplekar *et. al.,* 2011).

To date, no variation in *esxA* has been reported, yet in a recent study, 18 clinical isolates were reported to harbour the amino acid substitution (E68K) in *esxB* (Uplekar *et. al*., 2011). Since commercially available TB diagnosis assays utilize peptides from *esxA* and *esxB,* this particular *esxB* mutation may influence *esxB* peptide responses in IGRA assays, QuantiFERON Gold Test (Mazurek *et. al.,* 2001) and T-SPOT TB (Meier *et. al.,* 2005) in the diagnosis and vaccine development against TB.

# CHAPTER THREE

# Comparative Transcriptional Analysis of the

# ESX Gene Family

## 3.1 Chapter Summary

In this chapter, I describe experiments to characterize the transcription of ESX gene family members in different strains of *Mtb*. The experimental design challenge was to design specific primers or probes to detect single nucleotide polymorphisms, which are required to differentiate between some members of the ESX gene family. This is particularly a problem when targeting the SNPs within the QILSS and Mtb9.9 subfamilies. For some *esx* genes I was able to design a conventional Syto9 based RT-qPCR to compare transcription. This was performed in accordance with the published recommened Minimum Informaion of Quantitative Real-Time PCR Experiments (MIQE) guidelines (Bustin *et. al.,* 2009). Differences between strains were detected in the expression of *esxA, B, C, F, M* and *Q* genes. Expression was expressed relative to *sigA*. Attempts to develop an allele specific RT-qPCR assay were not successful for near identical members of the QILSS and Mtb9.9 subfamilies.

## 3.2 Materials and Method

### 3.2.1 Allele Specific PCR for SNP Detection in *esxI, J, K, L, M, N, O, P, V* and *W*

Several of the ESX genes only differ by limited numbers of SNPs and these have not been possible to transcriptionally analyze using standard microarrays because of the homology between gene sequences. For genes with only a single polymorphism distinguishing ESX members it was therefore necessary to clone PCR templates with and without the mutation in order to develop and evaluate an allele specific PCR.

#### 3.2.1.1   SNP Detection Strategy

The SNP detection strategy (Figure 3.1) was developed starting with amplification of the flanking region of the ESX gene of interest housing the SNPs. The amplified bands were then verified on a 1% agarose gel and gel purified. The purified products were then cloned into the TOPO Cloning vector, transformed into competent *E. coli* cells followed by plasmid isolation on the positive transformants. The specificity of the SNP primers were then confirmed by testing each primer pair with the gel purified cloned flanking region template. Thus, one would only expect amplification of the SNP specific to the

primer used in the reaction and not of any other ESX member from the QILSS and Mtb9.9 subfamilies.
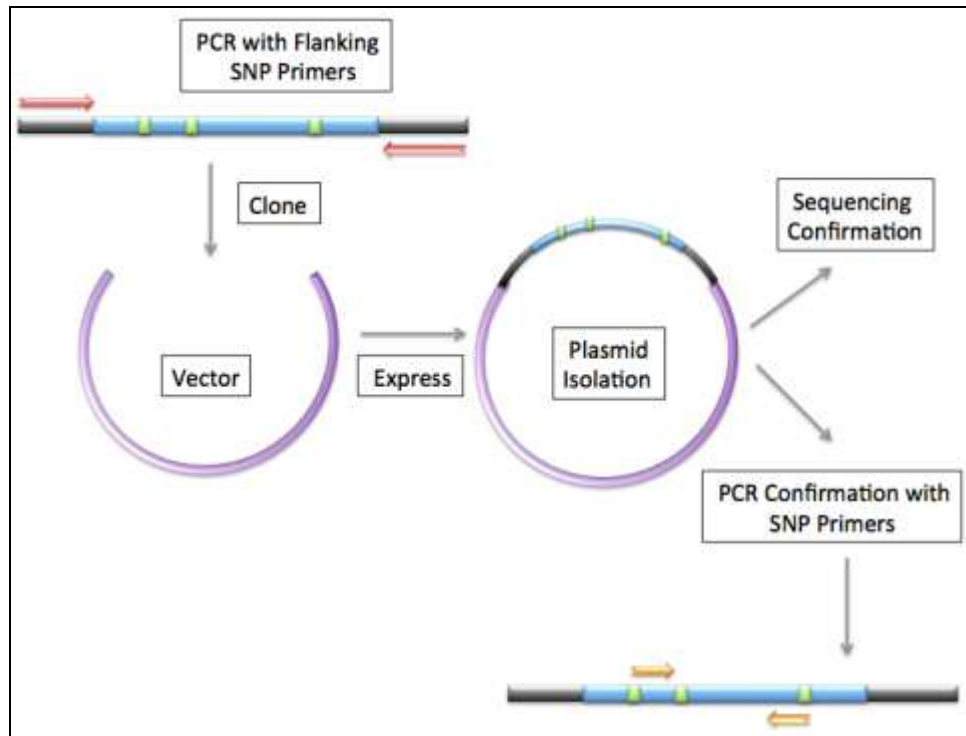


**Figure 3.1**    Experimental Design for SNP Primer Specificity Confirmation in QILSS and Mtb9.9 ESX members. Schematic is as follows: Red arrows represent Primers amplifying the region flanking the SNPs and Orange arrows represent primers targeting the SNPs of interest.

### 3.2.1.2    PCR of *esxI, J, K, L, M, N, O, P, V* and *W* SNPs

Individual primers were designed for the SNPs and region flanking the SNPs of interest using the Primer3 program (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) (Table 3.1a and b). The PCR reactions for the QILSS and Mtb9.9 ESX genes were amplified from H37Rv genomic DNA using the ESX flanking primers in a 20μl reaction volume. The PCR was performed with the Phusion High Fidelity PCR master mix (New England Biolabs, Finland, UK) containing 1μl of the template DNA, 10μl of the 2x Phusion mastermix, 5pM of each primer and 1.5μl of DMSO. The thermal cycling was performed in a GeneAmp PCR system 9700 PCR machine (Applied Biosystems, Foster City, CA, USA), with the following thermal profile: initial denaturation step of 98°C for 30 seconds, followed by 30 cycles of 98°C for 30 seconds, annealing at 55°C for 30 seconds and extension at 72°C for 30 seconds. The final extension time was 72°C for 7 minutes. Primer specificity was confirmed through gel electrophoresis of the amplicons with Sybr safe (Invitrogen, Carlsbad, CA). Bands of the expected sizes for the flanking region were selected for gel purification using the MiniElute Gel Extraction Kit (Qiagen) as per manufacturer's protocol and stored at -20 °C for further cloning experiments.

**Table 3.1a**     Allele Specific Primers for SNP Detection.

| Genes | Expected Product Size (bp) | Primer Sequence (5' – 3') |
|---|---|---|
| *esxN*snp | 177 | F: ATGACGATTAATTACCAGTT<br>R: CTCCTGGCAAGCCACCGAAC |
| *esxO*snp | 156 | F: GCCGCGGGTGACTTTTG<br>R: TCAGGCCCAGCTGGAGCCGA |
| *esxL*snp | 258 | F: TGTCGACGCTCACGGC<br>R: CAGGCCCAGCTGGAGCCGAC |
| *esxI*snp | 131 | F: CGGCGCCATGATCC<br>R: GCTGGGTAATGAACCCC |
| *esxV*snp | 105 | F: ACGCTCACGGCGCCATG<br>R: ACCGGCGCCGCCC |
| *esxK*snp | 164 | F: ACGTTTTATGACGGATCCGCA<br>R: CCATGGTGTCTAGCGAGGTC |
| *esxJ*snp | 90 | F: ATGCTGCACGGGGTG<br>R: TCAGCTGCTGAGGATCTGC |
| *esxP*snp | 190 | F: AACACGTTTTATGACGGAT<br>R: CGAAACGCCTGATTCATCT |
| *esxW*snp | 294 | F: GACCTCGCGTTTTATGACG<br>R: CAGCTGCTGAGGATCTGC |

F: Forward Primer; R: Reverse Primer

**Table 3.1b**     Primers for Amplification of ESX Flanking Regions.

| Genes | Expected Product Size (bp) | Primer Sequence (5' – 3') |
|---|---|---|
| *esxN*flank | 560 | F:  gccaacaactacgaacagca<br>R:  ggtgcgggtactctgttgat |
| *esxO*flank | 520 | F:  ccaacaactacgagcagcaa<br>R:  cttagcggaggcaccagag |
| *esxL*flank | 642 | F: cagcagatcctcagcagcta<br>R: actgaccagcaccgaagaac |
| *esxI*flank | 654 | F: gccaacaactacgaacagca<br>R: ctgtcggtttgtgccatgt |
| *esxV*flank | 487 | F: gccaacaactacgaacagca<br>R: aacttaggacccgggttagg |
| *esxK*flank | 340 | F: ggcgcagactgtcgttattt<br>R: aaaagtcactcgcggtcaac |
| *esxJ*flank | 380 | F: tttgaggccagcaattaacc<br>R: aaaagtcactcgcggtcaac |
| *esxP*flank | 400 | F: accagtgatcggcggtct<br>R: gctgctgaggatctgctg |
| *esxW*flank | 582 | F: acgtgcgtttaacaacttcg<br>R: aaaagtcactcgcggtcaac |

F: Forward Primer; R: Reverse Primer

### 3.2.1.3 Cloning of Flanking SNP Region

The gel purified flanking products were cloned using the Zero Blunt TOPO PCR Cloning Kit (Invitrogen, Carlsbad, CA) as per manufacturer's protocol. Ligated products were transformed using the One Shot TOP10 chemically competent *E. coli* cells (Invitrogen, Carlsbad, CA) as per manufacturer's protocol and selection of positive clones occurred on LB plates supplemented with Kanamycin (50μg/ml). Positive transformant colonies were selected and grown overnight in 2 ml LB medium supplemented with 50μg/ml kanamycin, from which the plasmid was isolated using the QIAprep Spin MiniPrep Kit (Qiagen), as per manufacturer's protocol. Plasmid Digestion was performed with EcoRI enzyme (Fermentas Life Sciences, USA) as per manufacturer's protocol and visualized on a 1.5% agarose gel stained with Sybr Safe (Invitrogen, Carlsbad, CA). PCR and Sanger sequencing according to protocols described in Chapter 2, Section 2.2.6, confirmed the SNP in each of the successfully cloned regions.

### 3.2.1.4 Confirmation of SNP Primer Specificity

A PCR based approach was used for the confirmation of primer specificity for the SNPs prior to RT-qPCR usage. The gel purified flanking products were used as the DNA templates in the 25μl GoTaq Green Master Mix (Promega Corp., USA) PCR reactions. These reactions were performed using each SNP primer pair against all the flanking products. Each reaction was composed of 12.5μl of 2x GoTaq Green master mix, 10pM of each SNP primer pair, 2.5μl of nuclease-free water and 5μl of the purified flanking products, in a GeneAmp PCR System 9700 PCR machine, respectively. The thermal cycling conditions included an initial denaturation step of 95°C for 2 minutes, followed by 30 cycles of 95°C for 30 seconds, annealing at 60°C for 30 seconds and extension at 72°C for 30 seconds. Final extension occurred at 72°C for 5 minutes followed by PCR product visualization on a 1% agarose gel using Sybr Safe (Invitrogen, Carlsbad, CA).

### 3.2.2 RT-qPCR Assay for *esxA, C, F, M* and *Q*

#### 3.2.2.1    RNA Extraction

RNA was extracted from 50ml cultures, grown to mid-exponential and early stationary phase. These cultures were centrifuged at 4000rpm for 10 minutes at room temperature, and the pellets immediately resuspended in 1ml of cold Trizol reagent (Invitrogen, Saint Aubin, France). The suspension was transferred to a 2ml screw-cap tube containing 0.4ml silica beads/zirconia (0.1mm in diameter) (Sigma-Aldrich, Saint Louis, MO), and the cells were mechanically disrupted at 7000rpm for 1 minute at room temperature using the MagNa Lyser (Roche Diagnostics South Africa, Randburg, SA), and cooled at -20℃ for 2 minutes. This process was repeated 5 times. The samples were periodically inverted for 5 minutes and centrifuged at maximum speed for 45 seconds to remove cell debris. The supernantant was transferred to 2.0ml tubes containing 300µl of chloroform (Sigma-Aldrich, Saint Louis, MO), inverted rapidly for 15 seconds, followed by a 2 minute incubation at room temperature. This was followed by centrifugation at 12000 rpm for 10 minutes at 4℃. The upper aqueous layer was transferred into a 2.0ml tube containing 500µl isopropanol, inverted rapidly for 15 seconds and incubated at room temperature for 10 minutes. Samples were pelleted at 12000rpm for 10 minutes at 4℃, and the pellets were washed with 1ml of 75% ethanol, centrifuged for 5 minutes and air-dried for 10 minutes at room temperature. Resultant pellets were resuspended in 36µl of cold DEPC water (Invitrogen, Saint Aubin, France), and the RNA concentration and integrity was determined using the Bio-Rad Experion RNA Standard Sens chip (Bio-Rad, Hercules, CA), as per manufacturer's protocol.

### 3.2.2.2 Genomic DNA Removal and cDNA Synthesis

Prior to cDNA synthesis, the RNA was treated with RNase-free DNase I (Fermentas Life Sciences, USA) following the manufacturer's instructions. For the treatment, 5µl of the 10x DNase I buffer, 5U of RNase-free DNase I were added to 5µg RNA and DEPC-treated water, totalling a 50µl volume. The mixture was incubated at 37ºC for 15 minutes. Nuclease activity was terminated with the addition of 5µl of 25mM EDTA (pH 8.0), and the mixture was incubated at 65ºC for 10 minutes.

For cDNA synthesis, the DNase I treated RNA was used in a total reaction volume of 40µl, with the iScript cDNA Synthesis Kit (Bio-Rad, Hercules, CA) as per manufacturer's instructions. The reaction mixture contained 8µl of the supplied 5x iScript Reaction Mix and 2µl of the iScript Reverse Transcriptase and nuclease-free water. The reaction mix was incubated at 25°C for 5 minutes, at 42°C for 30 minutes, 85°C for 5 minutes and finally chilled on ice. The cDNA was stored at -20ºC until use for real time PCR.

### 3.2.2.3 Primers for RT-qPCR

Primers were designed for the *esx* genes of interest using the DNAStar primer designing software (LaserGene 7.2, Madison, Wisconsin, USA) (Table 3.2), and purchased from Inqaba Biotec (Pretoria, South Africa). For the reference gene, *sigA* was chosen due to the abundant presence within live cells. The primer specificity was confirmed by gel electrophoresis of the amplified products.

**Table 3.2**     Primers for RT-qPCR Assay.

| ESX Genes | Oligonucleotides (5' - 3') |
| --- | --- |
| *esxA* | F: CAGGTTCTGCAGCGCGTTGTTC |
| | R: CGCAATCCAGGGAAATGTCAC |
| *esxB* | F: GGCAGAGATGAAGACCGATG |
| | R: CTCGTCGAGTTCCTGCTTCT |
| *esxC* | F: CGATCGCGTTGTCCAGCA |
| | R: AAGACACCGCCAGCAAAA |
| *esxF* | F: ACGCGGTCTCGTTCTGTTGATA |
| | R: GCAGATGTTGGGCGGGTGG |
| *esxM* | F: TACGCAGCTGTGGCTTTCGG |
| | R: ACCTCGCTAGACACCATGACCT |
| *esxQ* | F: GACCCCACCCTAATCGTTCC |
| | R: CCATGGAGGCTCTCGC |
| *sigA* | F: CCTACGCTACGTGGTGGATT |
| | R: TGGATTTCCAGCACCTTCTC |

F: Forward Primer; R: Reverse Primer

### 3.2.2.4 Constructon of External Standard Curves

The genomic DNA from H37Rv served as the known copy number standard for each gene targeted. Standard curves were established by 10 fold serial dilutions of the 100ng/µl genomic DNA to a dilution of $10^{-5}$. The standards were performed in duplicate to achieve a reliable standard curve range per gene target. A negative control ($H_2O$) was included to check for cross contamination in each run. The raw data analysis was performed using the BioRad CFX Manager version 3.0. The RT-qPCR assay was quantified according to the recommended hallmarks by the manufacturer (Bio-Rad, Hercules, CA). This optimisation was in agreement of the linear standard curve having an $R^2 > 0.980$ or r > 1-0.990), a high amplification efficiency in the range of 90-105% and consistency across replicate reactions.

The ratio of measured signals was used to calculate the threshold cycle ($C_T$) value. The standard curve (plot of $C_T$ value against copy number or log starting quantity of the standard) was generated using different dilutions of the standard in paralell for each gene.

### 3.2.2.5 RT-qPCR Assay

RT-qPCR was performed using the BioRad CFX96 Real-Time System instrument (Bio-Rad, Hercules, CA) in a total volume of 10µl, using 2µl of undiluted cDNA and 5pM of gene specific primers (Table 3.2), 12.5µl of the GoTaq Colorless Mastermix (Promega Corp., USA) and 50µM of Syto9 (Invitrogen, Saint Aubin, France). A negative control ($H_2O$) was included as an assessment for cross contamination in each run. The reactions were performed in duplicate for reproducibility. Cycling conditions was as follows: after the enzyme activation step at 50ºC for 10 minutes, 40 cycles of amplification was performed at 95ºC for 5 minutes, 95ºC for 10 seconds and 60ºC for 15 seconds. At the end of the annealing step, fluorescence was measured. A melt curve was generated with the following conditions: 95ºC for 10 seconds minute, 65ºC for 5 second and cooled at increments of 0.5ºC for 5 seconds. The gene targets (*esxA, B, C, F, M* and *Q*) and reference gene (*SigA*) were quantified using the established standard curves to confirm that the cDNA was amplified with similar efficiency. The raw data analysis was performed using the BioRad CFX96 Real-Time System instrument (Bio-Rad, Hercules,

CA). The ratio of measured signals was used to calculate the $C_T$ value. The results were expressed as the absolute mRNA copy number per 5µg/µl for each gene, using the absolute quantification method according to the following equation:

$$Copies = 10^{\frac{(C_T - b)}{m}}$$

I opted for the absolute quantification method, due to my interest in interpolating the quantity of *sigA* in the assayed samples. The ratio of the expression of the *esx* genes was expressed relative to *sigA*.

## 3.3 Results and Discussion

### 3.3.1 Allele Specific PCR SNP Detection in *esxI, J, K, L, M, N, O, P, V* and *W*

Results are shown for *esxO*. The PCR amplification, using the *esxO* flanking primer set, successfully produced the expectant product of 520 bp (Figure 3.2). The Phusion High Fidelity polymerase with proofreading activity was used in this reaction to reduce possible errors that may be introduced by amplification. The remaining members of the QILSS and Mtb9.9 subfamilies also were successfully amplified at the expectant product sizes (Table 3.1b) (results not shown). This gel purified flanking product was successfully cloned using the Zero Blunt TOPO PCR Cloning Kit and transformed into One Shot TOP10 chemically competent *E. coli* cells (Figure 3.3) and presence of the PCR product was confirmed by EcoR1 enzymatic digestion (Figure 3.4). Plasmid sequencing was also performed to confirm the identity of the cloned product as well as whether the product was inserted in correct frame.

The SNP primers were used in a PCR assay to attain the specificity on amplifying the regions containing the SNPs of interest. Therefore, the purified flanking PCR product from Figure 3.2 was used as the template in the reaction together with the *esxO* primer pair. In addition, the purified flanking PCR products from the remaining QILSS and Mtb9.9 members were used as template for the *esxO* primer set. Based on the products from Figure 3.5, the *esxO* SNP primers were found not to be specific due to amplification of the QILSS and Mtb9.9 members. This trend was also noted in combination with the other SNP primers for the other genes (results not shown).

Additional steps were implemented to optimise and improve the assay in order to obtain the expected results of allele specificity. This also included altering the $MgCl_2$ concentration in the PCR reactions, implementing a temperature gradient during the PCR cycling, increasing and decreasing the primer concentrations and PCR template. All these approaches yielded an increase in non-specific amplification of the QILSS and Mtb9.9 members and primer dimers (results not shown). Therefore, the attempts to improve the specificity of this assay were unsuccessful and I was unable to generate sufficient specificity to evaluate differential ESX gene expression.
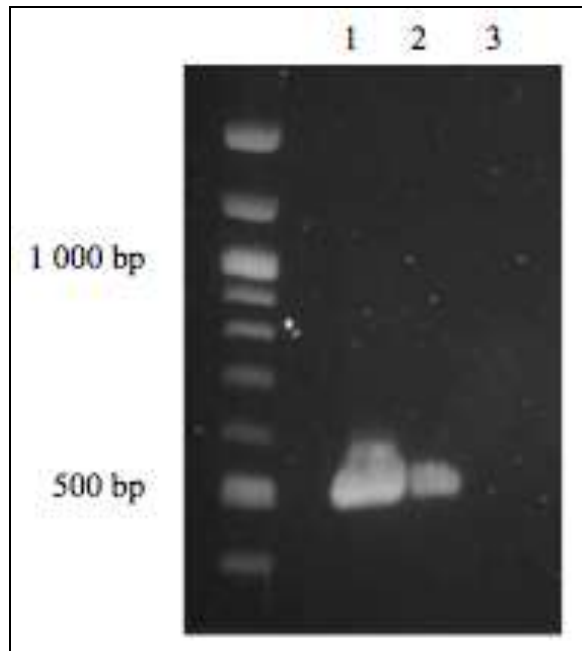
**Figure 3.2** PCR amplification of *esxO* gene with *esxO* flanking primers on a 1.5% (w/v) agarose gel with H37Rv genomic DNA. Lanes 1 and 2 represent the H37Rv 20μl and 5 μl of the amplified product that was loaded onto the gel, respectively. Lane 3 is the no DNA control. A 100bp molecular weight marker was run on the gel to verify amplified products.
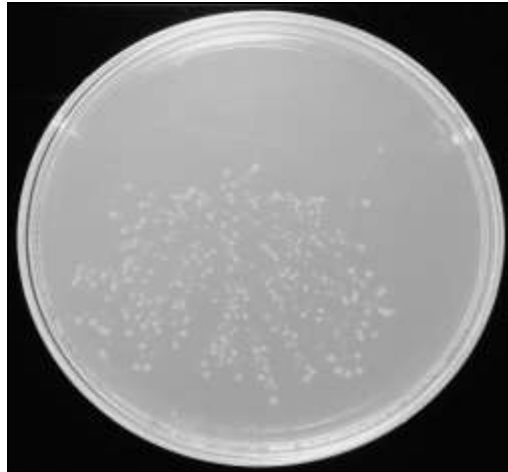
**Figure 3.3**    *esxO* positive transformant colonies grown on a LB plate supplemented with 50 μg/ml Kanamycin.



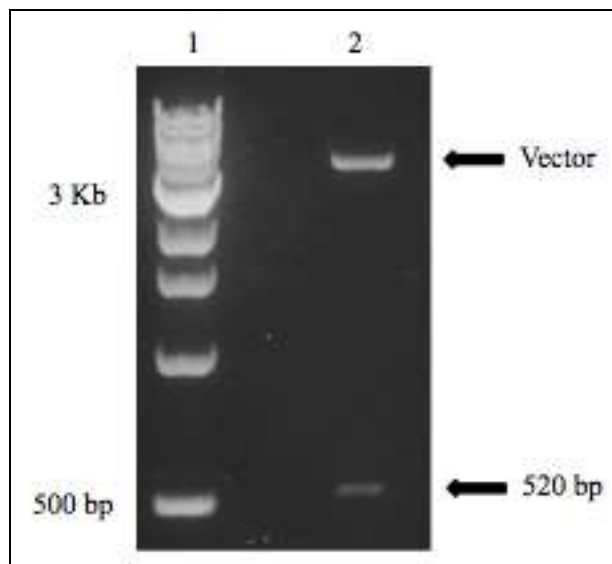**Figure 3.4**    An EcoR1 restriction digest of the pCR-Blunt II-TOPO vector verifying the vector and cloned *esxO* PCR product.
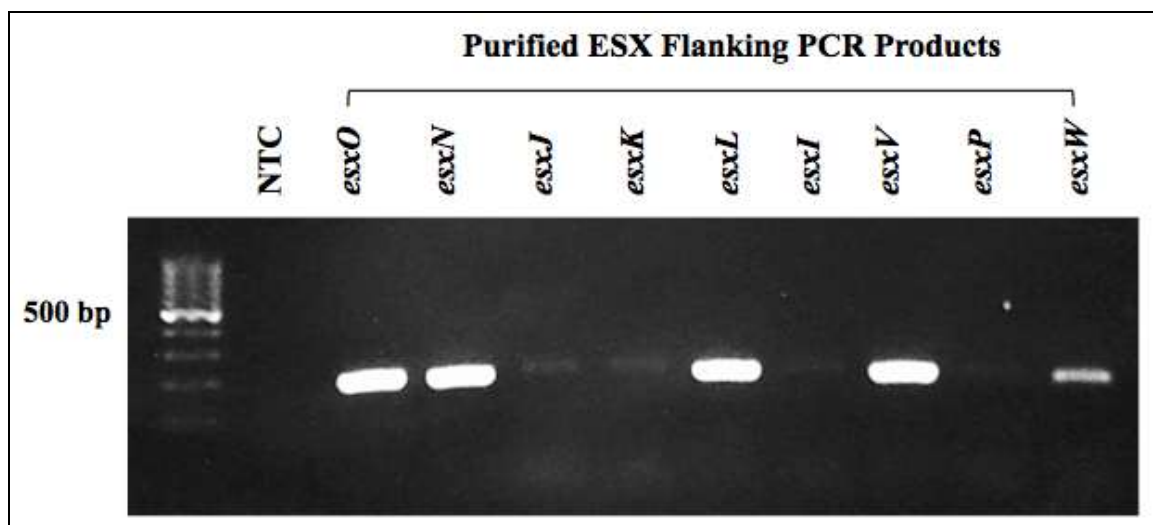
**Figure 3.5** PCR amplification using the *esxO* primer set in combination with purified flanking PCR products from the remaining QILSS and Mtb9.9 members on a 1.5% (w/v) agarose gel. NTC is the no template DNA control. A 100bp molecular weight marker was run on the gel to verify amplified products.

### 3.3.2 RT-qPCR Assay for *esxA, B, C, F, M* and *Q*

#### 3.3.2.1    RNA Integrity Confirmation

Figure 3.6 is the virtual gel for the qualitative assessment of RNA integrity using the Experion RNA StdSens analysis Kit (Bio-Rad, Hercules, CA). The internal 50bp lower marker was present in all samples, which served as the normaliser for the migration times of the samples in the different wells. Intact 16s and 23s bands are present across all samples (Figure 3.6). Similar virtual gel results were obtained for the remaining samples (results not shown). The RNA Quality Indicator (RQI) was used to estimate the level of degradation in the RNA samples. This was automatically calculated using the Experion in-house RQI algorithm software. For all the isolates tested, the RQI value was between the recommended range of 7-10. According to the manufacturer recommendations, this is an acceptable range for intact RNA, thus rendering my samples acceptable for RT-qPCR analysis.
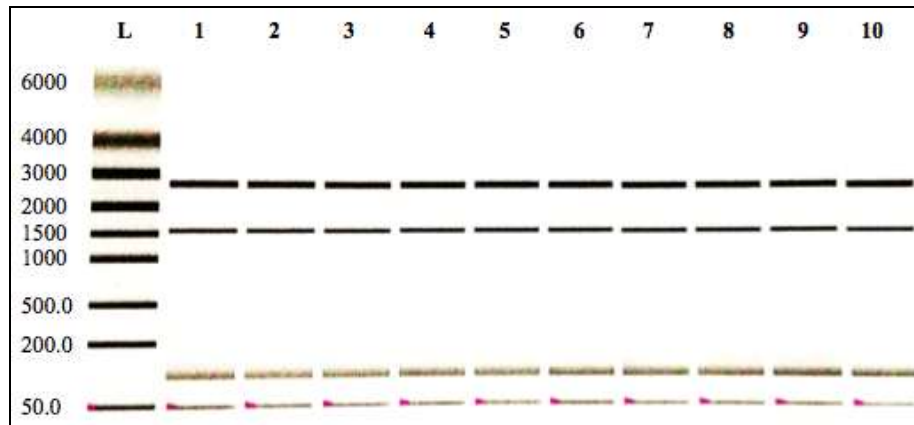


**Figure 3.6**    Virtual Gel generated from Bio-Rad Experion RNA Standard Sens chip. Lanes 1 to 12 contain the DNAse treated RNA. The RNA ladder is in the Lane L. The DNAse treated RNA was loaded onto the chip in the following order, Lane 1: H37Rv; Lane 2: H37Ra; Lane 3: BCG; Lane 4: 605; Lane 5: 443; Lane 6: 4207; Lane 7: 1528; Lane 8: 62; Lane 9: 26 and Lane 10: 299.

### 3.3.2.2 Standard Curves Generated with H37Rv Genomic DNA

The graphical representation for the serial dilutions of the concentrated genomic DNA, ranging from a maximum of 100ng, are presented in Figure 3.7. As displayed by the graphs, the $C_T$ values are spaced evenly, denoting the linear relationship of the dilution range. However, at the highest dilutions the $C_T$ values and DNA concentration developed into a non-linear relationship. This implied that the values were close to the quantification limit and those DNA concentrations were excluded from the standard curve (results not shown). Likewise, for the very high DNA concentration of $10^2$, a non-linear relationship also occurred between the $C_T$ value and DNA concentration. As a result, I aimed to ensure all quantification was carried out only in the linear range of each assay.

Standard curves were generated to quantify the initial amount of the target genes and the reference gene (*sigA*) in the samples, using H37Rv genomic DNA. The curves were utilized to confirm the amplification efficiency of the cDNA. Figure 3.7 illustrates the standard curves for the *esx* genes and the reference gene, which were repeated in duplicate for the different RT-qPCR runs. A linear relationship was revealed once the calculated means of the copy numbers were plotted against the means of the $C_T$ values exhibiting a single log change per $C_T$ value. The limits of detection of the assays, when the amount of template DNA was converted to gene copy number using Avogadro's number and the size of the *Mtb* genome, ranged from 2 to 20. For simplicity I have called the absolute quantification units copy number even though the units are strictly in template concentrations derived from the standard curves using DNA concentrations.
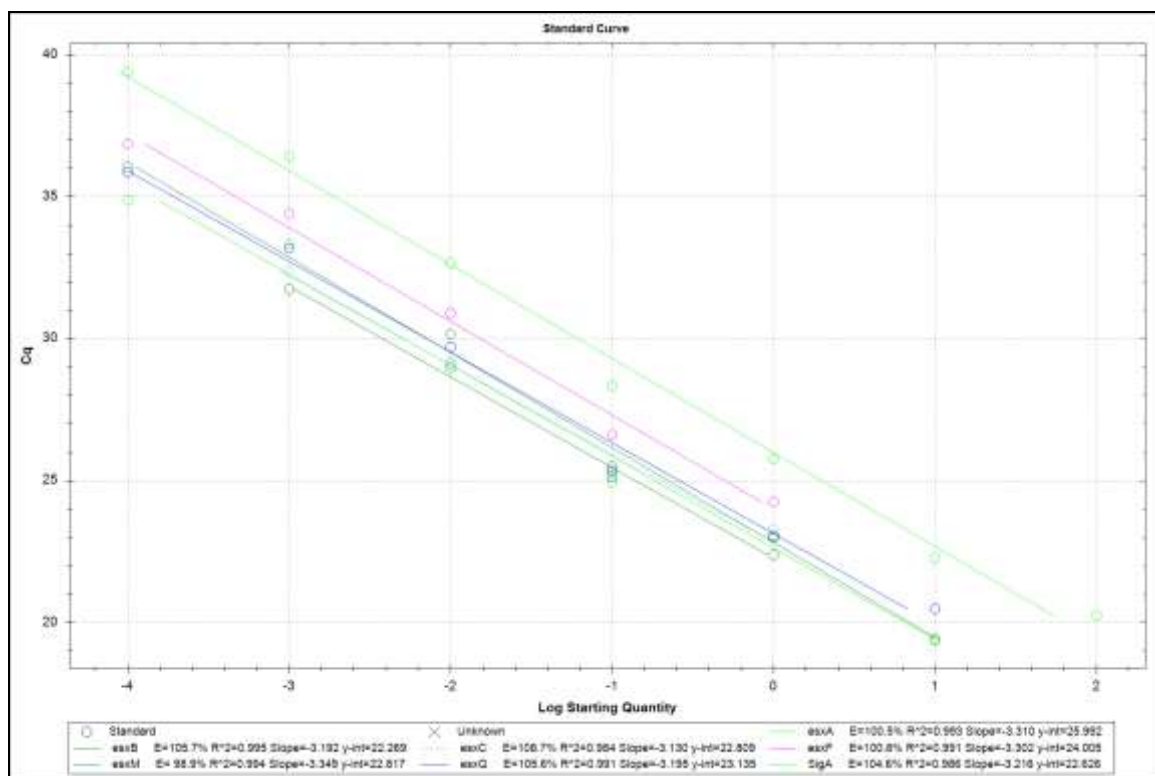
**Figure 3.7**  Standard curves of *esx* RT-qPCR assays showing cycle threshold for serial dilution of genomic DNA from 100ng per reaction for each gene target during 40 cycles of amplification.

### 3.3.2.3 *esx* RT-qPCR

Two independent biological replicates were performed and the experiments were repeated twice, and the RT-qPCR experiments were done in duplicates. Genes whose expression levels was less than 0.5 fold relative to *sigA* were considered as down-regulated or under expressed.

Absolute quantitation was performed to assess the copies of *sigA* expression for the selected isolates using the $C_T$ values. The *sigA* copy number was between $30x10^3$ - $46x10^3$ copies in each isolate sample (Figure 3.8) indicating expression levels were comparable across all isolates. *sigA* is the primary sigma factor that is essential for growth in *Mtb* (Manganelli *et. al.,* 2004). Due to the stability of *sigA* mRNA (half life >40) (Hu and Coates, 1999), and its wide use in RT-qPCR experiments (Manganelli *et. al.,* 1999; Dubnau *et. al.,* 2002), it was selected as the internal standard for normalization in the RT-qPCR experiment. Two previous studies have reported on the up-regulation of *sigA* during infection in human macrophages (Volpe *et. al.,* 2006) and its down-regulation in low aeration and stationary growth stages (Manganelli *et. al.,* 1999). Manganelli and colleagues speculated that in the low aeration condition and the stationary phase growth stage, the available energy to the bacteria may be low thus decreasing *sigA* expression. As a consequence, this decrease is reflected in a decrease in the mRNA pool relative to the total RNA within the cell (Manganelli *et. al.,* 1999). However, my experiments were conducted in well-aerated mid-log phase cultures, thus these fluctuations in *sigA* are unlikely to influence the results and the levels of the gene was very consistent across samples.
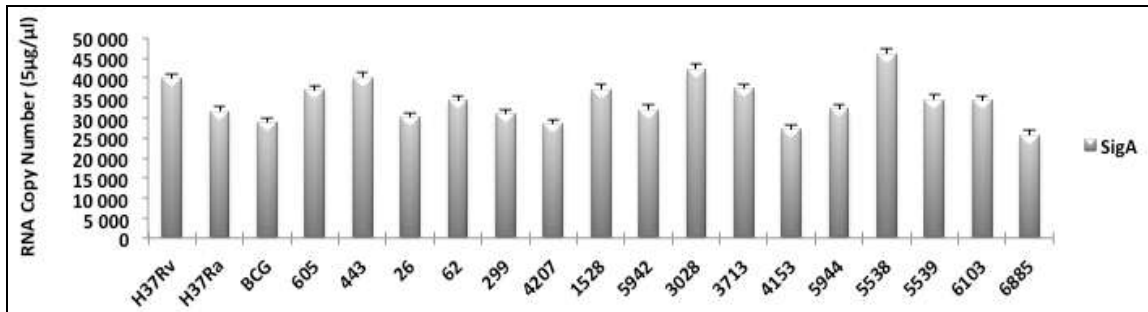
**Figure 3.8** RNA copy number of the reference gene (*sigA*) expression in selected isolates expressed as copy number per 5μg of total RNA. The error bars represent the standard error.

Represented by Figure 3.9 are the *esxA* and *esxB* expression levels relative to *sigA*. The *esxA* expression in H37Rv was higher than in the other isolates. The expression levels for *esxB* ranged from 0.5 to 1 between isolates. Amongst the other isolates that was variable levels of expression with several isolates such as 4207 (KZN Isolate) and 299 (Unknown Isolate) having lower levels of expression relative to *sigA*. Lowest levels of expression for *esxA* and *esxB* were observed in the Beijing Isolates. Interestingly in these isolates *esxB* transcripts were usually detectable at higher levels that *esxA*. The two negative controls H37Ra and BCG were below the limit of detection for both *esxA* and *B* as anticipated (Figure 3.9). Overall, variable expression was noted for *esxA* and *esxB* within these isolates under the experimental growth conditions. Evidently it would be interesting to look at gene expression in other conditions including intracellularly and in vivo to determine whether the variable expression levels I detected were maintained.

**Figure 3.9** Mean *esxA* and *esxB* expression levels in selected isolates relative to *sigA*. The error bars represent the standard error.

Figures 3.10 to 3.13 represent the other ESX gene expression levels per isolate relative to *sigA*. In Figure 3.10 variability in *esxC* expression was noted between the isolates. H37Ra, H37Rv and 7 isolates (605, 26, 62, 299, 4207, 1528 and 4153) had prominent expression while the remaining isolates expression for *esxC* was low. For *esxF* (Figure 3.11), only 5 isolates had marked expression (605, 26, 3713, 4153 and 6885). Comparison of *esxM* expression (Figure 3.12) between isolates revealed a generally lower level of expression except for H37Ra and isolates 443, 62 and 4207. In contrast *esxQ* (Figure 3.13) had higher levels of expression across isolates. Overall, *esxC, F, M* and *Q* was relatively down-regulated in BCG, 5942, 3028, 5944, 5538, 5539 and 6103.

97

**Figure 3.10** *esxC* expression levels in selected isolates relative to *sigA*. The error bars represent the standard error.



**Figure 3.11** *esxF* expression levels in selected isolates relative to *sigA*. The error bars represent the standard error.

98
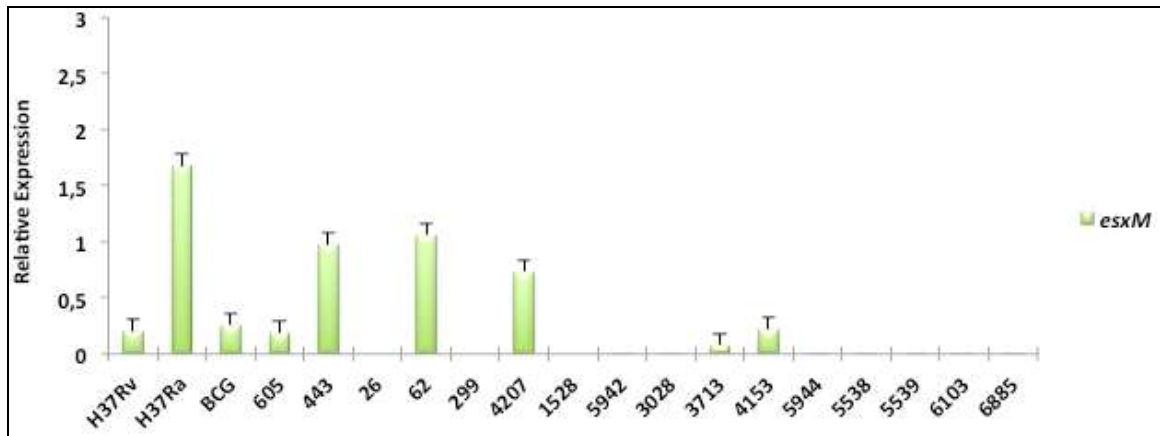
**Figure 3.12**    *esxM* expression levels in selected isolates relative to *sigA*. The error bars represent the standard error.
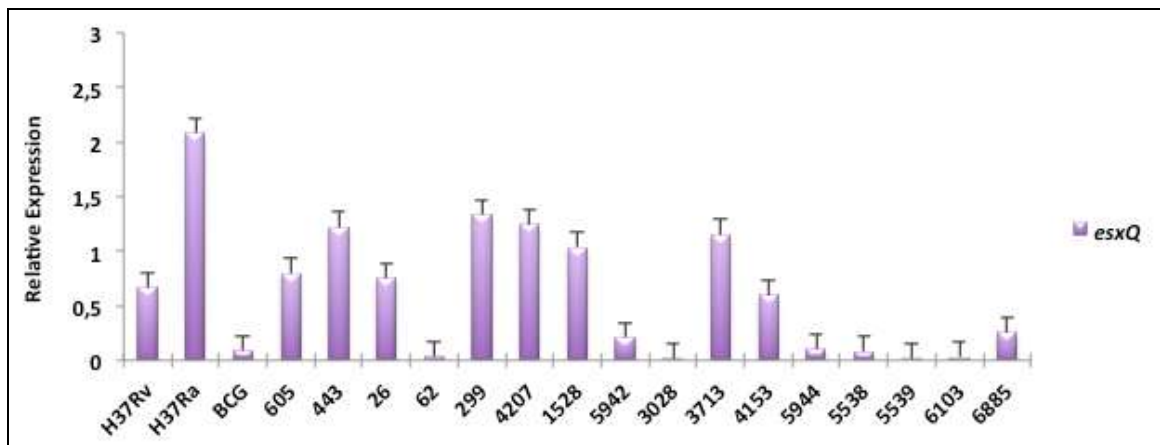


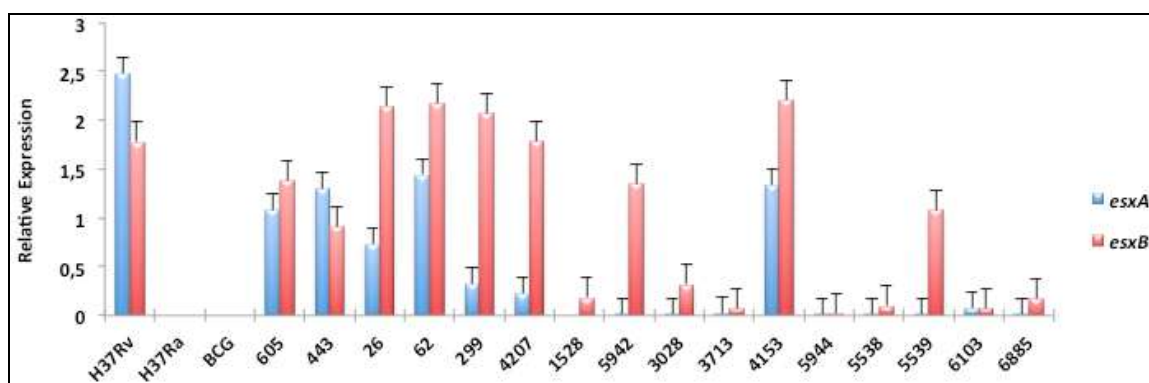**Figure 3.13**    *esxQ* expression levels in selected isolates relative to *sigA*. The error bars represent the standard error.

Given the variability amongst the clinical isolates I was interested to see if there were distinct patterns of gene expression for each of the three groups. Figures 3.14 to 3.16 are representations of ESX gene expression according to the 3 isolate groupings. When the levels of expression across all of the KZN (Figure 3.14), Beijing (Figure 3.15) and Other (Figure 3.16) isolates are compared there is clearly marked interstrain variability, however, there is also a striking difference when the groups are compared. The KZN and Other strains have in general higher levels of expression across all the gene members relative to the Beijing strains, hinting at some coordinated regulation of *esx* expression that differs between *M. tuberculosis* clades. As all of my comparisons are relative to *sigA*, it could be that the differences in expression of *sigA* in the different clades could account for this. I would need to exclude this by looking at expression relative to another reference gene, although the high level of expression of *esxC* and *esxQ* in two isolates points against this explanation.
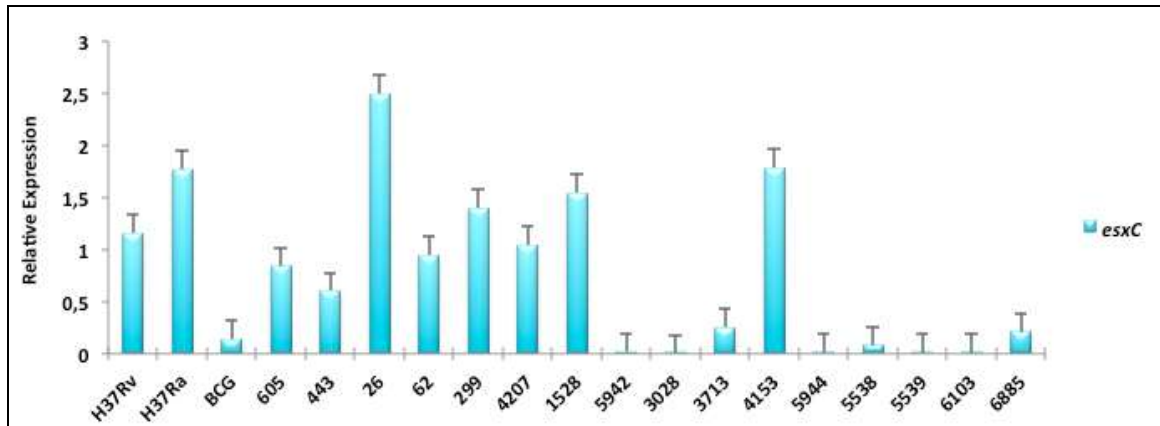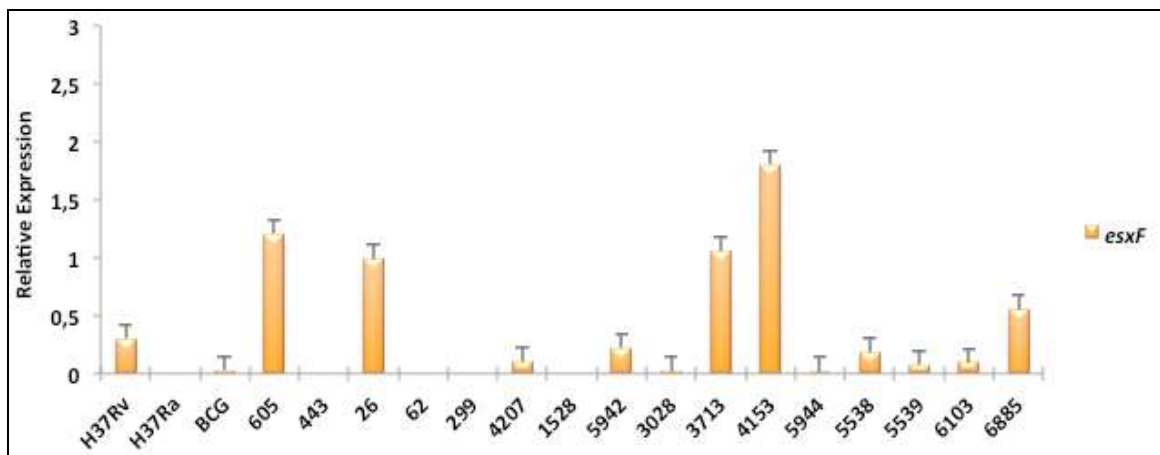


**Figure 3.14**    ESX expression levels in selected KZN Isolates relative to *sigA*. The error bars represent the standard error.

**Figure 3.15**    ESX expression levels in selected Beijing Isolates relative to *sigA*. The error bars represent the standard error.
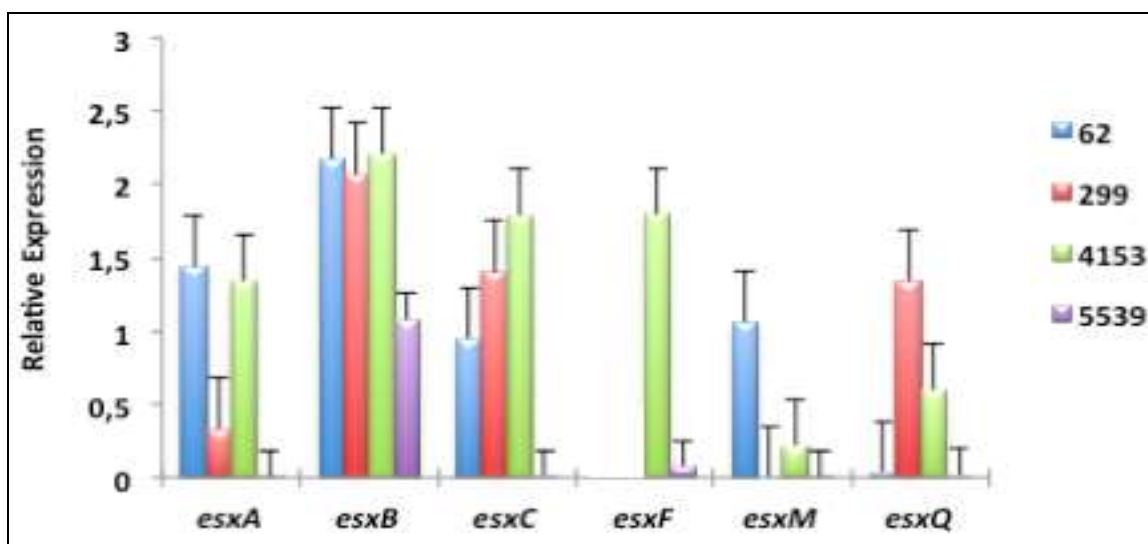


**Figure 3.16**    ESX expression levels in selected Other Isolates relative to *sigA*. The error bars represent the standard error.
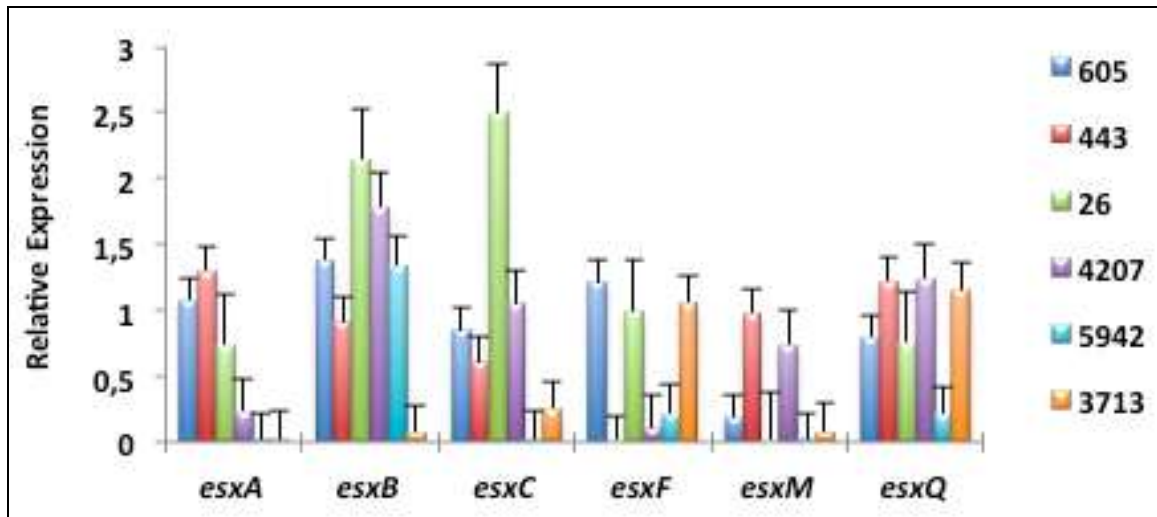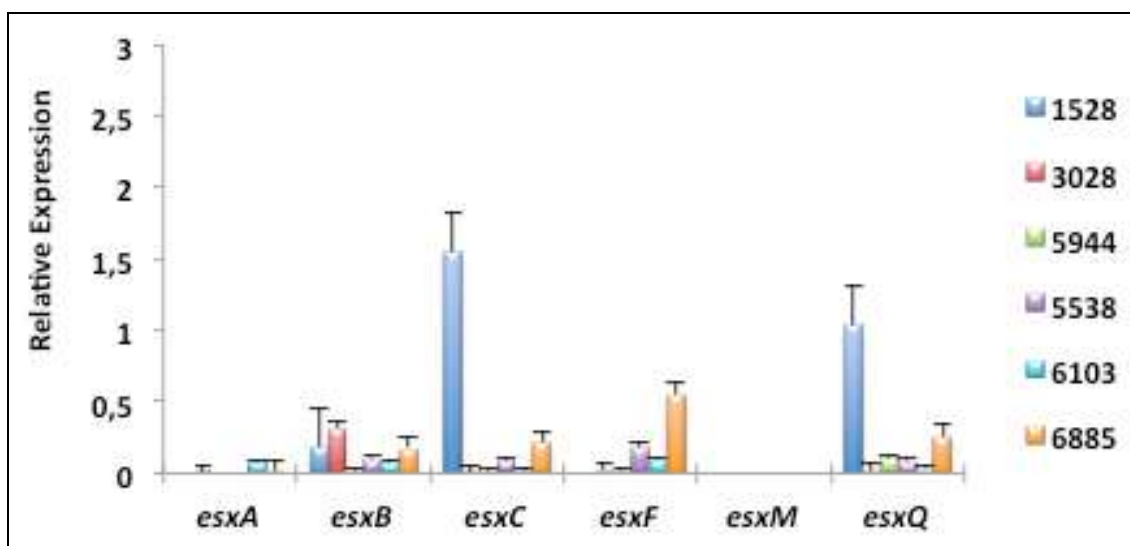
## 3.4 Conclusion

This initial screening of primer specificity prior to performing the RT-PCR assay is indicative of the difficulty posed by the sequence similarity of members from the QILSS and Mtb9.9 subfamilies. The cross-reactivity of the primers with different ESX members is a huge limitation in attaining further expression data for these genes. However, this assay is a good quality check for specificity in genes that display sequence similarity.

The expression of *esxA, B, C, F, M* and *Q* was detected at very low levels, but small gene variation in expression levels was noted across the three isolate groupings. Beijing isolates are poor expressors of the ESX genes, since the lowest expression levels was observed amongst this isolate group. The amount of mRNA produced is a reflection of the level of protein production that is coded by the gene (Bustin, 2002). This is indicative of a change in gene regulation for the ESX genes may confer an advantageous selective phenotype on the Beijing genotype.

Generally, the expression levels of all the isolates for *esxA, B, C, F, M* and *Q* were very low. But, differential expression did occur across all isolates. In 2011, Bukka and colleagues analyzed the RNA expression patterns in sputum. Amongst all the ESX transcripts, they identified that *esxIJ* and *esxKL* (members of the QILSS and Mtb9.9 subfamilies) were differentially expressed under varying growth conditions. The transcripts for broth culture growth for *esxIJ* were 4 fold lower than the *esxOP* transcripts that were 8 fold higher than the *sigA* reference transcript. Similarly, in *esxKL,* lower expression was observed. They also reported growth defects in the mutants generated from these genes (Bukka *et. al.,* 2011). However, my study did not include the remainder ESX genes for RT-qPCR analysis. Apart from the study by Bukka and colleagues, very little data is available on the transcriptional levels of the other ESX members in *Mtb.*

# CHAPTER FOUR

# ESX Protein and Expression Profiles for *M. tuberculosis* Clinical and Laboratory Isolates

## 4.1 Chapter Summary

This chapter describes a proteomics approach using SDS-PAGE, immunodetection and MALDI-TOF mass spectrometry to determine if difference at the protein expression level for ESAT-6 and CFP-10 could be identified in different clinical isolates of *M. tuberculosis*. The main objective was to quantitatively compare ESX protein expression profiles within and between *Mtb* clinical and laboratory isolates.

## 4.2 Materials and Method

### 4.2.1 ESX Protein Extraction

#### 4.2.1.1   Preparation of Culture Filtrate and Whole Cell Lysates:

For this component of the experiments, 20 isolates were selected to represent the 3 groupings used in this study, namely KZN, Beijing and Other. As represented in Table 4.1, the first 10 isolates were selected from Table 2.1a and 2.1b and the remainder were kind gifts from the strain collection maintained by Mr. K. Maharaj at K-RITH.

Isolation of both the culture filtrate and whole cell lysate was carried out as described previously (Smith *et. al.,* 2008). The cultures were grown in 7H9 media to the mid-log growth phase. Each culture was washed, diluted 10 times in Sauton's medium (Larsen *et. al.*, 2007) and cultured further to the mid-log phase. The step was repeated for an additional 2 days. The cells were pelleted by centrifugation at 12000rpm for 20 minutes at room temperature. The culture supernatant was collected and protease inhibitor cocktail tablets were added (Roche Diagnostics South Africa, Randburg, SA) to inhibit protease activity. This soluble fraction (culture supernatant with inhibitor cocktail) was concentrated with the Amicon Ultra-4 centrifugal filter device (Merck Millipore, Germany) with a molecular weight cutoff of 3,000. This concentrated culture filtrate fraction was used for further protein analysis. The whole cell lysates (pellet from the centrifuged culture) was washed once with PBS buffer (50mM Potassium Phosphate, 150mM NaCl, pH 7.2) and centrifuged at 5000rpm for 10 minutes at room temperature. The pellet was resuspended in PBS buffer supplemented with the protease inhibitor cocktail (Roche Diagnostics South Africa, Randburg, SA) followed by mechanical

disruption using the MagNa Lyser (Roche Diagnostics South Africa, Randburg, SA). Thereafter, the cell lysate was centrifuged at 6000rpm for 5 minutes at room temperature and the supernatant was concentrated with the Amicon Ultra-4 centrifugal filter device (Merck Millipore, Germany) with a molecular weight cutoff of 3,000. This served as the whole cell lysate fraction to be used in the SDS-PAGE assay.

**Table 4.1**    Bacterial strains used for the Protein component of this study.

| Isolates: | RFLP / Strains: | Received From: |
|---|---|---|
| **H37Rv** | Reference Strain | Medical Micro. (UKZN) |
| **H37Ra** | Avirulent Strain | Medical Micro. (UKZN) |
| **BCG** | Reference Strain | Medical Micro. (UKZN) |
| **605** | KZN | Medical Micro. (UKZN) |
| **443** | KZN | Medical Micro. (UKZN) |
| **4207** | KZN | Medical Micro. (UKZN) |
| **1528** | Beijing | Medical Micro. (UKZN) |
| **62** | Unknown | Medical Micro. (UKZN) |
| **26** | KZN | Medical Micro. (UKZN) |
| **299** | Unknown | Medical Micro. (UKZN) |
| **3028** | Beijing | K. Maharaj (K-RITH) |
| **4153** | Unknown | K. Maharaj (K-RITH) |
| **5539** | Unknown | K. Maharaj (K-RITH) |
| **3713** | KZN | K. Maharaj (K-RITH) |
| **5538** | Beijing | K. Maharaj (K-RITH) |
| **6103** | Beijing | K. Maharaj (K-RITH) |
| **3717** | KZN | K. Maharaj (K-RITH) |
| **5944** | Beijing | K. Maharaj (K-RITH) |
| **6885** | Beijing | K. Maharaj (K-RITH) |
| **5942** | KZN | K. Maharaj (K-RITH) |

#### 4.2.1.2 Total Protein Concentrations

Total protein concentrations were determined spectrophotometrically using the Pierce BCA Protein Assay Kit (Pierce Biotechnology, Rockford, USA), according to the manufacturer's instructions. Briefly, a fresh set of Bovine Serum Albumin (BSA) Protein Standards was prepared from the content of one BSA ampule. This was diluted into several clean vials according to Table 4.2. The working reagent was prepared in a ratio of 50:1 of BCA reagent A with BCA reagent B, respectively. In addition, a 1:8 ratio of the sample to working reagent was prepared. The samples and standards were incubated on a heating block at 37°C for 30 minutes. The optical densities were measured using the FLUOstar Optima Microplate Fluorometer plate reader (BMG LABTECH, Germany) at an $OD_{562}$, and the standard curve was plotted, from which the concentrations for the protein extracts were calculated.

**Table 4.2**    Preparation of diluted BSA standards for the protein concentration determination, using the BCA protein assay kit.

| Vial | Volume of Dilute (µl) | Volume of BSA (µl) | Final BSA Concentration (µg/ml) |
|------|------------------------|----------------------|----------------------------------|
| A | 0 | 300 of BSA Stock | 2000 |
| B | 125 | 375 of BSA Stock | 1500 |
| C | 325 | 325 of BSA Stock | 1000 |
| D | 175 | 175 of Vial B | 750 |
| E | 325 | 325 of Vial C | 500 |
| F | 325 | 325 of Vial E | 250 |
| G | 325 | 325 of Vial F | 125 |
| H | 400 | 100 of Vial G | 25 |
| I | 400 | 0 | 0 = Blank |

### 4.2.1.3    SDS-PAGE Gel Electrophoresis

The Bio-Rad Mini-PROTEAN TGX precast any kDa gels (Bio-Rad, CA, USA) and the Mini-PROTEAN Tetra Cell Electrophoresis Module (Bio-Rad, CA, USA) were used for the SDS-PAGE separation of the protein extracts. Assembly of the gels and the gel unit was according to the manufacturer's instructions. The gel unit was filled with pre-cooled tank buffer (0.025M Tris, 0.192M glycine, 0.1% (w/v) SDS, pH 8.3). The concentrated 3,000 cutoff protein extracts contained an equal volume of sample buffer (125 mM Tris-HCl, 4% (w/v) SDS, 20% (v/v) glycerol, 0.2M Dithiothreitol (DTT), 0.02% (w/v) bromophenol blue, pH 6.8). Prior to loading onto the gels, the extracts were boiled for 5 minutes, to completely denature the proteins. Electrophoresis was carried out at a final current of 33mA per gel, until the bromophenol blue marker migrated to the bottom of the gel (approximately 28 minutes). Gels were run in duplicate, whereby one gel was stained with agitation over-night in Acqua Stain (Vacutec, South Africa) then replaced with distilled water and further agitated until protein bands were visual and the background sufficiently low. The second gel was used for western blotting.

### 4.2.1.4    Electro Blot Transfer and Immunodetection:

The Bio-Rad mini polyvinylidene (PVDF) transfer pack and the Bio-Rad Trans-Blot Turbo system was used for the protein transfer (Bio-Rad, CA, USA), as per manufacturer's instructions. Transfer was performed by selecting the Mixed MW Turbo option on the transfer unit (25V, 2.5A for 7 minutes).  The PVDF membrane was blocked with agitation in TBS-T buffer (20mM Tris-Cl, 500mM NaCl, 0.1% (v/v) Tween-20, pH 7.5).  Thereafter, the membrane was incubated with the primary antibody diluted in the TBS-T buffer for 1 hour at 37°C. After an additional wash with TBS-T buffer, the membrane was incubated with the horseradish peroxidase-conjugated secondary antibody (AbD Serotec, Oxford, UK) for 1 hour at room temperature with slow agitation. Detection of the secondary antibody was via chemiluminescence using the SuperSignal West Dura Extended Duration Substrate (Pierce, USA) and detected via the CCD camera using the Bio-Rad ChemiDoc MP (Bio-Rad, CA, USA). The CFP-10 and ESAT-6 primary antibodies were kind gifts from Dr. Roland Brosch (Institute Pasteur, Paris, France), and were generated by immunizing rabbits with whole CFP-10 and ESAT-6.

108

The secondary goat anti mouse monoclonal antibodies was obtained from the AbD Serotec (Oxford, UK). Anti-GroEL monoclonal antibody (Hsp-65) was obtained from Abcam (UK) and used in 1:5000 dilution. The CFP-10 primary antibody was used in 1:5000 dilution, while the ESAT-6 primary was used in 1:40 dilution, respectively.

## 4.2.2 MALDI-TOF Mass Spectrometery

### 4.2.2.1   MALDI-TOF Mass Spectrometry Protein Extraction

All *Mtb* isolates and lab strains were cultured to mid log phase of growth in Sauton's liquid medium, of which 1.2ml was used for protein extraction. This culture was centrifuged at maximum speed for 2 minutes, followed by resuspending the pellet in 300µl of deionised MS grade water. To this suspension, 900µl absolute ethanol was added, proceeded by a 10 minute room temperature incubation. Cells were pelleted at maximum speed for 2 minutes and resuspended in 500µl of deionised water followed by an additional centrifugation at maximum speed for 2 minutes. The pelleted cells were resuspended in 50µl of deionised water and heat inactivated by boiling for 30 minutes. After being cooled to room temperature, 1200µl of pre-cooled 100% ethanol was added to the suspension, followed by centrifugation at maximum speed for 2 minutes. The supernatant was discarded and the pellet was air-dried. Zirconia/silica beads (Sigma-Aldrich, Saint Louis, MO) were added to the pellet together with 50µl of 100% acetonitrile. This suspension was bead-beated three times using the MagNa Lyser (Roche Diagnostics South Africa, Randburg, SA) at 7000rpm for 1 minute. Thereafter, 50µl of 70% (v/v) formic acid was added to the suspension and vortexed for 5 seconds. The suspension was centrifuged at maximum speed for 2 minutes and the supernatant as stored at -20°C for further MALDI-TOF analysis.

### 4.2.2.2  MALDI-TOF Protein Analysis

1µl of each protein extract was spotted onto the MTP 384 ground steel plate (Bruker Daltronics, Billerica, MA) using the dried droplet method and allowed to evaporate at room temperature. Firstly, 0.5µl of matrix, saturated HCCA (4-hydroxy-α-cyanocinnamic acid) (Bruker Daltronics, Billerica, MA) in 1ml of 0.1% TFA solution was deposited onto selected wells on the steel plate and allowed to evaporated. This was followed by spotting 1µl of the protein extract onto the already dried matrix and also allowed to evaporate at room temperature. Using the Autoflex III MALDI-rTOF instrument, spectra were obtained in linear mode. For the spectrometer, positive voltage polarity was used. The ion source voltage 1 (IS1) was set at 19.885kV, (IS2) was set at 18.35kV and the lens voltage at 6kV. The pulsed ion extraction delay (PIE) was 150ns with a mass range of 5000-12 000m/z.  A total of 1000 shots at 100Hz were summed for each acquired spot, where one sample was analyzed every 30 seconds. Spectra acquisition was performed using the Flex Analysis software (Bruker Daltronics, Billerica, MA). This experiment was repeated in duplicate on three separate occasions, and the spectra corresponding to the isolates were collected and overlaid.

## 4.3 Results and Discussion

### 4.3.1 Total Protein Concentration

The Bicinchoninic acid (BCA) protein assay method was used to quantify protein concentrations of the protein extracts. This assay exhibits a strong linear absorbance with increasing protein concentration, from which a standard curve was plotted from the average of duplicate protein extracts for both the supernatant and whole cell lysate fractions (Figure 4.1). From the standard curve, the protein concentration for each isolate was calculated and found to be greater than 10μg/μl for all samples (Table 4.3). Therefore, this concentration of protein was used for the SDS-PAGE gels and immune-detection assay.
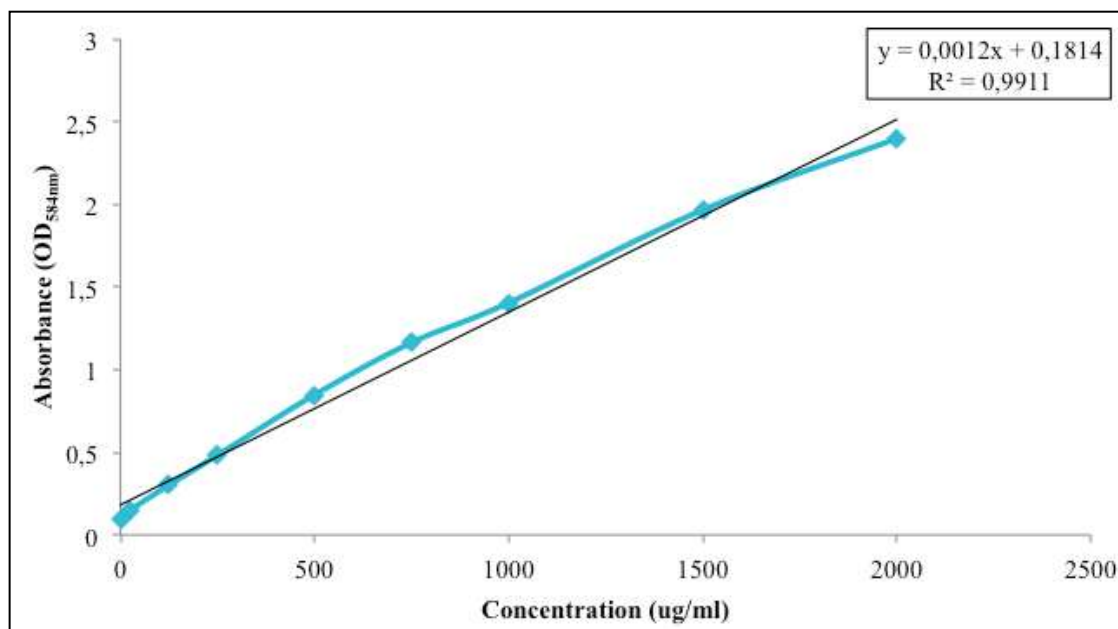


**Figure 4.1**   Standard Curve of Bovine Serum Albumin (BSA) standards (0.5-2000μg/ml). The plots represent the following: Fit Curve (▬) and Raw Data (▬).

**Table 4.3**     Protein Concentrations of *Mtb* Clinical and Laboratory Isolates as determined from a standard curve constructed using the BCA assay.

| Isolates: | OD$_{562nm}$: | Protein Concentration (µg/40µl): |
|---|---|---|
| H37Rv | 2.554 | 11.50 |
| H37Ra | 3.562 | 23.36 |
| BCG | 3.478 | 22.37 |
| 605 | 3.228 | 19.43 |
| 443 | 3.815 | 26.33 |
| 4207 | 3.869 | 26.97 |
| 1528 | 3.379 | 21.20 |
| 62 | 3.836 | 26.92 |
| 26 | 3.572 | 23.47 |
| 299 | 3.446 | 21.99 |
| 3028 | 3.739 | 25.44 |
| 4153 | 3.222 | 19.36 |
| 5539 | 2.93 | 15.92 |
| 3713 | 3.118 | 18.13 |
| 5538 | 3.442 | 21.94 |
| 6103 | 3.123 | 18.19 |
| 3717 | 3.609 | 23.91 |
| 5944 | 3.196 | 19.05 |
| 6885 | 3.798 | 26.13 |
| 5942 | 3.335 | 20.69 |

### 4.3.2 SDS-PAGE *esxAB* Immuno-detection

Following the protein extraction of the culture filtrate and whole cell lysates, SDS protein profiles were established for the selected isolates.

In Figures 4.2 and 4.3, the isolates represent those from the UKZN departmental collection. The protein profiles for these isolates were similar to the isolates from the K-RITH collection (results not shown), and a 65 kDa protein band was prominently observed in all the culture filtrate extracts (4.2a and 4.3a). Protein profiles for the whole cell lysates were consistent and showed a wide range of protein bands in contrast to the culture filtrates. The stained gel images gave reproducible and consistent results for the standardized amount of 10µg/µl protein loaded onto the gel

I then examined both the culture filtrate and whole cell lysate fractions with antibodies specific for ESAT-6 and CFP-10 to detect the presence of these proteins (Figure 4.4). In Figure 4.4, ESAT-6 and CFP-10 was only detected in the whole cell lysate fractions and not in the culture filtrate at the expected molecular weights of 6 kDa and 14.9 kDa, respectively. Blots were also probed with the GroEL antibody, which served as the control for autolysis and a reference to compare ESAT-6 and CFP-10 band intensities. As observed in Figure 4.4, GroEL was found only in the whole cell lysate fraction at the expected 65 KDa molecular weight, thus verifying that the presence of ESAT-6 and CFP-10 in the same fraction but absence in the cell lysate was not as a consequence of cell lysis or leakage. The blots were also overexposed to compare the amounts of released GroEL, but no significant differences in cytolysis were observed (data not shown). My finding is in agreement with previous studies that reported on similar GroEL observations (Stanley *et. al.,* 2003; DiGuiseppe *et. al.,* 2009; Champion *et. al.,* 2012).

Figure 4.4 gives a result compatible with the known deletion of RD1 in BCG (Harboe *et. al.,* 1996; Behr *et. al.*, 1999; Gordon *et. al.,* 1999; Colangeli *et. al.,* 2000; Wards *et. al.,* 2000). Likewise, this absence of bands was seen in H37Ra where ESAT-6 and CFP-10 are not or weakly expressed, however, the expression of GroEL was low in comparison to H37Rv and the other isolates. In 2008, Frigui and colleagues reported on a point mutation

(S219L) in the DNA binding region of the PhoP regulator of H37Ra which impacted on the secretion and reduced immunity of ESAT-6 (Frigui *et. al.,* 2008).

Despite the similarity in the protein profiles for all the isolates (Figures 4.2 and 4.3), there are noticeable differences between the expression levels in the culture filtrates and whole cell lysates blots (Figure 4.4). An example of this is noted in the Unknown Isolate 62 (Figure 4.4), where the expression of ESAT-6 and CFP-10 was increased in the whole cell lysate fraction compared to the other isolates. Likewise, extremely low-level protein expression was observed in the Beijing isolates (1528, 5944, 5538, 6103, 6885 and 3028) and one KZN isolate (5942). The lowest levels of expression were detected in all the Beijing isolates for ESAT-6. In addition, this trend was also evident in a KZN (3713) and Unknown (5539) isolates (Figure 4.4). Levels of CFP-10 immunodetection in the Beijing isolates appeared to be higher than for ESAT-6, in keeping with the expression levels for *esxA* and *esxB* determined by transcriptomics (Figure 3.9).

Variability in GroEL levels was detected for H37Ra and several Beijing Isolates, including 1528 and 6885, respectively. This result was in agreement to the earlier study by Pheiffer and colleagues in 2005, where they showed using Western Blot analysis in Beijing strains, decreased expression of the GroEL protein compared to other isolates and H37Rv, thus concluding this result may be unique to Beijing Isolates (Pheiffer *et. al.,* 2005). GroEL is therefore not an ideal control for these experiments, and ideally it would be necessary to carry out further immunoblot analysis with additional antibodies.
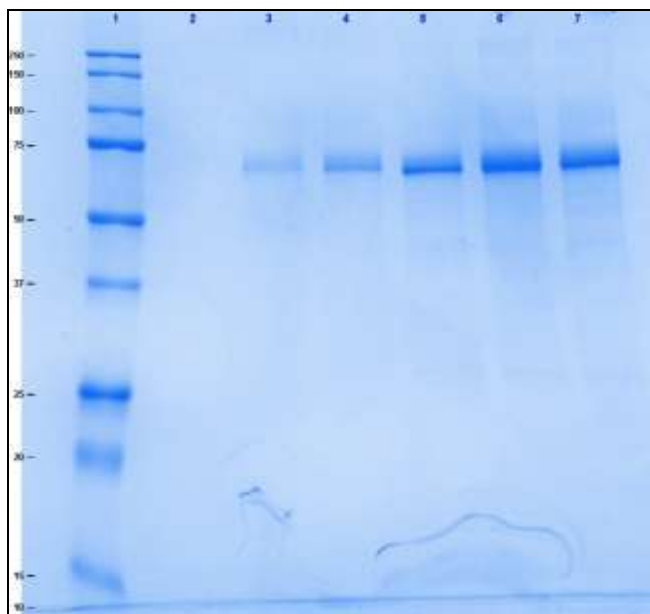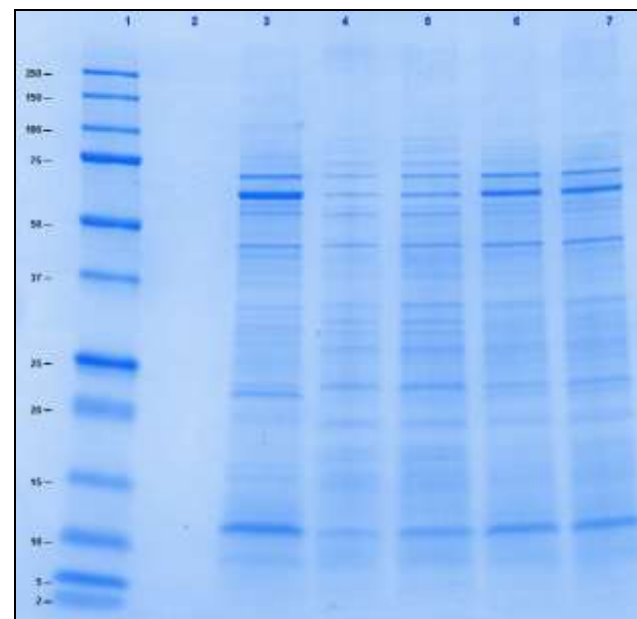
**(a)**                                                                                    **(b)**

**Figure 4.2**    Protein profiles of *Mtb* clinical isolates analysed on the Mini-Protean TGX precast any kDa gels. The gels are represented as follows: **(a)** Culture filtrate extract with 3 kDa cutoff; **(b)** Whole cell lysate with 3 kDa cutoff. Lane 1: BioRad Precision Plus Dual Xtra Standards Molecular Weight Marker. Lane 2: empty. Lanes 3-7: Represent the *Mtb* Clinical Isolates and Laboratory strains. Lane 3: H37Rv; Lane 4: H37Ra; Lane 5: BCG; Lane 6: Isolate 605; Lane 7: Isolate 443.
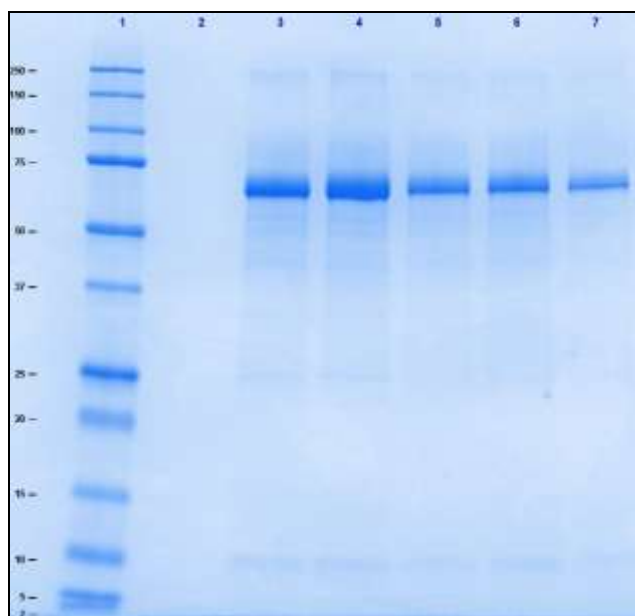
**(a)**  **(b)**
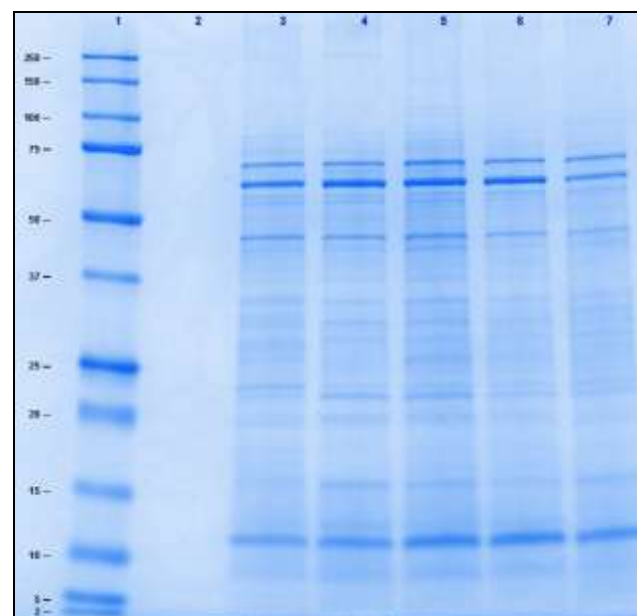
**Figure 4.3**　Protein profiles of *Mtb* clinical isolates analysed on the Mini-Protean TGX precast any kDa gels. The gels are represented as follows: **(a)** Culture filtrate extract with 3 kDa cutoff; **(b)** Whole cell lysate with 3 kDa cutoff. Lane 1: BioRad Precision Plus Dual Xtra Standards Molecular Weight Marker. Lane 2: empty. Lanes 3-7: Represent the *Mtb* Clinical Isolates. Lane 3: Isolate 26; Lane 4: Isolate 62; Lane 5: Isolate 299; Lane 6: Isolate 4207; Lane 7: Isolate 1528.

**Figure 4.4**    Immunoblot analysis, using the CFP-10, ESAT-6 and GroEL antibodies, of whole cell lysate (wcl) and culture filtrate (cf) protein extracts of *Mtb* Clinical and Laboratory cultures.

### 4.3.3 MALDI-TOF

Initially, the whole cell lysate and culture filtrate preparations were analysed using MALDI-TOF MS. I initially tried the protein extraction method in Section 4.2.1.1, lacking the addition of ethanol and formic acid, but no protein spectrum was detected (results not shown). Therefore, the specific extraction method in Section 4.2.2, was appl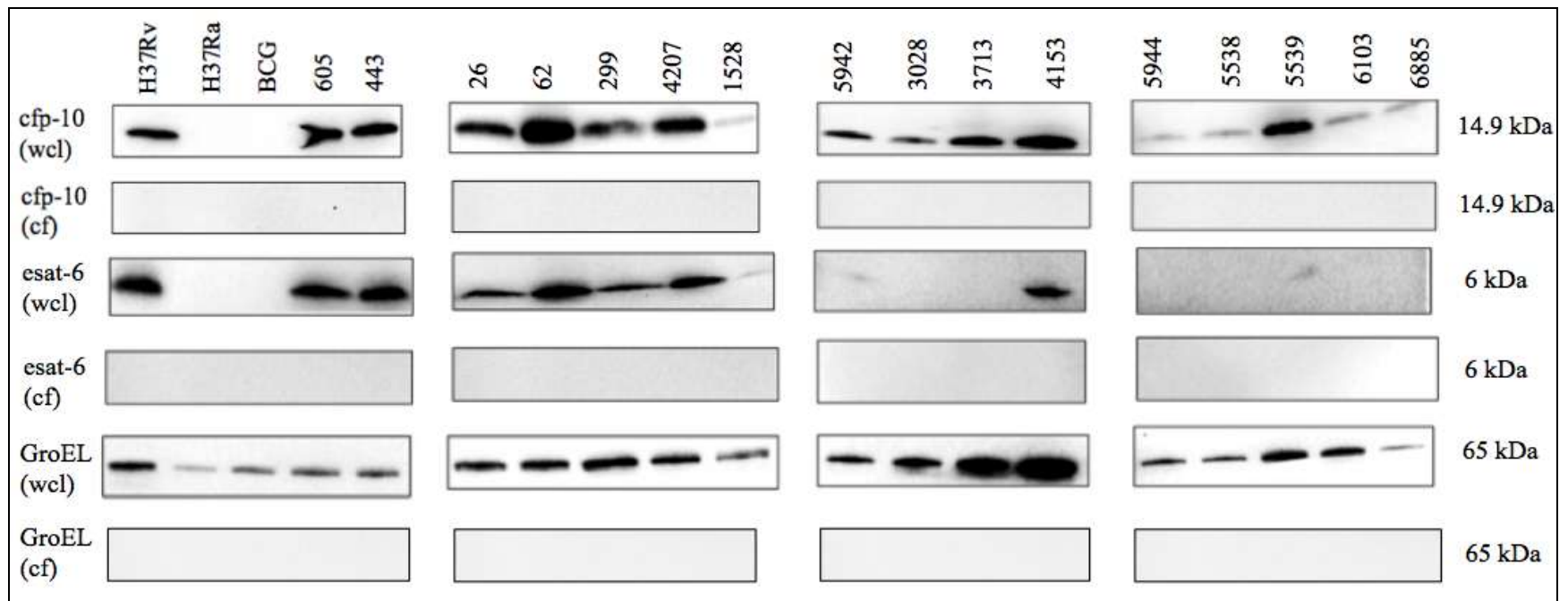ied. The aims were to apply a "Top-Down" approach (identification of small intact proteins without proteolytic digestion) to determine if I could identify differences in expression of ESAT-6 and CFP-10 and other ESX members in selected isolates. I targeted the whole cell fraction as most *esxA* and *B* is found in this fraction rather than the filtrate fraction.

Figure 4.5 is a representation of the protein spectral profiles for H37Rv, H37Ra and BCG. I reasoned that if I could see a difference in the protein spectra between H37Rv, H37Ra and BCG, it would potentially identify a peak that could be used to compare clinical isolates. This approach is only semiquantitative, but I reasoned that it would serve as a secondary method of protein detection to confirm the results of the immunoblot assays described above. BCG and H37Ra would act as my control for the experiment, due to the deletion of ESAT-6 and CFP-10 or their lack of expression in these two strains.

Spectra are shown in Figure 4.5a, for my 3 reference strains. Remarkably, a single peak (10632 Da) was detected in the H37Rv protein spectrum, but was absent in both H37Ra (Figure 4.5b) and BCG (Figure 4.5c). This peak is within range of the alleged size reported by Champion and colleagues in 2012, of 10609 Da for CFP-10 (Champion *et. al.,* 2012). The overall intensities of the peaks varied between runs so an overlay was used that clearly shows that other peaks are conserved between strains but the identified peak was absent in H37Ra and BCG (Figure 4.5d). Intriguingly, there was no corresponding peak for ESAT-6 that is expected to be less than 10000 Da. Ionization of whole proteins is unpredictable and non quantitative in these types of MALDI-TOF experiments, with certain proteins being preferentially ionized. However, if the peak observed is CFP-10 then the absence of a paired peak corresponding to ESAT-6 is

compatible with my observations in immunoblot and transcriptomic experiments that CFP-10 is found in excess in whole cell preparations relative to ESAT-6. There are several explanations for this including the relative stability of the two proteins and the preferential export of ESAT-6 into the culture filtrate fraction.

The putative CFP-10 peak was present in the spectral protein overlays for the Bejing clinical isolates (Figure 4.6), the KZN clinical isolates (Figure 4.7) and 2of the Other clinical isolates (Figure 4.8). As the technique is only semi-quantitative it is difficult to make absolute comparisons between strains. However, H37Rv had a noticeably higher spectral intensity as compared to the other isolates, except for Figure 4.7, where an Other Isolate 62 displayed a higher spectral intensitiy profile than H37Rv. Interestingly, in the immunoblot experiments (Figure 4.4), I observed the strongest CFP-10 antibody response in Isolate 62 compared to the remainder of the isolates tested. This result correlates to the MALDI-TOF MS protein profile for CFP-10 (Figure 4.9a). In addition the qPCR experiments indicated a high level of expression of *esxAB* for H37Rv. The profiles of the Beijing strains also show that all the strains had a putative CFP-10 size peak that was less than the peak of approximately 7200 Da. In contrast for the KZN strains, the peak in the region of 10609 Da was less than their peaks at 7200 Da. As previously stated the technique is only semi-quantitative, but these results do support clade specific differences in protein regulation.

There are some important limitations to these experiments and it is acknowledged that these data are preliminary and experiments had to be stopped before completion due to equipment breakage at the Westville MALDI-TOF facility. Firstly, I have no guarantee that the peak is actually CFP-10. I tried running a commercial available ESAT-6 recombinant protein but was unsuccessful in obtaining a standard peak with this preparation. Ideally, I would also liked to have tried a similar standard for CFP-10, although post-translational modification could affect the masses of the naturally expressed proteins. For more accurate quantification, in future research it would be advisable to add a quantity of known protein to each sample preparation to act as a standard. In addition it should be possible to focus the mass spectroscopy to a narrower

bandwidth. Finally, further optimization of protein extraction could potentially reduce the background signal.
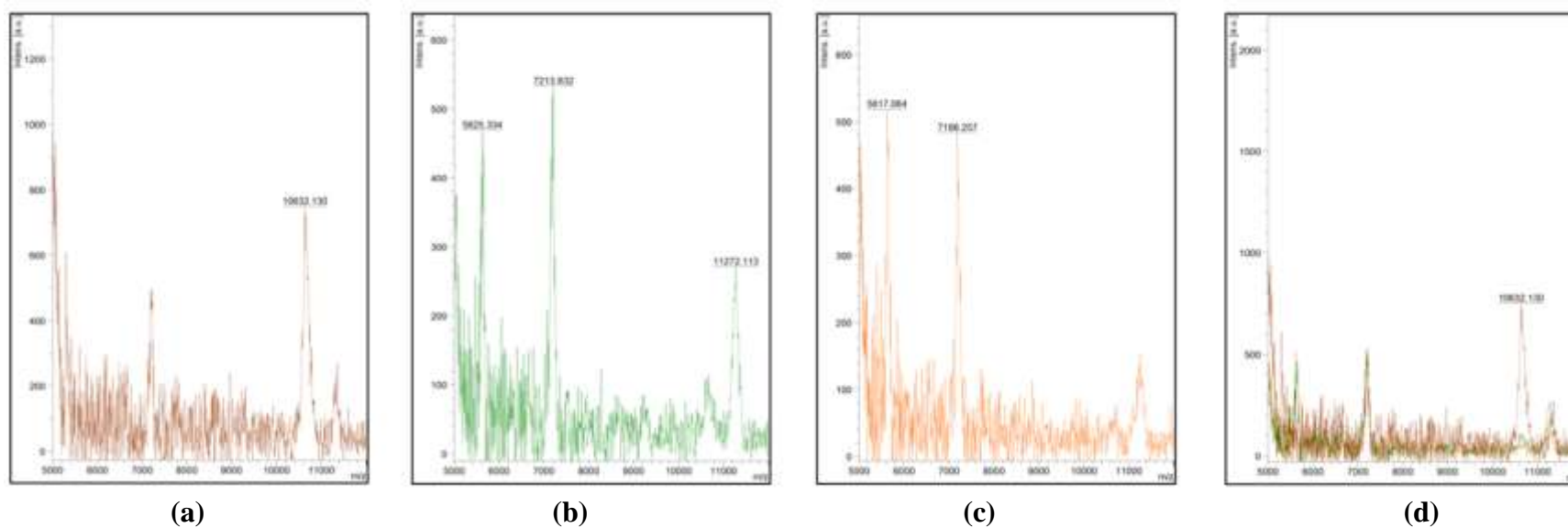
**Figure 4.5** MALDI-TOF Spectra of **(a)** H37Rv, **(b)** H37Ra, **(c)** BCG and **(d)** Overlay of H37Rv, H37Ra and BCG protein spectral profiles. The 10632 Da peak is present in H37Rv but is absent in H37Ra and BCG.
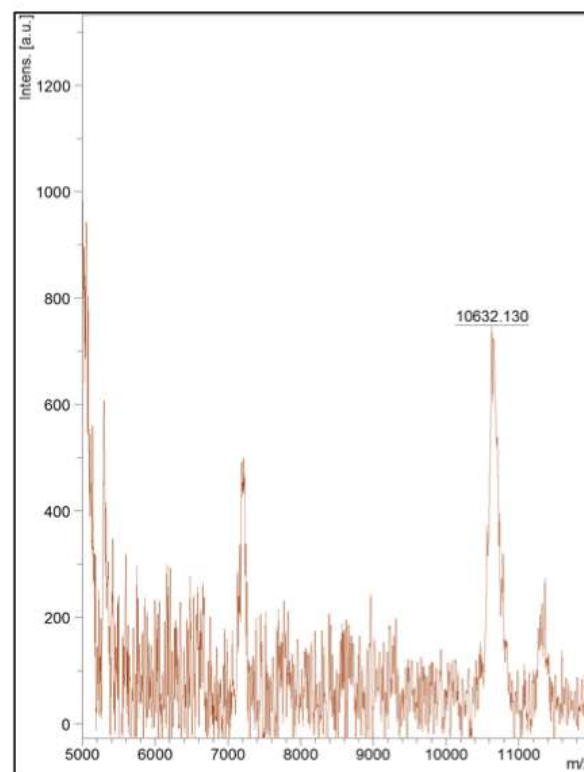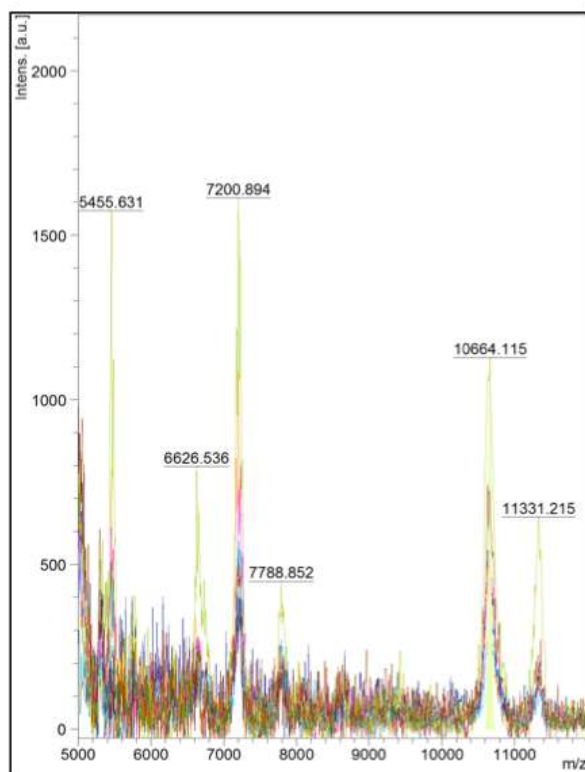
**Figure 4.6**    MALDI-TOF Spectra of **(a)** Overlay of all the Beijing Isolates protein spectral profiles and **(b)** H37Rv. A 10632 Da peak is present in H37Rv and 10664 Da peak is present for the highest peak for the Beijing Isolates.
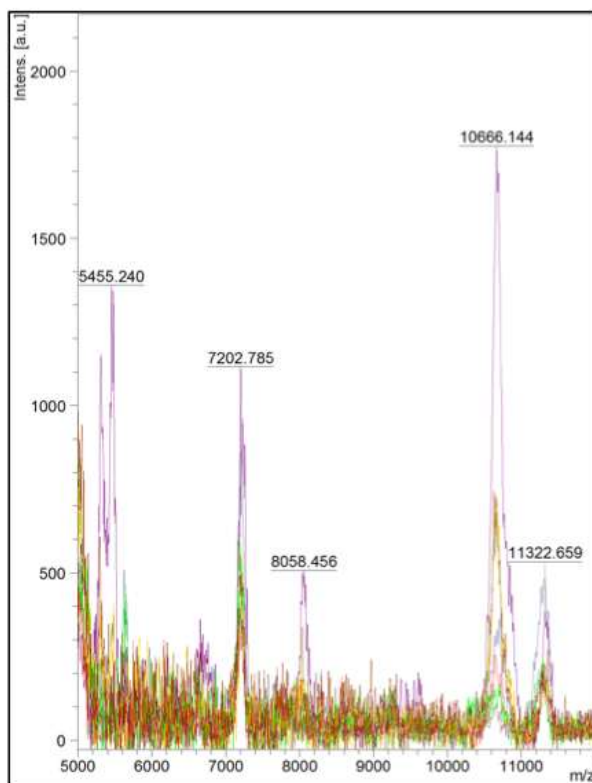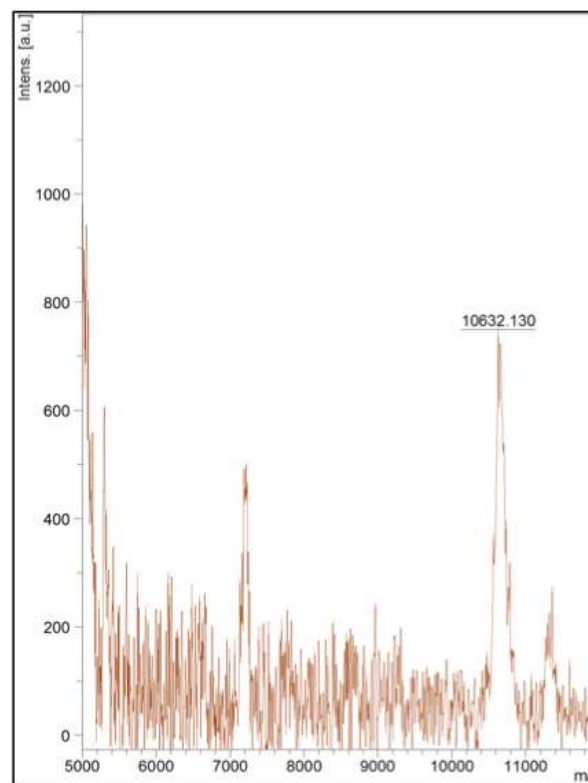
**Figure 4.7** MALDI-TOF Spectra of **(a)** Overlay of all the KZN Isolates protein spectral profiles and **(b)** H37Rv. A 10632 Da peak is present in H37Rv and 10666 Da peak is present for the highest peak for the KZN Isolates.
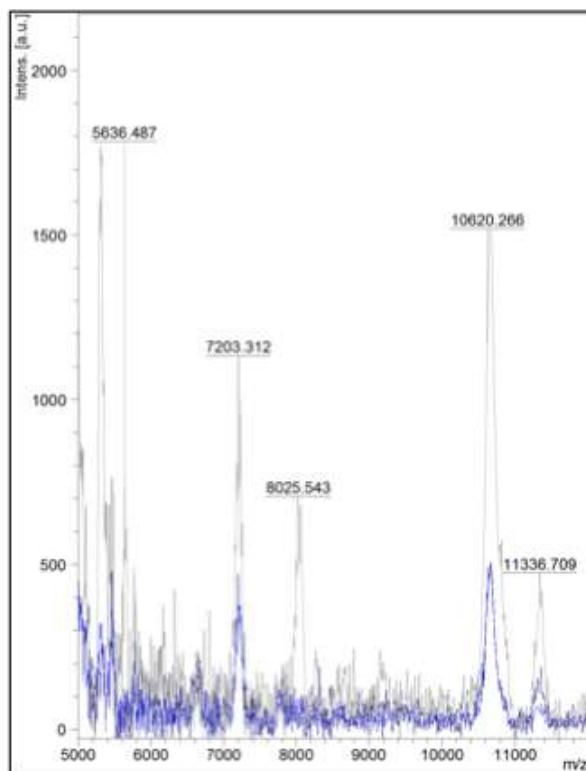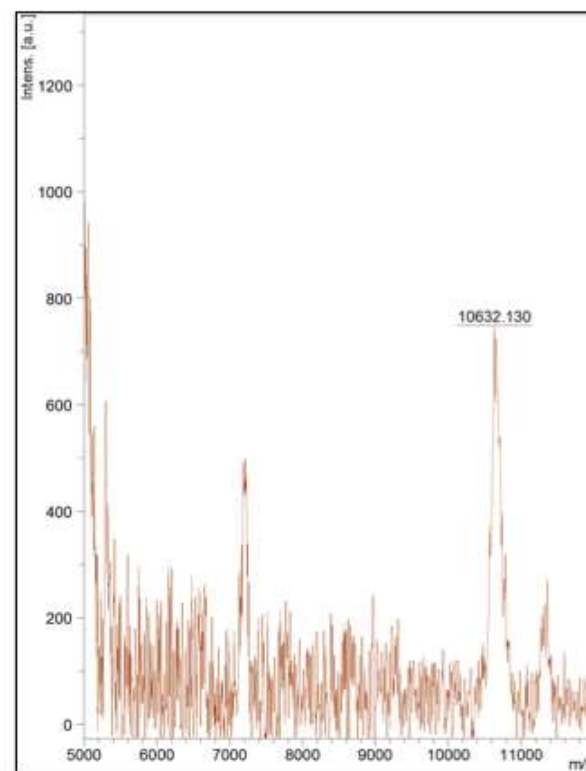
(a)                                                                                      (b)

**Figure 4.8**      MALDI-TOF Spectra of **(a)** Overlay of 2 Other Isolates protein spectral profiles and **(b)** H37Rv. A 10632 Da peak is present in H37Rv and 10620 Da peak is present for the highest peak for the Unknown Isolates.

124
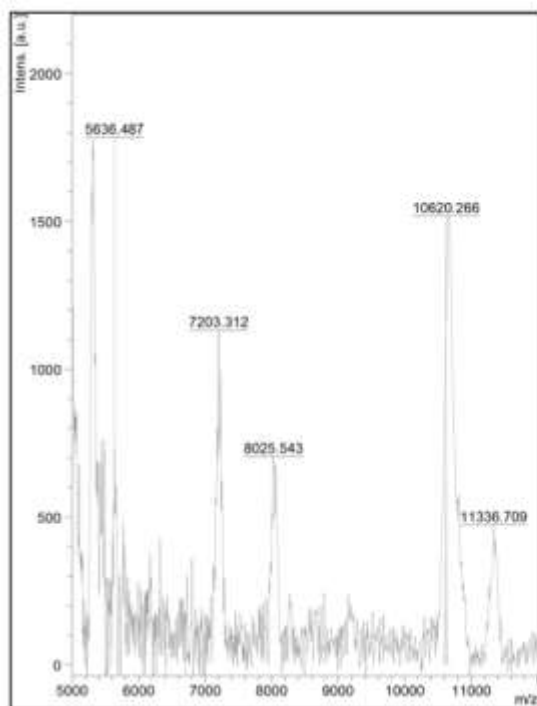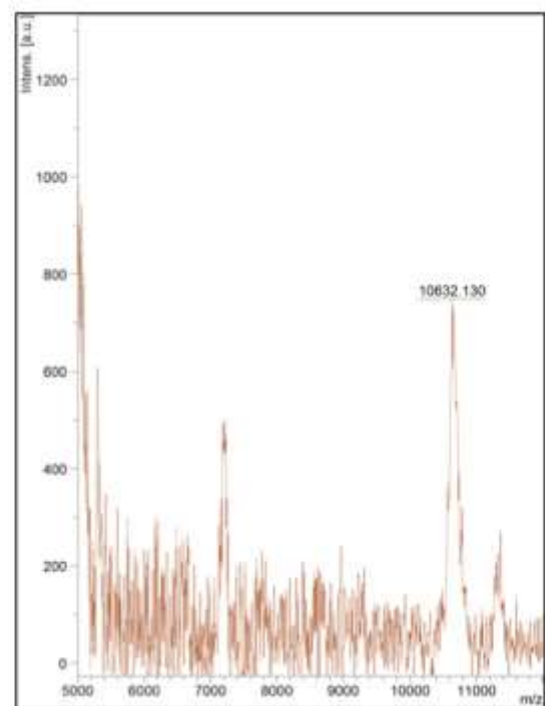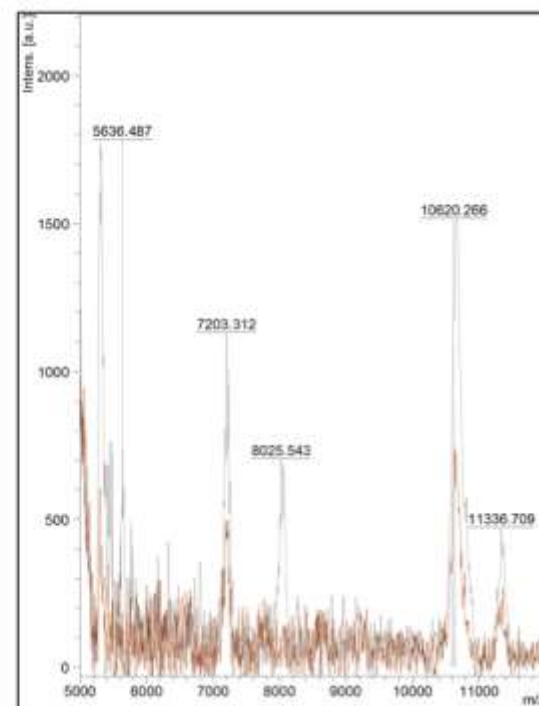
**Figure 4.9** MALDI-TOF Spectra of **(a)** Unkown Isolate 62 with **(b)** H37Rv and **(c)** Overlay of H37Rv and 62 protein spectral profiles. A 10632 Da peak is present in H37Rv and 10620 Da peak is present for Unknown Isolate 62.

## 4.4 Conclusion

In this study I investigated the protein profiles in selected *Mtb* clinical and laboratory isolates for the presence of ESAT-6 and CFP-10. With the immunoblots, I was able to detect differences in the proteins of interest. These proteins were detected only in the whole cell lysate fraction, whereas the culture filtrate fractions displayed no presence at all at the volumes of culture from which I concentrated proteins. Previous experiments have quantified the amount of protein in the supernatant fraction and have detected ESAT-6 and CFP-10. However, these experiments have required large-scale concentration of proteins. The previous works from the laboratories of Cole and Fortune have shown that most of these 2 ESX members are concentrated in the cell wall fraction. I therefore focussed on the cellular compartment reasoning that quantification of protein should be done in the fractionwhere the protein was most abundant.

I noted decreased expression of ESAT-6 and CFP-10 across all the Bejing Isolates. Interestingly, in some Beijing isolates, I noticed the reduced expression of GroEL, such as in Isolate 1528, and this has been previously suggested by Stewart and colleagues (Stewart *et. al.,* 2001). Ideally, I should have repeated the experiments with an antibody other than GroEL, where there is no difference in expression. However, I used careful protein quantification, gel staining and repeat experiments to ensure that the quantity of protein subject to imunne blotting didn't vary between strains. Confirmation of a previous finding by my immunoblotting is reassuring that the differences in expression that were detected are likely to reflect.

Even though optimization was not complete, I had preliminary data that suggested a modified whole colony MALDI-TOF MS could be used to directly detect and compare molecules of ESX protein size in whole cell lysate extracts. Where results were available, I was able to correlate the protein spectra profiles for an isolate to the immunoblot result in Figure 4.4. It is unclear why a peak corresponding to the mass of CFP-10 is so pronounced in my protein spectra, as usually one would only expect ribosomal proteins or other proteins to be the most pronounced peaks. Further work is evidently required to refine this technique to make it truly quantitative and discriminative.

126

Previous studies have reported on the implementation of this technology in the identification of bacterial strains routinely in a clinical laboratory setting (Bessède *et. al.,* 2011; Bizzini *et. al.,* 2010; Neville *et. al.,* 2011; Sogawa *et. al.,* 2011). However, a 2004 study by Hettick and colleagues demonstrated that using the whole colony MALDI-TOF MS method in mycobacteria does not always kill the mycobacterial colonies. Therefore, the implementation of this approach directly to identify a BSL3 organism is difficult (Hettick *et. al.,* 2004). But for proteomic comparisons of pathogenic bacteria grown under BSL3 conditions, and filter sterilized the technique is very promising.

Collectively, the results presented in this study section are in agreement with transcriptional analysis, which indicates there is variability in the expression of ESAT-6 and CFP-10 in cell culture between strains. Potentially this has implications for immunodiagnostic and vaccine design, although it would be important to determine if difference in protein expression also existed when bacteria are grown *in vivo* systems.

# CHAPTER FIVE

# ELISPOT Responses in Healthy and TB

# Positive Donors for QILSS and Mtb9.9

# Peptides

## 5.1   Chapter Summary

The main objective of this chapter was to characterize T-lymphocyte responses to ESX peptides from the QILLSS and Mtb9.9 proteins in patients infected with TB. Due to the concentration of non-synonymous SNPs in these two protein groups, I wanted to determine experimentally which regions of these protein families were immunogenic. It is expected that patients actively infected with *M. tuberculosis* would have had prolonged and significant exposure to *M. tuberculosis* and therefore would most likely to mount immune responses to a broad range of antigens. Having shown there is variable expression in other members of the ESX family of proteins, I hypothesized that the QILLSS and Mtb9.9 protein families may also be subject to similar variation. Given the limited genetic variability between individual members of the QILLSS and Mtb9.9 proteins I was particularly interested to determine if the variable residues would influence the immune responses. Overlapping peptides, incorporating the regions of variability were therefore assayed using an ELISPOT assay with PBMCs from healthy donors and patients actively infected with tuberculosis. Healthy patients who were ELISPOT negative for standard pools of EsxA and EsxB peptides, where found not to have responses to the QILSS and Mtb9.9 peptides suggesting there was no cross reactivity with non tuberculosis antigens. In contrast when QILSS and Mtb9.9 peptides were tested individually for immune responses in tuberculosis infected patients, strong but variable immune responses were detected for specific regions of the ESX subfamilies.

## 5.2 Materials and Method

### 5.2.1 Study Partcipants

PBMCs were isolated from a total of 43 individuals. Sixteen were healthy volunteer donors routinely donating blood (SANBS). Buffy coats, a by-product of red cell donation, were collected and used for PBMC extraction. Individuals present regularly to SANBS to donate blood and are evaluated with a symptom screen prior to each blood donation. In addition they have their temperature monitored. Any signs of fever or reported symptoms suggestive of concurrent illness will result in the individuals not being sampled on that day. All individuals are regularly checked for HIV infection. Although tuberculosis can present as an occult infection, it is unusual in HIV uninfected individuals. I therefore considered this group of individuals as having an extremely low probability of active tuberculosis. Subsequent testing with conventional ELISPOT assay for EsxA and EsxB immune responses subclassifed this group into latently infected and not latently infected. The remaining 27 TB positive patients came from an ongoing cohort study based at King Dinizulu Hospital. These patients were being initiated for MDR-TB treatment and were defined as having TB based on either a TB specific nucleic amplification assay or by isolation of *M. tuberculosis* using the standard MIGIT liquid culture system. A minimum of 20ml of blood was drawn from individuals that provided written informed consent and were processed at the K-RITH processing laboratory, from which a target of $30\times10^{6}$ PBMCs were provided for this study.

### 5.2.2 Peptide Design and Production

In this part of my study, the aim was to determine how the limited variability in the QILSS and Mtb9.9 protein families influenced the immune response. As an initial stage I wanted to fully characterize the sequences derived from H37Rv. A second planned stage would be to include the variability I identified amongst clinical isolates. Both these families only differ in sequence at several codons, and I wanted to determine if specific amino acid sequences influenced the antigenicity of these proteins. Overlapping peptides spanning the QILSS and Mtb9.9 proteins were designed for all the ESX members that are part of these subfamilies. Where there was no sequence variation a single peptide was designed. In the case of sequence variation multiple peptides were designed that captured each of the variable amino acid sequence in the protein family members (Figure 5.1a). A total of 67 peptides were designed (18 amino acid residues long with a 11 amino acid residue overlap) representing the QILSS and Mtb9.9 (Table 5.1) members. ESAT-6 and CFP-10 peptides were 15 amino acid residues in length with a 10 amino acid residue overlap (Table 5.2) and were those previously used in other studies (Lalvani *et. al.,* 2001a; Lalvani *et. al.,* 2001b; Vordermeier *et. al.,* 2001; Arlehamn *et. al.,* 2012). This length of peptide is thought to be optimal for CD4 responses. All peptides were designed using the Geneious software package and commercially synthesized by Pepceuticals Limited (Pepceuticals Limited, Leicestershire, England). Each individual peptide was dissolved in RPMI 1640 (BioWhittaker, Lonza, Belgium) with 20% dimethyl sulfoxide (DMSO) to a final stock concentration of 5mg/ml. These peptides were then used to stimulate fresh PBMCs at a final concentration of 10μg/ml. Similarly, the individual ESAT-6 and CFP-10 peptides were dissolved in RPMI 1640 containing 20% DMSO to stock concentration of 5mg/ml. The individual peptides were then pooled to obtain a peptide cocktail and each pool of ESAT-6 and CFP-10 was used at a final concentration of 8μg/ml in keeping with previous standard protocols for ELISPOT assays. This particular peptide pool was included in the ELISPOT assay as the known standard to act as a positive control and ensure that all individuals mounted a response against *Mtb* antigens and to compare with the strength of the responses of the QILSS and Mtb9.9 peptides.
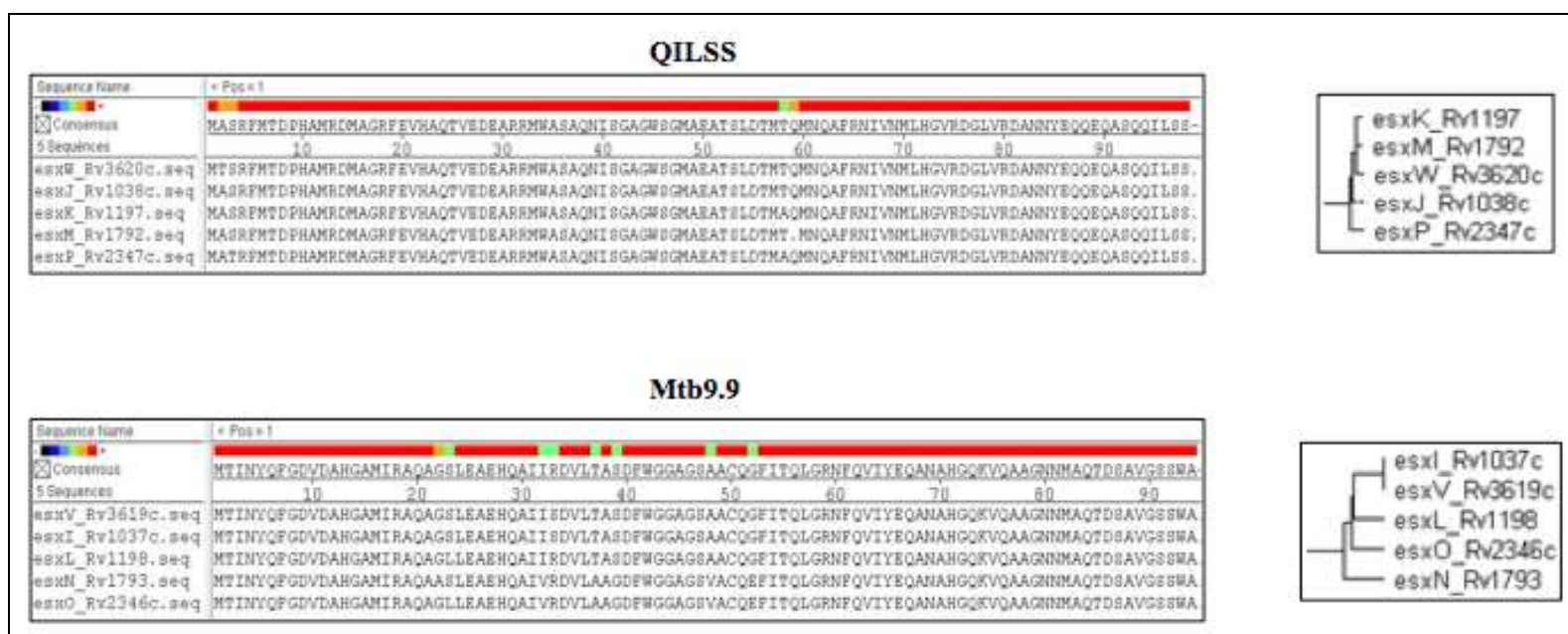
**Figure 5.1a** Protein and Phylogenetic tree alignments of QILSS and Mtb9.9 ESX members using the DNAStar MegAlign and Treeview software packages. Red: sequences are identical; Orange and Green: regions containing SNPs.

**Table 5.1**    Epitope Sequences and Amino Acid Range with SNP Permutation in red font for QILSS and Mtb9.9 Peptides.

| Number | Subfamily | Region ID | A.A. Range | SNP Loci | SNP Permutation | Peptide Name | Sequence |
|---|---|---|---|---|---|---|---|
| 1 | QILSS | 1 | 1-18 | 2, 3 | A, S | Q01.1 | MASRFMTDPHAMRDMAGR |
| 2 | QILSS | 1 | 1-18 | 2, 3 | T, S | Q01.2 | MTSRFMTDPHAMRDMAGR |
| 3 | QILSS | 1 | 1-18 | 2, 3 | A, T | Q01.3 | MATRFMTDPHAMRDMAGR |
| 4 | QILSS | 2 | 7-24 | None | None | Q02.1 | TDPHAMRDMAGRFEVHAQ |
| 5 | QILSS | 3 | 13-30 | None | None | Q03.1 | RDMAGRFEVHAQTVEDEA |
| 6 | QILSS | 4 | 19-36 | None | None | Q04.1 | FEVHAQTVEDEARRMWAS |
| 7 | QILSS | 5 | 25-42 | None | None | Q05.1 | TVEDEARRMWASAQNISG |
| 8 | QILSS | 6 | 31-48 | 48 | M | Q06.1 | RRMWASAQNISGAGWSGM |
| 9 | QILSS | 6 | 31-48 | 48 | Q | Q06.2 | RRMWASAQNISGAGWSGQ |
| 10 | QILSS | 7 | 37-54 | 48 | M | Q07.1 | AQNISGAGWSGMAEATSL |
| 11 | QILSS | 7 | 37-54 | 48 | Q | Q07.2 | AQNISGAGWSGQAEATSL |
| 12 | QILSS | 8 | 43-60 | 48, 58 | M, T | Q08.1 | AGWSGMAEATSLDTMTQM |
| 13 | QILSS | 8 | 43-60 | 48, 58 | M, A | Q08.2 | AGWSGMAEATSLDTMAQM |
| 14 | QILSS | 8 | 43-60 | 48, 58 | Q, T | Q08.3 | AGWSGQAEATSLDTMTQM |
| 15 | QILSS | 9 | 49-66 | 58 | A | Q09.1 | AEATSLDTMAQMNQAFRN |
| 16 | QILSS | 9 | 49-66 | 58 | T | Q09.2 | AEATSLDTMTQMNQAFRN |
| 17 | QILSS | 10 | 55-72 | 58 | A | Q10.1 | DTMAQMNQAFRNIVNMLH |
| 18 | QILSS | 10 | 55-72 | 58 | T | Q10.2 | DTMTQMNQAFRNIVNMLH |
| 19 | QILSS | 11 | 60-77 | None | None | Q11.1 | MNQAFRNIVNMLHGVRDG |

| 20 | QILSS | 12 | 67-84 | None | None | Q12.1 | IVNMLHGVRDGLVRDANN |
| 21 | QILSS | 13 | 73-90 | None | None | Q13.1 | GVRDGLVRDANNYEQQEQ |
| 22 | QILSS | 14 | 81-98 | None | None | Q14.1 | DANNYEQQEQASQQILSS |
| 34 | Mtb9.9 | 1 | 1-18 | 12 | A | M01.1 | MTINYQFGDVDAHGAMIR |
| 35 | Mtb9.9 | 1 | 1-18 | 12 | D | M01.2 | MTINYQFGDVDDHGAMIR |
| 36 | Mtb9.9 | 2 | 7-24 | 12, 20, 22, 23 | A, Q, A, S | M02.1 | FGDVDAHGAMIRAQAASL |
| 37 | Mtb9.9 | 2 | 7-24 | 12, 20, 22, 23 | A, Q, A, L | M02.2 | FGDVDAHGAMIRAQAALL |
| 38 | Mtb9.9 | 2 | 7-24 | 12, 20, 22, 23 | A, Q, G, S | M02.3 | FGDVDAHGAMIRAQAGSL |
| 39 | Mtb9.9 | 2 | 7-24 | 12, 20, 22, 23 | A, Q, G, L | M02.4 | FGDVDAHGAMIRAQAGLL |
| 40 | Mtb9.9 | 2 | 7-24 | 12, 20, 22, 23 | A, L, G, S | M02.5 | FGDVDAHGAMIRALAGSL |
| 41 | Mtb9.9 | 2 | 7-24 | 12, 20, 22, 23 | A, L, G, L | M02.6 | FGDVDAHGAMIRALAGLL |
| 42 | Mtb9.9 | 2 | 7-24 | 12, 20, 22, 23 | D, Q, G, L | M02.7 | FGDVDDHGAMIRAQAGLL |
| 43 | Mtb9.9 | 3 | 13-30 | 20, 22, 23 | Q, A, S | M03.1 | HGAMIRAQAASLEAEHQA |
| 44 | Mtb9.9 | 3 | 13-30 | 20, 22, 23 | Q, A, L, | M03.2 | HGAMIRAQAALLEAEHQA |
| 45 | Mtb9.9 | 3 | 13-30 | 20, 22, 23 | Q, G, S | M03.3 | HGAMIRAQAGSLEAEHQA |
| 46 | Mtb9.9 | 3 | 13-30 | 20, 22, 23 | Q, G, L | M03.4 | HGAMIRAQAGLLEAEHQA |
| 47 | Mtb9.9 | 3 | 13-30 | 20, 22, 23 | L, G, L | M03.5 | HGAMIRALAGLLEAEHQA |
| 48 | Mtb9.9 | 3 | 13-30 | 20, 22, 23 | L, G, S | M03.6 | HGAMIRALAGSLEAEHQA |
| 49 | Mtb9.9 | 4 | 19-36 | 20, 22, 23, 32, 33 | Q, A, S, V, R | M04.1 | AQAASLEAEHQAIVRDVL |
| 50 | Mtb9.9 | 4 | 19-36 | 20, 22, 23, 32, 33 | Q, G, L, I, R | M04.2 | AQAGLLEAEHQAIIRDVL |

| 51 | Mtb9.9 | 4 | 19-36 | 20, 22, 23, 32, 33 | Q, G, L, I, S | M04.3 | AQAGLLEAEHQAIISDVL |
| 52 | Mtb9.9 | 4 | 19-36 | 20, 22, 23, 32, 33 | Q, G, S. I, R | M04.4 | AQAGSLEAEHQAIIRDVL |
| 53 | Mtb9.9 | 4 | 19-36 | 20, 22, 23, 32, 33 | Q, G, S, I, S | M04.5 | AQAGSLEAEHQAIISDVL |
| 54 | Mtb9.9 | 4 | 19-36 | 20, 22, 23, 32, 33 | L, G, L, I, R | M04.6 | ALAGLLEAEHQAIIRDVL |
| 55 | Mtb9.9 | 4 | 19-36 | 20, 22, 23, 32, 33 | L, G, L, I, S | M04.7 | ALAGLLEAEHQAIISDVL |
| 56 | Mtb9.9 | 4 | 19-36 | 20, 22, 23, 32, 33 | L, G, L, V, R | M04.8 | ALAGLLEAEHQAIVRDVL |
| 57 | Mtb9.9 | 4 | 19-36 | 20, 22, 23, 32, 33 | L, G, S, I, S | M04.9 | ALAGSLEAEHQAIISDVL |
| 58 | Mtb9.9 | 5 | 25-42 | 32, 33, 37, 39 | I, R, T, S | M05.1 | EAEHQAIIRDVLTASDFW |
| 59 | Mtb9.9 | 5 | 25-42 | 32, 33, 37, 39 | I, S, T, S | M05.2 | EAEHQAIISDVLTASDFW |
| 60 | Mtb9.9 | 5 | 25-42 | 32, 33, 37, 39 | V, R, A, G | M05.3 | EAEHQAIVRDVLAAGDFW |
| 61 | Mtb9.9 | 5 | 25-42 | 32, 33, 37, 39 | V, S, A, G | M05.4 | EAEHQAIVSDVLAAGDFW |
| 62 | Mtb9.9 | 6 | 31-48 | 32, 33, 37, 39, 48 | I, R, T, S, A | M06.1 | IIRDVLTASDFWGGAGSA |
| 63 | Mtb9.9 | 6 | 31-48 | 32, 33, 37, 39, 48 | I, S, T, S, A | M06.2 | IISDVLTASDFWGGAGSA |

| 64 | Mtb9.9 | 6 | 31-48 | 32, 33, 37, 39, 48 | V, R, A, G, V | M06.3 | IVRDVLAAGDFWGGAGSV |
| 65 | Mtb9.9 | 6 | 31-48 | 32, 33, 37, 39, 48 | V, S, A, G, V | M06.4 | IVSDVLAAGDFWGGAGSV |
| 66 | Mtb9.9 | 7 | 37-54 | 37, 39, 48, 52 | T, S, A, G | M07.1 | TASDFWGGAGSAACQGFI |
| 67 | Mtb9.9 | 7 | 37-54 | 37, 39, 48, 52 | A, G, V, G | M07.2 | AAGDFWGGAGSVACQGFI |
| 68 | Mtb9.9 | 7 | 37-54 | 37, 39, 48, 52 | A, G, V, E | M07.3 | AAGDFWGGAGSVACQEFI |
| 69 | Mtb9.9 | 8 | 43-60 | 48, 52 | A, G | M08.1 | GGAGSAACQGFITQLGRN |
| 70 | Mtb9.9 | 8 | 43-60 | 48, 52 | V, G | M08.2 | GGAGSVACQGFITQLGRN |
| 71 | Mtb9.9 | 8 | 43-60 | 48, 52 | V, E | M08.3 | GGAGSVACQEFITQLGRN |
| 72 | Mtb9.9 | 9 | 49-66 | 52 | G | M09.1 | ACQGFITQLGRNFQVIYE |
| 73 | Mtb9.9 | 9 | 49-66 | 52 | E | M09.2 | ACQEFITQLGRNFQVIYE |
| 74 | Mtb9.9 | 10 | 55-72 | None | None | M10.1 | TQLGRNFQVIYEQANAHG |
| 75 | Mtb9.9 | 11 | 61-78 | None | None | M11.1 | FQVIYEQANAHGQKVQAA |
| 76 | Mtb9.9 | 12 | 68-85 | None | None | M12.1 | ANAHGQKVQAAGNNMAQT |
| 77 | Mtb9.9 | 13 | 74-91 | None | None | M13.1 | KVQAAGNNMAQTDSAVGS |
| 78 | Mtb9.9 | 14 | 77-94 | None | None | M14.1 | AAGNNMAQTDSAVGSSWA |

**Table 5.2**     List of ESAT-6 and CFP-10 Peptides and Peptide Sequences.

| Peptide Name | Sequence |
| --- | --- |
| ESAT-6 (01) | MTEQQWNFAGIEAAA |
| ESAT-6 (02) | WNFAGIEAAASAIQG |
| ESAT-6 (03) | IEAAASAIQGNVTSI |
| ESAT-6 (04) | SAIQGNVTSIHSLLD |
| ESAT-6 (05) | NVTSIHSLLDEGKQS |
| ESAT-6 (06) | HSLLDEGKQSLTKLA |
| ESAT-6 (07) | EGKQSLTKLAAAWGG |
| ESAT-6 (08) | LTKLAAAWGGSGSEA |
| ESAT-6 (09) | AAWGGSGSEAYQGVQ |
| ESAT-6 (10) | SGSEAYQGVQQKWDA |
| ESAT-6 (11) | YQGVQQKWDATATEL |
| ESAT-6 (12) | QKWDATATELNNALQ |
| ESAT-6 (13) | TATELNNALQNLART |
| ESAT-6 (14) | NNALQNLARTISEAG |
| ESAT-6 (15) | NLARTISEAGQAMAS |
| ESAT-6 (16) | ISEAGQAMASTEGNV |
| ESAT-6 (17) | QAMASTEGNVTGMFA |
| CFP-10 (01) | MAEMKTDAATLAQEA |
| CFP-10 (02) | TDAATLAQEAGNFER |
| CFP-10 (03) | LAQEAGNFERISGDL |
| CFP-10 (04) | GNFERISGDLKTQID |
| CFP-10 (05) | ISGDLKTQIDQVEST |
| CFP-10 (06) | KTQIDQVESTAGSLQ |
| CFP-10 (07) | QVESTAGSLQGQWRG |
| CFP-10 (08) | AGSLQGQWRGAAGTA |
| CFP-10 (09) | GQWRGAAGTAAQAAV |
| CFP-10 (10) | AAGTAAQAAVVRFQE |
| CFP-10 (11) | AQAAVVRFQEAANKQ |

| | |
|---|---|
| **CFP-10 (12)** | VRFQEAANKQKQELD |
| **CFP-10 (13)** | AANKQKQELDEISTN |
| **CFP-10 (14)** | KQELDEISTNIRQAG |
| **CFP-10 (15)** | EISTNIRQAGVQYSR |
| **CFP-10 (16)** | IRQAGVQYSRADEEQ |
| **CFP-10 (17)** | VQYSRADEEQQQALS |
| **CFP-10 (18)** | ADEEQQQALSSQMGF |

## 5.2.3 Lymphocyte Separation for ELISPOT

Peripheral blood mononuclear cells (PBMC) were separated from 50ml of blood by an initial centrifugation at 1600 rpm for 10 minutes. The plasma layer was removed and added in a 1:1 ratio to PBS supplemented with antibiotics. The mixture was layered onto histopaque (Sigma-Aldrich, Saint Louis, MO) followed by a 30 minute centrifugation at 1600rpm. Thereafter, the PBMC layer was carefully removed with a Pasteur pipette into a sterile 50ml conical tube to which an equal volume of PBS, supplemented with antibiotics, was added. This mixture was centrifuged for 10 minutes at 1600rpm and the resultant pellet was resuspened in 20ml of R10 media (RPMI 1640 media supplemented with 2mM-L-glutamine, 100mg/mL ampicillin and 50 mg/mL gentamicin, and 10% heat-inactivated fetal calf serum) (Sigma-Aldrich, Saint Louis, MO). Cells were manually counted in a haemocytometer in a trypan blue suspension, and 20 million cells of freshly isolated PBMCs were used for the ELISPOT assay.

## 5.2.4 ELISPOT

For the assay, a 96-well multiscreen-IP sterile plate (Millipore, USA) was pre-coated with 5µl of anti-human IFN-γ monoclonal antibody (MAb) 1-DIK (Mabtech, Sweden) and incubated overnight at 4°C. The pre-coated plate contents were decanted before use. The plate was washed 6 times with 150µl of Dulbecco's Phosphate Buffered Saline (DPBS) (BioWhittaker, Lonza, Belgium), blotted dry and 100µl of R10 media was added to each well. According to the 96-well plate design (Figure 5.1b), the first 6 wells on the plate were reserved for the positive and negative controls represented in triplicate. The negative control lacked the peptide and the positive control was stimulated with 10µg/ml of Phytohemagglutinin (PHA). The ESAT-6 and CFP-10 peptide pools were included on the plate at a final concentration of 8µg/ml. Peptides were added into the appropriate wells at a concentration of 10µg/ml. PBMCs were plated at 200 000 cells/well and the plates were incubated overnight in a 5% $CO_2$ incubator at 37°C. After the incubation period, the plates were washed 6 times with DPBS, and 100µl of 1µg/ml of biotinylated anti-human IFN-γ MAb (Mabtech, Sweden) was added into each well and incubated in the dark for 90 minutes at room temperature. The plates were washed and 100µl of a 1:1000 dilution of Streptavidine-alkaline phosphatase conjugate (Mabtech, Sweden) was added to each well. Following a 45 minute incubation in the dark at room temperature, the plates were washed and 100µl of 5-Bromo-4-Chloro-3-Indolyphosphate Toluidine salt (BCIP) in combination with 100µl of Nitro-Blue Tetrazolium (NBT) in Tris buffer solution was added to each well. After the development time of approximately 8 minutes, the reaction was stopped by washing the plates under tap water and blotted dry. The plates were read on an automated ELISPOT reader (Autoimmune Diagnostics-Germany). Responses were defined based on the size and intensity of the spot. The average of the negative control wells were calculated and the actual response was based on the peptide response subtracted from the average of the negative control wells. An ELISPOT response was considered positive based on several criteria. Firstly, the average number of spots in test wells, (ESAT-6 and or CFP-10) were twice that of the NC wells. Secondly, the average number of spots in test wells was greater than 4 per 200 000 PBMCs. The criterion used for the positive control (PHA) was a minimum of 100 SFUs to total

139

saturation of wells ('black out'). Assays that did not meet this criterion were deemed as failed. ELISPOT results were identified as indeterminate based on the response (back ground) observed in the NC wells. NC wells that had high background (equivalent or higher than the responses observed in test wells) were considered as indeterminate.

The assay was performed blinded to the status of the patients.

| PHA | PHA | PHA | NC | NC | NC | Q01.1 | Q01.2 | Q01.3 | Q02.1 | Q3.1 | Q4.1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Q5.1 | Q06.1 | Q06.2 | Q07.1 | Q07.2 | Q08.1 | Q08.2 | Q08.3 | Q09.1 | Q09.2 | Q10.1 | Q10.2 |
| Q011.1 | Q12.1 | Q13.1 | Q14.1 | M01.1 | M01.2 | M02.1 | M02.2 | M02.3 | M02.4 | M02.5 | M02.6 |
| M02.7 | M03.1 | M03.2 | M03.3 | M03.4 | M03.4 | M03.5 | M03.6 | M04.1 | M04.2 | M04.3 | M04.4 |
| M04.5 | M04.6 | M04.7 | M04.8 | M04.9 | M05.1 | M05.2 | M05.3 | M05.4 | M06.1 | M06.2 | M06.3 |
| M06.4 | M07.1 | M07.2 | M07.3 | M08.1 | M08.2 | M08.3 | M09.1 | M09.2 | M10.1 | M11.1 | M12.1 |
| M13.1 | M14.1 | ESAT-6 Pool | CFP-10 Pool | | | | | | | | |
| | | | | | | | | | | | |

**Figure 5.1b** Diagrammatic representation of 96-well plate layout for ELISPOT assay. The first 6 wells were reserved for the standard controls (PHA and negative control). Blue: QILSS peptides; Green: Mtb9.9 peptides; Yellow: ESAT-6 and CFP-10 pooled peptides. The remainder of the plate remained empty for the assay.

### 5.3 Results and Discussion

### 5.3.1 Immune Responses to QILSS and Mtb9.9 Proteins in Healthy Volunteers

The graphs representing the results of the ELISPOT assays using both Mtb9.9 and QILSS peptides for 16 healthy donors are shown in Figure 5.2 and 5.3. Of the 16 subjects, 2 were excluded as a consequence of a failed ELISPOT assay (no positive control response). Of the remaining 14 total donors 3 had strong responses to both ESAT-6 and CFP-10 peptides suggestive of latent infection, and their results are shown in Figure 5.2.

Among the 3 patients who were latently infected 2 were responsive to some of the QILSS peptides whilst 3 responded to some of the Mtb9.9 peptides (Figure 5.2a and b), although all of the responses were only weakly positive. The number of these patients is evidently too few to comment on the frequency of responders to Mtb9.9 and QILSS peptides in latent infection (and not the focus of these experiments). However, all three patients did mount a response to one or more of the peptides spanning the conserved C-terminal region of Mtb9.9 (Figure 5.2b). This region was not responsible for stronger immune responses in patients with active tuberculosis.

It would have been interesting to evaluate responses in a larger number of latently infected individuals and compare these results more formally to actively infected TB patients to determine if the QILSS and Mtb9.9 protein might be differentially expressed at different stages of infection. This was beyond the scope of this study and would require a larger cohort.

The strongest immune response amongst this small group of patients was to peptide Q6.1, albeit in a single individual. This is in keeping with a study that examined the immune responses to *esxJ* (Rv1083c) and found that a peptide spanning the exact same region of the QILSS family (codon 31 to 45) was the only core antigenic region in latently infected individuals (Figure 5.2a). A recent study conducted by Sette and colleagues involved investigating the T- cell responses for TB antigens at genome wide level in latently infected individuals (Arlehamn *et. al.,* 2012). Their study involved the usage of PBMCs from latently infected subjects that were tested in ELISPOT assays with approximately 20000 peptides predicted to bind to a series of

common HLA class II antigens. This study identified 3 antigenic islands encoding clusters of high responder epitopes, which included *esxH*, *esxG*, *esxR*, *esxQ*, *esxA* and *esxB*. Pertinently the only other *esx* loci associated with significant responder frequencies were the Mtb9.9 family. The screen didn't use overlapping peptides spanning the whole coding sequence but identified two peptides that elicited responses in more than 10% of subjects, confirming that Tb9.9s are latent tuberculosis antigens.

Donors that were ESAT-6 and CFP-10 ELISPOT negative for these peptides are represented in Figure 5.2. For the healthy donors that were EsxAB negative (Figure 5.3a and b) only low level responses to Mtb9.9 and QILSS peptides were detected in a minority of subjects, and only at threshold levels indicating there was no broad cross reactivity to other antigens that the subjects may have been exposed too. It was noted that some patients displayed no responses at all to certain QILSS and Mtb9.9 peptides. This was observed in peptides Q4.1-Q7.2; Q9.1-Q10.1 and Q13.1-Q14.1 (Figure 5.3a, and in peptides M2.3-M5.3; M6.2-M6.4; M7.3-M8.3; M11.1-M12.1 and M14.1 (Figure 5.3b). Although thedifferences between peptides, low-level responders and zero responders may be solely due to background noise it does raise the possibility that the low level responses to some of the peptides could represent specific immune responses. This could be due to exposure to *M. tuberculosis* or to other cross-reactive antigens (Mitchell *et. al.,* 2012). This seems unlikely given that the latently infected individuals as defined by EsxA and ExB responses had more robust reponses to QILSS and Mtb9.9 peptides, but it hints at the possibility that particular subjects could be latently infected but not detected by ESAT-6 or CFP-10 peptides.
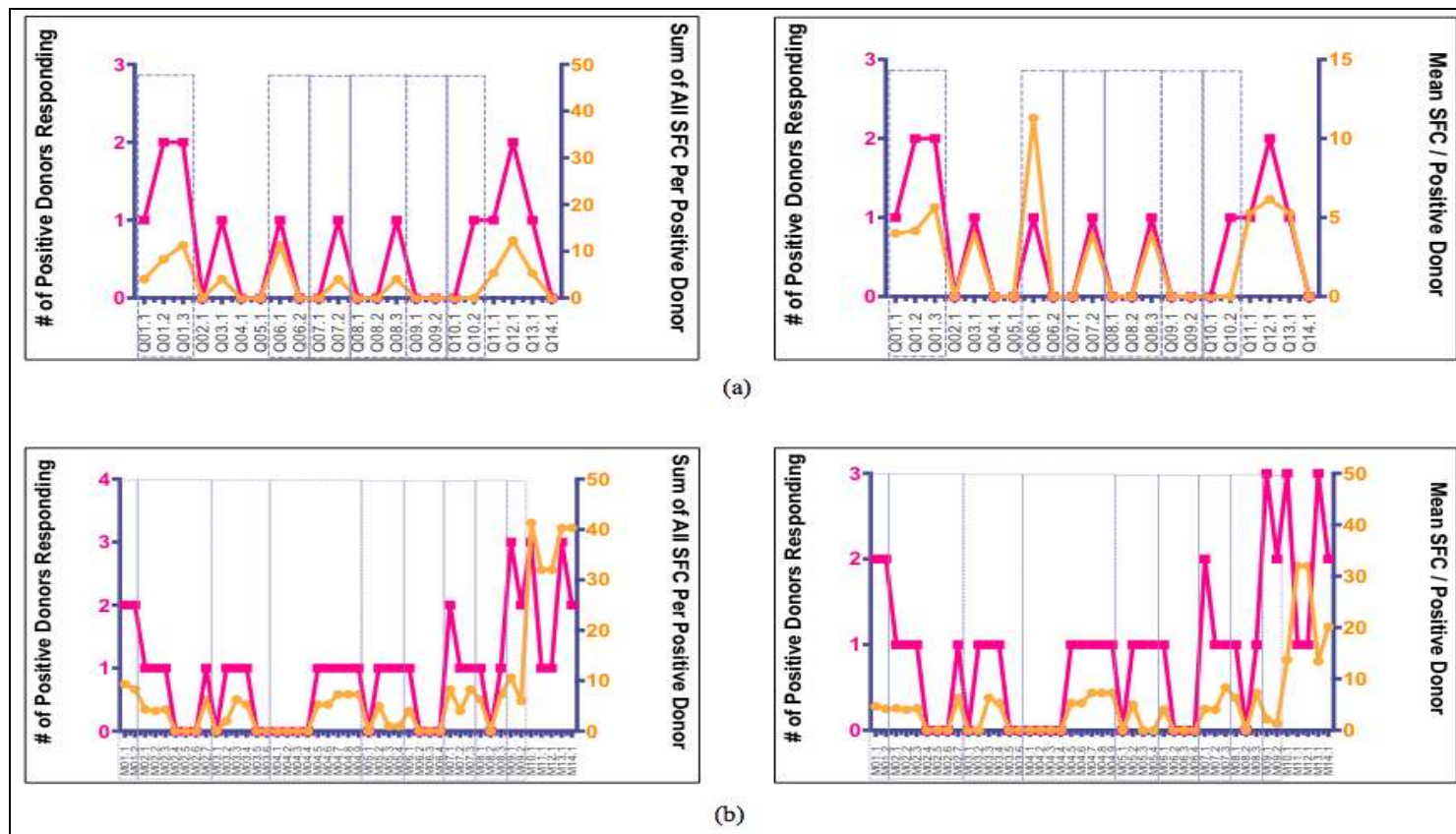
**Figure 5.2**     Graph of Healthy Donors (EsxAB ELISPOT +ve) for **(a)** QILSS and **(b)** Mtb9.9 Peptide Positive Responses. The pink component of the graph represents the number of positive donors responding and the orange is indicative of the SFC per positive donor, respectively. Dashed boxes group separate peptides that cover the exact same region of the protein but differ by one or more amino acid from conserved peptides shared by all proteins.

**Figure 5.3** Graphs of Healthy Donors (EsxAB ELISPOT -ve) for **(a)** QILSS and **(b)** Mtb9.9 Peptide Peptide Positive Responses. The pink component of the graph represents the number of positive donors responding and the orange is indicative of the SFC per positive donor, respectively. Dashed boxes separate peptides that cover the exact same region of the protein but differ by one or more amino acid from conserved peptides shared by all proteins.
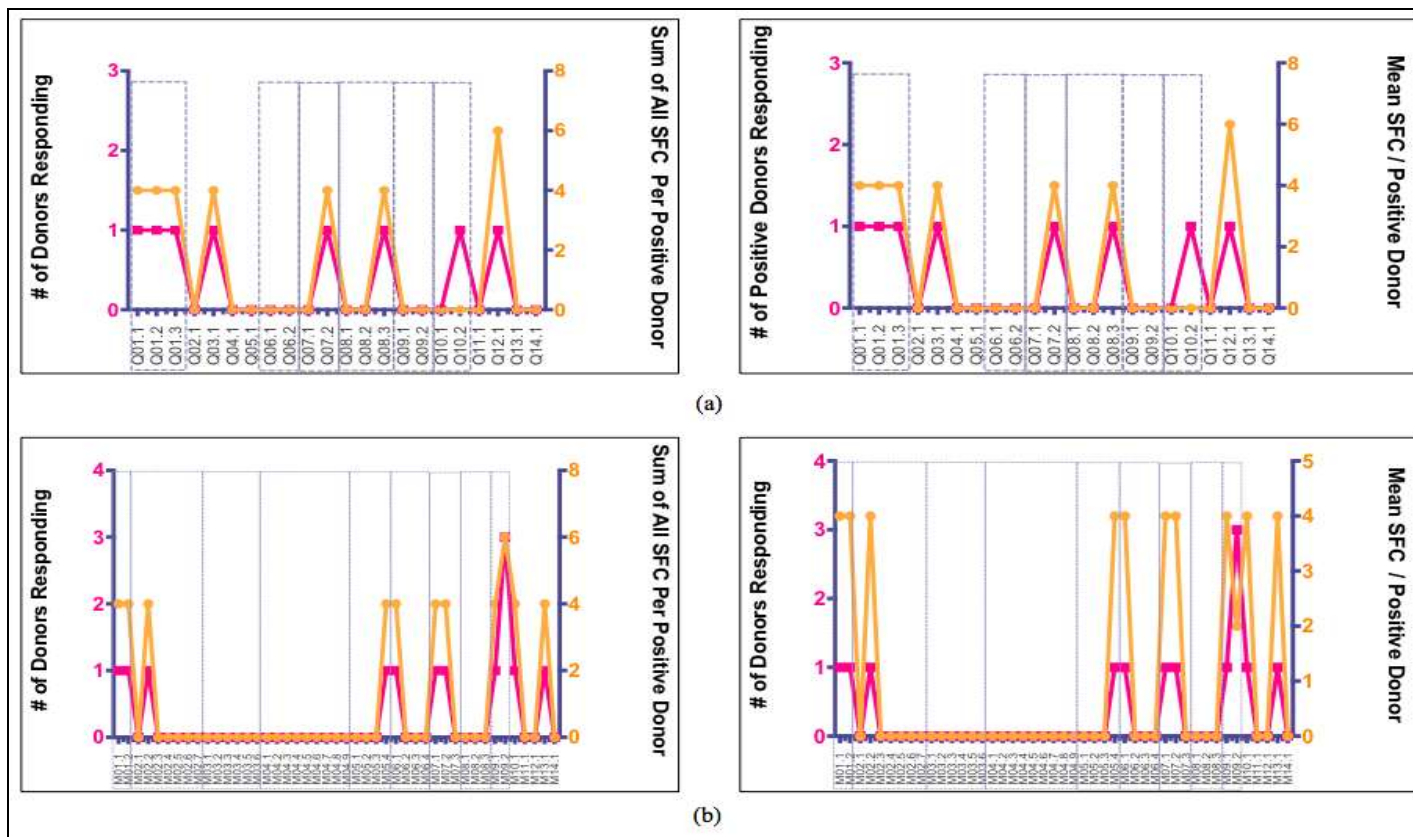
**5.3.2 Immune Responses to QILSS and Mtb9.9 Proteins in Active Tuberculosis Patients**

I thereafter looked at peptide responses in actively infected tuberculosis patients. A total of 27 PBMC samples were obtained from patients recruited in an ongoing study currently being conducted at King Dinizulu hospital.

The number of donors recognizing the QILSS (Figure 5.3a) and Mtb9.9 (Figure 5.4b) peptides is represented as well as the magnitude of the ELISPOT responses. Of the 27 subjects, 21 (77%) had strong responses to ESAT-6 and CFP-10 peptides, while the remaining 6 (22%) displayed no responses, respectively. Five of these six patients had no responses to any QILSS or Mtb9.9 peptides suggesting anergic responses to TB antigens previously observed in HIV infected individuals as well as in patients with severe tuberculosis.

The sum of all SFC per positive donor for QILSS and Mtb9.9 peptide responses ranged from 0 to 530 spot forming counts (SFC). No responses were observed for peptides Q7.1, Q14.1, M7.3 and M8.1, for both healthy (Figures 5.2b, 5.3b) and TB infected donors (Figure 5.4b). The TB infected donors exhibited noticeably lower responses from peptides Q5.1 to Q14.1 (Figure 5.4a) and suggest a more general pattern that the QILSS proteins are less immunogenic that the Mtb9.9 proteins as shown by both the number of responders and the strength of response. It is also interesting to point out that there were only 5 variant residues in the QILSS compared to 11 in the Mtb9.9 family, which is consistent with a relationship between genetic variability and immunogenicity compatible with immune selection. However, the response from multiple peptides across both proteins also confirms that both these ESX proteins have multiple epitopes recognised by the study population.

Although there were antigenic responses elicited by multiple peptides spanning both the QILSS and Tb9.9 proteins it is interesting to determine if there were regions that are immune-dominant in terms of either responder frequency or SFCs (Figure 5.4) and whether there is any correlation between immune-dominance and sequence variation. In the case of responder frequency, if immune-dominance is defined as a greater than 25%,

then peptide loci Q1, Q4 and Q11 are dominant for the QILSS (Figure 5.5b) of which two are sequence invariant. Using mean magnitude of response (greater than 20 SFUs) then Q1, Q2, Q3 are dominant, with only Q1 having sequence variation. In contrast for the Tb9.9 proteins M2, M3 and M5 are responder dominant and M2, M4, M5 and M9 are magnitude dominant and all of these span regions of sequence variation. Similar peptide-mapping experiments in cattle infected by *M. bovis* conducted by Jones and colleagues in 2010 demonstrated that the location of the major immunodominant epitopes also occured within regions of sequence diversity for both the QILSS and Mtb9.9 subfamilies (Jones *et. al.,* 2010a).

The key aim of this study was to determine if the variable regions of these two protein groups influenced immune recognition. For the exactly overlapping peptides (Figure 5.4a), there were some noticeable differences in immune responses. For the QILSS peptides Q1.1 to Q1.3 there were 8 of the 21 positive patients (38%) with similar number of responders for each peptide, but varying peptide responses ranging from 12 to 68 mean SFC with Q1.1 having a four fold increase compared to the other two. These peptides correspond to the N-terminal region of the QILSS and vary at 2 adjacent residues located at codon 2 and 3 (AS, TS, AT). However, the eccentric location of this variability on the peptide makes it less likely that the differences in the response intensity are due to epitope recognition. Conceivably N-terminal sequence variability could effect gene expression that would account for the observed differences in immune responses to the peptides. More convincingly there were no responders to peptide Q7.1 compared to 4 (14%) for Q7.2. These two peptides differ by a single substitution corresponding to codon 48. Interestingly, apart from Q2.1 and Q3.1 the remaining QILSS peptides had low peptide responses, although the number of responders observed for Q4.1 and Q11.1 was high with 9 and 7 TB positive donors respectively.

In this study as for QILSS, the Mtb9.9 peptides were designed overlapping the antigenic region for all the ESX members that are part of this subfamily (*esxN, I, L, O,* and *V*). The Mtb9.9 peptides with identical spans but different sequences (Figure 5.4b) revealed some peptides were more immunogenic than others. This was particularly evident for peptide

2.7, which is distinguished from all of the other matched peptides by a substitution equivalent to A12D mutation. In contrast to peptides M2.1 to M2.6 the mean SFC was five fold less. There are other adjacent mutations represented on this peptide equivalent to positions 20, 22 and 23. However, substitutions at these residues are isolated from A12D in the next set of peptides (M3.1 to M3.6) that span from codon 13 to 30. However, all of these peptides produce only low-level mean SFUs (mean SFU between 4 and 12), indicating it is the codon 12 substitution that probably accounts for the low-level immune responses of peptide M2.7. The A12D substitution was identified through genome sequencing in *esxL* and was found in isolates 0070 and 0080 that are part of the EAI *Mtb* clade (Figure 2.12). The work of Uplekar and colleagues also identified this mutation in *esxL* as well as another Tb9.9 protein *esxO*, and these also occurred in EAI clinical isolates (Uplekar *et. al.,* 2011).

Another example of variability in immune responses occurs amongst the peptides M4.1 to M4.9 (Figure 5.4b). Although this region spanning codons 19 to 36 is comparatively less responsive in terms of both responder number and mean size of response, one peptide M4.4 stands out as having more that 10% responder frequency as well as higher mean SFCs. This is the most variable region of the Tb9.9 proteins, and these peptides differ at 5 different sites. The M4.4 peptide recapitulates an L23S in *esxL*. This mutation is rare only occurring in 2 of the genome sequenced strains and one of the amplicon sequenced strains and in none of the strains characterized by Uplekar and colleagues (Uplekar *et. al.,* 2011).

Another region in the Mtb9.9 protein suggest an impact of sequence variation within the Mtb9.9 family on immune responses. Peptides M5.1 and M5.2 result in either a high frequency of responders or a high mean SFC in contrast to their counterparts M5.3 and M5.4 (Figure 5.4b). This is another region of high variability with four variant sites. Peptide M5.1 corresponds to the *esxL* sequence, and M5.2 representing *esxL* with an R33S mutation (associated with the LAM4 clade). This mutation at *esxL* also results in a sequence that is identical to *esxI* and *V* between codons 25 and 42. M5.3 corresponds to the *esxO* (which is identical to *esxN*) sequence that differs from *esxL* at four sites, with the M5.4 peptide corresponding to *esxO* with the same R33S mutation. In this case the

difference in immunogenicity between the two pairs of peptides appears not to be due to the mutation identified in clinical isolates but rather differences between members of the Tb9.9 family. This is most likely to be due to epitope modification but also raises the possibility that differences in gene expression or gene dosage could also play a role.

Mean or total SFCs do not always provide the best means of comparing immune responses and may conceal differences in immune responses to specific peptides. I therefore looked at responses on an individual patient level by plotting each patient's response to related peptides (Figure 5.5, 5.6 and 5.7). This revealed that sequence variation at other sites could be important. An example of this is seen in Figure 5.5e that charts the responses of 8 responders to either peptide Q9.1 or Q9.2. These peptides map to a region corresponding to codons 49 to 66 and differ by only a single residue at codon 58 corresponding to the amino acids A58 or T58 (Table 5.1). A58 is the H37Rv protein sequence for QILSS members *esxK* and *esxP* whereas T58 is found in *esxJ*, *esxM* and *esxW* (Tables 2.6 and 2.8). The site is also variable amongst clinical isolates with the KZN-LAM4 strains all having an A58T mutation in *esxK* (Figure 2.12). One isolate was found by amplicon sequencing to have a A58P in *esxK* and other isolates were found to have a A58T in *esxP*. Overall there was no difference in the magnitude of response between the two peptides either in terms of total or mean SFUs. However, only one patient (P032) had an equal response to both of the two peptides. Four patients had zero or low responses to Q9.1 but high responses to Q9.2 (Figure 5.5e). Conversely the remaining 3 patients had high responses to Q9.1 with lower or zero response to Q9.2. This is highly suggestive that sequence variation at codon 58 can influence immunogenicity, and is supported by the study of Jones and colleagues who found that the T58 peptide had a higher frequency of responders in cattle when compared to the A58 peptide (Jones *et. al.,* 2010a).
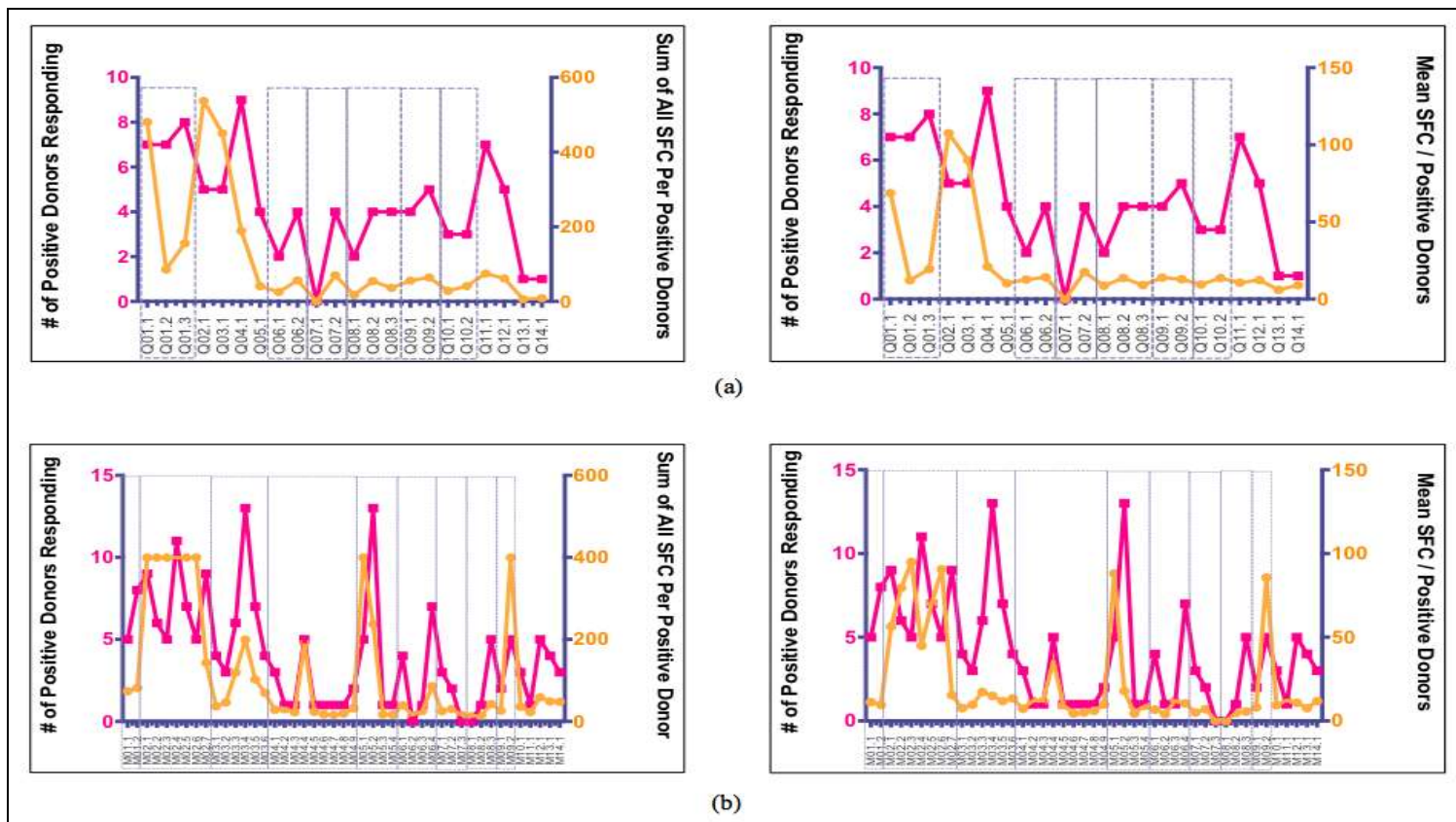
**Figure 5.4** Graphs of TB Actively Infected Donors for **(a)** QILSS and **(b)** Mtb9.9 Peptide Responses. The pink component of the graph represents the number of positive donors responding and the orange is indicative of the SFC per positive donor, respectively. Dashed boxes separate peptides that cover the exact same region of the protein but differ by one or more amino acid from conserved peptides shared by all proteins.
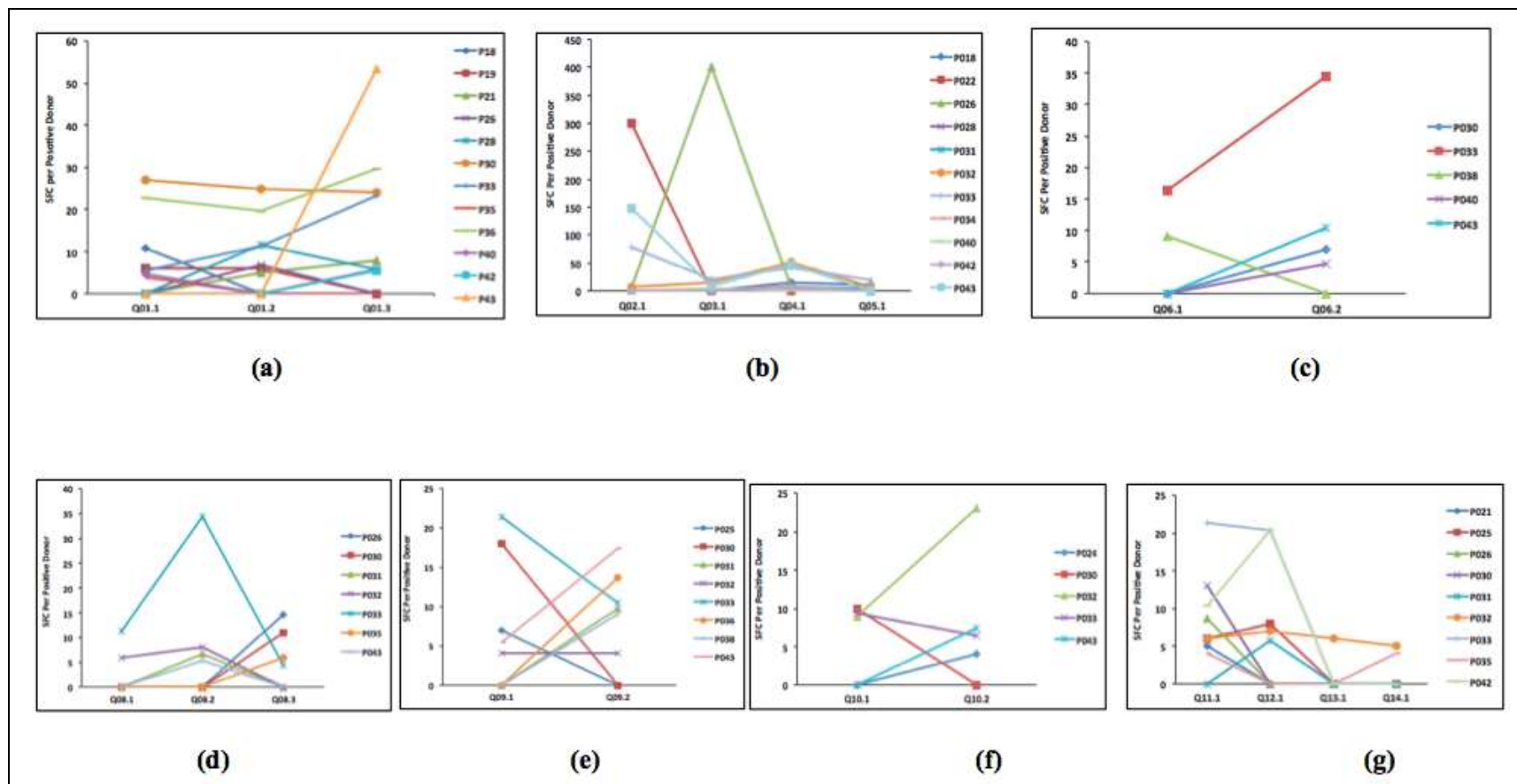
**Figure 5.5** QILSS Peptide Responses per TB Infected Donor for variable regions. Graphs **(a)** to **(g)** represent the overlapping peptides with the mean SFC responses.

151

**Figure 5.6** Mtb9.9 Peptide Responses per TB Infected Donor for variable regions. Graphs (**a**) to (**f**) represent the overlapping peptides with the mean SFC responses.

152

**Figure 5.7** Mtb9.9 Overlapping Peptide Responses per TB Infected Donor for variable regions. Graphs **(a)** to **(d)** represent the overlapping peptides with the mean SFC responses.

### 5.4 Conclusion

The importance of ESAT-6 and CFP-10 for immune-diagnostics and vaccine design has meant that the immune responses in humans to these two prot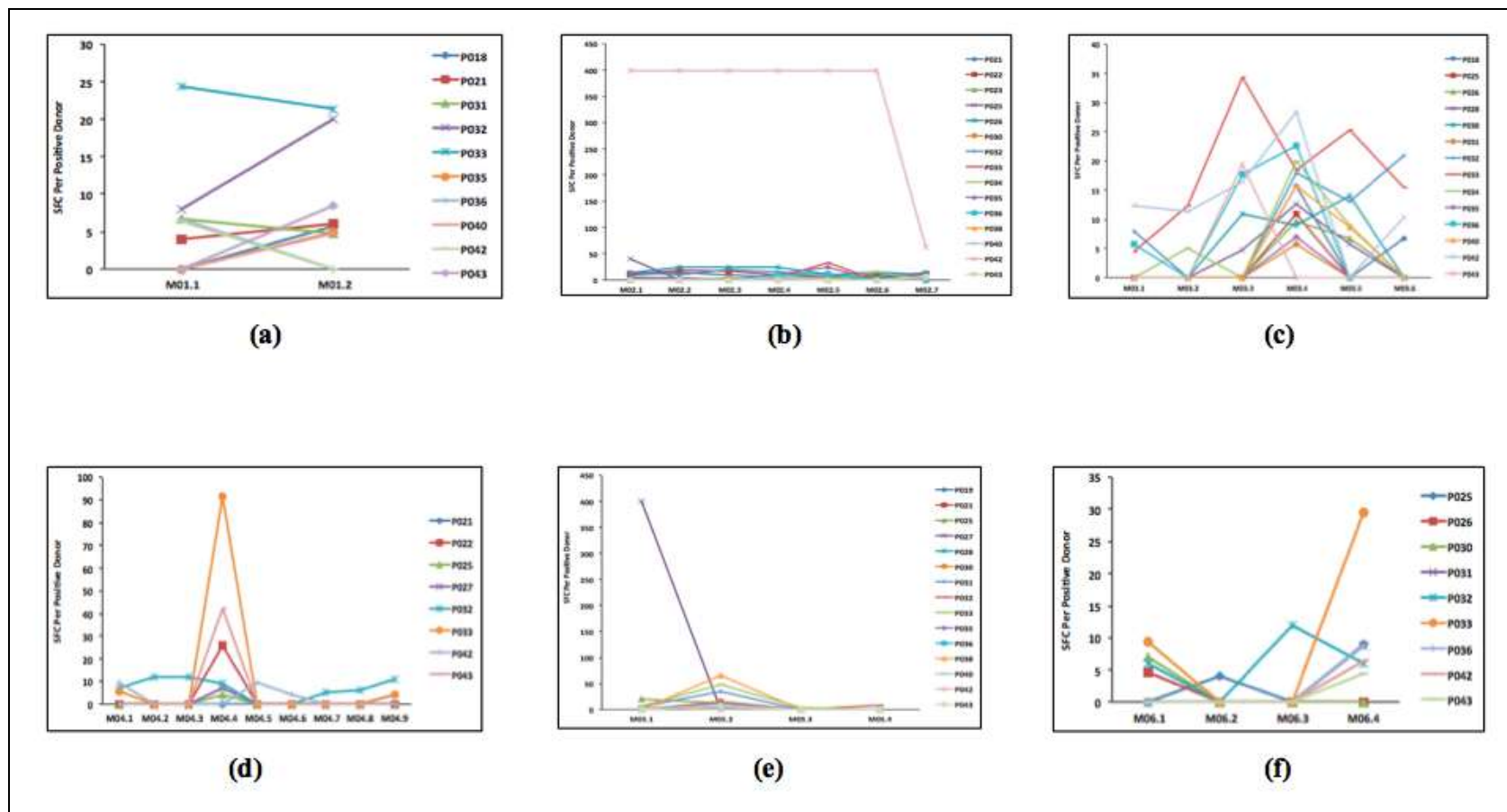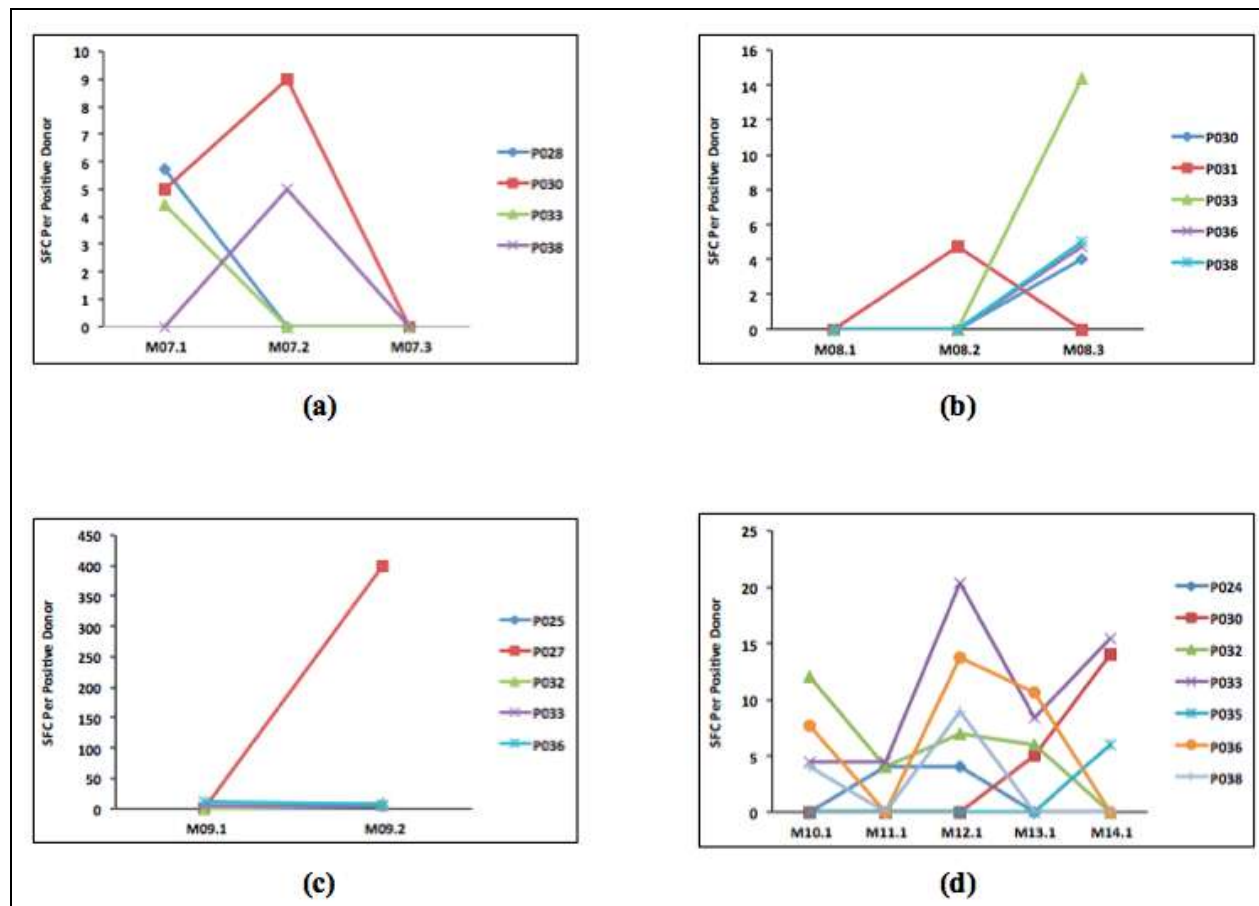eins have been exhaustively studied. It is therefore surprising that there have been few previous systematic attempts to immunologically characterize the other members of the ESX protein family. I focussed on the QILSS and Mtb9.9 proteins because of my observation that genetic heterogeneity was concentrated in these subsets of the broader ESX protein family. The results represent the first comprehensive immunological analysis of the QILSS and Mtb9.9 family in humans and show that both proteins are highly immunogenic and contain at least one epitope that is recognized by more that 40% of patients infected with tuberculosis. Whilst both protein families have multiple epitopes, the number or responders and the magnitude of the responses indicate that the Mtb9.9 family are more immunogenic than the QILSS proteins. The immunogenicity of QILSS and Mtb9.9 proteins relative to other proteins in the *Mtb* genome is highlighted by the data reported by Arlehamn and colleagues (Arlehamn *et. al.,* 2012) who identified members of these families in their whole genome screen even though they didn't use peptides covering the complete proteins. A possible source for this type of prominent peptide response could be as a result of the high degree of amino acid similarity for the QILSS and Mtb9.9 proteins resulting in high levels of protein expression as a result of expression from multiple genes.

Documenting the strong and broad immune responses to QILSS and Mtb9.9 proteins was important because it allowed me to evaluate how genetic diversity could play a role in generating different immune responses. I was interested to explore how immune responses differed to the individual QILSS and Mtb9.9 families because they are very similar. Although I have not shown variability in the regulation and expression of this protein set in different clinical isolates, the wide variation in expression of other ESX members suggests this may be the case. It is also difficult to imagine why *Mtb* would possess 5 copies of genes that are almost identical if they are not differentially expressed. At the same time I evaluated mutations that occurred in clinical isolates. The results provide strong evidence that the genetic diversity between the QILSS and Mtb9.9 protein

family members does result in difference in immune recognition. In addition, I showed that mutations that arise in clinical isolates can also change the strength of the immune response.

This is the first experimental evidence that sequence diversity in *Mtb* effects immunogenicity. It is remarkable, given the intense research with other pathogens into the effects of sequence diversity on immune responses and the importance of this type of research for vaccine design, that no other studies like this have been carried out. However, it is still necessary to carry out further studies in which the *Mtb* strain infecting an individual is characterized and this is then correlated with the immune response in that same individual.

One limitation in this study is the relatively small number of donor patients included for the ELISPOT assay. In addition, it was necessary to restrict the number of replicates and number of peptides that I used because patients with tuberculosis, even without HIV, do not yield high numbers of PBMCs from the volumes of blood that was drawn in the context of the clinical situation in which this study was carried out. However, having documented the immune-dominate regions of both the QILSS and Mtb9.9 proteins, and identified loci that could be responsible for immune variation it is now possible to target a reduced number of peptides in a follow up study with a greater number of patients with multiple replicates.

# CHAPTER SIX

# CONCLUSION AND FUTURE RESEARCH

A great and chief milestone in *Mtb* research history was the complete genomic sequencing of the H37Rv genome in 1998 (Cole *et. al.,* 1998). This genome sequence ultimately paved the way for many avenues of research not least in the study of *esx* genes. Cole and colleague's analysis of the genome sequence identified the full extent of the *esx* gene family, as well as describing five genomic islands that later came to be known as ESX-1 to ESX-5. They also accurately predicted that these islands encoded the secretory mechanism for the cognate *esx* genes that was later proven to be the case by scientists in his group (Cole *et. al.,* 2008). Since then considerable progress has been made in elucidating the function and mechanism of ESX-1, ESX-3 and to a lesser extent ESX-5. ESX-1 encodes and exports EsxA and EsxB, for which there is now good evidence that these two proteins are required for *Mtb* escape from the phagolysosome into the cytoplasm. ESX-3 was found to be essential for in vitro growth of *Mtb*, is under control of the divalent cation regulators *IdeR* and *Zur* and probably has a yet undefined role in iron and zinc homeostasis.

Less is known about ESX-5 that encodes the QILSS and Mtb9.9 genes *esxM* and *esxN*. There is now direct evidence for the secretion of *esxN*, based on immune-blotting experiments with an *esxN* antibody. Although there is no conclusive published data for the secretion of the other members of the QILSS and Mtb9.9 proteins by the ESX-5 type VII secretory apparatus, Bottai and colleagues noted 4 distinct spots when carrying out 2D immune-blotting of culture filtrate proteins with their *esxN* antibody (Bottai *et. al.,* 2012). The detection of these appropriately sized spots in the culture filtrate was dependent on an intact ESX-5 and hence suggestive that they were *esxN* homologues. Their and others work have also established that ESX-5 exports its encoded PPE and PE proteins which induce strong T-Cell immune responses. It has also been noted that immune responses to PPE and PE proteins not encoded by ESX-5 were also dependent on an intact *EccD5* transmembrane channel, suggesting a broader role in immune regulation perhaps via export of a wider class of PPE-PE proteins. These results were published after the initiation of my studies and I initially conceived a similar immune-regulator role of Esx-5 via the export of varied combinations of the QILSS and Mtb9.9 proteins, driven by both intragenic de-novo mutation as well as differential transcription.  It is also

pertinent that unlike ESX-1, the selective inactivation of the ESX-5 *esx* genes does not produce a significant attenuated phenotype, suggesting *esxM* and *esxN* are not virulence effectors like their counterparts from ESX-1 and may have a different role such as immune modulation.

The aim of the first part of this study was to evaluate heterogeneity in the QILSS and Mtb9.9 families and this is the first study assessing the ESX gene diversity and SNP variation in strains circulating within the South African population. In Chapter Two, my amplicon data suggests that *esxA, B, C, E, G, H, J, R, S* and *T* are highly conserved. However, with the inclusion of additional sequences from a collection of strains which were genome sequenced the number of invariant genes declined as nSNPs were also identified in *esxB*, *C*, *E*, *R*. It is likely that as more clinical isolates are sequenced the number of invariant genes will decline further and for a complete picture a broader global collection of strains will need to be characterized. This will be important if immune selection is to be evaluated in the context of different host genetic backgrounds.

The distribution of nSNPs was not uniform across the *esx* genes and this possibly could result from a lack of immune selective pressure by the host immune system on the invariant members. The concentration of nSNPs in the QILSS and Mtb9.9 genes as appose to the ESX-1 and ESX-3 *esx* members can be explained by the functional roles of the latter two systems. In both these systems deletion of *esx* genes leads to attenuation or lack of viability and therefore mutations in these genes are likely to carry a high fitness cost. In contrast, if ESX-5 has an immune-regulatory role, then mutations in the associated *esx* genes could carry a selective advantage. Therefore a further in depth investigation of the interaction with the host is required.

As previously stated a limiting factor in this study was the number of isolates used, hence the *esxA, G, H, J, S* and *T* genes cannot be inferred as being globally conserved. One distinct difference from the study by Comas and colleagues, is that I found *esxH* to be invariant, compared to their study that found to it to be hypervariable (Comas *et. al.,* 2010). They only looked at 26 strains and therefore their conclusions are not likely to be

valid, so further studies at a later stage should include a larger genetically diverse number of isolates with a broader worldwide representation of the strain diversity.

One intriguing observation that has come out of these sequencing studies is the sites of sequence variation leading to nSNPs occur predominantly at sites that are already variable amongst the QILSS and particularly the Mtb9.9 family. An example of this is the A58T, and A58P mutations that I observed and described in Chapter 5 in the context of immune recognition of QILSS proteins. The A58T mutation was observed in *esxK* and *esxP* and converted these proteins to have the same sequence as *esxW, J, M*. I also found 1 isolate with an A58P mutation at this residue. It is easy to imagine that this variation is an experimental artefact of sequencing highly similar genes. This is extremely unlikely as these types of mutations were seen both by amplicon and genome sequencing done in different laboratories. In addition similar gene conversion events were reported in the amplicon sequencing study carried out by Uplekar and colleagues (Uplekar *et. al.,* 2011). Of note, the Broad institute used both short and jumping libraries for sequencing. The jumping libraries were approximately 2000 to 3000bps in length and allow pair end sequences separated by approximately 2000bps to be more accurately positioned. This overcomes some of the difficulties of resolving diversity in repetitive regions that was encountered in earlier sequencing efforts.

Homologous recombination between the highly similar sequences of the QILSS and Mtb9.9 genes is the most likely mechanism generating this kind of diversity. My results presented in Chapter 5 provide evidence that at least for codon 58 of the QILSS family, immune selection could be driving these gene conversions. It is not clear how altering the number of copies of a single allele from three to four out of five (for example in the case of an A58T mutation in *esxK*) would significantly alter immunogenicity. It seems unlikely that such a relatively small proportional effect on gene number could translate into a significant change in immune phenotype. Alternatively, if the genes were transcribed differently between strains then it might represent an allelic switch.

Additional bioinformatic analysis on a larger number of sequences is required to more formally address this question of recombination events. Understanding how diversity in the immune-dominant ESX families is generated and maintained is clearly highly important for vaccine development. It is hoped this will ultimately contribute in the management of this disease burden in South Africa.

In Chapter Three, I was able to establish that variation in ESX transcription does exist amongst the clinical isolate groupings. The major findings of this study demonstrate (i) *esxA, B, C, F, M* and *Q* are variably expressed in clinical isolates; (ii) *esxC, F, M* and *Q* was down-regulated in BCG, 5942, 3028, 5944, 5538, 5539 and 6103; (iii) low expression for *esxA* and *esxB* was observed in the Beijing Isolates. There are a number of limitations to my approach. Firstly, I only assessed transcription in mid-log phase cultures and not in more relevant cell cultures or animal models. It was important to standardize conditions and restrict the passage number of these recent clinical isolates to ensure there was no adaptation to growth in culture media. I therefore decided to adopt a simple reproducible system that was designed to detect gross differences between strains. Using this approach I was able to demonstrate a consistent difference in the transcriptional profile in the Beijing family of strains. A second limitation is that I only targeted a limited number of *esx* members and this was based on the existence of sufficient sequence diversity to design a robust RT-qPCR assay. Unfortunately, this excluded the QILSS and Mtb9.9 genes the main focus of my studies, and attempts to design an allelic specific PCR were not successful. I considered further optimization or the use of a probe based assay but with the increasing availability of RNA-Seq I opted for the latter approach. These experiments are presently underway but have not been included in this thesis. Finally, I only examined the transcription of one gene in some of the *esx* linked pairs.

My findings are descriptive and would need to be completed in a wider number of conditions to determine whether *esxA* and *B* expression is universally suppressed in Beijing strains. Without further functional studies it is only possible to speculate on the mechanism behind the inter-strain differences in *esx* gene regulation. The fact that

transcriptional differences extent beyond the RD1 members hints at a wider derangement in gene regulation and raises the possibility that multiple *esx* gene pairs are regulated in a coordinated fashion.

Bacterial regulatory networks are complex and interconnected. The work of Gonzalo-Asensio and colleagues have shown that PhoP is a regulator of RD1 expression (Gonzalo-Asensio *et. al.,* 2008) but it also appears to influence the expression of GroEL and have some overlap with the DosS/DosT-DosR two-component regulatory system. Beijing strains are natural mutants of this two-component regulatory system as well as having lower levels of GroEL expression suggesting a future direction of research to understand the basis of variable RD1 *esx* expression (Stanely *et. al.,* 2003; DiGuiseppe *et. al.,* 2009; Champion *et. al.,* 2012). In addition further studies should focus on characterizing the expression of all *esx* genes, using whole-genome transcriptional profiling in a larger number of isolates and isolate groupings. Finally it will be critical to determine if there are any clinical implications of variation in *esxA* and *B* expression. Interferon gamma release assays (IGRA) based on ESX-1 peptides have been reported as performing adequately in China where Beijing strains predominant but there may be more subtle implications in terms of the efficacy of ESAT-6 or CFP-10 based vaccines.

It was reassuring that differences in transcription between clinical isolates could be confirmed by Western Blot analysis as seen in Chapter Four (Section 4.3.2). One clear limitation of my experiments was that the control protein I selected, GroEL appears to be down regulated in Beijing strains. This was not appreciated at the time I designed the experiments and it will be necessary to repeat the experiments with a different protein whose regulation does not vary in clinical isolates of *Mtb*.

In my experiments in Chapter Four, ESAT-6 was not detected in the culture filtrate protein extracts in any of the isolates. This could be as a consequence of insufficient expression of ESAT-6 in the case of the Beijing strains selected, but was probably due to the starting volume of culture. I reasoned that it was easier to standardize quantification

in the whole cell fraction of mid-log phase cultures and it proved a successful approach to identifying heterogeneity.

The protein spectra obtained using the MALDI-TOF MS in Chapter Four (Section 4.3.3), is the first report of a modified whole colony MALDI-TOF MS that directly detects ESX proteins in whole cell lysate extracts. I opted for the "Top Down" approach to establish protein spectra for the selected *Mtb* isolates. More optimization will be required to obtain higher resolution of the protein spectra. The proteomics analysis can further be explored using the "bottom-up" approach by trypic digestion to further characterize the putative CFP-10 protein peak as well as the additional peaks noted within the vicinity. An additional future consideration is to explore the identification of the other members of the ESX-1 component that may contribute to the unique protein profiles for *Mtb,* as well as the effect of the growth cycle and other conditions on CFP-10 expression. The use of other proteomic platforms is another possibility and might be suitable for analysing extracted proteins from conditions mimicking the host environment in animal or macrophage models.

My proteomics approach also confirmed the Immunoblot results seen in Chapter Four (Section 4.3.2). Overall, my study proves and substantiates the usefulness and benefit of proteomic technology in the laboratory setting to confirm expression levels independently of antibodies which have proved difficult to raise against ESX proteins. It could be optimized into the routine identification and quantification of *Mtb* for the presence of CFP-10 and perhaps other ESX members in whole cell protein extracts and other cellular fractions. This combined and collective approach will also consequently contribute to the understanding the pathogenesis of tuberculosis by the identification of differentially expressed proteins, thus assisting in understanding why certain strain families, such as the Beijing family, are phenotypically distinct from others.

The work described in Chapters 3 and 4 establishes that at least in mid-log phase culture there is variable expression of selected ESX proteins in clinical isolates of *Mtb*. I was restricted at the time of conducting the experiments to focussing on ESX members that

had either sufficient sequence diversity for transcriptional analysis, had available antibodies or were identifiable by proteomics. However, the sequence analysis described in Chapter One indicated that most variation was found in the QILSS and Mtb9.9 genes. I therefore focussed my immune analysis on these two sub-families. I was able to show that in the limited number of patients enrolled evidence that heterogeneity at the nucleotide level identified in clinical isolates can affect immune responses. To my knowledge this is some of the first data that shows the human response in tuberculosis may be modified by the infecting strain of *Mtb*. Critically I did not directly correlate the genotype, or transcriptional type of the infecting strain with the immune response in individual patients. The isolates from the study reported here have been saved and are currently being analysed. In the next phase, studies will focus on the immune-dominant variable regions in QILSS and Mtb9.9 identified in this study and perform multiple replicate plates in a sufficient number of patients to generate a more quantitative analysis. This will be combined with HLA typing and HLA peptide binding assays to isolate the influence of human genetics on the variability of responses.

The peptides used in my ELISPOT assay in Chapter Five, have the potential to be utilized as a diagnostic tools in future research, screening and identification of latently and actively infected individuals. The identification of differentially expressed antigens within and between strains will provide significant information in the design of serodiagnostic tests as well as vaccines. Overall, the results from this study establish that differences in sequence, transcriptional profiles and protein expression patterns in ESX exported proteins do exist between different clinical isolates and affect immunogenicity The indirect benefits and future applications of the results reported in this study could impact our understanding of bacterial pathogenesis, immunology and influence vaccine development.

# REFERENCES

1.  **Aagaard, C., Govaerts, M., Meikle, V., Vallecillo, A. J., Gutierrez-Pabello, J. A., Suarez Guemes, F., McNair, J., Cataldi, A., Espitia, C., Andersen, P. and Pollock, J. M.** 2006. Optimizing antigen cocktails for detection of *Mycobacterium bovis* in herds with different prevalences of bovine tuberculosis: ESAT6-CFP10 mixture shows optimal sensitivity and specificity. J. Clin. Microbiol. **44:**4326-4335.

2.  **Aagaard, C., Govaerts, M., Okkels, M., Andersen, P. and Pollock, J. M.** 2003. Genomic approach to identification of *Mycobacterium bovis* diagnostic antigens in cattle. J. Clin. Microbiol. **41:**3719-3728.

3.  **Abdallah, A. M., Bestebroer, J., Savage, N. D., de Punder, K., van Zon, M., Wilson, L., Korbee, C. J., van der Sar, A. M., Ottenhoff, T. H., van der Wel, N. N., Bitter, W. and Peters, P. J.** 2011. Mycobacterial secretion systems ESX-1 and ESX-5 play distinct roles in host cell death and inflammasome activation. J. Immunol. **187:**4744-4753.

4.  **Abdallah, A. M., Gey van Pittius, N. C., DiGiuseppe Champion, P. A., Cox, J., Luirink, J., Vandenbroucke-Grauls, C. M. J. E., Appelmelk, B. J. and Bitter, W.** 2007. Type VII secretion-mycobacteria show the way. Nat. Rev. Microbiol. **5:**883-891.

5.  **Abdallah, A. M., Savage, N. D., van Zon, M., Wilson, L., Vandenbroucke-Grauls, C. M., van der Wel, N. N., Ottenhoff, T. H. and Bitter, W.** 2008. The ESX-5 secretion system of *Mycobacterium marinum* modulates the macrophage response. J. Immunol. **181:**7166-7175.

6.  **Abdallah, A. M., Verboom, T., Hannes, F., Safi, M., Strong, M., Eisenberg, D., Musters, R. J. P., Vandenbroucke-Grauls, C. M. J. E., Appelmelk, B. J., Luirink, L. and Bitter, W.** 2006. A specific secretion system mediates PPE41 transport in pathogenic mycobacteria. Mol. Microbiol. **62:**667-679.

7.  **Abdallah, A. M., Verboom, T., Weerdenburg, E. M., Gey van Pittius, N. C., Mahasha, P. W., Jimenez, C., Parra, M., Cadieux, N., Brennan, M. J., Appelmelk, B. J. and Bitter, W.** 2009. PPE and PE_PGRS proteins of *Mycobacterium marinum* are transported via the type VII secretion system ESX-5. Mol. Microbiol. **73:**329-340.

8.  **Abebe, M., Doherty, T. M., Wassie, L., Aseffa, A., Bobosha, K., Demissie, A., Zewdie, M., Engers, H., Andersen, P., Kim, L., Huggett, J., Rook, G., Yamuah, L. K., Zumla, A. and VACSEL study group.** 2010. Expression of apoptosis-related genes in an Ethiopian cohort study correlates with tuberculosis clinical status. Eur. J. Immunol. **40:**291-301.

9.  **Alderson, M. R., Bement, T., Day, C. H., Zhu, L., Molesh, D., Skeiky, Y. A., Coler, R., Lewinsohn, D. M., Reed, S. G. and Dillon, D. C.** 2000. Expression cloning of an immunodominant family of *Mycobacterium tuberculosis* antigens using human CD4(+) T cells. J. Exp. Med. **191:**551-560.

10. **Amoudy HA, Al-Turab MB, Mustafa AS.** 2006. Identification of transcriptionally active open reading frames within the RD1 genomic segment of Mycobacterium tuberculosis. Med Princ Pract. **15:**137-144.

11. **Andersen, P., Andersen, A. B., Sørensen, A. L. and Nagai, S.** 1995. Recall of long lived immunity to *Mycobacterium tuberculosis* infection in mice. J. Immunol. **154:**3359-72.

12. **Andersen, P., Munk, M. E., Pollock, J. M. and Doherty, T. M.** 2000. Specific immunebased diagnosis of tuberculosis. Lancet. **356:**1099-104.

13. **Arbing, M. A., Kaufmann, M., Phan, T., Chan, S., Cascio, D. and Eisenberg, D.** 2010. The crystal structure of the *Mycobacterium tuberculosis* Rv3019c-Rv3020c ESX complex reveals a domain-swapped heterotetramer. Protein. Sci. **19:**1692-1703.

14. **Arlehamn, C. S., Sidney, J., Henderson, R., Greenbaum, J. A., James, E. A., Moutaftsi, M., Coler, R., McKinney, D. M., Park, D., Taplitz, R., Kwok, W. W., Grey, H., Peters, B. and Sette, A.** 2012. Dissecting mechanisms of immunodominance to the common tuberculosis antigens ESAT-6, CFP10, Rv2031c (*hspX*), Rv2654c (TB7.7), and Rv1038c (*EsxJ*). J. Immunol. **188:**5020-5031.

15. **Arpaia, N., Godec, J., Lau, L., Sivick, K. E., McLaughlin, L. M., Jones, M. B., Dracheva, T., Peterson, S. N., Monack, D. M. and Barton, G. M.** 2011. TLR signaling is required for *Salmonella typhimurium* virulence. Cell. **144:**675-688.

16. **Ashiru, O. T., Pillay, M. and Sturm, A. W.** 2010. Adhesion to and invasion of pulmonary epithelial cells by the F15/LAM4/KZN and Beijing strains of *Mycobacterium tuberculosis*. J. Med. Microbiol. **59:**528-533.

17. **Behr, M. A., Wilson, M. A., Gill, W. P., Salamon, H., Schoolnik, G. K., Rane, S. and Small, P. M.** 1999. Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science. **284:**1520-1523.

18. **Berthet, F., Rasmussen, P., Rosenkrands, I., Andersen, P. and Gicquel, B.** 1998. A *Mycobacterium tuberculosis* operon encoding ESAT-6 and a novel low- molecular-mass culture filtrate protein (CFP-10). Microbiol. **144:**3195-3203.

19. **Bertholet, S., Ireton, G. C., Kahn, M., Guderian, J., Mohamath, R., Stride, N., Laughlin, E. M., Baldwin, S. L., Vedvick, T. S., Coler, R. N. and Reed, S. G.** 2008. Identification of human T cell antigens for the development of vaccines against *Mycobacterium tuberculosis*. J. Immunol. **181:**7948-7957.

20. **Bessède, E., Angla-Gre, M., Delagarde, Y., Sep Hieng, S., Mènard, A. and Mègraud, F.** 2011. Matrix-assisted laserdesorption/ionization biotyper: experience in the routine of a University hospital. Clin. Microbiol. Infect. **17**:533-538.

21. **Betts, J. C., Lukey, P. T., Robb, L. C., McAdam, R. A. and Duncan, K.** 2002. Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. Mol. Microbiol. **43:**717-731.

22. **Bitter, W., Houben, E. N. G., Bottai, D., Brodin, P., Brown, E. J., Cox, J. S., Derbyshire, K., Fortune, S. M., Gao, L.-Y., Liu, J., Gey van Pittius, N. C., Pym, A. S., Rubin, E. J., Sherman, D. R., Cole, S. T. and Brosch, R.** 2009. Systematic Genetic Nomenclature for Type VII Secretion Systems. PloS. Pathog. **5:**e1000507.

23. **Bizzini, A., Durussel, C., Bille, J., Greub, G. and Prod'hom, G.** 2010. Performance of matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of bacterial strains routinely isolated in a clinical microbiology laboratory. J. Clin. Microbiol. **48:**1549 -1554.

24. **Blasco, B., Chen, J. M., Hartkoom, R., Sala, C., Uplekar, S., Rougemont, J., Pojer, F. and Cole, S. T.** 2012. Virulence Regulator EspR of *Mycobacterium tuberculosis* Is a Nucleoid-Associated Protein. PloS Pathog. **8:**e1002621.

25. **Bottai, D., Di Luca, M., Majlessi, L., Frigui, W., Simeone, R., Sayes, F., Bitter, W., Brennan, M. J., Leclerc, C., Batoni, G., Campa, M., Brosch, R. and Esin, S.** 2012. Disruption of the ESX-5 system of *Mycobacterium tuberculosis* causes loss of PPE protein secretion, reduction of cell wall integrity and strong attenuation. Mol. Microbiol. **83:**1195-1209.

26. **Brandt, L., Oettinger, T., Holm, A., Andersen, A. B and Andersen, P.** 1996. Key epitopes on the ESAT-6 antigen recognized in mice during the recall of protective immunity to *Mycobacterium tuberculosis.* J. Immunol. **157:**3527-3533.

27. **Brites, D. and Gagneux, S.** 2012. Old and new selective pressures on *Mycobacterium tuberculosis.* Infect. Genet. Evol. **12:**678-685.

28. **Brodin, P., Majlessi, L., Brosch, R., Smith, D., Bancroft, G., Clark, S., Williams, A., Leclerc, C. and Cole, S. T.** 2004. Enhanced protection against extrapulmonary tuberculosis with an attenuated live *Mycobacterium microti* vaccine inducing T-cell immunity against RD1 antigens. J. Infect. Dis. **190:**115-122.

29. **Brodin, P., Majlessi, L., Marsollier, L., de Jonge, M. I., Bottai, D., Demangel, C., Hinds, J., Neyrolles, O., Butcher, P. D., Leclerc, C., Cole, S. T. and Brosch, R.** 2006. Dissection of ESAT-6 system 1 of *Mycobacterium tuberculosis* and impact on immunogenicity and virulence. Infect. Immun. **74:**88-98.

30. **Brodin, P., Poquet, Y., Levillain, F., Peguillet, I., Larrouy-Maumus, G., Gilleron, M., Ewann, F., Christophe, T., Fenistein, D., Jang, J., Jang, M., Park, S., Rauzier, J., Carralot, J., Shrimpton, R., Genovesio, A., Gonzalo-Asensio, J. A., Puzo, G., Martin, C., Brosch, R., Stewart, G. R., Gicquel, B. and Neysolles, O.** 2010. High content phenotypic cell-based visual screen identifies *Mycobacterium tuberculosis* acyltrehalose-containing glycolipids involved in phagosome remodeling. PLoS Pathog. **6:**e1001100.

31. **Brosch, R., Gordon, S. V., Pym, A., Eiglmeier, K., Garnire, T. and Cole, S. T.** 2000. Comparative genomics of the mycobacteria. Int. J. Med. Microbiol. **290:**143-152.

32. **Brosch, R., Pym, A. S., Gordon, S. V. and Cole, S. T.** 2001. The evolution of mycobacterial pathogenicity: clues from comparative genomics. Trend. Microbiol. **9:**452-458.

33. **Broussard, G. W. and Ennis, D. G.** 2007. *Mycobacterium marinum* produces long- term chronic infections in medaka: a new animal model for studying human tuberculosis. Comp. Biochem. Physiol. **145:**45-54.

34. **Buddle, B. M., Ryan, T. J., Pollock, J. M., Andersen, P. and de Lisle, G. W.** 2001. Use of ESAT-6 in the interferon-gamma test for diagnosis of bovine tuberculosis following skin testing. Vet. Microbiol. **80:**37-46.

35. **Bukka, A., Price, C. T., Kernodle, D. S. and Graham, J. E.** 2011. *Mycobacterium tuberculosis* RNA Expression Patterns in Sputum Bacteria Indicate Secreted Esx Factors Contributing to Growth are Highly Expressed in Active Disease. Front Microbiol. **2:**266.

36. **Burts, M. L., DeDent, A. C. and Missiakas, D. M.** 2008. EsaC substrate for the ESAT-6 secretion pathway and its role in persistent infections of *Staphylococcus aureus*. Mol. Microbiol. **69:**736-746.

37. **Bustin, S. A.** 2002. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. J. Mol. Endocrinol. **29:**23-39.

38. **Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M. W., Shipley, G. L., Vondesompele, J and Wittwer, C. T.** 2009. The MIQE Guidelines: Minimum Information for Publication of Qutantitative Real-Time PCR Experiments. Clin. Chem. **55:**611-622.

39. **Callahan, B., Nguyen, K., Collins, A., Valdes, K., Caplow, M., Crossman, D. K., Steyn, A. J., Eisele, L. and Derbyshire, K. M.** 2010. Conservation of structure and protein-protein interactions mediated by the secreted mycobacterial proteins *EsxA*, *EsxB*, and *EspA*. J. Bacteriol. **192:**326-335.

40. **Calmette, A. and Plotz, H.** 1929. Protective inoculation against tuberculosis with BCG. Am. Rev. Tuberc. **19:**567-572.

41. **Cardoso, F. L. L., Antas, P. R. Z., Milagres, A. S., Geluk, A., Franken, K. L., Oliveira, E. B., Teixeira, H. C., Nogueira, S. A., Sarno, E. N., Klatser, P., Ottenhoff, T. H. and Sampaio, E. P.** 2002. T-cell responses to the *Mycobacterium tuberculosis*-specific antigen ESAT-6 in Brazilian tuberculosis patients. Infect. Immun. **70:**6707-6714.

42. **Carlsson, F., J. Kim, C. Dumitru, K. H. Barck, R. A. Carano, M. Sun, L. Diehl, and E. J. Brown.** 2010. Host-detrimental role of Esx-1-mediated inflammasome activation in mycobacterial infection. PLoS Pathog **6:**e1000895.

43. **Cascioferro, A., M. H. Daleke, M. Ventura, V. Dona, G. Delogu, G. Palu, W. Bitter, and R. Manganelli.** 2011. Functional dissection of the PE domain responsible for translocation of PE_PGRS33 across the mycobacterial cell wall. PLoS One **6:**e27713.

44. **Cerdeño-Tárraga, A. M., Efstratiou, A., Dover, L. G, Holden, M. T., Pallen, M., Bentley, S. D., Besra, G. S., Churcher, C., James, K. D., Zoysa, A., Chillingworth, T., Cronin, A., Dowd, L., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Moule, S., Quail, M. A., Rabbinowitsch, E., Rutherford, K. M., Thomson, N. R., Unwin, L., Whitehead, S., Barrell, B. G. and Parkhill, J.** 2003. The complete genome sequence and analysis of *Corynebacterium diphtheriae* NCTC13129. Nucleic Acids Res. **31:**6516-23.

45. **Chapman, A. L., Munkanta, M., Wilkinson, K. A., Pathan, A. A., Ewer, K., Ayles, H., Reece, W. H., Mwinga, A., Godfrey-Faussett, P. and Lalvani, A.** 2002. Rapid detection of active and latent tuberculosis infection in HIV-positive individuals by enumeration of *Mycobacterium tuberculosis*-specific T cells. AIDS. **16:**2285-2293.

46. **Colangeli, R., Spencer, J.S., Bifani, P., Williams, A., Lyaschenko, K., Keen, M. A., Hill, P. J., Belisle, J. and Gennaro, M. L.** 2000. MTSA-10, the product of the Rv3874 gene of *Mycobacteria tuberculosis*, elicits tuberculosis-specific, delayed-type hypersensitivity in guniea pigs. Infect. Immun. **68:**990-993.

47. **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, M. A., 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.

48. **Chakhaiyar, P., Nagalakshmi, Y., Aruna, B., Murthy, K. J., Katoch, V. M. and Hasnain, S. E.** 2004. Regions of high antigenicity within the hypothetical PPE major polymorphic tandem repeat open-reading frame, Rv2608, show a differential humoral response and a low T cell response in various categories of patients with tuberculosis. J. Infect. Dis. **190**:1237-1244.

49. **Champion, M., Williams, E. A., Kennedy, G. M. and Champion, P. A. D.** 2012. Direct detection of bacterial protein secretion using whole colony proteomics. Mol. Cell. Proteom. **11:**596-604.

50. **Choudhary, R. K., Mukhopadhyay, S., Chakhaiyar, P., Sharma, N., Murthy, K. J. R., Katoch, V. M. and Hasnain, S. E.** 2003. PPE antigen Rv2430c of *Mycobacterium tuberculosis* induces a strong B-cell response. Infect. Immun. **71**:6338-6343.

51. **Cockle, P. J., Gordon, S. V., Lalvani, A., Buddle, B. M., Hewinson, R. G. and Vordermeier, H. M.** 2002. Identification of Novel *Mycobacterium tuberculosis* Antigens with Potential as Diagnostic Reagents or Subunit Vaccine Candidates by Comparative Genomics. Infect. Immun. **70:**6996-7003.

52. **Comas, I., Chakravartti, J., Small, P. M., Galagan, J., Niemann, S., Kremer, K., Ernst, J. D. and Gagneux, S.** 2010. Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. Nat. Genet. **42:**498-503.

53. **Comas, I. and Gagneux, S.** 2011. A role for systems epidemiology in tuberculosis research. Trends Microbiol. **19:**492-500.

54. **Converse, S. E. and Cox, J. S.** 2005. A protein secretion pathway critical for *Mycobacterium tuberculosis* virulence is conserved and functional in *Mycobacterium smegmatis*. J. Bacteriol. **187:**1238-1245.

55. **Daleke, M. H., Cascioferro, A., de Punder, K., Ummels, R., Abdallah, A. M., van der Wel, N., Peters, P. J., Luirink, J., Manganelli, R. and Bitter, W.** 2011. Conserved Pro-Glu (PE) and Pro-Pro-Glu (PPE) protein domains target LipY lipases of pathogenic mycobacteria to the cell surface via the ESX-5 pathway. J. Biol. Chem. **286:**19024-19034.

56. **Daniel, T. M.** 2006. The history of tuberculosis. Respir. Med. **100:**1862-70.

57. **Das, C., Ghosh, T. S. and Mande, S. S.** 2011. Computational analysis of the ESX-1 region of *Mycobacterium tuberculosis*: insights into the mechanism of type VII secretion system. PLoS One. **6:**e27980.

58. **Davila, J., Zhang, L., Marrs, C. F., Durmaz, R. and Yang, Z.** 2010. Assessment of the genetic diversity of *Mycobacterium tuberculosis esxA, esxH,* and *fbpB* genes among clinical isolates and its implication for the future immunization by new tuberculosis subunit vaccines Ag85B-ESAT-6 and Ag85B-TB10.4. J. Biomed. Biotechnol. **2010:**208371.

59. **Davis, J. M., Clay, H., Lewis, J. L., Ghori, N., Herbomel, P. and Ramakrishnan, L.** 2002. Real-time visualization of mycobacterium-macrophage interactions leading to initiation of granuloma formation in zebrafish embryos. Immunity. **17:**693-702.

60.     **Deb, C., Daniel, J., Sirakova, T. D., Abomoelak, B., Dubey, V. S. and Kolattukudy, P. E.** 2006. A novel lipase belonging to the hormone-sensitive lipase family induced under starvation to utilize stored triacylglycerol in *Mycobacterium tuberculosis*. J. Biol. Chem. **281:**3866-3875.

61.     **DiGiuseppe Champion, P. A., Champion, M. M., Manzanillo, P. and Cox, J. S.** 2009. ESX-1 secreted virulence factors are recognized by multiple cytosolic AAA ATPases in pathogenic mycobacteria. Mol. Microbiol. **73:**950-962.

62.     **Dillon, D. C., Alderson, M. R., Day, C. H., Bement, T., Campos-Neto, A., Skeiky, Y. A., Vedvick, T., Badaro, R., Reed, S. G. and Houghton, R.** 2000. Molecular and immunological characterization of *Mycobacterium tuberculosis* CFP-10, an immunodiagnostic antigen missing in *Mycobacterium bovis BCG*. J. Clin. Microbiol. **38:**3285-3290.

63.     **Dillon, D. C., Alderson, M. R., Day, C. H., Lewinsohn, D. M., Coler, R., Bement, T., Campos-Neto, A., Skeiky, Y. A., Orme, I. M., Roberts, A., Steen, S., Dalemans, W., Badaro, R. and Reed, S. G.** 1999. Molecular characterization and human T-cell responses to a member of a novel *Mycobacterium tuberculosis* mtb39 gene family. Infect. Immun. **67:**2941-2950.

64.     **Doherty, T. M., Demissie, A., Olobo, J., Wolday, D., Britton, S., Eguale, T., Ravn, P. and Andersen, P.** 2002. Immune responses to the *Mycobacterium tuberculosis*-specific antigen ESAT-6 signal subclinical infection among contacts of tuberculosis patients. J. Clin. Microbiol. **40:**704-706.

65.     **Drummond, A. J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., Heled, J., Kearse, M., Markowitz, S., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T. and Wilson, A.** 2012. Geneious, **v5.6**, http://www.geneious.com.

66. **Dubnau, E., Fontan, P., Manganelli, R., Soares-Appel, S. and Smith, I.** 2002. *Mycobacterium tuberculosis* genes induced during infection of human macrophages. Infect. Immun. **70:**2787-2795.

67. **Elhay, M. J., Oettinger, T. and Andersen, P.** 1998. Delayed-type hypersensitivity responses to ESAT-6 and MPT64 from *Mycobacterium tuberculosis* in the guinea pig. Infect. Immun. **66:**3454-3456.

68. **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. J. Clin. Microbiol. **31:**406-409.

69. **Ewer, K., Deeks, J., Alvarez, L., Bryant, G., Waller, S., Andersen, P., Monk, P. and Lalvani, A.** 2003. Comparison of T-cell-based assay with tuberculin skin test for diagnosis of *Mycobacterium tuberculosis* infection in a school tuberculosis outbreak. Lancet. **361:**1168-1173.

70. **Filliol, I., Motiwala, A. S., Cavatore, M., Qi, W., Hazbón, M. H., Bobadilla del Valle, M., 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., Ponce de León, A., 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. J. Bact. **188:**759-772.

71. **Finn, R. D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S. R., Sonnhammer, E. L. and Bateman, A.** 2006. Pfam: clans, web tools and services. Nucleic Acids Res. **34**:D247-D251.

72.  **Fisher MA, Plikaytis BB, Shinnick TM.** 2002. Microarray analysis of the *Mycobacterium tuberculosis* transcriptional response to the acidic conditions found in phagosomes. J. Bacteriol. **184:**4025-4032.

73.  **Fletcher, H. A.** 2007. Correlates of immune protection from tuberculosis. Curr. Mol. Med. **7:**319-325.

74.  **Flynn, J. L., Goldstein, M. M., Chan, J., Triebold, K. J., Pfeffer, K., Lowenstein, C. J., Schreiber, R., Mak, T. W and Bloom, B. R.** 1995. Tumor necrosis factor-alpha is required in the protective immune response against *Mycobacterium tuberculosis* in mice. Immunity. **2:**561-572.

75.  **Fortune, S. M., Jaeger, A., Sarracino, D. A., Chase, M. R., Sassetti, C. M., Sherman, D. R., Bloom, B. R. and Rubin, E. J.** 2005. Mutually dependent secretion of proteins required for mycobacterial virulence. PNAS. **102:**10676-10691.

76.  **Frigui, W., Bottai, D., Majlessi, L., Monot, M., Josselin, E., Brodin, P., Gar nier, T., Gicquel, B., Martin, C., Leclerc, C., Cole, S. T. and Brosch, R.** 2008. Control of *M. tuberculosis* ESAT-6 Secretion and Specific T Cell Recognition by PhoP. PLoS Pathog. **4:**e33.

77.  **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. and Small, P. M.** 2006. Variable host–pathogen compatibility in *Mycobacterium tuberculosis*. PNAS. **103:**2869-2873.

78.  **Gao, L. Y., Guo, S., McLaughlin, B., Morisaki, H., Engel, J. N. and Brown, E. J.** 2004. A mycobacterial virulence gene cluster extending RD1 is required for cytolysis, bacterial spreading and ESAT-6 secretion. Mol. Microbiol. **53:**1677-1693.

79.  **Garufi, G., Butler, E. and Missiakas, D.** 2008. ESAT-6-like protein secretion in *Bacillus anthracis*. J. Bacteriol. **190:**7004-7011.

80.  **Gey van Pittius, N. C., Gamieldien, J., Hide, W., Brown, G., Siezen, R. and Beyers, A.** 2001. The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C Gram-positive bacteria. Genome Biol. **2:**research0044.1-research0044.18.

81.  **Gey van Pittius, N. C., Sampson, S. L., Lee, H., Kim, Y., van Helden, P. D. and Warren, R. M.** 2006. Evolution and expansion of the *Mycobacterium tuberculosis* PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions. BMC Evol. Biol. **6:**95.

82.  **Goldstone, R. M., Goonesekera, S. D., Bloom, B. R. and Sampson, S. L**. 2009. The transcriptional regulator Rv0485 modulates the expression of a pe and ppe gene pair and is required for *Mycobacterium tuberculosis* virulence. Infect. Immun. **77**:4654-4667.

83.  **Gonzalo-Asensio, J., Mostowy, S., Harders-Westerveen, J., Huygen, K., Hernandez-Pando, R., Thole, J., Behr, M., Gicquel, B. and Martín, C.** 2008. PhoP: A Missing Piece in the Intricate Puzzle of *Mycobacterium tuberculosis* Virulence. PLoS One. **3:**e3496.

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

85.  **Goter-Robinson, C., Derrick, S. C., Yang, A. L., Jeon, B. Y. and Morris, S. L.** 2006. Protection against an aerogenic *Mycobacterium tuberculosis* infection in BCG-immunized and DNA vaccinated mice is associated with early type I cytokine responses. Vaccine. **24:**3522-3529.

86. **Guinn, K. M., Hickey, M. J., Mathur, S. K., Zakel, K. L., Grotzke, J. E., Lewinsohn, D. M., Smith, S. and Sherman, D. R.** 2004. Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of *Mycobacterium tuberculosis*. Mol. Microbiol. **51:**359-370.

87. **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.   J. Infect. Dis. **193:**121-128.

88. **Gutacker, M., Smoot, J., Migliaccio, C., Ricklefs, S., Hua, S., Cousins, D., Graviss, E., Shashkina, E., Kreiswirth, B. and Musser, J.** 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.

89. **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*. PLOS Pathog. **1:**e5.

90. **Harboe, M., Oettinger, T., Wiker, H. G., Rosenkrands, I. and Andersen, P.** 1996. Evidence for occurence of the ESAT-6 protein in *Mycobacteria tuberculosis* and virulent *Mycobacterium bovis* and for its absence in *Mycobacterium bovis* BCG. Infect. Immun. **64:**16-22.

91. **Hawkey, P. M., Smith, E. G., Evans, J. T., Monk, P., Bryan, G., Mohamed, H. H., Bardhan, M. and Pugh, R. N.** 2003. Mycobacterial Interspersed Repetitive Unit Typing of *Mycobacterium tuberculosis* compared to IS*6110*-Based Restriction Fragment Length Polymorphism Analysis for investigation of apparently clustered cases of tuberculosis. J. Clin. Microbiol. **41:**3514-3520.

92.  **Hershberg, R., Lipatov, M., Small, P. M., Sheffer, H., Niemann, S., Homolka, S., Roach, J. C., Kremer, K., Petrov, D. A., Feldman, M. W and Gagneux, S.** 2008. High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. PLoS Biol. **6:**e311.

93.  **Hervas-Stubbs, S., Majlessi, L., Simsova, M., Morova, J., Rojas, M. J., Nouzé, C., Brodin, P., Sebo, P. and Leclerc, C.** 2006. High frequency of CD4+ T cells specific for the TB10.4 protein correlates with protection against *Mycobacterium tuberculosis* infection. Infect. Immun. **74:**3396-3407.

94.  **Hettick, J. M., Kashon, M. L., Simpson, J. P., Siegel, P. D., Mazurek, G. H. and Weissman, D. N.** 2004. Proteomic profiling of intact mycobacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Anal. Chem. **76:**5769-5776.

95.  **Hu, Y. and Coates, A. R.** 1999. Transcription of two sigma 70 homologue genes, sigA and sigB, in stationary phase *Mycobacterium tuberculosis*. J. Bacteriol. **181:**469-476.

96.  **Huard, R. C., Fabre, M., de Haas, P., Lazzarini, L. C., van Soolingen, D., Cousins, D. and Ho, J. L.** 2006. Novel Genetic Polymorphisms That Further Delineate the Phylogeny of the *Mycobacterium tuberculosis* Complex. J. Bacteriol. **188:**4271-4287.

97.  **Hussain, R., Talat, N., Shahid, F. and Dawood, G.** 2007. Longitudinal tracking of cytokines after acute exposure to tuberculosis: association of distinct cytokine patterns with protection and disease development. Clin. Vaccine Immunol. **14:**1578-1586.

98.  **Hsu, T., Hingley-Wilson, S. M., Chen, B., Chen, M., Dai, A. Z., Morin, P. M., Marks, C. B., Padiyar, J., Goulding, C., Gingery, M., Eisenberg, D., Russell, R. G., Derrick, S. C., Collins, F. M., Morris, S. L., King, C. H. and Jacobs, W. R., Jr.** 2003. The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. PNAS. **100:**12420-12425.

99.  **Ilghari, D., Lightbody, K. L., Veverka, V., Waters, L. C., Muskett, F. W., Renshaw, P. S. and Carr, M. D.** 2011. Solution structure of the *Mycobacterium tuberculosis* EsxG.EsxH complex: functional implications and comparisons with other M. tuberculosis Esx family complexes. J. Biol. Chem. **286:**29993-30002.

100. **Jones, G. J., Gordon, S. V., Hewinson, R. G. and Vordermeier, H. M.** 2010b. Screening of predicted secreted antigens from *Mycobacterium bovis* reveals the immunodominance of the ESAT-6 protein family. Infect. Immun. **78:**1326-1332.

101. **Jones, G. J., Hewinson, R. G. and Vordermeier, H. M.** 2010a. Screening of predicted secreted antigens from *Mycobacterium bovis* identifies potential novel differential diagnostic reagents. Clin. Vaccine Immunol. **17:**1344-1348.

102. **de Jonge, M. I., Pehau-Arnaudet, G., Fretz, M. M., Romain, F., Bottai, D., Brodin, P., Honore, N., Marchal, G., Jiskoot, W., Endgland, P., Cole, S. T. and Brosch, R.** 2007. ESAT-6 from *Mycobacterium tuberculosis* Dissociates from Its Putative Chaperone CFP-10 under Acidic Conditions and Exhibits Membrane-Lysing Activity. J. Bacteriol. **189:**6028-6034.

103. **Junqueira-Kipnis, A. P., Basaraba, R. J., Gruppo, V., Palanisamy, G., Turner, O. C., Hsu, T., Jacobs, W. R. Jr., Fulton, S. A., Reba, S. M., Boom, W. H. and Orme, I. M.** 2006. Mycobacteria lacking the RD1 region do not induce necrosis in the lungs of mice lacking interferon-gamma. Immunol. **119:**224-231.

104. **Kanaujia, G. V., Garcia, M. A., Bouley, D. M., Peters, R. and  Gennaro, M. L.** 2003. Detection of early secretory antigenic target-6 antibody for diagnosis of tuberculosis in non-human primates. Comp. Med. **53:**602-606.

105. **Kato-Maeda, M., Bifani, P. J., Kreiswirth, B. N. and Small, P. M.** 2001. The nature and consequence of genetic variability within *Mycobacterium tuberculosis.* J. Clin. Invest. **107:**553-537.

106.  **Kaufmann, S. H. E., Hussey, G. and Lambert, P. H.** 2010. New vaccines for tuberculosis. Lancet. **375:**2110-2119.

107.  **Kawamura, I.** 2006. Protective immunity against *Mycobacterium tuberculosis*. Kekkaku. **81:**687-691.

108.  **Khan, N., Alam, K., Nair, S., Valluri, V. L., Murthy, K. J. and Mukhopadhyay, S.** 2008. Association of strong immune responses to PPE protein Rv1168c with active tuberculosis. Clin. Vaccine Immunol. **15**:974-980.

109.  **Lalvani, A., Nagvenkar, P., Udwadia, Z., Pathan, A. A., Wilkinson, K. A., Shastri, J. S., Ewer, K., Hill, A. V., Mehta, A. and Rodrigues, C.** 2001a. Enumeration of T cells specific for RD1-encoded antigens suggests a high prevalence of latent *Mycobacterium tuberculosis* infection in healthy urban Indians. J. Infect. Dis. **183:**469-477.

110.  **Lalvani, A., Pathan, A. A., Durkan, H., Wilkinson, K. A., Whelan, A., Deeks, J. J., Reece, W. H., Latif, M., Pasvol, G. and Hill, A. V.** 2001b. Enhanced contact tracing and spatial tracking of *Mycobacterium tuberculosis* infection by enumeration of antigen-specific T cells. Lancet. **357:**2017–2021.

111.  **Lamichhane, G., Zignol, M., Blades, N. J., Geiman, D. E., Dougherty, A., Grosset, J., Broman, K. W. and Bishai, W. R.** 2003. A postgenomic method for predicting essential genes at subsaturation levels of mutagenesis: application to *Mycobacterium tuberculosis*. PNAS. **100:**7213-7218.

112.  **Larsen, M. H., Biermann, K., Tandberg, S., Hsu, T. and Jacobs, W. R. Jr.** 2007. Genetic Manipulation of *Mycobacterium tuberculosis*. Vol. 10A.1.1-10A1.8.

113. **Lechat, P., Hummel, L., Rousseau, S. and Moszer, I.** 2008. GenoList: an integrated environment for comparative analysis of microbial genomes. Nucleic Acids Res. **36**:D469-D474.
[http://genodb.pasteur.fr/cgi-bin/WebObjects/GenoList.woa/wa/goToMainPage]
[Accessed Online: 02 February 2013].

114. **Lew, J. M., Kapopoulou, A., Jones, L. M. and Cole, S. T.** 2011. TubercuList-10 years after, p. 1-7, Tuberculosis, vol. 91, Edinb.
[http://genolist.pasteur.fr/TubercuList/] [Accessed Online: 15 January 2013].

115. **Lewis, K. N., Liao, R., Guinn, K. M., Hickey, M. J., Smith, S., Behr, M. A. and Sherman, D. R.** 2003. Deletion of RD1 from *Mycobacterium tuberculosis* mimics bacille Calmette-Guérin attenuation. J. Infect. Dis. **187:**117-123.

116. **Lightbody, K. L., Ilghari, D., Waters, L. C., Carey, G., Bailey, M. A., Williamson, R. A., Renshaw, P. S. and Carr, M. D.** 2008. Molecular features governing the stability and specificity of functional complex formation by *M. tuberculosis* CFP-10/ESAT-6 family proteins. J. Biol. Chem **283:**17681-17690.

117. **Lopez-Vidal, Y., Leon-Rosales, S. P., Castanon-Arreola, M., Srangel-Frausto, M. S., Melendez-Herrada, E. and Sada-Diaz, E.** 2004. Response of IFN-γ and IgG to ESAT-6 and 38 kDa recombinant proteins and their peptides from *Mycobacterium tuberculosis* in tuberculosis patients and asymptomatic household contacts may indicate possible early stage infection in the latter. Arch. Med. Res. **35:**308-317.

118. **Louise, R., Skjøt, V., Agger, E. M. and Andersen, P.** 2001. Antigen discovery and tuberculosis vaccine development in the postgenomic era. Scand. J. Infect. Dis. **33:**643–647.

119. **Ma, J., Chen, T., Mandelin, J., Ceponis, A., Miller, N. E., Hukkanen, M., Ma, G. F. and Konttinen, Y. T.** 2003. Regulation of macrophage activation. Cell. Mol. Life Sci. **60:**2334–2346.

120. **Macedo, G. C., Bozzi, A., Weinreich, H. R., Bafica, A., Teixeira, H. C. and Oliveira, S. C.** 2011. Human T Cell and Antibody-Mediated Responses to the *Mycobacterium tuberculosis* Recombinant 85A, 85B, and ESAT-6 Antigens. Clin. Dev. Immunol. **2011:**1-10.

121. **MacGurn, J. A., Raghavan, S., Stanley, S. A. and Cox, J. S.** 2005. A non-RD1 gene cluster is required for Snm secretion in *Mycobacterium tuberculosis*. Mol. Microbiol. **57:**1653-1663.

122. **Maciag, A., Dainese, E., Rodriguez, G. M., Milano, A., Provvedi, R., Pasca, M. R,, Smith, I., Palù, G., Riccardi, G. and Manganelli, R.** 2007. Global analysis of the *Mycobacterium tuberculosis* Zur (FurB) regulon. J. Bacteriol. **189:**730-740.

123. **Maheiras, G. G., Sabo, P. J., Hickey, M. J., Devinder, C. S. and C. K. Stover.** 1996. Molecular analysis of genetic differences between *Mycobacterium bovis BCG* and virulent *M. bovis.* J. Bacteriol. **178:**1274-1282.

124. **Majlessi, L., Rojas, M. J., Brodin, P. and Leclerc, C.** 2003. CD8+ -T-cell responses of *Mycobacterium*-infected mice to a newly identified major histocompatibility complex class I-restricted epitope shared by proteins of the ESAT-6 family. Infect. Immun. **71:**7173-7177.

125. **Malen, H., Berven, F. S., Fladmark, K. E. and Wiker, H. G.** 2007. Comprehensive analysis of exported proteins from *Mycobacteriu tuberculosis* H37Rv. Proteomics. **7:**1702-1718.

126. **Malen, H., Softeland, T. and Wiker, H. G.** 2008. Antigen analysis of *Mycobacterium tuberculosis* H37Rv culture filtrate proteins. Scand J Immunol. **67:**245-252.

127. **Malik, A. N. and Godfrey-Fausset, P.** 2005. Effectes of genetic variability of *Mycobacterium tuberculosis* strains on the presentation of disease. Lancet Infect. Dis. **5:**174-183.

128. **Manganelli, R., Dubnau, E., Tyagi, S., Kramer, F. R. and Smith, I.** 1999. Differential expression of 10 sigma factor genes in *Mycobacterium tuberculosis*. Mol. Microbiol. **31:**715-724.

129. **Manganelli, R., Provvedi, R., Rodrigue, S., Beaucher, J., Gaudreau, L. and Smith, I.** 2004. Sigma factors and global gene regulation in *Mycobacterium tuberculosis*. J. Bacteriol. **186:**895-902.

130. **Manzanillo, P. S., Shiloh, M. U., Portnoy, D. A. and Cox, J. S.** 2012. *Mycobacterium tuberculosis* activates the DNA-dependent cytosolic surveillance pathway within macrophages. Cell Host Microbe. **11:**469-480.

131. **Marmiesse, M., Brodin, P., Buchrieser, C., Gutierrez, C., Simoes, N., Vincent, V., Glaser, P., Cole, S. T. and Brosch, R.** 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. Microbiol. **150:**483-496.

132. **Mattow J, Schaible UE, Schmidt F, Hagens K, Siejak F, Brestrich G, Haeselbarth G, Muller EC, Jungblut PR, Kaufmann SH.** 2003. Comparative proteome analysis of culture supernatant proteins from virulent Mycobacterium tuberculosis H37Rv and attenuated M. bovis BCG Copenhagen. Electrophoresis. **24:**3405-3420

133. **Mazurek, G. H., LoBue, P. A., Daley, C. L., Bernardo, J., Lardizabal, A. A., Bishai, W. R., Iademarco, M. F. and Rothel, J. S.** 2001. Comparison of a whole-blood interferon gamma assay with tuberculin skin testing for detecting latent *Mycobacterium tuberculosis* infection. JAMA. **286:**1740-1747.

134. **McLaughlin, B., Chon, J. S., MacGurn, J. A., Carlsson, F., Cheng, T. L., Cox, J. S. and Brown, E. J.** 2007. A mycobacterium ESX-1-secreted virulence factor with unique requirements for export. PLoS Pathog. **3:**e105

135. **Meher, A. K., Bal, N. C., Chary, K. V. and Arora, A.** 2006. *Mycobacterium tuberculosis* H37Rv ESAT-6-CFP-10 complex formation confers thermodynamic and biochemical stability. FEBS J. **273:**1445-1462.

136. **Meier, T., Eulenbruch, H. P., Wrighton-Smith, P., Enders, G. and Regnath, T.** 2005. Sensitivity of a new commercial enzyme-linked immunospot assay (T SPOT-TB) for diagnosis of tuberculosis in clinical practice. Eur. J. Clin. Microbiol. Infect. Dis. **24:**529-536.

137. **Mestre, O., Luo, T., Dos Vultos, T., Kremer, K., Murray, A., Namouchi, A., Jackson, C., Rauzier, J., Bifani, P., Warren, R., Rasolofo, V., Mei, J., Gao, Q. and Gicquel, B.** 2011. Phylogeny of *Mycobacterium tuberculosis* Beijing Strains Constructed from Polymorphisms in Genes Involved in DNA Replication, Recombination and Repair. PLoS One. **6:**e16020.

138. **Mishra, K. C., de Chastellier, C., Narayana, Y., Bifani, P., Brown, A. K., Besra, G. S., Katoch, V. M., Joshi, B., Balaji, K. N. and Kremer, L.** 2008. Functional role of the PE domain and immunogenicity of the *Mycobacterium tuberculosis* triacylglycerol hydrolase LipY. Infect. Immun. **76:**127-140.

139. **Mitchell, J. E., Chetty, S., Govender, P., Pillay, M., Jaggernath, M., Kasmar, A., Ndung'u, T., Klenerman, P., Walker, B. D. and Kasprowicz, V. O.** 2012. Prospective Monitoring Reveals Dynamic Levels of T Cell Immunity to *Mycobacterium tuberculosis* in HIV Infected Individuals. PLoS One. **7:**e37920.

140. **Mohan, V. P., Scanga, C. A., Yu, K., Scott, H. M., Tanaka, K. E., Tsang, E., Tsai, M. M., Flynn, J. L. and Chan, J.** 2001. Effects of tumor necrosis factor alpha on host immune response in chronic persistent tuberculosis: possible role for limiting pathology. Infect. Immun. **69:**1847-1855.

141. **Musser, J. M., Amin, A. and Ramaswamy, S.** 2000. Negligible genetic diversity of *Mycobacterium tuberculosis* host immune system protein targets: evidence of limited selective pressure. Genetics. **155:**7-16.

142. **Nei, M. and Gojobori, T.** 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:**418-426.

143. **Neville, S. A., Lecordier, A., Ziochos, H., Chater, M. J., Gosbell, I. B., Maley, M. W. and van Hal, S. J.** 2011. Utility of matrix-assisted laser desorption ionization-time of flight mass spectrometry following introduction for routine laboratory bacterial identification. J. Clin. Microbiol. **49:**2980-2984.

144. **Nicol, M. P. and Wilkinson, R. J.** 2008. The clinical consequences of strain diversity in *Mycobacterium tuberculosis*. Trans. R. Soc. Trop. Med. Hyg. **102:**955-565.

145. **Novikov, A., Cardone, M., Thompson, R., Shenderov, K., Kirschman, K. D., Mayer-Barber, K. D., Myers, T. G., Rabin, R. L., Trinchieri, G., Sher, A. and Feng, C. G.** 2011. *Mycobacterium tuberculosis* Triggers Host Type I IFN Signaling To Regulate IL-1β Production in Human Macrophages. J. Immunol. **187:**2540-2547.

146. **Ogawa, T., Uchida, H., Kusumoto, Y., Mori, Y., Yamamura, Y. and Hamada, S.** 1991. Increase in tumor necrosis factor alpha- and interleukin-6-secreting cells in peripheral blood mononuclear cells from subjects infected with *Mycobacterium tuberculosis*. Infect. Immun. **59:**3021–3025.

147. **Ohol, Y. M., Goetz, D. H., Chan, K., Shiloh, M. U., Craik, C. S. and Cox, J. S.** 2010. *Mycobacterium tuberculosis* MycP1 protease plays a dual role in regulation of ESX-1 secretion and virulence. Cell. Host Microbe. **7:**210 –220.

148. **Okkels, L. M., Brock, I., Follmann, F., Agger, E. M., Arend, S. M., Ottenhoff, T. H. M., Oftung, F., Rosenkrands, I. and Andersen, P.** 2003. PPE Protein (Rv3873) from DNA Segment RD1 of *Mycobacterium tuberculosis*: Strong Recognition of Both Specific T-Cell Epitopes and Epitopes Conserved within the PPE Family. 2003. Infect. Immun. **71:**6116-6123.

149. **Okkels, L. M. and P. Andersen.** 2004. Protein-Protein Interactions of Proteins from the ESAT-6 Family of *Mycobacterium tuberculosis*. J. Bacteriol. **186:**2487-2491.

150. **Pai, M., Riley, L. W. and Colford, J. M. Jr.** 2004. Interferon-gamma assays in the immunodiagnosis of tuberculosis: a systematic review. Lancet Infect. Dis. **4:**761-776.

151. **Pallen, M. J.** 2002. The ESAT-6/WXG100 superfamily - and a new Gram-positive secretion system? Trend Microbiol. **10:**209-212.

152. **Pheiffer, C., Betts, J. C., Flynn, H. R., Lukey, P. T. and van Helden, P.** 2005. Protein expression by a Beijing strain differs from that of another clinical isolate and *Mycobacterium tuberculosis* H37Rv. Micro. **151:**1139-1150.

153. **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. Clin. Infect. Dis. **45:**1409-1414.

154. **Pillay, N.** 2010. Development of a genotypic test for the detection of KZN strain of *Mycobacterium tuberculosis*. University of KwaZulu-Natal, Durban, South Africa. (unpublished MSc *Thesis*)

155. **van Pinxteren, L. A., Ravn, P., Agger, E. M., Pollock, J. and Andersen, P.** 2000. Diagnosis of tuberculosis based on the two specific antigens ESAT-6 and CFP10. Clin. Diagn. Lab. Immunol. **7:**155-160.

156. **Pollock, J. M. and Andersen, P.** 1997. Predominant recognition of the ESAT-6 protein in the first phase of infection with *Mycobacterium bovis* in cattle. Infect. Immun. **65:**2587-2592.

157. **Prouty, M. G., Correa, N. E., Barker, L. P., Jagadeeswaran, P. and Klose, K. E.** 2003. Zebrafish *Mycobacterium marinum* model for mycobacterial pathogenesis. FEMS Microbiol. Lett. **225:**177-182.

158. **Pym, A. S., Brodin, P., Majlessi, L., Brosch, R., Demangel, C., Williams, A., Griffiths, K. E., Marchal, G., Leclerc, C. and Cole, S. T.** 2003. Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. Nat. Med. **9:**533-539.

159. **Raghavan, S., Manzanillo, P., Chan, K, Dovey, C. and Cox, J. S.** 2008. Secreted transcription factor controls *Mycobacterium tuberculosis* virulence. Nature. **454:**717–721.

160. **Ramakrishnan, L., Valdivia, R. H., McKerrow, J. H. and Falkow, S.** 1997. *Mycobacterium marinum* causes both long-term subclinical infection and acute disease in the leopard frog (*Rana pipiens*). Infect. Immun. **65:**767-773.

161. **Ravn, P., Demissie, A., Eguale, T., Wondwosson, H., Lein, D., Amoudy, H. A., Mustafa, A. S., Jensen, A. K., Holm, A., Rosenkrands, I., Oftung, F., Olobo, J., von Reyn, F. and Andersen, P.** 1999. Human T cell responses to the Esat-6 antigen from *Mycobacterium tuberculosis*. J. Infect. Dis. **179:**637-645.

162. **Reddy, T. B., Riley, R., Wymore, F., Montgomery, P., DeCaprio, D., Engels, R., Gellesch, M., Hubble, J., Jen, D., Jin, H., Koehrsen, M., Larson, L., Mao, M., Nitzberg, M., Sisk, P., Stolte, C., Weiner, B., White, J., Zachariah, Z. K., Sherlock, G., Galagan, J. E., Ball, C. A. and Schoolnik, G. K.** 2009. TB database: an integrated platform for tuberculosis research. Nucleic Acids Res. **37**(**Database issue):**D499-508. [http://www.tbdb.org] [Accessed Online: 18 June 2013].

163. **Renshaw, P. S., Panagiotidou, P., Whelan, A., Gordon, S. V., Hewinson, R. G., Williamson, R. A. and Carr, M. D.** 2002. Conclusive Evidence That the Major T-cell Antigens of the *Mycobacterium tuberculosis* Complex ESAT-6 and CFP-10 Form a Tight, 1:1 Complex and Characterization of the Structural Properties of ESAT-6, CFP-10, and the ESAT-6 CFP-10 Complex. IMPLICATIONS FOR PATHOGENESIS AND VIRULENCE. J. Biol. Chem. **277:**21598-21603.

164. **Rindi, L., Lari, N. and Garzelli, C.** 1999. Search for genes potentially involved in *Mycobacterium tuberculosis* virulence by mRNA differential display. Biochem. Biophys. Res. Commun. **258:**94-101.

165. **Rindi, L., Lari, N. and Garzelli, C.** 2001. Genes of *Mycobacterium tuberculosis* H37Rv downregulated in the attenuated strain H37Ra are restricted to *M. tuberculosis* complex species. New Microbiol. **24:**289-294.

166. **Rodrigue, S., Provvedi, R., Jacques, P. E., Gaudreau, L. and Manganelli, R.** 2006. The sigma factors of *Mycobacterium tuberculosis*. FEMS Microbiol. Rev. **30:**926-941.

167. **Rodriguez, G. M., Voskuil, M. I., Gold, B., Schoolnik, G.K. and Smith, I.** 2002. ideR, An essential gene in mycobacterium tuberculosis: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. Infect. Immun. **70:**3371-3381.

168. **Romano, M., Rindi, L., Korf, H., Bonanni, D., Adnet, P. Y., Jurion, F., Garzelli, C. and Huygen, K.** 2008. Immunogenicity and protective efficacy of tuberculosis subunit vaccines expressing PPE44 (Rv2770c). Vaccine. **26:**6053-6063.

169. **van der Sar, A. M., Musters, R. J., van Eeden, F. J., Appelmelk, B. J., Vandenbroucke-Grauls, C. M. and Bitter, W.** 2003. Zebrafish embryos as a model host for the real time analysis of *Salmonella typhimurium* infections. Cell. Microbiol. **5:**601-611.

170. **Sassetti, C. M., Boyd, D. H. and Rubin, E. J.** 2003b. Genes required for mycobacterial growth defined by high density mutagenesis. Mol. Microbiol. **48:**77-84.

171. **Sassetti, C. M. and Rubin, E. J.** 2003a. Genetic requirements for mycobacterial survival during infection. PNAS. **100:**12989-12994.

172. **Schürch, A. C., Kremer, K., Hendriks, A. C. A., Freyee, B., McEvoy, C. R. E., van Crevel, R., Boeree, M. J., van Helden, P., Warren, R. M., Siezen, R. J. and van Soolingen, D.** 2011. SNP/RD Typing of *Mycobacterium tuberculosis* Beijing Strains Reveals Local and Worldwide Disseminated Clonal Complexes. PLoS One. **6:**e28365.

173. **Serafini, A., Boldrin, F., Palu, G. and Manganelli, R.** 2009. Characterization of a *Mycobacterium tuberculosis* ESX-3 conditional mutant: essentiality and rescue by iron and zinc. J. Bacteriol. **191:**6340-6344.

174. **Siegrist, M. S., Unnikrishnan, M., McConnell, M. J., Borowsky, M., Cheng, T. Y., Siddiqi, N., Fortune, S. M., Moody, D. B. and Rubin, E. J.** 2009. Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition. PNAS. **106:**18792-18797.

175. **Simeone, R., Bobard, A., Lippmann, J., Bitter, W., Majlessi, L., Brosch, R. and Enninga, J.** 2012. Phagosomal Rupture by *Mycobacterium tuberculosis* Results in Toxicity and Host Cell Death. PLoS Pathog. **8:**e1002507.

176. **Skjøt, R. L. V., Brock, I., Arend, S. M., Munk, M. E., Theisen, M., Ottenhoff, T. H. M. and Andersen, P.** 2002. Epitope Mapping of the Immunodominant Antigen TB10.4 and the Two Homologous Proteins TB10.3 and TB12.9, Which Constitute a Subfamily of the esat-6 Gene Family. 2002. Infect. Immun. **70:**5446-5453.

177. **Skjøt, R., Oettinger, T., Rosenkrands, I., Ravn, P., Brock, I., Jacobsen, S. and Andersen, P.** 2000. Comparative evaluation of low-molecular-mass proteins from *Mycobacterium tuberculosis* identifies members of the ESAT-6 family as immunodominant T-cell antigens. Infect. Immun. **68:**214-220.

178. **Smith, J., Manoranjan, J., Pan, M., Bohsali, A., Xu, J., Liu, J., McDonald, K. L., Szyk, A., LaRonde-LeBlanc, N. and Gao, L. Y.** 2008. Evidence for pore formation in host cell membranes by ESX-1-secreted ESAT-6 and its role in *Mycobacterium marinum* escape from the vacuole. Infect. Immun. **76**:5478-5487.

179. **Sogawa, K., Watanabe, M., Sato, K., Segawa, S., Ishii, C., Miyabe, A., Murata, S., Saito, T. and Nomura, F.** 2011. Use of the MALDI BioTyper system with MALDI-TOF mass spectrometry for rapid identification of microorganisms. Anal. Bioanal. Chem. **400:**1905-1911.

180. **van Soolingen, D., Dehaas, P. E., Hermans, P. W. and Vanembden, J. D.** 1994. DNA fingerprinting of *Mycobacterium tuberculosis*. Meth. Enzymol. **235:**196-205.

181. **Sørensen, A. L., Nagai, S., Houen, G., Andersen, P. and Andersen, A. B.** 1995. Purification and characterization of a low-molecular-mass T-cell antigen secreted by *Mycobacterium tuberculosis.* Infect. Immun. **63:**1710-1717.

182. **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. PNAS. **94:**9869-9874.

183. **Stanley, S. A., Johndrow, J. E., Manzanillo, P. and Cox, J. S.** 2007. The type I IFN response to infection with *Mycobacterium tuberculosis* requires ESX-1 mediated secretion and contributes to pathogenesis. J. Immunol. **178:**3143-3152.

184. **Stanley, S. A., Raghavan, S., Hwang, W. W. and Cox, J. S.** 2003. Acute infection and macrophage subversion by *Mycobacterium tuberculosis* require a specialized secretion system. PNAS. **100:**13001-13006.

185. **Stewart, G. R., Snewin, V. A., Walzl, G., Hussell, T., Tormay, P., O'Gaora, P., Goyal, M., Betts, J., Brown, I. N. and Young, D. B.** 2001. Overexpression of heat-shock proteins reduces survival of *Mycobacterium tuberculosis* in the chronic phase of infection. Nat. Med. **7:**732-737.

186. **Stewart, G. R., Wernisch, L., Stabler, R., Mangan, J. A., Hinds, J., Laing, K. G., Young, D. B. and Butcher, P. D.** 2002. Dissection of the heat-shock response in Mycobacterium tuberculosis using mutants and microarrays. Microbiol. **148:**3129-3138.

187. **Stoop, E. J., Schipper, T., Huber, S. K., Nezhinsky, A. E., Verbeek, F. J., Gurcha, S. S., Besra, G. S., Vandenbroucke-Grauls, C. M., Bitter, W. and van der Sar, A. M.** 2011. Zebrafish embryo screen for mycobacterial genes involved in the initiation of granuloma formation reveals a newly identified ESX-1 component. Dis. Model Mech. **4:**526-536.

188. **Sullivan, B. M., Jobe, O., Lazarevic, V., Vasquez, K., Bronson, R., Glimcher, L. H. and Kramnik, I.** 2005. Increased susceptibility of mice lacking T-bet to infection with *Mycobacterium tuberculosis* correlates with increased IL-10 and decreased IFN-γ production. J. Immunol. **175:**4593-4602.

189. **Swaim, L. E., Connolly, L. E., Volkman, H. E., Humbert, O., Born, D. E. and Ramakrishnan, L.** 2006. *Mycobacterium marinum* infection of adult zebrafish causes caseating granulomatous tuberculosis and is moderated by adaptive immunity. Infect. Immun. **74:**6108-6117.

190. **Talaat, A. M., Lyons, R., Howard, S. T. and Johnston, S. A.** 2004. The temporal expression profile of *Mycobacterium tuberculosis* infection in mice. PNAS. **101:**4602-4607.

191. **Talaat, A. M., Reimschuessel, R., Wasserman, S. S. and Trucksis, M.** 1998. Goldfish, *Carassius auratus*, a novel animal model for the study of Mycobacterium marinum pathogenesis. Infect. Immun. **66:**2938-2942.

192. **Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S.** 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol. Biol. Evol. **28:**2731-2739.

193. **Tebianian, M., Hoseinib, A. Z., Ebrahimia, S. M., Memarnejadianc, A., Mokarrama, A. R., Mahdavib, M., Sohrabid, N. and Taghizadeha, M.** 2011. Cloning, expression, and immunogenicity of novel fusion protein of *Mycobacterium tuberculosis* based on ESAT-6 and truncated C-terminal fragment of HSP70. Biologicals. **39:**143-148.

194. **Tekaia, F., Gordon, S. V., Garnier, T., Brosch, R., Barrell, B. and Cole, S. T.** 1999. Analysis of the proteome of *Mycobacterium tuberculosis* in silico. Tuber. Lung Dis. **79:**329-342.

195. **Trede, N. S., Langenau, D. M., Traver, D., Look, A. T. and Zon, L. I.** 2004. The use of zebrafish to understand immunity. Immunity. **20:**367-379.

196. **Ulrichs, T., Munk, M. E., Mollenkopf, H., Behr-Perst, S., Colangeli, R., Gennaro, M. L., and Kaufmann, S. H.** 1998. Differential T cell responses to *Mycobacterium tuberculosis* ESAT6 in tuberculosis patients and healthy donors. Eur. J. Immunol. **28:**3949-3958.

197. **Uplekar, S., Heym, B., Friocourt, V., Rougemont, J. and Cole, S. T.** 2011. Comparative genomics of *esx* genes from clinical isolates *Mycobacterium tuberculosis* provides evidence for gene conversion and epitope variation. Infect. Immun. **79:**4042-4049.

198. **Volkman, H. E., Clay, H., Beery, D., Chang, J. C., Sherman, D. R. and Ramakrishnan, L.** 2004. Tuberculous granuloma formation is enhanced by a *Mycobacterium* virulence determinant. PLoS Biol. **2:**e367.

199. **Volpe, E., Cappelli, G., Grassi, M., Martino, A., Serafino, A., Colizzi, V., Sanarico, N. and Mariani, F.** 2006. Gene expression profiling of human macrophages at late time of infection with *Mycobacterium tuberculosis*. Immunol. **118:**449-460.

200. **Vordermeier, H. M., Whelan, A., Cockle, P. J., Farrant, L, Palmer, N. and Hewinson, R. G.** 2001. Use of synthetic peptides derived from the antigens ESAT-6 and CFP-10 for differential diagnosis of bovine tuberculosis in cattle. Clin. Diagn. Lab. Immunol. **8:**571-578.

201. **Wards, B. J., de Lisle, G. W. and Collins, D. M.** 2000. An Esat-6 knockout mutant of *Mycobacterium bovis* produced by homologuous recombination will contribute to the development of a live tuberculosis vaccine. Tuber. Lung. Dis. **80:**185-189.

202. **Watson, R. O., Manzanillo, P. S. and Cox, J. S.** 2012. Extracellular *M. tuberculosis* DNA tragets bacteria for autophagy by activating the host DNA-Sensing pathway. Cell. **150:**803-815.

203. **WHO.** 2012. Global Tuberculosis Report. [Accessed Online: 4 February 2013].

204. **WHO.** 2013. Immunization, Vaccines and Biologicals. [Accessed Online: 4 February 2013].

205. **Winter, S. E., Thiennimitr, P., Winter, M. G., Butler, B. P., Huseby, D. L., Crawford, R. W., Russell, J. M., Bevins, C. L., Adams, L. G., Tsolis, R. M., Roth, J. R. and Bäumler, A. J.** 2010. Gut inflammation provides a respiratory electron acceptor for *Salmonella*. Nature. **467:**426-429.

206. **Xu, J., Laine, O., Masciocchi, M., Manoranjan, J., Smith, J., Du, S. J., Edwards, N., Zhu, X., Fenselau, C., and Gao, L. Y.** 2007. A unique Mycobacterium ESX-1 protein co-secretes with CFP-10/ESAT-6 and is necessary for inhibiting phagosome maturation. Mol. Microbiol. **66:**787-800.

207. **Young, D. B.** 2003. Building a better tuberculosis vaccine. Nat. Med. **9:**503-504.