Phylo-epidemiological and pathogenic diversity of Mycobacterium tuberculosis strains in London with implications for vaccine develpoment

Velji, Preya

PHYLO-EPIDEMIOLOGICAL AND PATHOGENIC DIVERSITY OF *MYCOBACTERIUM TUBERCULOSIS* STRAINS IN LONDON WITH IMPLICATIONS FOR VACCINE DEVELOPMENT

Preya Velji

DEGREE IN DOCTORATE OF PHILOSOPHY

Acknowledgements

I would like to thank Professor Francis Drobniewski, Dr. Tim Brown, and Professor Phil Marsh for their supervision, continued guidance and support throughout the degree.

I would like to thank the team at Health Protection Agency, Mycobacterium Reference Unit for their support, in particular Dr. Vladyslav Nikolayevskyy and Dr. Marjorie Pion for their great help and advice about the molecular and tissue culture experiments.

I would like to thanks the team at Health Protection Agency, Centre for Emergency Preparedness and Response for their help and support with the guinea pig experiments.

Finally I would like to thank my family and husband for their invaluable support and understanding throughout the degree and thesis write-up period.

ABSTRACT

Approximately one-third of the global population is infected with tuberculosis causing approximately 1.7 million deaths. Currently, the BCG vaccine is used to protect against TB, but it cannot prevent primary infection or reactivation of latent infection. Ideally a vaccine should protect against a diverse array of *Mycobacterium tuberculosis* strains and promote a strong, long-lasting T_H1 cell-mediated immune response. Whilst evaluating the efficiency of novel vaccines using laboratory control strains (*M. tuberculosis* H37Rv, H37Ra and *M. bovis*-BCG), it is important to test efficacy against a representative panel of wild-type circulating strains. In England 42.2% of TB cases are reported in London and the diversity of nationalities generates a diverse pool of strains consisting of globally representative TB strains. The aim of the study was to construct a representative panel of strains for vaccine evaluation studies and general TB research.

Common *M. tuberculosis* strains were identified by performing molecular MIRU-VNTR and spoligotyping on 2363 isolates from TB cases reported in London during a one-year period. Epidemiological analysis demonstrated there were representatives from 13 global regions, including high TB burden countries. An algorithm was designed to select strains for a preliminary panel based on associations between MTBC families in clusters of more common strains, the country of birth and VNTR sub-clusters. The preliminary panel contained 42 MTBC strains belonging to 10 MTBC families from patients born in 17 countries.

Results of phylogenetic analysis of all 2363 isolates was used to select a smaller panel of strains from the preliminary panel to represent MTBC lineages to investigate if wild-type strains were phenotypically similar. The final panel included five strains from each of the Baker *et al*., 2004 *M. tuberculosis* lineages (*M. tuberculosis* Beijing, LAM10, two CAS, EAI5 strains representing lineage I, II, III, IV, respectively) and an *M. africanum* strain.

In vitro tissue culture experiments demonstrated significantly higher growth of the Beijing strain compared to the other wild-type and laboratory strains. Higher growth rates of this strain were also observed in a cell-free culture system. Aerosol challenge of guinea pigs with wild-type strains showed a quicker dissemination of the EAI5

strain from the lung to the spleen 16 days post-challenge, but significantly higher c.f.u. count of the Beijing strain in the spleen 56 days post-challenge. Collectively, the data demonstrated that there are phenotypic differences between wild-type circulating MTBC strains.

CONTENTS

List of figures

List of tables

Abbreviations

Ag85	antigen 85
AGI-30	all-glass impinger-30
ATCC	American Type Culture Collection
avidin-HRP	avidin-horseradish peroxidase
b.p.	base pairs
BCG	Bacillus Calmette–Guérin
BSC	biological safety cabinets
c.f.u.	colony forming units
CAS	Central Asian
CBA	Columbia blood agar
cDNA	complementary DNA
CEPR	Centre for Emergency Preparedness and Response
CFP-10	culture-filtrate proteins-10
CO_2	carbon dioxide
COB	country of birth
Ct	threshold cycle
DC	dendritic cells
DC-SIGN	dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
dNTPs	deoxynucleotide triphosphates
DofH	Department of Health
DPBS	Dulbecco's phosphate-buffered saline
DR	direct repeat
DVR	direct variable repeat
DVR-PCR	direct variable repeat-polymerase chain reaction
EAI	East-African Indian
ECL	enhanced chemiluminescent
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
ESAT-6	early secreting antigen target-6
ETRs	exact tandem repeats

EU	European Union
fbpA	gene encoding for fibronectin-binding protein A
FBS	foetal bovine serum
FP9	fowlpox virus
g	grams
G-C	guanine-cytosine
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HGDI	Hunter-Gaston Discrimination Index
HIV	human immunodeficiency virus
hly	gene encoding for listeriolysin
hrs	hours
HPA	Health Protection Agency
HPA,MRU	Health Protection Agency, Mycobacterium Reference Unit
ID	identification
IFN-γ	interferon-gamma
IGRA	interferon gamma release assays
IL	interleukin
inhA	gene encoding for enoyl-acyl carrier protein reductase
ino1	gene encoding for inositol-1-phosphate synthase
IR	interspersed repeats
IR	intergenic region
IS	insertion sequence
ITS	internal transcribed spacer
katG	gene encoding for catalase-peroxidase
kb	kilobases
kDa	kiloDalton
LAM	lipoarabinomannan
LAM 1-10	Latin-American and Mediterranean 1-10
log	logarithm
LPS	lipopolysaccharide
LSPs	large-sequence polymorphisms
Ltd.	Limited
ManLAM	mannosylated lipoarabinomannan
MDC	macrophage-derived chemokine
MDR	multi-drug resistant

$MgCl_2$	magnesium chloride
MGIT	Mycobacteria Growth Indicator Tube
MHC-II	major histocompatibility complex class II
mins	minutes
MIRUs	mycobacterial interspersed repetitive units
ml	millilitres
M	molar
mM	millimolar
mm	millimetres
µl	microlitres
µm	micrometre
metB	gene involved in methionine biosynthesis
MOI	multiplicity of infection
MPTRs	major polymorphic tandem repeats
MTBC	*M. tuberculosis* complex
MVA	modified vaccinia virus Ankara
NaOH	sodium hydroxide
NCTC	National Collection of Type Cultures
NH_4	ammonium
Nº	number
nsSNPs	nonsynonymous SNPs
OD	optical density
oxyR	gene encoding for a homolog of the oxidative stress response protein
oxyS	gene encoding small non-coding RNA in response to oxidative stress
panC *panD*	genes involved in pantothenate biosynthesis
PBMC	peripheral blood mononuclear cells
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
pg/ml	picogram/millilitre
PMA	phorbol 12-myristate 13-acetate
proC	gene involved in proline biosynthesis
purC	gene involved in purine biosynthesis
QUB	Queen's University Belfast
R^2	coefficient of correlation

RAPET	rapid PCR-based epidemiological typing
rBCG	recombinant BCG
RD1	region of deletion-1
rDNA	ribosomal DNA
RDs	regions of differences
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RPMI	Roswell Park Memorial Institute
rpoB	gene encoding for RNA polymerase β
rRNA	ribosomal ribonucleic acid
SDS	sodium dodecyl sulphate
secs	seconds
SNP	single-nucleotide polymorphisms
SP-A	surfactant protein A
SP-D	surfactant protein D
SpolDB4	spoligotyping database 4
sSNPs	synonymous SNPs
SSPE	saline sodium phosphate EDTA
T	Tuscany
t(½)	half-life
TB	tuberculosis
TbD1	*M. tuberculosis* specific deletion 1
TE	Tris-EDTA
THP-1	human acute monocytic leukaemia cell line
TNF-α	tumour necrosis factor-alpha
TPA	tissue plasminogen activator signal sequence
TR	tandem repeats
trpD	gene encoding tryptophan biosynthetic enzyme, anthranilate phosphoribosyltransferase
TST	tuberculin skin test
U	units
Ub	ubiquitin
UK	United Kingdom
ureC	gene encoding for urease C
USA	United States of America

v/v	volume/volume
VNTRs	variable number tandem repeats
w/v	weight/volume
WHO	World Health Organisation
%	percent/percentage
×g	relative centrifugal force
Δ	gene deletion
::	gene insertion

Publications from this thesis

Velji, P., V. Nikolayevskyy, T. Brown, F. Drobniewski (2009). "Discriminatory ability of hypervariable number tandem repeat loci in population-based analysis of *Mycobacterium tuberculosis* strains, London, United Kingdom." Emerging Infectious Diseases 15 (10): 1609-16. *(Manuscript attached at the end of thesis)*

Nikolayevskyy, V., P. Velji, T. Brown, F. Drobniewski (2010). "Exploring associations between *Mycobacterium tuberculosis* strain type and bacterial phenotype." Emerging Infectious Diseases 16 (2). *(Accepted and to be published)*

Presentations from this thesis

27th Annual Conference of the European Society of Mycobacteriology, London, UK. July 2006. Nikolayevskyy, V., T. Brown, P. Velji, F. Drobniewski. "London-wide prospective ***M. tuberculosis*** genotyping project 2005-2006: discriminatory power and epidemiological relevance of multilocus VNTR typing." *(Oral presentation)*

Health Protection 2006, Warwick, UK. September 2006. Velji, P., V. Nikolayevskyy, T. Brown, P. D. Marsh, M. D. Yates, F. Drobniewski. "Investigating the diversity of *Mycobacterium tuberculosis* strains circulating within London with implications for vaccine development." *(Oral presentation)*

18th European Congress of Clinical Microbiology and Infectious Diseases. April 2008. Nikolayevskyy, V., P. Velji, T. Brown, F. Drobniewski**. "**Biodiversity of ***M. tuberculosis*** complex strains circulating in Greater London area: implications for prospective epidemiology and phylogenesis." *(Oral presentation)*

29th Annual Conference of the European Society of Mycobacteriology, Plovdiv, Bulgaria. July 2008. Velji, P., V. Nikolayevskyy, T. Brown, F. Drobniewski. "Factors influencing fragment analysis of highly discriminatory variable number tandem repeat (VNTR) loci in *Mycobacterium tuberculosis*." *(Poster presentation)*

CHAPTER 1

INTRODUCTION

1.1 TUBERCULOSIS – THE DISEASE

1.1.1 Epidemiology

Globally, tuberculosis (TB) remains one of the major infectious causes of mortality resulting annually in 2 million deaths (WHO 2009). According to WHO reports, approximately one-third of the global population (nearly 2 billion people) is infected with the bacterium causing TB, with 9 million new cases reported annually leading to approximately 1.6 million to 1.7 million deaths (WHO 2007). Global rates of active TB are increasing at nearly 1 percent (%) per year (Dye 2006). The high TB burden areas include Asia (the South-East Asia and Western Pacific regions) and the African Region, which account for 55% and 31% of global cases (WHO 2009). The Americas, European and Eastern Mediterranean regions accounted for smaller proportions of the global cases.

In the United Kingdom (UK), the Health Protection Agency (HPA) reported 8417 TB cases in 2007, which is a large increase from the 6726 cases reported in 2000 (HPA 2008). The TB incidence varies significantly between regions, the lowest incidence being in Northern Ireland and the highest in England, which accounted for 7742 of the 8417 cases; even more astoundingly, 42.2% of the cases reported in England were from patients living in London (HPA 2007). In 2007, 3265 cases were notified in the city giving an incidence of 43.2/100,000 in the population.

A comparison of TB notifications in London showed that the highest proportions (20.6, 24.4, and 20.5%) of patients were from the North East, North West and South East areas of London, respectively (HPA 2007). Approximately 75% of people with TB in London were born abroad, mainly from high TB incidence countries within the Indian subcontinent and African regions (Anderson *et al.,* 2007). With patients from so many different countries, it has previously been demonstrated that there are a plethora of *M. tuberculosis* strains circulating in London originating within migrant patients from different countries (Dale *et al.,* 2005).

The most common causative bacterium of TB in humans is *Mycobacterium tuberculosis*, which belongs to the *M. tuberculosis* complex (MTBC) group. Other species within the group include *M. bovis*, *M. caprae*, *M. africanum*, *M. microti*, and *M. canettii*. The number of MTBC isolates reported by reference laboratories in the UK increased by over 50% between 1994 and 2003 (3264 isolates in 1994, to 4944 isolates in 2003) (HPA 2008).

1.1.2 Infection and immune response

Tuberculosis is spread via aerosols, which contain TB bacilli, when infected individuals with pulmonary TB cough or sneeze and other individuals inhale the droplets containing bacteria. In order to cause infection only a small number of bacilli need to be inhaled. An individual with untreated active TB can infect 10 to 15 other people on average per year if untreated (WHO 2007). Transmission is more likely between members of a family, in schools, prisons and hospitals due to the prolonged exposure to the same infectious individuals. The disease is associated with poverty and it has been observed that the ideal conditions for rapid spread of TB are crowded working and living conditions, as are seen in institutions such as prisons, hospitals, crowded refugee shelters and camps (Drobniewski *et al.,* 2000; Drobniewski *et al.,* 2002; Ruddy *et al.,* 2005; Huang *et al.,* 2007; Schmid *et al.,* 2008). Improved social conditions, good institutional cross infection measures and successful treatment programmes aid in the reduction of TB incidence. Global migration from high incidence countries is a significant influential factor in the rapid spread of TB worldwide as approximately 50% of TB cases in industrialised countries are found in migrants or those born abroad (HPA 2006).

Individuals who are latently infected with TB bacteria are asymptomatic; infection can be detected using tuberculin skin test (TST) or by ex-vivo interferon gamma release assays (IGRA). Approximately 10% of infected patients will go on to develop active disease, but up to half will develop active TB if severely immunocompromised such as co-infection with human immunodeficiency virus (HIV).

M. tuberculosis infection is controlled by the cellular immune system; where this fails the patient develops the classic symptoms of fever, weight loss and persistent coughing with phlegm. In some cases, the immune system is able to destroy all TB bacilli, but in the majority of cases some of the TB bacilli will lay dormant for many

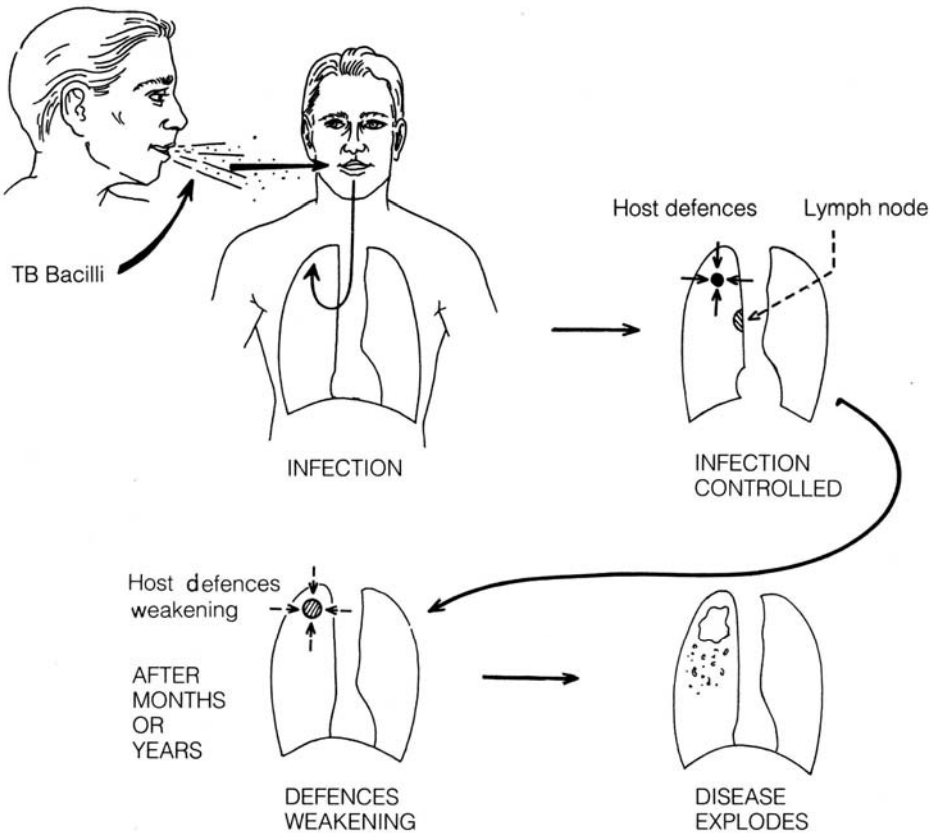
years whilst the immune system is able to keep them under control (Crofton *et al.,* 1992). Figure 1.1 illustrates the process of infection and the progression to disease.

Wallgren, 1948, defined four stages in the infection with TB, the timescales involved and the outcomes (Wallgren 1948). Three to eight weeks after bacilli have been inhaled and enter the alveoli, the mycobacteria circulate via the lymphatic system to lymph nodes in the lung. This first stage marks the point at which tuberculin reactivity occurs. The second stage involves haematogenous dissemination of bacteria to other parts of the lung and to other organs. This can last for three months and may lead to TB meningitis or miliary TB that can be acute or fatal. The next stage, which can be delayed for up to two years, lasts for approximately three to seven months involving dissemination of bacteria into the pleural space and inflammation of the pleural surfaces causing severe chest pain. In the majority of people that are infected with TB, the disease does not progress to the final stage.

The process of infection is now known to be much more complex involving components of the mycobacterial cell and the host immune system. The bacteria alone do not cause disease as the overall immunopathology is a combination of host and pathogen interactions. Evolution has not only allowed *M. tuberculosis* to survive the host immune system but to also persist for decades in many infected individuals with latent infection (Cole *et al.,* 1998; Lillebaek *et al.,* 2002; Hsu *et al.,* 2003).

Recent reviews by Smith, 2003 and Doherty and Andersen, 2005 describe current concepts of the process of infection (Smith 2003; Doherty *et al.,* 2005). Figure 1.2 illustrates the disease process described by Doherty and Andersen, 2005. The detailed events, including which cells and when they are involved, according to Smith, 2003, will be described in reference to points 1 to 4. When the TB bacilli in droplets enter the alveoli (point 1; Figure 1.2), macrophages are the first cells that the bacilli come into contact with, but it is also possible that bacilli are ingested by alveolar epithelial type II pneumocytes, which are more abundant than macrophages (Bermudez *et al.,* 1996; Mehta *et al.,* 1996). Dendritic cells (DC) are of great importance as well because they have improved antigen presenting and migratory properties ideal for T cell activation with *M. tuberculosis* antigens and for bacterial dissemination, respectively, than macrophages.

Figure 1.1 *The progression to disease after infection with TB bacilli. Taken from Crofton et al., 1992 (Crofton et al., 1992).*



However, more is understood about the role of macrophages. The primary contact of TB bacilli with the host immune system is via the alveolar macrophages and surfactants. When bacterial cells come into contact with mannose and possibly complement receptors on macrophages, phagocytosis of bacteria takes place and key surfactant proteins affect this process in different ways. Phospholipids are a major component of surfactants, but there is also 5% to 10% protein present and the primary role of surfactants is to decrease the alveolar lining surface tension (Hawgood *et al.,* 1990).

Surfactant proteins A (SP-A) and D (SP-D) are two surfactants that have been identified as being involved in host immunity (Tenner *et al.,* 1989; Persson *et al.,* 1990; Voorhout *et al.,* 1992; Haagsman 1994; Gaynor *et al.,* 1995; Ferguson *et al.,* 1999). Both SP-A and SP-D are produced and secreted by alveolar type II epithelial cells (Hawgood *et al.,* 1990; Voorhout *et al.,* 1992). The major constituent of surfactant is the nonserum protein, SP-A, which is involved in surfactant lipid homeostasis, and oligosaccharides of SP-A come together to produce a binding site

with increased valency for various molecules including bacterial cell surface molecules and receptors on the surface of alveolar type II cells and macrophages (Hawgood *et al.,* 1990; Haagsman 1994). Whilst SP-A can upregulate mannose receptor activity to enhance binding and phagocytosis of bacilli, SP-D blocks mannosyl oligosaccharide residues on the bacilli cell surface to attempt to stop any interactions with the macrophage mannose receptor, consequently inhibiting phagocytosis (Tenner *et al.,* 1989; Persson *et al.,* 1990; Gaynor *et al.,* 1995; Ferguson *et al.,* 1999).

Once intracellular pathogens such as *M. tuberculosis* enter the macrophage, they exist in an endocytic vacuole known as a phagosome. However, whilst the phagosome usually fuses with lysosomes in an endocytic pathway to create unfavourable surroundings for the pathogen (for example by creating an acidic pH or producing toxic peptides, reactive oxygen intermediates and lysosomal enzymes), pathogenic *M. tuberculosis* prevent fusion of phagosomes with lysosomes that prevent the development of endosomes (Armstrong *et al.,* 1975; Frehel *et al.,* 1986).

Normally after macrophages are infected and activated, proteases in the vesicles, which contain the bacteria, degrade proteins belonging to pathogens to produce fragments of peptides, which are transported to the cell surface via major histocompatibility complex class II (MHC-II) proteins so the peptides are presented to $CD4^+$ T cells, which in their naïve state can differentiate into two types of T cells, T_H1 and T_H2 cells, affecting the type of adaptive immune response elicited against the bacterial infection (Janeway *et al.,* 2005). However, in the case of *M. tuberculosis* infection of macrophages, to support the studies which have demonstrated that the development of endosomes is inhibited, decreased MHC-II protein expression and decreased antigen presentation has been observed, meaning $CD4^+$ T cells would not recognise infected macrophages (Noss *et al.,* 2001).

In the lungs, infected macrophages produce chemokines to attract cells such as neutrophils, lymphocytes and monocytes in their inactive forms (point 2; Figure 1.2), and in this form these cells are not sufficiently effective for eradication of bacteria (Fenton *et al.,* 1996; van Crevel *et al.,* 2002). Macrophage-derived giant cells and lymphocytes are components of granulomatous focal lesions that start to form, possibly aiding in the containment of the mycobacterial cells. A caseous centre of the

granuloma forms as cellular immunity develops and infected macrophages are killed. Around the caseous centre is a zone containing cells that were attracted to the infected macrophages (Fenton *et al.,* 1996). There is decreased oxygen, an acidic pH and toxic fatty acids in the caseous tissue so bacteria are probably unable to multiply, but some bacilli may survive and remain dormant. At this point the infection is latent but can persist for the rest of the host's life, with the host not displaying any symptoms or transmitting to other individuals. If the individual has efficient host immunity, infection may be permanently detained (point 3; Figure 1.2), leading to the eventual evolution of granulomas as small fibrous/calcified lesions (Fenton *et al.,* 1996).

On the other end of the scale, if an infected individual is not able to control the initial lung infection or the immune system of latently infected individuals becomes weaker (point 4; Figure 1.2) for various reasons, including HIV co-infection, treatment with immunosuppressive drugs, aging or malnutrition, the centre of the granuloma may liquefy providing revived mycobacterial bacilli with the rich medium required for replication (Converse *et al.,* 1996). Active pulmonary or extrapulmonary TB can develop as bacteria escape the granuloma and spread to other regions of the lung or other tissues through the blood and lymphatic system, respectively.

The exact mechanisms by which *M. tuberculosis* evade the host immune system are still not completely understood. Previous research has elucidated some of the key immunological defence mechanisms involved. Figure 1.3 shows some of the possible ways in which *M. tuberculosis* influences the immune system favouring its survival (Doherty *et al.,* 2005). Macrophages serve as a habitat for TB bacilli over a long period of time and in order to make the environment within the cell more favourable, the bacteria influences gene expression within the macrophage. There is upregulation of genes encoding molecules that are essential for chemotaxis to attract other cells to the infection site, for example, interleukin (IL)-8, which aids in the recruitment of neutrophils, and IL-1β that plays a role in granulomatous inflammation (Volpe *et al.,* 2006). In their study, Volpe *et al.,* 2006, also noted upregulation of macrophage-derived chemokine (MDC), another chemoattractant that is important for many reasons including its influence on T helper type 2 cell activation, and functions of DCs and natural killer cells.

Figure 1.2 *A diagram showing the events that take place within the lungs after an individual is infected with TB bacilli. The events occurring at each of the numbered points in the diagram are referred to in Section 1.1.2. Diagram adapted from Doherty and Andersen, 2005 (Doherty et al., 2005).*

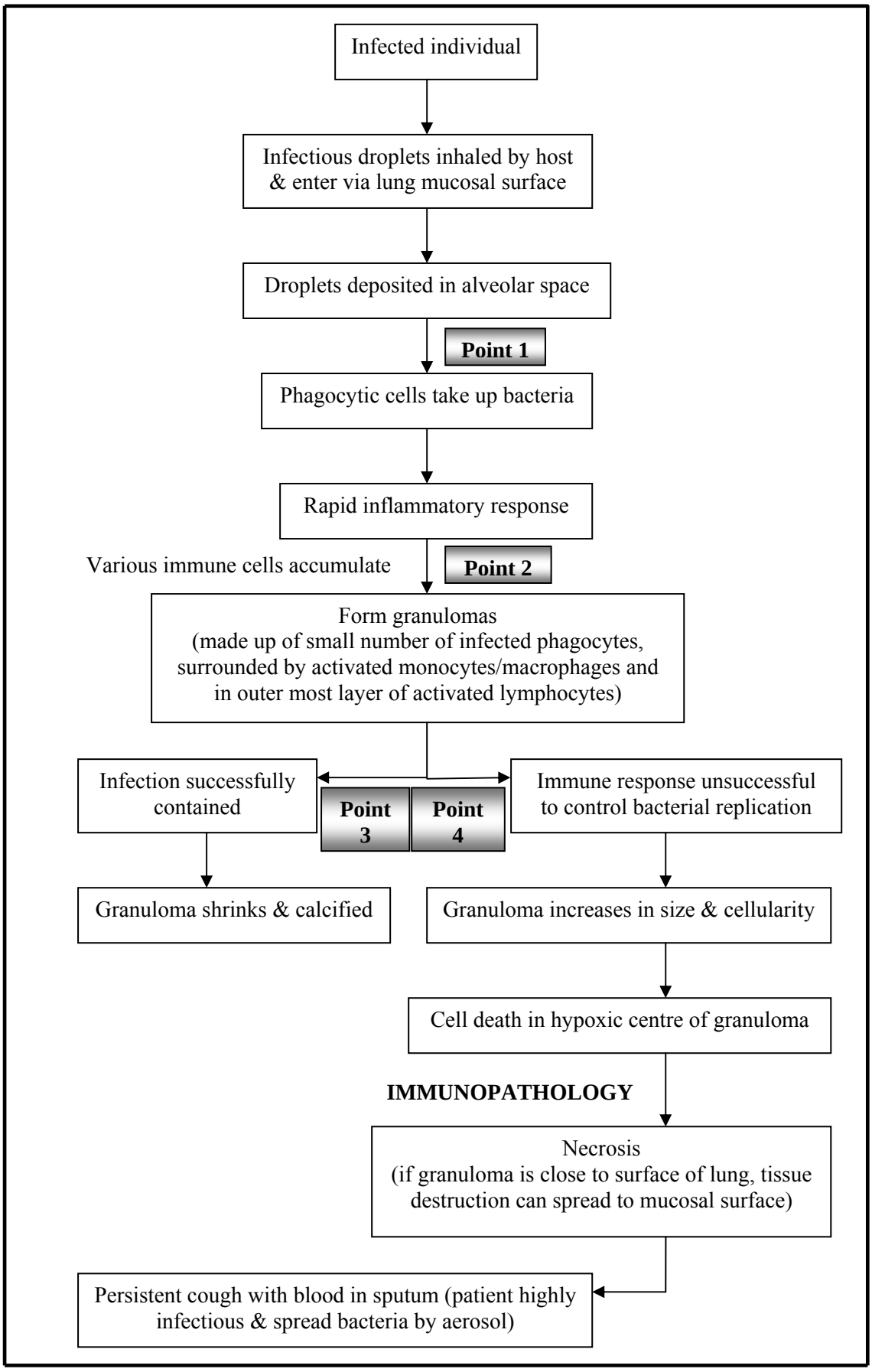
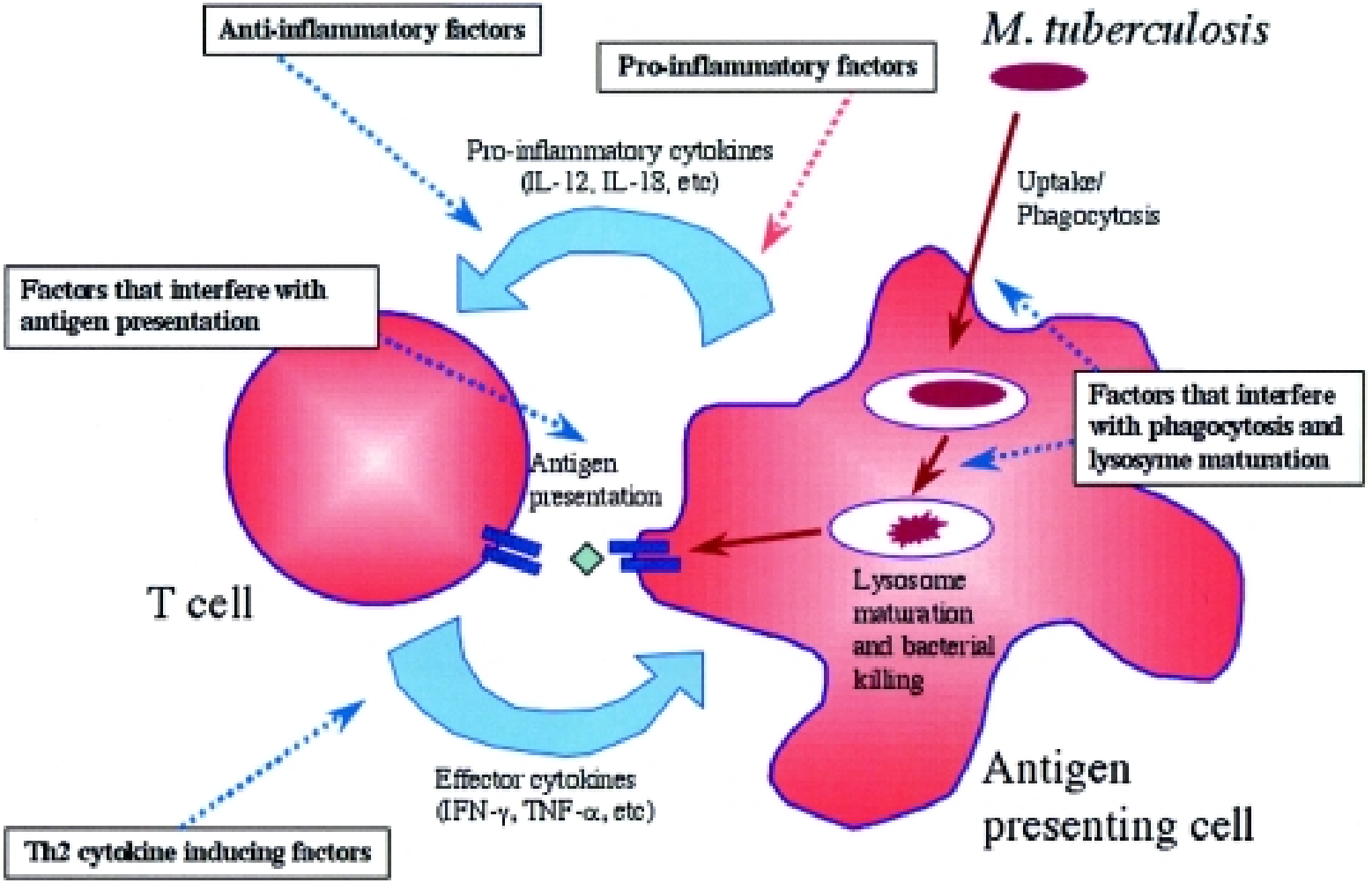


Figure 1.3 *A diagram showing how the immune system may kill M. tuberculosis via phagocytosis, and features that may interfere with the response. Taken from Doherty and Andersen, 2005 (Doherty et al., 2005).*



Whilst chemotaxis helps in the recruitment of cells to the site of infection, which would eventually lead to the formation of granulomas, a good inflammatory response is also essential. This was demonstrated by Sugawara *et al.,* 2002, who showed that the inflammatory host response was strongly stimulated by long chain fatty acids expressed on the mycobacterium cell wall, such as mycolic acid (Sugawara *et al.,* 2002). These long chain fatty acids have multiple roles as they can encourage cytokine production, including those that regulate production of interferon-gamma (IFN)-γ, and can regulate apoptosis or survival of infected macrophages (Ryll *et al.,* 2001; Nuzzo *et al.,* 2002).

Increased production of IFN-γ is insufficient for eradicating mycobacterial cells as the downstream effects of IFN-γ are inhibited, by the bacteria or components on the bacterial cell wall, which modulate the effects of cytokine so it is advantageous for the bacteria (Ting *et al.,* 1999). It seems that the necessary immune response to induce IFN-γ production is inhibited (Lienhardt *et al.,* 2002). In the study by Lienhardt *et al.,* 2002, IFN-γ, which is produced by T_H1 cells, and IL-4, produced by T_H2 cells, were measured *in vivo*; there was a low Th1 and high Th2 response in TB patients demonstrating a Th1/ Th2 response imbalance that allows progression of TB to the disease state with a consequential poor clinical outcome. A Th1 immune response is vital for protection against TB infection and studies have shown that the cytokine IL-12, which is secreted by infected DCs and macrophages, promote a Th1 response which leads to the secretion of other pro-inflammatory cytokines (Flynn *et al.,* 1995). There is an increased production of pro-inflammatory cytokines, tumour necrosis factor-alpha (TNF-α) and IL-6, by macrophages after infection with virulent *M. tuberculosis* and these cytokines work to reduce mycobacterial burden (Ragno *et al.,* 2001; Volpe *et al.,* 2006; Bhatt *et al.,* 2007). The role of these two cytokines has been further discussed in Chapter 6 of the present study as cytokine production was investigated after infecting a tissue culture cell line with different MTBC strains.

Other studies have shown that although a strong Th1 response may be elicited by *M. tuberculosis* this does not always lead to protection against the bacteria (Hovav *et al.,* 2003). Species within the MTBC specifically have a 27 kiloDalton (27-kDa) mycobacterial antigen and immune response to this lipoprotein induces secretion of IFN-γ, inducing a strong Th1 response. That there is no evident protective response seems to indicate that this lipoprotein is in fact a decoy that can be used to ultimately

protect the pathogen. However there are lipoproteins that, whilst stimulating an immune response to protect the host, in fact have adverse effects. It has been suggested that the MTBC species specific 19-kDa antigen is expressed particularly to act as a competitive inhibitor; therefore, it would have a higher affinity for antigen presenting molecules than other antigens that would truly elicit an effective immune response (Yeremeev *et al.,* 2000). Neyrolles *et al.,* 2001 also found that trafficking of the 19-kDa lipoprotein within infected macrophages was completely isolated from live mycobacterial cells (Neyrolles *et al.,* 2001).

Mycobacterial cells need to be able to survive within the macrophage and obtain nutrients required for its survival. One mechanism by which mycobacterial cells survive is by disrupting maturation of the phagosome in which it resides, but fusion with other vesicles like lysosomes, membrane remodelling and trafficking can still take place (Russell *et al.,* 1996). Therefore, *M. tuberculosis* bacteria can acquire essential nutrients, whilst exporting their own proteins. As well as disrupting maturation of phagosomes, DCs are specifically targeted by *M. tuberculosis* as interactions between dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin (DC-SIGN) and mycobacterial cell wall component, mannosylated lipoarabinomannan (ManLAM), means that the bacteria are internalised (Geijtenbeek *et al.,* 2003). *M. tuberculosis* negatively influences DC maturation and positively influences production of IL-10, which is an anti-inflammatory cytokine that promotes immunosuppression, ultimately aiding in bacterial cell survival.

Research so far has provided great insight into the process of infection by *Mycobacterium* species and its clinical progression. There is also increased understanding of how the mycobacteria are able to survive within the host and overcome the most challenging issues of the host immune system, even during its latent phase. It is highly possible that there are many other mechanisms and cytokines waiting to be discovered by which this pathogen is able to not only survive but to infect other individuals contributing to the potential virulence of MTBC strains. There are additional factors that play a role in the virulence of mycobacteria.

1.1.3 Mycobacterial virulence factors

As mentioned in Section 1.1.2, mycobacteria can direct an immune response to increase chances of survival (Russell *et al.,* 1996; Geijtenbeek *et al.,* 2003). The

degree to which particular strains are able to direct the immune response to self-protect, therefore reflecting the virulence of strains, can potentially be measured by quantifying the amount of cytokines, such as IFN-γ, TNF-α, IL-2, -12 and -6, or monitoring the growth of mycobacteria in *in vitro* models such as monocytic THP-1 cell line, therefore providing potential markers for virulence of mycobacterial strains (Fremond *et al.,* 2004; Theus *et al.,* 2004; Theus *et al.,* 2005; Park *et al.,* 2006; Lee *et al.,* 2007; Sow *et al.,* 2007). Identifying any differences in the virulence of *M. tuberculosis* strains, by monitoring the growth of strains in a THP-1 cell line model and in a cell-free culture system, and also measuring the production of TNF-α, IL-10, -1β and -6 have been demonstrated in the present study and is discussed in greater detail in Chapter 6.

In vivo models have provided a good base for comparing virulence of mycobacterial strains and a common *in vivo* marker for virulence is extrapulmonary dissemination of TB bacilli as there are histological similarities between primary pulmonary lesions observed in guinea pigs and humans (Bhatia *et al.,* 1961; Prabhakar *et al.,* 1987). Many studies have used guinea pigs as models for comparing virulence of laboratory and wild-type, clinical strains and more importantly for evaluating the performance of novel vaccines after aerosol challenge of guinea pigs with TB culture (Bhatia *et al.,* 1961; Prabhakar *et al.,* 1987; Williams *et al.,* 2000; Williams *et al.,* 2005; Williams *et al.,* 2005; Vipond *et al.,* 2006). One particular study by Williams *et al*., 2005 demonstrated that the number of mycobacterial bacilli in the spleen at the early time point of 16 days post-challenge of guinea pigs with the TB culture provided a good marker of how virulent the strain is because an increased mycobacterial load in the spleen would indicate the faster dissemination of bacilli from the lungs (Williams *et al.,* 2005). Virulence in this study was compared using the percentage weight gain data, lung and spleen mycobacterial load 16 and 56 days post-challenge and by observing the histopathology of lung and spleen tissue. The same protocol was adopted for the present study and is discussed in greater detail in Chapter 8.

Other factors of virulence that have also been successfully identified, but not included in the present study, include differences in single-nucleotide polymorphisms (SNPs) between virulent *M. bovis* strains and all Bacillus Calmette–Guérin (BCG) strains (Garcia Pelayo *et al.,* 2009). Garcia Pelayo *et al*., 2009 identified 115 nonsynonymous

SNPs (nsSNPs) which play a role in various functions affecting the virulence of strains, including central metabolism and transcriptional factors.

M. tuberculosis bind selectively to fibronectin and this binding can be blocked by antibodies against members of the antigen 85 (Ag85) complex of fibronectin-binding proteins, which include Ag85A, Ag85B, and Ag85C, that are secreted and stored in the cell wall of *M. tuberculosis* (Abou-Zeid *et al.,* 1988; Ratliff *et al.,* 1988; Ratliff *et al.,* 1993). Patti *et al*., 1994 demonstrated that the degree to which pathogens bind to fibronectin enhances the virulence of the pathogens as the specific binding could aid in the adherence and dissemination of the mycobacteria from the lungs to other organs and tissues (Patti *et al.,* 1994). For this reason, proteins from the Ag85 complex family have been identified as potential virulence factors and mutating genes, *fbpA*, and *fbpB*, which express the proteins Ag85A and Ag85B, respectively, has helped to identify that *fbpA* has a function towards the pathogenesis of *M. tuberculosis* H37Rv as loss of expression led to decreased growth in human or mouse macrophage-like cell lines (Armitige *et al.,* 2000).

Further investigation of the mycobacterial cell wall chemistry has led to the identification of cell wall components that pose as factors of virulence. Microbial glycolipids, such as phenolic glycolipid-I from *M. leprae*, have a role in the virulence and pathogenicity by searching for toxic oxygen radicals (Neill *et al.,* 1988; Chan *et al.,* 1989). Lipoarabinomannan (LAM) is a complex cell wall-associated glycolipid produced abundantly by *M. tuberculosis* and *M. leprae* and is made up of mannose and arabinose saccharide units linked to a phosphatidylinositol moiety, which helps to attach LAM to the mycobacterial cell cytoplasmic membrane (Hunter *et al.,* 1986). The LAM component is a suggested mycobacterial virulence factor as LAM provides protection against the antimicrobial activities of mononuclear phagocytes (Chan *et al.,* 1989).

1.1.4 Current immunisation to TB and future novel vaccines

The current vaccine against TB is the BCG vaccine, which has existed for over 80 years, and which comes from the original *M. bovis* isolate at the Institut Pasteur in Lille after passaging numerous times from 1909 to 1921 to produce an attenuated vaccine strain (WHO 2004). Although reconstituted vaccines contain dead and viable *M. bovis* bacilli leading to variations between administered doses, the BCG vaccine is

currently the most widely used TB vaccine and in countries that are included in the national childhood immunization programme, the vaccine is administered to over 80% of neonates and children.

In the UK the BCG immunisation programme commenced in 1953 with the aim of protecting young people that just started working; the initial target of the vaccine was 14 year old children that were just about to leave school (DofH 2006). By the 1960s, TB was commonly identified in immigrants from the countries of high TB prevalence so the BCG vaccine was administered to children of the new immigrants, even if they were born in the UK. The universal school immunisation programme continued until 2005, when it was decided to suspend it as the TB rates in the British population had declined. Currently, the vaccination programme focuses on infants between the ages of 0 to 12 months that are born in a high incidence area or that have parents and grandparents that were born in a high incidence country (DofH 2006).

Whilst the BCG vaccine provides protection in children against disseminated TB and meningitis, it has limited impact on TB transmission as the vaccine is not able to prevent primary infection with mycobacterial bacilli and more importantly, it cannot stop reactivation of latent pulmonary infection, which is the main source of transmission in communities (Rodrigues *et al.,* 1993; WHO 2004). *In vivo* experiments in guinea pigs have demonstrated that the BCG vaccination may be more protective against *M. bovis* infection than infection with *M. tuberculosis* H37Rv (Williams *et al.,* 2000).

The global efficacy of BCG has varied from 0% to 80% in different studies with higher efficacy in industrialised countries compared to non-industrialised countries closer to the equator (Fine 1995; Fine *et al.,* 1998; Brewer 2000). One reason proposed for the variation seen in vaccine efficiency was the difference in the exposure of individuals to environmental mycobacteria in the different countries because the bacilli in the environment might stimulate the immune system and thereby prevent an appropriate response to the BCG vaccine (Fine 1995; Fine *et al.,* 1998). There have also been variations in the effectiveness of the BCG vaccine amongst schoolchildren in the UK with protective levels ranging from 0% to between 70 and 80% (Sutherland *et al.,* 1987; Rodrigues *et al.,* 1991).

The BCG vaccine provides protection for approximately 10 to 15 years after which protection may wane, although there is limited data on this (Sterne *et al.,* 1998). As part of the BCG vaccination programme, it is not recommended to administer the vaccine to individuals over 16 years of age unless the person belongs to a high risk group, which includes laboratory and healthcare staff, or for the purposes of traveling abroad (DofH 2006). There is no evidence that repeat BCG vaccination increases or prolongs protection.

Whilst the use of BCG vaccine is still recommended, there are advancements in research leading to a greater understanding of the complex interactions involved in the immunological response of the host and the bacteria, and subsequently increased research into novel vaccines against TB. Since the sequencing of *M. tuberculosis* H37Rv genome and its annotation by Cole *et al.,* 1998, there has been a rapid expansion in the knowledge of the genetics involved in the survival and evolution of mycobacteria (Cole *et al.,* 1998). Ideally an effective vaccine should be able to protect individuals against all of the circulating *M. tuberculosis* strains and promote a strong and long-lasting T_H1 cell-mediated immune response (refer to Section 1.1.2).

Currently, there are major projects in operation to develop more effective TB vaccines. European collaborative projects, like the European Union (EU) TB Vaccine Cluster project, aim to develop new TB vaccines that are more effective than BCG. Such projects involve a step-wise evaluation of these novel vaccines starting by using mice, followed by evaluation in guinea pigs, and finally lead candidates may be tested in non-human primates, and/or humans. After observing the protection against virulent challenges and immunogenicity following the vaccination of mice with the novel vaccines, vaccine candidates that hold potential have been selected for evaluation in guinea pigs (Williams *et al.,* 2005). Some of the vaccine candidates include recombinant protein (designed to over-express antigenic targets of *M. tuberculosis* that would elicit a protective immune response against TB infection), improved BCG strains expressing additional antigens, or the development of attenuated strains of *M. tuberculosis* and *M. microti* strains (Collins *et al.,* 2001; Williams *et al.,* 2005). Prime-boost vaccine strategies were also evaluated and involved immunising guinea pigs with BCG followed by a booster, which may be a sub-unit antigen expressed as a protein or delivered by a viral vector such as modified vaccinia virus Ankara (MVA) (McShane *et al.,* 2004; McShane *et al.,* 2005).

The prime-boost strategy, DNA prime MVA/Ag85A boost, which involved firstly immunising guinea pigs with DNA expressing Ag85A followed by MVA, provided significantly better protection in guinea pigs infected with *M. tuberculosis* H37Rv than in guinea pigs belonging to the saline control group (Williams *et al.,* 2005). In this study, protection against TB infection had been determined in terms of prolonged survival of guinea pigs infected with TB after administering the DNA prime MVA/Ag85A boost, by the reduced number of mycobacterial bacilli in the spleen demonstrating slower dissemination of bacilli, and in terms of percentage lung consolidation demonstrating reduced lung pathology upon infection after being vaccinated using the prime-boost strategy.

Other studies have investigated the potential of new vaccines, including attenuated, live vaccines, vaccines designed to increase the efficacy of the present BCG vaccine, subunit vaccines designed to supplement or boost the BCG vaccine, and deoxyribonucleic acid (DNA) vaccines (Hubbard *et al.,* 1992; Olsen *et al.,* 2003; Pym *et al.,* 2003; Gupta *et al.,* 2007). A summary of the different vaccines and their strategies have been summarised in Table 1.1, some of which have already been described previously in this section.

Whilst live, attenuated vaccines offer great potential for protection against TB, it is a challenge to obtain a suitable balance between the level of attenuation and maintenance of immunogenicity (Gupta *et al.,* 2007). Attenuated *M. tuberculosis* mutants present species-specific antigens, meaning these strains could potentially promote a more improved protection against TB infection than BCG (Sambandamurthy *et al.,* 2002; Sambandamurthy *et al.,* 2005). Protection against mycobacterial infection has been elicited by *M. vaccae* in mice as there are common antigens between *M. tuberculosis* and *M. vaccae* and a live, attenuated form of *M. microti* administered orally to mice provided better protection against infection with virulent *M. tuberculosis* than BCG (Hernandez-Pando *et al.,* 1997; Manabe *et al.,* 2002).

Table 1.1 *A table summarising the novel vaccine candidates that have been discussed in Section 1.1.4 and additional vaccines that have been studied. Vaccines have been categorised as live, attenuated vaccines, recombinant BCG vaccines, subunit vaccines or DNA vaccines. Also included is a reference to the source and the study of the vaccines. The antigens described in the table are targets of M. tuberculosis that would elicit a protective immune response against TB infection.*

Type of vaccine	Name of vaccine (brief description)		Reference
Live, attenuated	*M. microti* ATCC 19422		(Manabe *et al.*, 2002)
	Heat inactivated *M. vaccae*		(Hernandez-Pando *et al.*, 1997)
	MVA expressing Ag85A	(used as a booster after BCG vaccination)	(McShane *et al.*, 2004)
	MVA & fowlpox virus, FP9, both expressing Ag85A	(used as a booster after BCG vaccination)	(Williams *et al.*, 2005)
	M. tuberculosis Δ*RD1*		(Hsu *et al.*, 2003)
	M. tuberculosis Δ*purC*	purine biosynthesis gene	(Jackson *et al.*, 1999)
	M. tuberculosis Δ*metB*	methionine biosynthesis gene	(Smith *et al.*, 2001)
	M. tuberculosis Δ*proC*	proline biosynthesis gene	
	M. tuberculosis Δ*trpD*	tryptophan biosynthesis gene	
	M. tuberculosis Δ*panCpanD*	pantothenate biosynthesis gene	(Sambandamurthy *et al.*, 2002)
Recombinant BCG	BCG :: RD1-2F9	expresses ESAT-6/CFP-10	(Pym *et al.*, 2003)
	rBCG30 expressing Ag85A		(Horwitz *et al.*, 2000)
	Δ*ureC hly*$^+$ rBCG	(urease deficient mutant expressing listeriolysin O)	(Grode *et al.*, 2005)
	rBCG38 Tice	(expresses the 38 kDa antigen of Ag85 complex)	(Castanon-Arreola *et al.*, 2005)
	rBCG 19T	(expresses the 19 kDa antigen of Ag85 complex)	(Rao *et al.*, 2003)
	rBCG 38T	(expresses the 38 kDa antigen of Ag85 complex)	
	rBCG E6T	(expresses ESAT-6)	
Subunit	pJI23 (DNA-64)	(gene expressing MPT64 antigen)	(Kamath *et al.*, 1999)
	pJI30 (DNA-85B)	(gene expressing Ag85B)	
	pJIE6 (DNA-E6)	(gene expressing ESAT-6)	
	Ag85B/ESAT-6 fusion protein		(Olsen *et al.*, 2004; Langermans *et al.*, 2005)
	Ag85B/TB 10.4 fusion protein		(Dietrich *et al.*, 2005)
	single recombinant polyprotein Mtb72F	(polyprotein made up of Mtb 32 and Mtb 39 proteins)	(Skeiky *et al.*, 2004)
DNA	Mtb72F DNA	delivered as naked DNA	(Skeiky *et al.*, 2004)
	plasmid DNA Mtb 8.4	(expresses an immunoreactive T cell antigen)	(Coler *et al.*, 2001)
	MTB41-DNA	(antigen eliciting a Th1 response)	(Skeiky *et al.*, 2000)
	Ag85A DNA		(Tanghe *et al.*, 2000)
	pE6/85	(expressing ESAT-6/Ag85B)	(Derrick *et al.*, 2004)
	MTB Ag *ESat-6*	all genes express *M. tuberculosis* antigens and were fused with tissue plasminogen activator (TPA) signal sequence & ubiquitin (Ub)	(Delogu *et al.*, 2002)
	MTB Ag *mpt64*		
	MTB Ag *MPT63*		
	MTB Ag *mpt8e*		
	MTB Ag *Ag85b*		
	MTB Ag *katG*		
	MTB Ag *mtb12*		
	MTB Ag *mtb8.4*		
	MTB Ag *MTB39*		
	MTB Ag *1818C*		

Attempts to improve the efficacy of the BCG vaccine have been made by adding *M. tuberculosis* genes, which encode proteins known to mediate protection when used as subunit vaccines (Dietrich *et al.,* 2006). The recombinant BCG vaccine, rBCG30, expresses and secretes the protein Ag85B and has improved protection against TB in guinea pig models (Horwitz *et al.,* 2000). Increased protection was also observed in animal models administered with a recombinant vaccine expressing region of deletion-1 (RD1), but the disadvantage was that the vaccine strain became more virulent (Pym *et al.,* 2003). The locus RD1 was selected because this locus contains the genes for Early Secreting Antigen Target-6 (ESAT-6) and Culture-filtrate proteins-10 (CFP-10), which are both immune targets.

The subunit vaccines are based on the availability of proteins that are secreted by mycobacterial bacilli, for recognition of infected macrophages at an early stage in the immune response, thereby controlling growth and multiplication of mycobacterial cells (Dietrich *et al.,* 2006). Similar targets to those used for preparing the recombinant BCG vaccines have also been adopted for some of the subunit vaccines being investigated. A significantly higher degree of protection is provided by ESAT-6 as there is strong recognition of this protein by T cells and also by Ag85A and Ag85B, which are also secreted by *M. tuberculosis* and are crucial targets for a protective human T-cell response against the pathogen (Boesen *et al.,* 1995; Kamath *et al.,* 1999). In addition to subunit vaccines with single proteins, fusion protein subunit vaccines have also been evaluated, including a vaccine containing ESAT-6 and Ag85B (Olsen *et al.,* 2004; Dietrich *et al.,* 2005; Langermans *et al.,* 2005; Dietrich *et al.,* 2006). Administration of this vaccine promoted a strong and highly protective immune response against TB.

Promotion of $CD4^+$ and $CD8^+$ T cell responses against TB can be encouraged by using DNA vaccines and protection against TB infection has been observed after delivery of mycobacterial antigens, for instance Ag85 and ESAT-6, to mice in the form of naked DNA (Kamath *et al.,* 1999; Britton *et al.,* 2003; Olsen *et al.,* 2003). Administration of chimeric DNA or co-delivery of multiple DNA plasmids has increased the effectiveness of DNA vaccines and proved to reduce the numbers of mycobacterial cells in the lungs of aerosol-challenged mice (Tanghe *et al.,* 2000; Delogu *et al.,* 2002; Derrick *et al.,* 2004).

Whilst designing vaccines that positively influence the immune response, there have been attempts to design vaccines that would physically stop the maturation of phagosomes containing the mycobacterial bacilli. An example is a urease-deficient BCG mutant that would express the listeriolysin O gene from *Listeria monocytogenes* (a bacterium that cannot prevent phagosome maturation because of urease-deficiency) (Campos *et al.,* 1996; Grode *et al.,* 2005). There are also attempts to try and attenuate *M. tuberculosis* instead of genetically modified designs of BCG as more human TB infections are caused by the former species.

The evaluation strategy for novel TB vaccine candidates is crucial. As well as testing the performance of vaccines using standard laboratory control strains like *M. tuberculosis* H37Rv and H37Ra (the avirulent laboratory strain) and *M. bovis* BCG, it is also important to test their efficiency against a representative panel of wild-type *M. tuberculosis* strains that are circulating in the population, as these are the strains that a vaccine should ideally protect populations against.

1.2 MOLECULAR TYPING OF *M. TUBERCULOSIS* STRAINS

1.2.1 The genome – conservation & variability

The genome size of *M. tuberculosis* H37Rv is approximately 4.4×10^6 base pairs (b.p.) and is characterised by its very high guanine and cytosine content of 65.6% (Cole *et al.,* 1998). In order to use information about the genome effectively it is important to identify and understand regions of conservation and variability. There is evolutionary conservation in the genome of MTBC and the species within MTBC are closely related genetically due to a high degree of DNA homology between strains (Imaeda 1985).

Frothingham *et al.,* 1994, focused on the 16S-to-23S ribosomal DNA (rDNA) internal transcribed spacer (ITS), which has a high rate of nucleotide substitution. However, their studies demonstrated that strains from the different MTBC species and *M. bovis* BCG had the same ITS sequence showing that the 16S rDNA sequence was conserved, which was corroborated by the identification of conserved domains in the 16S ribosomal ribonucleic acid (rRNA) molecule (Kirschner *et al.,* 1993; Frothingham *et al.,* 1994). Whilst this region remains conserved within the MTBC, there are a minimum of seven possible sites within the ITS that are specific for MTBC

(Frothingham *et al.,* 1994). Another major factor that contributes to variability in the genome of most human pathogens is the degree of horizontal gene transfer, which is absent in modern *M. tuberculosis* strains (Alland *et al.,* 2003). Conversely, it has been reported that the *M. tuberculosis* genome appears to be a composite assembly due to horizontal gene transfer that occurred prior to an evolutionary bottleneck approximately 35,000 years ago, after which there was a clonal expansion of *M. tuberculosis* strains (Supply *et al.,* 2003; Gutierrez *et al.,* 2005).

Compared to other bacterial pathogens affecting humans, there is a significantly lower rate of mutations, more specifically silent nucleotide substitutions, in *M. tuberculosis*, suggesting that evolutionary change in this organism occurs slowly (Sreevatsan *et al.,* 1997). Although the *M. tuberculosis* genome is conserved, it does not mean that regions of variability or polymorphism do not exist. The polymorphic regions of most interest are those that include repeated units of identical sequence. There are interspersed repeats (IR), including direct and insertion sequence-like repeats, and tandem repeats (TR), including direct repeats that are uninterrupted head-to-tail (Mathema *et al.,* 2006). Reorganisation of the genome is more commonly noted with unstable transposable elements leading to deletions, inversions, duplications and transpositions. Insertion sequences (IS) are transposable elements and although it has been identified as being a rare event, transposition of IS elements has been observed and these elements are a major contributing factor in producing polymorphisms in the genome, which would potentially aid in strain change and differentiation (van Soolingen *et al.,* 1991; Cole *et al.,* 1998).

Hermans *et al.,* 1991, identified that a particular IS, IS*987*, was positioned within a direct variable repeat (DVR) containing locus and that the sequence of this locus only seemed to be found exclusively in MTBC species (Hermans *et al.,* 1991). This was when the true potential of the chromosomal direct repeat (DR) locus was recognised as it displayed polymorphism with regard to its length and composition. Each DVR is composed of a DR and a spacer region, which is a short non-repetitive sequence. The sizes of these are 36 b.p. and 35 to 41 b.p., respectively. Due to particular events involving rearrangement of the genes, some spacers may be deleted (van Embden *et al.,* 2000). Two such events are homologous recombination between distant or adjacent chromosomal DRs, leading to variation in a single or a few DVRs, and transposition of an IS, IS*6110*, which is almost always present in the DR locus.

Supply *et al.,* 1997, reported a new group of repeated sequences, within the intercistronic region of the senX3-regX3 mycobacterial two-component system encoding operon, called mycobacterial interspersed repetitive units (MIRUs) (Supply *et al.,* 1997). Following on from this identification, Magdalena *et al.,* 1998, amplified the *senX3-regX3* intergenic region (IR) on various MTBC strains and, based on the size of the polymerase chain reaction (PCR) fragment, found that the number of MIRUs varied between strains (Magdalena *et al.,* 1998).

Prior to molecular typing techniques, phage typing was the most commonly used method for differentiating between strains in the MTBC (Bates *et al.,* 1967). However, whilst the technique was useful for typing strains involved in laboratory cross-contamination and from outbreaks, the technique was not sufficiently discriminative as phage types were limited and the procedure was difficult to perform.

The variations within the *M. tuberculosis* genome, which have been described briefly, in combination with the rapid advancements in molecular techniques, provided the foundations for identifying genetic markers that show sufficient variation and polymorphism to differentiate between strains belonging to MTBC and broaden our understanding of the evolutionary and epidemiological patterns of species within the MTBC.

Three molecular techniques, including one requiring prolonged cultivation, that are currently widely in use are discussed in detail along with applications of the techniques and the advantages and disadvantages of their use with regard to various applications. Molecular amplification-based methods seem to be more attractive than those requiring prolonged TB cultivation as minimal amounts of DNA are required.

1.2.2 IS*6110* restriction fragment length polymorphism

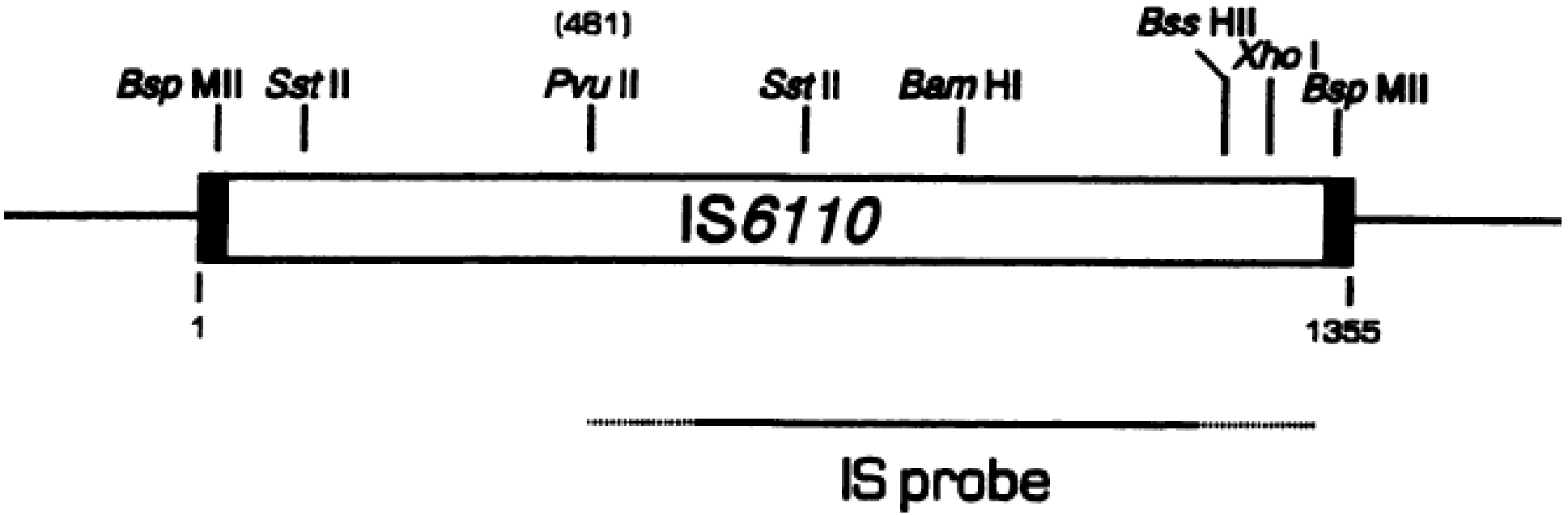
The IS element, IS*6110*, which has a sequence of 1355 b.p., is the most commonly used element for analysis, usually in a restriction fragment length polymorphism (RFLP) format. This element was first recognised by Thierry *et al.,* 1990, when it was also concluded that this element was specific to species belonging to the MTBC (Thierry *et al.,* 1990). Thierry *et al.,* 1990 noticed that there were similarities between IS*6110* and other elements belonging to the enterobacterial IS3 family, and this was

supported by McAdam *et al.,* 1990, who identified IS*986* as being part of the IS3 family and noted the high degree of similarity between the sequences of IS*986* and IS*6110*, therefore concluding that IS*6110* belonged to the IS3 family (McAdam *et al.,* 1990).

Strains from MTBC have varying numbers of IS*6110* copies, ranging from 0 to 25 copies, that are incorporated into the chromosome at different sites creating a combined polymorphism of copy number and position which makes IS*6110* a useful element for fingerprinting (van Soolingen *et al.,* 1991; van Soolingen *et al.,* 1993). The possibility of random integration of the element has been investigated using the IS*6110*-RFLP technique described later in this section, and it was reported that the insertions were not completely random i.e. there were preferential insertion sites (McHugh *et al.,* 1998). There have been many studies demonstrating that one of the preferential sites of integration is the DR region and a study by Legrand *et al.,* 2001 showed the insertion of an IS*6110* copy in between spacers 31 and 32 in the DR region (Legrand *et al.,* 2001). The DR region is not the only hotspot as other sites such as the genomic region between Rv1754c and Rv1762c and the *dnaA-dnaN* IR have also been recognised as preferential sites of integration (Kurepina *et al.,* 1998; Sampson *et al.,* 1999).

The IS*6110*-RFLP method has been described in great detail by van Embden *et al.,* 1993 (van Embden *et al.,* 1993). In brief, the method involves a single cleavage of each IS*6110* element present on the genome using the restriction enzyme *PvuII* that cleaves at the 481 b.p. position of the 1355 b.p. element as shown in Figure 1.4. Agarose gel electrophoresis separates the resulting fragments with sizes ranging from 0.9 to 10 kilobases (kb). Using an IS*6110* probe allows visualisation of fragments as the probe hybridises to the right of the *PvuII* cleavage site, as indicated in Figure 1.4. This gives a fingerprint of the strain that can be compared to reference fingerprints with known band sizes. Currently, IS*6110*-RFLP remains the gold standard technique for typing MTBC strains, which is why it has been discussed in great detail even though this technique was not used in the present study.

Figure 1.4 *A map of the 1355 b.p. IS6110 region with the cleavage site for PvuII restriction enzyme at the 481 b.p. position and the position at which the IS probe hybridises. Taken from Embden et al., 1993 (van Embden et al., 1993).*



1.2.3 Spoligotyping

As mentioned in Section 1.2.1, species from the MTBC contain a specific chromosomal region featuring multiple 36 b.p. DRs interspersed by 35 to 41 b.p. spacers, which are unique. The possibility of using the DR region as a tool for typing was first pursued by Groenen *et al.,* 1993, using direct variable repeat-polymerase chain reaction (DVR-PCR) (Groenen *et al.,* 1993). Their study showed that MTBC strains with no epidemiological link could be differentiated effectively using a single PCR procedure. Whilst DVR-PCR allowed detection of *M. tuberculosis* as well as typing of strains, another simpler procedure, which was designed for a similar purpose, was introduced. Kamerbeek *et al.,* 1997, developed a technique based on the hybridization patterns, which were thought to be strain-dependent, of amplified PCR product with various spacer oligonucleotides and, for this reason, the method was called spacer oligotyping or, more commonly, spoligotyping (Kamerbeek *et al.,* 1997).

DNA sequence comparison has ultimately led to the discovery of 94 different spacer sequences with sequences for each displaying no homology with DNA sequences external to the DR region (van Embden *et al.,* 2000). However, only 43 spacers, which were identified originally, are used for spoligotyping as the addition of extra spacers does not markedly improve discrimination (van Embden *et al.,* 2000; Sebban *et al.,* 2002). The polymorphic nature of the DR region is probably due to homologous recombination or transposition of the IS*6110* element (van Embden *et al.,* 2000). Both events would lead to deletion of spacers, which provides the basis for spoligotyping.

Spoligotyping as described by Kamerbeek *et al.,* 1997 involves the use of membranes that have 43 lines of covalently bound oligonucleotides, to which amplified product from the DR region will hybridise depending on which spacers are present and absent in the region (Kamerbeek *et al.,* 1997). The hybridisation patterns for each strain are visualised using chemiluminescence producing a profile with black rectangles or blank spaces representing presence and absence of spacers, respectively. According to their studies, Kamerbeek *et al.,* 1997, found that spoligotyping patterns were unique for the majority of the strains used in their study and further to this, observations showed that strains from particular outbreaks had the same pattern showing the potential of spoligotyping.

1.2.4 Variable Number Tandem Repeat, Exact Tandem Repeat & Mycobacterial Interspersed Repetitive Units

Variable number tandem repeats (VNTRs) are minisatellites with 10 to 100 b.p. repeats and are found in great number in the genome of most bacteria (Mathema *et al.,* 2006). VNTRs are found in regulatory and intergenic regions and within open reading frames and VNTRs in the human genome are particularly useful tools for forensic analysis, paternity testing and genetic mapping and identification. The polymorphic nature of this region meant that it would be useful for typing *M. tuberculosis* strains.

Initially, Hermans *et al.,* 1992, identified multiple copies of DNA sequences in the *M. tuberculosis* genome with a 10 b.p. sequence that was repeated tandemly and separated by distinctive spacers, 5 b.p. in length (Hermans *et al.,* 1992). As there was slight variation in the sequence these repeats were called major polymorphic tandem repeats (MPTRs). The same study identified that the polymorphism would be especially useful for epidemiological studies as the location of MPTRs was very variable. The presence of MPTRs was confirmed by Frothingham and Meeker-O'Connell, 1998, who reported the existence of 11 loci with tandem repeats in the genome of *M. tuberculosis* (Frothingham *et al.,* 1998). These loci included 5 MPTRs and 6 exact tandem repeats (ETRs). The difference being that MPTRs contained repeats that were 15 b.p. in length but still had significant sequence variation between copies next to each other, whilst ETRs contained longer repeats and the sequences between adjacent repeats were identical. All ETR loci displayed length polymorphisms that were representative of deletions and insertions of tandem repeats, but this degree of polymorphism was not observed with MPTR loci.

As mentioned in Section 1.2.1, MIRUs were found to exist in the region of the senX3-regX3 mycobacterial two-component system encoding operon and in the study by Supply *et al.,* 1997, it was observed that the sizes of MIRU tandem repeats varied (Supply *et al.,* 1997; Magdalena *et al.,* 1998). This is the main difference between the VNTRs and MIRUs (Mostrom *et al.,* 2002). The length of MIRUs range from 40 to 100 b.p. and 41 MIRUs have been identified as shown on the *M. tuberculosis* H37Rv genome in Figure 1.5 (Supply *et al.,* 2000). Even though there are many MIRUS, only 12 showed variations not only in the number of copies of tandem repeats but also in

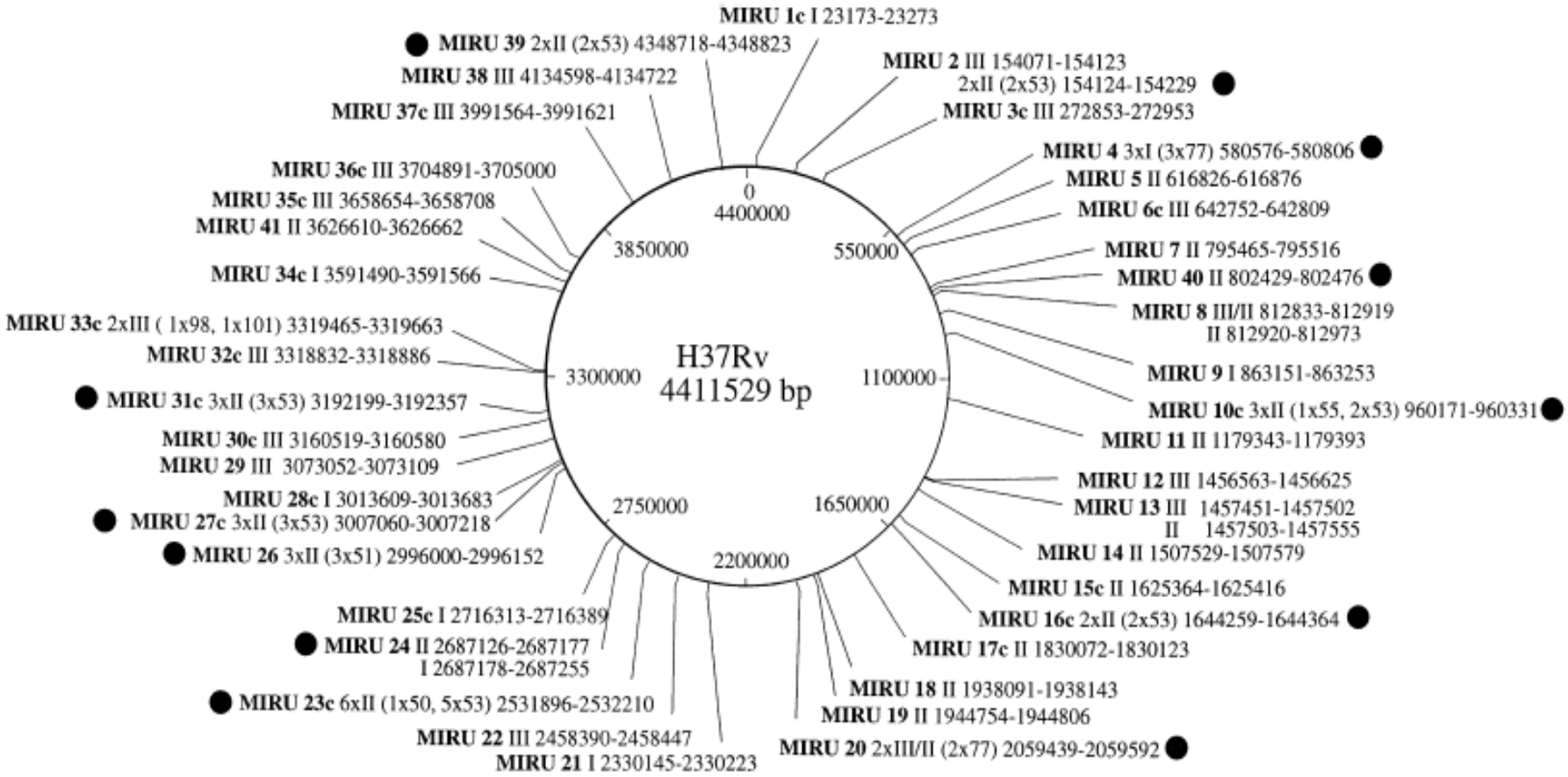
the sequences between MIRUs. From the chosen 12 MIRUs, MIRU-4 and -31 were the same as ETR-D and -E, correspondingly, which were identified by Frothingham and Meeker-O'Connell, 1998.

Further research into VNTRs yielded the identification of 7 new variable VNTR loci designated Queen's University Belfast (QUB) 5, 11a, 11b, 15, 18, 23 and 26 (Skuce *et al.,* 2002). As novel loci were discovered, in an attempt to simplify the situation a uniform nomenclature system was proposed by Smittipat *et al.,* 2005 (Smittipat *et al.,* 2005). Table 1.2 shows the nomenclature system as defined by Smittipat *et al.,* 2005, with the alternative names that are given to some of the VNTR loci by different researchers. Although not all loci have been included in the table, the selected loci are relevant to the present study. MTPRs have not been included in the original table by Smittipat *et al.,* 2005, presumably because the original study by Frothingham and Meeker-O'Connell, 1998, showed that there was not a significant amount of polymorphism in the loci suitable for analysis (Frothingham *et al.,* 1998; Smittipat *et al.,* 2005). Table 1.3 shows the three sets of loci that were used along with the nomenclature that had been adopted for the present study.

The use of various MIRU and VNTR loci means that the technique is widely known as MIRU-VNTR typing. Originally, MIRU-VNTR typing involved the determination of number of repeats by using primers specific to each locus to amplify the region in separate PCR reactions, after which agarose gel electrophoresis was performed to simultaneously confirm presence of PCR products and to determine the molecular weight. The number of repeats was calculated using previous knowledge of the molecular weight of expected products for various numbers of repeats in specific loci.

The technique can be performed using an automated system, which involves the use of fluorescently labelled primers for amplification of loci making performance of multiplex PCR reactions possible and allowing large-scale genotyping (Supply *et al.,* 2001). Using the panel for 12 MIRU-VNTR (see Table 1.3) as an example, the number of repeats in each locus would be represented in a 12-digit format. The automated system involves the use of capillary electrophoresis (or comparable systems) instead of conventional agarose gel electrophoresis and for clustering studies capillary electrophoresis has proved to be more useful (Yokoyama *et al.,* 2006).

Figure 1.5 *The M. tuberculosis H37Rv genome showing the positions of the 41 MIRU loci, where the black dots represent the 12 variable MIRUs. Taken from Supply et al., 2000 (Supply et al., 2000).*



1.2.5 Stability of genetic elements & discrimination power of IS*6110*-RFLP, spoligotyping and MIRU-VNTR typing

The stability of genetic markers and their discriminatory power are two of the main factors influencing the application of IS*6110*-RFLP, spoligotyping and MIRU-VNTR for typing.

Several studies have focused on the stability of IS*6110* element with conflicting results. IS*6110* elements are stable over a short time scale of a few months but when looking over longer periods of time, a significant rate of transposition is observed (Tanaka *et al.,* 2000). Strains with a larger number of IS*6110* copies are more polymorphic than strains with low copy number (Tanaka *et al.,* 2000; Tanaka *et al.,* 2001; Maguire *et al.,* 2002).

In a detailed study carried out in the Netherlands by de Boer, *et al.,* 1999, IS*6110*-RFLP fingerprint profiles were different between initial and follow-up isolates; changes were more frequent in individuals with extrapulmonary TB than in people with both pulmonary and extrapulmonary TB (de Boer *et al.,* 1999). From their study they calculated the half-life (t(½)) of IS*6110* element as being 3.2 years. In a study, which focused on serial isolates (with a known time interval between isolates) from patients in San Francisco, the t(½) was 2 years (Yeh *et al.,* 1998; de Boer *et al.,* 1999). A suggested explanation for the difference between the results of the IS*6110* element t(½) in the Netherlands and San Francisco studies was the lack of bias in the serial isolates included in the Netherlands study as there was no prior knowledge of the time interval between serial isolates.

A study by Warren *et al.,* 2002, showed that the IS*6110* t(½) was 8.74 years (Warren *et al.,* 2002). Yeh *et al.,* 1998, suggested that there was a rapid rate of transposition as IS*6110*-RFLP patterns changed in a 3 year time gap (Yeh *et al.,* 1998). A possible explanation for the differences in the calculated t(½) between all of the studies is that rates were considerably affected by the period between TB onset and the point where sputum was sampled, and the potential for re-infection in a high incidence country indicated the complexity of estimating the true degree of stability.

Table 1.2 *The nomenclature for VNTR loci, along with the alternative names assigned by different researchers. (Produced using Table 2 in Smittipat et al., 2005 (Smittipat et al., 2005).*

Tandem repeat locus	Alternative nomenclature
ETR-A	loci not included in the table by Smittipat *et al.,* 2005 but included in this table for the purposes of the present study
VNTR-2163A	
VNTR-2163B	
VNTR-1982	
MIRU-40	
VNTR-2059	MIRU-20
VNTR-0154	MIRU-2
VNTR-0577	ETR-C
VNTR-0580	ETR-D, MIRU-4
VNTR-0960	MIRU-10
VNTR-1644	MIRU-16
VNTR-2347	
VNTR-2461	ETR-B
VNTR-2531	MIRU-23
VNTR-2687	MIRU-24
VNTR-2996	MIRU-26
VNTR-3007	MIRU-27, QUB-5
VNTR-3155	QUB-15
VNTR-3192	ETR-E, MIRU-31
VNTR-3232	QUB-3232
VNTR-3239	ETR-F
VNTR-3336	QUB-3336
VNTR-4052	QUB-26
VNTR-4348	MIRU-39

Table 1.3 *A table showing the three sets of loci with their nomenclature as used in the present study.*

MIRU-VNTR loci	12 MIRU	15 MIRU-ETR	Extra 7 loci
MIRU-2	✔	✔	
MIRU-4	✔	✔	
MIRU-10	✔	✔	
MIRU-16	✔	✔	
MIRU-20	✔	✔	
MIRU-23	✔	✔	
MIRU-24	✔	✔	
MIRU-26	✔	✔	
MIRU-27	✔	✔	
MIRU-31	✔	✔	
MIRU-39	✔	✔	
MIRU-40	✔	✔	
ETR-A		✔	
ETR-B		✔	
ETR-C		✔	
VNTR-2163A			✔
VNTR-2163B			✔
VNTR-1982			✔
VNTR-2347			✔
VNTR-3232			✔
VNTR-3336			✔
VNTR-4052			✔

Niemann *et al.,* 1999, observed that whilst there was variation in IS*6110*-RFLP patterns between initial and follow-up isolates, there were no changes in the spoligotyping patterns indicating that the rate of change is much lower in the DR region making this genetic marker more stable than IS*6110* (Niemann *et al.,* 1999). A similar scenario was reported by Goguet de la Salmoniere *et al.,* 1997, when the same spoligotyping pattern was obtained for all strains of *M. bovis* BCG used in their study (Goguet de la Salmoniere *et al.,* 1997). Although rare, coexistence of strains with two different spoligotyping patterns in a patient has been observed: the second isolate differed from the initial isolate due to a deletion that had occurred during a single episode of TB within the patient (Kamerbeek *et al.,* 1997; van Embden *et al.,* 2000). However, except for this one deletion, spoligotypes were shown to be stable. The existence of an international spoligotyping database, SpolDB4, means that comparison studies using this typing technique are possible (Brudey *et al.,* 2006).

As is the case with spoligotyping, MIRU-VNTR is reliable for follow-up of patients, in particular in patients that have chronic infection over a long period of time (Savine *et al.,* 2002). The *in vivo* stability of the 12 MIRU-VNTR loci has been observed and the estimated length of stability has been reported to be at least 18 months, whilst MIRUs are reasonably stable for 30 years in axenic culture environments where only one organism is present (Supply *et al.,* 2000; Mazars *et al.,* 2001). In a study involving typing of BCG strains that had been cultivated for more than 30 years, there was polymorphism in locus MIRU-4, whilst the remaining 11 loci were monomorphic indicating that there is slow evolution of these minisatellite-like structures (at least in BCG) (Supply *et al.,* 2000). The polymorphisms due to repeat units being added and deleted are believed to lead to long term stability of loci.

The MIRU-VNTR loci have a variable range of alleles; for instance, in MIRU-2 and -24 mostly 1 or 2 copies are observed, therefore, these loci are not very polymorphic in comparison with VNTR-3232, -3820 and -4052 (Smittipat *et al.,* 2005). More than 9 allelic variants have been observed in these loci and VNTR-3820 had the highest variability with numbers of repeats varying from 3 to 32. There are external factors that influence the variability of particular MIRU-VNTR loci; for instance, MIRU-20 appears to be more polymorphic in some studies and not in others (Smittipat *et al.,* 2005). Influential factors may include the source of samples geographically and possibly the inherent genetic diversity. Therefore when selecting the panel of loci,

factors such as the predominance of certain clones in specific geographical locations and the variability within different genetic groups of MTBC strains needs to be taken into account (Mathema *et al.,* 2006).

However, selecting different VNTRs for panels with different applications may not be ideal as this diminishes advantages of using a universal MIRU-VNTR panel for interlaboratory analysis and adopting a technique that is cost effective and manageable with regard to labour. An assessment of the reproducibility and interlaboratory variation of IS*6110*-RFLP patterns showed that patterns with smaller numbers of bands were reproducible but not for strains with higher numbers of IS*6110* copies and complex patterns (Braden *et al.,* 2002). In contrast, a comparative study showed high reproducibility using MIRU-VNTR demonstrating its potential for interlaboratory analysis (Kremer *et al.,* 2005).

There have been numerous studies to compare the results of different molecular typing methods and there are clear conclusions on the discriminative abilities of techniques (Goguet de la Salmoniere *et al.,* 1997; Kremer *et al.,* 1999; Barlow *et al.,* 2001; Lee *et al.,* 2002; Hawkey *et al.,* 2003; Gopaul *et al.,* 2006). *M. tuberculosis* isolates with a low IS*6110* copy number of 1 to 4 have been excluded because of the lower degree of discrimination and where there is a low copy number of IS*6110*, the degree of differentiation offered by spoligotyping is higher than it would be if there was a high IS*6110* copy number (Goguet de la Salmoniere *et al.,* 1997; Maguire *et al.,* 2002). Using both of these methods together gives greater discrimination for genotyping (Kremer *et al.,* 1999).

The loci included in the 12 MIRU-VNTR panel, described by Supply *et al.,* 2000, have varying discriminatory powers with the most discriminative loci being MIRU-10, -23, -26, -31, -40 and ETR-A (when compared to ETR-B and -C), whilst MIRU-4, -16, -24 and -39 were found to be moderately discriminative, and MIRU-2, -20 and -27 were of poor discrimination (Sola *et al.,* 2003). The discriminative power of MIRU-VNTR is proportional to the number of loci in a panel and is similar to that of IS*6110*-RFLP when there is a high copy number of IS*6110*, but more discriminatory when the IS*6110* copy number is low (Barlow *et al.,* 2001; Lee *et al.,* 2002; Hawkey *et al.,* 2003; Gopaul *et al.,* 2006; Mathema *et al.,* 2006). In a study by Mazars *et al.,* 2001, the use of 12 MIRU-VNTR yielded 2 to 8 MIRU-VNTR alleles, meaning there

were over 16 million different combinations showing that the resolution of MIRU-VNTR was close to that of IS*6110*-RFLP (Mazars *et al.,* 2001). However it has been reported that 12 MIRU-VNTR is less discriminatory than IS*6110*-RFLP typing of high copy number isolates. Combining 12 MIRU-VNTR and spoligotyping seems to provide more discrimination than IS*6110*-RFLP on its own (Roring *et al.,* 2002). Some studies have shown that whilst patterns for 12 MIRU-VNTR are more distinct than those obtained by IS*6110*-RFLP and spoligotyping, specificity is maximised when multiple typing methods are used (Cowan *et al.,* 2002).

1.2.6 Deletion mapping

Differences in the genomes of *M. tuberculosis* strains H37Rv and CDC1551 have been observed and are believed to be caused by SNPs, (i.e. differences in single base pairs), and large-sequence polymorphisms (LSPs) that show differences in sequences greater than 10 b.p. (Fleischmann *et al.,* 2002). In addition to variation due to SNPs, LSPs have been identified as the source of most insertion-deletion events (Brosch *et al.,* 2001). Various regions of variability due to insertion and deletion have been identified. These deletions occur in aggregations rather than randomly along the genome and the irreversible nature of deletions forms the basis of deletion mapping, which can be utilised for identifying strains within the MTBC (Brosch *et al.,* 2002; Tsolaki *et al.,* 2004).

Such studies have already been performed and yielded very promising results. Brosch *et al.,* 2002 investigated 20 variable regions in MTBC strains and results showed that these polymorphisms were not independent events but due to genetic occurrences that were irreversible in common ancestral strains (Brosch *et al.,* 2002). The main region of interest from their study was the *M. tuberculosis* specific deletion 1 (TbD1). The presence or absence of this region was indicative of ancestral and modern strains, respectively. Also identified in this study were regions of differences (RDs), one of which includes RD9 and if a deletion is recognised in this region it determines that a strain is neither ancestral nor modern *M. tuberculosis*, but one of the other species in the MTBC, namely *M. africanum*, *M. microti*, and *M. bovis*. The presence or absence of other RD regions allows further sub-speciation within the MTBC as indicated in Figure 1.6 (Brosch *et al.,* 2002).

MTBC strain typing using the identified regions of possible deletion is very simple and straightforward to perform as it involves amplification of regions of interest using

specific primers and running the resulting products on an agarose gel. This reveals bands representing larger and lower molecular weights corresponding to the presence and absence of the region, respectively. Although deletion mapping was not included in the present study it is important to acknowledge its importance in MTBC phylogenetic analysis, which will be discussed in Section 1.3.2.

1.3 APPLICATIONS OF MOLECULAR TYPING METHODS

1.3.1 Epidemiological studies

There are a large range of applications for IS*6110*-RFLP, spoligotyping and MIRU-VNTR typing and now deletion mapping. These techniques can give a deeper insight into the epidemiology of *M. tuberculosis* (using IS*6110*-RFLP, spoligotyping and MIRU-VNTR typing) and its phylogeny (using IS*6110*-RFLP, spoligotyping and MIRU-VNTR typing and deletion mapping). The varying discriminatory power of the typing techniques makes certain methods more favourable than others for specific applications. With regard to MIRU-VNTR, there are applications that are especially useful for individual laboratories and will be discussed.

Due to its high discriminatory power, IS*6110*-RFLP has been the most widely used approach for large scale national and international molecular epidemiology studies and outbreak investigations. IS*6110*-RFLP was favoured as the international standard procedure that was recognised by all laboratories especially for TB epidemiological studies and remains the gold standard technique due to the combination of a very high degree of polymorphism and stability of the IS*6110* element (van Soolingen *et al.,* 1991; van Soolingen *et al.,* 1993; Li *et al.,* 2005; Kik *et al.,* 2008; Devaux *et al.,* 2009). For large scale epidemiological studies, whilst interlaboratory comparison of IS*6110*-RFLP patterns is possible, quality control is essential and has to be carefully monitored.

IS*6110*-RFLP has been used to understand epidemiological transmission of TB all over the world. One particular study of interest is that performed by Maguire *et al.,* 2002, in London to investigate recent transmission events of *M. tuberculosis* (Maguire *et al.,* 2002). Isolates were typed using IS*6110*-RFLP and patient epidemiological data collated. This study demonstrated that TB occurrence in London, in comparison with the other cities, was not mainly due to recent transmission. The majority of TB in

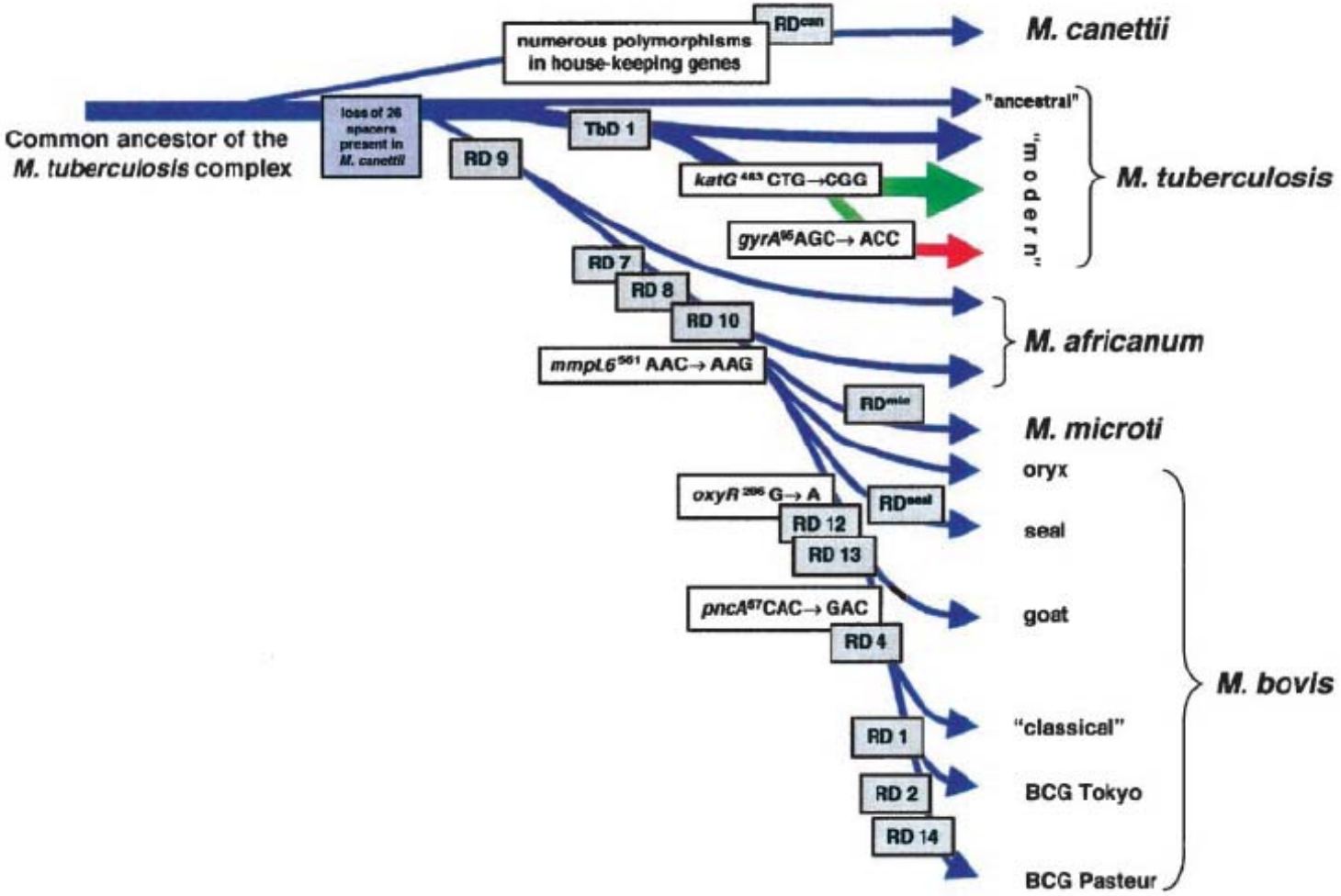
London was due to either infection being imported by immigrants that moved to London recently or by reactivation of previous TB infections (Maguire *et al.,* 2002).

However, there are limitations to the uses of IS*6110*-RFLP as the fewer bands there are within strains the harder it is to interpret, compare patterns and discriminate between strains even when using computerised data image analysis. IS*6110*-RFLP is labour-intensive especially when typing large numbers of isolates, making spoligotyping and MIRU-VNTR more attractive, and only using IS*6110*-RFLP to confirm the identity of the organism (Kremer *et al.,* 2002). A study investigating the potential of IS*6110*-RFLP and spoligotyping for epidemiological studies showed that the discriminatory power of spoligotyping was not sufficient to make it a method that could be used on its own and that IS*6110*-RFLP was more discriminatory for outbreak studies (Wilson *et al.,* 1998).

The potential of the loci included in MIRU-VNTR as genetic markers for epidemiological studies has been recognised and evaluated (Mazars *et al.,* 2001). MIRU-VNTR profiles of isolates that are epidemiologically linked have been shown to cluster proving that stability of this genetic marker is sufficient for identifying outbreaks. The added advantages of using MIRU-VNTR are the high throughput of isolates that can be typed and the ease with which numerical profiles can be compared. In some epidemiology studies, a two-step approach has been used with IS*6110*-RFLP as the primary technique followed by MIRU-VNTR (Kwara *et al.,* 2003). It has even been suggested that using spoligotyping in conjunction with MIRU-VNTR would be appropriate for studying epidemiology of TB (Sola *et al.,* 2001).

When used collectively, molecular typing techniques have proved useful for identifying possible sources of outbreaks, which were previously not known, in specific social scenarios (Yaganehdoost *et al.,* 1999; Drobniewski *et al.,* 2003). In another case study, strain typing was used to recognize and characterise strains, which had originated in hospitals, and their transmission. In this particular study molecular markers were identified to verify an outbreak of multi-drug resistance (MDR) TB and clarifying the history of the order in which drug resistance was acquired as the progress of the outbreak could be monitored by molecular typing (Bifani *et al.,* 1996).

Figure 1.6 *A diagram showing the regions that would be deleted (in the boxes) from the genome to identify strains within the MTBC. Taken from Brosch et al., 2002 (Brosch et al., 2002).*



Epidemiology studies using molecular typing has helped to identify the source of infection in order to take the most appropriate measures to control TB infection. In New York, for instance, the continuous spread of the *M. tuberculosis* Beijing strain led to an escalation in incidence of multiple drug resistance strains of *M. tuberculosis* and as soon as this was recognised, appropriate measures were enforced to curb the spread of the strain (Bifani *et al.,* 1996; Moss *et al.,* 1997). Molecular typing is also valuable for detected sources of TB infection on a smaller scale, for instance, in hospitals as described by Allix *et al.,* 2004 (Allix *et al.,* 2004). This study described a scenario in which patients had visited the same hospital during a 1 year period and had TB infection in which isolates had identical IS*6110*-RFLP patterns and confirmation was obtained using MIRU-VNTR typing. Their visit to the hospital was traced back and it was discovered that they had been to the hospital pneumological surgery department for bronchoscopic examination on the same day. The TB infection was bronchoscopy-related.

1.3.2 Phylogenetical studies

Whilst molecular typing techniques are useful for studying the epidemiology of TB transmission, there is also significant potential for clarifying phylogenetic relationships of clinical isolates and numerous studies have been carried out in an attempt to decipher the phylogenetic structure and relationships within MTBC. Before continuing, a brief description of SNP has to be included as major breakthroughs in phylogenetic studies have resulted from the use of SNP analysis.

M. tuberculosis genome comparisons have yielded polymorphisms at various nucleotides, which is a source of more genetic markers to distinguish between *M. tuberculosis* strains. There exists nsSNPs and synonymous SNPs (sSNPs), the main difference being that nsSNPs lead to changes in amino acids, whilst sSNPs do not i.e. they are neutral. It is the sSNPs that form the basis of studies into evolutionary relationships, as sequences of specific regions would help to identify these neutral polymorphisms between strains.

Using IS*6110*-RFLP to assign *M. tuberculosis* isolates into lineages has shown that there are only two specific lineages determined by high and low IS*6110* copy number (Mazars *et al.,* 2001). Using sSNP has proved otherwise. In the study by Gutacker *et al*., 2002, the phylogenetic relationship of *M. tuberculosis* isolates from various global

sources was investigated using 230 sSNP resulting in eight clear-cut lineages (Gutacker *et al.,* 2002). Sreevatsan *et al.,* 1997, assigned these eight lineages, I to VIII, into three principle genetic groups (Sreevatsan *et al.,* 1997). The largest genetic group contained isolates from lineage III to VI, whilst the smallest group was *M. bovis* in a lineage on its own.

Delving deeper into the matter of how important IS*6110* copy number is for phylogeny studies, Gutacker *et al.,* 2002 has shown that the association between IS*6110* copy number and phylogenetic lineages is more complex and that the copy number alone is not useful for phylogeny studies (Gutacker *et al.,* 2002). This was confirmed by Gutacker *et al.,* 2006, which specifically looked at variations in IS*6110*-RFLP, spoligotyping and MIRU-VNTR amongst lineages defined by sSNP (Gutacker *et al.,* 2006).

In their study, Gutacker *et al.,* 2006, again found a similar phylogenetic trend except this time nine genetic lineages were identified, the extra one being II.A (Gutacker *et al.,* 2006). When looking for associations between the nine lineages and the number of IS*6110* copies, IS*6110* elements that had been inserted in the *mmpS1* gene could be genetically associated with lineage IV, showing that these particular strains were related by descent. Isolates with less than 6 copies could be associated with different lineages, whilst isolates with just 1 copy were distributed amongst lineages I, II.A, and IV. This again contradicts the definition of two lineages using the high copy number and low copy number categorisation.

Comparison of spoligotypes with phylogenetic lineages shed some light on the genetic relationship between principle genetic groups (Gutacker *et al.,* 2006). As a whole, population principal genetic group 2/3 isolates were genetically related but could be distinguished from principal genetic group 1 organisms as there was no overlap of spoligotypes between group 1 and group 2/3. A similar trend was observed when MIRU-VNTR profiles for isolates were compared with the phylogenetic lineages. This indicates that spoligotyping and MIRU-VNTR typing hold great potential for phylogenetic studies. When looking at the data using all three typing methods and sSNP, there was a strong indication that strains within the 9 lineages were clonally related by descent, as they had a common ancestor (Gutacker *et al.,* 2006). The influence of various genetic markers in phylogenetic studies has been confirmed by

Warren *et al.,* 2004, who suggested that clonal expansion occurred via sequential acquisition of additional IS*6110* copies, deletions of specific DR sequences and VNTR sequence expansion and contraction (Warren *et al.,* 2004).

Baker *et al.,* 2004, used various typing methods to investigate the phylogeny of MTBC including sSNP, TbD1 deletion mapping, IS*6110*-RFLP and spoligotyping. A phylogenetic tree, shown in Figure 1.7a, was produced from data obtained using sSNPs and the maximum parsimony method and the different branches of the tree represented unique sSNP combinations (Baker *et al.,* 2004). Isolates were defined into four distinct *M. tuberculosis* lineages, I to IV, and a closely related *M. bovis* lineage.

Correlations and discrepancies between this study and that performed by Gutacker *et al.,* 2006 have been identified and reviewed by Mathema *et al.,* 2006 (Mathema *et al.,* 2006). Lineage I in Baker *et al.,* 2004, corresponds to lineage II in Gutacker *et al.,* 2006, lineage II is a combination of lineage III, IV, V and VI, lineage III is associated with II.A and lineage IV to lineage I. In both studies, *M. bovis* is a separate lineage. A study by Gagneux *et al*., 2006, showed agreement with lineages I, II, III and IV defined by Baker *et al*., 2004 (Gagneux *et al.,* 2006). In addition to the four lineages, two lineages for *M. africanum* were identified by Gagneux *et al*., 2006 as there had been few *M. africanum* strains in the study by Baker *et al*., 2004.

Baker *et al.,* 2004 also looked at the association of the absence and presence of TbD1 with particular lineages. TbD1 was present in lineage IV and *M. bovis*, *M. microti* and *M. africanum* whilst it was absent in the remaining lineages (Figure 1.7b) (Baker *et al.,* 2004). IS*6110*-RFLP and spoligotyping allowed the assignation of various strain families into lineages (Figure 1.7c). The various strains of *M. tuberculosis* agreed with the parsimony tree originally defined using sSNP. The introduction of the spoligotyping database, SpolDB4, will contribute greatly towards understanding the phylogeny of MTBC as it will allow better identification of isolates to then investigate the dissemination of a strain and its history (Brudey *et al.,* 2006).

The possibility of using 12 MIRU and 3 ETR as a rapid tool for phylogenetic classification of *M. tuberculosis* and *M. bovis* using the five lineages observed by Baker *et al.,* 2004 was investigated by Gibson *et al.,* 2005 (Gibson *et al.,* 2005). This study generated specific panels of MIRU-VNTR loci that were specific for the lineages suggesting that MIRU-VNTR could be used for classification of *M.*

tuberculosis and *M. bovis* and investigating phylogeny and evolution, whilst providing information on sublineages and molecular clocks.

Results from a study that combined MIRU-VNTR and spoligotyping methods confirmed a scenario that was proposed by Brosch *et al.,* 2002, from results obtained from deletion analysis using TbD1 and various RD regions (Sola *et al.,* 2003). This scenario involved a common MTBC ancestor from which the first divergence would have led to the species *M. canettii* followed by the separation of *M. tuberculosis* East-African Indian (EAI) strains (Brosch *et al.,* 2002; Sola *et al.,* 2003). *M. africanum* and *M. bovis* would have been next to be separated proceeded by separation that would distinguish *M. tuberculosis* Beijing, which is distinctive from the Central Asian 1 (CAS1) family, and non-Beijing strains. Latin-American and Mediterranean (LAM), X and Haarlem strains are thought to have deviated later. A similar scenario was proposed by Ferdinand *et al.,* 2004, using spoligotyping and MIRU-VNTR, more specifically MIRU-24, which the study reported was the most appropriate for classification (Ferdinand *et al.,* 2004).

Even though there may be discrepancies between the results obtained from phylogenetical studies, it seems to be a general conclusion that distinct lineages can be defined and that *M. tuberculosis* is clonal where horizontal gene transfer does not occur. Mathema *et al.,* 2006, suggested reasons for the differences in studies including evolutionary convergence that is specific to certain loci which would affect results particularly if only one genetic marker is used, or locus-specific transition, which would again affect results when using a single marker (Mathema *et al.,* 2006). These potential problems can be overcome by using a combination of molecular methods, including sSNP, deletion mapping of more than one region, IS*6110*-RFLP, spoligotyping and MIRU-VNTR.

Molecular typing techniques are of great value for phylogenetic analysis of MTBC strains as they add further resolution to lineages that have already been defined by sSNP. Phylogenetic analysis can help to identify and assign strains that have been identified in epidemiological studies to further the knowledge of the association between the MTBC strain genotypes.

1.3.3 Other applications

Molecular typing techniques can be applied to clinical situations that are most commonly experienced in laboratories, including detection of laboratory cross-contamination and mixed tuberculosis infection. In order to obtain results rapidly, but still ensuring that results are reliable, typing methods based on PCR amplification are more suitable. Using a combination of spoligotyping and MIRU-VNTR means that results are obtained quicker, whilst still having the discriminatory level of IS*6110*-RFLP.

Laboratory cross-contamination is a potential problem when diagnosing TB. Ideally the rate of cross-contamination in a laboratory should be less than 1%, but reports have shown contamination rates of 0.1 to 65%, i.e. in some laboratories it is a serious issue that needs to be addressed (Bhattacharya *et al.,* 1998; de *et al.,* 1999; Ruddy *et al.,* 2002). This is where molecular typing techniques come into play as they allow reliable detection of cross-contamination. Identifying false-positive cultures of *M. tuberculosis* as early as possible ultimately means the prevention of inappropriate treatment of TB in the patient. Whilst IS*6110*-RFLP can be used to confirm results of spoligotyping and MIRU-VNTR, this method is not ideal as it is time consuming and would delay detection.

For example, in a study on cross-contamination carried out at the Health Protection Agency, Mycobacterium Reference Unit (HPA, MRU), IS*6110*-RFLP was used to confirm results using another technique, rapid PCR-based epidemiological typing (RAPET) (Drobniewski *et al.,* 2003). This study showed that there were three types of cross-contamination including laboratory cross-contamination, bronchoscopic-related and ward-based incidents. In another case study, specimens from 6 different patients, which had been decontaminated on the same day, were all positive for *M. tuberculosis* (Allix *et al.,* 2004). Upon MIRU-VNTR typing it was suggested that there was possible laboratory contamination as profiles for the majority of isolates were identical, whilst isolates from patients that were smear negative had different MIRU-VNTR profiles.

Although mixed cultures can be detected using IS*6110*-RFLP and spoligotyping, the patterns can be complex to interpret (Pavlic *et al.,* 1999; Allix *et al.,* 2004). MIRU-VNTR is more appealing for detecting mixed cultures as the simultaneous existence

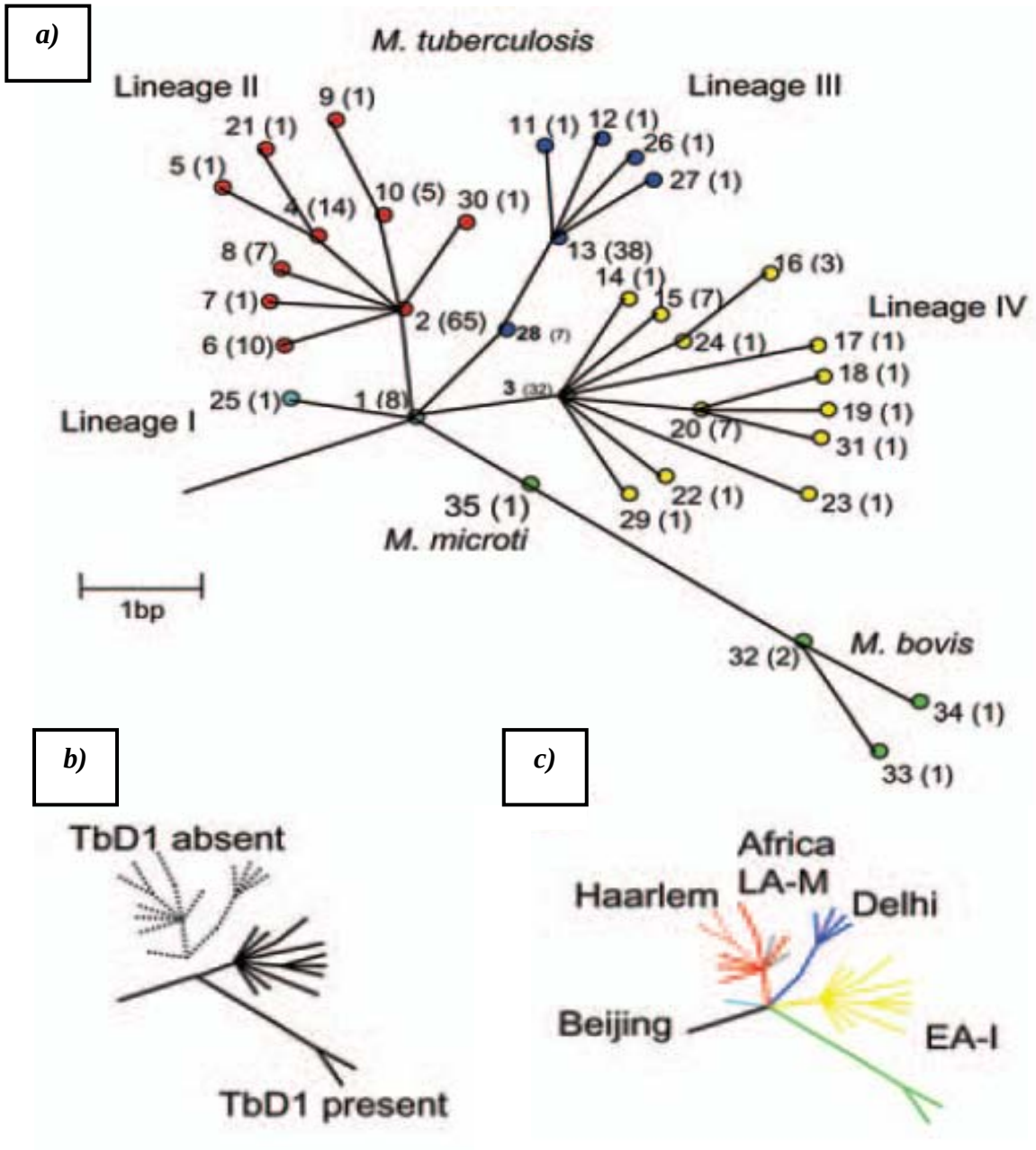
of 2 alleles in individual loci can be more easily identified. In addition, the alleles can be distinguished from stutter peaks that are commonly observed. Mixed cultures can be differentiated from laboratory cross contamination especially if follow-up isolates are obtained from the patient in question (Allix *et al.,* 2004).

1.4 PROJECT AIMS

The aims of the present study were to:-

- identify the various common and unique clinical *M. tuberculosis* strains circulating within the ethnically diverse population in London (http://www.londonshealth.gov.uk/pdf/HINL2004/hilfullreport2004.pdf) by doing cluster analysis using molecular typing data for all MTBC isolates from TB cases reported within London during a one year time period from the 1st April 2005 to 31st March 2006,
- understand the global representation of *M. tuberculosis* strains in the study population of London by performing epidemiological analysis using molecular typing and country of birth data,
- construct a preliminary panel of strains that would be as representative of the global population of TB strains as possible and include a broad spectrum of *M. tuberculosis* strains against which any vaccine should be effective; the strains in this panel of *M. tuberculosis* strains would be useful for selecting strains for future vaccine evaluation studies and general TB research,
- understand if TB strains in the preliminary panel are indeed different from each other by testing the null hypothesis that strains are phenotypically similar by performing a series of initial *in vitro* experiments followed by *in vivo* experiments on a smaller panel of strains, representing the six MTBC phylogenetic lineages defined by Baker *et al*., 2004 and Gagneux *et al*., 2006, derived from the preliminary panel.

Figure 1.7 *The phylogeny of M. tuberculosis, where (a) shows a maximum parsimony tree of M. tuberculosis and M. bovis lineages based on the combinations of sSNPS, (b) shows how absence and presence of TbD1 is related to the lineages, and (c) shows the relationship between the defined phylogenies and the various strains including Beijing, Harleem, Africa, Latin America-Mediterranean, Delhi, and East Africa-India (EA-I). Taken from Baker et al., 2004 (Baker et al., 2004).*



CHAPTER 2

MATERIALS & METHODS

2.1 PREPARATION OF SOLUTIONS, MEDIA & REAGENTS

2.1.1 Spoligotyping stock solutions

20% sodium dodecyl sulphate (SDS) was prepared by dissolving 100 grams (g) SDS (BDH Laboratory Supplies, Poole, UK) in molecular grade water (Sigma-Aldrich, Poole, UK) to make a final volume of 500 millilitres (ml).

0.5M Ethylenediaminetetraacetic acid disodium salt dihydrate 99+% (EDTA) (pH 8.0) was prepared by dissolving 93.0g EDTA (Sigma-Aldrich) in distilled water to make a final volume of 500ml. The pH was adjusted to 8.0 using sodium hydroxide (NaOH) pellets (BDH Laboratory Supplies).

All stock solutions were stored at room temperature.

2.1.2 Solutions for spoligotyping that are made fresh each time

2× saline sodium phosphate EDTA (SSPE)/0.1% SDS was prepared by adding 30ml 20× SSPE buffer (Sigma-Aldrich), 1.5ml 20% SDS and distilled water to make a final volume of 300ml per membrane. SDS was not added to concentrated SSPE as this would cause it to precipitate.

2× SSPE/0.5% SDS was prepared by adding 100ml 20× SSPE buffer, 25ml 20% SDS and distilled water to make a final volume of 1000ml per membrane.

2× SSPE was prepared by diluting 50ml 20× SSPE in distilled water to make a final volume of 500ml per membrane.

1% SDS was prepared by diluting 10ml 20% SDS in distilled water to make a final volume of 200ml per membrane.

20 millimolar (mM) EDTA was prepared by diluting 4ml 0.5 molar (M) EDTA in distilled water to make a final volume of 100ml per membrane.

2.1.3 Media for culturing *M. tuberculosis* strains

Middlebrook 7H9 medium with 0.05% Tween-80 was prepared by dissolving 4.7g BD Difco Middlebrook 7H9 broth (Becton, Dickinson and Company, New Jersey, United States of America; USA) in 900ml distilled water and sterilising the medium by autoclaving at 121°C for 10 minutes (mins). Tween-80 solution was added after

autoclaving. In a Class II cabinet, 5ml 10% Tween-80 solution (Sigma-Aldrich) was added to the sterile medium at a final concentration of 0.05% in 1000ml. The Middlebrook 7H9 medium with 0.05% Tween was filter sterilised using a 0.2 micrometre (μm) filter unit (Nalgene Labware, New York, USA) and stored at 20°C. Before use, depending on the volume of medium required for an experiment, BD BBL Middlebrook OADC Enrichment (Becton, Dickinson and Company) was added at a medium:OADC ratio of 9:1. OADC supplement was not added to media when it was made initially to give the media a longer shelf life.

Middlebrook 7H11 agar plates were prepared by dissolving 20.5g Middlebrook 7H11 Agar Base (Sigma-Aldrich) in 900ml distilled water and adding 5ml glycerol (Sigma-Aldrich). The medium was sterilised by autoclaving and then left to cool down to 50°C. Only then was 100ml OADC added to the medium and mixed using a magnetic stirrer. The liquid agar was dispensed into OPTILUX Petri Dishes 100×20 millimetres (mm) Style (Becton, Dickinson and Company) so that there was approximately 35ml medium per plate.

2.1.4 Reagents for DNA extraction

Tris-EDTA (TE) buffer (10mM Tris, 1mM EDTA pH8.0) was prepared by measuring 200 microlitres (μl) Tris-EDTA Buffer (Sigma-Aldrich) and using molecular grade water to make the volume up to 20ml. This reagent was stored at room temperature.

Lysozyme (100mg/ml) was prepared by dissolving 1g lysozyme from chicken egg white (Sigma-Aldrich) in 10ml molecular grade water, after which 1ml aliquots were prepared in 1.5ml microcentrifuge tubes (Alpha Laboratories Limited (Ltd.), Hampshire, UK), which were stored at -20°C.

Buffer AW1 and AW2 from the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) were prepared by adding the volume of ethanol 96-100% (Sigma-Aldrich) specified on the bottle. These reagents were stored at room temperature.

2.1.5 Media & solutions for tissue culture

Complete Roswell Park Memorial Institute (RPMI) 1640 medium was prepared aseptically in a Class II safety cabinet with a final concentration of 10% Standard Quality Foetal Bovine Serum (FBS; PAA Laboratories, Wagram, Austria), 100 units (U) Penicillin/ 100 ug Streptomycin (Invitrogen Ltd., Paisley, UK), 20 mM L-glutamine (Invitrogen), and 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic

acid (HEPES) (Gibco Invitrogen, California, USA). Media were stored at 20°C ready for use.

1× DPBS was prepared by diluting 100ml 10× Dulbecco's phosphate-buffered saline (DPBS) without calcium and magnesium (Gibco, Invitrogen) with 900ml distilled water. The full volume was filter sterilised using a 0.2μm filter unit and stored at 20°C ready for use.

2.2 MOLECULAR TYPING OF *M. TUBERCULOSIS* STRAINS

2.2.1 Preparation of *M. tuberculosis* DNA extracts

Culturing of *M. tuberculosis* specimens and DNA extraction was performed at the HPA, MRU. During a 12 month period, between 1st April 2005 and 31st March 2006, there were a total of 2363 isolates, which had been received from 30 hospitals and/or laboratories situated within the M25 region of London, and had been routinely identified as MTBC.

Crude DNA extraction was performed in biological safety cabinets (BSC) within a Containment Level 3 laboratory, following all safety guidelines. In brief, 100μl of each culture was aliquoted into individually labelled 1.5ml microcentrifuge tubes. An equal volume of chloroform (VWR International, Leicestershire, UK) was added to each tube. After ensuring the cap was screwed on tightly, each tube was vortexed for 30 seconds (secs). Tubes were not opened at this stage due to aerosol production during vortexing. Tubes containing culture/chloroform mixture were heated at 80°C for 20 mins after which tubes were stored at -20°C ready for use.

2.2.2 DNA amplification for MIRU-VNTR typing

Typing was performed on the crude DNA extracts from all isolates, analysing the 15 MIRU and 3 ETR minisatellite regions spanning the *M. tuberculosis* genome and included the loci MIRU-2, -4, -10, -16, -20, -23, -24, -26, -27, -31, -39, -40, and ETR-A, -B, -C. WellRED dye labelled forward primers (Sigma Proligo, Poole, UK) and unlabelled reverse primers (Invitrogen Ltd.) were already in use within the laboratory and were designed using previous studies by Frothingham and O'Connell, 1998 and Supply *et al.,* 2001 with some modifications during optimisation (Frothingham *et al.,* 1998; Supply *et al.,* 2001). Sequences of all primers and labelling dyes are included in Table 2.1.

Table 2.1 *The sequences for all primers used for MIRU-VNTR typing.*

Primer	Forward sequence	Reverse sequence	Dye used for labelling
MIRU-2	5'- CAG GTG CCC TAT CTG CTG ACG-3'	5'-GTT GCG TCC GGC ATA CCA AC-3'	WellRED D3
MIRU-4	5'-GTC AAA CAG GTC ACA ACG AGA GGA A-3'	5'-CCT CCA CAA TCA ACA CAC TGG TCA T-3'	WellRED D2
MIRU-10	5'-ACC GTC TTA TCG GAC TGC ACT ATC AA-3'	5'- CAC CTT GGT GAT CAG CTA CCT CGA T-3'	WellRED D4
MIRU-16	5'-CGG GTC CAG TCC AAC TAC CTC AAT-3'	5'-GAT CCT CCT GAT TGC CCT GAC CTA-3'	WellRED D2
MIRU-20	5'-CCC CTT CGA GTT AGT ATC GTC GGT T-3'	5'-CAA TCA CCG TTA CAT CGA CGT CAT C-3'	WellRED D2
MIRU-23	5'-CGA ATT CTT CGG TGG TCT CGA GT-3'	5'-ACC GTC TGA CTC ATG GTG TCC AA-3'	WellRED D4
MIRU-24	5'-GAA GGC TAT CCG TCG ATC GGT T-3'	5'-GGG CGA GTT GAG CTC ACA GAA C-3'	WellRED D3
MIRU-26	5'-GCG GAT AGG TCT ACC GTC GAA ATC-3'	5'- TCC GGG TCA TAC AGC ATG ATC A-3'	WellRED D4
MIRU-27	5'- TCT GCT TGC CAG TAA GAG CCA-3'	5'-GTG ATG GTG ACT TCG GTG CCT T-3'	WellRED D3
MIRU-31	5'-CGT CGA AGA GAG CCT CAT CAA TCA T-3'	5'- AAC CTG CTG ACC GAT GGC AAT ATC-3'	WellRED D3
MIRU-39	5'- CGG TCA AGT TCA GCA CCT TCT ACA TC-3'	5'- GCG TCC GTA CTT CCG GTT CAG-3'	WellRED D2
MIRU-40	5'-GAT TCC AAC AAG ACG CAG ATC AAG A-3'	5'-TCA GGT CTT TCT CTC ACG CTC TCG-3'	WellRED D3
ETR-A	5'-AAA TCG GTC CCA TCA CCT TCT TAT-3'	5'-CGA AGC CTG GGG TGC CCG CGA TTT-3'	WellRED D2
ETR-B	5'-GCG AAC ACC AGG ACA GCA TCA TG-3'	5'-GGC ATG CCG GTG ATC GAG TGG-3'	WellRED D4
ETR-C	5'-GTG AGT CGC TGC AGA ACC TGC AG-3'	5'-GGC GTC TTG ACC TCC ACG AGT G-3'	WellRED D4
VNTR 2163b	5'-CGT AAG GGG GAT GCG GGA AAT AGG-3'	5'-CGA AGT GAA TGG TGG CAT-3'	WellRED D2
VNTR 2347	5'-GCC AGC CGC CGT GCA TAA ACC T-3'	5'-GCC AGC CGC CGT GCA TAA ACC T-3'	WellRED D2
VNTR 3232	5'-CAC TAG TTG TTG CGG CGA TGG T-3'	5'-AGC CAC CCG GTG TGC CTT GTA TGA C-3'	WellRED D3
VNTR 2163a	5'-CCC GGG GCG CTC GTG ATG-3'	5'-CAC TAG TTG TTG CGG CGA TGG T-3'	WellRED D4
VNTR 1982	5'-GGA ATG GCT ACG GAA GGA ATA CTC-3'	5'-AAG GGC GGC ATT GTG TTC C-3'	WellRED D2
VNTR 3336	5'-GAT CGG GTG CAG TGG TTT CAG GTG-3'	5'-CCC GGG GCG CTC GTG ATG-3'	WellRED D3
VNTR 4052	5'-AAC GCT CAG CTG TCG GAT-3'	5'-CGG CGG CAC CCT GGA GTC TGG-3'	WellRED D4

All forward primers were labelled for capillary gel electrophoresis and unlabelled for agarose gel electrophoresis; reverse primers were unlabelled

A 10ml 2× reaction buffer was prepared using 10× ammonium (NH_4) reaction buffer, 50mM magnesium chloride ($MgCl_2$) and 100mM deoxynucleotide triphosphates (dNTPs), (all from Bioline, London, UK) and molecular grade water. Final concentrations were 2× NH_4 reaction buffer, 3mM $MgCl_2$ and 0.4mM of each dNTPs. Aliquots of 1.5ml were stored at -20°C ready for use.

PCR were set up as either simplex or duplex reactions so 9 primer sets, containing forward and reverse primers for each locus, were prepared using molecular grade water giving a final concentration of each primer of 0.5 µM. Primer set 1 contained MIRU-4 and -16 primers; set 2 contained MIRU-39 and ETR-A primers; set 3 contained MIRU-20 primers; set 4 contained MIRU-2 and-24 primers; set 5 contained MIRU-31 and -40 primers; set 6 contained ETR-C primers; set 7 contained MIRU-10 and -23 primers; set 8 contained MIRU-27 and ETR-B primers; set 9 contained MIRU-26 primers. Primer mixes were stored at -20°C ready for use.

Amplification of the different loci in each DNA extract was conducted as follows: 10µl PCR reactions were prepared in 0.2ml 96 well Thermal Cycler Plates (Alpha Laboratories), for each primer set, and contained 4.33µl 2× reaction buffer, 0.08µl 99.9% dimethyl sulphoxide (DMSO) (Sigma-Aldrich), 4.50µl primer set, and 0.09µl 5U/µl BIOTAQ DNA polymerase (Bioline). For optimisation and amplification of MIRU-4 and –16 loci, 0.27µl 50mM $MgCl_2$ was added to each reaction. Crude DNA extracts were allowed to thaw and any debris was spun down by centrifugation at 8050×*g* for 2 mins and then 1µl was mixed into each reaction by pipetting up and down.

All reactions were spun down by centrifuging plates at 250×*g* for 1 minute in an IEC Centra CL3R (Thermo Life Sciences, Basingstoke, UK). A GeneAmp PCR System 9700 (Applied Biosystems, Warrington, UK) was used to amplify each locus. The following cycle was used for amplification: 95°C for 180 secs; 35 cycles of 95°C for 30 secs, 60°C for 30 secs, and 72°C for 60 secs; 72°C for 300 secs. During amplification, PCR fragments were labelled with dyes so fragment sizes could be detected using capillary electrophoresis.

For agarose gel electrophoresis, DNA amplification procedures were exactly the same except simplex PCR reactions were set up to amplify loci individually using unlabelled forward and reverse primers with the same primer sequences.

2.2.3 Capillary gel electrophoresis

Capillary gel electrophoresis allowed identification of amplified product according to their molecular weight and the dye with which products were labelled during amplification. Analysis of PCR products was performed using Beckman Coulter CEQ 8000 Genetic Analysis System (Beckman Coulter, Fullerton, USA). MIRU and ETR loci were analysed together in three separate capillaries. For each DNA extract, amplified products from MIRU-2, -4, -10, -16, -23, -24 were analysed in capillary A, products from MIRU-27, -31, -39, -40, ETR-A and -B were analysed in capillary B and products from MIRU-20, -26 and ETR-C were analysed in capillary C.

Due to the varying intensities of labelling dyes, PCR products were diluted and pooled so resulting intensities of peaks on the chromatogram, following electrophoresis, were similar. PCR products that were labelled with Dye 2 were diluted 1 in 10 using molecular grade water in the PCR plate. PCR products labelled with Dye 3 and 4 were diluted 1 in 100 using the diluted Dye 2 fragments as the diluent. Pooled PCR products were mixed by pipetting up and down and centrifuged for 1 minute at 250×*g*.

A Sample Microtiter plate (Beckman Coulter) was prepared so each well contained 25µl sample loading solution (Beckman Coulter) and 0.1µl DNA Size Standard 600 (Beckman Coulter) for analysing MIRU and ETR loci. For each DNA extract, 1µl of pooled PCR product was added to separate wells. The prepared sample plate was loaded onto the analysis system according to the manufacturer's instructions along with a Beckman Coulter 96 well plate filled with Separation buffer (Beckman Coulter) and a 10ml CEQ Separation gel cartridge (Beckman Coulter).

After capillary gel electrophoresis was completed, the automatically generated raw data traces, displaying peaks for each locus, were analysed using the calling tables in Appendix 1 and 2, to generate corresponding traces displaying peaks that were annotated with the locus name, molecular weight and calculated number of repeats.

Manual agarose gel electrophoresis was performed if there was an absent peak for any of the loci.

Reproducibility of data was verified by repeating MIRU-VNTR typing for all loci on randomly selected isolates.

2.2.4 Agarose gel electrophoresis

For those isolates with loci that had missing peaks after capillary gel electrophoresis, the loci were amplified from the isolate as described in Section 2.2.2, using unlabelled forward primers; DMSO was not used in the PCR reactions and a negative control for each locus primer set was included using 1µl molecular grade water instead of DNA template.

A 1.2% weight/volume (w/v) agarose gel (Agarose LE Analytical grade; Promega, Southampton, UK) was used to resolve 5µl of each PCR product against a 1000 b.p. HyperLadder IV (Bioline). The gel was viewed on an ultraviolet transilluminator (UVP inc, Cambridge, UK) and an image captured using a digital camera (Canon PowerShot A620, Surrey, UK). Product sizes were determined by comparing bands with the ladder and using the calling tables (Appendix 1 and 2) to calculate the number of repeats in the locus of interest.

In some cases, amplified ETR-A fragments had to be resolved against the 2000 b.p. ladder or a custom size ladder comprised of amplified ETR-A fragments as the molecular weight of some products exceeded 1000 b.p. Also, during capillary gel electrophoresis of amplified MIRU-4 fragments, there were peaks with molecular weights considerably lower than expected indicating that MIRU-4 was more polymorphic with possible deletions. Whilst these polymorphisms were detected by capillary electrophoresis, detection was not as easy with agarose gel electrophoresis as the latter method provides lower resolution. Therefore, a custom size ladder for MIRU-4 and another ladder for MIRU-4 with the various deletions was prepared.

2.2.5 Preparation of custom size ladders for ETR-A and MIRU-4

From the data obtained after capillary gel electrophoresis, extracts with a known number of repeats in ETR-A, MIRU-4 and MIRU-4 with the various deletions were identified. Seven extracts were selected for each of the three ladders with different

numbers of repeats to give a broad range of values and ensure that there were not too many bands close together during electrophoresis.

The locus, MIRU-4 or ETR-A, was amplified from each isolate in separate 40µl reactions, containing 20µl 2× reaction buffer, 10µl unlabelled primer mix for either ETR-A or MIRU-4, 1µl BIOTAQ polymerase and 2µl DNA. Amplification conditions were as follows: 95°C for 180 secs; 30 cycles of 95°C for 30 secs, 55°C for 30 secs, and 75°C for 60 secs; 72°C for 300 secs. After amplification, 5µl of each product was resolved against a 1000 b.p. ladder, as described in Section 2.2.4, to confirm the presence of PCR products and to approximate product sizes. The full volume of PCR product was combined to produce three separate custom size ladders for ETR-A, MIRU-4 and MIRU-4 with deletions that were stored at -20°C. When ladders were loaded onto gels, HyperLadder IV was also loaded in adjacent wells.

2.2.6 MIRU-VNTR typing using more discriminative VNTR loci

All MIRU-VNTR typing data was entered into a Microsoft Access database, which was designed by Dr. Tim Brown and Dr. Vladyslav Nikolayevskyy, HPA, MRU, and was directly linked to BioNumerics version 3.00. Firstly MIRU-VNTR profiles produced by capillary gel electrophoresis were exported directly into Microsoft Excel format. Scripts within BioNumerics were used to create a table with one row per extract and columns for extract identification and each locus. The table format was used to export data into a Microsoft Access database, where additional typing data from agarose gel electrophoresis were manually added. Cluster analysis was performed in BioNumerics, using the n-1 method, to identify all DNA extracts whose MIRU-VNTR profiles were identical and formed clusters, which by definition contained 2 or more isolates, and which would be typed using the additional 7 VNTR loci.

The more discriminatory 7 VNTR loci which were evaluated included VNTR-2163B, -2347, -3232, -2163A, -1982, -3336 and -4052. Serial isolates were also typed using these VNTR loci as part of another smaller study investigating the reproducibility of the VNTR loci, which will be described in Section 2.2.8 and in Chapter 4. Each locus was amplified from the DNA extracts as described in Section 2.2.2 and six primer sets were prepared as follows. Set 1 contained VNTR-2163B and -2347 primers; set 2 contained VNTR-3232 primers; set 3 contained VNTR-2163A primers; set 4

contained VNTR-1982 primers; set 5 contained VNTR-3336 primers; set 6 contained VNTR-4052 primers. Sequences of all primers and dyes used for labelling are included in Table 2.1. PCR reactions were set up as described previously, except Diamond DNA polymerase (Bioline) was used instead, as VNTR loci seemed to require a polymerase with higher specificity for effective amplification. The amplification cycle was: 95°C for 180 secs; 35 cycles of 95°C for 30 secs, 60°C for 30 secs, and 72°C for 120 secs; 72°C for 300 secs.

Capillary gel electrophoresis was performed as described in Section 2.2.3 and VNTR loci analysed in two separate capillaries where VNTR-2163B, -2347, -3232, -2163A were analysed in capillary A and VNTR-1982, -3336, -4052 were analysed in capillary B. When loading the Sample Microtiter plate, as well as using 0.1µl DNA Size Standard 600, 0.1µl MapMarker D1 labelled 640-1000 (BioVentures, Inc., Murfreesboro, USA) was added to each well as some VNTR fragment sizes were expected to exceed 600 b.p. As for MIRU and ETR loci, if there were any missing peaks after capillary gel electrophoresis, agarose gel electrophoresis was performed as described in Section 2.2.4, but using a 2000 b.p. HyperLadder II standard (Bioline) instead. The calling table for the VNTR loci can be found in Appendix 2. The additional typing data were entered into the same Access database.

2.2.7 Spoligotyping

Spoligotyping was performed on all of the 2363 crude DNA extracts as follows: firstly, 1ml mastermix containing 5% DMSO, 2× NH_4 reaction buffer, 3mM $MgCl_2$, 0.4mM of each dNTP and 0.5nmol/ml of each primer, biotinylated DRa (5'-CCG AGA GGG GAC GGA AAC-3') and DRb (5'-GGT TTT GGG TCT GAC GAC-3') was prepared and stored at -20°C ready for use. Primers were obtained from Isogen Life Science, Maarssen, The Netherlands, and sequences were as described by Goguet de la Salmoniere *et al.,* 1997 (Goguet de la Salmoniere *et al.,* 1997).

Spoligotyping was performed as described in Kamerbeek *et al*., 1997 (Kamerbeek *et al.,* 1997). Each spoligo-membrane (Isogen Life Science) accommodated forty samples including two positive controls, *M. tuberculosis* strain H37Rv and *M. bovis* BCG P3 (Isogen Life Science), which were included for each membrane. PCR reactions, containing 18µl diluted mastermix and 2µl DNA, were set up in 0.2ml 96 well Thermal Cycler Plates. The DR region was amplified using the following

conditions: 94°C for 300 secs; 30 cycles of 94°C for 30 secs, 55°C for 60 secs, and 72°C for 60 secs; 72°C for 300 secs. As DNA was amplified, reverse strands were biotin labelled. During amplification, solutions required for hybridisation and detection steps were made up as described in Section 2.1.1 and 2.1.2.

Hybridised DNA was detected using chemiluminescence: equal volumes of enhanced chemiluminescent (ECL) Detection Reagents 1 and 2 (Amersham Biosciences) were mixed to make a total volume of 20ml for each membrane. Membranes were incubated with ECL reagent for 1 minute before being transferred into clear plastic covers ensuring there were no air bubbles. A light sensitive Hyperfilm ECL (Amersham Biosciences) was exposed to the membrane for 2 mins, before it was developed and fixed using GBX developer and replenisher and GBX fixer and replenisher (Kodak, New York, USA). Depending on whether the intensity of the signal was faint or too dark, membranes were used again to expose another film for a longer or shorter period of time, respectively.

Membranes were re-used for more spoligotyping by stripping PCR products from the previous analysis using stringent washing conditions. Membranes were washed twice in 100ml 1% SDS in the hybridisation bottles at 80°C for 30 mins and then once in 100ml 20mM EDTA at room temperature on the rocker. Each membrane was stored in 100ml 20mM EDTA ready for the next hybridisation. Both solutions were prepared as described in Section 2.1.1 and 2.1.2.

Each spoligotype profile was scanned into specifically formatted strips with corresponding black and white dots for present and absent spacers, respectively, and since BioNumerics was directly linked to the main Microsoft Access database, spoligotyping data automatically appeared in the Access database as a 43 digit binary code. In BioNumerics, the binary codes were converted into 15 digit octal codes using a script within BioNumerics, which was based on the study by Dale *et al*., 2001, to make the spoligotyping digital profiles easier to use for further analysis (Dale *et al.*, 2001). The octal codes, were used to assign a spoligofamily to each isolate using an online tool (http://cgi2.cs.rpi.edu/~bennek/SPOTCLUST.html), based on the SpolDB3-based probabilistic approach described by Vitol *et al*., 2006 (Vitol *et al.*, 2006).

2.2.8 Evaluation of MIRU-VNTR typing using VNTR-1982 and -3232

The stability of VNTR-1982 and -3232 and the reproducibility of typing data, under different amplification and fragment detection conditions, were investigated. The details of this study, including the optimised methodology, are described in greater detail in Chapter 4. In brief, DNA extracts from 16 MTBC isolates, which had been included for quality control purposes during MIRU-VNTR typing and had complete data for 12 MIRU, 3 ETR and 7 VNTR loci, were selected to cover a range of repeats in MIRU-26, ETR-B and VNTR-1982 and -3232 loci. These loci were amplified using different polymerases including Bioline BIOTAQ and Diamond polymerase and Qiagen HotStartTaq DNA polymerase and HotStartTaq *Plus* DNA polymerase. Products were resolved on a 1.2% (w/v) agarose gel to ensure PCR products were present, and then analysed using capillary gel electrophoresis. During this analysis the effect on detected fragment sizes was evaluated using different parameters that involved changes in capillary temperature and the time allowed for the denaturing and separation steps.

Reproducibility of typing using all MIRU, ETR and VNTR loci, which were included for this study, was also investigated in a blinded study, by checking for consistency between MIRU-VNTR profiles of serial isolates (i.e. multiple isolates from the same patient taken at successive time intervals).

2.3 SELECTING STRAINS OF PARTICULAR INTEREST

2.3.1 Epidemiology

Demographic data, including gender, date of birth and country of birth, was available from the London TB Register for the patients included in the present study. We had ethical approval to use the demographic data of patients for the purposes of research. To perform epidemiological analysis in the present study, the country of birth data was used and was available for the majority of extracts. In addition, the drug sensitivities for the first line drugs, including isoniazid, ethambutol, rifampicin, pyrazinamide and streptomycin, were available at the HPA, MRU.

In order to understand the degree to which London TB isolates reflected the global TB burden, the number of isolates in each spoligofamily was plotted onto a global map using the country of birth data to assign region of origin for each patient. The global

regions were as defined by WHO; to avoid any bias, serial isolates were not taken into consideration i.e. one isolate per patient was included for analysis.

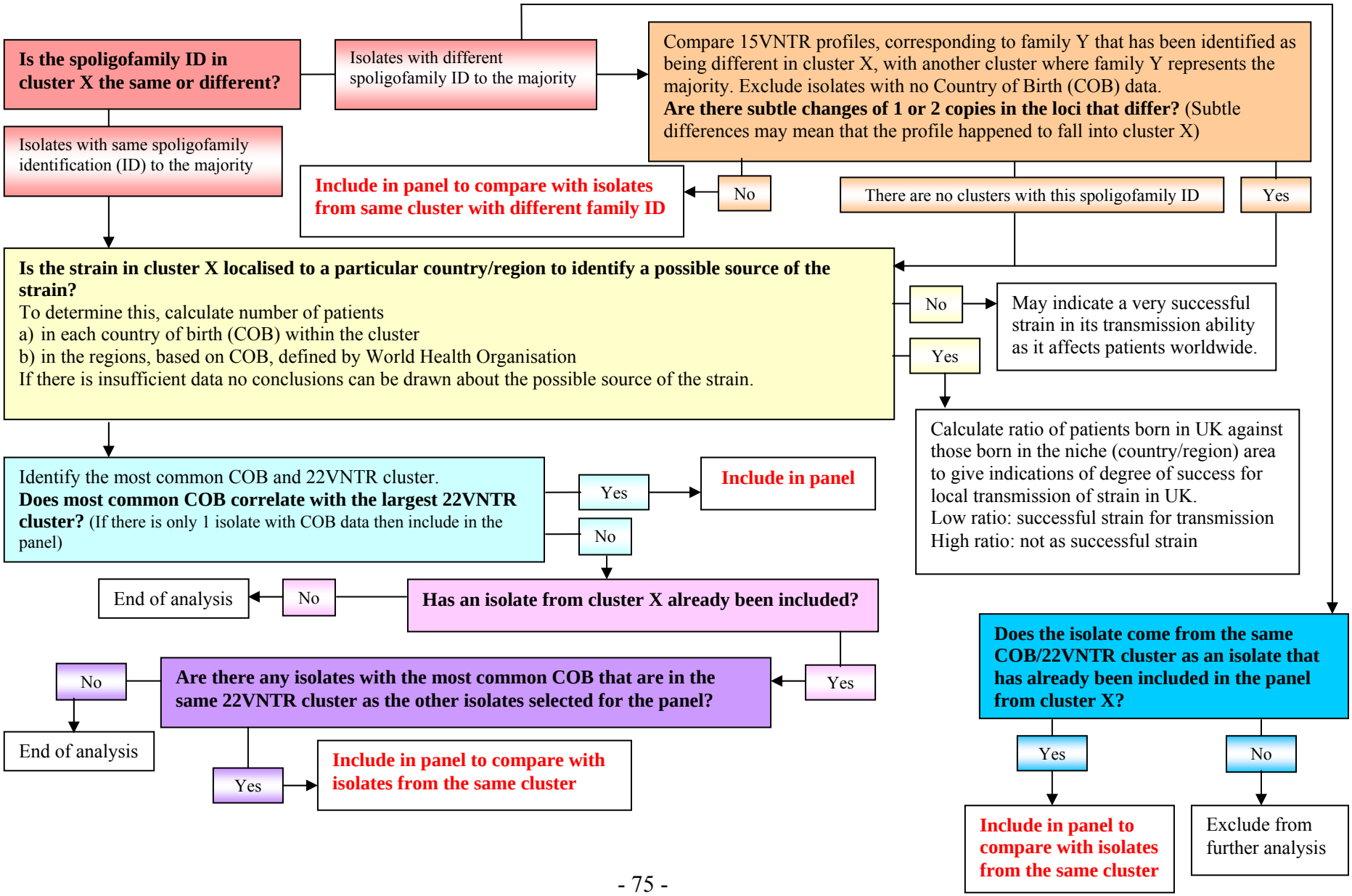
2.3.2 Selection of strains for phenotypic analysis

The molecular typing, cluster analysis and epidemiology data were used to select strains for Panel A. For further details of the selection analysis see Section 3.2.4. Clusters containing more than nine isolates were included for selection of strain and an algorithm was designed to select particular strains of interest and is shown in Figure 2.1. This algorithm selected strains based mainly on correlations, within the cluster, between the most common spoligofamily, the most common country of birth and the most common sub-cluster, which were identified using profiles for the more discriminatory 7 VNTR loci that split the MIRU and ETR clusters.

Strains were also selected if they did not fit any general trends observed within the MIRU and ETR clusters. Should there be any isolates within the cluster that had a different spoligofamily to other isolates within the cluster, the MIRU and ETR profile of this isolate was compared to isolates of the same spoligofamily from other clusters. If the profiles differed by more than 2 copies in any of loci, the isolate was included in the panel. Isolates taken from patients that had a country of birth and did not fit the general country of birth and spoligofamily identification trend in the cluster, were also included in the panel. If isolates had the same spoligofamily identification as strains already selected for the panel but with a different country of birth, it was also included in the panel to broaden the range of different *M. tuberculosis* strains.

Using the selection criteria described, 42 strains that had particular properties were selected. This panel will be referred to as Panel A from here onwards and details of the strains in this panel can be seen in Table 3.2.

Figure 2.1 *The algorithm that was applied to 12 MIRU and 3 ETR profile clusters to select particular strains of interest for inclusion in Panel A.*



2.3.3 Phenotypic analysis

The 42 Panel A strains were analysed *in vitro* for the following properties: immune response and growth of strains in tissue culture models and growth rates of strains in Middlebrook 7H9 culture systems.

As part of the preparation of a panel of strains representing the maximum diversity against which a vaccine would provide protection, *in vivo* phenotypic experiments were planned in guinea pig models. So that sufficient reproducibility could be achieved to critically test the null hypothesis that TB strains were fundamentally the same, a smaller number of strains derived from the original 42 strains in Panel A were selected to create a core second panel (Panel B) based on the different phylogenetic lineages within MTBC. The selection of strains for Panel B is described in more detail in Section 3.3. Also included in Panel B was a wild-type *M. africanum* strain. Laboratory strains, included *M. tuberculosis* H37Rv (one from HPA, MRU and the other from HPA, Centre for Emergency Preparedness and Response; CEPR, Porton Down), *M. tuberculosis* H37Ra and the vaccine strain *M. bovis* BCG. All Panel B and laboratory strains were included for phenotypic experiments.

2.4 PREPARATION OF MYCOBACTERIAL STOCK CULTURES

2.4.1 Culturing *M. tuberculosis* strains

In order to have sufficient uniform seed-stock for all phenotypic experiments, a stock of frozen cultures was prepared of the Panel B strains (obtained from the HPA, MRU archives), and all laboratory strains (both *M. tuberculosis* H37Rv strains, and *M. tuberculosis* H37Ra, and *M. bovis* BCG). All manipulations with cultures were carried out in Class I biological safety cabinets following safety guidelines.

Using 1ml graduated sterile plastic pasteur pipettes (Alpha Laboratories) for each strain, a pyruvate and glycerol slope (Media for Mycobacteria, Cardiff, UK) and a 50ml Falcon conical tube (Becton, Dickinson and Company) containing no more than 10 sterile 5mm diameter glass beads (Sigma-Aldrich) and 10ml Middlebrook 7H9 with 0.05% Tween-80 medium, were each inoculated with 2 drops of the original culture from archive. The medium was prepared as described in Section 2.1.3. Even though Tween would reduce clumping of the mycobacteria, the liquid cultures were vortexed so the glass beads would break up clumps as well (after vortexing, it was

important NOT to open the tubes as there was aerosol production during the process). All slopes and liquid cultures were incubated at 37°C. The liquid cultures were vortexed every 2 days to ensure that clumping was reduced as much as possible.

After 2 weeks the liquid cultures were sub-cultured, except 15ml Middlebrook 7H9 with 0.05% Tween-80 medium was used. The fresh liquid cultures were incubated at 37°C for 3 weeks, vortexing the cultures every 2 days.

2.4.2 Archiving *M. tuberculosis* cultures

After 3 weeks, when mycobacterial cells should potentially still be in the logarithmic phase of the bacterial growth curve, the cultures were removed from incubation and 1.5ml of each culture was transferred into labelled sterile 2.0ml cryovial tubes (Simport, Beloeil, Canada) so there were 10 cryovial tubes per strain, which were then stored at -80°C, ready for use in phenotypic experiments.

2.5 QUANTIFICATION OF *M. TUBERCULOSIS*

2.5.1 Developing a quantification assay for MTBC cultures

The accuracy of quantification of MTBC cultures using colony forming units (c.f.u.) is compromised by the tendency for mycobacterial cells to form clumps; the process is also relatively slow but considered the gold standard quantification procedure. An alternative methodology was sought to accurately quantify MTBC cultures and the development of the MTBC DNA quantification assay is described in detail in Section 5.2.

The detailed procedure for preparation of the *M. tuberculosis* DNA standard is described in Section 5.2. Briefly, after 4 weeks, the mycobacterial growth on slopes inoculated with *M. tuberculosis* H37Rv was re-suspended in TE buffer and incubated in an 80^{o}C waterbath for 50 mins to kill the mycobacteria. Bacterial cells were lysed with lysozyme. Proteinase K and Buffer AL (both provided in the Qiagen DNeasy Blood & Tissue kit) was added to the lysed bacterial cells, to help break down any cell debris, after which 96-100% ethanol (Sigma-Aldrich) was added to the tube. Purification of the extracted DNA was performed using the protocol and reagents provided with the Qiagen DNeasy Blood & Tissue kit.

The concentration of double-stranded DNA, in ng/µl, was calculated at 260nm, which was then used to calculate the number of genomes per µl in the DNA extract. Using this value, DNA standards containing 1×10^7, 1×10^6, 1×10^5, 1×10^4, 1×10^3, 1×10^2, 1×10^1 genomes/µl were prepared and stored at -20°C ready for use.

2.5.2 Comparing quantification methods for MTBC cultures

Liquid cultures of 30ml volume were prepared, as described in Section 2.4.1, for the Panel B strains and incubated at 37°C. At regular intervals of 3 or 4 days, the optical density at 600nm (OD_{600}) was measured for all cultures, and the number of mycobacteria was quantified using real-time PCR and c.f.u. The details of these procedures are described in Sections 5.3. For real-time PCR, primers amplifying the ribonucleic acid polymerase β-subunit-encoding gene (*rpoB*) region of the *M. tuberculosis* genome were used (refer to Section 5.2.2 for details on evaluation of primers that could be used for real-time PCR). The results of the comparison of these quantification methods is detailed in Section 5.3.4, but briefly, it was concluded that OD_{600} of 0.2 could be used to indicate mycobacterial cells being within the logarithmic phase of the growth curve and real-time PCR could be used as a more rapid method for quantifying MTBC cultures.

2.6 *IN VITRO* TISSUE CULTURE EXPERIMENTS

2.6.1 Preparation of mycobacteria cells for infection experiments

Tissue culture experiments were performed using the ten Panel B strains selected in Section 2.3.3 (refer to details of strains in Section 3.3), including *M. tuberculosis* Beijing, *M. tuberculosis* LAM10, two *M. tuberculosis* CAS, *M. tuberculosis* EAI5, and *M. africanum*, *M. tuberculosis* H37Rv (MRU), *M. tuberculosis* H37Rv (Porton Down), *M. tuberculosis* H37Ra, and *M. bovis* BCG. These experiments were performed to provide some insight into how human macrophage-like cells are affected after infection with the different strains.

Ten days before setting up a tissue culture infection experiment, one of the archived tubes of frozen cultures for each of the strains was thawed at room temperature and the full volume used to inoculate a 50ml falcon tube, containing 15ml Middlebrook 7H9 with 0.05% Tween-80 media and beads. After vortexing, the cultures were incubated at 37°C for eight days, but every two days the cultures were vortexed.

Two days before the tissue culture infection experiment, the OD_{600} was measured for each of the cultures using the procedure described in Section 2.5.2 and detailed in Section 5.3.1 and recorded on the falcon tube. At the same time, a drop was placed on a section of Columbia blood agar (CBA) plate (Oxoid Ltd., Cambridge, UK) and spread using a plastic loop to ensure the purity of cultures. If OD_{600} was above 0.2 for any of the cultures, then an aliquot of the culture was diluted accordingly, using fresh media, in new 50ml falcon tubes containing glass beads, so that the final volume of sub-culture was 15ml (OD_{600} 0.2). All cultures and CBA plates were incubated at 37°C.

After overnight incubation, the CBA purity plates were checked for contamination and then re-incubated. If there was contamination, a culture from any previous experiment was used to prepare 15ml culture at OD_{600} 0.2. A drop of all cultures that had been diluted the previous day was also plated onto CBA plates. Cultures and plates were incubated until required for the infection experiment on the following day.

2.6.2 Preparation of monocytic THP-1 cells for infection experiments

The human monocytic cell line, THP-1, was obtained from the American Type Culture Collection (ATCC, Virginia, USA). All tissue culture procedures were carried out under aseptic conditions in a Class II safety cabinet, which was decontaminated with a 30-minute ultraviolet exposure before use and sprayed with 70% ethanol before and after use. All items, including safety gloves, were decontaminated by spraying 70% ethanol before going inside the cabinet. Sterile consumables were used for all procedures including all 5ml, 10ml, and 25ml pipettes (Sarstedt Ltd., Nümbrecht, Germany), all 75ml cell culture flasks (Becton, Dickinson and Company), and 50ml falcon tubes. Cell culture flasks were sprayed carefully with ethanol ensuring no ethanol spray went on the vented cap.

The THP-1 cells were thawed on arrival and cultured in complete RPMI 1640 medium, pre-warmed to 37°C, in 75ml cell culture flasks. Complete RPMI medium was prepared as described in Section 2.1.5. Cells were incubated at 37°C, 5% carbon dioxide (CO_2). The THP-1 cells were not activated therefore the cells were proliferative and non-adherent. Every 2 or 3 days, depending on the density of the cells in the culture flasks, the THP-1 cells were sub-cultured by firstly pooling all

THP-1 cells from all flasks by centrifuging cells down in a single 50ml polystyrene conical centrifuge tube (Becton, Dickinson and Company) for 5 mins at 1100×*g* and then resuspending cells in 10ml fresh complete RPMI medium. Five culture flasks containing 10ml fresh complete RPMI medium were inoculated aseptically with 2ml of the pooled THP-1 cells.

Two days prior to setting up an infection experiment, THP-1 cells were counted, activated and plated. Firstly, all cells were pooled in a small volume of fresh complete RPMI medium as described above. For cell counting, 30µl of cell culture was diluted 1 in 2 using Trypan blue solution (Sigma-Aldrich). After Trypan blue staining, live THP-1 cells looked bright, clear and round when observed under a Motic Inverted Microscope (Motic Incorporation Ltd., Richmond, Canada), whilst dead THP-1 cells took up the stain and were blue. Trypan blue/THP-1 cell mixture was transferred to fill one of ten chambers of a KOVA Glasstic Slide 10 with Grid Chamber (Hycor Biomedical, Ltd., Penicuik, UK), holding approximately 6.6µl. One well of the slide contained a large 3 by 3 grid and in each square of this grid was a smaller 3 by 3 grid. The full calculation for establishing the number of THP-1 cells per ml is detailed in Appendix 3. The calculated value was used to dilute or concentrate the THP-1 culture accordingly so there were 300,000 cells per ml. In this study, 6ml of THP-1 cells, at the correct concentration, were required for each of the ten strains, and for uninfected, lipopolysaccharide (LPS) and IFN-γ controls.

THP-1 cells were activated by adding phorbol 12-myristate 13-acetate (PMA; Sigma-Aldrich) at a final concentration of 500ng/ml, so the cells differentiated from monocytic, proliferative, non-adherent THP-1 cells into macrophage-like, non-proliferative, adherent cells. A volume of 1ml PMA-activated THP-1 cells (corresponding to 300,000 cells) was plated into 12 wells in each of seven Costar 24-well flat bottom cell culture plates (Corning, New-York, USA), leaving a column of empty wells between the two strains per plate to avoid cross-contamination during further experimental procedures. The plates were incubated at 37°C, 5% CO_2 for 48 hours (hrs).

2.6.3 Infection of THP-1 cells with *M. tuberculosis*

On day 0, two days after the mycobacterial cultures and macrophage-like THP-1 cells were prepared, the THP-1 cells were infected with MTBC at a ratio of 1:1; numbers of MTBC were quantified using real-time PCR.

The tubes containing the culture for each strain were vortexed, and left to stand for 5 mins. The OD_{600} was measured for each of the cultures and two aliquots of 500µl from each culture were transferred into two labelled 1.5ml microcentrifuge tubes. The microcentrifuge tubes were centrifuged at 12,000×*g* for 10 mins, the supernatant from each tube safely discarded and the pelletted mycobacterial cells were resuspended in 100µl TE buffer. DNA extraction and purification procedures were then performed as described in Section 2.5.2 and detailed in Section 5.2 and the volume of Buffer AE for eluting DNA was 40µl.

Real-time PCR reactions were set up using *rpoB* primers, DNA standards (ranging from 1×10^1 to 1×10^7 genomes/µl) and the two DNA extracts of unknown concentration per strain, and then the number of genomes/ml culture was calculated (the detailed procedures for setting up PCR and calculating genomes/ml is described in Section 5.2.2, except DNA had been extracted from 500µl culture instead of 1ml). As there were duplicate DNA extracts per culture, the final number of genomes/ml was an average of the duplicates. In order to infect THP-1 cells with bacteria at a multiplicity of infection (MOI) of 1, the volume of each culture required to infect 300,000 THP-1 cells with 300,000 mycobacterial cells was calculated. During the PCR step, Middlebrook 7H11 plates were inoculated with serial dilutions of each culture and all plates incubated at 37°C for 3 to 4 weeks, to obtain c.f.u. data (refer to Section 5.3.3 for the detailed procedures for the serial dilution and plating).

Once the volume of culture required was established for each strain culture, the original cultures in the 50ml falcon tubes were vortexed and the tubes left to stand for 5 mins. During this time, the 24-well tissue culture plates containing activated THP-1 cells were observed under the inverted microscope to ensure that the cells had adhered to the bottom of the wells and that the RPMI medium was not contaminated. As there were 6 wells per strain, the appropriate volume of each of the 10 strain cultures was added to each of the 6 wells of the labelled 24-well plate aseptically. A drop of each culture was also plated onto CBA plates.

As well as infecting activated THP-1 cells with each of the ten different strains, three sets of controls were prepared. The first control was uninfected activated THP-1 cells. For this control, each of the 6 wells were inoculated with a volume of Middlebrook 7H9 medium (with OADC) equivalent to the largest volume of culture that was used for infecting THP-1 cells. This control enabled the identification of any effects that the medium may have on the THP-1 cells. The second control was set up by adding 25µl of 50µg/ml LPS from *Escherichia coli* 055:B5 (Sigma-Aldrich) to each of the 6 wells. The third control was prepared by adding 25µl of 50µg/ml IFN-γ (Miltenyi Biotech, Bergisch Gladbach, Germany) to each of the 6 wells. Using LPS and IFN-γ to activate THP-1 cells can potentially act as positive controls as they induce the production of certain cytokines and these controls are particularly important for measuring cytokine production using assays such as Enzyme-Linked ImmunoSorbent Assay (ELISA). All of the plates were incubated at 37°C, 5% CO_2, overnight, for infection of the THP-1 cells.

The following morning, day 1 post-infection, any unincorporated mycobacterial cells were removed by washing the infected THP-1 cells. The RPMI medium from each well was transferred into separate sterile microcentrifuge tubes. Each well was washed by adding 500µl 1× DPBS (prepared as described in Section 2.1.5) gently against the side of the well and then transferring the full volume into corresponding microcentrifuge tubes containing the old RPMI medium. The infected THP-1 cells in the wells were kept wet by gently adding 1ml 1× DPBS to each well.

Any THP-1 cells that may have become detached from the bottom of the well and aspirated were spun down by centrifugation at 300×*g* for 10 mins, (which avoids spinning down bacterial cells). The supernatant was discarded safely and the 1ml 1× DPBS was transferred from the 24-well plate and into corresponding microcentrifuge tubes. Again, THP-1 cells were spun down, during which time 1ml fresh complete RPMI was added to each of the wells. THP-1 pellets at the bottom of microcentrifuge tubes were resuspended by taking a small volume of RPMI, approximately 80µl, from the appropriate well and transferring the resuspended cells back in the well. The 24-well plates were incubated for 5 to 6 hrs, allowing the intracellular mycobacteria to elicit an effect, if any, before samples were taken for the day 1 time point.

2.6.4 Sample collection of mycobacteria infected THP-1 cells

Samples were taken at day 1, 4 and 7 post-infection. At every time point, for each strain and control, RPMI medium was transferred from two wells and into separate 1.5ml microcentrifuge tubes. All THP-1 and mycobacterial cells were spun down by centrifugation at 8000×*g* for 10 mins, during which time 300µl 1× DPBS was added to each of the wells from which RPMI had been removed. After centrifugation, the supernatant was transferred into separate microcentrifuge tubes, which were then stored at -80°C ready for ELISA.

The 300µl 1× DPBS was transferred from the wells and into the microcentrifuge tubes containing the pelletted cells. Addition of 100µl of 0.25% Trypsin, 0.25% (1×) with EDTA (Invitrogen Ltd.) for 3 mins allowed the THP-1 cells to detach from the bottom of the wells. The trypsin-EDTA was then inactivated by adding 500µl RPMI. The full volume was transferred into the appropriate microcentrifuge tubes and if some cells could still be seen at the bottom of the well, a small volume of liquid from the microcentrifuge tube was used to scrape the cells and aspirate them back into the microcentrifuge tubes. The cells in the microcentrifuge tubes were spun down by centrifugation, this time discarding the supernatant and centrifugation was repeated. The tubes containing the dried pellets of cells were stored at -80°C ready for analysis.

The experimental procedures described in Sections 2.6.1 to 2.6.4 were repeated two more times so there were supernatants and pelletted cells, stored at -80°C, from three independent experiments.

2.6.5 Measuring cytokine levels in supernatant

The levels of human TNF-α, human IL-10, human IL-1β, and human IL-6 cytokines were measured in all of the supernatants, using ELISA Ready-SET-Go! Kits (eBioscience, California, USA). The assay was performed using the protocols provided with the kits, which included the 96-well ELISA plates and all reagents except wash buffer, which was a solution of 1× DPBS (without calcium and magnesium) with 0.02% Tween-20 solution (Sigma-Aldrich).

In a single day, the assays for each of the 4 different cytokines could be performed using supernatant from one independent THP-1 infection experiment as the 96-well ELISA plate could accommodate these supernatants and all recommended standards.

The day before performing the assay, ELISA plates (one plate per cytokine) were coated with capture antibodies, which were diluted as directed in the protocols using Coating Buffer that was prepared as directed using the powdered buffer provided. Aliquots of 100µl diluted capture antibody were dispensed into each well of the plates, which were then sealed and incubated overnight at 4°C. If there was any cytokine in the supernatant that was complementary to the capture antibody it would specifically bind to it.

The following morning, supernatant from one of the three independent experiments was left at room temperature to thaw. The capture antibody was aspirated from all of the plates, which were washed 5 times with 300µl Wash Buffer per well, allowing 1 minute for soaking between washes to wash away excess capture antibody. All plate washes were performed using a plate washer (Wellwash 4 Mk2; Thermo Scientific). Afterwards, 200µl 1× Assay Diluent, prepared by diluting 5× Assay Diluent with distilled water, was added to each of the wells and all plates incubated at room temperature for 1 hour. The 1× Assay Diluent was aspirated from the wells and the plates washed as described previously.

The top concentration of standards, which were specific to each cytokine assay, were prepared by diluting the stock standard from the kit with 1× Assay Diluent using the volumes detailed in the protocol, as these vary between cytokine assay and also between kits for the same cytokine. For the standard curve, six 2-fold serial dilutions of the top standards were prepared so the final volume of diluted standards was 400µl. For each cytokine plate, there were seven standards and a blank, which was the 1× Assay Diluent.

One plate was set up at a time to minimise the time gap between the addition of standards and the last sample in the final well. Firstly, 100µl of the top standard was added to each of the top two wells of the first two columns, so there were duplicate wells for each standard. Two 100µl aliquots of the next dilution down from the top standard were added into the next two wells and so forth with the remaining standards. Aliquots of 1× Assay Diluent were added to the bottom two wells. In the remaining wells, 100µl of the supernatant was added with 1 well per sample, so all supernatants from one experiment could be analysed simultaneously. All plates were

sealed and incubated for 2 hrs at room temperature to allow any cytokine in the supernatant to bind to the capture antibody.

After incubation, standards and supernatants were aspirated from the wells and washed 5 times as described previously to leave only the bound cytokine molecules. These molecules were detected by adding 100µl cytokine specific Detection antibody, which was diluted in 1× Assay Diluent as directed in the protocol, to each of the wells. ELISA plates were sealed and incubated for 1 hour at room temperature. The Detection antibody was then aspirated and the plates washed as before.

A detection enzyme was added to each well to bind to the Detection antibody. For this assay Avidin-horseradish peroxidase (Avidin-HRP) was used and this reagent was the same for all cytokine assays. Avidin-HRP was diluted using 1× Assay Diluent and the volumes detailed in the protocol, after which 100µl was quickly added to each well, as the enzyme was light sensitive. The plates were sealed and incubated in the dark at room temperature for 30 mins. The Avidin-HRP was aspirated and washed as before except there were 7 washes instead of 5 to make sure any unbound Avidin-HRP was washed away completely.

A volume of 100µl Substrate solution was added to each well and the plate incubated in the dark at room temperature for 15 mins. As any bound Avidin-HRP reacted with substrate it caused a colour change from colourless to blue if the cytokine was present. The extent to which there was colour development showed how much cytokine there was present initially. As the reaction between the Avidin-HRP and substrate is light sensitive, as soon as the 15 mins incubation was over, 50µl Stop Solution was added to each well to stop the reaction and to avoid false readings. The Stop solution caused a colour change from blue to yellow. The OD at 450nm (OD_{450}) was measured for each well in all cytokine plates using a Multiskan Microplate Reader (Thermo Scientific).

The logarithm (log) of OD_{450} values was calculated. The mean log OD_{450} value was taken for each of the standards, as there were duplicate readings, to produce the standard curve of log OD_{450} value of standards versus the corresponding log concentration values, which were calculated using the top standard concentration, in picogram/millilitre (pg/ml), provided in the kit's protocol. The top standard

concentrations varied between cytokine assays and between kits. The log cytokine concentration in the supernatants was established by extrapolation against the standard curve and then the actual concentration calculated using $10^{\log \text{ concentration}}$. During the initial experiment, when samples were taken at each time point, there were duplicate samples per time point. Therefore, the mean cytokine concentration was taken. The cytokine concentrations for each of the ten strains and the three controls were plotted against days post infection to identify any trends.

The ELISA assays were repeated using the same protocol for supernatant collected from the two repeated independent infection experiments described in Section 2.6.4. If there were any supernatants in which the calculated cytokine concentration was above the concentration of the top standard, the ELISA had to be repeated as the standard curve was only valid for the concentration stated in the protocol and any OD_{450} values above this could not be reliably extrapolated. The supernatants were diluted 1 in 4 using complete RPMI as the diluent before being added to the ELISA plate well. When calculating the cytokine concentration, the dilution factor was taken into account.

2.6.6 Measuring fold enhancement and growth of mycobacteria in THP-1 cells

Real-time PCR was used to calculate the growth of mycobacteria inside THP-1 cells at the three different time points. Firstly, DNA was extracted from the dried pellets of cells that were prepared during sample collection. The DNA extraction protocol differed from previous protocols, as DNA was extracted from both THP-1 and mycobacterial cells that had initially been taken up by THP-1 cells.

The pellets of cells were resuspended in 200µl 1× DPBS by pipetting up and down gently after which 20µl proteinase K, 22µl 100mg/ml lysozyme and 200µl Buffer AL was added to the cells and mixed thoroughly by vortexing. The tubes were incubated in a 56^{o}C waterbath for 15 mins, after which 200µl ethanol (96-100%) was added to each tube and mixed thoroughly to make a homogenous solution. The DNA purification steps were performed using the Qiagen DNeasy Blood & Tissue Kit and the procedures described in Section 5.2.1. The DNA was stored at -20°C ready for real-time PCR.

Real-time PCR reactions were set up using the volumes detailed in Section 5.2.2. The set up was slightly different because no DNA standards were used. For each DNA extract two sets of three reactions were set up, one set of reactions was prepared using *rpoB* primers and the other set of reactions was set up using primers for actin (forward: 5'-TCA CCC ACA CTG TGC CCA TCT ACG A-3'; reverse: 5'-CAG CGG AAC CGC TCA TTG CCA ATG G-3'). Setting up reactions with these primers meant that the threshold cycle (Ct) values could be used to calculate the amount of mycobacteria relative to the THP-1 cells present as it is not known to what extent the THP-1 cells would survive after infection with mycobacteria. However, it is important to note that as DNA levels are being assayed there would be no discrimination between viable and dead mycobacteria and also between dead or living THP-1 cells.

After real-time PCR, the fold enhancement relative to day 1, for each strain, was calculated using the $2^{-\Delta\Delta Ct}$ mathematical model (Pfaffl 2001):

i. calculating ΔCt for each time point

$$= Ct (rpoB) - Ct (actin)$$

ii. calculating ΔΔCt between day 1 and all other time points

$$= \Delta Ct (day\ 4) - \Delta Ct (day\ 1) \quad \text{and}$$

$$= \Delta Ct (day\ 7) - \Delta Ct (day\ 1)$$

iii. calculating fold enhancement (the efficiency of both pairs of the primers is 2 so this value was used for calculations) for each ΔΔCt

$$= 2^{-[\Delta\Delta Ct(dy4 - dy1)]} \text{ and}$$

$$= 2^{-[\Delta\Delta Ct(dy7 - dy1)]}$$

The fold enhancement at day 4 and 7 for each strain was plotted as a bar graph to make comparisons of the fold enhancements between strains.

2.7 *IN VITRO* GROWTH EXPERIMENTS

2.7.1 Inoculation of MGIT tubes

As well as studying the growth of mycobacteria in tissue culture systems, the growth of the ten strains was investigated using another *in vitro* system, which allowed real-time observation of growth curves. The MTBC cultures that were used for infecting activated macrophage-like THP-1 cells, had already been quantified using real-time PCR, therefore the BBL Mycobacteria Growth Indicator Tube (MGIT) 7ml tubes (Becton, Dickinson and Company) were inoculated with the same cultures on the

same day as the tissue culture experiments were set up. The MGIT tubes contained 7ml modified Middlebrook 7H9 medium and, before inoculation, 800µl OADC was added to each of the tubes required for each experiment.

As there were no previous data on the number of mycobacteria that should be used for inoculation, MGIT tubes were set up with 10 fold differences in the number of mycobacteria, with the highest number being 600,000 genomes and the lowest 6 genomes. In order to inoculate a MGIT tube with 600,000 mycobacteria, two times the volume of culture required per well for the tissue culture experiments was added to the tube. The same volume of culture was then used to complete a set of five 10-fold serial dilutions, before inoculating MGIT tubes with 60,000, 6,000, 600, 60 and 6 mycobacteria. For each strain, two sets of six MGIT tubes were inoculated with the different number of mycobacteria so there were duplicate sets of MGIT tubes. Negative controls were set up in exactly the same way, using the same Middlebrook 7H9 medium that had been used to inoculate the non-infected THP-1 cells.

After inoculation, MGIT tubes were registered onto the EpiCenter that was the dedicated computer linked to the BACTEC MGIT 960 Mycobacterial Detection System (Becton, Dickinson and Company), which was an automated system into which the tubes were placed for incubation. The temperature inside the MGIT 960 was controlled and maintained at 37^{o}C to allow mycobacteria to grow. Fixed to the bottom of each tube was silicone, embedded with a fluorescent growth indicator, which was sensitive to dissolved oxygen in the medium. As the mycobacteria grew during incubation in the MGIT 960, the oxygen in the tubes was utilised and this oxygen depletion caused an increase in fluorescence. Every hour, the MGIT 960 took fluorescence readings from all of the tubes simultaneously. The algorithm within the software (TB eXiST version 5.53) automatically calculated the growth units in accordance to the detected fluorescence. Real-time growth curves displaying growth units versus time, in hrs, were automatically generated using the TB eXiST software.

2.7.2 Analysis of growth curves

Once all of the growth curves had reached a plateau, the raw growth unit data for all of the strains and negative controls, at each of the different inoculation sizes, was exported into Microsoft Excel, where growth curves were plotted again. These plots were used to identify the two time points between which the growth curve was at

logarithmic phase (i.e. the time point between which the diagonal line of the curve was the straightest). Once these time points had been established, the raw data for the relevant curve were used to find the exact growth unit at the two time points.

The growth rate at mid-logarithmic phase was then calculated by dividing the difference in growth units at each time point by the difference between the two time points. In each experiment, for every strain, the mean growth rate of the duplicate values at the different inoculum sizes was calculated.

2.8 AEROBIOLOGY STUDIES

2.8.1 Preparation of cultures

The stability of the core five Panel B *M. tuberculosis* strains as circulating aerosols was tested using an all-glass impinger-30 (AGI-30) sampling device. Cultures with 1×10^7 c.f.u. were required. One frozen aliquot of archived culture for *M. tuberculosis* Beijing, *M. tuberculosis* LAM10, two *M. tuberculosis* CAS, and *M. tuberculosis* EAI5 was thawed at room temperature and the full volume used to inoculate a CBA plate to check for purity, and 50ml falcon tubes, containing 15ml Middlebrook 7H9 medium with 0.05% Tween-80 media and glass beads. After vortexing, the cultures and CBA plates were incubated at 37°C. Every day the plates were checked for growth and the cultures were vortexed to break up the clumps. After a week, subcultures were prepared by inoculating fresh Middlebrook medium and CBA plates (to check for purity) and incubated for 10 days to ensure that the bacteria were in the exponential phase of growth.

Instead of using c.f.u., quantification of mycobacteria in each culture was performed using real-time PCR procedures as previously described in Section 2.6.3. For each strain, two aliquots of the volume of culture required to obtain 1×10^7 mycobacterial genomes were transferred into 2.0ml microcentrifuge tubes and the bacterial cells pelletted by centrifugation at 12,000×*g* for 10 mins to remove media containing Tween. Tween is a detergent and may affect the aerosolisation of the cultures. For this reason, bacterial cells were resuspended in 1.5ml Middlebrook 7H9 media without Tween and without OADC, as the cultures were used immediately. All culturing procedures were performed at HPA, MRU.

2.8.2 AGI-30 sampling

The AGI-30 sampling was performed in collaboration with HPA, CEPR, based in Porton Down, UK, and the procedures for actual AGI-30 sampling were performed by Simon Clark and colleagues in CEPR using a Class III flexible film isolator (Bell Isolation Systems Ltd, Scotland, UK).

For each strain, the full volume of culture, was diluted 1 in 10 using sterile distilled water, so the final 15ml diluted culture contained 1×10^6 genomes. The cultures were aerosolized using a 3-jet collison nebulizer (BGI, Incorporated, Massachusetts**,** USA) and then fed through the Henderson apparatus (CEPR), through which air was circulated. The diameter of the aerosol particles was on average 2.0μm, with a range of 0.5-7.0μm. The aerosols traveled through the apparatus and whilst the Henderson apparatus is usually used to deliver aerosols to animals, for the purposes of this experiment the aerosols were delivered to four AGI-30 samplers (Ace Glass, Incorporated, New Jersey**,** USA), which were effectively glass vessels, containing 5ml sterile distilled water, that draws air containing TB bacilli from the Henderson apparatus via an inlet and then captures the aerosols in the water. Each strain was tested separately and the apparatus was washed through with sterile distilled water between strains. Comparing the numbers of mycobacteria in the collison and the AGI-30 samplers for each strain would allow to identify any differences in the stability of strains following aerosolisation.

For safety reasons, all further procedures were conducted at the HPA, CEPR facilities. Therefore, to ensure all quantification was performed using the same method, colleagues at CEPR and I plated out mycobacteria from the collison and AGI-30 samplers, a second aliquot of culture that had been prepared in Section 2.8.1 was also plated onto Middlebrook 7H11 agar plates (Biomerieux UK Ltd., Hampshire, UK). For all collison and AGI-30 samples, 100μl of neat and diluted samples were plated; two 10-fold serial dilutions in duplicate were prepared. All plates were incubated at 37°C for four weeks after which the c.f.u. was counted. The number of c.f.u. per ml was calculated by multiplying the c.f.u. count by 10 (as 100μl was plated) and this value was multiplied by the dilution factor.

2.9 *IN VIVO* GUINEA PIG EXPERIMENTS

2.9.1 Infecting guinea pigs with TB aerosol

The phenotypes of core Panel B *M. tuberculosis* strains, were tested using an *in vivo* model of guinea pigs. Similar experiments had been performed previously at HPA, CEPR, using laboratory-adapted strains, so it was of particular interest to compare the effects of infecting guinea pigs with wild-type strains. Simon Clark and colleagues from CEPR carried out the procedures for infection of guinea pigs.

Sixteen female Dunkin-Hartley guinea pigs, weighing between 650g and 750g (Harlan Laboratories, Inc., Loughborough, UK) were infected with each strain. For infection, the diluted cultures that had been used for aerobiology testing were used again. The AGI-30 sampler components were exchanged for a sow, which accommodated 8 guinea pigs per challenge. This meant there were 2 separate challenges per strain and between each challenge a sample of the collison was taken for c.f.u. counts to monitor the numbers of mycobacteria throughout the challenge.

The diluted culture was aerosolised in the collison nebulizer and as the aerosols travelled through the Henderson apparatus, they were delivered straight to the snout of the guinea pigs. Guinea pigs were challenged for 5 mins under the controlled conditions of the apparatus; meaning approximately 10 mycobacteria should be retained in the lungs of the guinea pigs. After 16 guinea pigs had been infected with one strain, the apparatus was washed through with sterile distilled water before another 16 animals were infected with the second strain and so forth.

After infection, all animals were kept under controlled and contained conditions. Their health was monitored and all guinea pigs were weighed every 5 days. Post-challenge, guinea pigs were culled after 16 days post-challenge (8 guinea pigs per strain), and then 56 days post-challenge. All animals were euthanised with overdoses of sodium pentabarbitone.

2.9.2 Tissue preparation for viable c.f.u. counts

After culling, the lung and spleen were aseptically dissected. These procedures were performed by Simon and colleagues at CEPR. All other lungs and spleens were stored at -20°C immediately ready for processing, but day 56 lungs and spleens were weighed, in grams, and a small segment of the organs was removed prior to storage,

for histopathology. A rotating blade macerator system (MSE homogeniser; Ystral GmbH, Ballrechten-Dottingen, Germany) was used to homogenise lung tissue in 10ml sterile distilled water and spleen tissue in 5ml water, and colleagues at CEPR and I plated out 100µl aliquots of neat homogenized lung and spleen tissue onto Middlebrook 7H11 plates. Appropriate dilutions were plated for the day 16 and 56 lung and spleen tissues. All plates were inoculated in duplicate. Plates were incubated at 37°C for four weeks after which the c.f.u. was counted. The number of c.f.u. per ml was calculated as described in Section 5.3.3.

2.9.3 Tissue preparation for histopathology

For histopathology, all slide preparation procedures were performed by colleagues at CEPR. Day 56 lung and spleen tissue segments, which had been removed before storing the rest of the tissue at -20°C, were fixed in 10% volume/volume (v/v) formalin (Surgipath Europe Ltd., Peterborough, UK). A $1cm^3$ section was taken from one end of each spleen and the top left, top right and bottom right lobe sections from each set of lungs were used for histopathological analysis. Tissue samples were processed to paraffin wax after which 5µm sections for all lung and spleen tissues were stained for routine examination with haematoxylin and eosin; encapsulated and calcified lesions were detected using van Giesen and Alizarin Red staining, respectively. All stains were provided by Surgipath Europe Ltd.

The examination of slides was blinded and the extent to which lesions were present was objectively evaluated semi-quantitatively by a CEPR board-certified veterinary pathologist, Dr Graham Hall. The slides for all lung and spleen tissues were scored according to the size of the lesions present and the degree of consolidation. Consolidation was observed by microscopy and showed a specific change in the morphology of the lesion. The greater the number of areas of consolidation, the higher the score. The criteria for scoring were as detailed in Table 2.2. For tissues in which lesions were detected, the degree to which lymphocytes had infiltrated these lesions was also subjectively scored. Numerical data were obtained by recording the number of encapsulated lesions, the number of foci of caseation and the number of calcified lesions.

Table 2.2 *The criteria used to score the histopathology slides taken from guinea pig lung and spleen after infection with M. tuberculosis.*

Score	Lesion size	Degree of consolidation
0	lesions not detected	
1	very few or very small	0-10 %
2	few or small	10-20 %
3	medium sized	20-33 %
4	moderate sized; areas of pneumonia	33-50 %
5	large; moderately extensive pneumonia	50-80 %
6	extensive pneumonia	> 80 %

CHAPTER 3

MOLECULAR TYPING, EPIDEMIOLOGY, PHYLOGENETIC ANALYSIS & SELECTION OF STRAINS FOR PHENOTYPIC EXPERIMENTS

3.1 INTRODUCTION

In 2007 there were 8417 reported cases of tuberculosis in the UK, with 92% of cases being reported in England alone (HPA 2008). London accounted for 39% of cases in England and 72% of TB patients in London were born outside of the UK, mainly in the Indian sub-continent and sub-Saharan Africa countries, with the highest TB rates being seen amongst the ethnic Pakistani, Indian and Bangladeshi groups and black Africans (HPA 2008).

According to an earlier study by Dale *et al*., 2005, in London using IS*6110* RFLP and spoligotyping data, there are clear differences in the distribution of MTBC families amidst patients from different countries (Dale *et al.,* 2005). For instance, in patients originating from East Africa and East Asia, the *M. tuberculosis* African and Beijing strains were most common and strains observed in Indian and Pakistani patients were markedly different from those observed in patients from Bangladesh. Despite the noticeable differences in the *M. tuberculosis* strains, which seem to predominantly infect specific ethnic communities in London, previous studies revealed limited transmission from infected immigrants to individuals from other communities in the London population (Dale *et al.,* 2005). It is hypothesised that higher transmission had occurred either in the country from the immigrant country of origin before entering London or within the communities where immigrants are in very close proximity to each other causing dissemination of particular strains within specific communities.

The main aim of the present study was to construct a preliminary panel of strains (Panel A) representative of the global TB population which would span a broad spectrum of *M. tuberculosis* strains that would be useful for future vaccine evaluation studies and general TB research. In addition to designing an algorithm to select *M. tuberculosis* strains for the preliminary panel, the present study also aimed to test the null hypothesis that *M. tuberculosis* strains are phenotypically similar to each other using a smaller panel of strains (Panel B derived from the preliminary panel A) to

represent the six MTBC phylogenetic lineages defined by Baker *et al*., 2004 and Gagneux *et al*., 2006. This chapter focuses on the selection process of strains for the preliminary panel A and then goes into detail to describe the rationale used to select the smaller panel of strains for Panel B.

3.2 MOLECULAR TYPING & EPIDEMIOLOGY

3.2.1 MIRU-VNTR typing

All cultures received by HPA, MRU between 1st April 2005 and 31st March 2006 from 30 hospitals and/or laboratories situated within the M25 boundary of London, and which had been identified as belonging to *M. tuberculosis* complex, were included for initial MIRU-VNTR typing. Typing, which included the analysis of the 15 MIRU and 3 ETR loci, was performed on crude DNA extracts from 2363 isolates, which all fitted the above criteria for selection.

The 15 loci were amplified in each of the extracts, and molecular weights of amplified fragments were calculated as described in Section 2.2. Molecular weight data were displayed as a raw data trace, in which red peaks represented the size standard and blue, black or green peaks represented molecular weights of different loci according to the dye with which they were labelled during amplification. Figure 3.1a shows a raw data trace, as displayed by the Genetic Analysis System, from capillary A in which MIRU-2, -4, -10, -16, -23, -24 were analysed (refer to Section 2.2.3). Raw data were automatically analysed to produce a second trace, in which each peak was annotated with the locus and number of repeats corresponding to the detected molecular weight. The analysed data trace in Figure 3.1b corresponds to the raw data trace in Figure 3.1a.

For some isolates, if after capillary gel electrophoresis no amplified product was detected and peaks were absent for certain loci; these loci were re-amplified individually and molecular weights identified manually by resolving products against a DNA ladder using agarose gel electrophoresis. The number of repeats was calculated using the calling tables in Appendix 1 and 2.

The methodology was modified for certain loci: during capillary gel electrophoresis, there were missing peaks for ETR-A possibly because fragment sizes exceeded 640

b.p., which was the highest molecular weight of the size standard. Whilst extending the separation time for electrophoresis meant higher molecular weight peaks were included in traces, a locus specific custom size ladder was prepared in order to make analysis easier using agarose gel electrophoresis for those extracts in which ETR-A peaks were absent. A custom size ladder for MIRU-4 was also prepared because during data analysis, there were some MIRU-4 fragments with molecular weights that were lower than expected, indicating deletions at this locus.

Extracts that had been successfully analysed by 15 MIRU-VNTR typing were used to select extracts that would be used to make locus specific custom size standards. For MIRU-4, seven extracts containing repeats ranging from 1 to 9, three extracts with 1, 2 and 3 repeats and a 49 b.p. deletion, and an extract with 5 repeats with a 50 b.p. deletion were selected. Seven extracts containing repeats in ETR-A ranging from 1 to 9 were selected. Loci were amplified and products were resolved on gels to ensure PCR product was present. PCR products were pooled to produce three custom size ladders, ETR-A, MIRU-4 and MIRU-4 with deletions. The ladders were used if it was difficult to determine the number of repeats when using commercial Hyperladders.

All 15 MIRU-VNTR profiles were collated into a Microsoft Access database that was linked to BioNumerics version 3.00, which contained the program for cluster analysis. Isolates were analysed blinded and for 265 patients there were multiple isolates. Only one profile per patient was included, to avoid bias during analysis, by removing profiles of serial isolates (i.e. isolates taken from the same patient more than once). The profiles for serial isolates were included as part of another smaller study to investigate the reproducibility of particular loci (see Chapter 4). Similarly, during capillary gel electrophoresis, some extracts had been included more than once for quality control purposes. These profiles were also removed. A total of 102 profiles were removed to leave 2261 profiles for cluster analysis. From these 2261 profiles, there was complete data for 2046 isolates.

The dendrogram produced after performing the cluster analysis is included as supplementary data on the compact disc. The dendrogram was produced using the profiles from 12 MIRU, 3 ETR and 7 VNTR typing (NOTE: MIRU-4 and -31 are referred to as ETR-D and –E in the dendrogram). Also included are the octal codes of isolates after spoligotyping along with the family identification. The dendrogram

revealed that 1225 profiles were placed into one of 235 clusters, whose cluster sizes varied from 2 to 53 profiles. There were 1036 profiles that did not fall into any cluster and were termed unique isolates. All isolates with 15 MIRU-VNTR profiles that fell into clusters were further analysed using the more discriminatory loci, including, VNTR-2163B, -2347, -3232, -2163A, -1982, -3336 and -4052 (see Section 4.1.5). The dendrogram in the supplementary data disc includes the results of cluster analysis using the profiles from 12 MIRU and 3 ETR typing and also from the 7 VNTR typing.

3.2.2 Spoligotyping

The crude DNA extracts from the 2363 isolates were subjected to spoligotyping and the resulting octal codes used to identify which family isolates of *M. tuberculosis* belonged to, as described in Section 2.2.7. Figure 3.2 shows a scanned image of a film after spoligotyping. Altogether, there were complete spoligotyping profiles for 2233 isolates (94.5% of the total). From these profiles, 656 different spoligotypes were observed, and included 198 clustered spoligotypes, with cluster sizes varying from 2 to 221 profiles, and 458 unique spoligotyping profiles.

To identify which family the octal codes belonged to, profiles were entered into the online SPOTCLUST program, which was based on the SpolDB3 model. This model included profiles for *M. africanum*, *M. bovis*-BCG, *M. microti*, and nine main *M. tuberculosis* families. Within some *M. tuberculosis* family groups there were sub-families denoted with a number after the family name. In total, there were 32 families in the SpolDB3 database and four additional groups, Family 33-36. Figure 3.2 shows a scanned image of a spoligotyping membrane with the corresponding family designations for spoligotyping profiles. From the 2233 profiles, all, except 4, were assigned to one or more of the families, with a probability greater than 0.5. From the assigned profiles, 88.4% were assigned to a single family with a probability of more than 0.9. Remaining profiles could not necessarily be assigned to a single spoligotyping family i.e. there was ambiguity leading to their assignment to more than one family with probabilities ranging from 0.5 to 0.9; 88% (n = 231) of these profiles were assigned to one of the following four groups: i) all profiles identified as *M. tuberculosis* Haarlem 3 all had secondary designations, ii) all strains assigned to *M. tuberculosis* LAM1 were also assigned to LAM9, iii) all profiles assigned to *M. tuberculosis* S family, except for 1, had secondary designations and iv) all profiles

assigned to *M. tuberculosis* X, except 1, were assigned to *M. tuberculosis* Tuscany-1 (T1).

Figure 3.3a shows a bar graph of the numbers of profiles that were assigned to each of the major MTBC families. There were no *M. microti* in this study population. From Figure 3.3a, it could be seen that the most prevalent family was the CAS family, which accounted for 24.7% of the total number of DNA extracts that were genotyped. Other major families included T, EAI, LAM and Haarlem. From those extracts identified as belonging to the T family, the majority, 84.7%, belonged to the T1 sub-family, whilst 57.6% of isolates identified as belonging to EAI family were assigned to the EAI5 sub-family. From isolates identified as belonging to the LAM families, the predominant sub-family was LAM10 and the most predominant sub-family in the Haarlem family was Haarlem3. There was a lower prevalence of *M. tuberculosis* Beijing and X families, comprising 5.7% and 4.6% of the total number of isolates, respectively, and even fewer members of other families including *M. tuberculosis* H37Rv, S, *M. bovis*-BCG, *M. africanum* and the additional unclassified families 33-36.

3.2.3 Epidemiology

There was country of birth data for 1381 patients, which constituted 61% of all patients included in the study. Apart from the UK, patients were born in 89 different countries, which were categorised into global regions, as defined by the WHO. Figure 3.3b shows the proportions of patients that were born in each region. Altogether, there were 1157 patients born outside of the UK; one group of 501 patients was born in the African countries, with Somalia contributing the majority of cases (n = 179), and 463 patients were born in the Indian subcontinent, with the majority of patients being born in India (n = 247). The remaining migrant TB patients were from Western Africa, Eastern and South East Asia, South-Central Asia, Middle Africa, Western, Northern and Southern Europe, and Southern Africa (7.3%, 5.0%, 2.9%, 2.6%, 2.5%, and 2.4%, respectively). Each of the other global regions contributed less than 2.0%.

The geographical distribution of the major *M. tuberculosis* complex families, defined by spoligotyping, was investigated and the results are illustrated in Figure 3.4. From this diagram, geographical associations were seen between families such as *M. tuberculosis* CAS, EAI, and Beijing, and particular global regions. The patients born in the Indian subcontinent were infected mainly with *M. tuberculosis* CAS and EAI

families. The CAS strains isolated from patients born in the Indian subcontinent comprised approximately 60% of all CAS strains.

From the 58 patients born in Eastern and South-East Asia, the majority were infected with *M. tuberculosis* Beijing (27 patients) and EAI (21 patients) families. The EAI family was also primarily found in immigrants to the UK originally born in the Indian Subcontinent and Eastern Africa. Isolates that had been identified as belonging to the *M. africanum* family were predominantly isolated from patients born in the Western African countries.

There was no specific association of isolates identified as belonging to the *M. tuberculosis* T family and a particular global region as 40, 67 and 48 isolates (from the 366 isolates identified as *M. tuberculosis* T) were isolated from patients originally born in the UK, Eastern Africa and the Indian Subcontinent, respectively. There were fewer *M. tuberculosis* T isolates from patients born in South Africa (4 isolates), Western Asia (5 isolates), Northern Africa (2 isolates) and South-Central Asia (4 isolates). There was underrepresentation of *M. tuberculosis* LAM and Haarlem families, first in the Indian Subcontinent with 4.5% and 3.4%, respectively, and secondly in South-East Asia comprising 5.4% and 6.9%, respectively.

3.2.4 Selection of strains for preliminary Panel A

As it was not possible to perform further phenotypic experiments on all 2363 isolates, a series of analyses were conducted to create preliminary panels of strains resulting in a final panel of MTBC strains that would be as representative of the global TB strain population as possible and against which any vaccine should prove effective. For this, we examined a number of genotypic and phenotypic criteria, selecting strains initially based on:

1) frequency in an unselected metropolitan strain family population setting (London, UK), with an ethnically diverse population, with representatives from most of the global regions (refer to Section 1.4)
2) supplemented with more infectious strains as demonstrated by greater success e.g. higher clustering rates, in unselected regional/national settings in London UK, Samara Region (Russia) and Estonia (Kruuner *et al.,* 2001; Ruddy *et al.,* 2004; Drobniewski *et al.,* 2005), and strains infecting individuals outside their typical ethnic groups. Our logic was that strains found commonly in a particular

population, e.g. Somalis, would tend to spread amongst other Somalis in their country of origin and where Somalis develop active TB in the UK any transmission would be more likely to spread within their own ethnic groups and so to other Somalis. Where such strains are seen to spill over into the host or other non-Somali migrant communities (e.g. into an East European patient) it may be more likely to be the consequence of more transient exposure and suggest that the strain might be more infectious

3) strains were added from our population-based phylogenetic analysis (refer to Section 3.3) to ensure that all the globally-identified phylogenetic groups were fully represented
4) strains were added if phylogenetic analyses suggested that strains were of enhanced virulence compared to standard laboratory strains or clinical isolates.

An initial panel (A) reflecting the global population of strain families was selected to include a broad spectrum of *M. tuberculosis* strains. Firstly, in order to gain a better understanding of the most common profiles, cluster analysis was performed on all isolates (one isolate per patient) using only the MIRU and ETR profiles as described in Section 3.2.1 giving 235 clusters altogether, with cluster size varying from 2 to 53 profiles (see supplementary data on disc). In clusters containing 9 isolates and less, the country of birth data was more limited and in some clusters there was only data for one or two of the patients, meaning no correlations would be identified when applying the algorithm in Figure 2.1. Following the application of this cluster inclusion criterion, there were 24 clusters from the initial 235 clusters containing more than 9 profiles.

Using spoligotyping data, the proportion of each cluster accounted for by different strain families was determined. Many clusters were dominated by one spoligotype family. The proportions of isolates, in each cluster, with the family assignations are detailed in Table 3.1. In 21 of the 24 clusters over 80% of isolates belonged to a single *M. tuberculosis* spoligofamily. From these 21 clusters, there were 11 clusters in which all isolates came from a single strain family strongly supporting a common transmission among the individuals in the group at some time point. Five of the 24 clusters, contained sub-families of the same family (e.g. cluster 18 comprised of T family strains, but there were sub-families of T1 and T3). For these 5 clusters containing sub-families, the design of the selection algorithm would be different to

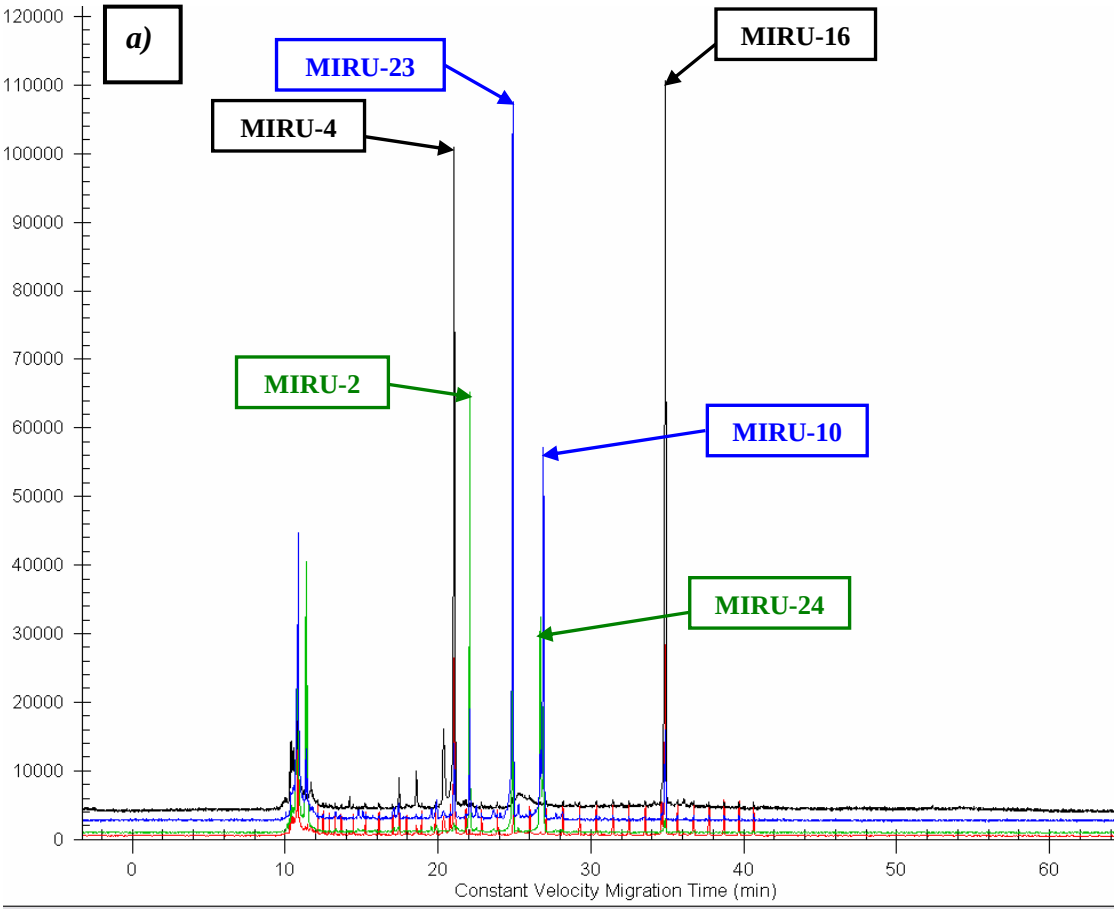
that in Figure 2.1 because in order to have the ideal representative panel of strains, isolates from each sub-family would need to be selected. The algorithm in Figure 2.1 was designed specifically to select a representative panel of strains from clusters that did not contain isolates from sub-families.

An algorithm was constructed to select strains of interest from the 19 clusters (refer to Figure 2.1). The algorithm was designed to help select strains based on correlations between the most common family in a cluster, the most common country of birth and the most common sub-cluster, which was assigned using the 7 VNTR profiles within the initial clusters. The common patterns that were considered were family designations and country of birth within a cluster.

Using the designed selection criteria, 42 strains of interest were selected from 16 of the 19 clusters (refer to Table 3.2). It was not possible to select strains from MRU clusters 8 and 38 as the initial cluster using 15 MIRU and ETR profiles were split by subsequent analysis using additional 7 VNTR profiles. When looking at the full 22 VNTR profiles for isolates in Cluster 8 and 38, 53.1% and 46.2% of isolates had different profiles to other isolates in their cluster, respectively, and were unique. This meant there were no correlations between the most common country of birth and 22 VNTR cluster designations in Cluster 8 and 38. Even though initially clusters with more than 9 isolates were included, there was country of birth data for only 1 of the 11 isolates in Cluster 148, which meant an isolate from this cluster could not be selected. Cluster 28 had to be included for analysis as there was country of birth data for 50% of the 10 isolates in this cluster, which had all been identified as *M. tuberculosis* X2. These 42 strains formed the first step toward a globally representative panel (Panel A).

The 42 strains selected for Panel A included 10 different families from patients born in 17 different countries. Two of the larger groups included in the panel were *M. tuberculosis* CAS (14 strains isolated from patients born in 8 countries) and Beijing (9 strains also isolated from patients born in 8 countries although different countries to the CAS strains). Strains belonging to *M. tuberculosis* Haarlem or S families, *M. africanum* or Families 33-36 were not included in Panel A because clusters containing these families did not fit the initial criteria for inclusion.

Figure 3.1 *(a) A raw data trace produced by the Beckman Coulter CEQ 8000 Genetic Analysis System, for capillary A, in which MIRU-2, -4, -10, -16, -23, -24 loci (identified by the labels) were analysed, and (b) a corresponding trace of analysed data in which the corresponding peaks to Figure 3.1a are labelled and which were automatically annotated with the molecular weight, locus name and number of repeats.*
(Red peaks represent the size standard)



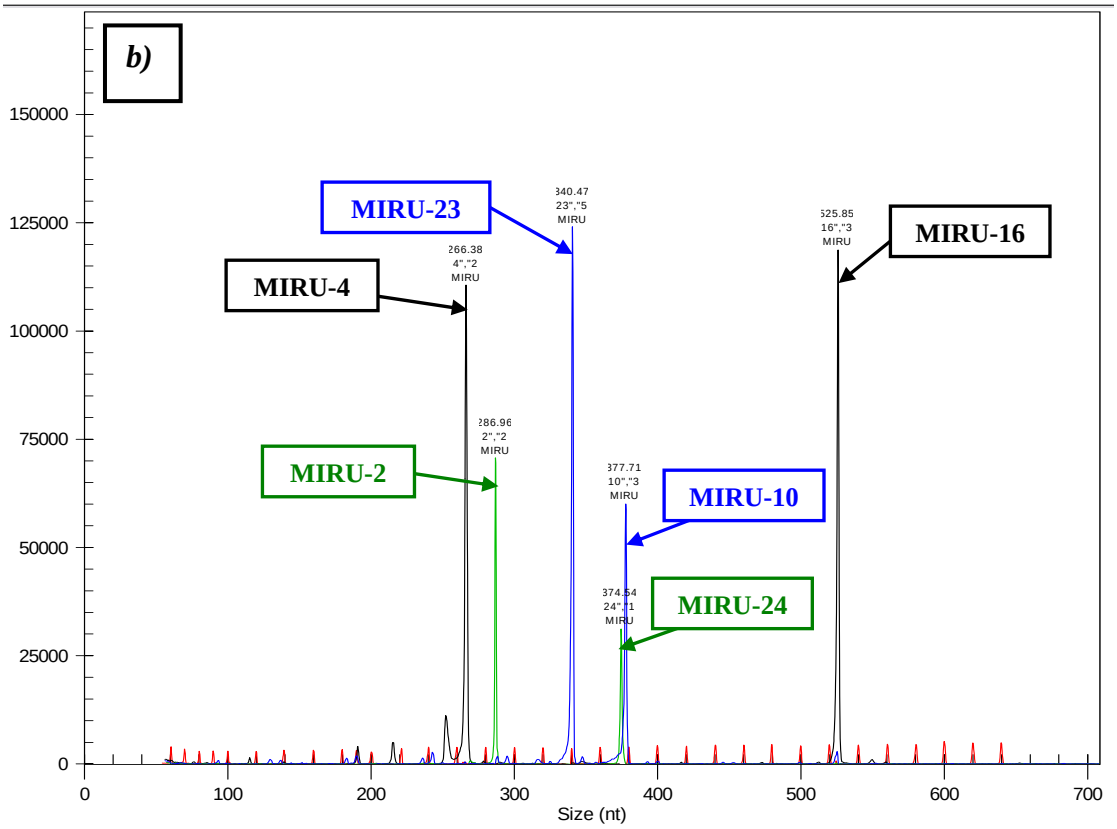


Figure 3.2 *A scanned image showing spoligotyping profiles that were obtained for the controls (profile 1 and 2) and 38 of the 2363 extracts (profiles 3 to 40). Absence and presence of spacers were displayed as empty spaces (binary code 0) and black squares (binary code 1), respectively. After converting binary codes to octal codes, the M. tuberculosis family for each profile was identified using an online SPOTCLUST program. Below are examples of some of the identifications that were made.*

Profile 1	*M. tuberculosis* H37Rv control
Profile 2	*M. bovis* BCG P3 control
Profile 3, 7	*M. tuberculosis* EAI5
Profile 6, 17, 20	*M. tuberculosis* EAI3
Profile 8, 10, 11, 13, 15	*M. tuberculosis* CAS
Profile 12	*M. tuberculosis* EAI4
Profile 14	*M. tuberculosis* LAM9
Profile 16, 37	*M. tuberculosis* Beijing
Profile 18, 19, 22	*M. tuberculosis* T1
Profile 21	*M. tuberculosis* LAM10

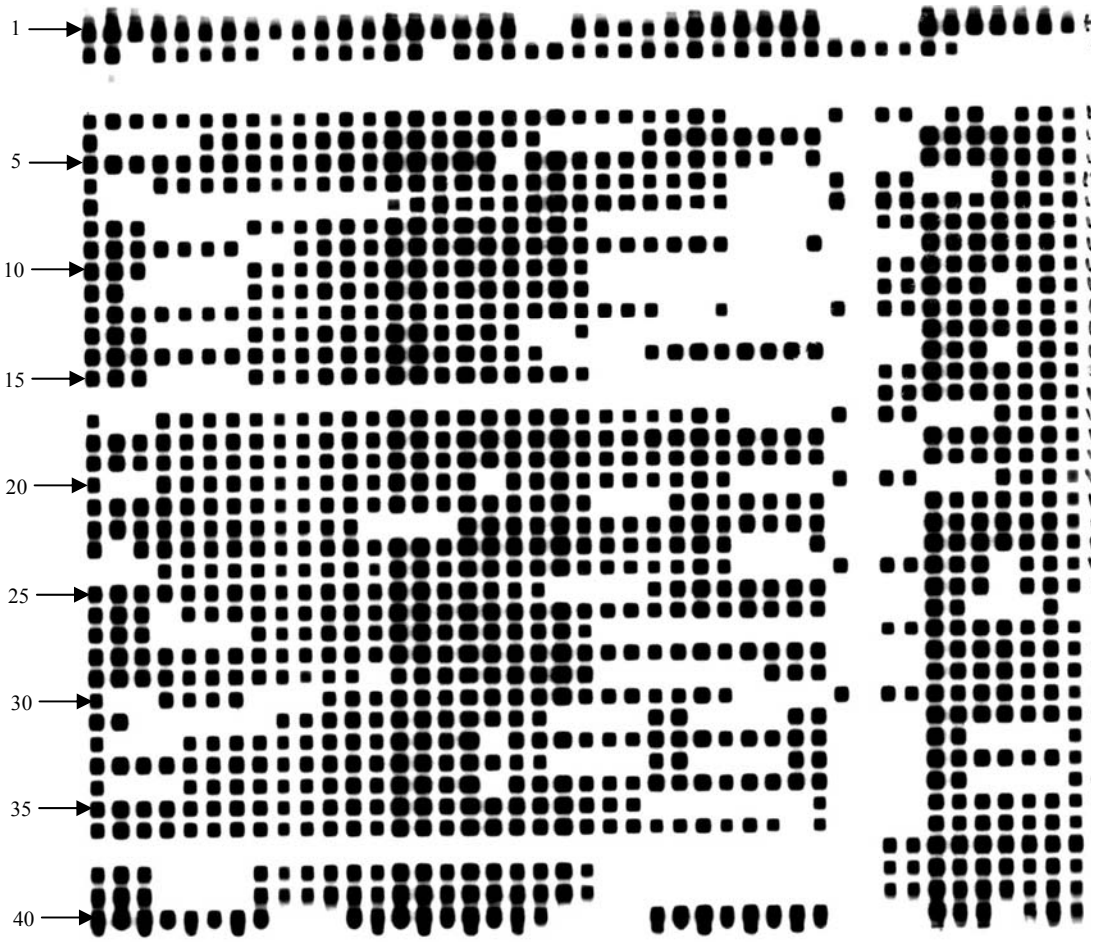
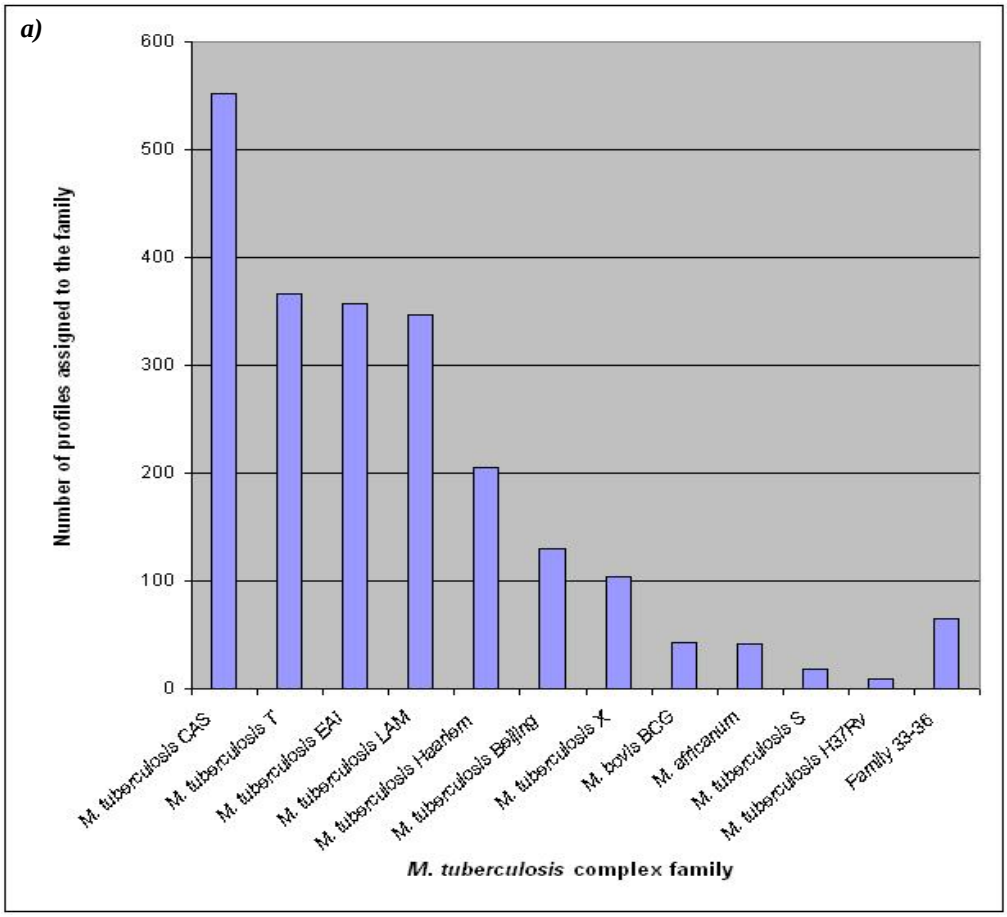


Figure 3.3 *(a) Histogram showing the number of spoligotyping profiles that were assigned to each of the main families; the numbers for M. tuberculosis T, Haarlem, X, EAI and LAM includes all sub-families, and (b) the proportion of patients born in global regions defined by the World Health Organisation.*



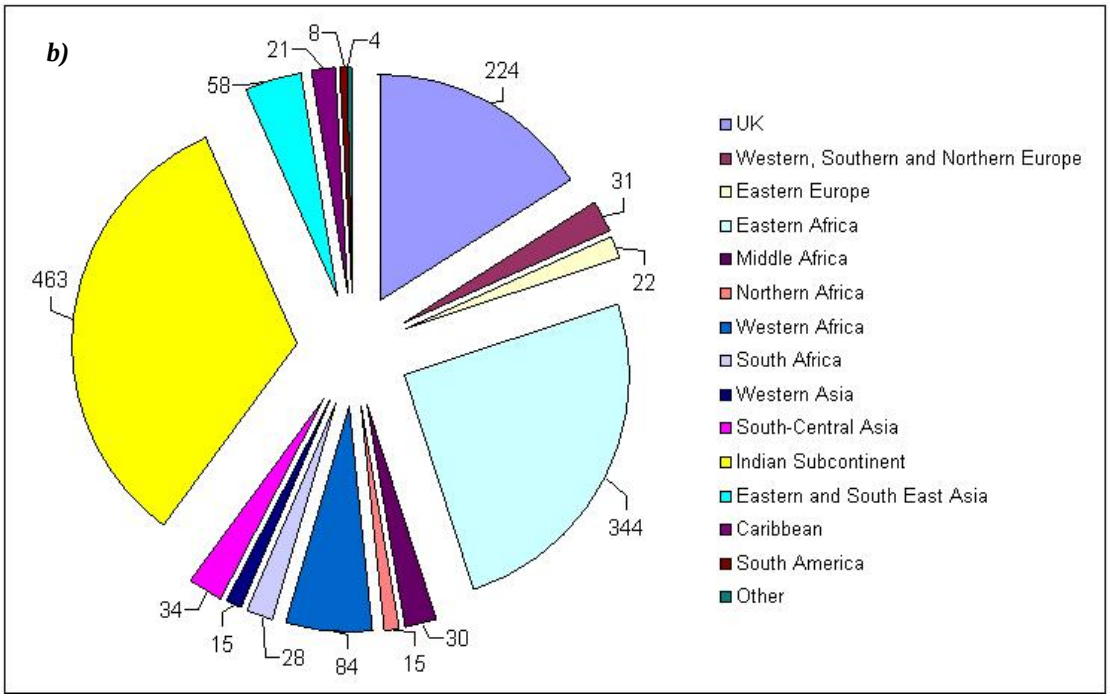
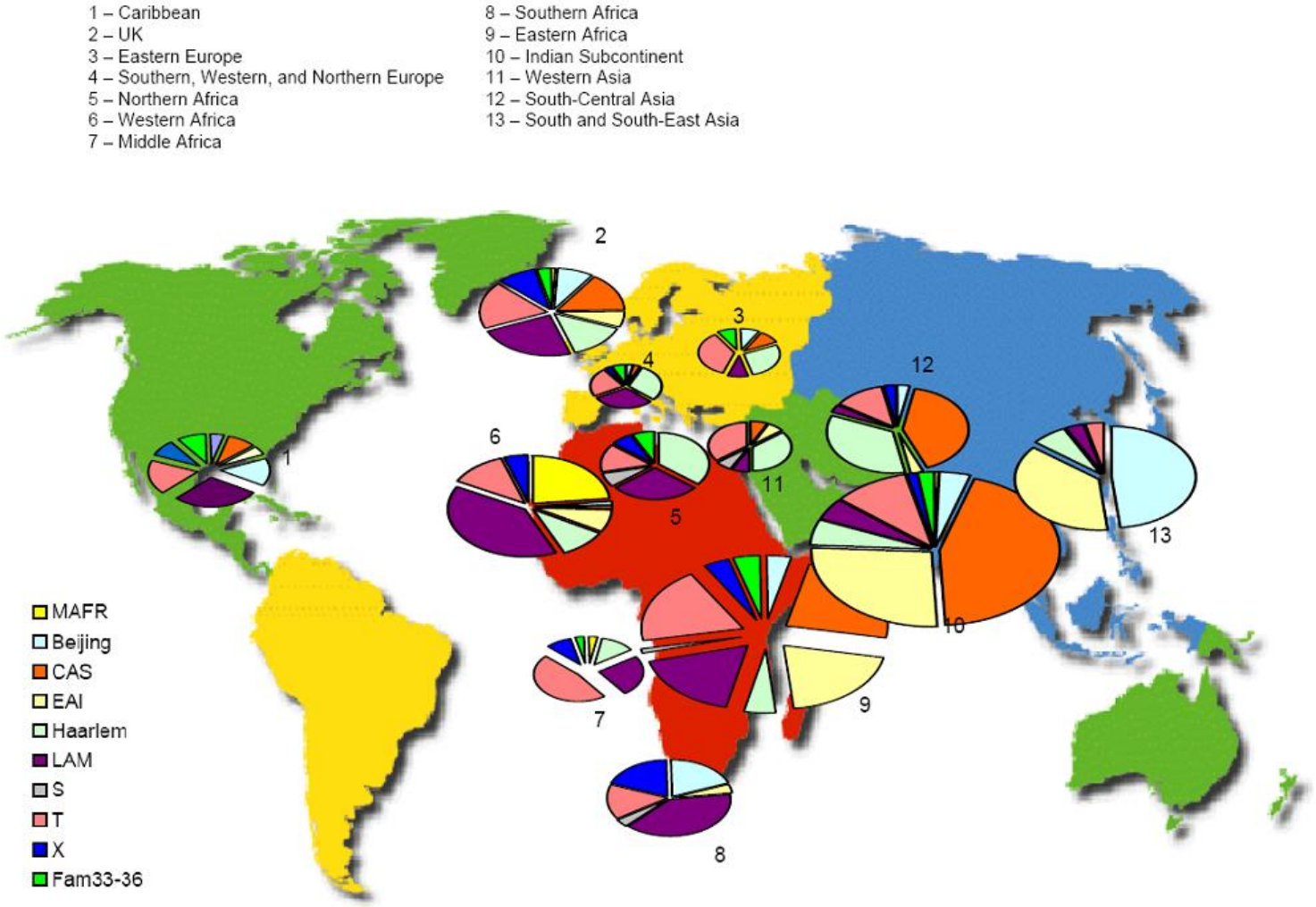


Figure 3.4 *The geographical distribution, using the global regions, defined by the World Health Organisation, of the major M. tuberculosis complex families that had been assigned to isolates using spoligotyping.*



If our underlying hypothesis was correct that these strains were more common or dominant in a population because of some underlying advantage compared to other strains of *M. tuberculosis* we would predict that in a range of experiments investigating different phenotypes we would be able to determine a difference or enhancement in the phenotype of one or more strain. An alternative way of considering this was for us to take a statistical approach in stating a null hypothesis that all TB strains are the same phenotypically and to then prove or disprove this null hypothesis.

The null hypothesis would be tested by comparing differences in phenotype or virulence *in vivo*. However, this could not be evaluated using all of the 42 strains in Panel A as this number would be impractical, too costly and was not necessary to prove or refute the underlying hypothesis. Previous studies had shown that the global phylogeny of TB strains fell into 6 major groups (including *M. bovis* and *M. africanum*) and so we determined whether the 42 strains in Panel A included strains spanning all the 6 phylogenetic groups. The *in vivo* experiments would need to include representatives of all of those groups.

3.3 PHYLOGENETIC ANALYSIS & SELECTION OF STRAINS

3.3.1 Phylogenetic analysis

Globally, TB strains can be ordered into 6 phylogenetic lineages (if *M. bovis* and *M. africanum* are included) (Baker *et al.,* 2004; Gagneux *et al.,* 2006). A study by Gibson *et al.,* 2005, demonstrated that MIRU and ETR codes could be used to classify *M. tuberculosis* and *M. bovis* strains into one of the five *M. tuberculosis* lineages (I, II, III, IV and *M. bovis*), initially defined by Baker *et al.,* 2004, and subsequently validated by Gagneux *et al*., 2006. Phylogenetic analysis involved assigning each of the 2261 isolates into each of the lineages based on the number of repeats in specific loci as detailed in Table 3.3. The proportion of families identified by spoligotyping was calculated within each of the lineages to check for any correlations between lineage and family. The results from this analysis are displayed in Figure 3.5.

Table 3.1 *The 24 clusters containing more that 9 profiles, after performing cluster analysis with MIRU and ETR profiles. The table includes details of the proportions of each of the M. tuberculosis families observed in each cluster. ∗ indicates the clusters containing sub-families and which were exempt from further analysis.*

MRU cluster number	Cluster size	MTBC spoligofamily	% of cluster
14	53	CAS	100
8	49	CAS	87.8
		Family35	4.1
		Beijing	4.1
		Haarlem 1	2.0
		S	2.0
1	43	LAM10	95.3
		T1	2.3
		No identification	2.3
18 ∗	33	T1	57.6
		T3	42.4
3	32	CAS	93.8
		No identification	6.2
27 ∗	31	LAM8	87.1
		LAM9	9.7
		T1	3.2
2 ∗	28	Haarlem 1	21.4
		Haarlem 2	7.1
		Haarlem 3	60.7
		T1	3.6
		X1	3.6
		No identification	3.6
42	27	LAM10	96.3
		X3	3.7
6	23	CAS	87.0
		Beijing	8.7
		EAI5	4.3
10	22	Beijing	100
201	21	*M. bovis* BCG	100
125	20	LAM1	95.0
		No identification	5.0
36	17	LAM3	100
100	16	T1	100
41	15	Beijing	100
12 ∗	14	Haarlem 1	28.6
		Haarlem 3	57.1
		T2	7.1
		X1	7.1
192	14	Beijing	100
38	13	EAI3	100
72	13	CAS	100
112	13	LAM10	100
130 ∗	12	LAM8	83.3
		LAM9	16.7
31	11	CAS	90.9
		No identification	9.1
148	11	LAM3	81.8
		X2	18.2
28	10	X2	100

Table 3.2 *The 42 strains that were included in Panel A, and strains that were then included in Panel B.*

MRU Cluster /Sub-cluster		*M. tuberculosis* spoligofamily	Country of Birth	Panel B
14	/1	CAS	Somalia	✔
	/unique	CAS	Bulgaria	
	/5	CAS	Hungary	
1	/1	T1	UK	
	/1	LAM10	UK	✔
	/1	LAM10	Nigeria	
3	/1	CAS	Jamaica	
	/unique	CAS	Qatar	
42	/1	LAM10	Nigeria	
	/2	X3	Nigeria	
	/1	LAM10	UK	
6	/4	Beijing	India	
	/4	EAI5	India	✔
	/4	CAS	Pakistan	
	/unique	CAS	Somalia	
	/6	CAS	India	
	/unique	CAS	Mozambique	
10	/unique	Beijing	China	
201	/unique	*M. bovis*-BCG	UK	
125	/1	LAM1	India	
	/1	LAM1	Poland	
	/unique	LAM1	Angola	
	/unique	LAM1	Gambia	
36	/1	LAM3	UK	
	/2	LAM3	Nigeria	
	/1	LAM3	Iran	
100	/1	T1	Jamaica	
41	/1	Beijing	UK	
	/3	Beijing	India	
	/unique	Beijing	Poland	
	/1	Beijing	Uganda	
192	/unique	Beijing	Hong Kong	
	/2	Beijing	Jamaica	
	/unique	Beijing	Malaysia	
72	/1	CAS	India	✔
	/2	CAS	Somalia	
	/1	CAS	Pakistan	
112	/1	LAM10	Pakistan	
31	/1	CAS	Somalia	
	/1	CAS	India	
28	/1	X2	UK	
	/unique	X2	Nigeria	
		Beijing	Estonia	✔

Table 3.3 *The defined MIRU loci and number of repeats that would define isolates into the five lineages, M. tuberculosis lineage I, II, III, IV and M. bovis according to Gibson et al., 2005.*

Lineage	**Locus associated with lineage (number of repeats)**		
I	MIRU-39 (3)	ETR-A (4)	ETR-C (4)
II	MIRU-16 (1, 2, 3)	MIRU-39 (2)	ETR-B (1, 2)
III	MIRU-23 (5)	ETR-C (2)	
IV	MIRU-24 (2)	MIRU-26 (2)	
M. bovis	MIRU-10 (2)	MIRU-40 (2)	ETR-C (5)

In total, 1223 isolates were ordered into the 5 *M. tuberculosis* lineages (lineages I to IV, which corresponded to East Asian, Euro-American, East African-Indian, and Indo-Oceanic lineages, respectively, and the *M. bovis* lineage) and 210 isolates, constituting 14.7%, remained undefined using the algorithm by Gibson *et al*., 2005. From the 1223 isolates, 1174 isolates were assigned to a single lineage, whilst the remaining isolates were assigned to two lineages. Lineage II contained the largest proportion of isolates (38.5% of the total number of isolates), while lineage III and IV were assigned 20.4% and 18.6% of isolates, respectively. The smaller lineages, lineage I and *M. bovis*, had similar proportions of isolates, 2.2% and 2.3%, respectively. From those isolates assigned to two lineages, lineage II overlapped with three other lineages including lineage III (17 strains), lineage IV (7 isolates) and *M. bovis* lineage (2 strains), whilst 23 isolates were assigned to both lineage I and IV. Using the algorithm, isolates identified as *M. africanum* by spoligotyping could not be assigned to any of the five lineages.

There was good correlation between families identified by spoligotyping and the lineages, especially for lineages I, III, IV, and the *M. bovis* lineage, which were primarily made up of isolates identified as *M. tuberculosis* Beijing, CAS, EAI and *M. bovis-BCG*, respectively. Lineage II was more diverse in the different families that were found in this lineage. Within lineage II were isolates belonging to T, LAM, Haarlem, X, and S families, which together constituted over 90% of isolates assigned to lineage II.

As the approach identified in Gibson *et al*., 2005 was unable to ascribe the appropriate lineage in 259 isolates, lineage designations were completed by performing SNP analysis, using genes described by Baker *et al*., 2004, including *oxyR* C37T, *katG* C87A, *rpoB* T2646G, and *rpoB* C3243T. This was performed by colleagues in HPA, MRU. From the 49 isolates originally assigned to more that one lineage, all strains

were assigned to one of the two lineages using SNP analysis. From the 210 isolates that were not defined at all, isolates identified as *M. tuberculosis* Beijing, CAS, and *M. bovis* BCG by spoligotyping were assigned to lineage I, III and *M. bovis*, respectively, by SNP analysis. The recently evolved modern families, Haarlem, LAM, X, T, and S were assigned to lineage II, and all but 3 isolates identified as EAI family were assigned to lineage IV. All *M. africanum* isolates, except one, were assigned to lineage I.

3.3.2 Selection of strains for Panel B

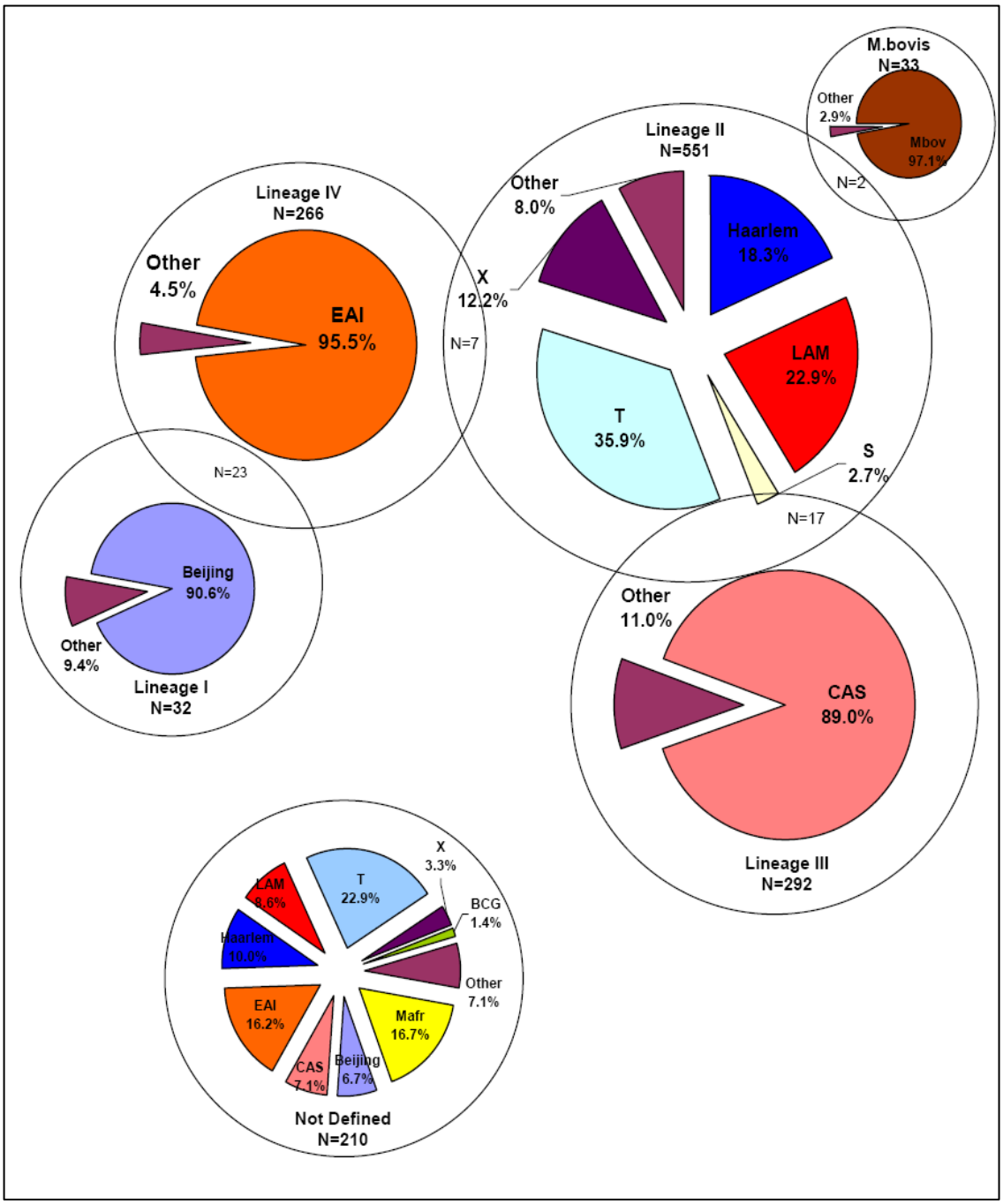
After performing the phylogenetic analysis it was decided to select strains primarily from Panel A but also to include some strains of interest identified in the scientific literature. *M. tuberculosis* Beijing, CAS and EAI strains were selected to represent lineages I, III, and IV. A LAM strain was chosen to represent lineage II because it was one of the larger families within this lineage. The final strains included in Panel B, the details of which can also be found in Table 3.2, were:

1) *M. tuberculosis* Beijing strain. Although Beijing strains had been included in Panel A, the Estonian Beijing IX strain was of more clinical significance as the MDR strain had appeared in the Estonian population and had come to dominate the population after a period of 10 years (data not published). A Beijing IX strain isolated from a patient in Estonia and fully sensitive to all first line drugs, isoniazid, streptomycin, rifampicin, ethambutol and pyrazinamide, was used to avoid bias if there are phenotypic changes related to resistance. This strain had the same MIRU and ETR profile as that of isolates in one of the larger clusters, cluster 10 (refer to Table 3.1).
2) *M. tuberculosis* LAM strain. A LAM10 strain, from Panel A, belonging to cluster 1 that had been isolated from a patient born in the UK was selected. This strain had been shown to be responsible for the largest outbreak of isoniazid resistant TB in London and, subsequently, in the UK (Ruddy *et al.*, 2004). The selected strain was resistant to isoniazid, but sensitive to streptomycin, rifampicin, ethambutol and pyrazinamide.
3) *M. tuberculosis* CAS strain. The CAS strain constituted approximately 25% of the population that was studied and, therefore, was essential for inclusion in Panel B. After performing cluster analysis using only the MIRU and ETR profiles, the largest cluster was Cluster 14 with 53 isolates all belonging to CAS family. A total of 53.1% of patients, for whom country of birth data were available, were from

4) *M. tuberculosis* CAS strain. From epidemiology analysis (Section 3.2.3), it was observed that most patients from the Indian subcontinent were infected with *M. tuberculosis* CAS and EAI strains, and India was the country of origin for most patients in this region infected with CAS. Therefore, a CAS strain was selected from Panel A, belonging to Cluster 72, which apart from Cluster 14, was the only other cluster in which all isolates had been identified as belonging to the CAS family. In Cluster 72, 44.4% of patients were born in India. The selected strain was sensitive to all first line drugs.
5) *M. tuberculosis* EAI strain. An EAI strain was selected to represent lineage IV. This family is of epidemiological interest because this strain was also isolated from most patients that were born in the Indian subcontinent. In Panel A there was only one EAI5 strain, which belonged to cluster 6 and isolated from a patient born in India. This strain was fully sensitive to all first line drugs.
6) Wild-type *M. africanum*. Besides lineages I, II, III and IV defined by Baker *et al*., 2004, a separate key lineage for *M. africanum* had been defined by Gagneux *et al*., 2006 (Baker *et al.,* 2004; Gagneux *et al.,* 2006). As the initial rationale for selecting Panel B was identifying a strain from each of the major MTBC lineage, a strain was selected at random from the study population of 41 *M. africanum* strains.

Laboratory control strains *M. tuberculosis* H37Rv (supplied by HPA, MRU), *M. tuberculosis* H37Rv and *M. bovis* BCG (supplied by HPA, Porton Down), and *M. tuberculosis* H37Ra (HPA, National Collection of Type Cultures; NCTC) were included for phenotypic analysis.

Figure 3.5 *The numbers of isolates assigned to each of the M. tuberculosis complex lineages, defined by Baker et al., 2004, and Gagneux et al., 2006, using MIRU and ETR codes defined by Gibson et al., 2005. The diagram also shows the proportions of families, identified by spoligotyping, which made up each lineage.*



3.4 DISCUSSION

The initial part of the present study involved identification of the various circulating strains and London provided the most appropriate setting with such an ethnically diverse population, so isolates from all 30 hospitals and/or laboratories situated in London were included prospectively for molecular typing. Patients were born in 89 countries, which included some of the high-TB burden countries defined by the WHO, and provided a significant global biodiversity of *M. tuberculosis* strains.

Isolates were typed using 12 MIRU and 3 ETR loci primarily, which grouped all profiles that were the same (clustered) and filtered all unique isolates. The clustered isolates were included for secondary analysis using the more discriminatory panel of 7 additional VNTR loci to identify isolates that were truly identical to each other. The usefulness of the 7 VNTR loci was demonstrated by the reduction in clustering rates from 54.5%, when using the MIRU and ETR loci, to 22.2%.

Although typing using the MIRU, ETR and VNTR loci allowed identification of the various common and unique molecular types circulating within London, spoligotyping was performed to identify which *M. tuberculosis* family each isolate belonged to and to understand the biodiversity of MTBC isolates. The online SPOTCLUST tool was used to assign spoligotyping profiles to MTBC families (Filliol *et al.,* 2002) and all but four strains were assigned to one of 32 families in the database and four additional families, Family 33 to 36.

During analysis of profiles it was observed that two different octal code profiles could be identified as the same *M. tuberculosis* family and it was determined that the recognition rules were based on deletions in specific spacers (Filliol *et al.,* 2002). For some octal code profiles there were two family identifications and it was realised that the rules were sometimes not sufficiently specific to some families. For instance, strains that had been identified as belonging to *M. tuberculosis* LAM1 family were also given an identification of LAM9, but at a lower probability, as the recognition rules involved deletions in spacers 21 to 24 and 33 to 36 for both families, and the tool could not distinguish between the codes.

After all of the molecular typing was completed, the global representation of *M. tuberculosis* strains in the study population of London was investigated. This was fulfilled by performing epidemiological analysis, which involved plotting proportions of identified MTBC families on a global map using the country of birth data that was available for 1381 patients. There were strong associations between country of birth of patients and *M. tuberculosis* families: for example the Beijing family was associated with patients born in South and South-East Asia, whilst the CAS strains were more predominant amongst patients originally from the Indian subcontinent and South-Central Asia. The EAI strain was slightly more widespread as it was commonly found in patients from the Indian subcontinent, South-East Asia and Eastern Africa, whilst the *M. africanum* strains were significantly associated with patients born in Western Africa.

Similar associations have been previously reported in recent studies and provide evidence of a clonal MTBC population in some high TB-burden countries and suggest higher TB transmission rates within these countries (Behr *et al*., 2004; Filliol *et al.,* 2002; Vitol *et al.,* 2006; Gagneux *et al.,* 2007). The lack of any definite association between country of birth and *M. tuberculosis* families S, T, Haarlem, and X, in this study, may suggest that the distribution of these families is fairly uniform globally. However, a more realistic explanation may be the poor definition of these families with the present recognition rules, as these families are still evolving and have a fairly recent origin, perhaps from the Indian subcontinent (Singh *et al.,* 2004; Gutierrez *et al.,* 2006; Vitol *et al.,* 2006). This epidemiological analysis demonstrated that the population of isolates included representatives from all MTBC families and almost all global regions, although there were few representatives from North and South America.

The clusters produced when using the MIRU and ETR profiles, along with the country of birth, family assignations and 7 VNTR profiles attached to the isolates within clusters, were used to select the strains for Panel A. The cluster analysis identified the more prevalent strains circulating within London and it seemed logical to base part of the selection of strains for vaccine evaluation on prevalence of strains as there is likely to be higher success in the transmission of these strains. The clusters that would be used for further analysis were selected by including clusters with more than 9 isolates. Whilst the main reason for these criteria was to include the more

prevalent strains, the decision was also influenced by limited country of birth data for clusters smaller than 10 isolates.

As strain prevalence was a factor for selection, clusters containing sub-families were excluded. In order to select which particular isolate from each cluster would be selected, an algorithm was designed to take into account important factors such as the country of birth of patients and the most prevalent 7 VNTR cluster within the main MIRU and ETR profile based cluster.

The country of birth data provided a very good indication of possible sources of strains and has been demonstrated previously by Dale *et al*., 2005 (Dale *et al.,* 2005). The strong association between the MTBC family identification and country of its origin may offer valuable information for vaccine development because, whilst the ideal option would be to have one universal vaccine, a niche vaccine providing significant efficacy against certain TB families, like *M. tuberculosis* CAS and Beijing, in regions where they are found at high incidence may offer protection to a substantial proportion of the population. For the purposes of selecting strains for Panel A, the strain was selected if there was a correlation between the most common country of birth and the most prevalent 7 VNTR cluster.

Using the algorithm, 42 strains of interest were selected. It was important to establish if there were any phenotypic differences between circulating *M. tuberculosis* strains using *in vitro* experiments initially, with the subsequent possibility of performing *in vivo* experiments. It was not possible to perform statistically valid *in vivo* experiments on all of the 42 strains in Panel A, so 6 wild-type strains were selected after performing phylogenetic analysis to represent the major *M. tuberculosis* lineages defined by Baker *et al*., 2004 and Gagneux *et al*., 2006 (Baker *et al.,* 2004; Gagneux *et al.,* 2006) (refer to Section 3.3). When assigning isolates into the lineages defined by MIRU and ETR codes, whilst the vast majority of isolates were assigned to a single lineage a small proportion were assigned to two lineages or not assigned at all (Gibson *et al.,* 2005). Therefore, SNP analysis was used to complete the designation of isolates successfully.

Phylogenetic analysis revealed families that made up the vast majorities of each of the lineages with *M. tuberculosis* Beijing, CAS and EAI strains representing lineages I,

III, and IV, respectively. Even though lineage II was made up of multiple families it was decided to select a LAM strain because it was one of the larger families within this lineage and, as mentioned previously, there was a LAM10 strain that was of particular interest. The final panel of strains comprised *M. tuberculosis* Beijing IX strain from a patient born in Estonia (this strain was selected as a result of unpublished data reporting its clinical significance as the strain has gone from first appearance to dominating the TB strains found in the Estonian population within a decade), *M. tuberculosis* LAM10 strain from a patient born in the UK (this isoniazid resistant strain was selected as it was responsible for the London TB outbreak), two *M. tuberculosis* CAS strains from clusters that contained isolates all identified as CAS, *M. tuberculosis* EAI5 strain isolated from a patient born in India, and a randomly selected *M. africanum* strain.

In addition, laboratory control strains were included (two *M. tuberculosis* H37Rv strains, *M. tuberculosis* H37Ra, and *M. bovis* BCG). An *M. tuberculosis* H37Rv strain from HPA, Porton Down, was included as the data from *in vivo* experiments would be compared to data from similar experiments previously carried out using the H37Rv strain from Porton Down. Therefore, phenotypic *in vitro* experiments would be performed using H37Rv strains from both laboratories.

CHAPTER 4

REPRODUCIBILITY OF NOVEL *M. TUBERCULOSIS* MINI-SATELLITE VNTR LOCI

4.1 GENOTYPING USING VNTR LOCI

4.1.1 Introduction

Molecular genotyping can be used to investigate suspected outbreaks and detect laboratory cross-contamination in low to middle TB incidence areas. The highest level of discrimination is required for population-level genotyping when identifying clustered cases that are not apparently linked and if there is limited epidemiological data (Gopaul *et al.,* 2006). In order to differentiate between re-infection and reactivation of infection, to detect laboratory cross-contamination and to establish epidemiological links between patients, IS*6110*-RFLP has been used routinely, sometimes in conjunction with spoligotyping (Supply *et al.,* 2006).

Typing using multiple VNTR loci, which had been previously proposed by Frothingham and Meeker-O'Connell, 1998, has been modified to increase discrimination by including additional loci (Frothingham *et al.,* 1998; Supply *et al.,* 2001; Kwara *et al.,* 2003; Supply *et al.,* 2006). Discrimination, which is very important for prospective molecular epidemiological studies, depends considerably on the number of loci used and their selection.

Despite the proposal of a standardised VNTR typing panel of 12 MIRU, 3 ETR and 9 VNTR (VNTR-424, -2401, -3690, -4156, -2163B, -1955, -4052, -2347 and -3171) loci, it was not clear if this panel provided sufficient discriminatory power in regions where homogeneous MTBC families were prevalent (Kam *et al.,* 2006; Nikolayevskyy *et al.,* 2006; Supply *et al.,* 2006; Oelemann *et al.,* 2007). There are reports of additional VNTR loci with improved discrimination for prospective routine typing and to distinguish isolates within particular homogeneous groups of strains with highly conserved genes, such as the *M. tuberculosis* Beijing family (Gopaul *et al.,* 2006; Kam *et al.,* 2006; Nikolayevskyy *et al.,* 2006). In the previous chapter, the additional seven VNTR loci had been useful due to their increased discriminatory

power to identify isolates that were the same as each other when applying the algorithm in Figure 2.1 to the 12 MIRU and 3 ETR clusters.

Concerns about the stability and reproducibility of some hypervariable and discriminative loci such as VNTR-3232, -2163A, -3336, and -1982 have resulted in their exclusion from MTBC multilocus VNTR typing panels proposed for international standards (Kremer *et al.,* 2005; Supply *et al.,* 2006). For these reasons, it was decided to conduct a study to examine the stability of hypervariable loci and the parameters associated with reproducibility.

4.1.2 Evaluation of conditions affecting analysis of VNTR loci

Factors affecting reproducibility, including optimisation of enzyme choice, PCR and fragment analysis conditions to ensure reproducibility of results of the two hypervariable loci, VNTR-3232 and -1982, were analysed using a panel of 16 MTBC isolates, which had been previously characterised (refer to Section 2.2) and for which there was complete data for the 12 MIRU, 3 ETR and 7 VNTR loci. Isolates were selected to cover a range of repeats in MIRU-26, ETR-B and VNTR-1982 and -3232 loci.

As primers were labelled with the same dye, a single primer mix was prepared with forward and reverse primers for MIRU-26 and ETR-B so the final concentration of each primer was 0.5 µM. Separate mixes were prepared for VNTR-1982 and -3232 as they were labelled with different dyes. For each of the 16 extracts, MIRU-26, ETR-B, VNTR-1982 and -3232 were amplified exactly as described in Section 2.2.2 and 2.2.6 using 0.2ml 8-strip tubes (Alpha Laboratories) instead. Each locus was amplified using BIOTAQ and Diamond polymerase, with their respective amplification cycles (refer to Section 2.2.2 and 2.2.6). Loci were also amplified using HotStartTaq DNA polymerase and HotStartTaq *Plus* DNA polymerase (both supplied by Qiagen). Each 10µl reaction contained 1× PCR Buffer (Qiagen), 0.25U/µl of the relevant polymerase, 0.2µM dNTPs, 0.125µM of relevant primer, 5% DMSO. For HotStartTaq DNA polymerase, the DNA amplification cycle was as follows: 95°C for 15 mins; 35 cycles of 94°C for 30 secs, 60°C for 30 secs, and 72°C for 60 secs; 72°C for 10 mins. The amplification cycle was the same when using HotStartTaq *Plus* DNA polymerase, except the initial 95^{o}C activation time was reduced to 5 mins. The number of repeats at loci in PCR products was calculated manually by resolving 4µl

of each product on a 1.2% (w/v) agarose gel as described in Section 2.2.4 against a 2000 b.p. HyperLadder II standard.

Molecular weights of PCR products and the corresponding number of repeats were then calculated by denaturing capillary electrophoresis. PCR products were pooled and resolved against DNA Size Standard 600 and MapMarker D1 labelled 640-1000 as described in Section 2.2.3 and 2.2.6, except three different parameter sets were used to analyse fragments, as detailed in Table 4.1. Method 1 (Table 4.1) shows the original parameters used for MIRU-VNTR analysis and the other methods included variations of the different parameters, including changes in capillary temperature, and in durations of denaturation and separation steps. Raw data were automatically analysed and the peaks in analysed data traces were automatically annotated with the molecular weights, in base pairs, and the calling tables (Appendix 1 and 2) were used to manually calculate the number of repeats in each locus.

Table 4.1 *The parameters used to analyse each set of sixteen pooled PCR products on the Beckman Coulter CEQ 8000 Genetic Analysis System.*

		Method 1*	**Method 2**	**Method 3**
Capillary	Temperature (^{o}C)	60	60	50
Denature	Temperature (^{o}C)	90	90	90
	Duration (secs)	120	180	180
Inject	Voltage (kV)	2.0	2.0	2.0
	Duration (secs)	30	30	30
Separate	Voltage (kV)	6.0	6.0	6.0
	Duration (mins)	60	60	70

* this was the original method used for 15 MIRU-VNTR

Reproducibility of typing using all of the 12 MIRU, 3 ETR and 7 VNTR loci, which were used for the larger study, was also investigated in a blinded study, by checking for consistency between MIRU-VNTR profiles of serial isolates, within the 2363 isolates that had been originally typed (refer to Section 2.2.6). There were 265 sets of serial isolates (i.e. multiple isolates from the same patient taken at successive time intervals), where each set contained 2 to 6 isolates from the same patients at different time points with intervals varying from 3 days to 11 months.

4.1.3 Optimised VNTR loci typing conditions

The 16 isolates selected for this study had repeats in MIRU-26 ranging from 1 to 8; ETR-B ranging from 1 to 6; VNTR-1982 ranging from 2 to 13; VNTR-3232 ranging from 1 to 16. Factors potentially influencing the performance of hypervariable VNTR loci were evaluated by amplifying loci using four different polymerase enzymes and resolving PCR fragments under non-denaturing (agarose gel electrophoresis) and denaturing conditions (automated capillary electrophoresis using three different sets of separation conditions; Table 4.1).

The ability to correctly amplify different VNTR loci was dependent on the enzyme used (Table 4.2). These results were the average of data from 2 repeated experiments. All polymerases were efficient for amplifying MIRU-26 and ETR-B, indicated by the presence of bands on agarose gels and peaks during capillary gel electrophoresis. All polymerases except Bioline BIOTAQ were able to amplify VNTR-1982 and longer fragments were amplified more efficiently by Bioline Diamond and QIAGEN polymerases. Effective amplification of VNTR-3232 was achieved only with Bioline Diamond (15/16 strains, i.e. 93.8%); therefore, Diamond polymerase was used for the amplification of additional VNTR loci.

Table 4.2 *The number of DNA extracts (out of the selected 16) for which peaks were detected by different conditions for capillary electrophoresis after amplifying the loci with different polymerases. The figures in brackets represent the number of extracts whose calculated number of repeats were [higher] and (lower) than the expected value.*

Locus	Method (Refer to Table 2.1)	Bioline polymerases		Qiagen polymerases	
		BIOTAQ	Diamond	HotStartTaq	HotStartTaq Plus
MIRU-26	1	16	16	16 [1]	16 [1]
	2	15	16	16 [1]	16 [1]
	3	16	16	16	16
ETR-B	1	16	16	16 (1)	16 (1)
	2	15	16	16 (1)	16 (1)
	3	16	16	16 (2)	16 (2)
VNTR-1982	1	8	13	14	14
	2	9	13	12	14
	3	6	11	12	14
VNTR-3232	1	11	15	13	14
	2	10	15	14	14
	3	11 (3)	15 (7)	13 (6)	13 (4)

Three different methods were evaluated for capillary gel electrophoresis, including Method 1 (Table 4.1), which is normally used in the laboratory for analysis. Methods 2 and 3 (Table 4.1) were included to investigate the effects of increasing the time for denaturing PCR products at 90°C and decreasing the capillary temperature. For each locus, the detected molecular weights of products, which had been amplified using each of the four polymerases and then resolved using the different capillary gel electrophoresis conditions, were plotted and compared with expected molecular weights, which were also plotted on the same graphs. The graphs for each locus, MIRU-26, ETR-B, VNTR-1982 and -3232 are shown in Figure 4.1a to 4.1d.

Molecular weights of the MIRU-26 fragments were as expected for all but 2 repeats. For this allele, the molecular weight of fragments amplified using BIOTAQ and Diamond polymerases were approximately 344 b.p., as expected, but higher than expected when using the Qiagen polymerases meaning that the interpreted number of repeats was 1 repeat more than expected. The smaller ETR-B fragments with 1 and 2 repeats all had molecular weights of expected sizes when using Methods 1 and 2, but were lower than expected when the capillary temperature was decreased (Method 3); this did not affect the interpreted number of repeats. For the higher number of repeats (4, 5 and 6 repeats), independent of the polymerase used to amplify them, molecular weights of fragments when resolved using capillary gel electrophoresis Method 3, were lower than expected, which in some cases affected the interpretation of the data.

Irrespective of the polymerase used for amplification and the method used for capillary gel electrophoresis, the molecular weights of amplified VNTR-1982 fragments were all similar to expected values as were the interpreted numbers of repeats. Analysis using Methods 1 and 2 for capillary gel electrophoresis generated the expected molecular weights for VNTR-3232 fragments, whilst lower molecular weights were detected when using Method 3. The lower molecular weights for VNTR-3232 fragments with an expected 1 to 6 repeats did not influence the interpreted number of repeats, but for fragments with more than 7 repeats, the lower molecular weights produced, gave a lower repeat number than expected.

Figure 4.1 *The detected molecular weights of (a) MIRU-26, (b) ETR-B, (c) VNTR-1982, and (d) VNTR-3232 fragments (amplified using different polymerases) when capillary gel electrophoresis analysis was performed under the different conditions outlined in Table 2.1.*

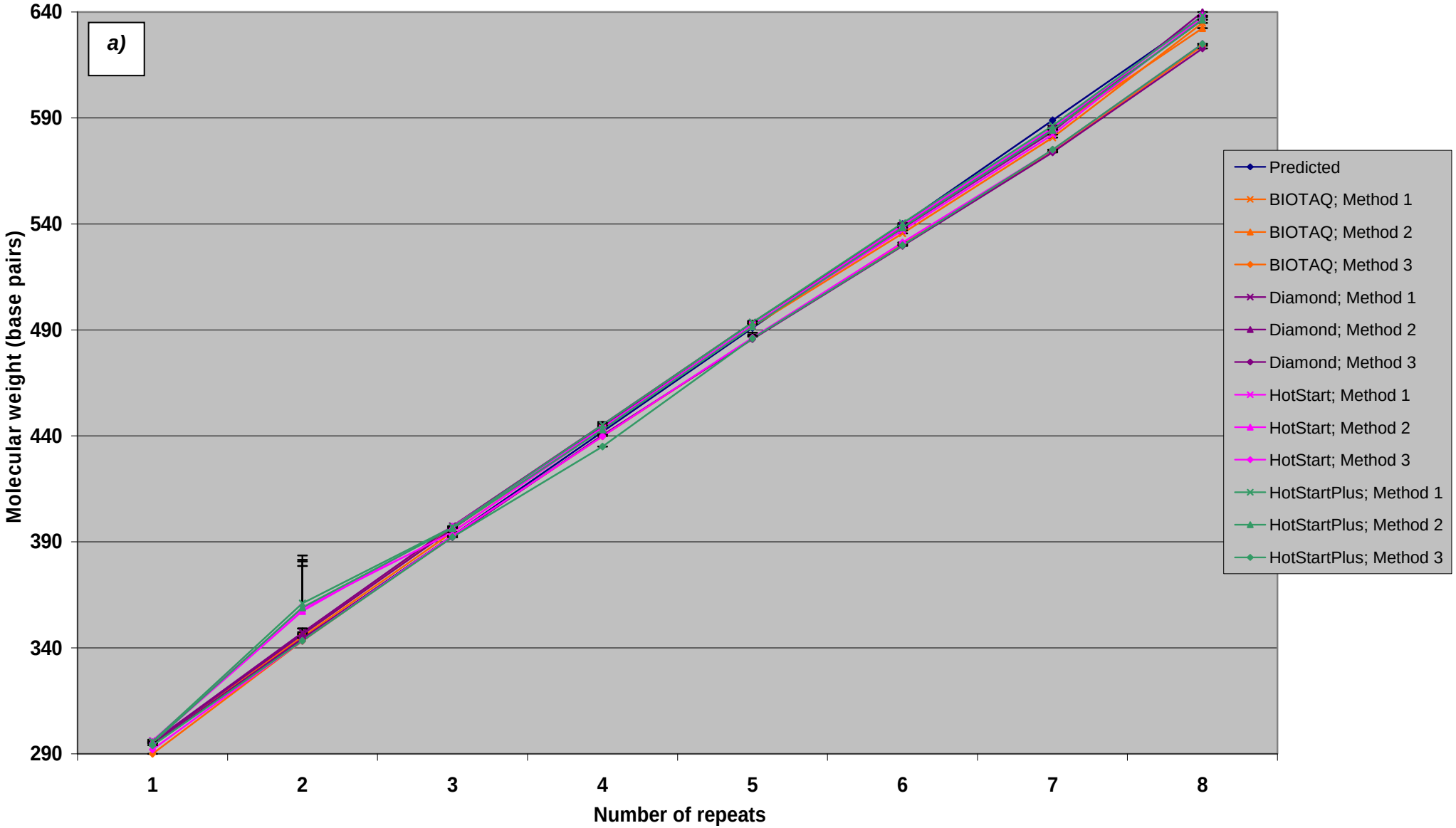


Figure 4.1 *continued*

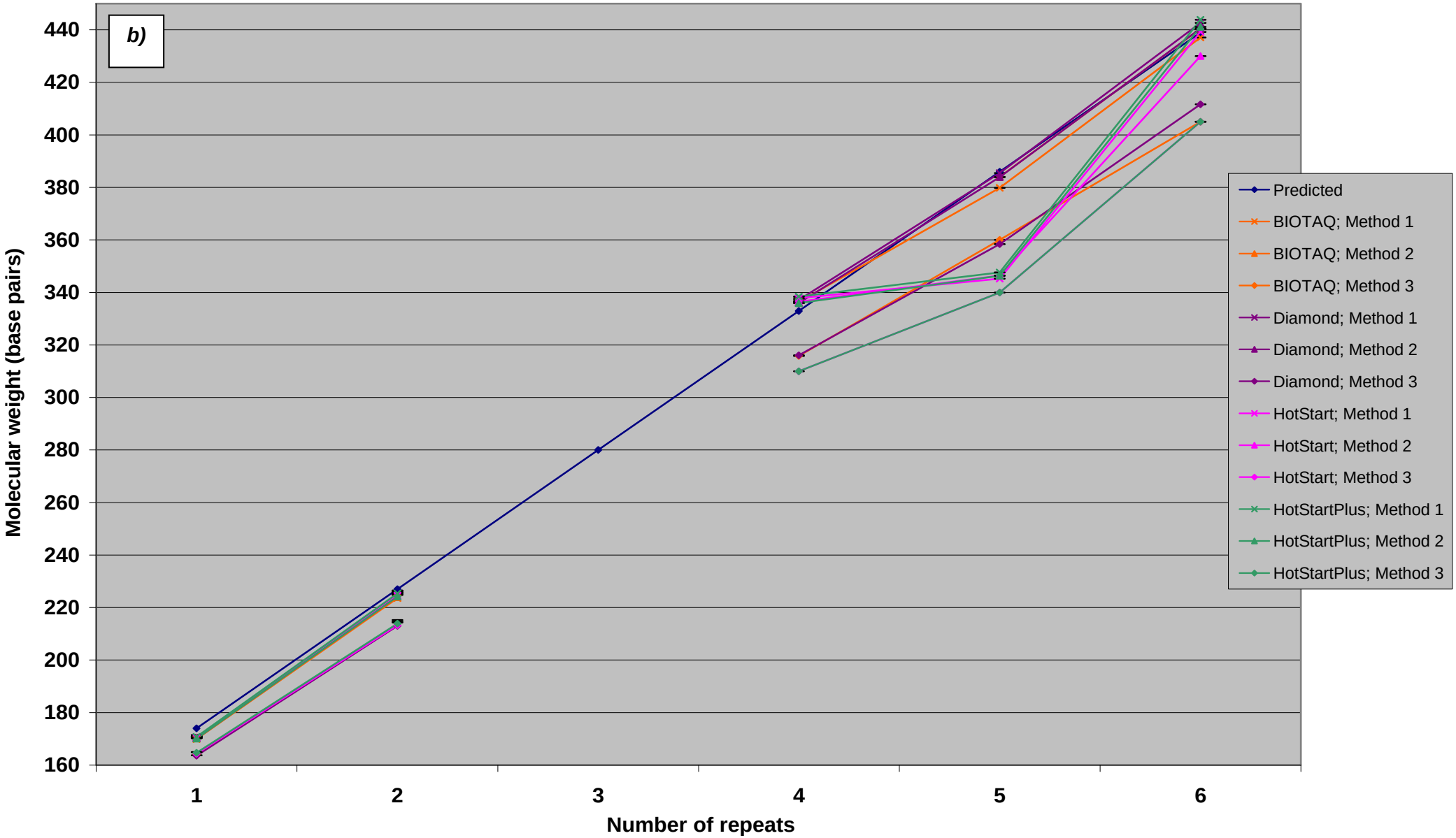


Figure 4.1 *continued*

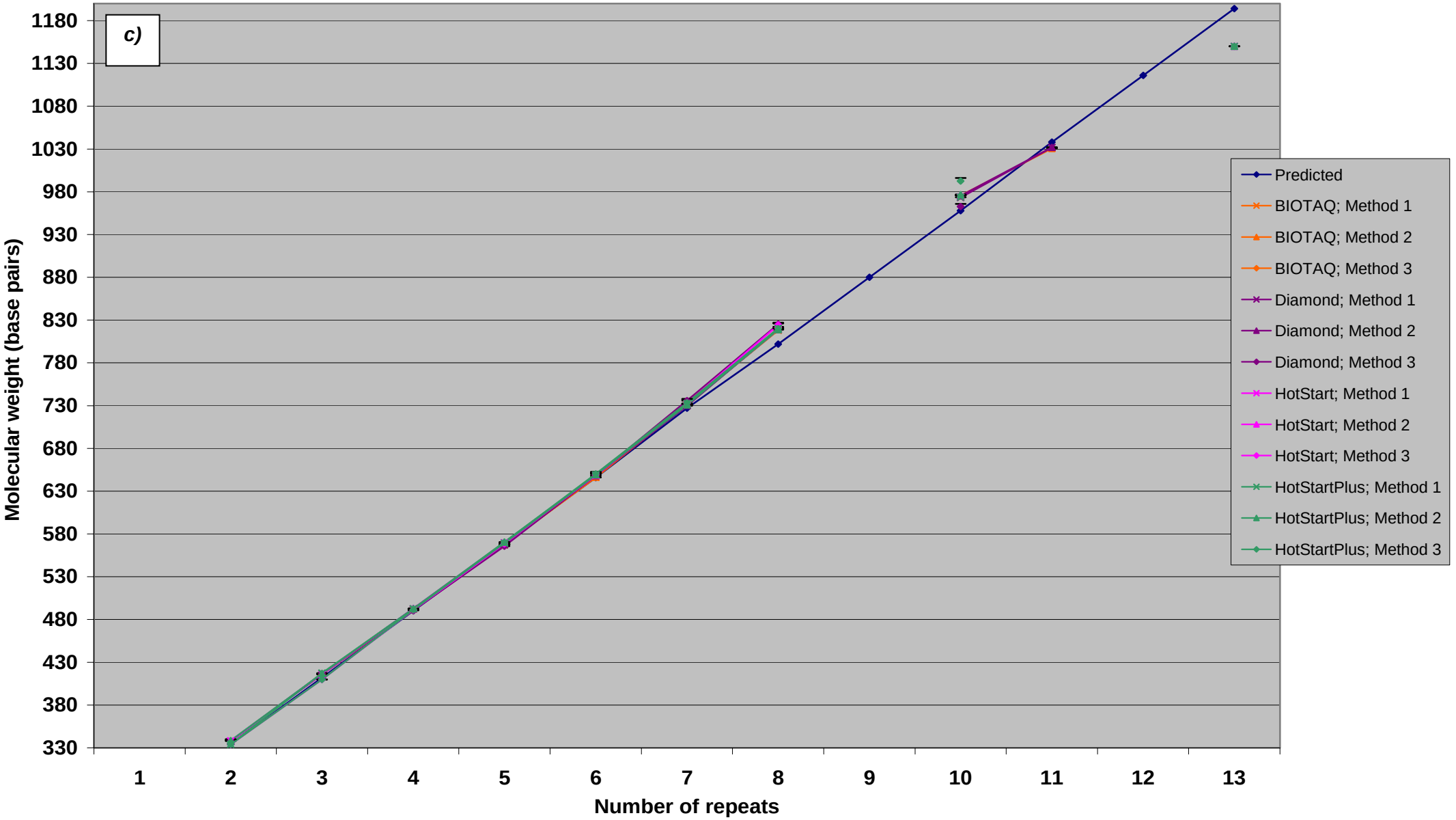
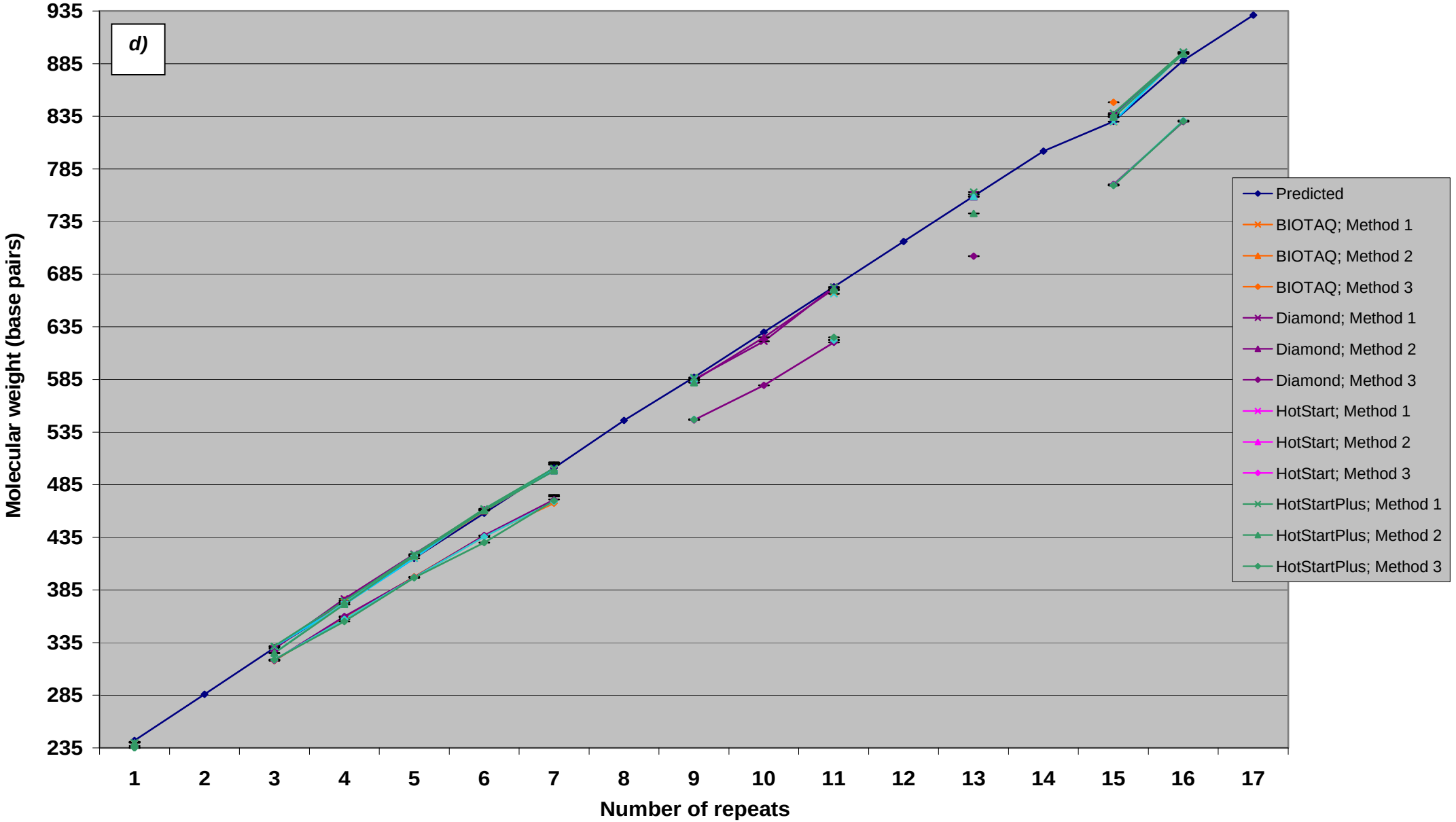


Figure 4.1 *continued*



4.1.4 Reproducibility of MIRU-VNTR typing

The stability and reproducibility was further investigated by performing MIRU-VNTR typing, using all 22 loci (12 MIRU, 3 ETR and 7 VNTR) on serial isolates (refer to Section 2.2.8). There were 265 sets of serial isolates with a total of 367 isolates. MIRU and ETR loci had been amplified using Bioline BIOTAQ polymerase and VNTR loci had been amplified using Bioline Diamond polymerase. Fragment analysis was performed using the optimised capillary gel electrophoresis conditions detailed in Method 1 (Table 2.1). Analysis was blinded.

There were no discrepancies in the MIRU-VNTR profiles between isolates in each set of serial isolates. In a proportion of serial isolates, genotyping results were validated by a colleague using both agarose and capillary gel electrophoresis and again no differences were found between different runs and different methods of fragment separation (Figures 4.2 and 4.3).

4.1.5 Genotyping using the additional 7 VNTR loci

Initially, for the main study, 2363 MTBC isolates were typed using 12 MIRU and 3 ETR loci. Cluster analysis was performed using the 12 MIRU and 3 ETR profiles. A cluster was defined as containing two or more isolates with the same 12 MIRU and 3 ETR profiles. All clustered isolates were then including for typing using the 7 additional VNTR loci (refer to Section 2.2.6). There was a clustering rate of 54.5% when using the MIRU and ETR profiles (refer to Section 3.1.1 for details). This clustering rate suggested that high rates of active TB transmission might currently be occurring in London.

From the 1225 isolates that had been clustered using MIRU and ETR profiles, 1196 isolates were subjected to secondary typing using the additional 7 VNTR loci (VNTR-2163B, -2347, -3232, -2163A, -1982, -3336 and -4052) and the optimised conditions described above. Improved resolution was achieved as strains that had been clustered initially were subdivided into new groups: 1730 isolates had unique genotyping patterns and the remaining 502 isolates were grouped into 158 clusters giving a new substantially lower clustering rate of 22.2% (see supplementary data on the disc for the dendrogram produced after cluster analysis). The addition of the 7 VNTR loci demonstrated that when using the 12 MIRU and 3 ETR loci there was an over-estimation in the transmission rate of strains.

Figure 4.2 *An agarose gel showing the stability of the locus VNTR-3336 and reproducibility of typing data by amplifying the locus in 2 serial isolates from 4 different patients, A, B, C and D, and then resolving products using non-denaturing agarose gel electrophoresis.*

1 – Patient A, Isolate 1, isolated 20/06/05; 8 copies

2 – Patient A, Isolate 2, isolated 11/07/05; 8 copies

3 – Patient B, Isolate 1, isolated 08/07/2005; 9 copies

4 – Patient B, Isolate 2, isolated 08/08/2005; 9 copies

5 – Patient C, Isolate 1, isolated 11/11/2005; 7 copies

6 – Patient C, Isolate 2, isolated 16/11/2005; 7 copies

7 – Patient D, Isolate 1, isolated 16/05/2005; 6 copies

8 – Patient D, Isolate 2, isolated 25/05/2005; 6 copies

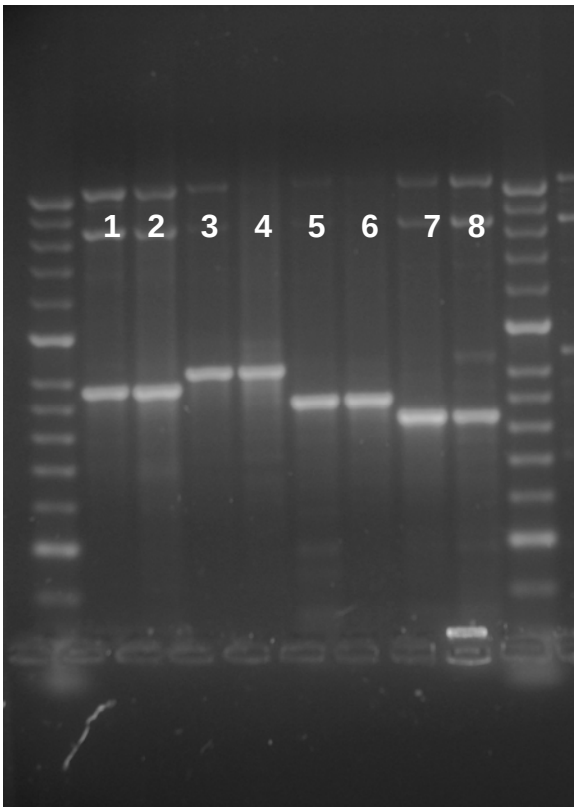
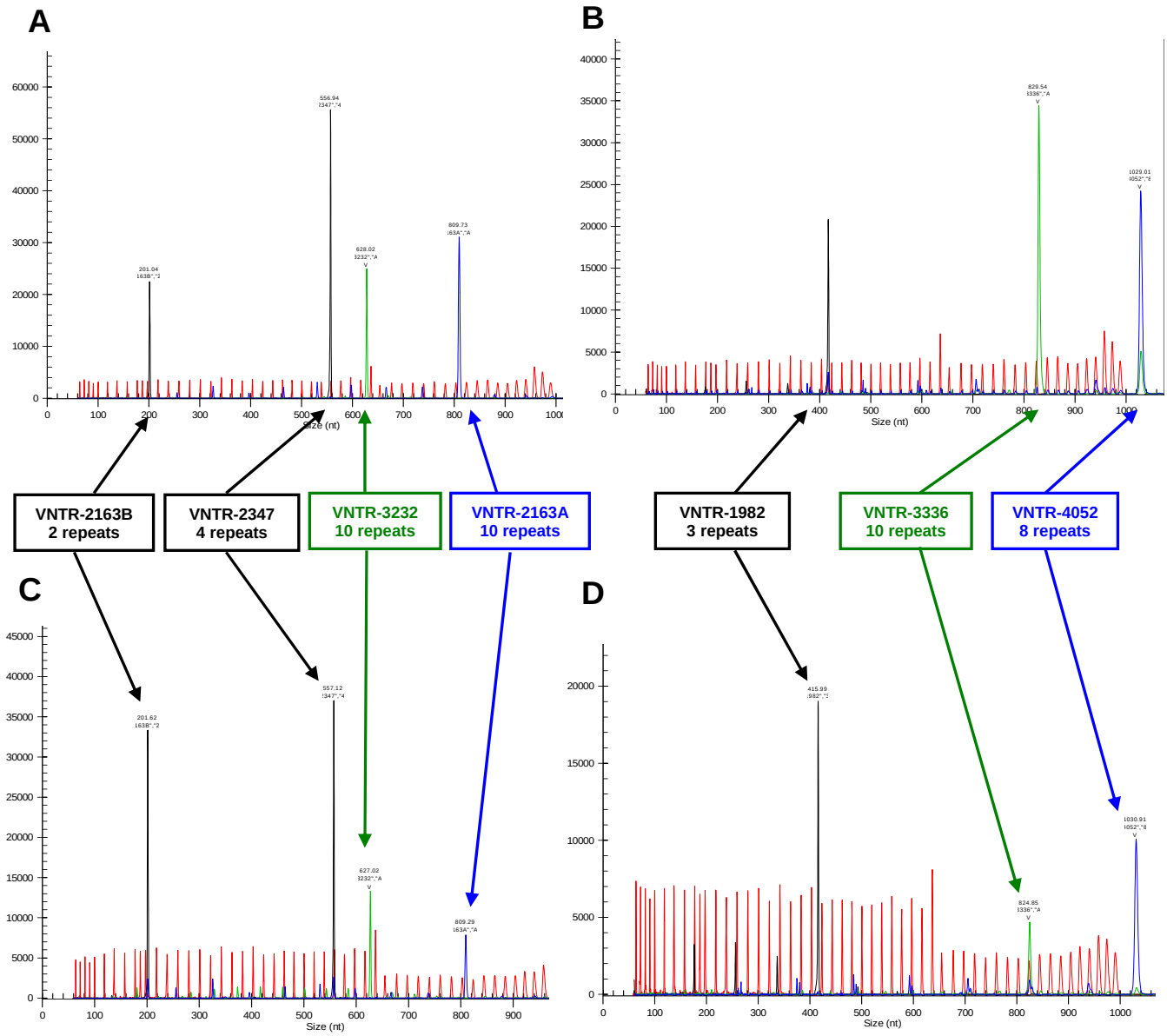


Figure 4.3 *Analysed data traces produced by the Beckman Coulter CEQ 8000 Genetic Analysis System. All traces were produced after amplifying the additional 7 VNTR loci from two isolates taken from the same patient at different time points with an 11-month interval. Traces A and B are from the first isolate and traces C and D from the second isolate. The data demonstrate the stability of these VNTR loci in serial isolates from the same patient and the reproducibility of typing using the optimised protocol for amplification and fragment analysis.*

Traces A & C – peaks representing VNTR-2163B, -2347, -3232, -2163A
Traces B & D – peaks representing VNTR-1982, -3336; -4052



The discriminative ability varied between all 22 loci with the Hunter-Gaston Discrimination Index (HGDI) values of MIRU loci ranged from 0.134 (MIRU-2 with 4 allelic variants) to 0.727 (MIRU-40 with 11 allelic variants). The HGDI values were calculated manually for all MIRU, ETR and VNTR loci using the principles described by Hunter and Gaston, 1988 (Hunter *et al.,* 1988). For the ETR loci, values ranged from 0.551 (ETR-B with 9 allelic variants) to 0.813 (ETR-A with 14 allelic variants). The HGDI values of VNTR loci were considerably more with values for VNTR-2163B, -2347, -3232, -2163A, -1982, -3336, and -4052 being more than 0.900 (0.985, 0.954, 0.994, 0.992, 0.989, 0.992, and 0.987, respectively). The allelic variants were higher for the VNTR loci indicative of their higher polymorphic nature (13, 6, 19, 18, 14, 15 and 10, respectively). As expected, none of the 22 loci were monomorphic in the current study. Loci included in the additional panel displayed higher variability than the panel of 12 MIRU and 3 ETR loci indicating their potential usefulness for prospective molecular genotyping.

4.2 DISCUSSION

Polymorphisms in rapidly evolving repetitive sequences like minisatellite VNTR are usually considered a valuable tool for practical epidemiological studies providing a very high degree of discrimination, which is an essential pre-requisite for prospective epidemiology where there would be little prior epidemiological data available. There are conflicting views about using hypervariable VNTR loci for typing, even though VNTR-3232 has been reported to have a particularly high discriminatory power in comparison with most VNTR, MIRU and ETR loci (Roring *et al.,* 2002; Gopaul *et al.,* 2006; Nikolayevskyy *et al.,* 2006).

The converse argument against its use is supported by the study of Supply *et al.,* 2006, who experienced problems when using VNTR-3232, including amplification of multiple alleles, the absence of PCR amplification products, data interpretation and interlaboratory reproducibility; due to the hypervariability experienced, this locus was excluded from their panel (Supply *et al.,* 2006). Amplification issues with this locus have also been reported by Kremer *et al.,* 2005 (Kremer *et al.,* 2005). Another locus with great potential for genotyping and for differentiating between strains is VNTR-1982, but this locus was excluded for similar reasons (Nikolayevskyy *et al.,* 2006; Supply *et al.,* 2006). Therefore, by identifying the conditions giving good

discrimination and those responsible for lack of reproducibility, optimal conditions were defined, which would enable molecular epidemiologists to use VNTR-1982 and -3232.

In the present study, the variability and reproducibility of these two loci was addressed, and MIRU-26 and ETR-B were included as controls because amplified fragments are stable during agarose and capillary gel electrophoresis and were also used as controls by Supply *et al.,* 2006 in their inter-laboratory analysis (Supply *et al.,* 2006).

The quality of bands and peaks for MIRU-26 and ETR-B were not affected by the polymerase used and the performance of all enzymes was 100%. However, peaks for VNTR-1982 and -3232 repeats were absent, particularly when using BIOTAQ polymerase and for fragments with an expected larger molecular weight. An improved performance was gained when using Diamond (VNTR-3232) and Diamond or Qiagen HotStart (VNTR-1982) polymerases (93.8% and 87.5%, respectively). For this reason, Bioline Diamond was recognised as the most suitable for amplifying the VNTR loci.

The variations in performance of polymerases for amplifying specific loci can be explained by their properties. BIOTAQ polymerase is a standard Taq, normally used for a wide variety of templates, whereas Diamond polymerase is a modified enzyme with a point mutation in its active site meaning it can amplify regions such as microsatellites, secondary structures, and templates with expected high guanine-cytosine percentage (G-C%) regions more effectively. The latter target for amplification is of particular interest when amplifying regions of the *M. tuberculosis* G-C rich genome. The Qiagen polymerases are chemically modified with a specificity similar to that of Diamond polymerase, which possibly explains their similar abilities in amplifying both VNTR loci. The buffer used with the Qiagen polymerases increases the specificity of binding of primers to the region of interest making these polymerases ideal for complex genomic DNA. HotStartTaq *Plus* polymerase is actually recommended by the supplier for typing studies.

Conditions influencing the detected sizes of PCR fragments that would influence interpreted copy number during capillary electrophoresis, namely parameters

affecting denaturation of PCR products, were investigated. Capillary temperature and the duration for DNA denaturation were altered as it was hypothesised that these parameters would affect the separation of DNA strands, therefore influencing how linear the DNA was, ultimately affecting the accuracy to which molecular weights of fragments are detected. Increasing the duration of the separation step allowed time for fragments larger than 1000 b.p. to be detected.

The parameters normally used for MIRU-VNTR analysis in the laboratory are detailed in Method 1 (Table 2.1). Method 2 was designed to examine the effects of increasing the time for denaturation. The molecular weights of fragments for each locus, obtained when using the original and modified Method 2, were similar and the offset values (differences between expected and experimental molecular weights) were similar.

Differences were observed when capillary temperature was decreased (Method 3; Table 2.1). Independent of the locus investigated and polymerase used, molecular weights of fragments were lower than expected, with offset values significantly larger in some cases, to the point where the calculated number of repeats differed from the expected values. Proof that the decreased detected molecular weights were the effects of decreased capillary temperature was obtained by amplifying a single locus from the same extract using each polymerase and then visually inspecting the fragments on an agarose gel; bands were at the same level when compared to each other and had the same molecular weights when compared to the molecular weight markers.

The denaturing conditions and capillary temperature were critical for reliable MIRU-VNTR analysis. So whilst optimised amplification of MIRU and ETR loci was with BIOTAQ polymerase, and Diamond polymerase for amplifying the VNTR loci, optimised conditions for analysis of PCR product were as detailed in Method 1; Table 2.1. The reproducibility and stability of the VNTR loci was also demonstrated by comparing 22 MIRU-VNTR profiles from serial isolates. The consistency between profiles of serial isolates from same patients showed that the amplification and fragment analysis conditions were indeed ideal for typing these loci and that these loci could be used for routine genotyping.

Using all 22 loci for typing gave the lowest clustering rate of 22.2% in MTBC strains obtained over one year from a single metropolitan setting (London), which is similar to proportions established in previous studies conducted in London in 1993 and 1995-1997, and similar to population-based studies in areas of low to middle TB incidence where RFLP was used alongside molecular genotyping methods (Maguire *et al.,* 2002; Oelemann *et al.,* 2007). This suggests, from the public health point of view, that TB transmission in London has remained stable over the last decade.

This present study determined that the MIRU and ETR panel are not sufficiently discriminative for tracking TB transmission. The higher allelic variation in the VNTR loci has also been observed in a previous study by Roring *et al*., 2002 (Roring *et al.,* 2002). For the purposes of genotyping and epidemiological analysis, this particular study was of interest as it suggested that using VNTR and ETR loci in combination could increase the discriminative power of a panel of loci (Roring *et al.,* 2002). The current observations regarding the applicability of the hypervariable VNTR loci, especially VNTR-3232, -3336, -2163A, and -1982, are in agreement with observations from previous studies, some of which demonstrated the use of VNTR loci for distinguishing between strains within the *M. tuberculosis* Beijing family (Drobniewski *et al.,* 2005; Iwamoto *et al.,* 2007; Wada *et al.,* 2007).

Our study suggests that hypervariable VNTR loci can be both discriminating and reproducible and, therefore, could be potentially be used successfully at multiple laboratories with consistent results providing that there is strict adherence to proposed reaction and PCR fragment separation conditions and the use of specific DNA polymerases.

CHAPTER 5

DEVELOPMENT OF A RAPID PCR-BASED QUANTIFICATION ASSAY OF *M. TUBERCULOISIS* COMPLEX CULTURES

5.1 INTRODUCTION

In cultures, mycobacteria tend to grow as clumps and when plating out bacteria for c.f.u. there is no certainty that a single colony has grown from a single mycobacterial cell, as might be the case with *Escherichia coli*, for example. Hence, there could be a huge margin for error in knowing how many mycobacteria are being used in experiments. Although it is a gold standard method for quantifying bacteria, there is a disadvantage with performing a conventional c.f.u. count with slow growing mycobacteria as it takes at least 3 weeks for colonies to grow meaning there would be a delay in setting up experiments. Whilst waiting for these colonies to grow, cultures are frozen and after thawing there would be loss of viability meaning the c.f.u. data may not necessarily be accurate. If the cultures were used directly after thawing, mycobacteria may not be in the exponential phase of the growth curve, meaning there would be a lag in the growth of bacteria during experiments.

For this reason it was decided to investigate the use of a potentially more accurate and quicker method to quantify the number of individual mycobacteria, so as to be able to set up phenotypic assays on the same day as quantification, without having to freeze-thaw cultures. The method under investigation was real-time PCR.

Real-time PCR is increasingly used because data are reproducible; it is a highly sensitive method and has the ability to perform quantification over a wide range (Chini *et al.,* 2007). This method has many applications including quantifying gene expression as demonstrated by Chini *et al*., 2007, where the expression of the *tst* gene responsible for producing toxic shock syndrome toxin-1 in methicillin-resistant *Staphylococcus aureus* was investigated (Chini *et al.,* 2007). This PCR assay has also been used for quantifying complementary DNA (cDNA) and measuring levels of various cytokines in a system (Whelan *et al.,* 2003).

In *M. tuberculosis*, the melting points of end PCR product, which is a feature of the real-time PCR cycles, were used to detect mutations in *rpoB* and the *katG* gene that encodes for the catalase-peroxidase enzyme (Garcia de Viedma *et al.,* 2002). In addition to genotyping assays already available for detecting rifampin and isoniazid resistance by focussing on the *rpoB* and *inhA* regions, the mutations in *rpoB* and *katG* genes conferred rifampin and isoniazid resistance, respectively, and real-time PCR have provided another genotypic system by which resistance to these drugs could be detected rapidly (Garcia de Viedma *et al.,* 2002; Miotto *et al.,* 2006; Miotto *et al.,* 2008).

More importantly, for the purposes of this study, real-time PCR has great potential for calculating the initial DNA levels, which would be equivalent to bacterial load. The IS*6110* region of the *M. tuberculosis* genome has been used to quantify bacteria in paraffin-embedded and formalin-fixed biopsy sections from lymph nodes and to positively prove that levels of *M. tuberculosis* DNA in sputum correlated with the numbers of acid-fast bacilli calculated by microscopy (Desjardin *et al.,* 1998; Ishige *et al.,* 1999). The use of other IS elements have been used to rapidly quantify mycobacterial DNA, including IS*900*, which enabled quantification of *M. avium* subspecies *paratuberculosis* in milk, and IS*2404*, which specifically quantified *M. ulcerans* in tissue excised from patients with Buruli ulcer (Rondini *et al.,* 2003; O'Mahony *et al.,* 2004).

For quantification of mycobacteria at the HPA, MRU, it was decided to perform the real-time PCR assay using the *rpoB* and *katG* genes as they are present as single copies, and they have been used in such assays previously (Garcia de Viedma *et al.,* 2002). In addition, as the present study was including all isolates belonging to the MTBC, the *rpoB* gene would be of most use, as this gene is present in all MTBC strains and other PCR assays have demonstrated that this gene can differentiate between mycobacteria belonging to MTBC and non-tuberculous mycobacteria (Kim *et al.,* 2004). The *oxyS* gene was also included in the analysis as this gene was not used as a target in any of the other PCR-based assays performed at the HPA, MRU, thereby reducing any chances of future cross-contamination problems.

This small study attempted to validate real-time PCR quantification data by comparing the growth curves from real-time PCR data with growth curves from

OD_{600} and c.f.u. data. In order to use real-time PCR, a DNA standard, containing a known number of *M. tuberculosis* H37Rv genome copies, had to be prepared and the performance of different primers evaluated using the DNA standard as the template.

5.2 DEVELOPMENT OF PCR-BASED QUANTIFICATION ASSAY

5.2.1 Preparation of *M. tuberculosis H37Rv* DNA standard

After 4 weeks, there was growth of mycobacteria on all pyruvate and glycerol slopes that had been inoculated whilst preparing frozen archive cultures (refer to Section 2.4). Bacterial growth from one of the pyruvate slopes that had been inoculated with *M. tuberculosis* H37Rv was transferred into a 1.5ml microcentrifuge tube and re-suspended in 100µl TE buffer (10mM Tris, 1mM EDTA pH8.0; prepared as described in Section 2.1.4). The re-suspended culture was incubated in an 80^{o}C waterbath for 50 mins to kill the mycobacteria. Even though this method of killing is effective, the remaining procedures for DNA extraction were still performed under the biological safety cabinet.

Bacterial cells were lysed by adding 5µl 100mg/ml lysozyme (prepared as instructed in Section 2.1.4) to the tube, which was vortexed and incubated for 2 hrs in a 37^{o}C hot block. After incubation, 25µl proteinase K and 200µl Buffer AL (both provided in the Qiagen DNeasy Blood & Tissue kit) was added to the lysed bacterial cells, to help break down any cell debris. The contents of the tube were mixed by vortexing and then the tube was incubated in a 56^{o}C hot block for 30 mins. Before the purification of extracted DNA, 200µl 96-100% ethanol (Sigma-Aldrich) was added to the tube and it was important to make sure that the contents were thoroughly mixed together by vortexing.

After this step it was safe to work with the extracted DNA on the bench. For the purposes of real-time PCR, DNA had to be as pure as possible. DNA purification was performed using the consumables, reagents, and protocol provided with the Qiagen DNeasy Blood & Tissue kit. Firstly, the DNA extract was transferred into a DNeasy Mini spin column that was resting in a 2ml collection tube. During centrifugation at 6000×*g* for 1 min, the DNA bound to the membrane within the column, whilst the flow-through was collected in the collection tube. After centrifugation, the flow-through was discarded and the column transferred into a new collection tube.

In order to purify the DNA, 500µl Buffer AW1, prepared as described in Section 2.1.4, was added to the tube, which was then centrifuged at 6000×*g* for 1 min. Again the flow-through was discarded and the column transferred into another clean collection tube. The second purification step involved adding 500µl Buffer AW2, prepared as described in Section 2.1.4, and centrifuging at 6000×*g* for 3 mins to ensure that the membrane in the column was dry. After the flow-through was discarded, the column was transferred into a clean 1.5ml microcentrifuge tube and left open so any residual ethanol could evaporate as the ethanol could affect the performance of the PCR. The initial pellet of harvested cells was large; therefore, 150µl Buffer AE was added directly onto the membrane and the tube was incubated for 1 min at room temperature. DNA was eluted and collected in the 1.5ml microcentrifuge tube by centrifugation at 6000×*g* for 1 min. Agarose gel electrophoresis was performed using 5µl DNA to ensure that extraction and purification was successful. The DNA could be seen as a smear when the agarose gel was viewed under the ultraviolet light.

The resulting concentration of double-stranded DNA, in ng/µl, was calculated at 260nm using a NanoDrop 1000 Spectrophotometer (Thermo Scientific) and its attached software. The number of genomes per µl in the DNA extract was calculated using the DNA concentration. The details of the calculation can be found in Appendix 4. Once the number of genomes/µl was established, Buffer AE, which had been used to elute the DNA during purification, was used to dilute a small volume of the standard DNA to obtain a 1×10^7 genomes/µl standard. Still using Buffer AE as the diluent, 10-fold serial dilutions were prepared, resulting in standards containing 1×10^6, 1×10^5, 1×10^4, 1×10^3, 1×10^2, 1×10^1 genomes/µl. The DNA standards were stored at -20°C ready for use.

5.2.2 Evaluation of *rpoB, katG* and *oxyS* primers

The simplest way to calculate the number of genomes in a DNA extract using real-time PCR was to quantify the number of single copy genes in a sample. Unlabelled forward and reverse primers were designed for three different genes, spanning the *M. tuberculosis* genome, which existed as single copies. As all real-time PCR was performed using the iQ5 Multicolor Real-Time PCR Detection System (Bio-Rad Laboratories, California, USA), primers were designed for *rpoB*, *katG* and *oxyS* genes

using the guidelines in the instruction manual. The primer sequences were: *rpoB* (forward: 5'-CCG CGA TCA AGG AGT TCT TC-3'; reverse: 5'-GCC GAT CAG ACC GAT GTT G-3'), *katG* (forward: 5'-GAC AAG GCG AAC CTG CTT AC-3'; reverse: 5'-CCC AGG TGA TAC CCA TGT C-3') and *oxyS* (forward: 5'-CGA AGG TGC AAG TCT GTT CC-3'; reverse: 5'-CAT CAC TGT CTT GGG TCT CG-3'). Primers were supplied by Invitrogen Ltd. and their performance evaluated.

The *rpoB*, *katG* and *oxyS* genes were amplified in separate reactions using the DNA standards, prepared in Section 5.2.1, ranging from 1×10^1 to 1×10^7 genomes/µl in 10-fold increments, as the template. The reaction for each primer set and DNA standard was prepared in triplicate in iQ 96-Well PCR Plates (Bio-Rad Laboratories) at a final volume of 20µl, which contained 10µl 2× iQ SYBR Green Supermix (Bio-Rad Laboratories), 0.1µl each of 200µM forward and 200µM reverse primers, 1µl DNA and 8.9µl molecular grade water. The SYBR Green Supermix contained 100mM potassium chloride, 40mM Tris-hydrochloric acid (pH8.4), 0.4mM of each dNTP, iTaq DNA polymerase (50 units/ml), 6mM $MgCl_2$, SYBR Green I, and 20nM fluorescein. Non-template controls were also set up for each primer set, using 1µl Buffer AE instead of DNA.

Reproducibility of the values obtained from the replicates was optimised by preparing master mixes of reagents wherever possible to reduce the amount of pipetting. Also, when dispensing the reagents into the wells of the PCR plate, it was important not to produce bubbles as they would affect the fluorescence readings during real-time PCR. Evaporation during amplification was reduced, as the tiniest change in volume could affect the readings, by using a clear Microseal 'B' Adhesive Seal (Bio-Rad Laboratories) to seal the plate being careful not to get any fingerprints on the seal as the fluorescence is read from the top through the seal. The reactions in the wells were spun down by centrifuging the plate at 250×*g* for 1 minute.

The following cycle was used for real-time PCR: 95°C for 180 secs; 50 cycles of 95°C for 30 secs, 65°C for 30 secs; 55°C for 60 secs. The cycle was programmed so fluorescence readings were acquired at the end of each of the 50 cycles because as the DNA is amplified the number of fluorescent labelled PCR product increases and this increase can be plotted using the fluorescence values. After the 55°C step, an additional step was inserted to obtain a melt curve. This step involved a 0.5°C

increase from 55°C to 95°C and at each increment a fluorescence reading was acquired.

After real-time PCR was complete the iQ5 Optical System Software version 2.0 (Bio-Rad Laboratories) was used to analyse the fluorescence readings. Graphs were automatically plotted for the amplification and melt-curve steps. Threshold cycle (Ct) values were automatically calculated by the software for all reactions and indicated the point at which there was enough DNA amplification to produce a detectable amount of fluorescence from the PCR product. Hence, if there was a lot of DNA in the initial reaction the Ct value would be lower as fewer amplification cycles would be needed for fluorescence to reach the detectable level. Conversely, if the DNA concentration was initially lower it would take more cycles to produce the same detectable level of fluorescence. The melt-curve graphs indicated whether or not there was a single PCR product in the reactions, which would be indicated by a single peak at the temperature when the double-stranded DNA separated. The temperature at which this occurred was dependent on the size of the DNA.

The graphs that were plotted during the amplification and melt-curve steps for each of the primers can be seen in Figure 5.1. There were clear differences in the quality of the primers that were used to amplify each of the genes. The replicates for *rpoB* were well defined so the average Ct value of replicates was calculated reliably. The curves representing the more concentrated DNA standards were separated from each other at regular intervals and this was corroborated numerically by the fact that there was a difference of 4 between Ct values for each DNA concentration.

Although the difference between Ct values for the more concentrated DNA standards was similar for the *katG* and *oxyS* primers, the curves for the replicates were not grouped together as clearly as for *rpoB*. This meant that the calculated average Ct value when using *katG* and *oxyS* primers was not reliable. The replicates for the DNA standard containing 1×10^1 genomes/µl were not reproducible when using all of the primer sets for amplification.

The graphs for melt-curves were good for all of the primer sets as there were clearly single peaks at the same temperature. The *rpoB* PCR products separated at approximately 92°C, *katG* products at approximately 89°C and *oxyS* products at

approximately 88°C. The differences in temperatures reflected the sizes of the PCR products as the amplified *rpoB* fragments were larger than the *katG* and *oxyS* products. The standard curves generated with the serially diluted standard DNA was linear over the data points with gradients of 3.79, 3.54, and 3.57 for primer sets *rpoB*, *katG* and *oxyS*, respectively, as shown in Figure 5.2. The efficiency of all primers was over 90%, which is supported by the coefficient of correlation (R^2) values that were above 0.99 for all standard curves. Similar R^2 values were observed in previous studies evaluating the use of real-time PCR using other primer targets (Desjardin *et al.,* 1998; Rondini *et al.,* 2003; O'Mahony *et al.,* 2004).

It was decided to use the forward and reverse primers for the *rpoB* gene for quantification of mycobacteria because the replicate reactions were more reproducible and the small difference in Ct values between the DNA concentrations meant that the assay would be more sensitive. Even though the replicates for lower DNA concentrations were not reproducible, the concentration of DNA that would be used for subsequent experiments would be higher, therefore this would not pose a problem. The quality of the standard curve, when using *rpoB* primers in the assay, further confirmed the decision to use these primers.

Using primers to amplify a gene that was present in the *M. tuberculosis* genome as a single copy meant that quantifying the number of *rpoB* copies in a DNA extract would be equivalent to quantifying the number of genomes. However, before using real-time PCR for quantifying the number of mycobacteria in cultures that would be used in phenotypic experiments, the technique had to be validated against current quantification methods, OD_{600} and c.f.u., both of which are influenced by the clumping of mycobacteria.

5.3 COMPARISON OF MTBC QUANTIFICATION METHODS

5.3.1 Quantification of MTBC cultures using OD_{600}

An archive tube of frozen culture for each of the 5 *M. tuberculosis* strains in Panel B (refer to Section 2.4) was thawed at room temperature and 2 drops were used to inoculate two 50ml falcon tubes, containing 30ml Middlebrook 7H9 with 0.05% Tween-80 medium and beads in each, so there were duplicate liquid sub-cultures for

each strain (refer to Section 2.4.1). After vortexing, the cultures were incubated at 37°C.

The day of inoculation was day 0. At day 3, 1ml of each culture was transferred into separate disposable plastic cuvettes (Alpha Laboratories Ltd.) and OD_{600} was measured with a cell density meter (WPA Biowave, Cambridge, UK), using 1ml medium as the blank to tare the density meter. The OD_{600} was plotted on a graph of time (days) versus OD_{600} (Figure 5.3a). The OD_{600} was measured regularly, every 3 or 4 days until the growth curves for each of the strains reached the stationary phase indicated by the plateux of the bacterial growth curve.

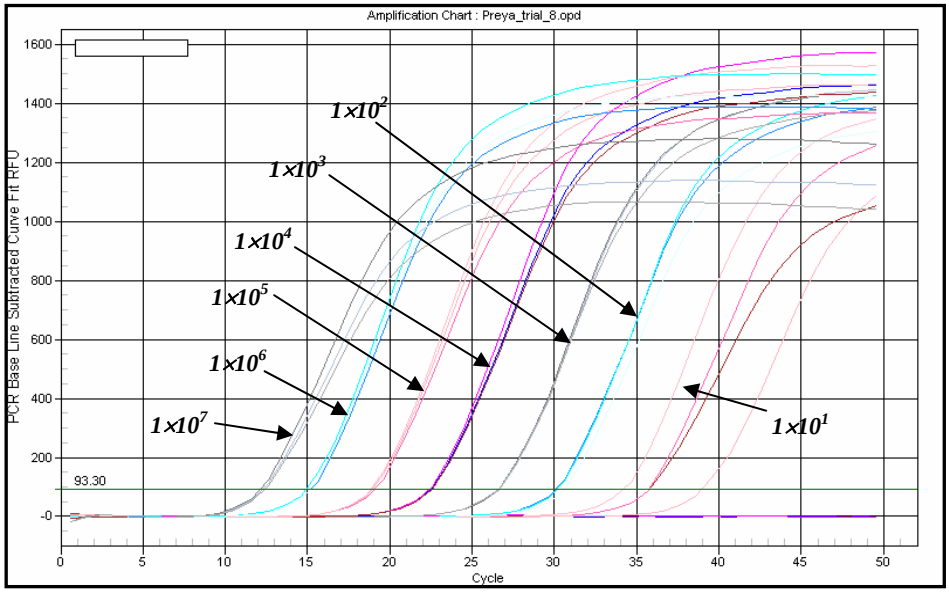
5.3.2 Quantification of MTBC cultures using real-time PCR assay

Each time the OD_{600} was measured at the various time points, a separate 1ml aliquot of each culture was transferred into a 1.5ml microcentrifuge tube. The bacterial cells were pelletted by centrifugation at 12,000×*g* for 10 mins and the supernatant was safely discarded. The pellet of mycobacteria was resuspended in 100µl TE buffer. The DNA extraction and purification procedure was as described in Section 5.2.1 except the volume of Buffer AE for eluting DNA was reduced to 30µl as the initial pellet of bacterial cells was much smaller. The DNA was stored at -20°C ready for quantification with real-time PCR.

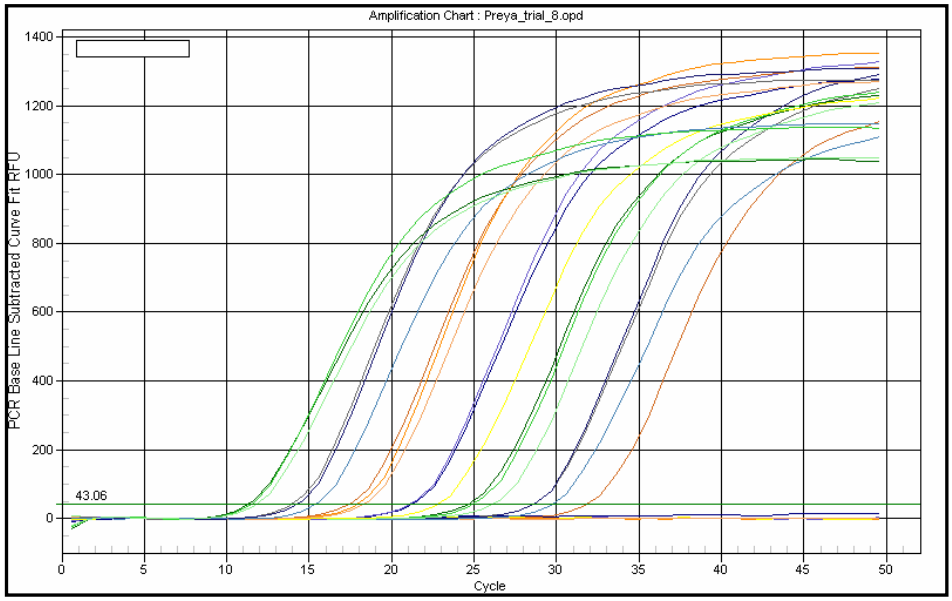
Using the standards prepared in Section 5.2.1, ranging from 1×10^1 to 1×10^7 genomes/µl, and the DNA of unknown concentration, real-time PCR reactions were set up using the *rpoB* primers exactly as described in Section 5.2.2. Just like for the standards and non-template control, three replicate reactions were prepared for each of the unknown samples. After the PCR and melt curve was complete, iQ5 software automatically analysed the data to calculate the starting concentration of genomes/µl in the unknown samples by extrapolating the Ct value of unknown samples against those of the DNA standards. The DNA standards were included in each real-time PCR run, as the Ct value of DNA standards in a previous run could not be used for extrapolation in subsequent runs.

Figure 5.1a *Graphs produced by the iQ5 Optical System Software after analysing fluorescence readings obtained during real-time PCR when using DNA standards ranging from 1×10^1 to 1×10^7 genomes/µl as templates and each of the three primer sets for the genes, i) rpoB, ii) katG and iii) oxyS. For each set of primers and DNA standards there were 3 replicate reactions. The replicate curves for each of the DNA standards are grouped more pronouncedly when using rpoB primers than the katG and oxyS as indicated in Figure 5.1a i).*

i)



ii)



iii)

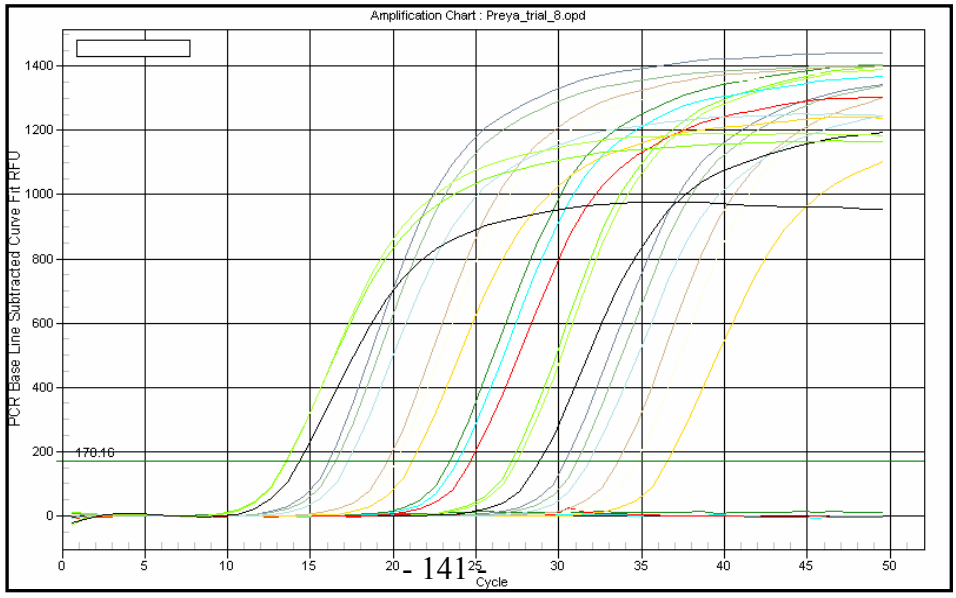
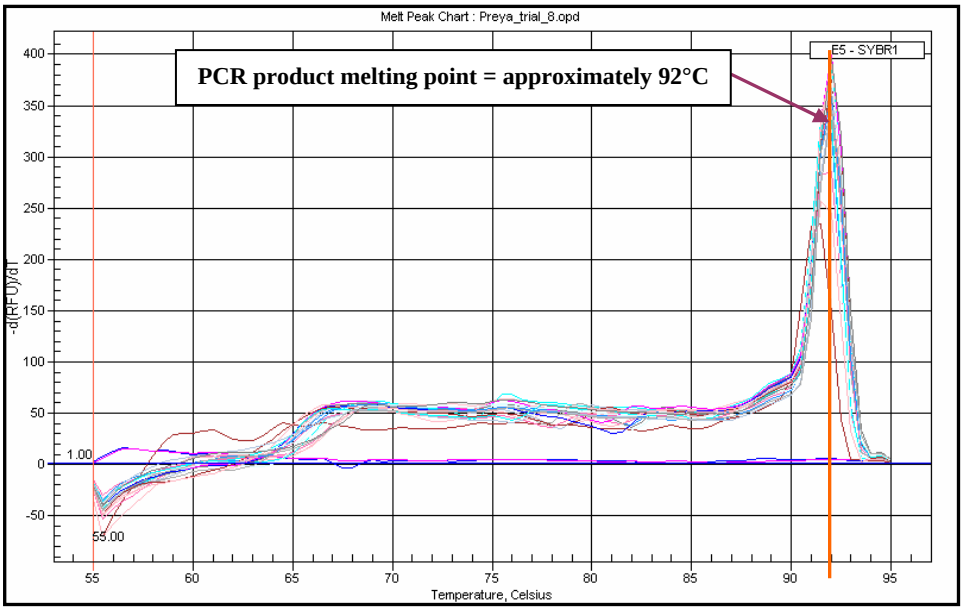
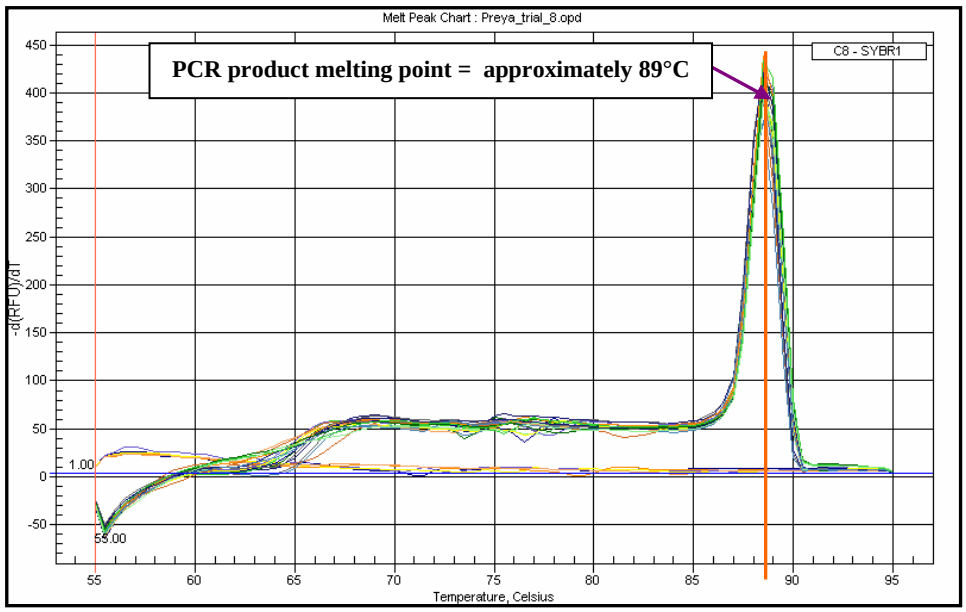


Figure 5.1b *Corresponding melt curves, to the graphs in Figure 5.1a, produced by the iQ5 Optical System Software, for PCR products, which had been amplified using DNA standards ranging from 1×10^{1} to 1×10^{7} genomes/µl in 10-fold increments as templates and each of the three primer sets for the genes, i) rpoB, ii) katG and iii) oxyS.*

i)



ii)



iii)

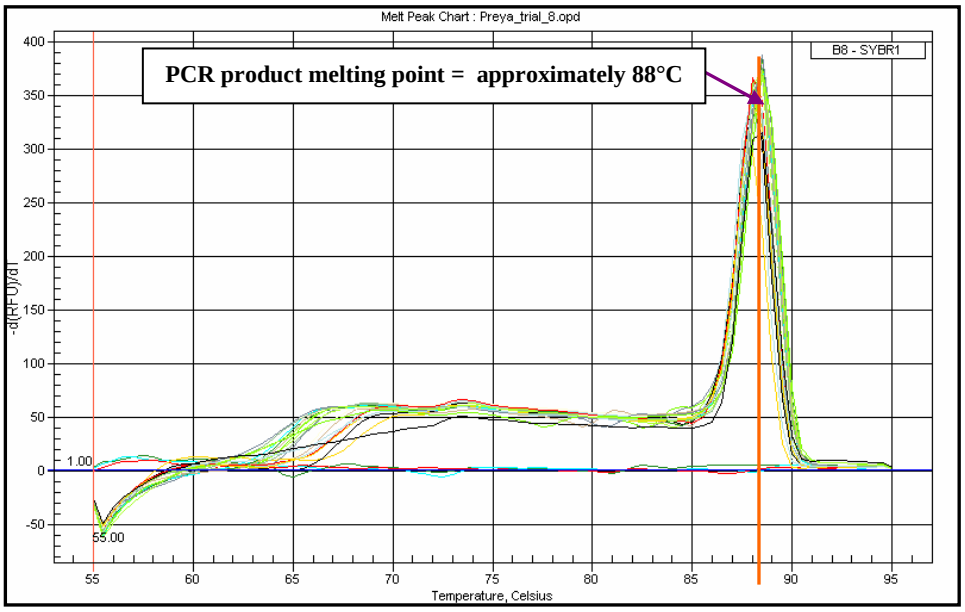
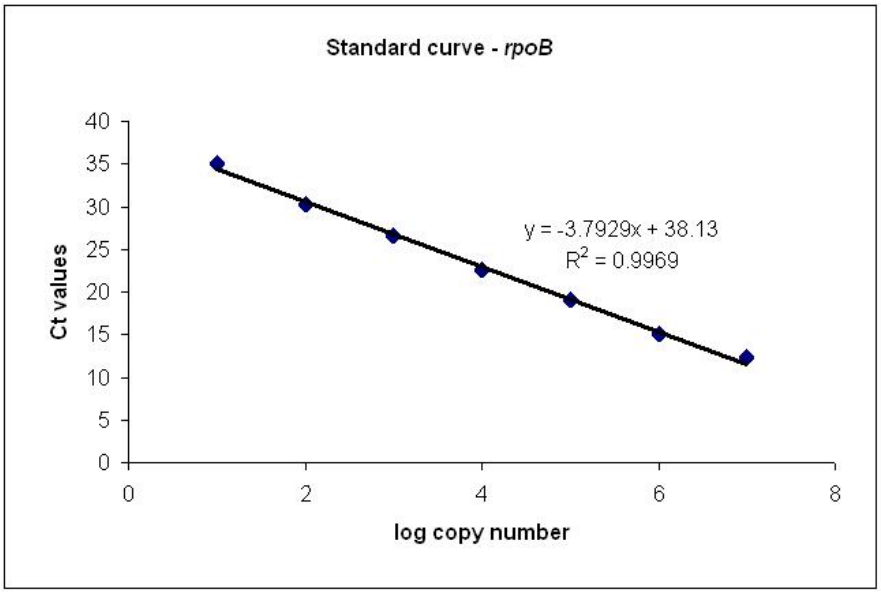
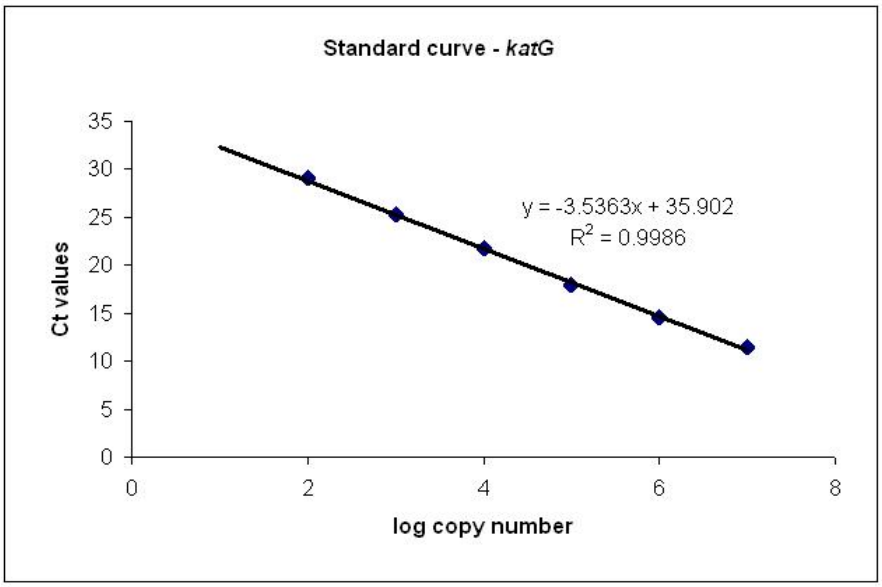


Figure 5.2 *The standard curves generated using Ct values after analysing DNA standards, containing known genome numbers, with real-time PCR assay using each primer pair for (a) rpoB, (b) katG and (c) oxyS. The standard curves also display the calculated regression line for each set of data points and the coefficient of correlation (R^2) for each standard curve.*

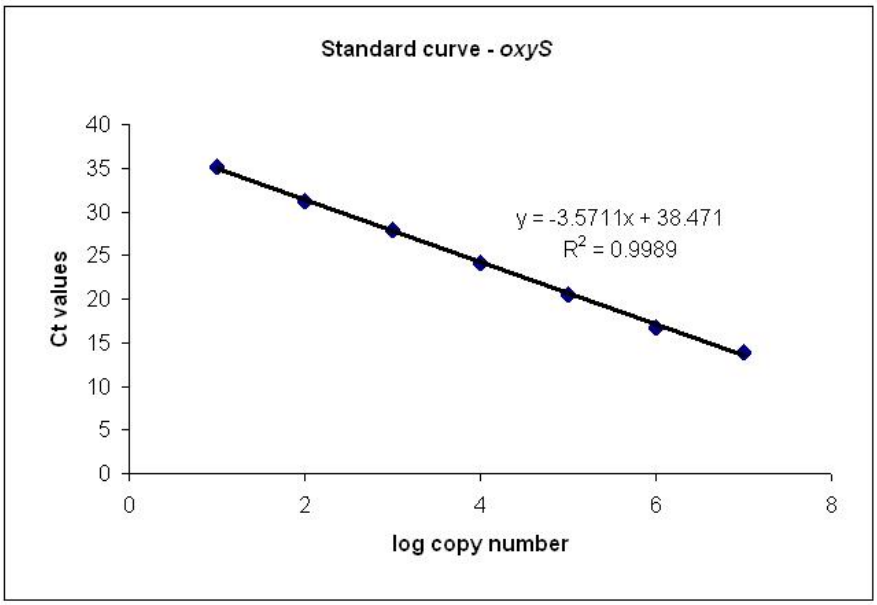
a)



b)



c)



The number of genomes/ml culture was calculated by multiplying the number of genomes/µl by the volume Buffer AE used for DNA elution. As the DNA had been extracted from 1ml culture initially, the calculation provided the number of genomes/ml culture. The calculated values were plotted on a graph of time (days) versus number (N^{o}) of genomes/ml (Figure 5.3b).

5.3.3 Quantification of MTBC cultures using c.f.u.

In addition to the OD_{600} and number of genomes data, c.f.u. data were also collected at each time point. Whilst taking aliquots of culture for reading OD_{600} and for DNA extraction, two separate 30µl aliquots, from each culture were diluted 1 in 10 in separate 1.5ml microcentrifuge tubes each containing 270µl Middlebrook 7H9 with 0.05% Tween-80 medium (without OADC). The contents of the tubes were mixed carefully by pipetting, before making five further 10-fold serial dilutions. As two 30µl aliquots were taken from each culture initially, there were two sets of dilutions and ultimately duplicate c.f.u. counts at every dilution for each culture. A volume of 50µl of each diluted culture was transferred onto half of a Middlebrook 7H11 agar plate and then carefully spread in that half of the plate. The agar plates had been prepared as described in Section 2.1.3. After inoculating the plates and letting them dry, all plates were sealed with parafilm, placed in a sealed bag and incubated at 37°C for 3 to 4 weeks depending on when colonies could be seen.

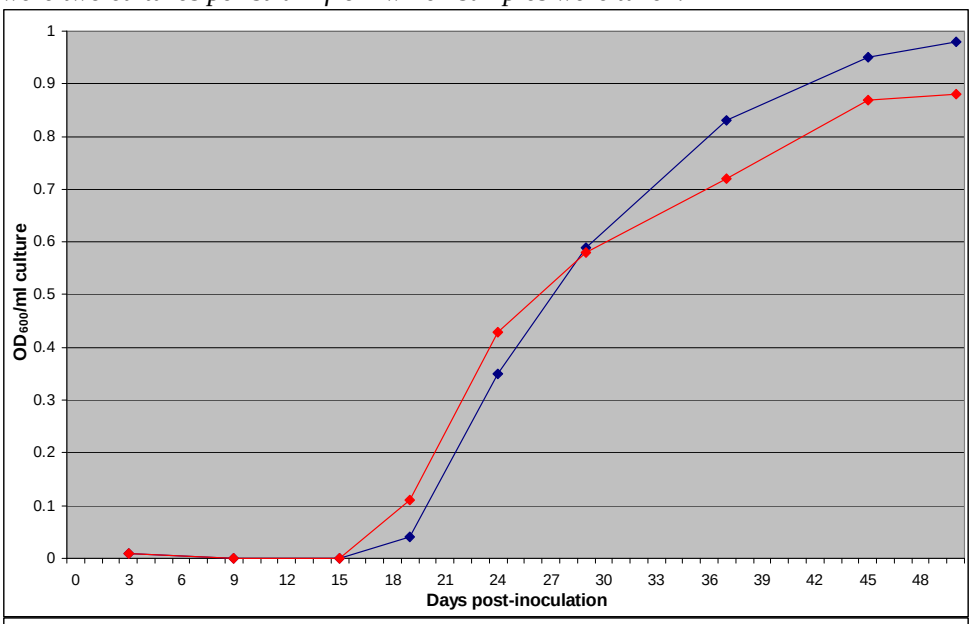
The colonies were counted on the plates and recorded. For each culture and each time point, the average number of c.f.u. was calculated using the data from a dilution factor at which there was the most reliable c.f.u. count (10-200 colonies). The number of colonies/ml culture was calculated by multiplying the average c.f.u. count by 20 (as 50µl of diluted culture was plated) and this value was multiplied by the dilution factor. The resulting values were plotted on a graph of time (days) versus N^{o} of colonies/ml (Figure 5.3c).

5.3.4 Results of comparing MTBC quantification methods

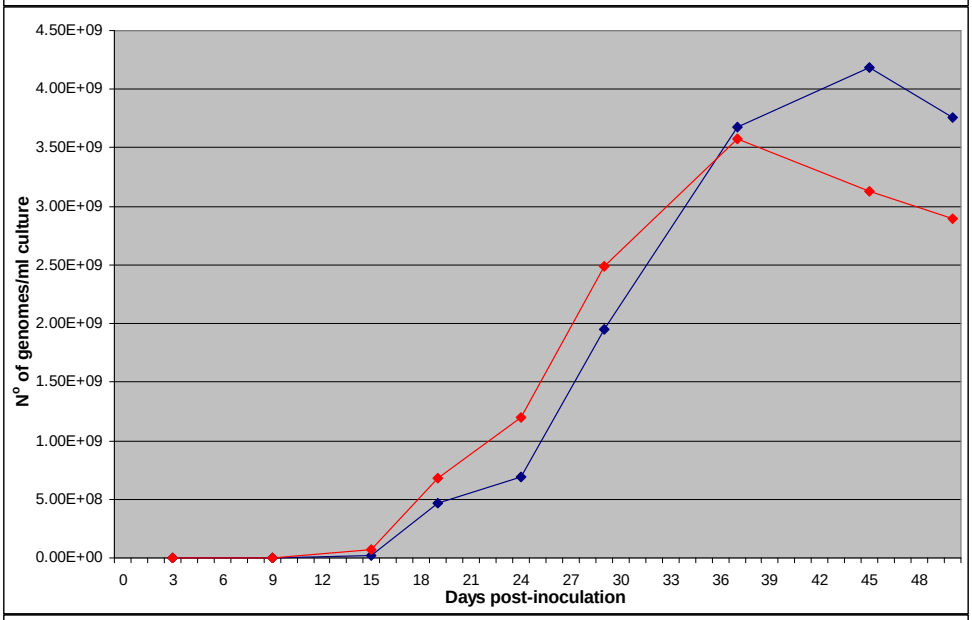
The three growth curves produced when plotting OD_{600}/ml culture, number of genomes/ml culture, and number of colonies/ml culture at each sampling day has been shown in Figure 5.3, for one of the Panel B strains, *M. tuberculosis* CAS, isolated from a patient born in India, as an example as the curves were similar for the other four strains in Panel B.

Figure 5.3 *The growth curves produced for M. tuberculosis CAS strain, isolated from a patient born in India, using the three different quantification methods including (a) OD_{600}, (b) number of genomes and (c) number of colonies. The blue and red curves represent each of the replicates, as there were two cultures per strain from which samples were taken.*

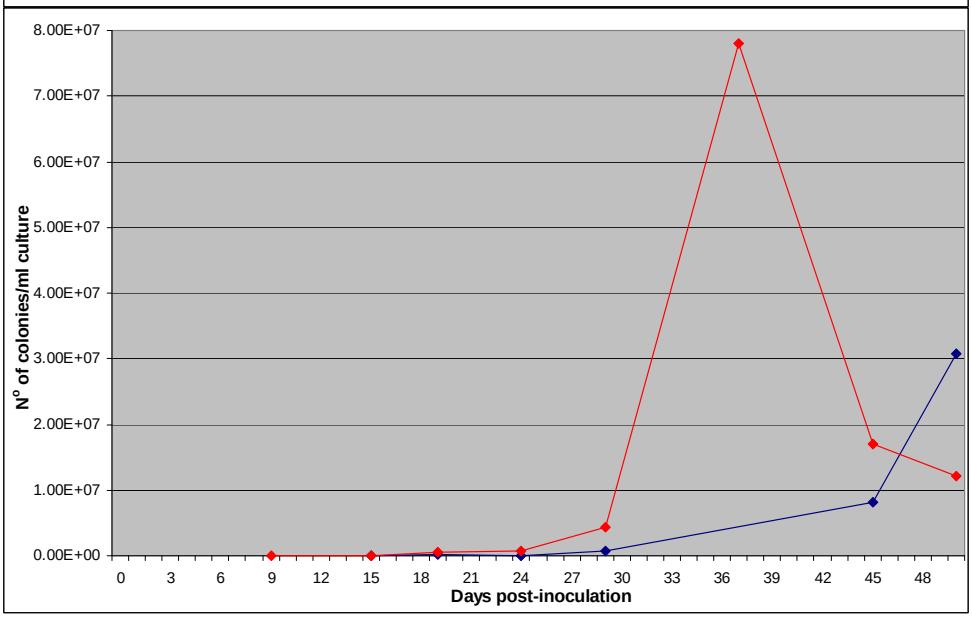
a)



b)



c)



As expected, the growth curve for OD_{600} for all of the strains followed the typical pattern with the lag phase at the beginning and then the exponential phase, followed by the plateaux indicating the mycobacteria had reached stationary phase. The same lag phase and exponential phase was seen when plotting the number of genomes/ml culture, but after plateaux, the typical horizontal line was not seen. Instead, just as the curve started to plateaux, the line became more haphazard with the points being up and down. This was observed with the genomes/ml culture growth curves for all of the *M. tuberculosis* strains.

The c.f.u. data from this experiment did not produce the complete standard bacterial growth curves as was observed with the OD_{600} and real-time PCR data. For some sampling points the data did not fit into the typical growth curve pattern. This was the case for each set of duplicate cultures for each *M. tuberculosis* strain. According to the OD_{600} and real-time PCR data the mycobacterial cells reached exponential phase by day 8 or 9 for each set of duplicate cultures for each of the strains; but the c.f.u. data for all of the cultures did not reflect this. For *M. tuberculosis* Beijing strain, cells entered the exponential phase at day 16, and were still actively dividing at day 19, but after this time point there was a sudden decrease in c.f.u. count, which was not observed with the curves produced using OD_{600} and real-time PCR data. This seemed to be a general trend with the c.f.u. data for all of the strains as the *M. tuberculosis* LAM10 mycobacterial cells entered the exponential phase at day 18 and while at day 23 the c.f.u. count increased, there was a sudden decrease at day 28 followed by another increase from day 36 to 49.

The *M. tuberculosis* CAS mycobacterial cells (isolated from the patient born in Somalia and India) entered the exponential phase at a much later stage (day 28 and 29, respectively) than the other strains. The cells were still actively dividing up until day 37 and 44, respectively, after which there was a decrease in the c.f.u. count. The growth curves produced using the c.f.u. data for *M. tuberculosis* EAI5 was similar to that of the *M. tuberculosis* CAS strain, isolated from a patient born in India, shown in Figure 5.3c, except mycobacterial cells entered the exponential phase slightly earlier at day 19 and continued to actively divide until day 29. After this time point there was a decrease in c.f.u. count instead of the typical plateaux seen with the growth curves produced using the OD_{600} and real-time PCR data.

For further validation of real-time PCR for mycobacterial quantification, any correlation between the genomes/ml culture data and the data from the other two quantification methods was investigated by plotting scatter graphs. The scatter graphs corresponding to the data for *M. tuberculosis* CAS, from the patient born in India, can be seen in Figure 5.4. There was good correlation between OD_{600}/ml culture and genomes/ml culture, as there was a linear relationship between the two measurements that could be defined by an equation (Figure 5.4a). Although there were linear relationships between OD_{600} and genomes/ml for the other strains, the equations were slightly different. As expected, after the lack of the typical growth curve pattern with c.f.u. data, there was no correlation between number of colonies/ml culture and genomes/ml culture for any of the strains (Figure 5.4b).

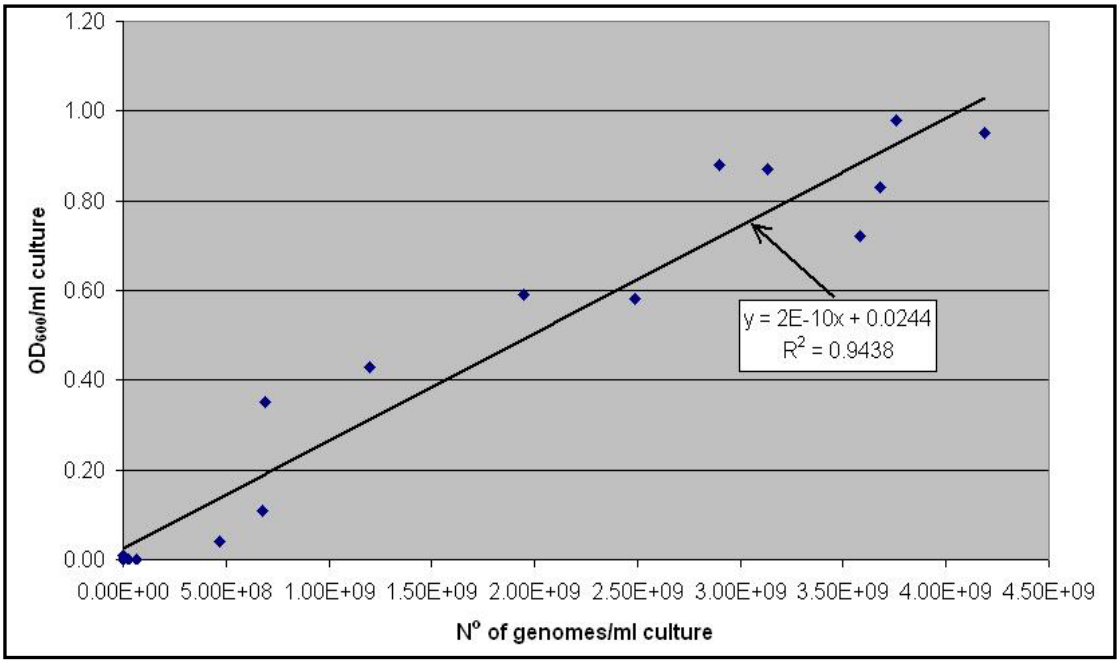
5.4 DISCUSSION

As the present study was focused on comparing the effects of the Panel B mycobacterial cells in *in vitro* and *in vivo* models, it was important to ensure that the same numbers of mycobacterial cells of each strain was used for infections to obtain comparable data and to exclude mycobacterial cell numbers used for experiments as a variable. Usually, c.f.u. count is used to quantify the numbers of mycobacterial cells, but cells tend to clump together meaning individual bacilli are not necessarily quantified. In addition, as mycobacteria are slow growing bacteria, it would take at least 3 weeks for colonies to grow on plates therefore delaying the set-up of experiments. For the purposes of the present study, it was necessary for cells to be actively growing in the exponential phase of the bacterial growth curve.

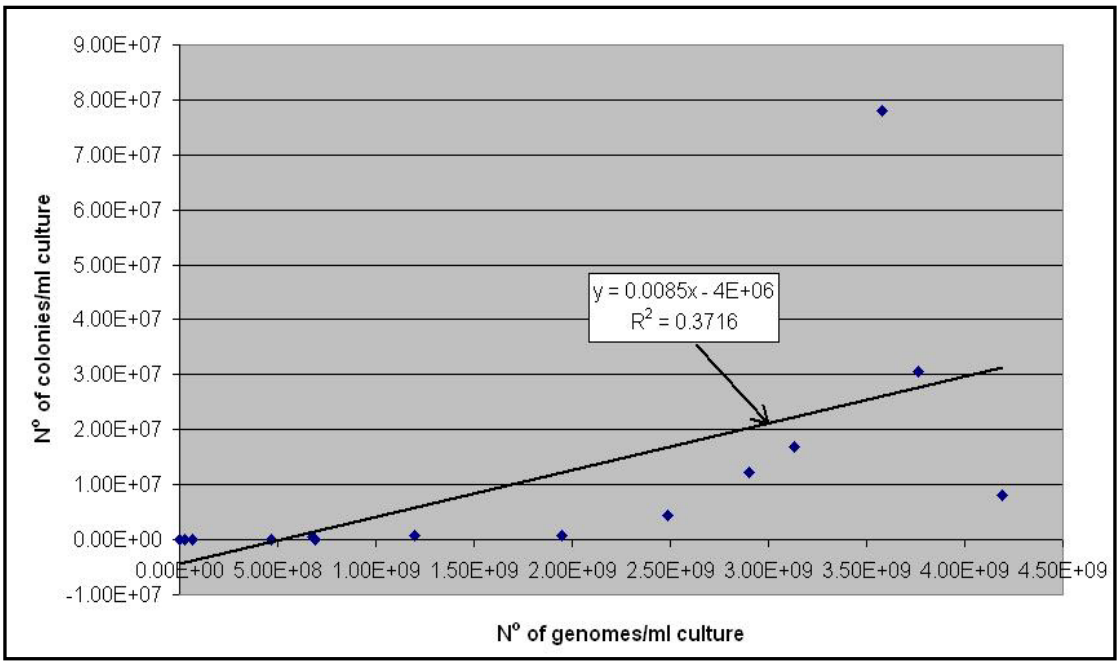
It was, therefore decided to investigate a more accurate and quicker method of real-time PCR to quantify the number of individual mycobacteria and to be able to set up phenotypic assays on the same day as quantification. Cultures of the Panel B *M. tuberculosis* strains were prepared and at regular intervals aliquots of the culture were quantified using real-time PCR to quantify number of genomes, c.f.u. count by plating cultures onto Middlebrook 7H11 plates and finally by measuring OD_{600}. The data from the latter two methods were used to validate the real-time PCR data against as these two methods have previously been used for quantifying bacteria.

Figure 5.4 *Graphs showing the correlation between (a) number of genomes/ml culture and OD_{600}/ml culture and (b) number of genomes/ml culture and number of colonies/ml culture for M. tuberculosis CAS strain, isolated from a patient born in India.*

a)



b)



The similarity between curves produced using the real-time PCR and OD_{600} data demonstrated that the lag phase and exponential phase of the bacterial growth curve could be followed using quantitative data from real-time PCR (Figure 5.3a and b). However, once mycobacteria reached the stationary phase, the growth curves produced using real-time PCR data, for all of the *M. tuberculosis* strains, did not follow the plateaux seen in typical growth curves. A possible explanation for this was that the cultures were getting old and there was mycobacterial cell death and degradation of mycobacterial DNA. For the purposes of carrying out phenotypic experiments in the present study, this was not significant, as actively growing and dividing mycobacteria, which would be in the exponential phase, would be required.

For this particular experiment, it was not as easy to validate the real-time PCR data against the c.f.u. data, as it had been with the OD_{600} data, because there was a sudden decrease in the c.f.u. count for all of the *M. tuberculosis* strains at some time points during the experimental time course. A possible explanation for this could be loss of viable mycobacterial cells in the cultures, but this is not corroborated by the real-time PCR data, which quantified the number of genomes. There was an increase in the number of genomes during the exponential phase of the growth curve, which was reflected by the growth curves produced using the OD_{600} data. However, as mentioned in Section 2.6.6, it is important to note that as with OD_{600}, quantifying the number of genomes does not discriminate between viable and dead mycobacteria therefore measuring DNA levels is not a measure of viability.

In addition, whilst there was a decrease in c.f.u. count with the *M. tuberculosis* LAM10 strain, there was an increase in the number of c.f.u. at the next sampling time points, demonstrating that there was no loss of viability in this culture. Another explanation could be an insufficient concentration of growth supplement (OADC) in the Middlebrook 7H11 plates as fresh plates were prepared at each sampling point or a need to improve the techniques for c.f.u. counting of mycobacteria.

There was good correlation between the OD_{600} and real-time PCR data for all of the *M. tuberculosis* strains and the growth curves produced using the real-time PCR quantification data followed the typical trend of a bacterial growth curve. When using OD_{600} and real-time PCR all viable and non-viable mycobacteria would have been included for quantification so both of these methods would be more ideal for bacterial

cells in the exponential phase of the bacterial growth curve. Converse to this, c.f.u. counting of bacterial cells only quantifies the viable cells in the culture.

For the purposes of the present study, whilst an OD_{600} of 0.2 was used as a marker of mycobacterial cells entering the exponential phase of the growth curve, as can be seen in the OD_{600} growth curve in Figure 5.3a, real-time PCR was used as a concurrently more rapid method for quantifying MTBC cultures for further *in vitro* and *in vivo* experiments. In order to avoid introducing variables into the experiments it was ensured that exactly the same DNA extraction and real-time PCR protocols were used for every independent experiment.

CHAPTER 6

IN VITRO PHENOTYPICAL ANALYSIS OF *M. TUBERCULOSIS* COMPLEX STRAINS

6.1 INTRODUCTION

Whilst the aim of the present study was to create a panel of *M. tuberculosis* strains for general TB research and vaccine evaluation studies, the preliminary panel A (containing 42 strains including a broad spectrum of *M. tuberculosis* strains and representatives from global population of strain families; refer to Sections 2.3.2 and 3.2.4), would have been too large for vaccine evaluation studies, therefore a smaller number of strains were selected for final analysis. As the final panel should ideally contain strains of different ranges of phenotypes and virulence, it was important to initially establish if there were any phenotypic differences between various circulating strains and also if there were any differences between circulating strains and laboratory strains, which are currently used for vaccine evaluation studies (Williams *et al.,* 2005; Vipond *et al.,* 2006).

A series of *in vitro* experiments were performed using six circulating strains (Panel B) selected from the five major lineages defined by Baker *et al*., 2004 and the *M. africanum* lineage identified by Gagneux *et al*., 2006, and four control strains including laboratory *M. tuberculosis* H37Rv from HPA, MRU and Porton Down, *M. tuberculosis* H37Ra from NCTC and vaccine *M. bovis* BCG (Baker *et al.,* 2004; Gagneux *et al.,* 2006).

The *in vitro* phenotypic experiments aimed to investigate variation in growth of strains in cell-free and *in vitro* tissue culture systems and other parameters such as cytokine production. Depending on the data from the experiments, virulence factors might be identified that would eventually play a very influential role during selection of strains for vaccine evaluation studies. Growth rates of all ten strains were obtained using a higher throughput MGIT 960 system (which gives better reproducibility of viable growth) and this data was compared with growth in an *in vitro* tissue culture system.

Currently, in the HPA, MRU the protocol for the MGIT 960 system provided by the manufacturer involves inoculating MGIT tubes with a drop of culture calibrated visually using a McFarland turbidity standard and using Pasteur pipettes; however, this means that the exact number of bacilli being used for inoculation is unknown. So as well as using the growth rate data to identify any phenotypic differences between strains for the purposes of the present study, the data provided additional insight into the effects of inoculum sizes on the general growth rates of mycobacterial cells in a cell-free culture system.

6.2 THE GROWTH RATES OF *M. TUBERCULOSIS* COMPLEX STRAINS IN MGIT 960 CULTURE SYSTEM

6.2.1 The effect of inoculum sizes on growth rates

In order to reliably relate growth rate data obtained when using the MGIT 960 culture system with results obtained from tissue culture experiments, the same cultures were used to set up both experiments on the same day. This meant there was growth rate data available from three independent experiments. For each strain, duplicate MGIT 7ml tubes were inoculated with each of the inoculum sizes, ranging from 6 genomes (real-time PCR had been used to quantify mycobacteria hence denotation of inoculum size as genomes) to 600,000 genomes in 10-fold increments. The tubes were incubated in the MGIT 960 system, which automatically produced growth curves real-time so it was easy to monitor when curves reached plateaux indicating that mycobacteria had reached the stationary phase. Figure 6.1a shows the growth curves produced by the MGIT 960 system automatically and Figure 6.1b shows the growth curves, produced using the raw numerical data from the MGIT 960 system, which were then used to calculate the growth rate at mid-logarithmic phase.

The average growth rate of duplicates at each inoculum size was calculated for the laboratory control strain, *M. tuberculosis* H37Rv (HPA, MRU) to identify which inoculum size produced the most reliable growth curves according to the manufacturer's recommendations. A graph showing the average growth rates at each of the inoculum sizes for *M. tuberculosis* H37Rv (HPA, MRU) is shown in Figure 6.2.

As the inoculum size increased the growth rate increased, reaching a peak of 310 growth units/hour at an inoculum size of 6,000 genomes, after which as the inoculum size increased further the growth rate decreased. The growth rates at the two extreme inoculum sizes of 600,000 and 6 genomes, was approximately three times slower than the peak growth rate. Ideally, the exponential phase of the growth curves generated by the MGIT 960 system should not commence too soon or too many days post-inoculation. According to the manufacturer's manual after inoculating MGIT tubes with *M. tuberculosis* culture, positivity should be achieved 6 to 10 days (144 to 240 hrs) post-inoculation. Referring to the growth curves in Figure 6.1b, the recommended positivity was achieved at inoculum sizes of 6,000 and 600 genomes. For this reason, the growth rates for all of the Panel B, laboratory control and vaccine strains was calculated at inoculum sizes of 6,000 and 600 genomes.

6.2.2 Growth rates of strains after inoculating MGIT tubes

At both of the inoculum sizes, 6,000 and 600 genomes, the fastest growth rate was observed with the *M. tuberculosis* Beijing strain (approximately 450 and 350 growth units/hour, respectively) (Figure 6.3). The *M. tuberculosis* CAS strain, isolated from the patient born in Somalia, had the slowest growth rate at an inoculum size of 6,000 genomes (the CAS strain grew approximately twice as slowly as the Beijing strain at 6,000 genome inoculum size). At an inoculum size of 600 genomes, *M. africanum* had the slowest growth rate (2.3 times slower than the Beijing strain at 600 genome inoculum size). At each of the inoculum sizes, there was an approximate 50 growth units/hour difference in growth rate between both of the CAS strains (isolated from the patients born in India and Somalia) and between the laboratory control strain *M. tuberculosis* H37Rv (from HPA, MRU and Porton Down), with the faster growth rates being observed with the CAS strain from India and the H37Rv strain from Porton Down.

Figure 6.1 *Graphs showing (a) an example of the growth curves produced by the MGIT detection system after inoculating MGIT tubes with 60 000, 6 000, 600, 60 and 6 genomes of M. tuberculosis H37Rv culture as indicated on the graph and (b) growth curves that were produced using raw numerical data from the MGIT detection system, to illustrate the effect of different inoculum sizes (600 000, 60 000, 6 000, 600, 60 and 6 genomes) of M. tuberculosis H37Rv culture on the bacterial growth curves and the time taken for mycobacterial cells to reach the exponential and stationary phases of the growth curves after inoculating 2 MGIT tubes (represented by 2 separate curves) with each of the inoculum sizes.*

a)

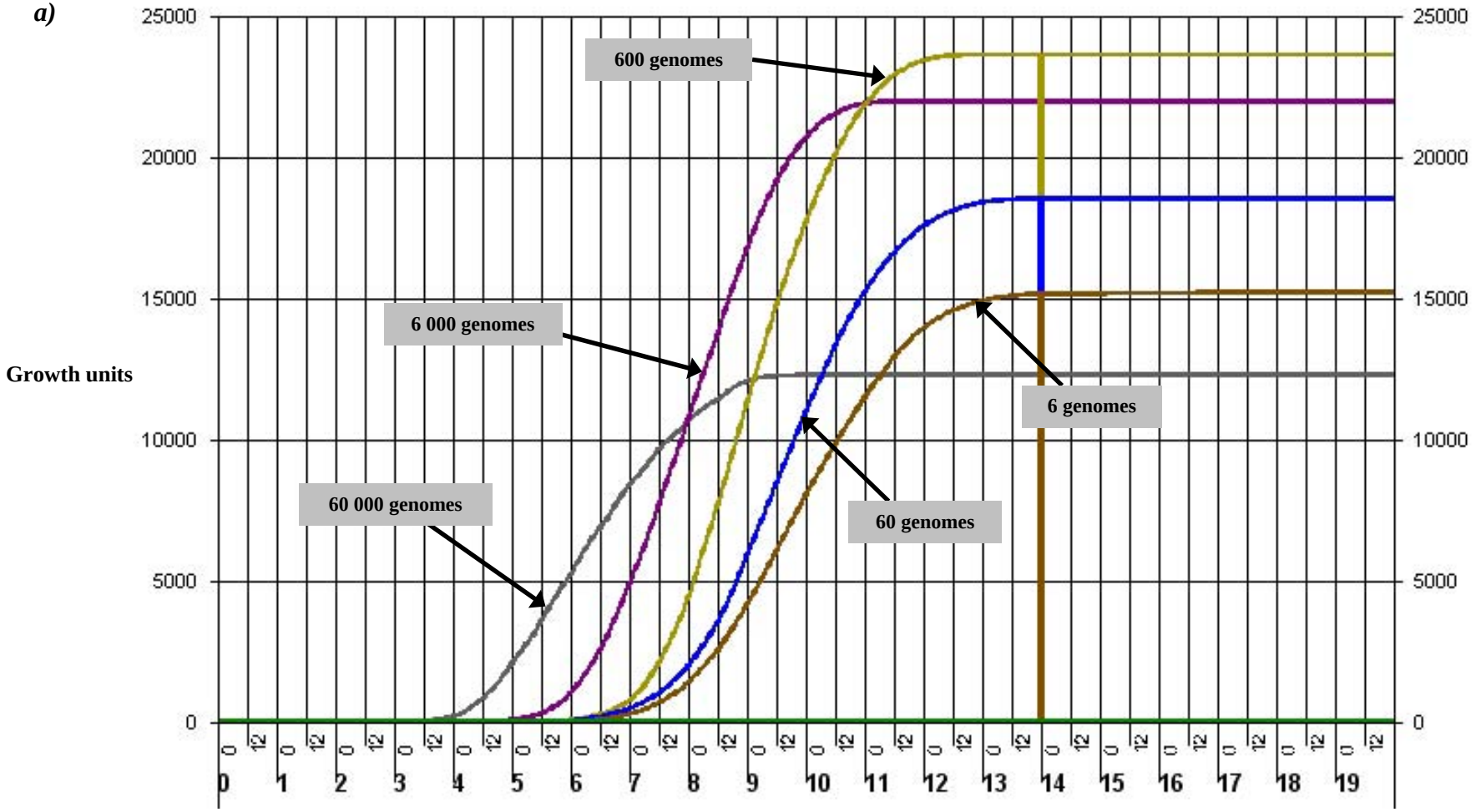


Figure 6.1 continued

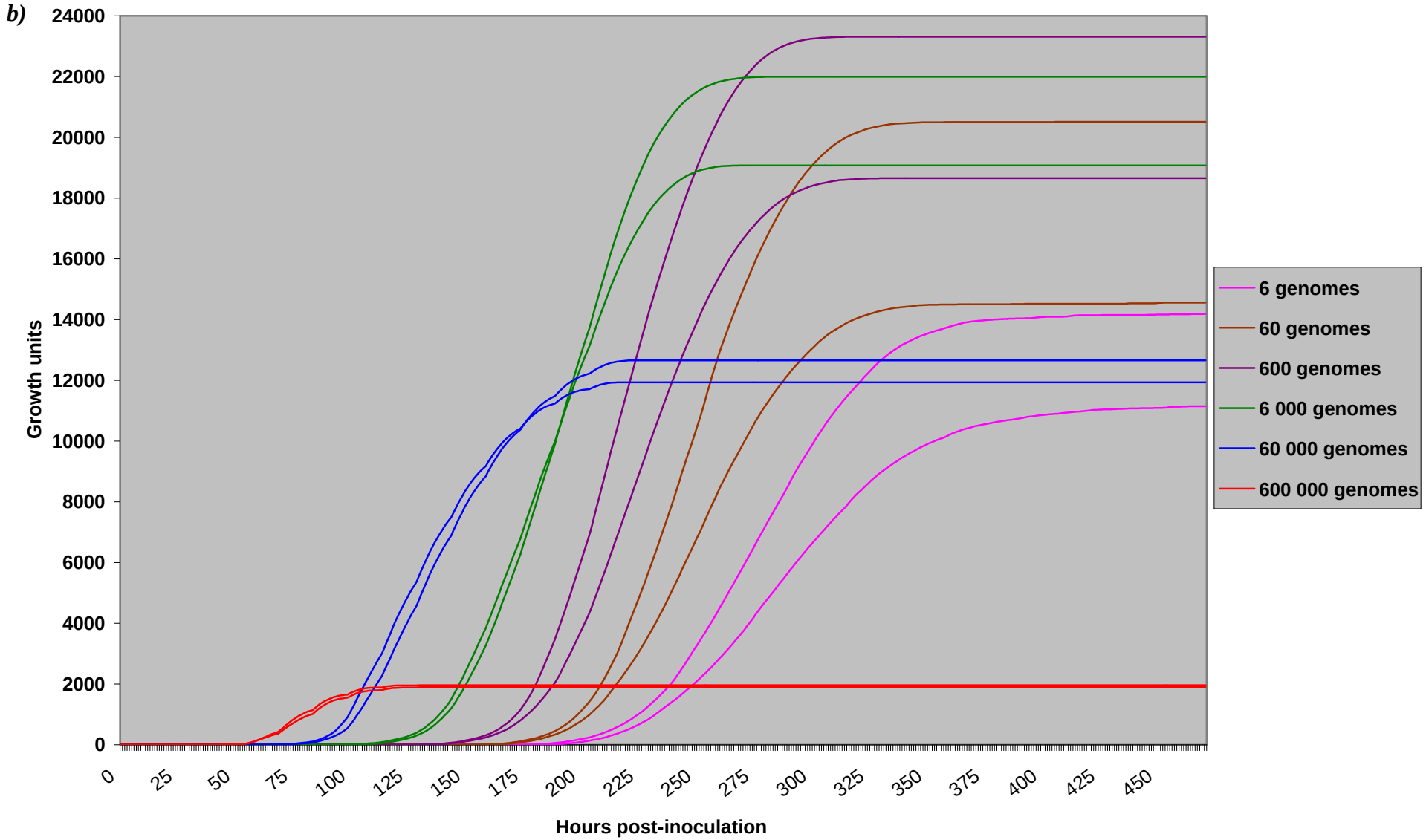


Figure 6.2 *The average growth rates (growth units/hour), of the three independent experiments, at the mid-logarithmic phase of growth curves produced by the MGIT 960 after inoculating MGIT 7ml tubes with different inoculation sizes of 600,000, 60,000, 6,000, 600, 60 and 6 genomes for M. tuberculosis H37Rv (from HPA, MRU).*

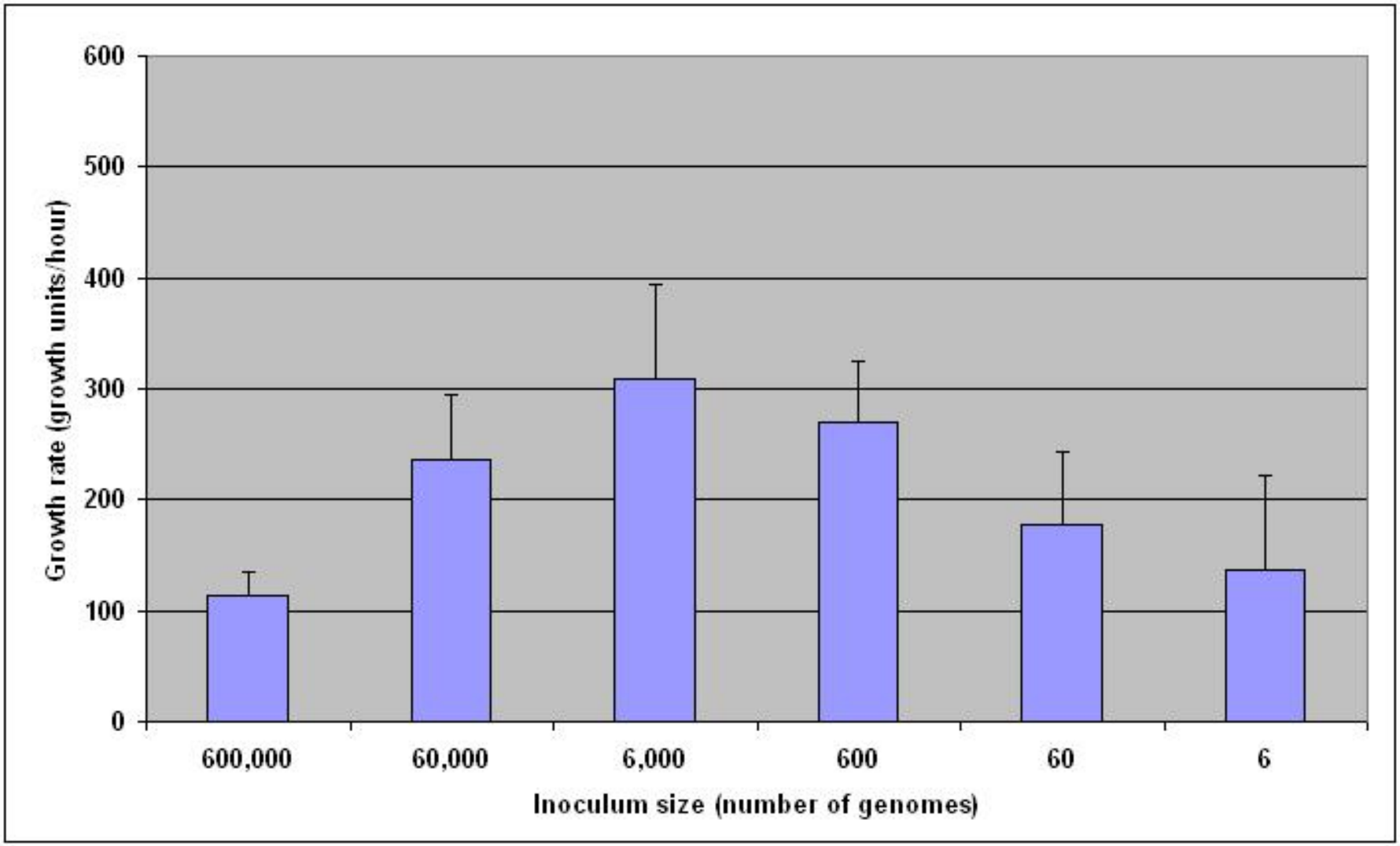
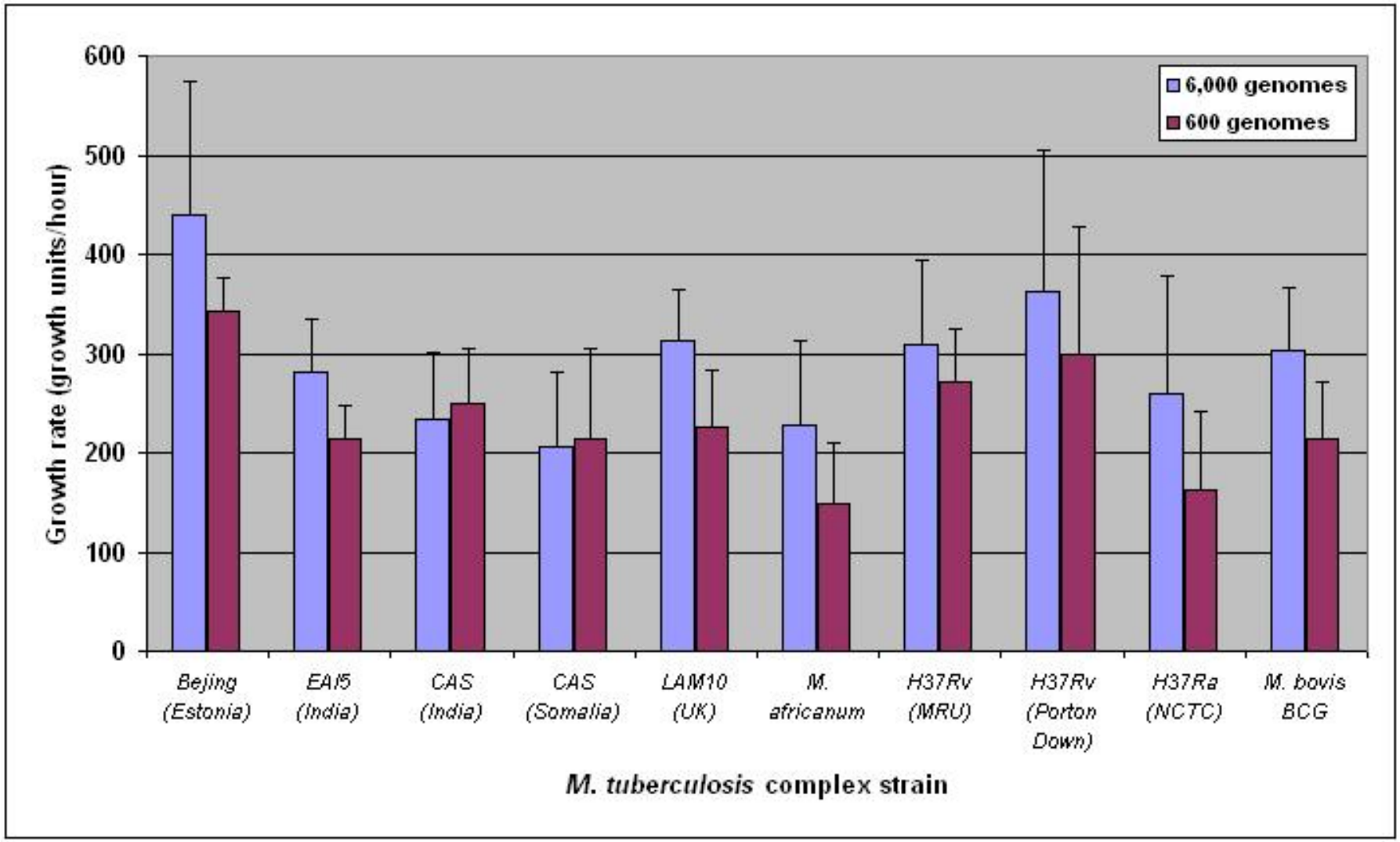


Figure 6.3 *The average growth rates (growth units/hour) for each of the Panel B, laboratory control and vaccine strains at inoculation sizes of 6,000 and 600 genomes.*



6.3 INFECTION OF THP-1 CELLS WITH *M. TUBERCULOSIS* COMPLEX STRAINS

6.3.1 Growth of TB strains in THP-1 cells after infection

For all strains, cultures from the same passage (passage 2) were used for tissue culture experiments that involved infecting PMA-activated, non-differentiating, macrophage-like THP-1 cells with mycobacteria, which had been grown to an OD_{600} 0.2 to ensure that they were actively growing at the time of infection, at an MOI of 1. Therefore 300,000 THP-1 cells were infected with 300,000 mycobacteria (quantified by real-time PCR). A passage number close to the original archive cultures (passage 2) was used to avoid any loss of virulence during culturing. Controls were also set up in which THP-1 cells were non-infected (Middlebrook 7H9 media was added instead of mycobacteria) and controls in which THP-1 cells were stimulated with LPS and IFN-γ.

For each strain and control, enough infections were set up so duplicate samples could be collected at day 1, 4 and 7 post-infection. Before day 1 sampling, any unincorporated mycobacteria were washed away so the effects of only intracellular mycobacteria were measured. Sampling involved collecting pellets of infected THP-1 cells (to measure the fold enhancement or multiplication of mycobacteria post-infection) and supernatant (to measure levels of cytokine production using ELISA assays. Three independent tissue culture infection experiments were set up.

DNA was extracted from all cells in the pellets and real-time PCR was performed using primers for actin to measure numbers of THP-1 cells and primers for *rpoB* gene to measure numbers of mycobacterial cells. These calculated levels, given by the Ct values during real-time PCR, were used to calculate the ratio of numbers of mycobacteria:THP-1 cells as apoptosis of PMA-activated THP-1 and human alveolar macrophages has been shown to differ between virulent and non-virulent strains (Keane *et al.,* 1997; Balcewicz-Sablinska *et al.,* 1998; Keane *et al.,* 2000; Riendeau *et al.,* 2003). Ratios of day 4:day 1 and day 7:day 1 were used as a statistic to demonstrate the fold enhancement or multiplication of mycobacteria after infecting THP-1 cells. Fold enhancement data was available from three independent experiments and is shown in Figure 6.4.

The lowest multiplication rate at both day 4 and 7 post-infection, relative to day 1 post-infection, was observed in THP-1 cells infected with the avirulent laboratory control strain, *M. tuberculosis* H37Ra. The fold enhancement values of both of the *M. tuberculosis* H37Rv strains was similar at day 4 and 7 but in comparison with H37Ra, whilst the fold enhancement at day 4 was the same, at day 7 the value was approximately 2 times higher. The multiplication rate of the vaccine strain, *M. bovis* BCG, was slightly higher than H37Ra at day 4 and 7 post-infection (2 and 1.4 times higher, respectively) (Figure 6.4).

There was clearly a higher multiplication rate of the Beijing strain in THP-1 cells than the other Panel B and laboratory control strains (Figure 6.4). Using the avirulent laboratory control strain as the baseline, at day 4 post-infection, the multiplication rate, relative to day 1 post-infection, was approximately 7 times higher, whilst at day 7 the rate was approximately 16 times higher. The markedly higher multiplication rate of the Beijing strain was observed in all of the repeated independent experiments. The increased multiplication rate of the Beijing strain in comparison to the other Panel B *M. tuberculosis* strains was also observed in primary macrophages isolated from peripheral blood mononuclear cells (PBMC) (data not shown).

The multiplication rates of the other Panel B *M. tuberculosis* strains were similar at both day 4 and 7 post-infection at an average fold enhancement of 2 and 13, respectively. Whilst the multiplication rate for *M. africanum* was the same as *M. tuberculosis* EAI5, LAM10 and both CAS strains at day 4 post-infection, the rate was two times higher than these four *M. tuberculosis* strains at day 7 post-infection (Figure 6.4).

6.3.2 Cytokine production after infecting THP-1 cells

The supernatant, which was collected at each of the time points, day 1, 4 and 7, was used to measure levels of cytokine produced during infection and activation of THP-1 cells. ELISA assays, to measure the amounts of human TNF-α, IL-10, IL-1β and IL-6, were performed using commercial kits from eBioscience (refer to Section 2.6.5). The average cytokine levels, in pg/ml, from the three independent infection experiments, calculated from the OD_{450} values obtained after reading ELISA plates, were plotted onto graphs (refer to Figure 6.5). The graphs contain data from the Panel B strains, *M. tuberculosis* H37Rv (HPA, MRU), and non-infected THP-1 cells.

Results of cytokine production after infecting THP-1 cells with *M. tuberculosis* H37Rv (HPA, Porton Down), *M. tuberculosis* H37Ra (NCTC), *M. bovis* BCG and stimulating THP-1 cells with LPS and IFN-γ will be discussed in this section.

In general, higher concentrations of human TNF-α production were observed with the largest amount being produced by THP-1 cells that had been activated by IFN-γ at day 1 (approximately 4500pg/ml). Generally, the lowest concentrations were observed with human IL-10 production, with the largest amount being produced by THP-1 cells infected with the laboratory strain, *M. tuberculosis* H37Rv from HPA, MRU at day 4 (approximately 14pg/ml).

At day 1 there was twice as much production of TNF-α after infecting THP-1 cells with the clinical TB strains than the non-infected THP-1 cells, whilst at day 4 and 7 TNF-α production was 1.3 times higher than the non-infected cells. In general, the lowest levels of TNF-α production was detected in non-infected THP-1 cells and cells activated by LPS; there was only a very slight decrease in TNF-α production from day 1 to 4 after which levels remained the same (Figure 6.5a). The Beijing strain produced the highest levels of TNF-α at day 1 post-infection, whilst THP-1 cells infected with *M. africanum* produced the lowest amount of TNF-α. At day 4 and 7 post-infection the production of TNF-α by Beijing infected THP-1 cells was lower than the other Panel B strains.

At day 1, the production of human IL-10 from non-infected THP-1 cells and THP-1 cells infected and activated with the different mycobacterial strains and IFN-γ and LPS ranged from 5 to 6pg/ml (Figure 6.5b). During the whole time course of 7 days IL-10 production by non-infected THP-1 cells and IFN-γ and LPS activated THP-1 cells was lower than the IL-10 production by THP-1 cells infected with the Panel B and laboratory control strains.

At day 1 and 7 the highest and lowest production of IL-10 was observed when THP-1 cells were infected with EAI5 and *M. africanum* strains, respectively (Figure 6.5b). At day 4, the highest IL-10 production was observed with both of the CAS strains. Between day 1 and 4 there was a general increase in IL-10 production, but the increase was more prominent (approximately 2.4 fold) in THP-1 cells infected with H37Rv (HPA, MRU) (Figure 6.5b). For the remaining strains the increase between

day 1 and 4 was 1.8 fold or less. Production of IL-10 between day 4 and 7 remained level for the majority of mycobacterial strains, except for H37Rv (MRU). For this strain, a slight decrease in IL-10 production was observed (0.8 fold).

Whilst at day 1 similar amounts of IL-10 was produced by the non-infected THP-1 cells and cells infected with the clinical TB strains, at day 4 and 7 1.6 times more IL-10 was produced by THP-1 cells infected with the TB strains than the non-infected cells. It was important to note that the levels of IL-10 produced by THP-1 cells infected with both of the CAS strains (isolated from patients born in India and Somalia) was the same at each of the time points during the 7 day experimental period (Figure 6.5b).

Production of human IL-1β by non-infected THP-1 cells and cells activated by LPS was notably lower than IL-1β production by THP-1 cells infected with each of the Panel B and laboratory control strains (Figure 6.5c). At day 1, 4 and 7 there was 7, 5 and 2.3 times more IL-1β produced by the clinical TB strains than the non-infected THP-1 cells, respectively. Production of IL-1β by THP-1 cells activated by IFN-γ remained constant at day 1, 4 and 7 at approximately 350pg/ml. For some strains, there was an increase in IL-1β production between day 1 and 4 followed by a decrease between day 4 and 7. One such example was IL-1β produced by THP-1 cells infected with *M. bovis* BCG as production increased from 227 to 361pg/ml between day 1 and 4 but then decreased to 340pg/ml at day 7. As was observed with THP-1 cells activated by IFN-γ, there was constant IL-1β production by THP-1 cells infected with H37Rv (HPA, MRU), but at approximately 400pg/ml. For some strains, like CAS (isolated from the patient born in India), production of IL-1β remained constant at approximately 420pg/ml between day 1 and 4 and then slightly decreased to 370pg/ml by day 7.

Comparing the production of IL-1β from THP-1 cells infected with the different Panel B strains showed that the CAS, isolated from the patient born in India, and *M. africanum* strains produced the highest and lowest levels of IL-1β, respectively, at day 1 and 4 post-infection. At day 7 THP-1 cells infected with the CAS strain, from India, produced the lowest levels of IL-1β, whilst the highest levels were produced by the LAM10 strain.

At day 1 post-infection, human IL-6 production ranged from 25pg/ml (non-infected THP-1 cells) to 220pg/ml (THP-1 cells activated with IFN-γ). Non-infected THP-1 cells produced the lowest levels of IL-6 at approximately 25pg/ml and remained at this level during the 7 day experiment period (Figure 6.5d). At day 1, 4 and 7 there was 2.5, 11.9 and 14.8 times more IL-6 produced by the clinical TB strains than the non-infected THP-1 cells, respectively. The highest levels of IL-6 production were observed at day 4 and 7 with THP-1 cells infected with the Beijing strain (839 and 895pg/ml, respectively). With the exception of non-infected THP-1 cells, for the other infected and activated THP-1 cells, there was a general increase in IL-6 production between day 1 and 4. The most notable increase was observed when THP-1 cells were infected with the five Panel B *M. tuberculosis* strains, H37Rv (HPA, Porton Down) and *M. bovis* BCG with an approximate 6 to 7 fold increase. From day 4 to 7, IL-6 production either remained constant or decreased. The levels remained constant for THP-1 cells infected with the Beijing, H37Rv (HPA, MRU) and EAI5 to name a few (Figure 6.5d). A decrease was observed in THP-1 cells infected with strains like *M. bovis* BCG, in which a 0.7 fold decrease was observed. At day 1, 4 and 7 post-infection, the highest levels of IL-6 were produced by THP-1 cells infected with the Beijing strain and the lowest levels produced by *M. africanum*

Figure 6.4 *Graph showing the average fold enhancement (growth of mycobacteria relative to THP-1 cells present) from the three independent experiments when the mycobacterial cell:THP-1 cell ratio of day 4 and day 7 post infection was compared with the ratio at day 1. This statistic was calculated for THP-1 cells infected with M. tuberculosis strains, CAS (from patients born in India and Somalia), Beijing, EAI5, and LAM10, M. africanum, H37Rv (from HPA, MRU and HPA, Porton Down), H37Ra, and M. bovis BCG.*

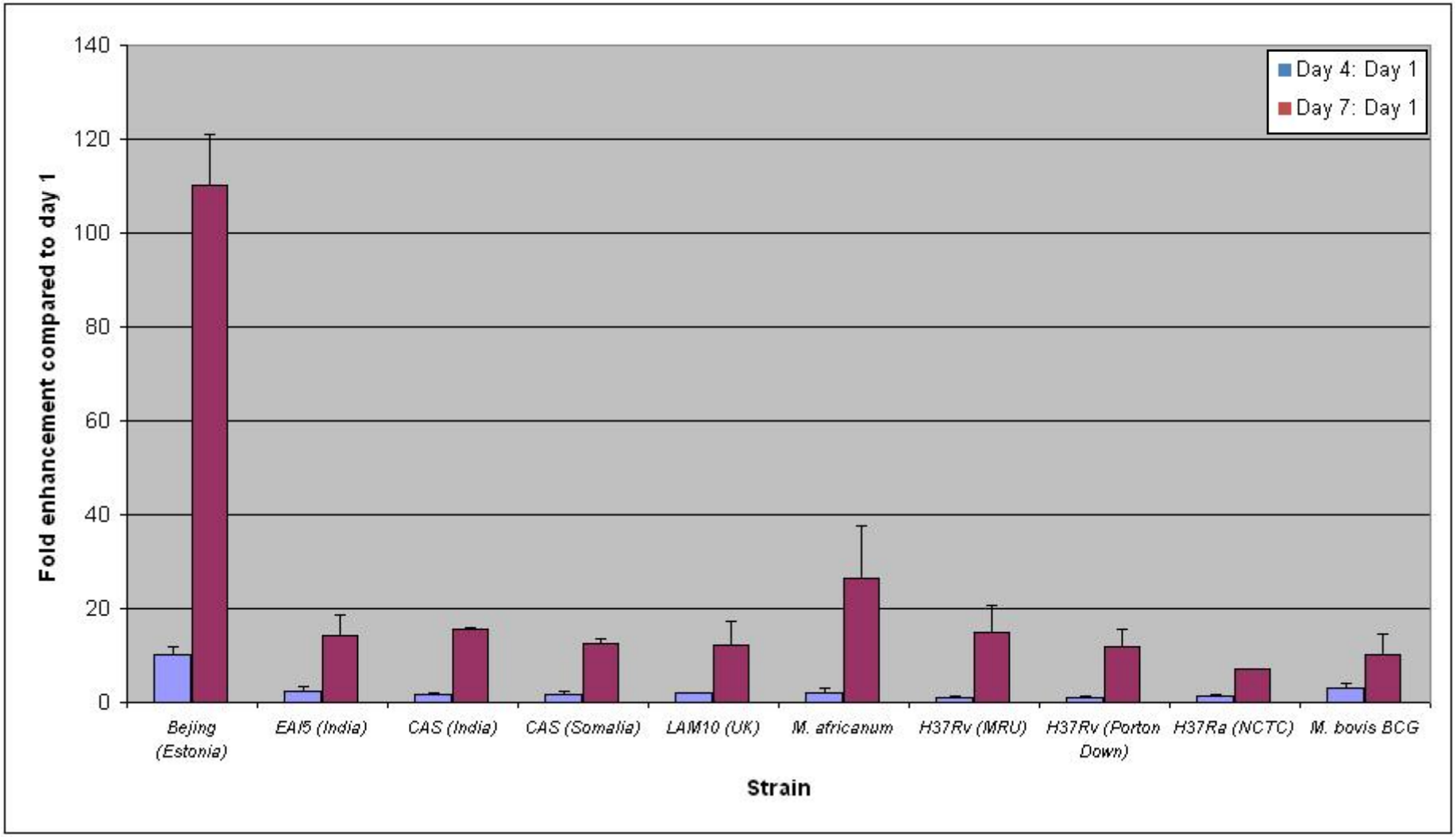


Figure 6.5 *Graphs showing the average concentrations (from the three independent experiments) of human (a) TNF-α, (b) IL-10, (c) IL-1β and (d) IL-6 at day 1, 4 and 7 after infecting macrophage-like THP-1 cells with M. tuberculosis strains, CAS (from patients born in India and Somalia), Beijing, EAI5, LAM10, M. africanum, and H37Rv (from HPA, MRU). Also shown is data from the negative control, non-infected THP-1 cells.*

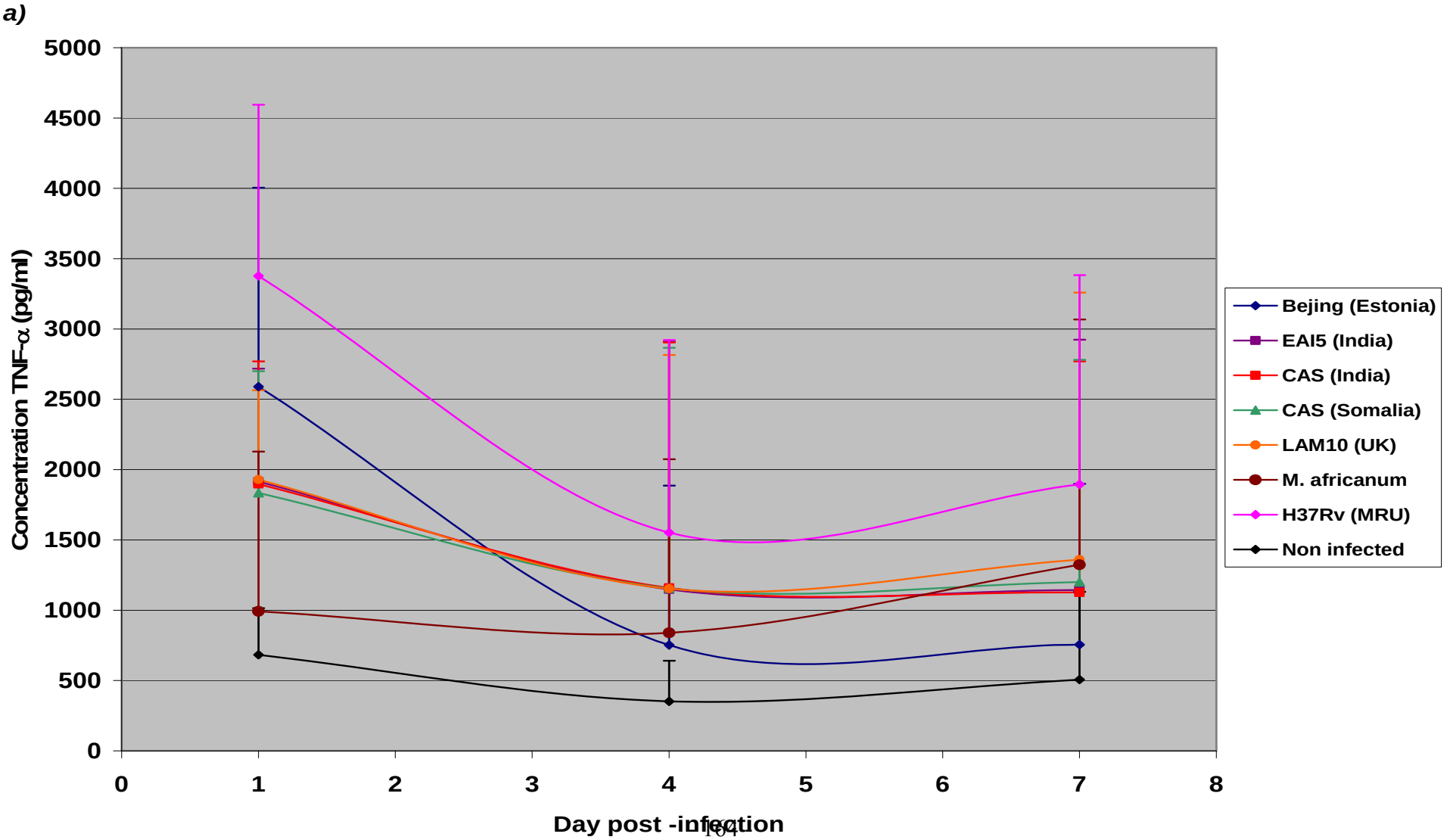


Figure 6.5 continued

b)

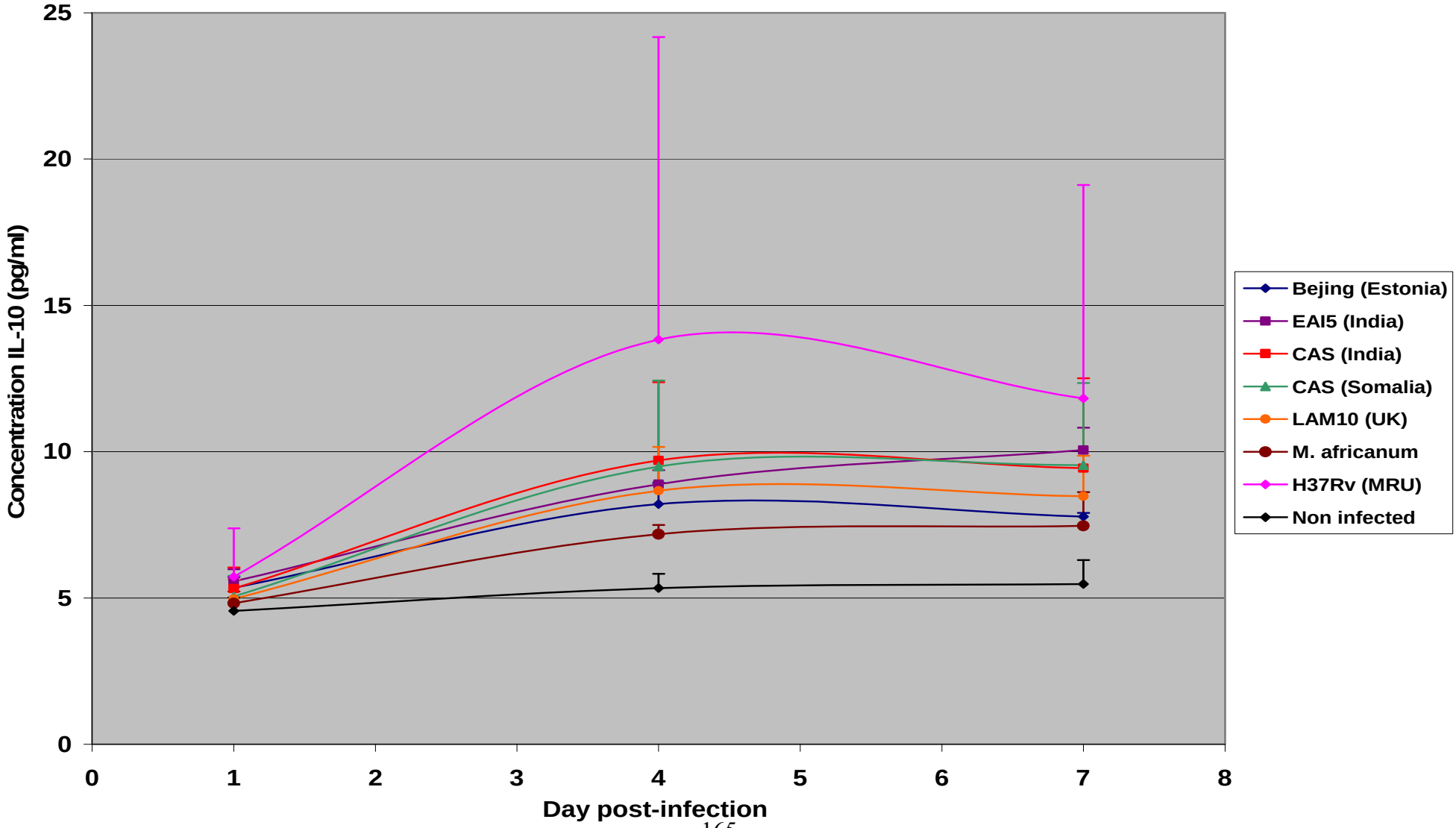


Figure 6.5 continued

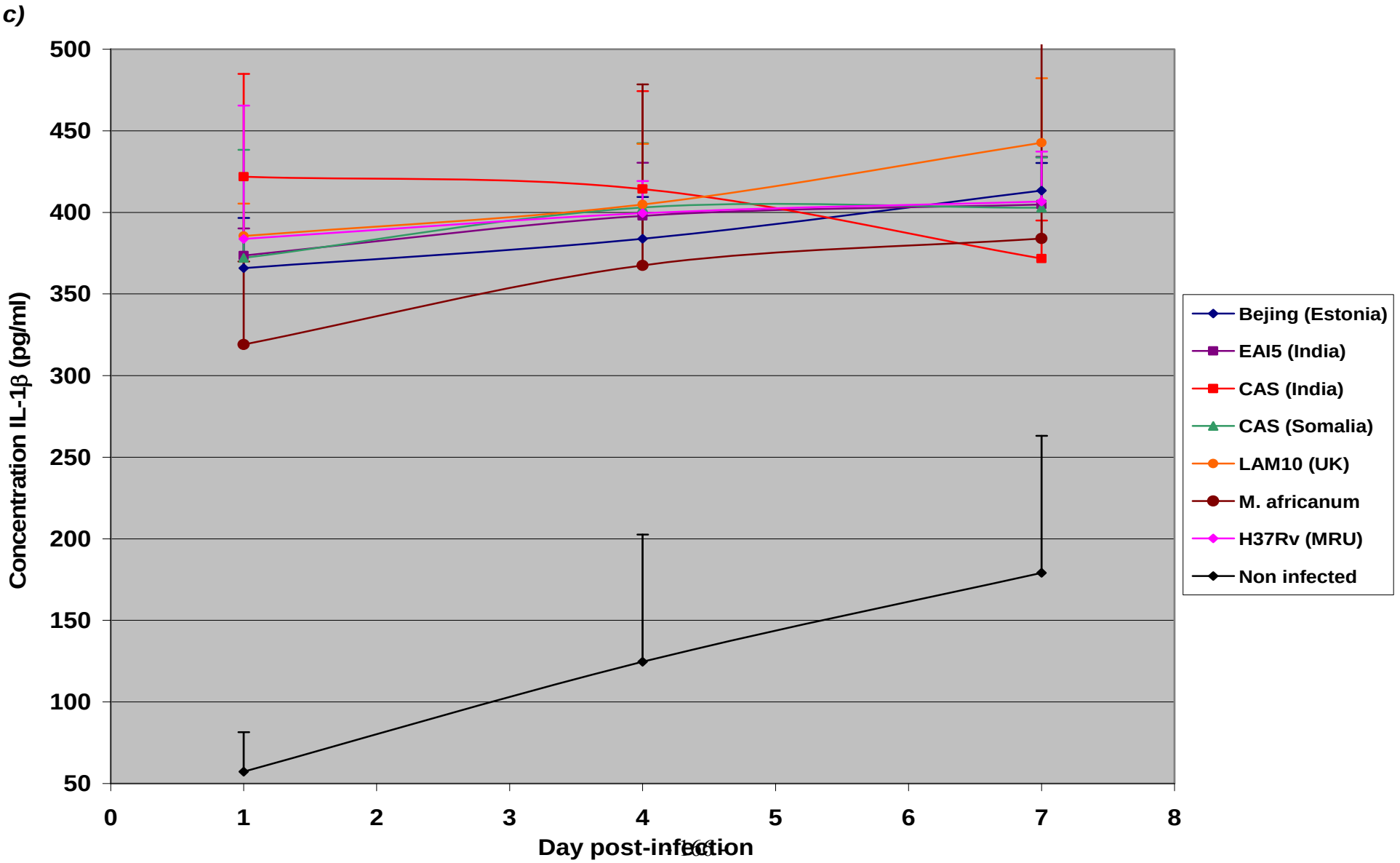
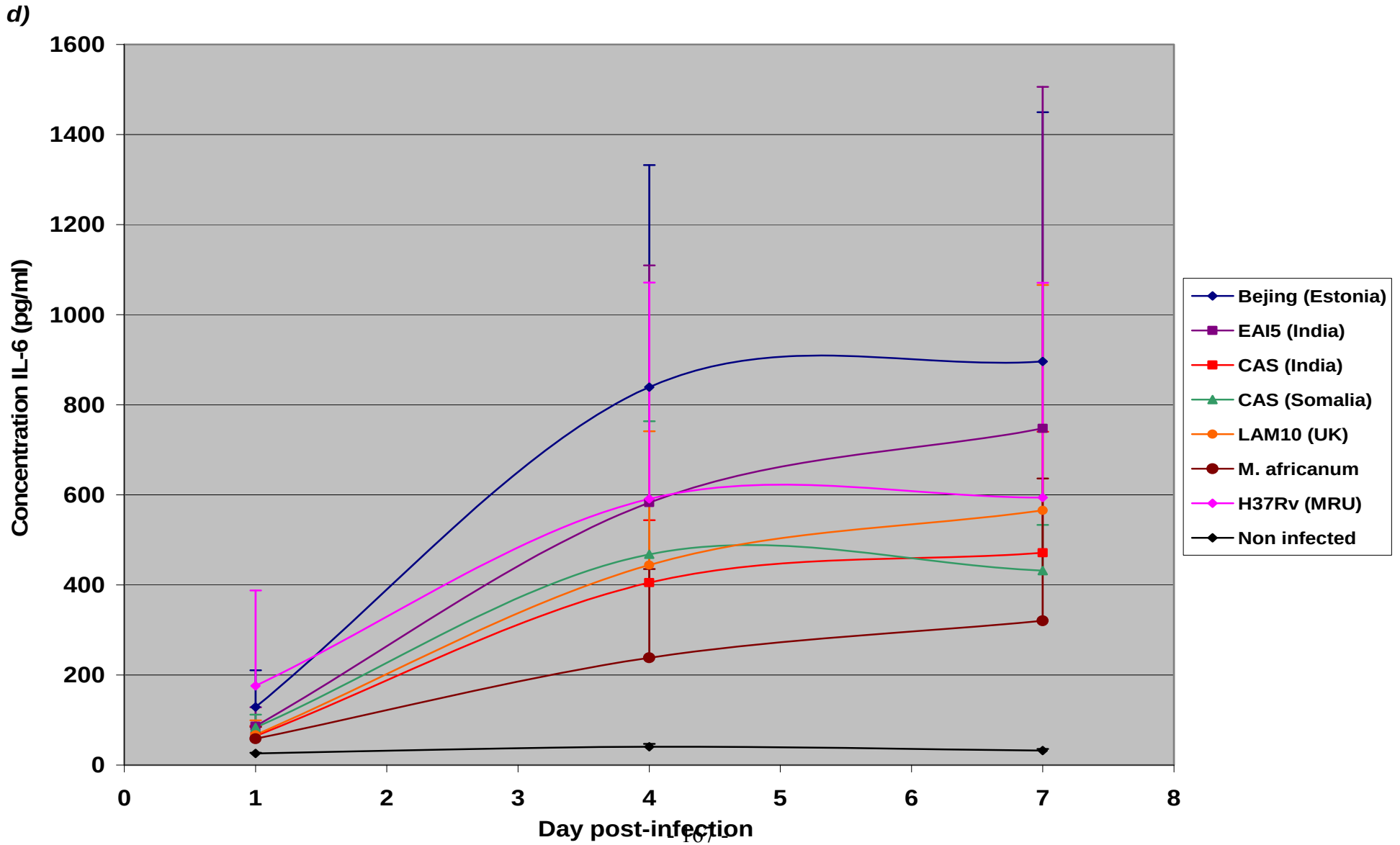


Figure 6.5 continued



6.4 DISCUSSION

As MTBC primarily infects alveolar macrophages before undergoing multiplication, a human macrophage-like model *in vitro* system (THP-1) was chosen to investigate variation in the phenotype of the different MTBC strains. Initially, THP-1 cells were proliferative and non-adherent but after activation with PMA, the cells changed from monocytes to macrophage-like cells, which were non-proliferative and adherent. Studies have demonstrated that activated THP-1 cells provided a reliable alternative to peripheral blood monocyte and bone-marrow derived macrophages to identify differences in the virulence of different *M. tuberculosis* strains (Theus *et al.,* 2004; Sow *et al.,* 2007). PMA-activated THP-1 cells have been used successfully for previous research into the effects of *M. tuberculosis* on macrophages and the various mechanisms involved in the survival of these bacilli once taken up by macrophages (Theus *et al.,* 2005; Lee *et al.,* 2007; Theus *et al.,* 2007).

Since mycobacterial cells are usually actively dividing during infection, THP-1 cells were infected with actively dividing mycobacterial cells. In addition, once activated, THP-1 cells survive for a week, reinforcing the importance of infecting THP-1 cells with actively dividing mycobacterial cells. From previous experiments and results, it was observed that mycobacteria were in the exponential phase of the bacterial growth curve, during which bacteria are actively dividing, at OD_{600} 0.2 (refer to Section 5.3). Therefore, cultures were prepared at the appropriate OD_{600} two days prior to infecting THP-1 cells. Macrophage-like THP-1 cells were infected at an MOI of 1 so 300 000 THP-1 cells were infected with an equal number of mycobacterial cells, which had been quantified using real-time PCR. The negative control was an equivalent volume of Middlebrook 7H9 medium.

After infecting THP-1 cells with the ten strains and activating THP-1 cells with LPS and IFN-γ, supernatant and cells were collected at days 1, 4 and 7 to monitor the effect of the strains, but cells could be collected at day 3 instead of day 4 (data not published). The infected THP-1 cells were pelletted and after DNA extraction, real-time PCR was performed to work out the fold enhancement of mycobacteria relative to the THP-1 cells present at day 4 and 7 post-infection when compared with day 1 post-infection. In essence the fold enhancement data showed the multiplication and growth of the different strains in an *in vitro* model.

Primers that targeted the *rpoB* region of the *M. tuberculosis* genome were used to quantify the amount of mycobacteria present and primers for actin were used to quantify the amount of THP-1 cells. It was decided to calculate fold enhancement of mycobacteria in this way as apoptosis of PMA-activated THP-1 cells is dependent on the virulence of the strains (Keane *et al.,* 1997; Balcewicz-Sablinska *et al.,* 1998; Keane *et al.,* 2000; Riendeau *et al.,* 2003). The data from these previous studies demonstrated that virulent strains like *M. tuberculosis* H37Rv, Erdmann and wild-type *M. bovis* induced decreased apoptosis of THP-1 and human alveolar macrophage cells *in vitro* than the low virulence strains like *M. tuberculosis* H37Ra, avirulent *M. bovis* BCG and *M. kansasii*.

Levels of human TNF-α, IL-10, IL-1β and IL-6 in the supernatants were measured using ELISA assays. It was decided to measure levels of these particular cytokines because of the role that they play in the immune response to *M. tuberculosis* infection (Bogdan *et al.,* 1991; Marques *et al.,* 1999; Stenger *et al.,* 2001; Fremond *et al.,* 2004; Theus *et al.,* 2005).

The production of human TNF-α, IL-10, IL-1β and IL-6 was notably less with the negative controls than levels produced upon infection with the strains in all of the independent experiments. Theus *et al*., 2005 also observed that there was an undetectable level of TNF-α and IL-10 production from non-infected PMA-activated THP-1 cells (Theus *et al.,* 2005). The data from the present study indicated that the medium used for culturing all of the strains did not affect the assay and that any differences seen in the production of cytokines was due to infection of THP-1 cells with the mycobacterial strain.

The other controls included the addition of LPS and IFN-γ to THP-1 cells. These were potentially positive controls, but LPS and IFN-γ have been used for different purposes in previous studies and their effect on cytokine production varied according to which cytokine was being measured. Once alveolar macrophages are infected with mycobacterial cells, macrophages are activated to increase cell-mediated Th1-type response, which leads to an increase in production of IFN-γ, IL-2 and IL-12. In some studies IFN-γ has been used to stimulate macrophage-like cells before infection with *M. tuberculosis* strain instead of stimulating cells with IFN-γ on its own (Park *et al.,*

2006; Sow *et al.,* 2007). However, the object of this study was to test the hypothesis that there were no phenotypic differences between the strains so IFN-γ was not added along with the strains to avoid introducing a variable.

LPS, a component of bacterial cell walls, has previously been used to compare levels of TNF-α and IL-6 after stimulating macrophages with medium, LPS and control strains, including *M. tuberculosis* H37Ra and H37Rv and vaccine *M. bovis* BCG strain (Fremond *et al.,* 2004). In this study the levels of TNF-α were similar when macrophages were stimulated with LPS and *M. tuberculosis* H37Rv, whilst there were notably higher levels of IL-6 when comparing LPS stimulated macrophages with those infected with the vaccine strain and *M. tuberculosis* H37Rv. However, whilst LPS was used as a potential positive control, the levels of all cytokines in the present study upon activating THP-1 cells with LPS were only slightly higher than the cytokine levels for non-infected THP-1 cells but markedly lower than cytokine levels when adding the different strains to the THP-1 cells. In another study by Marques *et al*., 1999 LPS proved to be a particularly strong inducer for TNF-α production (Marques *et al.,* 1999).

The data from the present study demonstrated that there was a markedly higher growth and multiplication of *M. tuberculosis* Beijing strain than the other Panel B, laboratory control and vaccine strains at day 4 and 7 indicating that there was continuously increased growth of the Beijing strain throughout the infection. The growth rate of a Beijing strain *in vitro* has been investigated previously by Theus *et al*., 2007. This study investigated any associations between intracellular growth of the *M. tuberculosis* Beijing strain and production of TNF-α and IL-10. It was concluded that there was an inverse relationship between TNF-α production and intracellular growth rate; hence the slower growing Beijing strains induced secretion of higher levels of TNF-α (Theus *et al.,* 2007). Following this principle it would be expected that there would be lower levels of TNF-α after infecting THP-1 cells with the Beijing in comparison with TNF-α levels produced after infecting THP-1 cells with the other strains that had been evaluated in the present study. However, whilst the highest multiplication rates were observed after infection with the Beijing strain, higher levels of TNF-α were produced by the THP-1 cells infected with the Beijing strain than THP-1 cells infected with the other four Panel B MTBC strains (two *M. tuberculosis* CAS strains, *M. tuberculosis* EAI5, *M. tuberculosis* LAM10 and *M.*

africanum). In the study by Theus *et al*., 2007 there was an inverse relationship between production of TNF-α and IL-10, which was the same as what had been observed in the other studies previously discussed and also in the present study.

Infection of alveolar macrophages with *M. tuberculosis* H37Rv and H37Ra induced IFN-γ production via the Th1 immune response pathway, enhancing the killing of intracellular mycobacteria by activating macrophages to become microbicidal and stimulating the recruitment of more macrophages which will eventually form granulomas (Bogdan *et al.,* 1991; Fenton *et al.,* 1997; Fremond *et al.,* 2004; Bhatt *et al.,* 2007).

Autoregulation within macrophages occurs as IFN-γ production can be blocked by production of immunosuppressive and anti-inflammatory IL-10, which is induced by the Th2 pathway (Fenton *et al.,* 1997; Bhatt *et al.,* 2007). The findings from these studies correlate with the present study as THP-1 cells activated with IFN-γ produced very little IL-10 as levels were very similar to those produced by non-infected THP-1 cells. Conversely, THP-1 cells activated with IFN-γ produced higher levels of TNF-α, IL-1β and IL-6. This inverse relationship in the production of these cytokines has been observed in previous studies and is discussed later.

Whilst IL-10 inhibits the production of IFN-γ, IL-10 also down-regulates production of TNF-α, which is a crucial cytokine for eliciting a pro-inflammatory response, and both IFN-γ and TNF-α, in conjunction, are important for controlling TB infection (Bogdan *et al.,* 1991; Marques *et al.,* 1999; Stenger *et al.,* 2001). Infection with *M. tuberculosis* generally leads to increased production of TNF-α by macrophages, which induces production of reactive nitrogen intermediates that work in synergy with IFN-γ to kill mycobacterial cells (Bhatt *et al.,* 2007). On its own, TNF-α helps to control secretion of other cytokines and stimulate apoptosis of infected macrophages to reduce bacterial burden.

In the present study, we focused on the analysis of the production of pro-inflammatory cytokines, TNF-α, IL-1β and IL-6 which had been shown to vary during virulent infection of macrophages (Ragno *et al.,* 2001; Volpe *et al.,* 2006). This study demonstrated that these pro-inflammatory cytokines played a particularly important role in the immune response against *M. tuberculosis* infection and it had

already been demonstrated that IL-1β was involved in granulomatous inflammation. A pronounced reduced level of IL-10 production in comparison with the significantly higher detected levels of TNF-α, IL-1β and IL-6 was observed in the present study, which correlated to the previous literature.

In the present study, the cytokine production data provided some insight into the possible immune response induced after infection of THP-1 cells with the Panel B wild-type strains. During the early stages in infection of THP-1 cells, at day 1, cells infected with the Beijing strain produced the highest levels of TNF-α and IL-6 indicating that there is an early pro-inflammatory response to this strain. Based on the principles described by Ragno *et al*., 2001 and Volpe *et al*., 2006 virulent strains induce higher production of pro-inflammatory cytokines showing that the Beijing strain is more virulent than the other Panel B strains.

Contrary to the production of TNF-α and IL-6 production observed with the Beijing strain, THP-1 cells infected with *M. africanum* produced the lowest levels of TNF-α and IL-6 indicating that this MTBC strain is not as virulent as the Beijing strain but this is not conclusive as more experiments need to be performed for confirmation. The lowest production of IL-1β was also observed with *M. africanum* further supporting the evidence that this strain is not as virulent as the other Panel B *M. tuberculosis* strains.

In addition to growth in an in vitro cellular model, growth in a MGIT cell-free model was compared using an ideal inoculum of 600 and 6,000 TB genomes. At both of these inoculum sizes the fastest growth rate was observed with the *M. tuberculosis* Beijing strain. The *M. tuberculosis* CAS strain, isolated from the patient born in Somalia, and *M. africanum* had the slowest growth rate at inoculum size of 6,000 and 600 genomes, respectively. Higher growth rates were observed with the Beijing strain in the tissue culture and cell-free culture system when compared to the Panel B strains.

A study by Park *et al*., 2006 used *M. tuberculosis* H37Rv and clinical isolates of known virulence to demonstrate that the fast growth rate of strains in an intracellular environment was a virulence factor during the early stages of *M. tuberculosis* infection and relating the growth rate to TNF-α production profiles showed that the

more virulent strains induced increased levels of TNF-α production (Park *et al.,* 2006). Our study confirms that the Estonian Beijing isolate which over a decade has come to dominate the TB population in patients in Estonia grows more quickly than other clinical and laboratory strains tested and initially produced higher levels of TNF-α.

Referring back to Section 3.2.4, one of the factors that had been used to select strains for Panel A had been partially based on the sizes of clusters. In the study by Theus *et al*., 2005, the element of cluster size was brought together with growth rate of strains and levels of cytokines produced after infecting activated THP-1 cells with clustered and unique clinical isolates (Theus *et al.,* 2005). Whilst MIRU and ETR profiles were used for cluster analysis in the present study, the study by Theus *et a*l., 2005 used IS*6110* RFLP patterns. There was a more rapid intracellular growth of unique isolates with greater levels of TNF-α production and lower levels of IL-10 production in comparison with clustered isolates, therefore during early infection there was a positive correlation between clustered isolates and IL-10 production but a negative correlation with TNF-α production (Theus *et al.,* 2005). Ultimately, the study determined that growth rate and production of IL-10 and TNF-α were good phenotypic markers for investigating epidemiologically relevant strains, whilst providing insight into virulence of *M. tuberculosis* strains as IL-10 antagonises the production of pro-inflammatory cytokines, IFN-γ, TNF-α and IL-12, disrupting the host immune defence against TB infection.

The collective data from all of the *in vitro* experiments disproved the initial null hypothesis that there were no differences between the different *M. tuberculosis* strains as it was obvious that the *M. tuberculosis* Beijing strain behaved differently to the other strains in its intracellular growth rate and the growth rate in a cell-free culture system, showing that this is a potentially virulent strain based on previous literature (Theus *et al.,* 2005; Park *et al.,* 2006; Theus *et al.,* 2007). The increased production of the pro-inflammatory cytokines, TNF-α and IL-6, compared to the other Panel B strains and the low levels of the anti-inflammatory cytokine, IL-10, produced after infecting THP-1 cells with the Beijing strain further demonstrated the potential virulent nature of the Beijing strain in comparison to the other Panel B strains.

CHAPTER 7

THE STABILITY OF AEROSOLS DERIVED FROM THE *M. TUBERCULOSIS* COMPLEX STRAINS

7.1 INTRODUCTION

Data from *in vitro* tissue culture experiments indicated that there were differences in the phenotypes of the circulating clinical strains included for analysis. The *M. tuberculosis* Beijing strain was included because it differed from the other circulating clinical, laboratory control and vaccine strains due to its higher growth in the activated THP-1 model. The remaining strains were selected based on the four main *M. tuberculosis* lineages defined by Baker *et al*., 2004. The final panel of strains, included *M. tuberculosis* Beijing (lineage I), LAM10 (lineage II), both CAS strains isolated from patients born in India and Somalia (lineage III) and EAI5 (lineage IV).

After performing *in vitro* experiments the next step would normally be to narrow down the panel of strains by selecting those of interest for analysis using low throughput *in vivo* experiments. For the purposes of this particular study, we tested the hypothesis that there were no differences in the phenotype or behaviour of the strains in *in vivo* experiments involving aerosol challenge of guinea pigs. During the preparation of the aerosols from cultures of each strain for guinea pig challenge studies, it was decided to perform All Glass Impinger-30 (AGI-30) sampling primarily to ensure that any differences seen between the *M. tuberculosis* strains, after infecting guinea pigs, were due to differences in virulence alone rather than differences in concentrations of delivered doses of mycobacteria and to validate that the Henderson apparatus, through which the aerosolised culture flowed, was working within the required parameters. However, these studies would also offer some insight into the stability of strains as aerosols because TB is transmitted via the respiratory route as aerosols produced whilst coughing and sneezing. As mentioned in Section 2.8.2 the AGI-30 sampler jar drew up air through an inlet from the Henderson apparatus and dissolved the aerosols containing TB bacilli in the 5ml sterile distilled water in the AGI-30 sampling jar.

As part of our overall aim we sought to investigate whether the stability of aerosols produced by these strains contributed to overall differences in phenotype. In previous studies, guinea pigs have been challenged with standard laboratory strains of *M. tuberculosis* H37Rv and *M. bovis*; these data provide a foundation from which to investigate the *in vivo* effects of wild-type strains that were circulating within London and which have more relevance for public health (Williams *et al.,* 2000; Williams *et al.,* 2005).

7.2 INVESTIGATION OF THE STABILITY OF AEROSOLS OF CIRCULATING *M. TUBERCULOSIS* COMPLEX STRAINS

The stability of each strain as an aerosol was tested using an AGI-30 sampling device. Cultures containing 1×10^7 c.f.u. had been used for previous guinea pig challenge procedures performed at Porton Down. However, as mycobacteria had been quantified using real-time PCR for the tissue culture experiments, this quantification method was also used to obtain 1.5ml of culture containing 1×10^7 mycobacterial genomes for each of the five strains (refer to Section 5.3 for the comparisons between using real-time PCR and c.f.u. for MTBC quantification). After dilution, the final working concentration of cells was 1×10^6 mycobacterial genomes. Cultures contained actively growing mycobacterial cells.

Cultures were aerosolised using a 3-jet collison nebulizer, after which aerosols were passed through the Henderson apparatus to be delivered to four AGI-30 samplers containing 5ml sterile distilled water into which the aerosols would be captured. Mycobacterial samples from the collison and AGI-30 samplers were serially diluted and plated onto Middlebrook 7H11 agar plates, incubated for 4 weeks and then colonies were counted. After taking into account the volume of culture plated and the dilution factor, the c.f.u./ml of mycobacteria in the collison were compared with c.f.u./ml from the AGI-30 samplers to identify if there were any differences in the survival of the strain in an aerosolised state.

Figure 7.1 shows the number of viable mycobacterial cells at each of the stages of the experiment, i.e. colony forming units were monitored in the collison nebulizer prior to aerosolisation for AGI-30 sampling and before each of the guinea pig challenges. A sample of the stock culture was also plated onto agar plates to obtain c.f.u./ml data

primarily to verify the mycobacterial concentration of the stock cultures and to ensure that the 1 in 10 dilution of the 5 strains, to prepare cultures at a working concentration for the collison nebuliser, was performed as accurately as possible.

From the plots for each of the five strains, the number of bacilli quantified using real-time PCR was approximately one $\log_{10}$ higher than the viable c.f.u. count. The lowest and highest viable c.f.u./ml was observed for *M. tuberculosis* EAI5 and LAM10, respectively, and the difference between the stock culture counts for these strains was one $\log_{10}$ c.f.u./ml between approximately $1\times10^{6.5}$ c.f.u./ml and $1\times10^{5.5}$ c.f.u./ml. The c.f.u./ml for all of the other strains was either just above or below 1×10^{6} c.f.u./ml.

The viable c.f.u. count at 0 mins represented the number of c.f.u. after diluting stock cultures 1 in 10. Theoretically, the final number of c.f.u. should be 1×10^{6} in 15ml with one $\log_{10}$ decrease between stock and diluted cultures. However, as the c.f.u. count of the stock culture was lower than 1×10^{7} c.f.u. in 1.5ml, the c.f.u. counts at 0 mins were lower than expected. Whilst the expected decrease was observed for *M. tuberculosis* LAM10 strain, there was a two $\log_{10}$ drop in c.f.u./ml for *M. tuberculosis* CAS isolated from the patient born in Somalia and the decrease for all other strains was slightly more than one $\log_{10}$.

The diluted cultures were firstly used for AGI-30 sampling in which cultures were aerosolised for 5 mins, during which time aerosols had been passed through the Henderson apparatus ready for collection in the AGI-30 samplers. After the AGI-30 sampling was complete, an aliquot of culture in the collison was plated for c.f.u. count before carrying on to challenge two sets of eight guinea pigs by the aerosol route. The c.f.u./ml counts for all of the strains were similar to the counts in the diluted cultures pre-challenge.

The final two points of each plot in Figure 7.1 represented the c.f.u./ml in the collison after each of the guinea pig challenges. For *M. tuberculosis* Beijing strain, the c.f.u./ml after each challenge was similar to that of the diluted culture and after AGI-30 sampling. For all of the other strains the c.f.u./ml was either similar or slightly higher than the counts of diluted cultures and after AGI-30 sampling. These trends indicated that the starting concentrations of mycobacteria were similar between the AGI-30 sampling, first challenge and then the second challenge.

Figure 7.2 shows the c.f.u. counts of the five *M. tuberculosis* strains that had been collected in the AGI-30 sampler after aerosolised diluted culture had been passed through the Henderson nebulizer. For all strains, except *M. tuberculosis* LAM10 strain, the c.f.u. counts in the four samplers were similar and on average there were approximately 1×10^2 c.f.u. in each of the samplers. For the LAM10 strain the number of c.f.u. varied between samplers, the largest difference being one $\log_{10}$ c.f.u./ml between approximately 1×10^3 c.f.u./ml, observed in replicate 3, and 1×10^2 c.f.u./ml, observed in replicates 2 and 4. The approximate average c.f.u. count from the LAM10 replicates was $1\times10^{2.5}$ c.f.u./ml, which is marginally higher than counts for the other strains. This increase may be due to a higher mycobacterial concentration in the stock culture for LAM10 than the other strains. The higher concentration in the collison may mean more mycobacterial bacilli passed through the apparatus and delivered to the AGI-30 samplers.

From these data, there is no indication that there is a difference in the ability of the five strains to survive in the aerosolised state, as a similar number of c.f.u. were recovered after being passed through the Henderson apparatus. Statistical analysis using the two sample t-test ($p<0.05$) confirmed that there was indeed no significant difference, between the *M. tuberculosis* strains, in the aerosols captured in the AGI-30 samplers. Importantly, all 5 strains retained their viability during the period of aerosolisation, which would mean that all guinea pigs would be challenged with an equivalent dose enabling a direct comparison to be made of their virulence in this model.

Figure 7.1 *Log_{10} c.f.u./ml of mycobacteria in cultures for each of the strains including M. tuberculosis Beijing (Estonia), CAS (India), CAS (Somalia), EAI5 (India) and LAM10 (UK). The log_{10} c.f.u./ml was taken for stock culture, in diluted culture at 0 mins, and then in the collison, after aerosolisation for AGI-30 sampling after 5 mins, and after aerosol challenge of each set of 8 guinea pigs at 10 and 15 mins. (b) Log_{10} c.f.u./ml of mycobacteria collected in the four AGI-30 samplers after the cultures for strains were aerosolised and passed through the Henderson apparatus.*

a)

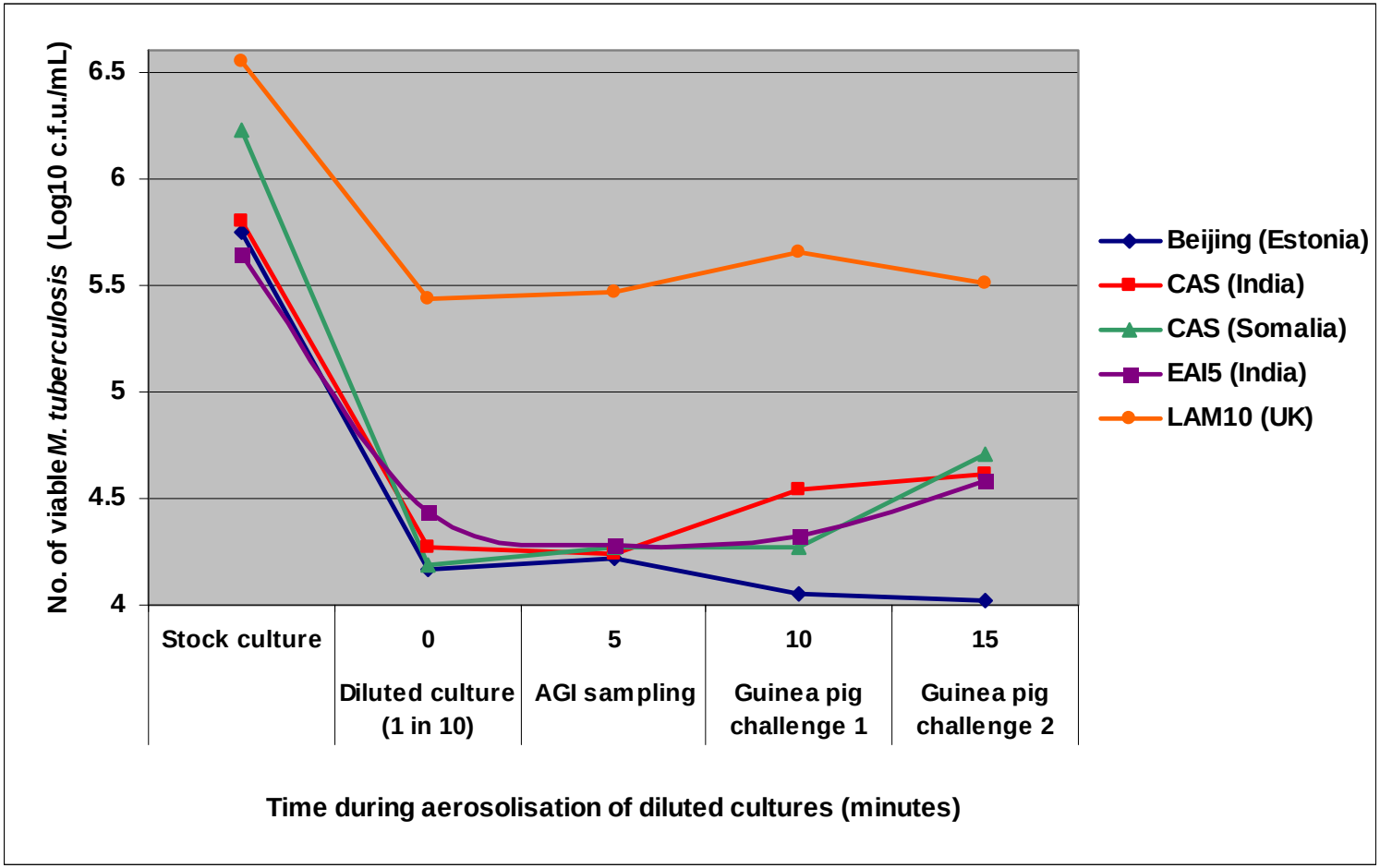


Figure 7.2 *Log_{10} c.f.u./ml of mycobacteria collected in the four AGI-30 samplers after the cultures for M. tuberculosis (a) Beijing (Estonia), (b) CAS (India), (c) CAS (Somalia), (d) EAI5 (India) and (e) LAM10 (UK) strains were aerosolised, passed through the Henderson apparatus and the aerosols were dissolved in 5ml distilled water.*

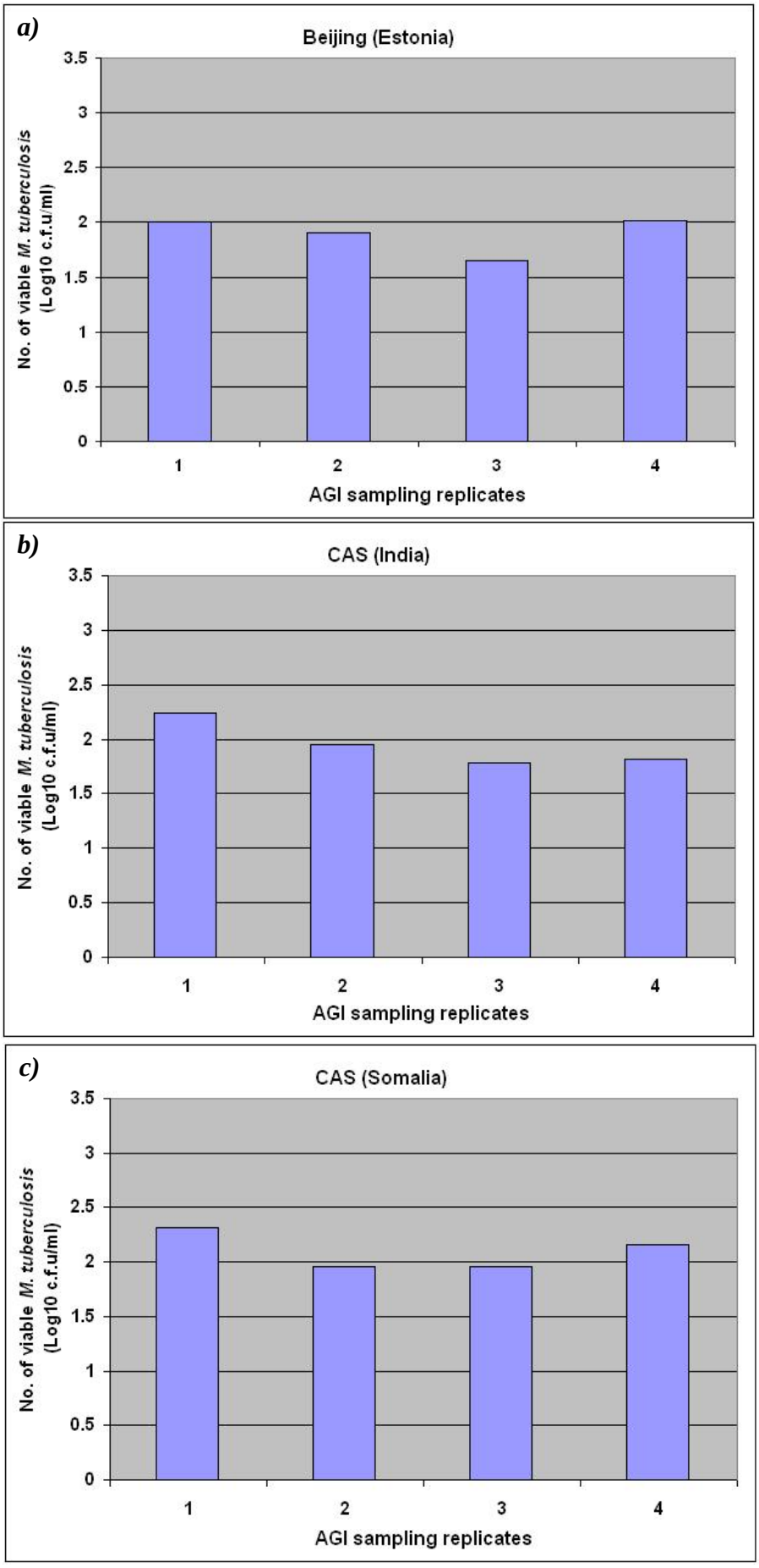
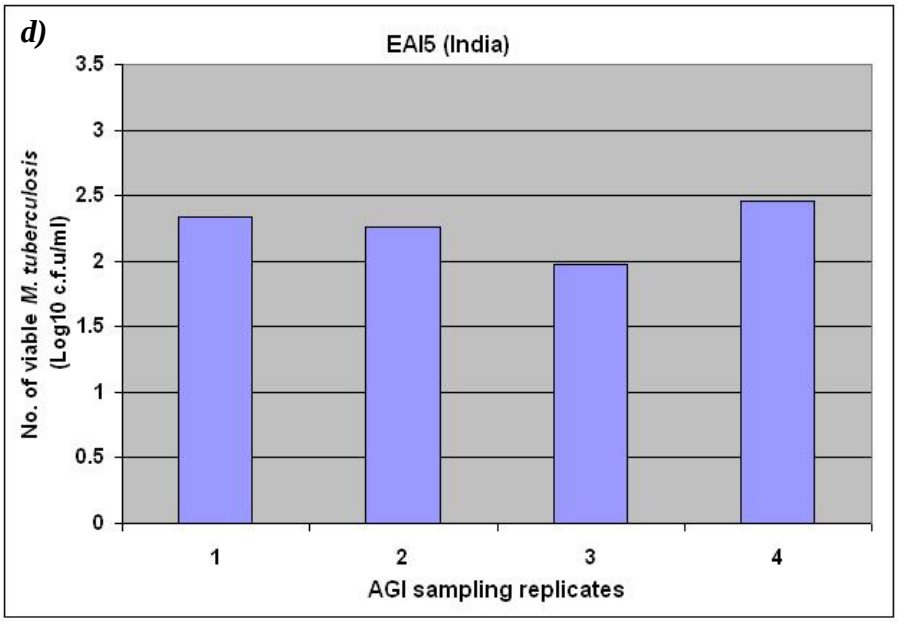
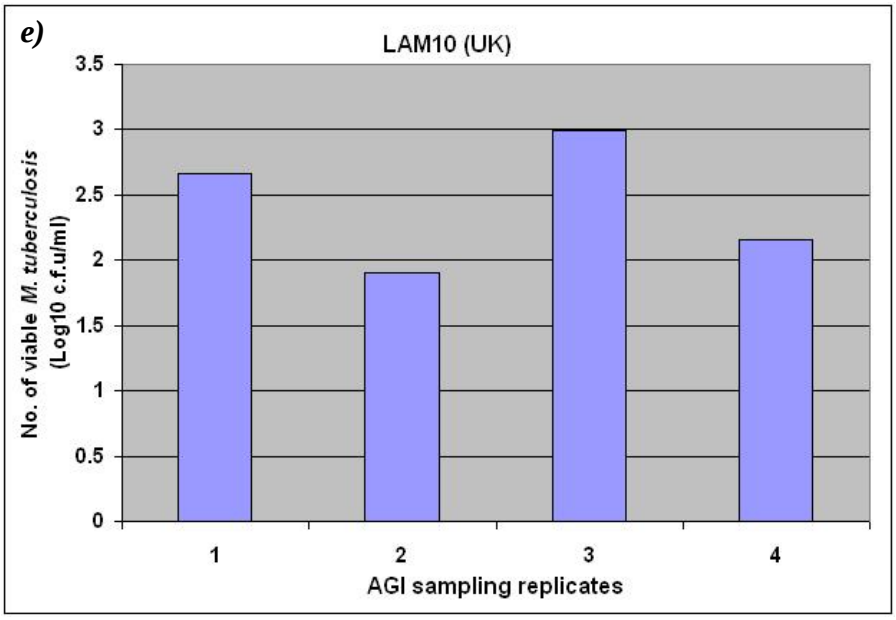


Figure 7.2 continued

d)



e)



7.3 DISCUSSION

TB bacilli are transmitted via aerosols and the prolonged stability of aerosols might aid in the successful transmission of certain strains, and be a potential property of more virulent strains. Whilst there are previous studies investigating the virulence of clinical strains *in vivo*, there was limited previous data showing correlations between aerosol stability and virulence of *M. tuberculosis* strains to prove or disprove this hypothesis (Williams *et al.,* 2005; Palanisamy *et al.,* 2009).

For the purposes of this study, *M. tuberculosis* cultures had been used, but in reality aerosols are produced from saliva or sputum as an infected individual coughs or sneezes. In a study by Lever *et al*., 2000, the survival of aerosols generated from saline containing mycobacterial bacilli was compared with survival of aerosols from artificial saliva and revealed that there were no differences in the rate of aerosol survival whether saliva or broth cultures were used for experiments (Lever *et al.,* 2000).

As AGI-30 sampling is a low throughput technique, and previous studies performed by staff at HPA, CEPR have produced historical aerosol survival data using *M. tuberculosis* H37Rv, strains were selected from those that had been included previously for tissue culture experiments. It was decided to compare the aerosol stability of cultures of circulating clinical strains from Panel B. After performing AGI-30 sampling using cultures for these five strains and analysing the data, there was no significant difference between strains in the c.f.u./ml collected in the AGI-30 samplers relative to the initial c.f.u./ml in the cultures. This showed that there was no significant difference in the stability of the tested strains in their aerosolised state. There have been very few studies that have focused on the stability of aerosols alone, as most studies have used aerosol production for infecting mice and guinea pigs to investigate the effects of vaccines *in vivo* (Williams *et al.,* 2000; Williams *et al.,* 2005; Williams *et al.,* 2005; Vipond *et al.,* 2006; Ordway *et al.,* 2007).

However, Lever *et al*., 2000, investigated the aerosol stability of different *Mycobacterium* species, as opposed to different *M. tuberculosis* strains, and demonstrated that despite the variation seen in human disease caused by these species, the survival of *M. tuberculosis*, *M. avium* and *M. intracellulare* in their aerosol state

was similar. These data partially support the results produced in the present study because despite the increased virulence of *M. tuberculosis* Beijing observed in the *in vitro* model, there were no differences in the stability of aerosols produced by the culture.

A crucial point is that the procedures used in the present study were performed in a controlled environment and whilst the data may provide some insight, the controlled conditions are artificial. In reality there are a plethora of other factors and variables that might influence the stability of aerosols, for example, ultraviolet light, wind speed, and temperature. The effects of these factors would be hard to investigate in an artificial model, which may explain the limited data available surrounding this topic. Nevertheless any research into aerosol stability may explain the observed epidemiological trends in global TB, for instance the dominance of the *M. tuberculosis* CAS strain in the Indian subcontinent. Is it possible that the climate in this region supports the prolonged stability of this strain? As it stands at the moment, data show that successful transmission of *Mycobacterium* species or strains is due to the prolonged contact of TB patients with other individuals that are in close proximity (Lever *et al.,* 2000). The similarities in the aerosol survival patterns for strains meant that any differences in virulence observed between strains during *in vivo* experiments was due to the strain and not due to the delivery of variable numbers of bacilli.

CHAPTER 8

IN VIVO PHENOTYPICAL ANALYSIS OF *M. TUBERCULOSIS* COMPLEX STRAINS

8.1 INTRODUCTION

In Section 6.3, the clinical strains in Panel B (*M. tuberculosis* Beijing, CAS India, CAS Somalia, EAI5, LAM10 and *M. africanum*) and laboratory control strains (*M. tuberculosis* H37Rv from HPA, MRU and Porton Down, *M. tuberculosis* H37Ra, *M. bovis* BCG) were used for *in vitro* experiments, infecting activated macrophage-like THP-1 cells, to test the null hypothesis that all strains had the same phenotype and behaved in a similar manner during infection. This hypothesis had been disproved after *in vitro* analysis showed there was a marked increase in the growth of *M. tuberculosis* Beijing strain compared to the other circulating and control strains.

The behaviour of strains in an *in vitro* assay may not necessarily resemble what would happen *in vivo*, therefore it was decided to test the same null hypothesis in an *in vivo* model. Mice have been used in previous TB studies, especially when investigating the *in vivo* immune response, as genetically engineered breeds can be more easily produced and there is a wider availability of immunological reagents for these animals (Lopez *et al.,* 2003; Flynn 2006). However, there is a lack of similarity between mouse and human pathology, both pulmonary and extrapulmonary, after TB infection. There are also important differences between murine and human immunology. In contrast, there are strong histological correlations between primary pulmonary lesions observed in guinea pigs and humans and there is good similarity in the extrapulmonary dissemination of TB bacilli in the two models (Bhatia *et al.,* 1961; Prabhakar *et al.,* 1987).

The extrapulmonary dissemination and pulmonary lesions are both crucial stages in human TB infection making guinea pigs a more attractive model for the purposes of the present study. Guinea pigs have successfully been used in previous *in vivo* studies for vaccine evaluation and investigating the infectivity and virulence of clinical and laboratory *M. tuberculosis* strains (Bhatia *et al.,* 1961; Prabhakar *et al.,* 1987;

Williams *et al.,* 2000; Williams *et al.,* 2005; Williams *et al.,* 2005; Vipond *et al.,* 2006).

Whilst all Panel B strains had been included in *in vitro* tissue culture experiments, they could not all be included in *in vivo* guinea pig experiments with a reasonable number of replicates as it would have been too costly with the limited numbers of guinea pigs available, it was decided to select *M. tuberculosis* LAM10, Beijing, EAI5 and two CAS strains to represent the four *M. tuberculosis* lineages out of the six major phylogenetic lineages to prove or disprove the initial null hypothesis (Baker *et al.,* 2004; Gagneux *et al.,* 2006). The same strains had been included for the AGI-30 sampling investigated in Chapter 7.

8.2 BODY WEIGHT OF GUINEA PIGS INFECTED WITH *M. TUBERCULOSIS* COMPLEX STRAINS

8.2.1 The effect on total body weight

In order to have a suitable number of replicates for each of the two time points at which guinea pigs would be culled, five groups of sixteen guinea pigs were aerosol challenged with each of the five clinical *M. tuberculosis* strains. Cultures for each of the *M. tuberculosis* strains were aerosolised using nebulizers and passed through a Henderson apparatus, but whereas for AGI-30 sampling aerosols were captured in sampling jars containing water, on this occasion aerosols were delivered straight to the snout of the guinea pigs so approximately 10 mycobacterial bacilli would be retained in the lungs of guinea pigs.

Post-challenge, guinea pigs were kept under controlled conditions and the body weight for all guinea pigs was assessed at regular intervals, which was possible as there was 100% survival of guinea pigs prior to culling. The average percentage weight change, for guinea pigs infected with the different strains, was plotted over time post-challenge and results of this analysis can be seen in Figure 8.1a. Variation in weight gain is linked to variation in pathogenicity of infecting strains. The area under the plotted curves for each of the strains was calculated using SigmaPlot (version 10.0) to represent the data more clearly and to visualise the effect of infecting guinea pigs with the different strains; results of this analysis can be seen in Figure 8.1b.

In general, there was an increase in the body weights of guinea pigs irrespective of which strain was used to infect animals, which was expected. However, a closer analysis of the curves in Figure 8.1a showed that throughout the whole time course, the increase in weight of guinea pigs infected with the Beijing strain was lower than the other four circulating strains, with the highest percentage weight gain being for guinea pigs infected with the CAS strain isolated from the patient born in India. The largest difference in body weight change between guinea pigs infected with Beijing and Indian CAS strains was approximately 8% between day 29 and 43, whilst the smallest difference was approximately 5% at day 23. The percentage weight change curve for guinea pigs infected with the CAS strain isolated from the patient born in Somalia was very similar to, but just below, the curve for the Indian CAS-infected guinea pigs.

Statistical analysis using the Mann-Whitney test ($p<0.05$) proved that there were significant differences in the percentage body weight change between guinea pigs infected with the Beijing and Indian CAS strains, at days 23, 29, 37, 43 and 57 post-challenge ($p=0.024$, 0.014, 0.018, 0.007, 0.041, respectively). At day 37 post-challenge there was a significant difference between animals infected with Beijing and Somalia CAS strains ($p=0.024$) and at day 43 there was a significant difference between guinea pigs infected with Indian CAS and LAM10 strains ($p=0.031$).

The differences in percentage weight change of guinea pigs infected with the different *M. tuberculosis* strains was further emphasised when looking at the overall change in body weight by calculating the area under the curves (Figure 8.1b). There was a 200 unit^2 difference between the area under the curves for percentage body weight change in guinea pigs infected with the Beijing and Indian CAS strain, with the smallest area occurring in animals infected with the Beijing strain. This difference was statistically significant ($p=0.007$) using the Mann-Whitney test. The values for area under the curves for the other strains were around 500 units^2 and although these values were intermediate to those for guinea pigs infected with the Beijing and Indian CAS strains, there was a significant difference between the area under the curves for the Indian CAS and LAM10 ($p=0.031$).

Due to the limited number of guinea pigs available for the present project, a group of guinea pigs infected with the virulent laboratory control strain *M. tuberculosis* H37Rv was not included in the present study. However, there was data for the laboratory control strain available from previous studies performed at HPA, CEPR (Movahedzadeh *et al.,* 2008; Vipond *et al.,* 2008). There was also data for a negative control group of guinea pigs infected with an *M. tuberculosis* double mutant strain, which had deletions in the inositol-1-phosphate synthase gene, *ino1*, and *trpD*, a gene encoding the tryptophan biosynthetic enzyme, anthranilate phosphoribosyltransferase (Movahedzadeh *et al.,* 2008). The data from the study by Movahedzadeh *et al*., 2008 demonstrated that this nutritional mutant strain was severely attenuated after infecting guinea pigs therefore the data produced after using this strain to infect guinea pigs served as a guide to represent what an unvaccinated-unchallenged weight data set would look like.

The weight of guinea pigs infected with the *M. tuberculosis* double mutant increased steadily over a 60 day period post-challenge and by day 60 there was a 113.6% change in body weight (data not shown). However, the same was not observed in guinea pigs infected with the positive laboratory control strain, *M. tuberculosis* H37Rv (data not shown). There was an increase in body weight of the guinea pigs infected with *M. tuberculosis* H37Rv until day 36 post-challenge, but there was only an 8.0% body weight change between day 0 and 36. After day 36 the body weight started to decrease so the final percentage body weight change at day 63 post-challenge was 0.85%.

Although this data was from a previous study it provided an insight into any differences in the percentage body weight change between guinea pigs infected with the laboratory *M. tuberculosis* H37Rv strain and the five clinical *M. tuberculosis* strains included in the present guinea pig study. The overall percentage body weight change 56 days post-challenge for each of the clinical *M. tuberculosis* strains, *M. tuberculosis* Beijing, CAS (India), CAS (Somalia), EAI5 and LAM10 was 19.4%, 24.6%, 23.6%, 20.8% and 20.6%, respectively.

8.2.2 The effect on lung weight to body weight ratio

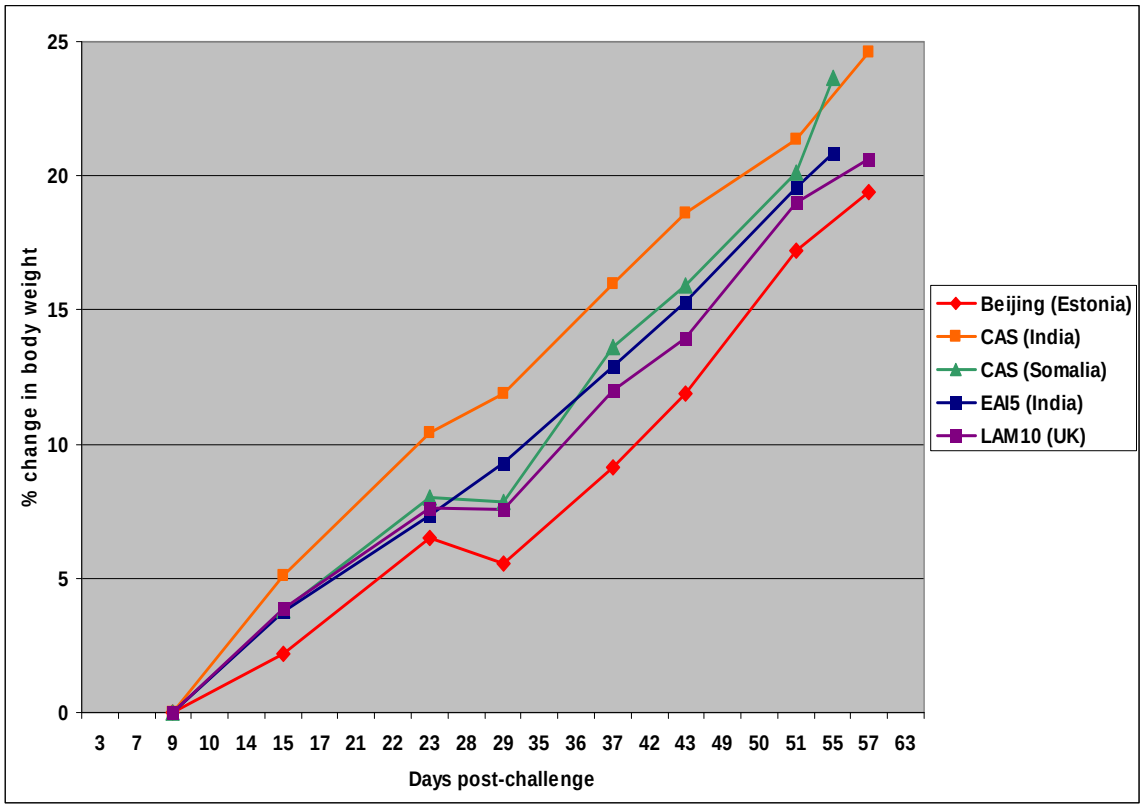
Each of the five clinical *M. tuberculosis* strains was used to infect sixteen guinea pigs and at two time points, day 16 and 56, eight guinea pigs from each group were culled. At necropsy at day 56 in addition to recording the total body weight, the weight of the lungs was recorded. The lung weight to body weight ratio was calculated and values for each guinea pig plotted along with the mean ratio for each group (Figure 8.2).

When looking for differences between the circulating strains, the mean lung to body weight ratio for guinea pigs infected with the Beijing strain was higher than the means for the other four circulating strains. The lowest mean ratio was observed with the groups of guinea pigs infected with both of the CAS strains, with the mean ratio for animals infected with Somalia CAS being slightly lower than the Indian CAS. After performing the Mann-Whitney statistical analysis test ($p < 0.05$), it was concluded that the lung to body weight ratio of guinea pigs infected with the Beijing strain was significantly higher than the ratios obtained when guinea pigs had been infected with Indian and Somalia CAS strains (p=0.041 and 0.024, respectively). In addition, the ratios of guinea pigs infected with EAI5 were significantly higher than the ratios of animals infected with the Somalia CAS (p=0.018).

There was lung to body weight ratio data from the historical controls in previous studies described in Section 8.2.1 where guinea pigs were challenged with *M. tuberculosis* H37Rv and the *M. tuberculosis* double mutant strain (Movahedzadeh *et al.,* 2008; Vipond *et al.,* 2008). There were some differences when comparing the lung to body weight ratios of guinea pigs infected with the circulating strains in the present study with the data from previous experiments. The lung to body weight ratios of guinea pigs challenged with the double mutant strain (negative control; mean lung to body weight ratio=4.16; data not shown) was approximately three times lower than the ratios of the guinea pigs that had been infected with the circulating and laboratory strains. In contrast, the mean lung to body weight ratio of the group of guinea pigs infected with *M. tuberculosis* H37Rv (mean lung to body weight ratio=15.5; data not shown) was approximately 0.7 times higher than the ratios of the circulating strains.

Figure 8.1 *Graphs showing (a) the mean percentage change in body weight of guinea pigs post-challenge with each of the five M. tuberculosis strains (Beijing; Indian CAS; Somalia CAS; EAI5; LAM10) and (b) the area under each of the percentage growth change in body weight curves, in Figure 8.1a, for the five circulating strains.*

a)



b)

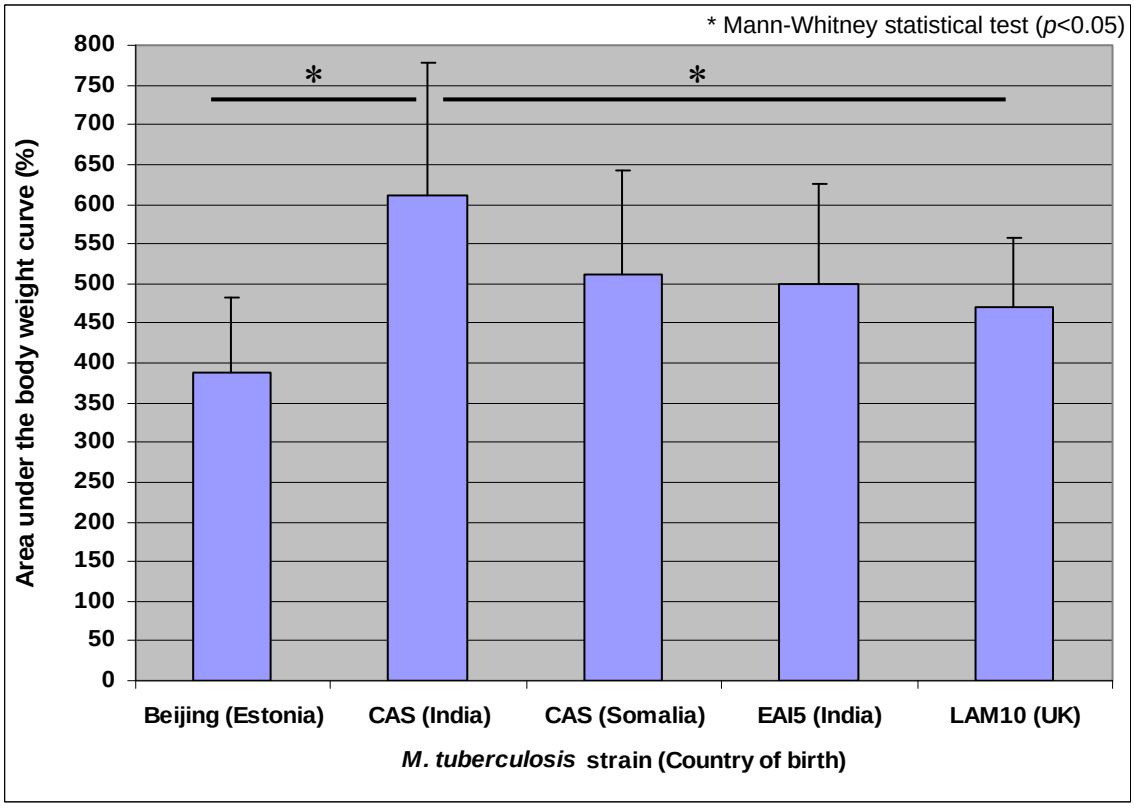
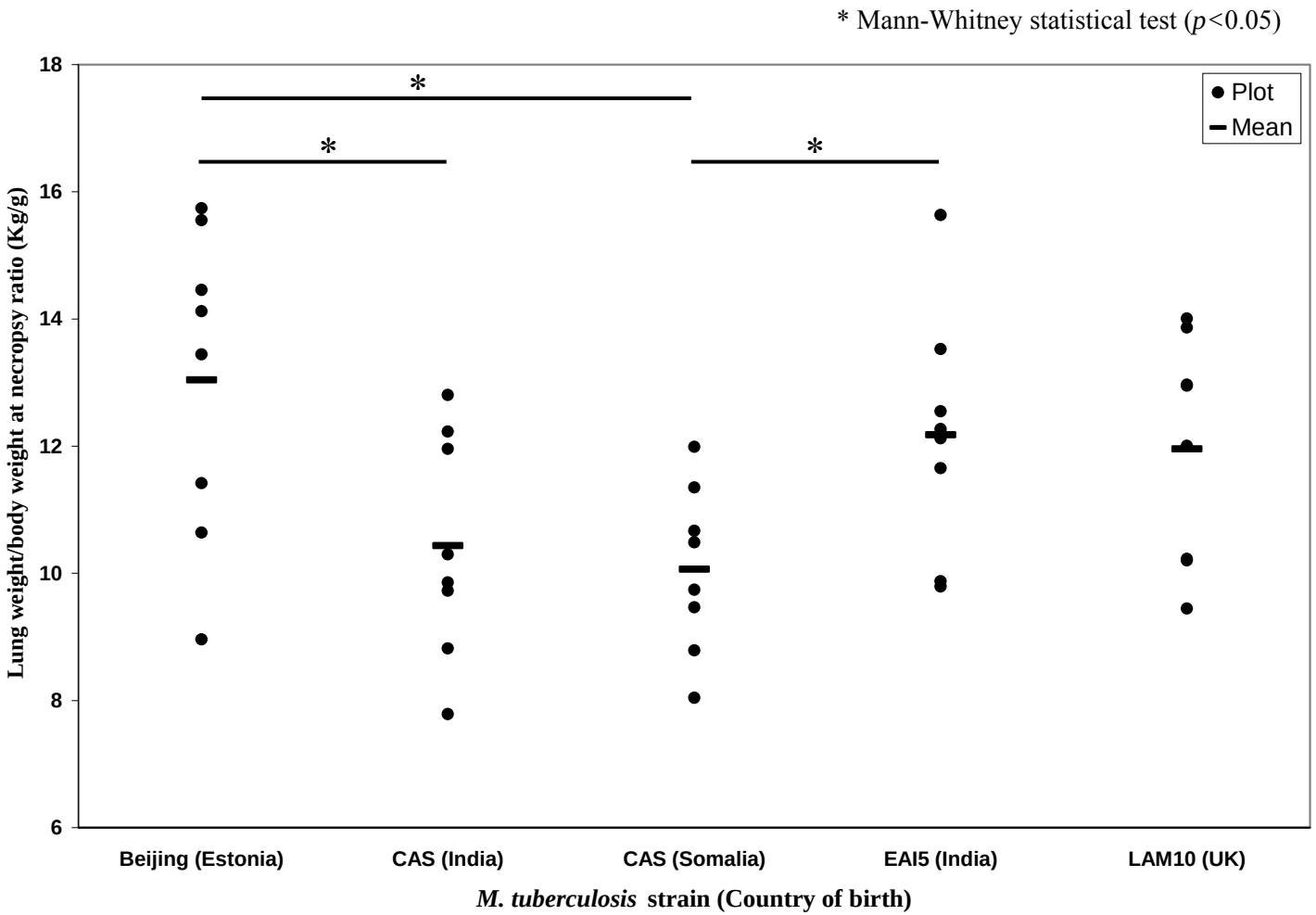


Figure 8.2 *A graph showing the lung to body weight ratio at the time of necropsy 56 days post-challenge of guinea pigs with each of the five M. tuberculosis strains (Beijing, Indian CAS, Somalia CAS, EAI5, LAM10), M. tuberculosis H37Rv challenged control guinea pigs and control guinea pigs [both sets of control data were provided by Vipond et al., 2008 and Movahedzadeh et al., 2008]. Each point represents lung to body weight ratio for an individual guinea pig and horizontal lines represent the mean lung to body weight ratio in each group.*



8.3 THE DISSEMINATION OF MYCOBACTERIA FROM THE LUNGS TO THE SPLEEN OF GUINEA PIGS

8.3.1 Mycobacterial load at day 16 post-challenge

Eight guinea pigs from each group that had been aerosol challenged with each of the five *M. tuberculosis* strains were culled at 16 days post-challenge. The lung and spleen from each animal was dissected and after homogenisation, samples were plated onto Middlebrook 7H11 agar plates. After plates had been incubated for 4 weeks, the number of colonies in all plates was counted and for each animal the c.f.u./ml in the lungs and spleen was calculated and plotted onto graphs along with the mean c.f.u./ml for each group of eight guinea pigs (Figure 8.3). Calculating c.f.u./ml in lungs and spleen at the early time point of 16 days post-challenge provided an insight into the infectivity of the *M. tuberculosis* strains *in vivo*.

In general, the c.f.u./ml in the lungs for each group of guinea pigs was higher than the c.f.u./ml in the spleen. When looking at the c.f.u./ml in the lungs, it was observed that there were similar numbers of c.f.u./ml in guinea pigs that had been infected with the Beijing, Somalia CAS and EAI5 strains with approximately $1\times10^{3.5}$ c.f.u./ml, whilst lower counts were observed in the guinea pig groups infected with the Indian CAS and LAM10 strains with approximately $1\times10^{3.2}$ c.f.u./ml.

From those groups of guinea pigs with higher numbers of c.f.u. in lungs, animals infected with the EAI5 strain had the highest c.f.u./ml in the spleen demonstrating a higher rate of dissemination. The spleen counts in the groups infected with the Beijing and Somalia CAS strains were low at approximately 1×10^{1} c.f.u./ml, showing a slower rate of dissemination. The lowest c.f.u./ml in the spleen was observed for the guinea pigs infected with the LAM10 strain.

In general, at the early time point, it could be concluded that guinea pigs had been successfully infected with TB. There was a slow rate of dissemination of mycobacteria from the lungs to the spleen in guinea pigs infected with *M. tuberculosis* Beijing, Indian and Somalia CAS, and LAM10 strains. The faster rate of dissemination 16 days after infecting guinea pigs with the EAI5 strain could mean that this strain was more infective than the other strains, demonstrating that there are differences in the behaviour of the strains in an *in vivo* model at an early time point.

8.3.2 Mycobacterial load at day 56 post-challenge

The remaining guinea pigs in each group were culled at a later time point of 56 days post-challenge to assess the development of infection. The dissected lung and spleen tissue was treated in exactly the same way as tissues dissected at day 16 post-challenge (Figure 8.4). There were clear differences in the way the strains behaved as infection had clearly progressed further.

There was an increase of mycobacterial load in lungs for guinea pigs that had been infected with the Beijing and LAM10, indicating that these strains had more potential to persist during infection. The increase in mycobacterial load in the lungs was greater in those guinea pigs infected with LAM10. For the other groups of guinea pigs infected with the CAS and EAI5 strains, there was a decrease in mycobacterial load in the lungs (compared to day 16; Figure 8.3) as infection progressed, which suggested that the guinea pigs were able to control infection with these strains more effectively than infection caused by the Beijing and LAM10 strains. The decrease in mycobacterial load was more prominent in guinea pigs infected with the EAI5 strain and after performing the Mann-Whitney statistical analysis test ($p < 0.05$) there was a significant difference between the groups of guinea pigs infected with EAI5 and LAM10 (p=0.018).

There was a general increase in mycobacterial load in the spleen for all groups of guinea pigs as infection had progressed to day 56 post-infection. The largest increase in mycobacterial load, of approximately $1\times10^{3.2}$ c.f.u./ml, in the spleen was observed in guinea pigs infected with the Beijing strain, indicating that this strain was more virulent than the others as there was increased dissemination of mycobacteria from the lung to the spleen during the progression of infection. A marked increase in bacterial load of approximately $1\times10^{2.5}$ c.f.u./ml was observed in spleens isolated from guinea pigs infected with the LAM10 strain. The increased load in animals infected with these two strains is reflected by the increased mycobacterial load in the lungs from these guinea pigs, which reinforced the initial suggestion that these strains were more persistent and potentially more virulent. The mycobacterial load in the spleen was significantly higher in guinea pigs infected with the Beijing strain than in guinea pigs infected with the Indian and Somalia CAS and EAI5 strains (p=0.003, 0.041, 0.031, respectively). Also there was a significantly higher mycobacterial load in the spleens

of guinea pigs infected with LAM10 strain than those infected with the India CAS strain (p=0.024).

The smallest increase in the mycobacterial load in the spleen was seen in guinea pigs infected with EAI5 strain, which is interesting as the spleen mycobacterial load 16 days post-challenge was higher than the spleen mycobacterial load of guinea pigs infected with the other four strains. This suggested that whilst dissemination of the EAI5 strain was quicker than the other strains in the early stages of infection, as infection progressed, the immune system of the guinea pigs was able to control infection as there was no significant increase in the dissemination of mycobacteria from the lung to spleen 56 days post-challenge.

8.4 HISTOPATHOLOGY OF TISSUE 56 DAYS POST-CHALLENGE

8.4.1 Histopathology of spleen tissue

At 56 days post-challenge, as well as determining the mycobacterial load in spleen and lung tissue and weight of lungs, sections of lung and spleen were taken and stained with haematoxylin and eosin for routine examination, with van Giesen staining method for detecting encapsulated lesions and with Alizarin Red stain for identifying calcified lesions. Slides were examined by a veterinary pathologist and all lung and spleen sections were scored according to the size of lesions present and the degree of consolidation, which was displayed by a specific morphological change of the lesion. The scores were assigned according to the criteria shown in Table 2.2; Section 2.9.3. The number of foci of caseation and calcified lesions were also recorded for lung tissue. The histopathology scores for the spleen and lung tissue and the numbers of caseated foci and calcified lesions in lung tissue from guinea pigs infected with the five different *M. tuberculosis* strains is shown in Figure 8.5.

In the spleen sections, the least pathology (average score=1) was observed in guinea pigs infected with the Indian CAS strain (refer to Figure 8.5a; refer to Figure 8.6bi for a histopathology image from one of the spleen sections taken from a guinea pig challenged with the Indian CAS strain). This degree of pathology when compared with the other strains correlated with the mycobacterial load in the spleen 56 days post-challenge. The highest observed pathology (average score=4) was in animals infected with the Beijing strain (refer to Figure 8.6ai for a histopathology image from

one of the spleen sections taken from a guinea pig challenged with the Beijing strain), with significant pathology also being observed for the Somalia CAS and EAI5 strains (refer to Figure 8.6ci, and di for a histopathology image from one of the spleen sections taken from a guinea pig challenged with the Somalia CAS and EAI5 strains, respectively). Again, this correlated with the mycobacterial load in the spleen of these animals. Surprisingly, despite the fact that there was a higher mycobacterial load in the spleens taken from guinea pigs infected with the LAM10 strain than the Somalia CAS and EAI5 strains, the pathology score was lower (average score=3) (refer to Figure 8.6ei for a histopathology image from one of the sections taken from a guinea pig challenged with the LAM10 strain).

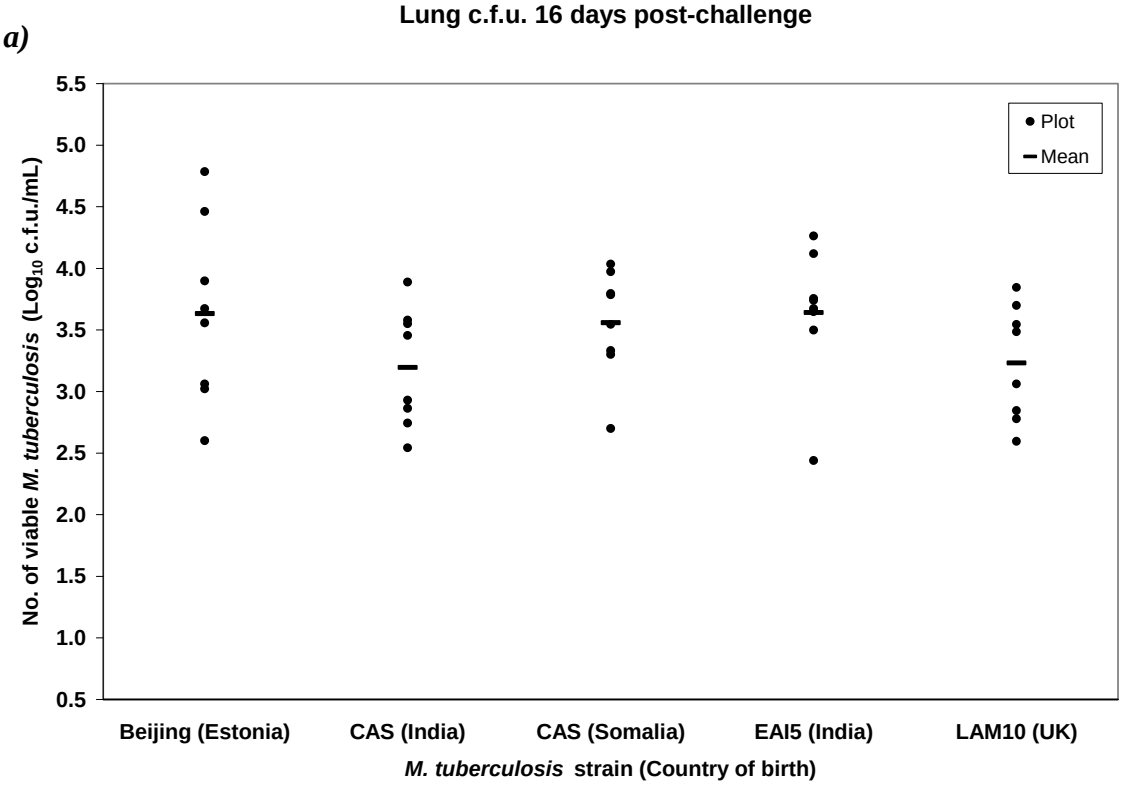
8.4.2 Histopathology of lung tissue

The overall score for the degree of consolidation in the lung tissue was higher than in the spleen tissue. As with the spleen tissue, the lowest score was given to lung sections taken from guinea pigs infected with the Indian CAS strain (refer to Figure 8.6bii for a histopathology image of one of the lung sections taken from a guinea pig challenged with the Indian CAS strain). The score for Indian CAS strain was in accordance with the mycobacterial load observed in the lung tissue. A similar score was assigned to lung sections taken from guinea pigs infected with LAM10 strains, which as with the spleen sections, did not correlate with the mycobacterial load observed in lungs at day 56 post-challenge. Refer to Figure 8.6eii for a histopathology image of one of the lung sections taken from a guinea pig challenged with the LAM10 strain.

As with the spleen sections, the higher scores were assigned to lung sections taken from guinea pigs infected with the Beijing, Somalia CAS and EAI5 strains (refer to Figure 8.6aii, cii, and dii for a histopathology image of one of the lung sections taken from a guinea pig challenged with the Beijing, Somalia CAS and EAI5 strains, respectively). These higher scores correlated with the mycobacterial load in lungs isolated from guinea pigs infected with Beijing and Somalia CAS.

Figure 8.3 *Graphs showing the number of c.f.u./ml in (a) the lungs and (b) the spleen, which were dissected 16 days post-challenge from five groups of 8 guinea pigs infected with M. tuberculosis strains (Beijing, Indian CAS, Somalia CAS, EAI5, LAM10). Each point represents the c.f.u. count for individual guinea pigs and horizontal lines represent the mean c.f.u. counts in each group.*
(No significant difference between strains when performing Mann-Whitney statistical test $p<0.05$)

a)



b)

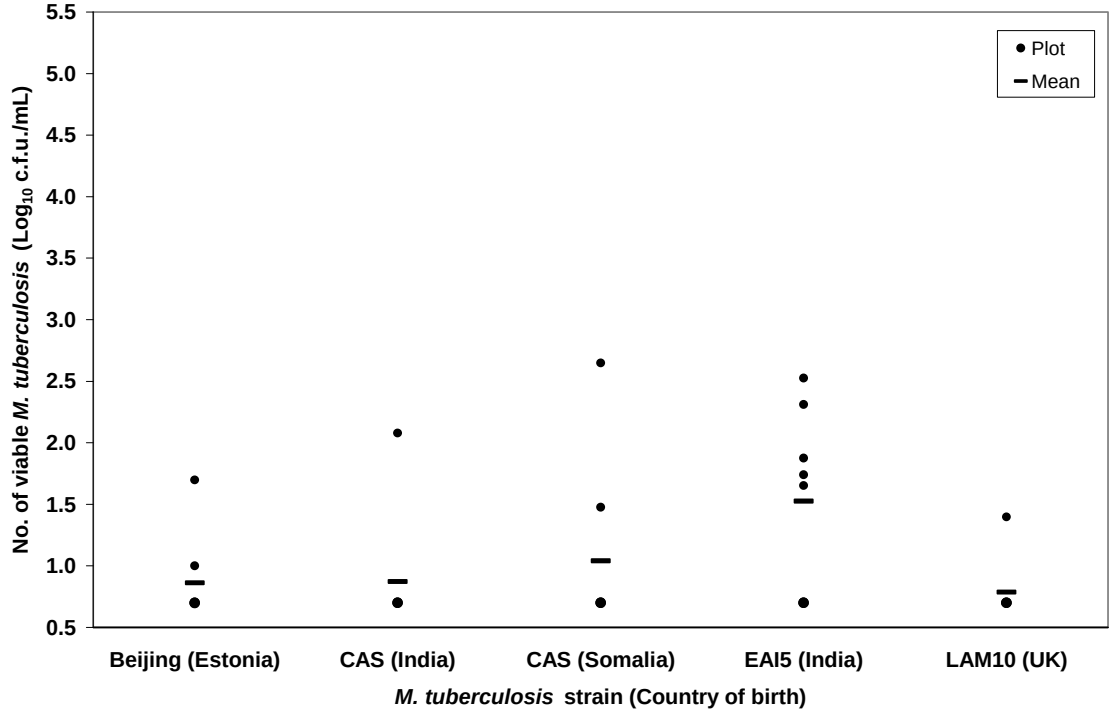
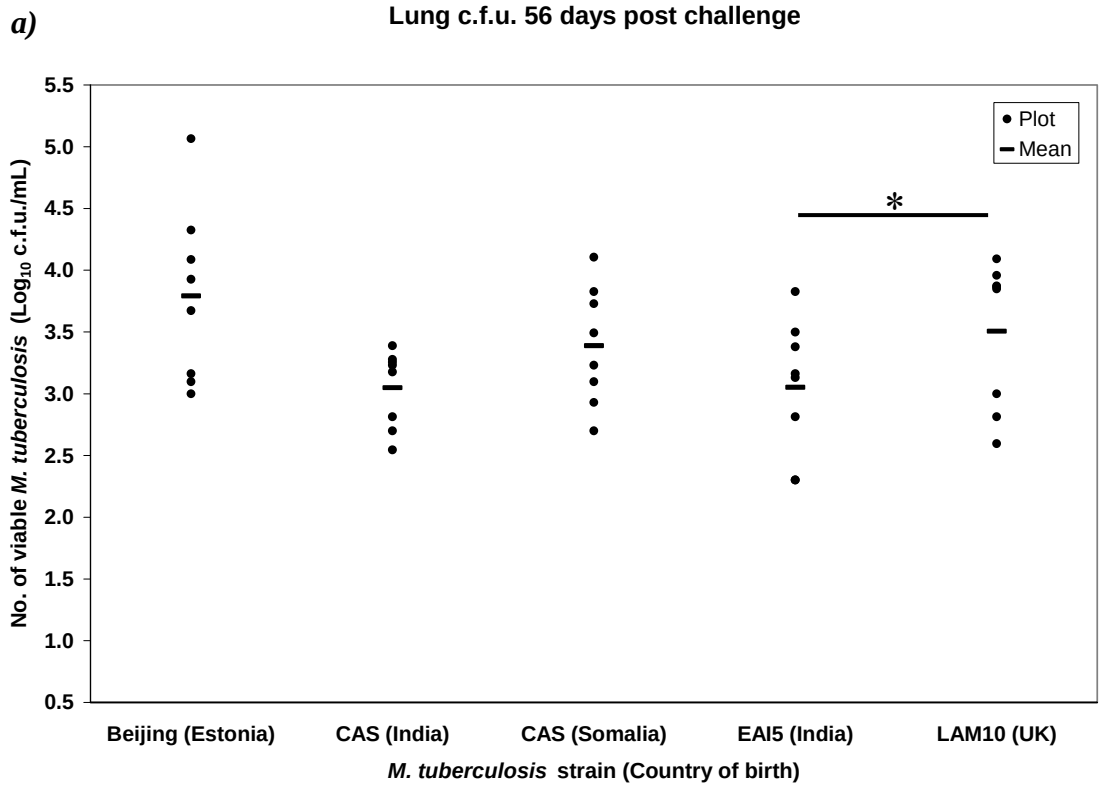


Figure 8.4 *Graphs showing the number of c.f.u./ml in (a) the lungs and (b) the spleen, which were dissected 56 days post-challenge from five groups of 8 guinea pigs infected with M. tuberculosis strains (Beijing, Indian CAS, Somalia CAS, EAI5, LAM10). Each point represents the c.f.u. count for individual guinea pigs and horizontal lines represent the mean c.f.u. counts in each group.*

* Mann-Whitney statistical test ($p<0.05$)

a)



b)

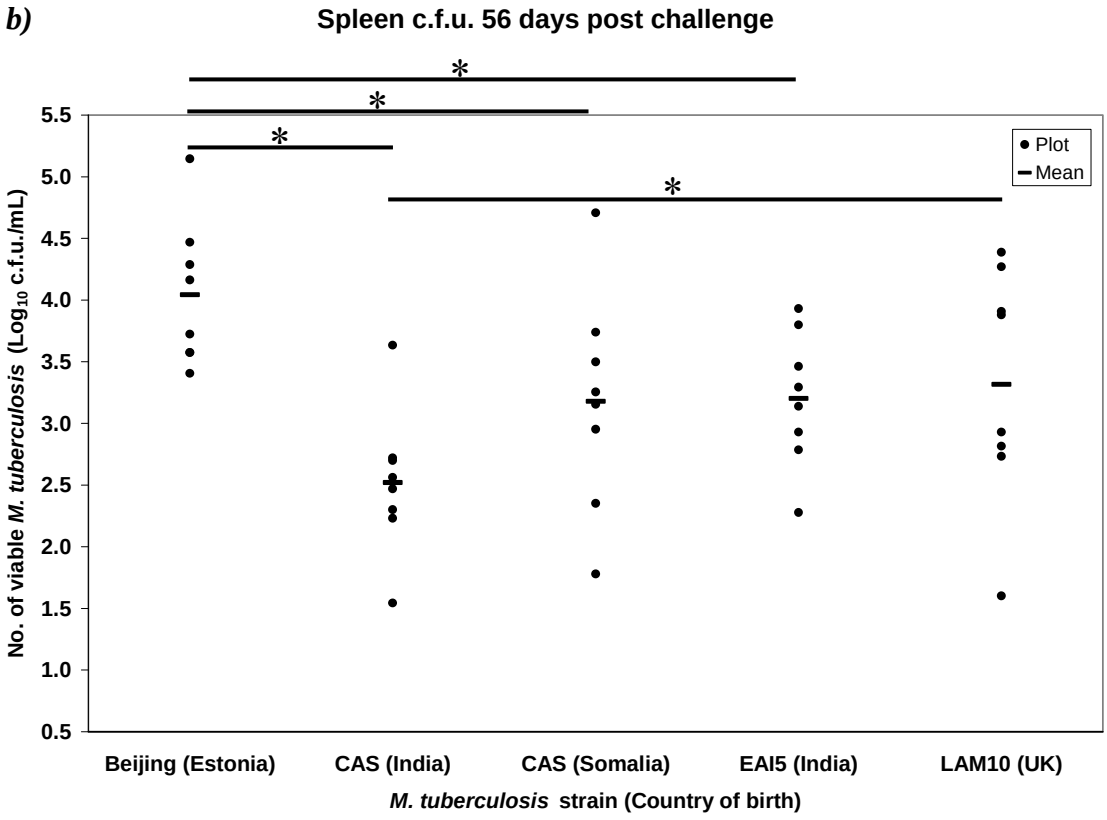
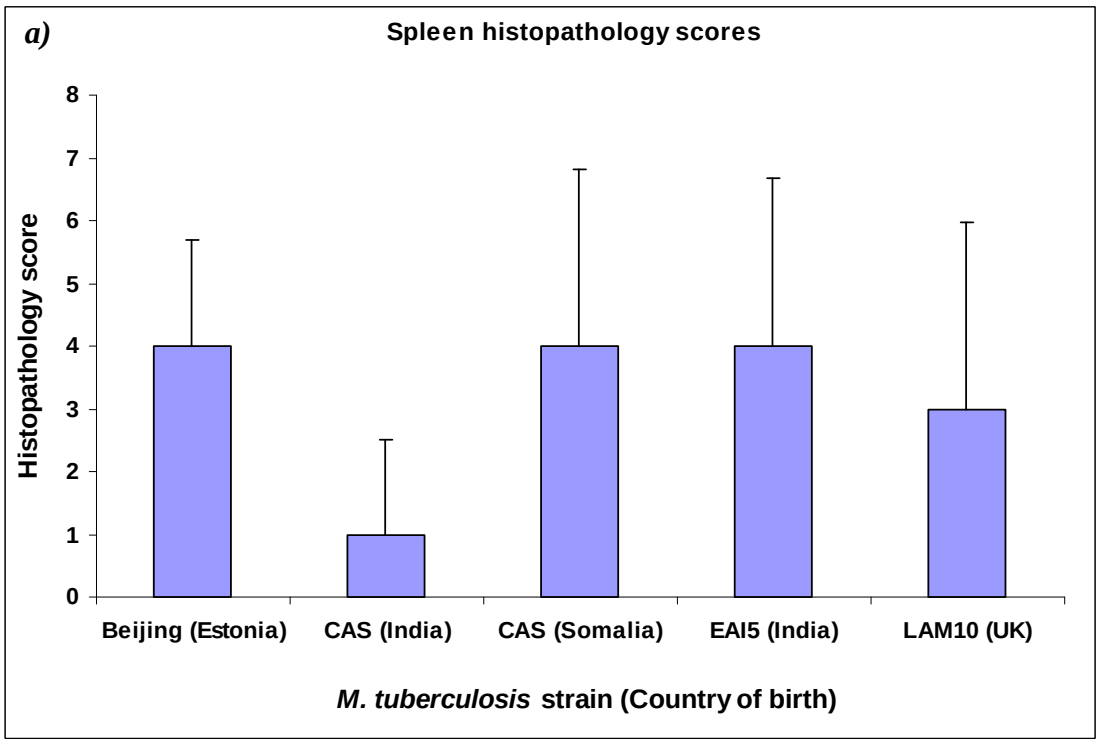


Figure 8.5 *Graphs showing (a) the average histopathology scores, with standard deviation bars, for each group of 8 guinea pigs whose spleens were dissected 56 days post-challenge with the five M. tuberculosis strains and (b) the average histopathology score for the same groups of guinea pigs whose lungs were dissected 56 days post-challenge with the five strains, but also shown is a breakdown of the scores for consolidation, caseation/necrosis, and calcification seen during examination.*



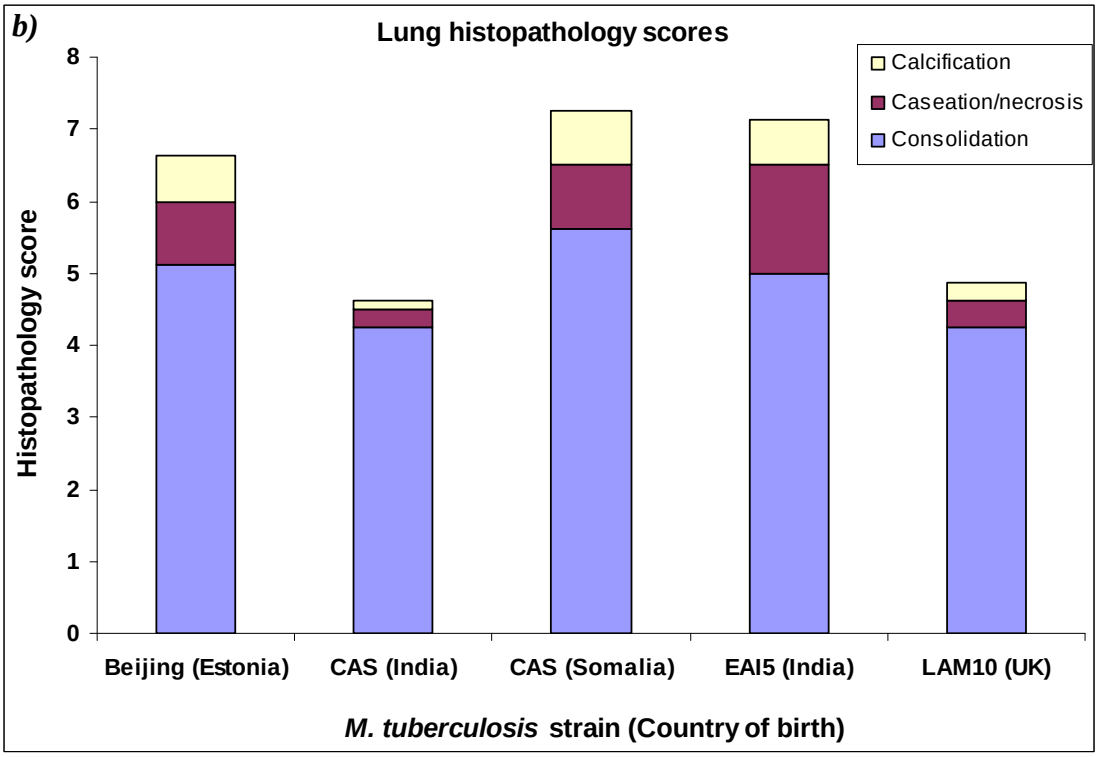
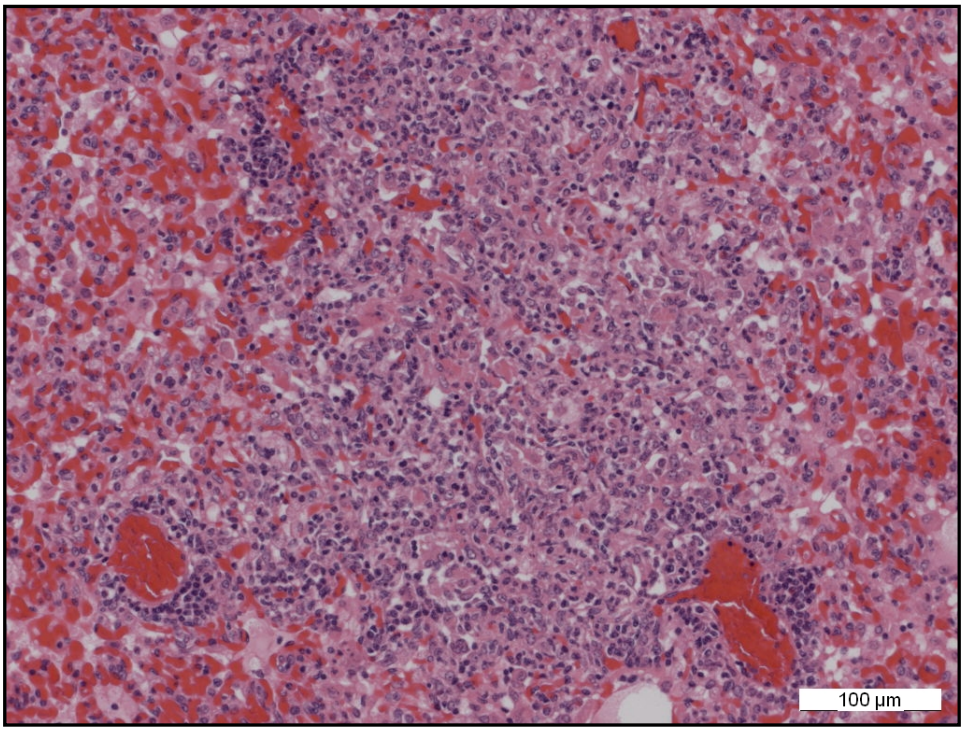


Figure 8.6 *Histopathology images of guinea pig i) spleen and ii) lung sections following dissection 56 days post-challenge with (a) M. tuberculosis Beijing (Estonia); (b) M. tuberculosis CAS (India); (c) M. tuberculosis CAS (Somalia); (d) M. tuberculosis EAI5 (India); and (e) M. tuberculosis LAM10 (UK) strains and staining with haematoxylin and eosin. (Refer to the scale bar on individual images for measurements and refer to Table 2.2 for the criteria used for histopathology scoring of sections).*
NOTE: Histopathology scores in Figure 8.5 are averages of all sections taken from guinea pigs in each of the five groups challenged with each of the five *M. tuberculosis* strains and the scores assigned in Figure 8.6 are for the images shown; therefore there may be differences in the scores.

a) M. tuberculosis Beijing (Estonia)

i) Image of a spleen section illustrating a histopathology score of 4 with 11 moderate to large sized lesions.

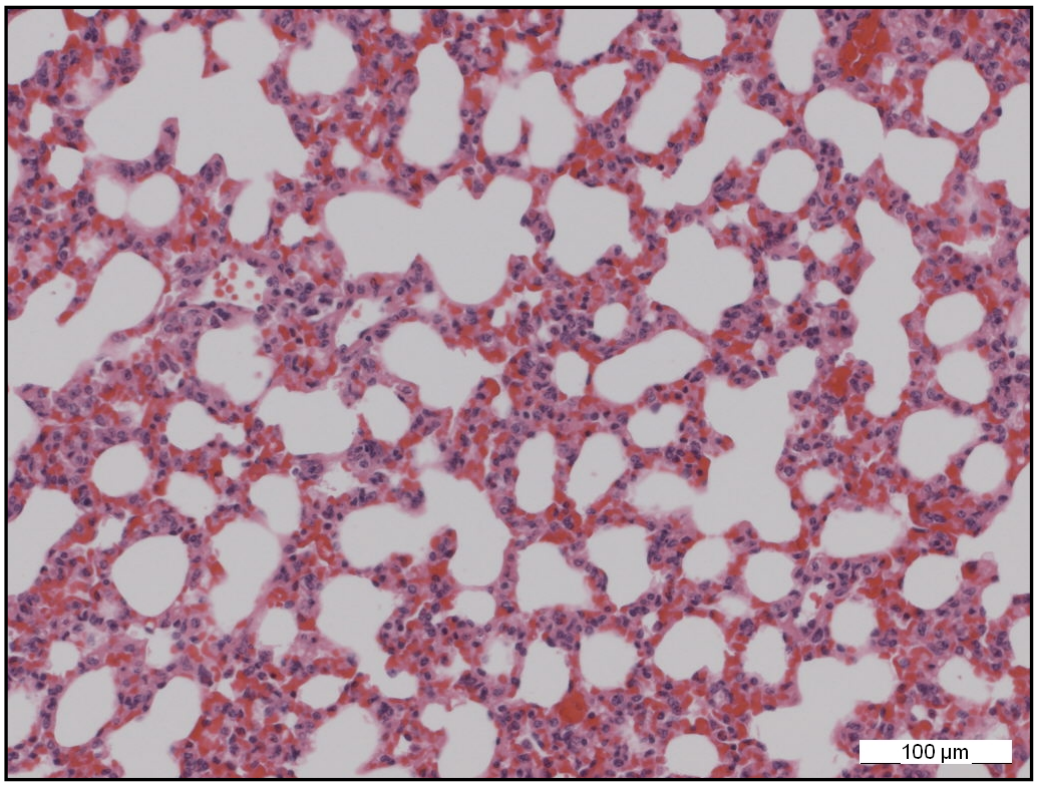
ii) Image of a lung section illustrating epithelioid macrophages and evenly distributed lymphocytes surrounded by peripheral blood vessels (with lymphocyte cuffs), all of which together formed an unorganised granuloma.



b) M. tuberculosis CAS (India)

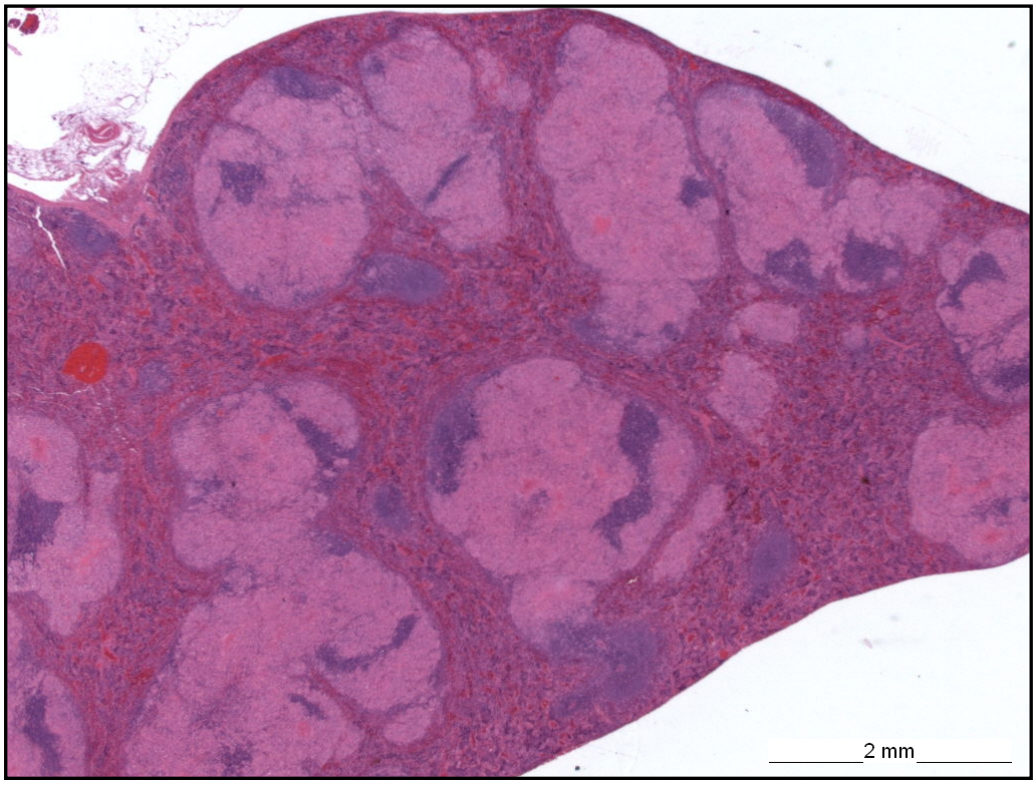
i) Image of a spleen section illustrating a histopathology score of 0 with no detected lesions or abnormalities.

ii) Image of a lung section illustrating a histopathology score of 0 as no lesions or abnormalities were detected.

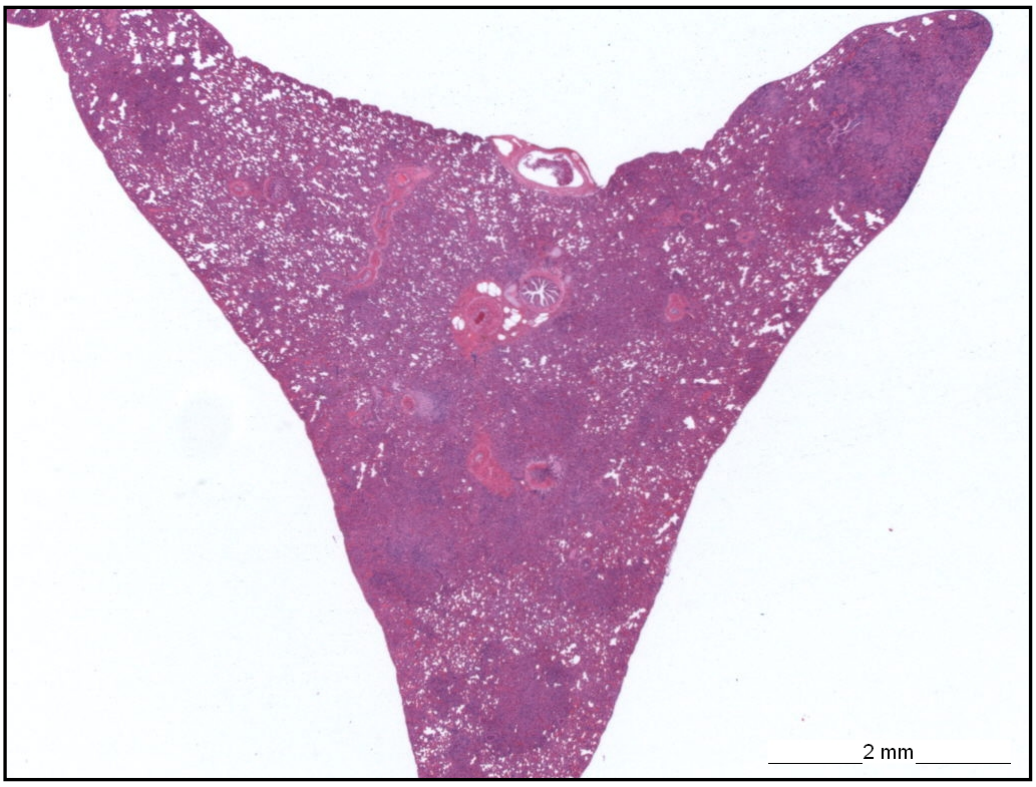


c) M. tuberculosis CAS (Somalia)

i) Image of a spleen section illustrating a histopathology score of 9 with more than 30 large sized lesions some of which are caseated.

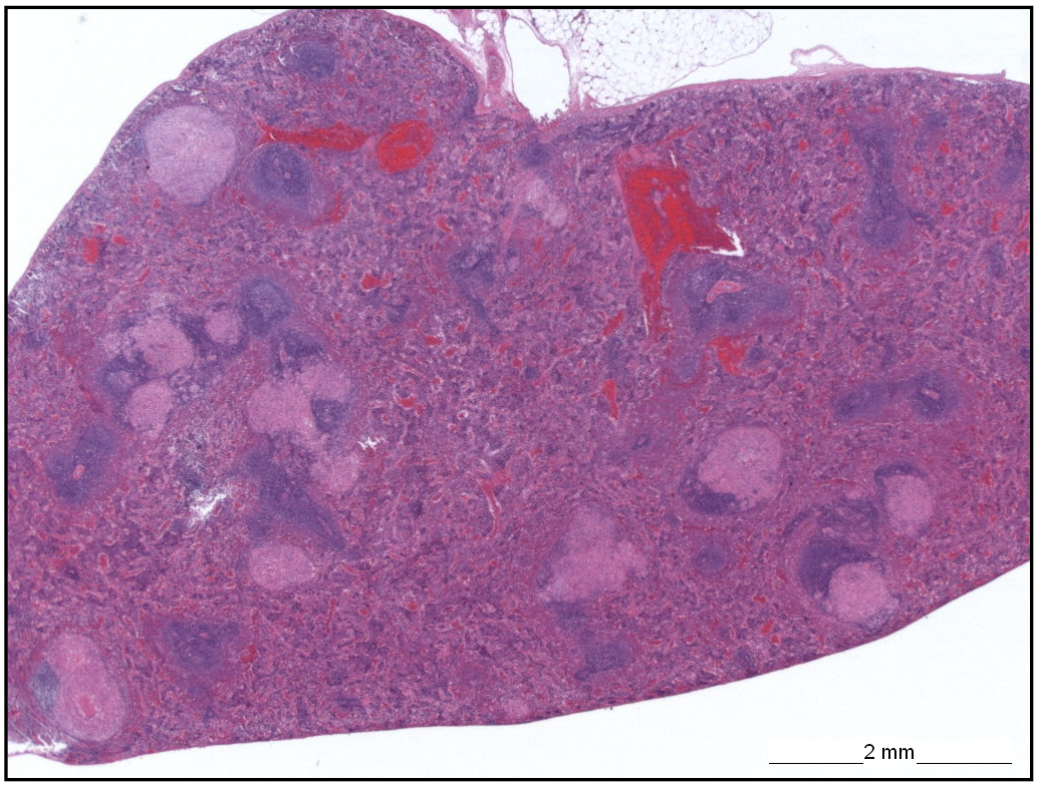


ii) Image of a lung section illustrating a histopathology score of 3 with medium sized lesions and 20-33% consolidation.

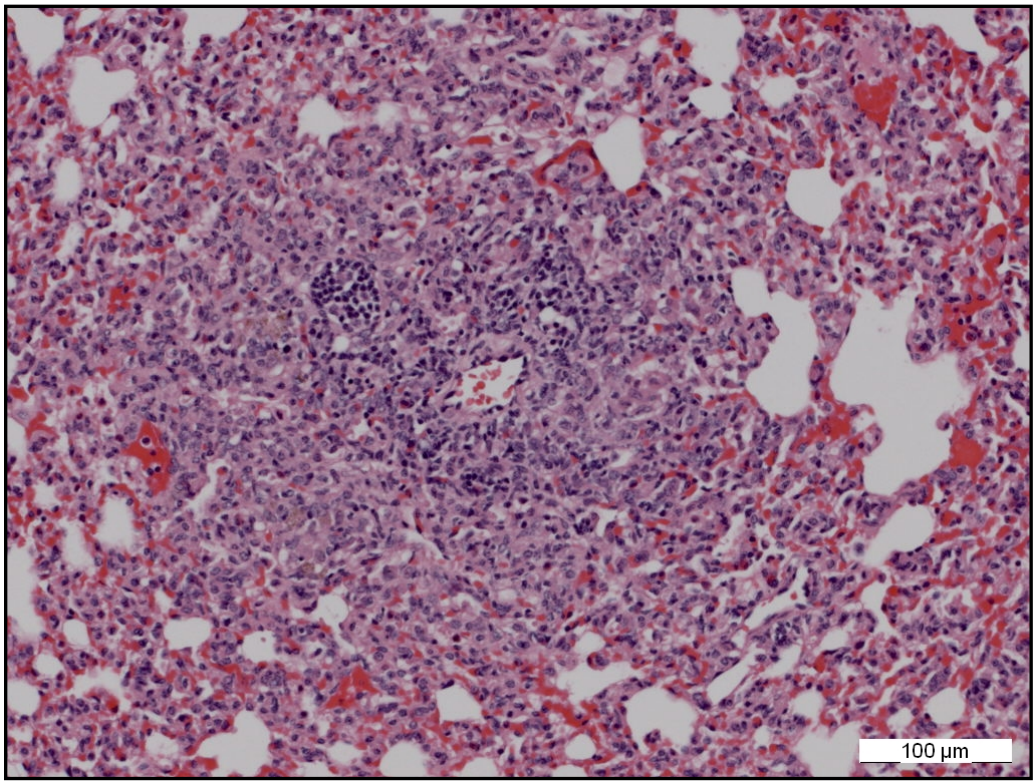


d) M. tuberculosis EAI5 (India)

i) Image of a spleen section illustrating a histopathology score of 5 with 22 lesions.

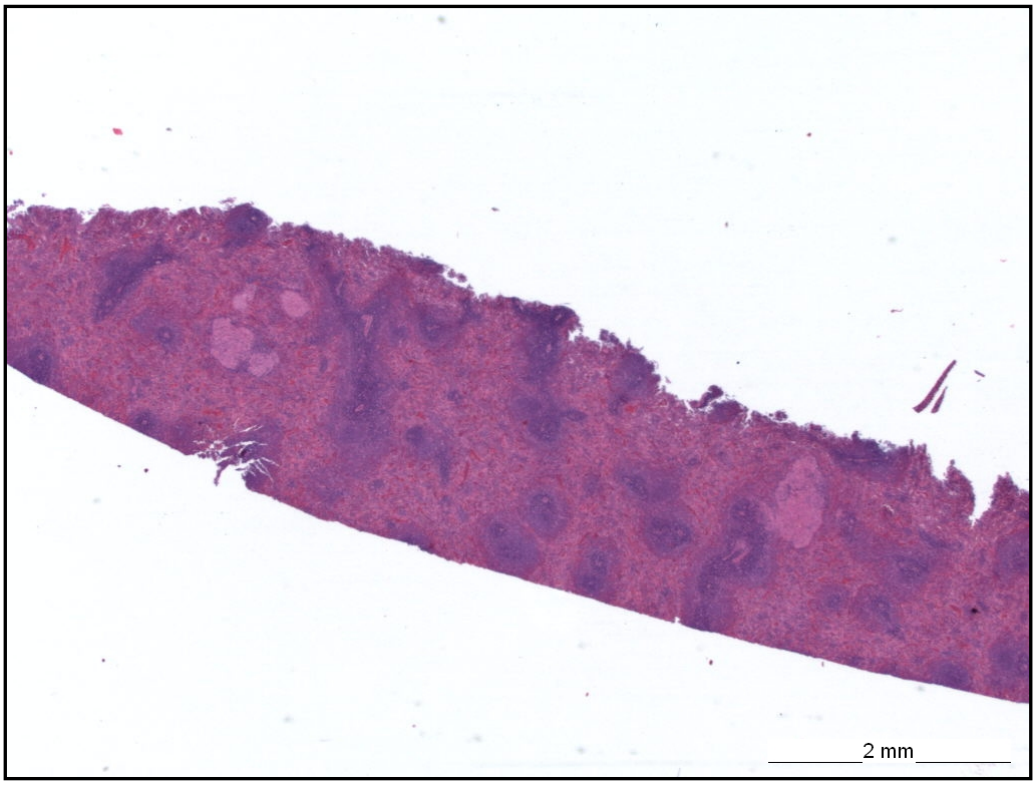


ii) Image of a lung section illustrating a histopathology score of 1 with a small unorganised granuloma.

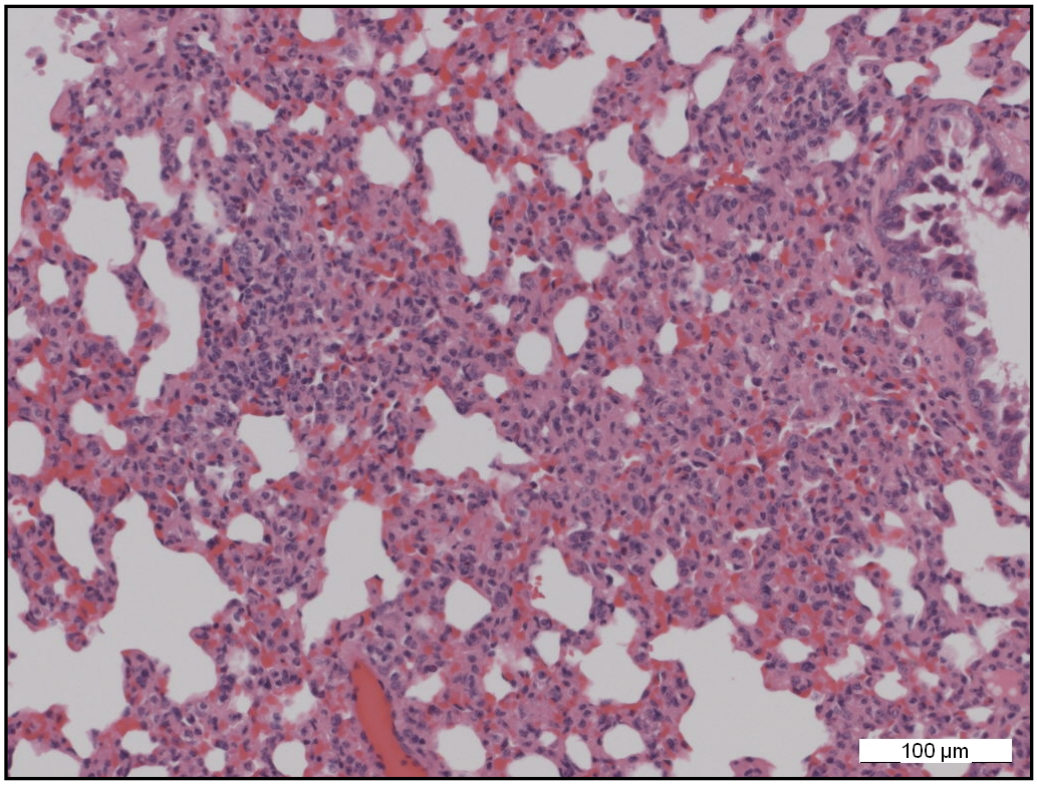


e) M. tuberculosis LAM10 (UK)

i) Image of a spleen section illustrating a histopathology score of 2 with 6 small non-caseated lesions.



ii) Image of a lung section illustrating two small unorganised granuloma with a histopathology score of 1.



There were no significant differences between the *M. tuberculosis* strains in the extent of caseation/necrosis and calcification seen in the lung sections (refer to Figure 8.5b). There was less caseation/necrosis and calcification in lung sections taken from guinea pigs infected with the Indian CAS and LAM10 strains than the other strains. This means that mycobacterial load would not necessarily correlate with the degree of caseation/necrosis and calcification as there was a lower load in lungs isolated from the guinea pigs infected with the Indian CAS than in lungs from animals infected with the LAM10 strain. Similarly, the mycobacterial load in lungs dissected from guinea pigs infected with EAI5 strain was as low as Indian CAS, but there was more caseation/necrosis and calcification seen in the EAI5 infected lung sections than any of the other strains. So while there are fewer *M. tuberculosis* EAI5 bacilli present, they may cause more pathological damage to the infected tissue. There was a similar level of caseation/necrosis and calcification seen in lung sections taken from guinea pigs infected with the Beijing and Somalia CAS strains.

8.5 DISCUSSION

The protocol that was used in the present study for the aerosol challenge of guinea pigs had previously been evaluated for the purpose of comparing infectivity of *M. tuberculosis* strains and whilst testing the null hypothesis, the data offered additional information about the degree to which strains were pathogenic (Williams *et al.,* 2005). In the study by Williams *et al*., 2005, a higher dose of mycobacterial bacilli had been administered to the guinea pigs via aerosols delivered to the snouts of the guinea pigs, but the study had shown that the validity of the assay remained with the lower dose used in the present study and that such a dose would more closely resemble natural doses. However, for the purposes of comparing data from the present study with data from previous studies the administered dose is an important factor to take into consideration.

The weight of guinea pigs was assessed post-challenge and this can be used as a sensitive read-out of the extent of infection. The lower increase in weight gain of guinea pigs infected with the Beijing strain when compared with the data for guinea pigs infected with the other four strains suggested that this strain was causing more severe disease than the other strains.

Supplementing the weight change data with the lung to body weight ratio data further demonstrated that there was a difference in the effect of the Beijing strain on disease progression in guinea pigs. Comparing the data for the five strains with the data from the non-infected guinea pigs proved that the weight gain patterns of guinea pigs infected with the five strains was caused by infection with the strains. Comparison with the previous data from guinea pigs infected with *M. tuberculosis* H37Rv showed that there were some differences in the weight gain patterns between circulating and laboratory strains (Movahedzadeh *et al.,* 2008; Vipond *et al.,* 2008). However, as mentioned earlier it is important to remember that the dose of H37Rv administered to the guinea pigs was higher than the dose administered for the five circulating strains making direct comparison difficult.

Mycobacterial load was evaluated in lungs and spleens from half of the guinea pigs from each group infected with the different strains at the early time point of day 16 post-challenge. Day 16 post-challenge is a very early time-point in guinea pigs and although it is expected that little to no dissemination from the lungs to the spleen would take place, the presence of any mycobacteria in the spleen 16 days post-challenge has been used as a marker of infectivity and pathogenicity of *M. tuberculosis* strains (Williams *et al.,* 2005). Other studies have also demonstrated that virulence of clinical isolates of *M. tuberculosis* strains can be assessed by calculating the number of viable mycobacterial cells in the spleen (Prabhakar *et al.,* 1987; Balasubramanian *et al.,* 1992).

Using this principle and the c.f.u. data from lungs and spleen at day 16 obtained from the present study, we would conclude that the Beijing, Somalia CAS and EAI5 strains all shared some of the markers of higher pathogenicity compared to the Indian CAS and LAM10 strains (reflected in the very low dissemination of bacilli from the lung to the spleen in the latter two strains).

Data for mycobacterial burden in the lungs and spleen was also obtained at a later time point of 56 days post-challenge to further assess virulence in terms of persistence relative to day 16 post-challenge. The pathogenicity of the strains was also evaluated. The pathology of lung and spleen tissue involved looking at the degree of consolidation, caseation/necrosis and calcification. Consolidation scores are an important marker for loss of lung tissue functionality, and general pathology of tissue

post-challenge offered valuable information when considering the virulence of *M. tuberculosis* strains (Dunn *et al.,* 1995; Williams *et al.,* 2005).

Compared to 16 days post-challenge, at the later time point, there was a significant increase in extrapulmonary dissemination of mycobacterial bacilli from the lungs to the spleen of guinea pigs infected with the Beijing strain, whilst there was still strong replication of mycobacteria in the lungs. The histopathology scores for spleen and lungs reflected the mycobacterial loads observed in the organs taken from animals infected with the Beijing strain showing that there was reduced functionality of tissues in these organs 56 days post-challenge. There was also increased necrosis and calcification observed in the lungs of guinea pigs infected with the Beijing strain. Collectively, these data suggested that the Beijing strain was more pathogenic and virulent when compared with the other strains.

The Beijing strain was selected not only to represent *M. tuberculosis* lineage I, but because this Estonian strain was of particular interest for clinical purposes as an MDR strain of the same molecular genotype (12 MIRU and 3 ETR profile) had changed from first appearing in the Estonian population to now dominating the TB population, over a period of 10 years. If the fully sensitive Beijing strain could illicit a significantly more pathogenic effect *in vivo*, this might partly explain the success of this strain in Estonia and reflect on how the MDR strain might have gained dominance.

In a study by Palanisamy *et al*., 2009, the virulence of different clinical *M. tuberculosis* strains was demonstrated using a guinea pig model. The panel of strains included MDR and drug sensitive Beijing and non-Beijing strains (Palanisamy *et al.,* 2009). The study showed that the MDR strains drug sensitive Beijing and non-Beijing strains did not grow as well as the laboratory control strain, *M. tuberculosis* H37Rv, in the lungs, whilst there was higher growth of bacilli observed with the drug sensitive Beijing and non-Beijing strains when compared to *M. tuberculosis* H37Rv. In addition there was a higher lung histopathology score for the drug sensitive strains than *M. tuberculosis* H37Rv and the MDR strains. Palanisamy *et al*., 2009 concluded that clinical *M. tuberculosis* strains caused different degrees of pathology and that the disease seen with the clinical strains differed to the disease caused by *M. tuberculosis*

H37Rv. The same observations were noted in the present study after performing both *in vitro* and *in vivo* analysis of the clinical *M. tuberculosis* strains in Panel B.

A lower mycobacterial burden was seen in the lungs when the Somali CAS strain was used to infect guinea pigs compared to the Beijing strain. There was also significantly less dissemination, but the histopathology scores were very similar, possibly suggesting that not many Somalia CAS bacilli are required to cause pathology. A similar scenario in mycobacterial load in the lungs and spleen was observed for the isoniazid resistant LAM10 strain, which is responsible for the North London, UK, TB outbreak. However, the degree of consolidation, caseation/necrosis, and calcification was lower than that of the Somalia CAS strain, indicating that whilst virulence of the LAM10 strain may be expressed by extrapulmonary dissemination, there was reduced pathology post-challenge.

When comparing the data obtained from guinea pigs infected with the Indian CAS strain with data from the other strains, there were very low levels of replication in the lungs and very low levels of dissemination at day 16, but at day 56, whilst there was still a similar mycobacterial load in the lungs, there was very little dissemination in comparison to the other strains. The histopathology scores and levels of necrosis and calcification in tissue taken from guinea pigs infected with the Indian CAS strain was lowest out of all of the strains, supporting the conclusion that this may be the least virulent of the five strains that were evaluated. It is important to note that the two CAS strains seem to behave differently, indicating that even strains within the same *M. tuberculosis* family vary.

Besides the study by Palanisamy *et al*., 2009, there have been very few studies investigating the effects of clinical *M. tuberculosis* strains in an *in vivo* model. Lopez *et al*., 2003 investigated the pathogenesis of *M. tuberculosis* strains from four different genotype families (Beijing, Somali, Haarlam and Canetti) from different geographical locations using a mouse model (Lopez *et al.,* 2003). The laboratory control strain, *M. tuberculosis* H37Rv, had been included for comparison. There was a higher rate of mortality in mice that had been infected with the Beijing strain and a higher mycobacterial burden was observed in the lungs with the Beijing strain than the other strains that were evaluated. The higher bacillary load in the lungs was also

observed in the present study. This study concluded that after infecting mice with the Beijing strain, a non-protective immune response was propagated.

As was concluded after performing the *in vitro* experiments in the present study, results from the *in vivo* analysis of the *M. tuberculosis* strains belonging to the different phylogenetic lineages disproved the null hypothesis and clearly showed that there were differences in the way in which strains behaved during infection. Lopez *et al*., 2003, arrived at a similar conclusion, which stated that "genetically different *M. tuberculosis* strains evoked markedly different immunopathological event" (Lopez *et al.*, 2003).

CHAPTER 9

GENERAL DISCUSSION

9.1 SUMMARY OF PRESENT STUDY

The main aim of the present study was to construct a panel of MTBC strains that would be as representative of the global population of TB strains as possible and include a broad spectrum of *M. tuberculosis* strains against which any vaccine should be effective. This panel of strains would be useful for future vaccine evaluation studies and general TB research.

The first step towards creating this representative panel involved the identification of common and unique clinical, wild-type *M. tuberculosis* strains circulating within the city of London, UK, which consists of an ethnically diverse population and could be used as a model of global TB diversity. Molecular typing using a discriminative panel of MIRU, ETR and VNTR loci was performed on MTBC isolates from TB cases reported within London during a one year time period. Cluster analysis of all MIRU-VNTR profiles was carried out to identify common and unique strains. After excluding serial isolates (multiple isolates from the same patient) to avoid bias, cluster analysis of 2261 profiles yielded 235 clusters, with the largest cluster consisting of 53 isolates and the smallest cluster consisting of 2 isolates, and 1036 unique isolates.

In order to ensure there was sufficient global representation of the *M. tuberculosis* strains included in the present study epidemiological analysis was performed. All isolates that had been typed using MIRU-VNTR analysis were also subjected to spoligotyping, which helped to identify which MTBC family isolates belonged to. The spoligotyping data along with country of birth data for the patients was used for epidemiological analysis. The majority of patients were born in East Africa and the Indian Subcontinent, but there were patients that had also been born in other high burden regions including Western Africa, Eastern and South East Asia, South-Central Asia, Middle Africa and Southern Africa. Geographical associations were identified between some *M. tuberculosis* families and particular global regions, for example, patients born in the Indian subcontinent were infected mainly with *M. tuberculosis*

CAS and EAI families, whilst the majority of patients born in South-East Asia were infected with *M. tuberculosis* Beijing and EAI families.

The data from the present study was corroborated by similar associations identified in previous studies between the EAI and CAS families and the Middle Eastern and Central Asian region, and between the Beijing strain and the Far Eastern regions (Filliol *et al.,* 2002; Vitol *et al.,* 2006). Studies have also identified associations between *M. africanum* and countries such as Ghana and Nigeria within the Western Africa region as was identified in the present study (Vitol *et al.,* 2006; Gagneux *et al.,* 2007). In the present study *M. tuberculosis* Beijing strains were also isolated from patients born in Eastern Europe. Previous studies have identified the prevalence of the Beijing family in Eastern European countries like Russia and Estonia (Kruuner *et al.,* 2001; Toungoussova *et al.,* 2002; Drobniewski *et al.,* 2005; Gagneux *et al.,* 2007). The epidemiological analysis in the present study revealed that there were representatives from all MTBC families and almost all global regions.

As it would be impossible to perform phenotypic analyses on all isolates circulating within London, a preliminary panel of MTBC strains had to be selected, to include a broad spectrum of *M. tuberculosis* strains, which was as representative of the global TB strain population as possible, and against which any vaccine should prove effective. Cluster analysis helped to identify the most dominant strains circulating in London, and based on the hypothesis that infectious strains are more successful (e.g. higher clustering rates), an algorithm was designed to identify associations between clusters, MTBC family identification and the country of birth of patients, to select MTBC strains of interest for the preliminary panel (Panel A). This panel also included strains that did not fit in with the general patterns identified in the clusters reflecting more transient contact and therefore potentially higher strain infectivity. Panel A contained 42 strains, which were made up of 10 different MTBC families from patients born in 17 different countries.

It was predicted that investigating the phenotype of strains would allow the identification of differences between strains if the initial hypothesis was correct that strains were more common or dominant in a population because of some underlying advantage in comparison with other TB strains. An experimental approach was taken

to prove or disprove the null hypothesis that all TB strains are phenotypically the same.

A smaller panel of MTBC strains (Panel B) was selected from the preliminary panel of strains in Panel A by performing phylogenetic analysis on all MTBC isolates that had been included for molecular typing. Isolates were successfully assigned to one of six lineages (four *M. tuberculosis* and *M. bovis* lineages defined by Baker *et al*., 2004 and the *M. africanum* lineage defined by Gagneux *et al*., 2006) using MIRU-VNTR profiles and supplementary SNP analysis data. Strains for Panel B were selected from each of the six lineages. In summary, Panel B included a *M. tuberculosis* Beijing strain (from a patient born in Estonia), a LAM10 strain (from a patient born in the UK), two CAS strains (from patients born in Somalia and India), an EAI5 strain (from a patient born in India) and an *M. africanum* strain. For phenotypic experiments, also included were the laboratory control strains: *M. tuberculosis* H37Rv (supplied by HPA, MRU), *M. tuberculosis* H37Rv and *M. bovis* BCG (supplied by HPA, Porton Down), and *M. tuberculosis* H37Ra (HPA, NCTC).

The Panel B and laboratory control strains were included for *in vitro* phenotypic analyses in tissue culture models and cell-free mycobacterial culture systems. Firstly, PMA-activated macrophage-like THP-1 cells were infected with each of the strains after which the growth of intracellular mycobacteria and cytokine production profiles were investigated. The growth rates of the strains were investigated in mycobacterial cell-free culture systems (7ml MGIT Middlebrook 7H9 tubes).

The data from these *in vitro* experiments disproved the initial null hypothesis as the *M. tuberculosis* Beijing strain behaved differently to the other strains. The intracellular growth rate and the growth rate in the cell-free culture system of the Beijing strain was higher than the other strains, which meant this strain was potentially more virulent using the principles described previously (Theus *et al.,* 2005; Park *et al.,* 2006; Theus *et al.,* 2007). The increased production of the pro-inflammatory cytokines, TNF-α, compared to the other Panel B strains and the low levels of the anti-inflammatory cytokine, IL-10, produced after infecting THP-1 cells with the Beijing strain further demonstrated the virulent nature of the Beijing strain in comparison to the other strains as had been established in previous studies (Park *et al.,* 2006; Theus *et al.,* 2007).

The association between the intracellular growth of the *M. tuberculosis* Beijing strain and production of the cytokines, TNF-α and IL-10, has been investigated previously (Theus *et al.,* 2007). There was an inverse relationship between intracellular growth rate of the Beijing strain and production of TNF-α; therefore strains with an increased intracellular growth rate produced less TNF-α. Whereas an inverse relationship was identified between intracellular growth rate and TNF-α production by Theus *et al*., 2007, a study by Park *et al*., 2006, demonstrated that the more virulent strains both grew better and produced more TNF-α (Park *et al.,* 2006; Theus *et al.,* 2007).

In the present study, the highest multiplication rates were observed after infecting THP-1 cells with the Beijing strain and higher levels of TNF-α were produced by the THP-1 cells infected with the Beijing strain than THP-1 cells infected with the other four Panel B MTBC strains (two *M. tuberculosis* CAS strains, *M. tuberculosis* EAI5, *M. tuberculosis* LAM10 and *M. africanum*).

The panel of strains was subsequently narrowed down by further selecting strains of interest for *in vivo* analyses in order to test the original null hypothesis. Strains representing the four *M. tuberculosis* lineages out of the six major phylogenetic lineages were selected (*M. tuberculosis* LAM10, Beijing, EAI5 and the two CAS strains).

The *in vivo* experiments involved aerosol challenge of guinea pigs. During the aerosolisation of mycobacterial cultures it was decided to perform AGI-30 sampling, primarily to ensure that any observations after infecting guinea pigs with each of the strains were due to differences in virulence of the *M. tuberculosis* strain alone, rather than physical differences of the bacteria, but it was recognised that these data may offer insights into the stability of strains as aerosols. There were no significant differences between strains in the c.f.u./ml collected in the AGI-30 samplers relative to the initial c.f.u./ml in the cultures indicating that there was no significant difference in the stability of the tested strains in their aerosolised state.

In vivo analyses of the *M. tuberculosis* strains in the guinea pig model demonstrated that there were differences between the wild-type clinical *M. tuberculosis* strains as well as differences between the wild-type clinical strains and laboratory control strain,

M. tuberculosis H37Rv (e.g. weight data). The percentage weight change in guinea pigs infected with the Beijing strain was significantly lower than the other clinical strains after using the Mann-Whitney statistical test. The lung to body weight ratio at the time of necropsy was significantly higher for the Beijing strain than both of the CAS strains and the ratios observed after guinea pigs were infected with laboratory control strain was higher than all of the clinical strains.

The largest increase in lung and spleen mycobacterial load was observed in guinea pigs infected with the Beijing strain indicating that this strain had more potential to persist during infection and was more virulent than the other strains due to the increased dissemination of mycobacteria from the lung to the spleen during TB infection.

In conclusion, the data from the *in vitro* and *in vivo* analyses in the present study, disproved the null hypothesis and showed that there were phenotypic differences between the wild-type *M. tuberculosis* strains that were of clinical relevance and also differences between the wild-type and laboratory control strains. These conclusions indicated that there is a need to perform future vaccine evaluation studies not only using standardised laboratory challenge strains (such as H37Rv and Erdmann) but also with wild-type circulating strains from relevant global regions. This will be particularly relevant when a lead vaccine candidate is identified, and which will require more extensive evaluation before being introduced for mass vaccination and possibly before primate studies are instituted. Also, if evaluation studies showed that a particular vaccine was more effective against a *M. tuberculosis* strain isolated from a patient born in a high TB burden region it could lead to a successful niche vaccine.

9.2 FUTURE WORK

All of the strains in the preliminary panel (Panel A) need to undergo *in vitro* evaluation by calculating the growth rates of strains using the MGIT 960 culture system and inoculum sizes of 600 and 6,000 genomes, calculating the fold enhancement of strain growth after infecting PMA-activated THP-1 cells and measuring the production of human TNF-α, IL-10, IL-1β and IL-6 by infected THP-1 cells. In order to better monitor the production of each cytokine earlier, more

extensive time points need to be investigated, because the highest levels of the cytokine, TNF-α, were produced by day 1 post-infection.

The data from THP-1 cells for *in vitro* experiments needs to be corroborated using primary cell lines such as bone-marrow derived macrophages or primary macrophages from PBMCs. Other variables also need to be investigated. In the present study an MOI of 1 was used to infect the THP-1 cells with the mycobacterial strains. Different MOIs need to be evaluated. During further experiments the effect of infecting macrophage cell-lines could be monitored by observing the cells using inverted microscopy during the time course of infection.

The data from the *in vitro* experiments performed on the entire Panel A strains would help to narrow down this panel to include the most virulent MTBC strains. This would provide insight into potential correlations between the virulence of a particular strain and the dominance of the strain in the population. The dominance of a strain would be determined using the size of clusters to which the strains were initially assigned.

In the present study Panel B included a strain from each of the lineages defined by Baker *et al*., 2004 and Gagneux *et al*., 2006. However, there was only one strain per lineage, so to establish if there are true differences between the lineages, more strains need to be selected to represent each of the lineages and *in vitro* experiments need to be repeated to confirm the results and conclusions in the present study.

As well as testing the null hypothesis that there are no differences between the strains from the different lineages, the hypothesis needs to be tested on strains within the same lineage to confirm if strains within lineages are phenotypically the same. In addition, the phenotypic differences between the fully drug sensitive and MDR *M. tuberculosis* Beijing strains needs to be evaluated as the Estonian Beijing strain that has dominated the population in Estonia is an MDR strain, whilst the strain included in the present study was a fully drug sensitive equivalent strain with the same MIRU and ETR profiles.

In addition to evaluating laboratory control strains, *M. tuberculosis* H37Rv and H37Ra, and the vaccine strain, other laboratory control strains like the virulent

Erdmann strain needs to be included alongside the clinical wild-type strains to conclusively establish that the clinical strains are different to laboratory control and vaccine strains.

REFERENCES

Abou-Zeid, C., T. L. Ratliff, H. G. Wiker, M. Harboe, J. Bennedsen and G. A. Rook (1988). "Characterization of fibronectin-binding antigens released by *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG." Infect Immun **56**(12): 3046-51.

Alland, D., T. S. Whittam, M. B. Murray, M. D. Cave, M. H. Hazbon, K. Dix, M. Kokoris, A. Duesterhoeft, J. A. Eisen, C. M. Fraser and R. D. Fleischmann (2003). "Modeling bacterial evolution with comparative-genome-based marker systems: application to *Mycobacterium tuberculosis* evolution and pathogenesis." J Bacteriol **185**(11): 3392-9.

Allix, C., P. Supply and M. Fauville-Dufaux (2004). "Utility of fast mycobacterial interspersed repetitive unit-variable number tandem repeat genotyping in clinical mycobacteriological analysis." Clin Infect Dis **39**(6): 783-9.

Anderson, S. R., H. Maguire and J. Carless (2007). "Tuberculosis in London: a decade and a half of no decline [corrected]." Thorax **62**(2): 162-7.

Armitige, L. Y., C. Jagannath, A. R. Wanger and S. J. Norris (2000). "Disruption of the genes encoding antigen 85A and antigen 85B of *Mycobacterium tuberculosis* H37Rv: effect on growth in culture and in macrophages." Infect Immun **68**(2): 767-78.

Armstrong, J. A. and P. D. Hart (1975). "Phagosome-lysosome interactions in cultured macrophages infected with virulent tubercle bacilli. Reversal of the usual nonfusion pattern and observations on bacterial survival." J Exp Med **142**(1): 1-16.

Baker, L., T. Brown, M. C. Maiden and F. Drobniewski (2004). "Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*." Emerg Infect Dis **10**(9): 1568-77.

Balasubramanian, V., W. Guo-Zhi, E. Wiegeshaus and D. Smith (1992). "Virulence of *Mycobacterium tuberculosis* for guinea pigs: a quantitative modification of the assay developed by Mitchison." Tuber Lung Dis **73**(5): 268-72.

Balcewicz-Sablinska, M. K., J. Keane, H. Kornfeld and H. G. Remold (1998). "Pathogenic *Mycobacterium tuberculosis* evades apoptosis of host macrophages by release of TNF-R2, resulting in inactivation of TNF-alpha." J Immunol **161**(5): 2636-41.

Barlow, R. E., D. M. Gascoyne-Binzi, S. H. Gillespie, A. Dickens, S. Qamer and P. M. Hawkey (2001). "Comparison of variable number tandem repeat and IS*6110*-restriction fragment length polymorphism analyses for discrimination of high- and low-copy-number IS*6110 Mycobacterium tuberculosis* isolates." J Clin Microbiol **39**(7): 2453-7.

Bates, J. H. and J. K. Fitzhugh (1967). "Subdivision of the species *M. tuberculosis* by mycobacteriophage typing." Am Rev Resp Dis **96**: 7-10.

Bermudez, L. E. and J. Goodman (1996). "*Mycobacterium tuberculosis* invades and replicates within type II alveolar cells." Infect Immun **64**(4): 1400-6.

Bhatia, A. L., A. Csillag, D. A. Mitchison, J. B. Selkon, P. R. Somasundaram and T. V. Subbaiah (1961). "The virulence in the guinea-pig of tubercle bacilli isolated before treatment from South Indian patients with pulmonary tuberculosis. 2. Comparison with virulence of tubercle bacilli from British patients." Bull World Health Organ **25**: 313-22.

Bhatt, K. and P. Salgame (2007). "Host innate immune response to *Mycobacterium tuberculosis*." J Clin Immunol **27**(4): 347-62.

Bhattacharya, M., S. Dietrich, L. Mosher, F. Siddiqui, B. E. Reisberg, W. S. Paul and J. R. Warren (1998). "Cross-contamination of specimens with *Mycobacterium tuberculosis*: clinical significance, causes, and prevention." Am J Clin Pathol **109**(3): 324-30.

Bifani, P. J., B. B. Plikaytis, V. Kapur, K. Stockbauer, X. Pan, M. L. Lutfey, S. L. Moghazeh, W. Eisner, T. M. Daniel, M. H. Kaplan, J. T. Crawford, J. M. Musser and B. N. Kreiswirth (1996). "Origin and interstate spread of a New York City multidrug-resistant *Mycobacterium tuberculosis* clone family." Jama **275**(6): 452-7.

Boesen, H., B. N. Jensen, T. Wilcke and P. Andersen (1995). "Human T-cell responses to secreted antigen fractions of *Mycobacterium tuberculosis*." Infect Immun **63**(4): 1491-7.

Bogdan, C., Y. Vodovotz and C. Nathan (1991). "Macrophage deactivation by interleukin 10." J Exp Med **174**(6): 1549-55.

Braden, C. R., J. T. Crawford and B. A. Schable (2002). "Quality assessment of *Mycobacterium tuberculosis* genotyping in a large laboratory network." Emerg Infect Dis **8**(11): 1210-5.

Brewer, T. F. (2000). "Preventing tuberculosis with bacillus Calmette-Guerin vaccine: a meta-analysis of the literature." Clin Infect Dis **31 Suppl 3**: S64-7.

Britton, W. J. and U. Palendira (2003). "Improving vaccines against tuberculosis." Immunol Cell Biol **81**(1): 34-45.

Brosch, R., S. V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeier, T. Garnier, C. Gutierrez, G. Hewinson, K. Kremer, L. M. Parsons, A. S. Pym, S. Samper, D. van Soolingen and S. T. Cole (2002). "A new evolutionary scenario for the *Mycobacterium tuberculosis* complex." Proc Natl Acad Sci U S A **99**(6): 3684-9.

Brosch, R., A. S. Pym, S. V. Gordon and S. T. Cole (2001). "The evolution of mycobacterial pathogenicity: clues from comparative genomics." Trends Microbiol **9**(9): 452-8.

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

Campos, J. M., J. P. Simonetti, M. V. Pone, L. A. Carvalho, A. C. Pereira and J. R. Garrido (1996). "Disseminated Bacillus Calmette-Guerin infection in HIV-infected children: case report and review." Pediatr AIDS HIV Infect **7**(6): 429-32.

Castanon-Arreola, M., Y. Lopez-Vidal, C. Espitia-Pinzon and R. Hernandez-Pando (2005). "A new vaccine against tuberculosis shows greater protection in a mouse model with progressive pulmonary tuberculosis." Tuberculosis (Edinb) **85**(1-2): 115-26.

Chan, J., T. Fujiwara, P. Brennan, M. McNeil, S. J. Turco, J. C. Sibille, M. Snapper, P. Aisen and B. R. Bloom (1989). "Microbial glycolipids: possible virulence factors that scavenge oxygen radicals." Proc Natl Acad Sci U S A **86**(7): 2453-7.

Chini, V., A. Foka, G. Dimitracopoulos and I. Spiliopoulou (2007). "Absolute and relative real-time PCR in the quantification of *tst* gene expression among methicillin-resistant *Staphylococcus aureus*: evaluation by two mathematical models." Lett Appl Microbiol **45**(5): 479-84.

Cole, S. T. and B. G. Barrell (1998). "Analysis of the genome of *Mycobacterium tuberculosis* H37Rv." Novartis Found Symp **217**: 160-72; discussion 172-7.

Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, 3rd, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead and B. G. Barrell (1998). "Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence." Nature **393**(6685): 537-44.

Coler, R. N., A. Campos-Neto, P. Ovendale, F. H. Day, S. P. Fling, L. Zhu, N. Serbina, J. L. Flynn, S. G. Reed and M. R. Alderson (2001). "Vaccination with the T cell antigen Mtb 8.4 protects against challenge with *Mycobacterium tuberculosis*." J Immunol **166**(10): 6227-35.

Collins, H. L. and S. H. Kaufmann (2001). "Prospects for better tuberculosis vaccines." Lancet Infect Dis **1**(1): 21-8.

Converse, P. J., A. M. Dannenberg, Jr., J. E. Estep, K. Sugisaki, Y. Abe, B. H. Schofield and M. L. Pitt (1996). "Cavitary tuberculosis produced in rabbits by aerosolized virulent tubercle bacilli." Infect Immun **64**(11): 4776-87.

Cowan, L. S., L. Mosher, L. Diem, J. P. Massey and J. T. Crawford (2002). "Variable-number tandem repeat typing of *Mycobacterium tuberculosis* isolates with low copy numbers of IS*6110* by using mycobacterial interspersed repetitive units." J Clin Microbiol **40**(5): 1592-602.

Crofton, J., N. W. Horne and F. J. W. Miller (1992). The battle between the tubercle bacilli and the patient. Clinical tuberculosis. London and Oxford, Macmillan Education Ltd**:** 7-15.

Dale, J. W., G. H. Bothamley, F. Drobniewski, S. H. Gillespie, T. D. McHugh and R. Pitman (2005). "Origins and properties of *Mycobacterium tuberculosis* isolates in London." J Med Microbiol **54**(Pt 6): 575-82.

Dale, J. W., D. Brittain, A. A. Cataldi, D. Cousins, J. T. Crawford, J. Driscoll, H. Heersma, T. Lillebaek, T. Quitugua, N. Rastogi, R. A. Skuce, C. Sola, D. Van Soolingen and V. Vincent (2001). "Spacer oligonucleotide typing of bacteria of the *Mycobacterium tuberculosis* complex: recommendations for standardised nomenclature." Int J Tuberc Lung Dis **5**(3): 216-9.

de Boer, A. S., M. W. Borgdorff, P. E. de Haas, N. J. Nagelkerke, J. D. van Embden and D. van Soolingen (1999). "Analysis of rate of change of IS*6110* RFLP patterns of *Mycobacterium tuberculosis* based on serial patient isolates." J Infect Dis **180**(4): 1238-44.

de, C. R. M., H. Soini, G. C. Roscanni, M. Jaques, M. C. Villares and J. M. Musser (1999). "Extensive cross-contamination of specimens with *Mycobacterium tuberculosis* in a reference laboratory." J Clin Microbiol **37**(4): 916-9.

Delogu, G., A. Li, C. Repique, F. Collins and S. L. Morris (2002). "DNA vaccine combinations expressing either tissue plasminogen activator signal sequence fusion proteins or ubiquitin-conjugated antigens induce sustained protective immunity in a mouse model of pulmonary tuberculosis." Infect Immun **70**(1): 292-302.

Derrick, S. C., A. L. Yang and S. L. Morris (2004). "A polyvalent DNA vaccine expressing an ESAT6-Ag85B fusion protein protects mice against a primary infection with *Mycobacterium tuberculosis* and boosts BCG-induced protective immunity." Vaccine **23**(6): 780-8.

Desjardin, L. E., Y. Chen, M. D. Perkins, L. Teixeira, M. D. Cave and K. D. Eisenach (1998). "Comparison of the ABI 7700 system (TaqMan) and competitive PCR for quantification of IS*6110* DNA in sputum during treatment of tuberculosis." J Clin Microbiol **36**(7): 1964-8.

Devaux, I., K. Kremer, H. Heersma and D. Van Soolingen (2009). "Clusters of multidrug-resistant *Mycobacterium tuberculosis* cases, Europe." Emerg Infect Dis **15**(7): 1052-60.

Dietrich, J., C. Aagaard, R. Leah, A. W. Olsen, A. Stryhn, T. M. Doherty and P. Andersen (2005). "Exchanging ESAT6 with TB10.4 in an Ag85B fusion molecule-based tuberculosis subunit vaccine: efficient protection and ESAT6-based sensitive monitoring of vaccine efficacy." J Immunol **174**(10): 6332-9.

Dietrich, J., C. V. Lundberg and P. Andersen (2006). "TB vaccine strategies--what is needed to solve a complex problem?" Tuberculosis (Edinb) **86**(3-4): 163-8.

DofH (2006). Tuberculosis. Immunisation against infectious diseases**:** 391-408.

Doherty, T. M. and P. Andersen (2005). "Vaccines for tuberculosis: novel concepts and recent progress." Clin Microbiol Rev **18**(4): 687-702.

Drobniewski, F., Y. Balabanova, V. Nikolayevsky, M. Ruddy, S. Kuznetzov, S. Zakharova, A. Melentyev and I. Fedorin (2005). "Drug-resistant tuberculosis, clinical virulence, and the dominance of the Beijing strain family in Russia." JAMA **293**(22): 2726-31.

Drobniewski, F., Y. Balabanova, M. Ruddy, L. Weldon, K. Jeltkova, T. Brown, N. Malomanova, E. Elizarova, A. Melentyey, E. Mutovkin, S. Zhakharova and I. Fedorin (2002). "Rifampin- and multidrug-resistant tuberculosis in Russian civilians and prison inmates: dominance of the beijing strain family." Emerg Infect Dis **8**(11): 1320-6.

Drobniewski, F. A., A. Gibson, M. Ruddy and M. D. Yates (2003). "Evaluation and utilization as a public health tool of a national molecular epidemiological tuberculosis outbreak database within the United Kingdom from 1997 to 2001." J Clin Microbiol **41**(5): 1861-8.

Drobniewski, F. A. and N. Q. Verlander (2000). "Tuberculosis and the role of war in the modern era." Int J Tuberc Lung Dis **4**(12): 1120-5.

Dunn, P. L. and R. J. North (1995). "Virulence ranking of some *Mycobacterium tuberculosis* and *Mycobacterium bovis* strains according to their ability to multiply in the lungs, induce lung pathology, and cause mortality in mice." Infect Immun **63**(9): 3428-37.

Dye, C. (2006). "Global epidemiology of tuberculosis." Lancet **367**(9514): 938-40.

Fenton, M. J. and M. W. Vermeulen (1996). "Immunopathology of tuberculosis: roles of macrophages and monocytes." Infect Immun **64**(3): 683-90.

Fenton, M. J., M. W. Vermeulen, S. Kim, M. Burdick, R. M. Strieter and H. Kornfeld (1997). "Induction of gamma interferon production in human alveolar macrophages by *Mycobacterium tuberculosis*." Infect Immun **65**(12): 5149-56.

Ferdinand, S., G. Valetudie, C. Sola and N. Rastogi (2004). "Data mining of *Mycobacterium tuberculosis* complex genotyping results using mycobacterial interspersed repetitive units validates the clonal structure of spoligotyping-defined families." Res Microbiol **155**(8): 647-54.

Ferguson, J. S., D. R. Voelker, F. X. McCormack and L. S. Schlesinger (1999). "Surfactant protein D binds to *Mycobacterium tuberculosis* bacilli and lipoarabinomannan via carbohydrate-lectin interactions resulting in reduced phagocytosis of the bacteria by macrophages." J Immunol **163**(1): 312-21.

Filliol, I., J. R. Driscoll, D. Van Soolingen, B. N. Kreiswirth, K. Kremer, G. Valetudie, D. D. Anh, R. Barlow, D. Banerjee, P. J. Bifani, K. Brudey, A. Cataldi, R. C. Cooksey, D. V. Cousins, J. W. Dale, O. A. Dellagostin, F. Drobniewski, G. Engelmann, S. Ferdinand, D. Gascoyne-Binzi, M. Gordon, M. C. Gutierrez, W. H. Haas, H. Heersma, G. Kallenius, E. Kassa-Kelembho, T. Koivula, H. M. Ly, A. Makristathis, C. Mammina, G. Martin, P. Mostrom, I. Mokrousov, V. Narbonne, O. Narvskaya, A. Nastasi, S. N. Niobe-Eyangoh, J. W. Pape, V. Rasolofo-Razanamparany, M. Ridell, M. L. Rossetti, F. Stauffer, P. N. Suffys, H. Takiff, J. Texier-Maugein, V. Vincent, J. H. De Waard, C. Sola and N. Rastogi (2002). "Global distribution of *Mycobacterium tuberculosis* spoligotypes." Emerg Infect Dis **8**(11): 1347-9.

Fine, P. E. (1995). "Variation in protection by BCG: implications of and for heterologous immunity." Lancet **346**(8986): 1339-45.

Fine, P. E. and E. Vynnycky (1998). "The effect of heterologous immunity upon the apparent efficacy of (e.g. BCG) vaccines." Vaccine **16**(20): 1923-8.

Fleischmann, R. D., D. Alland, J. A. Eisen, L. Carpenter, O. White, J. Peterson, R. DeBoy, R. Dodson, M. Gwinn, D. Haft, E. Hickey, J. F. Kolonay, W. C. Nelson, L. A. Umayam, M. Ermolaeva, S. L. Salzberg, A. Delcher, T. Utterback, J. Weidman, H. Khouri, J. Gill, A. Mikula, W. Bishai, W. R. Jacobs Jr, Jr., J. C. Venter and C. M. Fraser (2002). "Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains." J Bacteriol **184**(19): 5479-90.

Flynn, J. L. (2006). "Lessons from experimental *Mycobacterium tuberculosis* infections." Microbes Infect **8**(4): 1179-88.

Flynn, J. L., M. M. Goldstein, K. J. Triebold, J. Sypek, S. Wolf and B. R. Bloom (1995). "IL-12 increases resistance of BALB/c mice to *Mycobacterium tuberculosis* infection." J Immunol **155**(5): 2515-24.

Frehel, C., C. de Chastellier, T. Lang and N. Rastogi (1986). "Evidence for inhibition of fusion of lysosomal and prelysosomal compartments with phagosomes in macrophages infected with pathogenic *Mycobacterium avium*." Infect Immun **52**(1): 252-62.

Fremond, C. M., V. Yeremeev, D. M. Nicolle, M. Jacobs, V. F. Quesniaux and B. Ryffel (2004). "Fatal *Mycobacterium tuberculosis* infection despite adaptive immune response in the absence of MyD88." J Clin Invest **114**(12): 1790-9.

Frothingham, R., H. G. Hills and K. H. Wilson (1994). "Extensive DNA sequence conservation throughout the *Mycobacterium tuberculosis* complex." J Clin Microbiol **32**(7): 1639-43.

Frothingham, R. and W. A. Meeker-O'Connell (1998). "Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats." Microbiology **144 (Pt 5)**: 1189-96.

Gagneux, S., M. V. Burgos, K. DeRiemer, A. Encisco, S. Munoz, P. C. Hopewell, P. M. Small and A. S. Pym (2006). "Impact of bacterial genetics on the transmission of isoniazid-resistant *Mycobacterium tuberculosis*." PLoS Pathog **2**(6): e61.

Gagneux, S., K. DeRiemer, T. Van, M. Kato-Maeda, B. C. de Jong, S. Narayanan, M. Nicol, S. Niemann, K. Kremer, M. C. Gutierrez, M. Hilty, P. C. Hopewell and P. M. Small (2006). "Variable host-pathogen compatibility in *Mycobacterium tuberculosis*." Proc Natl Acad Sci U S A **103**(8): 2869-73.

Gagneux, S. and P. M. Small (2007). "Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development." Lancet Infect Dis **7**(5): 328-37.

Garcia de Viedma, D., M. del Sol Diaz Infantes, F. Lasala, F. Chaves, L. Alcala and E. Bouza (2002). "New real-time PCR able to detect in a single tube multiple rifampin resistance mutations and high-level isoniazid resistance mutations in *Mycobacterium tuberculosis*." J Clin Microbiol **40**(3): 988-95.

Garcia Pelayo, M. C., S. Uplekar, A. Keniry, P. Mendoza Lopez, T. Garnier, J. Nunez Garcia, L. Boschiroli, X. Zhou, J. Parkhill, N. Smith, R. G. Hewinson, S. T. Cole and S. V. Gordon (2009). "A comprehensive survey of single nucleotide polymorphisms (SNPs) across *Mycobacterium bovis* strains and *M. bovis* BCG vaccine strains refines the genealogy and defines a minimal set of SNPs that separate virulent *M. bovis* strains and *M. bovis* BCG strains." Infect Immun **77**(5): 2230-8.

Gaynor, C. D., F. X. McCormack, D. R. Voelker, S. E. McGowan and L. S. Schlesinger (1995). "Pulmonary surfactant protein A mediates enhanced phagocytosis of *Mycobacterium tuberculosis* by a direct interaction with human macrophages." J Immunol **155**(11): 5343-51.

Geijtenbeek, T. B., S. J. Van Vliet, E. A. Koppel, M. Sanchez-Hernandez, C. M. Vandenbroucke-Grauls, B. Appelmelk and Y. Van Kooyk (2003). "Mycobacteria target DC-SIGN to suppress dendritic cell function." J Exp Med **197**(1): 7-17.

Gibson, A., T. Brown, L. Baker and F. Drobniewski (2005). "Can 15-locus mycobacterial interspersed repetitive unit-variable-number tandem repeat analysis provide insight into the evolution of *Mycobacterium tuberculosis*?" Appl Environ Microbiol **71**(12): 8207-13.

Goguet de la Salmoniere, Y. O., H. M. Li, G. Torrea, A. Bunschoten, J. van Embden and B. Gicquel (1997). "Evaluation of spoligotyping in a study of the transmission of *Mycobacterium tuberculosis*." J Clin Microbiol **35**(9): 2210-4.

Gopaul, K. K., T. J. Brown, A. L. Gibson, M. D. Yates and F. A. Drobniewski (2006). "Progression toward an improved DNA amplification-based typing technique in the study of *Mycobacterium tuberculosis* epidemiology." J Clin Microbiol **44**(7): 2492-8.

Grode, L., P. Seiler, S. Baumann, J. Hess, V. Brinkmann, A. Nasser Eddine, P. Mann, C. Goosmann, S. Bandermann, D. Smith, G. J. Bancroft, J. M. Reyrat, D. van Soolingen, B. Raupach and S. H. Kaufmann (2005). "Increased vaccine efficacy against tuberculosis of recombinant *Mycobacterium bovis* bacille Calmette-Guerin mutants that secrete listeriolysin." J Clin Invest **115**(9): 2472-9.

Groenen, P. M., A. E. Bunschoten, D. van Soolingen and J. D. van Embden (1993). "Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method." Mol Microbiol **10**(5): 1057-65.

Gupta, U. D., V. M. Katoch and D. N. McMurray (2007). "Current status of TB vaccines." Vaccine **25**(19): 3742-51.

Gutacker, M. M., B. Mathema, H. Soini, E. Shashkina, B. N. Kreiswirth, E. A. Graviss and J. M. Musser (2006). "Single-nucleotide polymorphism-based population genetic analysis of *Mycobacterium tuberculosis* strains from 4 geographic sites." J Infect Dis **193**(1): 121-8.

Gutacker, M. M., J. C. Smoot, C. A. Migliaccio, S. M. Ricklefs, S. Hua, D. V. Cousins, E. A. Graviss, E. Shashkina, B. N. Kreiswirth and J. M. Musser (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**(4): 1533-43.

Gutierrez, M. C., N. Ahmed, E. Willery, S. Narayanan, S. E. Hasnain, D. S. Chauhan, V. M. Katoch, V. Vincent, C. Locht and P. Supply (2006). "Predominance of ancestral lineages of *Mycobacterium tuberculosis* in India." Emerg Infect Dis **12**(9): 1367-74.

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

Haagsman, H. P. (1994). "Surfactant proteins A and D." Biochem Soc Trans **22**(1): 100-6.

Hawgood, S. and J. A. Clements (1990). "Pulmonary surfactant and its apoproteins." J Clin Invest **86**(1): 1-6.

Hawkey, P. M., E. G. Smith, J. T. Evans, P. Monk, G. Bryan, H. H. Mohamed, M. Bardhan and R. N. Pugh (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**(8): 3514-20.

Hermans, P. W., D. van Soolingen, E. M. Bik, P. E. de Haas, J. W. Dale and J. D. van Embden (1991). "Insertion element IS*987* from *Mycobacterium bovis* BCG is located in a hot-spot integration region for insertion elements in *Mycobacterium tuberculosis* complex strains." Infect Immun **59**(8): 2695-705.

Hermans, P. W., D. van Soolingen and J. D. van Embden (1992). "Characterization of a major polymorphic tandem repeat in *Mycobacterium tuberculosis* and its potential use in the epidemiology of *Mycobacterium kansasii* and *Mycobacterium gordonae*." J Bacteriol **174**(12): 4157-65.

Hernandez-Pando, R., L. Pavon, K. Arriaga, H. Orozco, V. Madrid-Marina and G. Rook (1997). "Pathogenesis of tuberculosis in mice exposed to low and high doses of an environmental mycobacterial saprophyte before infection." Infect Immun **65**(8): 3317-27.

Horwitz, M. A., G. Harth, B. J. Dillon and S. Maslesa-Galic (2000). "Recombinant bacillus calmette-guerin (BCG) vaccines expressing the *Mycobacterium tuberculosis* 30-kDa major secretory protein induce greater protective immunity against tuberculosis than conventional BCG vaccines in a highly susceptible animal model." Proc Natl Acad Sci U S A **97**(25): 13853-8.

Hovav, A. H., J. Mullerad, L. Davidovitch, Y. Fishman, F. Bigi, A. Cataldi and H. Bercovier (2003). "The *Mycobacterium tuberculosis* recombinant 27-kilodalton lipoprotein induces a strong Th1-type immune response deleterious to protection." Infect Immun **71**(6): 3146-54.

HPA (2006). "Migrant Health: Infectious diseases in non-UK born populations in England, Wales and Northern Ireland. A baseline report – 2006." London: Health Protection Agency Centre for Infections. 2006.

HPA (2007). "Enhanced Tuberculosis Surveillance, Enhanced Surveillance of Mycobacterial Infections." Health Protection Agency, Centre for Infections **2007**.

HPA (2008). "Tuberculosis in the UK: Annual report on tuberculosis surveillance in the UK 2008." London: Health Protection Agency Centre for Infections, October 2008.

HPA (2008). "The UK Mycobacterial surveillance network report 1994 - 2003; 10 years of MycobNet." Tuberculosis Section Health Protection Agency Centre for Infections **London: Health Protection Agency Centre for Infections, October 2008**.

Hsu, T., S. M. Hingley-Wilson, B. Chen, M. Chen, A. Z. Dai, P. M. Morin, C. B. Marks, J. Padiyar, C. Goulding, M. Gingery, D. Eisenberg, R. G. Russell, S. C. Derrick, F. M. Collins, S. L. Morris, C. H. King and W. R. Jacobs, 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." Proc Natl Acad Sci U S A **100**(21): 12420-5.

Huang, H. Y., R. Jou, C. Y. Chiang, W. C. Liu, H. J. Chiu and J. J. Lee (2007). "Nosocomial transmission of tuberculosis in two hospitals for mentally handicapped patients." J Formos Med Assoc **106**(12): 999-1006.

Hubbard, R. D., C. M. Flory and F. M. Collins (1992). "Immunization of mice with mycobacterial culture filtrate proteins." Clin Exp Immunol **87**(1): 94-8.

Hunter, P. R. and M. A. Gaston (1988). "Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity." J Clin Microbiol **26**(11): 2465-6.

Hunter, S. W., H. Gaylord and P. J. Brennan (1986). "Structure and antigenicity of the phosphorylated lipopolysaccharide antigens from the leprosy and tubercle bacilli." J Biol Chem **261**(26): 12345-51.

Imaeda, T. (1985). "Deoxyribonucleic Acid Relatedness Among Selected Strains of *Mycobacterium tuberculosis*, *Mycobacterium bovis*, *Mycobacterium bovis* BCG, *Mycobacterium microti*, and *Mycobacterium africanum*." Int J System Bacteriol **35**(2): 147-150.

Ishige, I., Y. Usui, T. Takemura and Y. Eishi (1999). "Quantitative PCR of mycobacterial and propionibacterial DNA in lymph nodes of Japanese patients with sarcoidosis." Lancet **354**(9173): 120-3.

Iwamoto, T., S. Yoshida, K. Suzuki, M. Tomita, R. Fujiyama, N. Tanaka, Y. Kawakami and M. Ito (2007). "Hypervariable loci that enhance the discriminatory ability of newly proposed 15-loci and 24-loci variable-number tandem repeat typing method on *Mycobacterium tuberculosis* strains predominated by the Beijing family." FEMS Microbiol Lett **270**(1): 67-74.

Jackson, M., S. W. Phalen, M. Lagranderie, D. Ensergueix, P. Chavarot, G. Marchal, D. N. McMurray, B. Gicquel and C. Guilhot (1999). "Persistence and protective efficacy of a *Mycobacterium tuberculosis* auxotroph vaccine." Infect Immun **67**(6): 2867-73.

Janeway, C. A., P. Travers, M. Walport and M. Shlomchik (2005). Antigen Presentation to T Lymphocytes. Immunobiology : the immune system in health and disease. E. Lawrence, Garland Science Publishing**:** 169-201.

Kam, K. M., C. W. Yip, L. W. Tse, K. L. Leung, K. L. Wong, W. M. Ko and W. S. Wong (2006). "Optimization of variable number tandem repeat typing set for differentiating *Mycobacterium tuberculosis* strains in the Beijing family." FEMS Microbiol Lett **256**(2): 258-65.

Kamath, A. T., C. G. Feng, M. Macdonald, H. Briscoe and W. J. Britton (1999). "Differential protective efficacy of DNA vaccines expressing secreted proteins of *Mycobacterium tuberculosis*." Infect Immun **67**(4): 1702-7.

Kamerbeek, J., L. Schouls, A. Kolk, M. van Agterveld, D. van Soolingen, S. Kuijper, A. Bunschoten, H. Molhuizen, R. Shaw, M. Goyal and J. van Embden (1997). "Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology." J Clin Microbiol **35**(4): 907-14.

Keane, J., M. K. Balcewicz-Sablinska, H. G. Remold, G. L. Chupp, B. B. Meek, M. J. Fenton and H. Kornfeld (1997). "Infection by *Mycobacterium tuberculosis* promotes human alveolar macrophage apoptosis." Infect Immun **65**(1): 298-304.

Keane, J., H. G. Remold and H. Kornfeld (2000). "Virulent *Mycobacterium tuberculosis* strains evade apoptosis of infected alveolar macrophages." J Immunol **164**(4): 2016-20.

Kik, S. V., S. Verver, D. van Soolingen, P. E. de Haas, F. G. Cobelens, K. Kremer, H. van Deutekom and M. W. Borgdorff (2008). "Tuberculosis outbreaks predicted by characteristics of first patients in a DNA fingerprint cluster." Am J Respir Crit Care Med **178**(1): 96-104.

Kim, B. J., S. K. Hong, K. H. Lee, Y. J. Yun, E. C. Kim, Y. G. Park, G. H. Bai and Y. H. Kook (2004). "Differential identification of *Mycobacterium tuberculosis* complex and nontuberculous mycobacteria by duplex PCR assay using the RNA polymerase gene (*rpoB*)." J Clin Microbiol **42**(3): 1308-12.

Kirschner, P., B. Springer, U. Vogel, A. Meier, A. Wrede, M. Kiekenbeck, F. C. Bange and E. C. Bottger (1993). "Genotypic identification of mycobacteria by nucleic acid sequence determination: report of a 2-year experience in a clinical laboratory." J Clin Microbiol **31**(11): 2882-9.

Kremer, K., C. Arnold, A. Cataldi, M. C. Gutierrez, W. H. Haas, S. Panaiotov, R. A. Skuce, P. Supply, A. G. van der Zanden and D. van Soolingen (2005). "Discriminatory power and reproducibility of novel DNA typing methods for *Mycobacterium tuberculosis* complex strains." J Clin Microbiol **43**(11): 5628-38.

Kremer, K., A. van den Brandt, N. E. Kurepina, J. Glynn, P. J. Bifani and D. van Soolingen (2002). "Definition of the *Mycobacterium tuberculosis* Beijing genotype." European Union Concerted Action meeting(New genetic markers and techniques for the epidemiology and control of tuberculosis): Cascais, Portugal.

Kremer, K., D. van Soolingen, R. Frothingham, W. H. Haas, P. W. Hermans, C. Martin, P. Palittapongarnpim, B. B. Plikaytis, L. W. Riley, M. A. Yakrus, J. M. Musser and J. D. van Embden (1999). "Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility." J Clin Microbiol **37**(8): 2607-18.

Kruuner, A., S. E. Hoffner, H. Sillastu, M. Danilovits, K. Levina, S. B. Svenson, S. Ghebremichael, T. Koivula and G. Kallenius (2001). "Spread of drug-resistant pulmonary tuberculosis in Estonia." J Clin Microbiol **39**(9): 3339-45.

Kurepina, N. E., S. Sreevatsan, B. B. Plikaytis, P. J. Bifani, N. D. Connell, R. J. Donnelly, D. van Sooligen, J. M. Musser and B. N. Kreiswirth (1998). "Characterization of the phylogenetic distribution and chromosomal insertion sites of five IS*6110* elements in *Mycobacterium tuberculosis*: non-random integration in the *dnaA-dnaN* region." Tuber Lung Dis **79**(1): 31-42.

Kwara, A., R. Schiro, L. S. Cowan, N. E. Hyslop, M. F. Wiser, S. Roahen Harrison, P. Kissinger, L. Diem and J. T. Crawford (2003). "Evaluation of the epidemiologic utility of secondary typing methods for differentiation of *Mycobacterium tuberculosis* isolates." J Clin Microbiol **41**(6): 2683-5.

Langermans, J. A., T. M. Doherty, R. A. Vervenne, T. van der Laan, K. Lyashchenko, R. Greenwald, E. M. Agger, C. Aagaard, H. Weiler, D. van Soolingen, W. Dalemans, A. W. Thomas and P. Andersen (2005). "Protection of macaques against *Mycobacterium tuberculosis* infection by a subunit vaccine based on a fusion protein of antigen 85B and ESAT-6." Vaccine **23**(21): 2740-50.

Lee, A. S., L. L. Tang, I. H. Lim, R. Bellamy and S. Y. Wong (2002). "Discrimination of single-copy IS*6110* DNA fingerprints of *Mycobacterium tuberculosis* isolates by high-resolution minisatellite-based typing." J Clin Microbiol **40**(2): 657-9.

Lee, K. S., V. S. Dubey, P. E. Kolattukudy, C. H. Song, A. R. Shin, S. B. Jung, C. S. Yang, S. Y. Kim, E. K. Jo, J. K. Park and H. J. Kim (2007). "Diacyltrehalose of *Mycobacterium tuberculosis* inhibits lipopolysaccharide- and mycobacteria-induced proinflammatory cytokine production in human monocytic cells." FEMS Microbiol Lett **267**(1): 121-8.

Legrand, E., I. Filliol, C. Sola and N. Rastogi (2001). "Use of spoligotyping to study the evolution of the direct repeat locus by IS*6110* transposition in *Mycobacterium tuberculosis*." J Clin Microbiol **39**(4): 1595-9.

Lever, M. S., A. Williams and A. M. Bennett (2000). "Survival of mycobacterial species in aerosols generated from artificial saliva." Lett Appl Microbiol **31**(3): 238-41.

Li, W. M., S. M. Wang, C. Y. Li, Y. H. Liu, G. M. Shen, X. X. Zhang, T. G. Niu, Q. Gao, D. van Soolingen, K. Kremer and H. J. Duanmu (2005). "Molecular epidemiology of *Mycobacterium tuberculosis* in China: a nationwide random survey in 2000." Int J Tuberc Lung Dis **9**(12): 1314-9.

Lienhardt, C., A. Azzurri, A. Amedei, K. Fielding, J. Sillah, O. Y. Sow, B. Bah, M. Benagiano, A. Diallo, R. Manetti, K. Manneh, P. Gustafson, S. Bennett, M. M. D'Elios, K. McAdam and G. D. Prete (2002). "Active tuberculosis in Africa is associated with reduced Th1 and increased Th2 activity *in vivo*." Eur. J. Immunol **32**(6): 1605-1613.

Lillebaek, T., A. Dirksen, I. Baess, B. Strunge, V. O. Thomsen and A. B. Andersen (2002). "Molecular evidence of endogenous reactivation of *Mycobacterium tuberculosis* after 33 years of latent infection." J Infect Dis **185**(3): 401-4.

Lopez, B., D. Aguilar, H. Orozco, M. Burger, C. Espitia, V. Ritacco, L. Barrera, K. Kremer, R. Hernandez-Pando, K. Huygen and D. van Soolingen (2003). "A marked difference in pathogenesis and immune response induced by different *Mycobacterium tuberculosis* genotypes." Clin Exp Immunol **133**(1): 30-7.

Magdalena, J., A. Vachee, P. Supply and C. Locht (1998). "Identification of a new DNA region specific for members of *Mycobacterium tuberculosis* complex." J Clin Microbiol **36**(4): 937-43.

Maguire, H., J. W. Dale, T. D. McHugh, P. D. Butcher, S. H. Gillespie, A. Costetsos, H. Al-Ghusein, R. Holland, A. Dickens, L. Marston, P. Wilson, R. Pitman, D. Strachan, F. A. Drobniewski and D. K. Banerjee (2002). "Molecular epidemiology of tuberculosis in London 1995-7 showing low rate of active transmission." Thorax **57**(7): 617-22.

Manabe, Y. C., C. P. Scott and W. R. Bishai (2002). "Naturally attenuated, orally administered *Mycobacterium microti* as a tuberculosis vaccine is better than subcutaneous *Mycobacterium bovis* BCG." Infect Immun **70**(3): 1566-70.

Marques, L. J., L. Zheng, N. Poulakis, J. Guzman and U. Costabel (1999). "Pentoxifylline inhibits TNF-alpha production from human alveolar macrophages." Am J Respir Crit Care Med **159**(2): 508-11.

Mathema, B., N. E. Kurepina, P. J. Bifani and B. N. Kreiswirth (2006). "Molecular epidemiology of tuberculosis: current insights." Clin Microbiol Rev **19**(4): 658-85.

Mazars, E., S. Lesjean, A. L. Banuls, M. Gilbert, V. Vincent, B. Gicquel, M. Tibayrenc, C. Locht and P. Supply (2001). "High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology." Proc Natl Acad Sci U S A **98**(4): 1901-6.

McAdam, R. A., P. W. Hermans, D. van Soolingen, Z. F. Zainuddin, D. Catty, J. D. van Embden and J. W. Dale (1990). "Characterization of a *Mycobacterium tuberculosis* insertion sequence belonging to the IS3 family." Mol Microbiol **4**(9): 1607-13.

McHugh, T. D. and S. H. Gillespie (1998). "Nonrandom association of IS*6110* and *Mycobacterium tuberculosis*: implications for molecular epidemiological studies." J Clin Microbiol **36**(5): 1410-3.

McShane, H., A. A. Pathan, C. R. Sander, N. P. Goonetilleke, H. A. Fletcher and A. V. Hill (2005). "Boosting BCG with MVA85A: the first candidate subunit vaccine for tuberculosis in clinical trials." Tuberculosis (Edinb) **85**(1-2): 47-52.

McShane, H., A. A. Pathan, C. R. Sander, S. M. Keating, S. C. Gilbert, K. Huygen, H. A. Fletcher and A. V. Hill (2004). "Recombinant modified vaccinia virus Ankara expressing antigen 85A boosts BCG-primed and naturally acquired antimycobacterial immunity in humans." Nat Med **10**(11): 1240-4.

Mehta, P. K., C. H. King, E. H. White, J. J. Murtagh, Jr. and F. D. Quinn (1996). "Comparison of in vitro models for the study of *Mycobacterium tuberculosis* invasion and intracellular replication." Infect Immun **64**(7): 2673-9.

Miotto, P., F. Piana, D. M. Cirillo and G. B. Migliori (2008). "Genotype MTBDRplus: a further step toward rapid identification of drug-resistant *Mycobacterium tuberculosis*." J Clin Microbiol **46**(1): 393-4.

Miotto, P., F. Piana, V. Penati, F. Canducci, G. B. Migliori and D. M. Cirillo (2006). "Use of genotype MTBDR assay for molecular detection of rifampin and isoniazid resistance in *Mycobacterium tuberculosis* clinical strains isolated in Italy." J Clin Microbiol **44**(7): 2485-91.

Moss, A. R., D. Alland, E. Telzak, D. Hewlett, Jr., V. Sharp, P. Chiliade, V. LaBombardi, D. Kabus, B. Hanna, L. Palumbo, K. Brudney, A. Weltman, K. Stoeckle, K. Chirgwin, M. Simberkoff, S. Moghazeh, W. Eisner, M. Lutfey and B. Kreiswirth (1997). "A city-wide outbreak of a multiple-drug-resistant strain of *Mycobacterium tuberculosis* in New York." Int J Tuberc Lung Dis **1**(2): 115-21.

Mostrom, P., M. Gordon, C. Sola, M. Ridell and N. Rastogi (2002). "Methods used in the molecular epidemiology of tuberculosis." Clin Microbiol Infect **8**(11): 694-704.

Movahedzadeh, F., A. Williams, S. Clark, G. Hatch, D. Smith, A. ten Bokum, T. Parish, J. Bacon and N. Stoker (2008). "Construction of a severely attenuated mutant of *Mycobacterium tuberculosis* for reducing risk to laboratory workers." Tuberculosis (Edinb) **88**(5): 375-81.

Neill, M. A. and S. J. Klebanoff (1988). "The effect of phenolic glycolipid-1 from *Mycobacterium leprae* on the antimicrobial activity of human macrophages." J Exp Med **167**(1): 30-42.

Neyrolles, O., K. Gould, M. P. Gares, S. Brett, R. Janssen, P. O'Gaora, J. L. Herrmann, M. C. Prevost, E. Perret, J. E. Thole and D. Young (2001). "Lipoprotein access to MHC class I presentation during infection of murine macrophages with live mycobacteria." J Immunol **166**(1): 447-57.

Niemann, S., E. Richter and S. Rusch-Gerdes (1999). "Stability of *Mycobacterium tuberculosis* IS*6110* restriction fragment length polymorphism patterns and spoligotypes determined by analyzing serial isolates from patients with drug-resistant tuberculosis." J Clin Microbiol **37**(2): 409-12.

Nikolayevskyy, V., K. Gopaul, Y. Balabanova, T. Brown, I. Fedorin and F. Drobniewski (2006). "Differentiation of tuberculosis strains in a population with mainly Beijing-family strains." Emerg Infect Dis **12**(9): 1406-13.

Noss, E. H., R. K. Pai, T. J. Sellati, J. D. Radolf, J. Belisle, D. T. Golenbock, W. H. Boom and C. V. Harding (2001). "Toll-like receptor 2-dependent inhibition of macrophage class II MHC expression and antigen processing by 19-kDa lipoprotein of *Mycobacterium tuberculosis*." J Immunol **167**(2): 910-8.

Nuzzo, I., M. Galdiero, C. Bentivoglio, R. Galdiero and C. Romano Carratelli (2002). "Apoptosis modulation by mycolic acid, tuberculostearic acid and trehalose 6,6'-dimycolate." J Infect **44**(4): 229-35.

O'Mahony, J. and C. Hill (2004). "Rapid real-time PCR assay for detection and quantitation of *Mycobacterium avium* subsp. *paratuberculosis* DNA in artificially contaminated milk." Appl Environ Microbiol **70**(8): 4561-8.

Oelemann, M. C., R. Diel, V. Vatin, W. Haas, S. Rusch-Gerdes, C. Locht, S. Niemann and P. Supply (2007). "Assessment of an optimized mycobacterial interspersed repetitive- unit-variable-number tandem-repeat typing system combined with spoligotyping for population-based molecular epidemiology studies of tuberculosis." J Clin Microbiol **45**(3): 691-7.

Olsen, A. W. and P. Andersen (2003). "A novel TB vaccine; strategies to combat a complex pathogen." Immunol Lett **85**(2): 207-11.

Olsen, A. W., A. Williams, L. M. Okkels, G. Hatch and P. Andersen (2004). "Protective effect of a tuberculosis subunit vaccine based on a fusion of antigen 85B and ESAT-6 in the aerosol guinea pig model." Infect Immun **72**(10): 6148-50.

Ordway, D., G. Palanisamy, M. Henao-Tamayo, E. E. Smith, C. Shanley, I. M. Orme and R. J. Basaraba (2007). "The cellular immune response to *Mycobacterium tuberculosis* infection in the guinea pig." J Immunol **179**(4): 2532-41.

Palanisamy, G. S., N. Duteau, K. D. Eisenach, D. M. Cave, S. A. Theus, B. N. Kreiswirth, R. J. Basaraba and I. M. Orme (2009). "Clinical strains of *Mycobacterium tuberculosis* display a wide range of virulence in guinea pigs." Tuberculosis (Edinb).

Park, J. S., M. H. Tamayo, M. Gonzalez-Juarrero, I. M. Orme and D. J. Ordway (2006). "Virulent clinical isolates of *Mycobacterium tuberculosis* grow rapidly and induce cellular necrosis but minimal apoptosis in murine macrophages." J Leukoc Biol **79**(1): 80-6.

Patti, J. M., B. L. Allen, M. J. McGavin and M. Hook (1994). "MSCRAMM-mediated adherence of microorganisms to host tissues." Annu Rev Microbiol **48**: 585-617.

Pavlic, M., F. Allerberger, M. P. Dierich and W. M. Prodinger (1999). "Simultaneous infection with two drug-susceptible *Mycobacterium tuberculosis* strains in an immunocompetent host." J Clin Microbiol **37**(12): 4156-7.

Persson, A., D. Chang and E. Crouch (1990). "Surfactant protein D is a divalent cation-dependent carbohydrate-binding protein." J Biol Chem **265**(10): 5755-60.

Pfaffl, M. W. (2001). "A new mathematical model for relative quantification in real-time RT-PCR." Nucleic Acids Res **29**(9): e45.

Prabhakar, R., P. Venkataraman, R. S. Vallishayee, P. Reeser, S. Musa, R. Hashim, Y. Kim, C. Dimmer, E. Wiegeshaus, M. L. Edwards and *et al.,*. (1987). "Virulence for guinea pigs of tubercle bacilli isolated from the sputum of participants in the BCG trial, Chingleput District, South India." Tubercle **68**(1): 3-17.

Pym, A. S., P. Brodin, L. Majlessi, R. Brosch, C. Demangel, A. Williams, K. E. Griffiths, G. Marchal, C. Leclerc and S. T. Cole (2003). "Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis." Nat Med **9**(5): 533-9.

Ragno, S., M. Romano, S. Howell, D. J. Pappin, P. J. Jenner and M. J. Colston (2001). "Changes in gene expression in macrophages infected with *Mycobacterium tuberculosis*: a combined transcriptomic and proteomic approach." Immunology **104**(1): 99-108.

Rao, V., N. Dhar and A. K. Tyagi (2003). "Modulation of host immune responses by overexpression of immunodominant antigens of *Mycobacterium tuberculosis* in bacille Calmette-Guerin." Scand J Immunol **58**(4): 449-61.

Ratliff, T. L., R. McCarthy, W. B. Telle and E. J. Brown (1993). "Purification of a mycobacterial adhesin for fibronectin." Infect Immun **61**(5): 1889-94.

Ratliff, T. L., J. A. McGarr, C. Abou-Zeid, G. A. Rook, J. L. Stanford, J. Aslanzadeh and E. J. Brown (1988). "Attachment of mycobacteria to fibronectin-coated surfaces." J Gen Microbiol **134**(5): 1307-13.

Riendeau, C. J. and H. Kornfeld (2003). "THP-1 cell apoptosis in response to Mycobacterial infection." Infect Immun **71**(1): 254-9.

Rodrigues, L. C., V. K. Diwan and J. G. Wheeler (1993). "Protective effect of BCG against tuberculous meningitis and miliary tuberculosis: a meta-analysis." Int J Epidemiol **22**(6): 1154-8.

Rodrigues, L. C., O. Noel Gill and P. G. Smith (1991). "BCG vaccination in the first year of life protects children of Indian subcontinent ethnic origin against tuberculosis in England." J Epidemiol Community Health **45**(1): 78-80.

Rondini, S., E. Mensah-Quainoo, H. Troll, T. Bodmer and G. Pluschke (2003). "Development and application of real-time PCR assay for quantification of *Mycobacterium ulcerans* DNA." J Clin Microbiol **41**(9): 4231-7.

Roring, S., A. Scott, D. Brittain, I. Walker, G. Hewinson, S. Neill and R. Skuce (2002). "Development of variable-number tandem repeat typing of *Mycobacterium bovis*: comparison of results with those obtained by using existing exact tandem repeats and spoligotyping." J Clin Microbiol **40**(6): 2126-33.

Ruddy, M., Y. Balabanova, C. Graham, I. Fedorin, N. Malomanova, E. Elisarova, S. Kuznetznov, G. Gusarova, S. Zakharova, A. Melentyev, E. Krukova, V. Golishevskaya, V. Erokhin, I. Dorozhkova and F. Drobniewski (2005). "Rates of drug resistance and risk factor analysis in civilian and prison patients with tuberculosis in Samara Region, Russia." Thorax **60**(2): 130-5.

Ruddy, M., T. D. McHugh, J. W. Dale, D. Banerjee, H. Maguire, P. Wilson, F. Drobniewski, P. Butcher and S. H. Gillespie (2002). "Estimation of the rate of unrecognized cross-contamination with *Mycobacterium tuberculosis* in London microbiology laboratories." J Clin Microbiol **40**(11): 4100-4.

Ruddy, M. C., A. P. Davies, M. D. Yates, S. Yates, S. Balasegaram, Y. Drabu, B. Patel, S. Lozewicz, S. Sen, M. Bahl, E. James, M. Lipman, G. Duckworth, J. M. Watson, M. Piper, F. A. Drobniewski and H. Maguire (2004). "Outbreak of isoniazid resistant tuberculosis in north London." Thorax **59**(4): 279-85.

Russell, D. G., J. Dant and S. Sturgill-Koszycki (1996). "*Mycobacterium avium*- and *Mycobacterium tuberculosis*-containing vacuoles are dynamic, fusion-competent vesicles that are accessible to glycosphingolipids from the host cell plasmalemma." J Immunol **156**(12): 4764-73.

Ryll, R., Y. Kumazawa and I. Yano (2001). "Immunological properties of trehalose dimycolate (cord factor) and other mycolic acid-containing glycolipids--a review." Microbiol Immunol **45**(12): 801-11.

Sambandamurthy, V. K. and W. R. Jacobs, Jr. (2005). "Live attenuated mutants of *Mycobacterium tuberculosis* as candidate vaccines against tuberculosis." Microbes Infect **7**(5-6): 955-61.

Sambandamurthy, V. K., X. Wang, B. Chen, R. G. Russell, S. Derrick, F. M. Collins, S. L. Morris and W. R. Jacobs, Jr. (2002). "A pantothenate auxotroph of *Mycobacterium tuberculosis* is highly attenuated and protects mice against tuberculosis." Nat Med **8**(10): 1171-4.

Sampson, S. L., R. M. Warren, M. Richardson, G. D. van der Spuy and P. D. van Helden (1999). "Disruption of coding regions by IS*6110* insertion in *Mycobacterium tuberculosis*." Tuber Lung Dis **79**(6): 349-59.

Savine, E., R. M. Warren, G. D. van der Spuy, N. Beyers, P. D. van Helden, C. Locht and P. Supply (2002). "Stability of variable-number tandem repeats of mycobacterial interspersed repetitive units from 12 loci in serial isolates of *Mycobacterium tuberculosis*." J Clin Microbiol **40**(12): 4561-6.

Schmid, D., R. Fretz, H. W. Kuo, R. Rumetshofer, S. Meusburger, E. Magnet, G. Hurbe, A. Indra, W. Ruppitsch, A. T. Pietzka and F. Allerberger (2008). "An outbreak of multidrug-resistant tuberculosis among refugees in Austria, 2005-2006." Int J Tuberc Lung Dis **12**(10): 1190-5.

Sebban, M., I. Mokrousov, N. Rastogi and C. Sola (2002). "A data-mining approach to spacer oligonucleotide typing of *Mycobacterium tuberculosis*." Bioinformatics **18**(2): 235-43.

Singh, U. B., N. Suresh, N. V. Bhanu, J. Arora, H. Pant, S. Sinha, R. C. Aggarwal, S. Singh, J. N. Pande, C. Sola, N. Rastogi and P. Seth (2004). "Predominant tuberculosis spoligotypes, Delhi, India." Emerg Infect Dis **10**(6): 1138-42.

Skeiky, Y. A., M. R. Alderson, P. J. Ovendale, J. A. Guderian, L. Brandt, D. C. Dillon, A. Campos-Neto, Y. Lobet, W. Dalemans, I. M. Orme and S. G. Reed (2004). "Differential immune responses and protective efficacy induced by components of a tuberculosis polyprotein vaccine, Mtb72F, delivered as naked DNA or recombinant protein." J Immunol **172**(12): 7618-28.

Skeiky, Y. A., P. J. Ovendale, S. Jen, M. R. Alderson, D. C. Dillon, S. Smith, C. B. Wilson, I. M. Orme, S. G. Reed and A. Campos-Neto (2000). "T cell expression cloning of a *Mycobacterium tuberculosis* gene encoding a protective antigen associated with the early control of infection." J Immunol **165**(12): 7140-9.

Skuce, R. A., T. P. McCorry, J. F. McCarroll, S. M. Roring, A. N. Scott, D. Brittain, S. L. Hughes, R. G. Hewinson and S. D. Neill (2002). "Discrimination of *Mycobacterium tuberculosis* complex bacteria using novel VNTR-PCR targets." Microbiology **148**(Pt 2): 519-28.

Smith, D. A., T. Parish, N. G. Stoker and G. J. Bancroft (2001). "Characterization of auxotrophic mutants of *Mycobacterium tuberculosis* and their potential as vaccine candidates." Infect Immun **69**(2): 1142-50.

Smith, I. (2003). "*Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence." Clin Microbiol Rev **16**(3): 463-96.

Smittipat, N., P. Billamas, M. Palittapongarnpim, A. Thong-On, M. M. Temu, P. Thanakijcharoen, O. Karnkawinpong and P. Palittapongarnpim (2005). "Polymorphism of variable-number tandem repeats at multiple loci in *Mycobacterium tuberculosis*." J Clin Microbiol **43**(10): 5034-43.

Sola, C., S. Ferdinand, C. Mammina, A. Nastasi and N. Rastogi (2001). "Genetic diversity of *Mycobacterium tuberculosis* in Sicily based on spoligotyping and variable number of tandem DNA repeats and comparison with a spoligotyping database for population-based analysis." J Clin Microbiol **39**(4): 1559-65.

Sola, C., I. Filliol, E. Legrand, S. Lesjean, C. Locht, P. Supply and N. Rastogi (2003). "Genotyping of the *Mycobacterium tuberculosis* complex using MIRUs: association with VNTR and spoligotyping for molecular epidemiology and evolutionary genetics." Infect Genet Evol **3**(2): 125-33.

Sow, F. B., W. C. Florence, A. R. Satoskar, L. S. Schlesinger, B. S. Zwilling and W. P. Lafuse (2007). "Expression and localization of hepcidin in macrophages: a role in host defense against tuberculosis." J Leukoc Biol **82**(4): 934-45.

Sreevatsan, S., X. Pan, K. E. Stockbauer, N. D. Connell, B. N. Kreiswirth, T. S. Whittam and J. M. Musser (1997). "Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination." Proc Natl Acad Sci U S A **94**(18): 9869-74.

Stenger, S. and M. Rollinghoff (2001). "Role of cytokines in the innate immune response to intracellular pathogens." Ann Rheum Dis **60 Suppl 3**: iii43-6.

Sterne, J. A., L. C. Rodrigues and I. N. Guedes (1998). "Does the efficacy of BCG decline with time since vaccination?" Int J Tuberc Lung Dis **2**(3): 200-7.

Sugawara, I., T. Udagawa, S. C. Hua, M. Reza-Gholizadeh, K. Otomo, Y. Saito and H. Yamada (2002). "Pulmonary granulomas of guinea pigs induced by inhalation exposure of heat-treated BCG Pasteur, purified trehalose dimycolate and methyl ketomycolate." J Med Microbiol **51**(2): 131-7.

Supply, P., C. Allix, S. Lesjean, M. Cardoso-Oelemann, S. Rusch-Gerdes, E. Willery, E. Savine, P. de Haas, H. van Deutekom, S. Roring, P. Bifani, N. Kurepina, B. Kreiswirth, C. Sola, N. Rastogi, V. Vatin, M. C. Gutierrez, M. Fauville, S. Niemann, R. Skuce, K. Kremer, C. Locht and D. van Soolingen (2006). "Proposal for Standardization of Optimized Mycobacterial Interspersed Repetitive Unit-Variable-Number Tandem Repeat Typing of *Mycobacterium tuberculosis*." J Clin Microbiol **44**(12): 4498-4510.

Supply, P., S. Lesjean, E. Savine, K. Kremer, D. van Soolingen and C. Locht (2001). "Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units." J Clin Microbiol **39**(10): 3563-71.

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

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

Supply, P., R. M. Warren, A. L. Banuls, S. Lesjean, G. D. Van Der Spuy, L. A. Lewis, M. Tibayrenc, P. D. Van Helden and C. Locht (2003). "Linkage disequilibrium between minisatellite loci supports clonal evolution of *Mycobacterium tuberculosis* in a high tuberculosis incidence area." Mol Microbiol **47**(2): 529-38.

Sutherland, I. and V. H. Springett (1987). "Effectiveness of BCG vaccination in England and Wales in 1983." Tubercle **68**(2): 81-92.

Tanaka, M. M. and N. A. Rosenberg (2001). "Optimal estimation of transposition rates of insertion sequences for molecular epidemiology." Stat Med **20**(16): 2409-20.

Tanaka, M. M., P. M. Small, H. Salamon and M. W. Feldman (2000). "The dynamics of repeated elements: applications to the epidemiology of tuberculosis." Proc Natl Acad Sci U S A **97**(7): 3532-7.

Tanghe, A., O. Denis, B. Lambrecht, V. Motte, T. van den Berg and K. Huygen (2000). "Tuberculosis DNA vaccine encoding Ag85A is immunogenic and protective when administered by intramuscular needle injection but not by epidermal gene gun bombardment." Infect Immun **68**(7): 3854-60.

Tenner, A. J., S. L. Robinson, J. Borchelt and J. R. Wright (1989). "Human pulmonary surfactant protein (SP-A), a protein structurally homologous to C1q, can enhance FcR- and CR1-mediated phagocytosis." J Biol Chem **264**(23): 13923-8.

Theus, S., K. Eisenach, N. Fomukong, R. F. Silver and M. D. Cave (2007). "Beijing family *Mycobacterium tuberculosis* strains differ in their intracellular growth in THP-1 macrophages." Int J Tuberc Lung Dis **11**(10): 1087-93.

Theus, S. A., M. D. Cave and K. D. Eisenach (2004). "Activated THP-1 cells: an attractive model for the assessment of intracellular growth rates of *Mycobacterium tuberculosis* isolates." Infect Immun **72**(2): 1169-73.

Theus, S. A., M. D. Cave and K. D. Eisenach (2005). "Intracellular macrophage growth rates and cytokine profiles of *Mycobacterium tuberculosis* strains with different transmission dynamics." J Infect Dis **191**(3): 453-60.

Thierry, D., A. Brisson-Noel, V. Vincent-Levy-Frebault, S. Nguyen, J. L. Guesdon and B. Gicquel (1990). "Characterization of a *Mycobacterium tuberculosis* insertion sequence, IS*6110*, and its application in diagnosis." J Clin Microbiol **28**(12): 2668-73.

Ting, L. M., A. C. Kim, A. Cattamanchi and J. D. Ernst (1999). "*Mycobacterium tuberculosis* inhibits IFN-gamma transcriptional responses without inhibiting activation of STAT1." J Immunol **163**(7): 3898-906.

Toungoussova, O. S., P. Sandven, A. O. Mariandyshev, N. I. Nizovtseva, G. Bjune and D. A. Caugant (2002). "Spread of drug-resistant *Mycobacterium tuberculosis* strains of the Beijing genotype in the Archangel Oblast, Russia." J Clin Microbiol **40**(6): 1930-7.

Tsolaki, A. G., A. E. Hirsh, K. DeRiemer, J. A. Enciso, M. Z. Wong, M. Hannan, Y. O. Goguet de la Salmoniere, K. Aman, M. Kato-Maeda and P. M. Small (2004). "Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains." Proc Natl Acad Sci U S A **101**(14): 4865-70.

van Crevel, R., T. H. Ottenhoff and J. W. van der Meer (2002). "Innate immunity to *Mycobacterium tuberculosis*." Clin Microbiol Rev **15**(2): 294-309.

van Embden, J. D., M. D. Cave, J. T. Crawford, J. W. Dale, K. D. Eisenach, B. Gicquel, P. Hermans, C. Martin, R. McAdam, T. M. Shinnick and *et al.*,. (1993). "Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology." J Clin Microbiol **31**(2): 406-9.

van Embden, J. D., T. van Gorkom, K. Kremer, R. Jansen, B. A. van Der Zeijst and L. M. Schouls (2000). "Genetic variation and evolutionary origin of the direct repeat locus of *Mycobacterium tuberculosis* complex bacteria." J Bacteriol **182**(9): 2393-401.

van Soolingen, D., P. E. de Haas, P. W. Hermans, P. M. Groenen and J. D. van Embden (1993). "Comparison of various repetitive DNA elements as genetic markers for strain differentiation and epidemiology of *Mycobacterium tuberculosis*." J Clin Microbiol **31**(8): 1987-95.

van Soolingen, D., P. W. Hermans, P. E. de Haas, D. R. Soll and J. D. van Embden (1991). "Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis." J Clin Microbiol **29**(11): 2578-86.

Vipond, J., M. L. Cross, M. R. Lambeth, S. Clark, F. E. Aldwell and A. Williams (2008). "Immunogenicity of orally-delivered lipid-formulated BCG vaccines and protection against *Mycobacterium tuberculosis* infection." Microbes Infect **10**(14-15): 1577-81.

Vipond, J., R. Vipond, E. Allen-Vercoe, S. O. Clark, G. J. Hatch, K. E. Gooch, J. Bacon, T. Hampshire, H. Shuttleworth, N. P. Minton, K. Blake, A. Williams and P. D. Marsh (2006). "Selection of novel TB vaccine candidates and their evaluation as DNA vaccines against aerosol challenge." Vaccine **24**(37-39): 6340-50.

Vitol, I., J. Driscoll, B. Kreiswirth, N. Kurepina and K. P. Bennett (2006). "Identifying *Mycobacterium tuberculosis* complex strain families using spoligotypes." Infect Genet Evol **6**(6): 491-504.

Volpe, E., G. Cappelli, M. Grassi, A. Martino, A. Serafino, V. Colizzi, N. Sanarico and F. Mariani (2006). "Gene expression profiling of human macrophages at late time of infection with *Mycobacterium tuberculosis*." Immunology **118**(4): 449-60.

Voorhout, W. F., T. Veenendaal, Y. Kuroki, Y. Ogasawara, L. M. van Golde and H. J. Geuze (1992). "Immunocytochemical localization of surfactant protein D (SP-D) in type II cells, Clara cells, and alveolar macrophages of rat lung." J Histochem Cytochem **40**(10): 1589-97.

Wada, T., S. Maeda, A. Hase and K. Kobayashi (2007). "Evaluation of variable numbers of tandem repeat as molecular epidemiological markers of *Mycobacterium tuberculosis* in Japan." J Med Microbiol **56**(Pt 8): 1052-7.

Wallgren, A. (1948). "The time-table of tuberculosis." Tubercle **29**(11): 245-51.

Warren, R. M., G. D. van der Spuy, M. Richardson, N. Beyers, M. W. Borgdorff, M. A. Behr and P. D. van Helden (2002). "Calculation of the stability of the IS*6110* banding pattern in patients with persistent *Mycobacterium tuberculosis* disease." J Clin Microbiol **40**(5): 1705-8.

Warren, R. M., T. C. Victor, E. M. Streicher, M. Richardson, G. D. van der Spuy, R. Johnson, V. N. Chihota, C. Locht, P. Supply and P. D. van Helden (2004). "Clonal expansion of a globally disseminated lineage of *Mycobacterium tuberculosis* with low IS*6110* copy numbers." J Clin Microbiol **42**(12): 5774-82.

Whelan, J. A., N. B. Russell and M. A. Whelan (2003). "A method for the absolute quantification of cDNA using real-time PCR." J Immunol Methods **278**(1-2): 261-9.

WHO (2004). "Weekly epidemiological record." Geneva: World Health Organisation.

WHO (2007). "Global tuberculosis control: surveillance, planning, financing." WHO report 2007 **Geneva: World Health Organisation**.

WHO (2009). "Global tuberculosis control: surveillance, planning, financing." WHO Report 2009 **Geneva: World Health Organisation**.

Williams, A., A. Davies, P. D. Marsh, M. A. Chambers and R. G. Hewinson (2000). "Comparison of the protective efficacy of bacille calmette-Guerin vaccination against aerosol challenge with *Mycobacterium tuberculosis* and *Mycobacterium bovis*." Clin Infect Dis **30 Suppl 3**: S299-301.

Williams, A., N. P. Goonetilleke, H. McShane, S. O. Clark, G. Hatch, S. C. Gilbert and A. V. Hill (2005). "Boosting with poxviruses enhances *Mycobacterium bovis* BCG efficacy against tuberculosis in guinea pigs." Infect Immun **73**(6): 3814-6.

Williams, A., G. J. Hatch, S. O. Clark, K. E. Gooch, K. A. Hatch, G. A. Hall, K. Huygen, T. H. Ottenhoff, K. L. Franken, P. Andersen, T. M. Doherty, S. H. Kaufmann, L. Grode, P. Seiler, C. Martin, B. Gicquel, S. T. Cole, P. Brodin, A. S. Pym, W. Dalemans, J. Cohen, Y. Lobet, N. Goonetilleke, H. McShane, A. Hill, T. Parish, D. Smith, N. G. Stoker, D. B. Lowrie, G. Kallenius, S. Svenson, A. Pawlowski, K. Blake and P. D. Marsh (2005). "Evaluation of vaccines in the EU TB Vaccine Cluster using a guinea pig aerosol infection model of tuberculosis." Tuberculosis (Edinb) **85**(1-2): 29-38.

Williams, A., B. W. James, J. Bacon, K. A. Hatch, G. J. Hatch, G. A. Hall and P. D. Marsh (2005). "An assay to compare the infectivity of *Mycobacterium tuberculosis* isolates based on aerosol infection of guinea pigs and assessment of bacteriology." Tuberculosis (Edinb) **85**(3): 177-84.

Wilson, S. M., S. Goss and F. Drobniewski (1998). "Evaluation of strategies for molecular fingerprinting for use in the routine work of a Mycobacterium reference unit." J Clin Microbiol **36**(11): 3385-8.

Yaganehdoost, A., E. A. Graviss, M. W. Ross, G. J. Adams, S. Ramaswamy, A. Wanger, R. Frothingham, H. Soini and J. M. Musser (1999). "Complex transmission dynamics of clonally related virulent *Mycobacterium tuberculosis* associated with barhopping by predominantly human immunodeficiency virus-positive gay men." J Infect Dis **180**(4): 1245-51.

Yeh, R. W., A. Ponce de Leon, C. B. Agasino, J. A. Hahn, C. L. Daley, P. C. Hopewell and P. M. Small (1998). "Stability of *Mycobacterium tuberculosis* DNA genotypes." J Infect Dis **177**(4): 1107-11.

Yeremeev, V. V., I. V. Lyadova, B. V. Nikonenko, A. S. Apt, C. Abou-Zeid, J. Inwald and D. B. Young (2000). "The 19-kD antigen and protective immunity in a murine model of tuberculosis." Clin Exp Immunol **120**(2): 274-9.

Yokoyama, E., K. Kishida, M. Uchimura and S. Ichinohe (2006). "Comparison between agarose gel electrophoresis and capillary electrophoresis for variable numbers of tandem repeat typing of *Mycobacterium tuberculosis*." J Microbiol Methods **65**(3): 425-31.

APPENDIX 1

The expected molecular weights, in base pairs, of amplified fragments with various copy numbers in MIRU loci.

Copy number	Molecular weight of fragment (base pairs)											
	2	**4**	**10**	**16**	**20**	**23**	**24**	**26**	**27**	**31**	**39**	**40**
0	189	103	219	367	215	78	325	244	272	106	194	229
1	238	189	272	420	292	131	375	295	325	159	243	280
2	287	264	325	473	369	183	425	344	378	212	292	331
3	336	339	378	526	446	235	475	393	431	265	341	382
4	385	414	431	579	523	287	525	442	484	318	390	433
5	434	489	484	632	600	339	575	491	537	371	439	484
6	483	564	537	685	677	391	625	540	590	424	488	535
7	532	638	590	738		443		589	643	477	537	586
8	581	713	643			495		638		530	586	637
9	630	788	696			547		687		583	635	688
10		863	749			599		736		636		739
11			802			651		785				790
12								834				
13								883				
14								932				

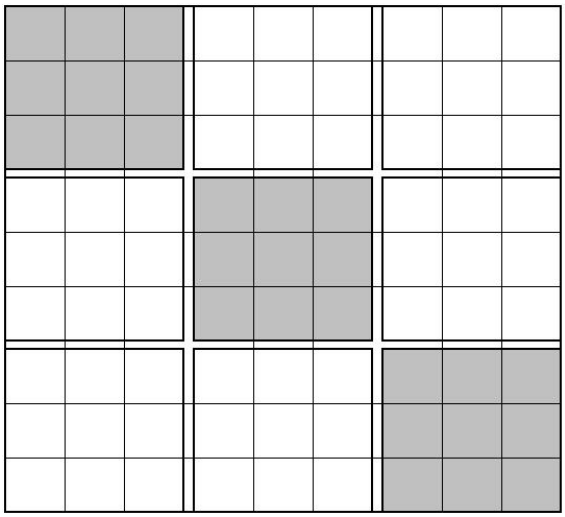
APPENDIX 2

The expected molecular weights, in base pairs, of amplified fragments with various copy numbers in ETR and VNTR loci.

Copy number	Molecular weight of fragment (base pairs)									
	A	**B**	**C**	**2163B**	**2347**	**3232**	**2163A**	**1982**	**3336**	**4052**
0	195	121					117	178	270	
1	270	174	133	133	385	242	186	256	325	266
2	346	227	187	202	441	286	255	334	380	376
3	422	280	242	273	501	330	324	412	435	488
4	499	333	297	343	556	372	393	490	490	596
5	570	386	350	412	611	415	462	568	545	710
6	645	439	405	480	667	458	531	646	600	820
7	720	492	460	548	724	501	600	727	655	935
8	795	545	515	617	781	546	669	802	710	1050
9	870	598	570	686	838	587	738	880	765	1165
10	945	651	625	755		630	807	958	820	1280
11	1020			824		673	876	1038	875	
12				892		716	945	1126	930	
13				963		759	1014	1272	985	
14				1034		802	1083	1350	1040	
15				1105		839	1152	1428	1095	
16				1176		888	1221		1150	
17						931	1290		1205	
18						974	1359		1260	
19						1017	1428			
20						1060	1497			
21						1103	1566			
22						1150	1635			
23						1200	1704			
24						1247	1773			
25							1842			
26							1911			
27							1980			
28							2049			
29							2118			
30							2187			
31							2256			
32							2325			
33							2394			
34							2463			

APPENDIX 3

The calculation used to perform a cell count of THP-1 cells using a Kova slide II.



Kova slide II

The number of bright, clear round cells (*x*) were counted in the smaller grids shaded in grey, making sure not to include cells that were in the spaces between the smaller grids.

The following calculation was done to establish the number of THP-1 cells per ml, which had been pooled in the falcon tube:-

$x \times 3$ = the number of cells in the whole large grid

= the number of cells/μl as 1μl covers the large grid

To account for the 1 in 2 dilution with Trypan blue solution:-

Number of cells/μl × 2 = actual number of cells/μl

Therefore:-

Number of cells/ml = actual number of cells/μl × 1000

APPENDIX 4

The calculation used to determine the number of genome molecules/µl.

Total *M. tuberculosis* genome	4,411,529 base pairs (b.p.)
Total guanidine + cytosine	2,894,585 b.p. (single strand) 5,789,170 b.p. (double strand)
Total adenosine + tyrosine	1,516,944 b.p. (single strand) 3,033,888 b.p. (double strand)

Compound molecular weight for base pairs:-

adenosine base	331.2 g/mol	cytosine base	307.2 g/mol
guanidine base	347.2 g/mol	tyrosine base	322.2 g/mol

guanidine + cytosine (307.2+347.2) ÷ 2 = 327.2 g/mol
adenosine + tyrosine (331.2+322.2) ÷ 2 = 326.7 g/mol

Compound molecular weight for *M. tuberculosis* genome:-

guanidine + cytosine 5,789,170 b.p. × 327.2 g/mol = 1,894,216,424 g/mol
adenosine + tyrosine 3,033,888 b.p. × 326.7 g/mol = 991,171,209.6 g/mol

Total = 2,885,387,633.6 g/mol

However when 2 bases bind a water (H_2O) molecule is lost.

The compound molecular weight for H_2O is 18 g/mol.

There are 4,411,529 b.p., but when calculating the g/mol H_2O lost the calculation is:-

4,411,528 b.p. × 18 g/mol = 79,407,504 g/mol

Therefore the actual g/mol of genome = 2,885,387,633.6 - 79,407,504
= 2,805,980,129.6 g/mol

Avogadro's constant = 6.02×10^{23} molecules/mol

The DNA concentration in the extract measured by nanodrop
= 124.1 ng/µl (1.241×10^{-7} g/µl)

2,805,980,129.6 g = 1 mol

1.241×10^{-7} g/µl = (1.241×10^{-7} g/µl ÷ 2,805,980,129.6 g) mol
= $4.422697035 \times 10^{-17}$ mol in 1µl

1 mol = 6.02×10^{23} molecules

$4.422697035 \times 10^{-17}$ mol = ($4.422697035 \times 10^{-17}$ mol × 6.02×10^{23} molecules)
= 26,624,636.15 molecules in 1µl

Therefore the DNA standard contains 2.7×10^7 genome molecules/µl.

Discriminatory Ability of Hypervariable Variable Number Tandem Repeat Loci in Population-based Analysis of *Mycobacterium tuberculosis* Strains, London, UK

Preya Velji,[1] Vladyslav Nikolayevskyy,[1] Timothy Brown, and Francis Drobniewski

To address conflicting results about the stability of variable number tandem repeat (VNTR) loci and their value in prospective molecular epidemiology of *Mycobacterium tuberculosis*, we conducted a large prospective population-based analysis of all *M. tuberculosis* strains in a metropolitan setting. Optimal and reproducible conditions for reliable PCR and fragment analysis, comprising enzymes, denaturing conditions, and capillary temperature, were identified for a panel of hypervariable loci, including 3232, 2163a, 1982, and 4052. A total of 2,261 individual *M. tuberculosis* isolates and 265 sets of serial isolates were analyzed by using a standardized 15-loci VNTR panel, then an optimized hypervariable loci panel. The discriminative ability of loci varied substantially; locus VNTR 3232 varied the most, with 19 allelic variants and Hunter-Gaston index value of 0.909 . Hypervariable loci should be included in standardized panels because they can provide consistent comparable results at multiple settings, provided the proposed conditions are adhered to.

Globally, tuberculosis (TB) accounts for almost 2 million deaths each year (*1*). Although TB notification rates in the United Kingdom (13.8/100,000 in 2007) remain low, rates differ substantially by region: London (43.2/100,000) accounts for ≈40% of all TB cases registered in the United Kingdom, and ≈75% of TB patients in London were born abroad (*2*). Rates of drug resistance also are higher in London than in the rest of the United Kingdom: 8.6% of isolates are isoniazid resistant, and 1.2% are multidrug resistant (UK Health Protection Agency; www.hpa.org.uk).

In settings where incidence of TB is low or moderate, molecular genotyping is used to investigate suspected TB outbreaks, laboratory cross-contamination, and reactivation and (at a population level) to identify clustered cases that are not apparently linked; for the latter purpose, the highest possible level of discrimination is required (*3*). For these purposes, insertion sequence (IS) *6110* restriction fragment length polymorphism (RFLP) analysis—often supplemented with spoligotyping and, more recently, with variable number tandem repeat (VNTR) typing—is used routinely.

The highest levels of epidemiologic discrimination of strains of the *Mycobacterium tuberculosis* complex (MTBC) can be achieved by using multilocus VNTR typing, but these results depend on the number and loci used, particularly for homogenous strain groups such as the Beijing family (*3*–*5*). This approach overcomes technical difficulties associated with IS*6110*-RFLP and is amenable to automation that results in a high throughput (*6*–*10*). A standardized panel of 15 + 9 VNTR loci (24 loci) has been proposed (*7*,*11*), but it is unclear whether sufficient discrimination would be seen when the panel is used in populations with a substantial prevalence of homogenous MTBC families (*4*,*5*,*12*). In addition, the discriminative power of VNTR loci may vary markedly among genetic families (*7*,*13*). Recent studies evaluating the discriminative power of VNTR typing have produced conflicting results that were generated by using convenience samples (small populations with low diversity or populations confined to a single geographic setting). These studies highlighted a need for larger population-based studies to identify discrimina-

Author affiliations: Barts and The London School of Medicine, Queen Mary, University of London, London, UK (P. Velji, V. Nikolayevskyy); and Health Protection Agency National Mycobacterium Reference Laboratory, London (T. Brown, F. Drobniewski)

DOI: 10.3201/eid1510.090463

[1]These authors contributed equally to this article.

tive VNTR loci and ascertain their applicability for various genetic groups.

Concerns about the stability and reproducibility of particularly useful hypervariable loci, such as 3232, 2163a, 3336, and 1982 (*3*–*5*,*14*), have been raised (*7*,*15*). As a result, they have been excluded from the proposed international panels for VNTR typing. For these reasons, we conducted a study to examine the stability of hypervariable loci and the parameters associated with reproducibility, to select loci suitable for prospective molecular epidemiologic studies, and to evaluate the discriminatory power of these loci at a population level in a metropolitan setting.

Materials and Methods

Bacterial Isolates

A total of 2,261 individual MTBC isolates (1 per patient) were included in this prospectively designed population study. These isolates represented 95.7% of the bacteriologically confirmed TB cases reported from the 30 London hospitals in the 12 months from April 2005 through March 2006. These isolates had been characterized by using spoligotyping, and all but 4 were assigned to 1 of 36 spoligotype families (*16*,*17*). Multiple isolates were available from 265 patients (11.7%), resulting in serial isolate sets of 2–6 isolates, which had been sampled at intervals of 3 days to 11 months (N = 632).

Multilocus VNTR Analysis

All extracts were typed by using 15 mycobacterial interspersed repetitive unit (MIRU)-VNTR loci as previously described (*3*). Isolates clustered when the 15 MIRU- Hunter-Gaston index value VNTR profiles we used were reanalyzed with an additional panel of VNTR loci 2163b, 2347, 3232, 2163a, 1982, 3336, and 4052 as previously described (*3*,*5*) after optimization of factors affecting reproducibility (see Hypervariable Loci Optimization). Variability or discrimination at a locus was assessed by using the Hunter-Gaston Discriminative Index (HGDI) (*18*). Loci with HGDI values <0.3, 0.3–0.6, and >0.6 were considered poorly, moderately, and highly discriminative, respectively (*19*).

Hypervariable Loci Optimization

We selected 16 previously characterized MTBC isolates to cover the complete range of repeat sizes at control loci MIRU 26 and exact tandem repeat (ETR)–B and experimental hypervariable loci VNTRs 1982 and 3232 (except 0 repeats for the locus 3232). For each of the 16 extracts, four 10-μL PCRs were conducted for each of the primer mixes in duplicate. Of these 4 reactions, the first was performed as described previously with BIOTAQ polymerase (Bioline, London, UK) (any enzyme in the given context means enzyme in conjunction with the buffer recommended and supplied by a manufacturer). Three other sets of PCRs were conducted under different amplification conditions (*1*).

Method 1

Diamond DNA polymerase (Bioline) was used (*9*). The PCR amplification cycle was 3 min at 95°C, followed by 35 cycles of 30 s at 95°C, 30 s at 60°C, and 2 min at 72°C, and 1 final cycle of 5 min at 72°C (*2*).

Method 2

HotStartTaq DNA polymerase (QIAGEN, Hilden Germany) was used. Each 10-μL reaction contained 1× PCR buffer (QIAGEN), 0.25 U/μL of the relevant polymerase, 0.2 μmol/L dNTPs, 0.125 μmol/L of relevant primer, and 5% dimethylsulfoxide. The DNA amplification cycle was 15 min at 95°C, followed by 35 cycles of 30 s at 94°C, 30 s at 60°C, and 1 min at 72°C, and a final cycle of 10 min at 72°C (*3*).

Method 3

HotStartTaq Plus DNA polymerase (QIAGEN) was used. The PCR mixture was the same as in method 2, and the amplification cycle was the same, except that the initial 95°C activation time was reduced to 5 min.

We manually calculated the number of repeats within each PCR product by resolving 4 μLof each product on a 1.2% (wt/vol) agarose gel (Agarose LE Analytical grade; Promega, Southampton, UK) against a 2,000-bp HyperLadder II standard (Bioline). The number of repeats at each locus also was calculated by sizing in a denaturing capillary electrophoresis system using a CEQ 8000 instrument with a DNA Size Standard 600 (Beckman Coulter, High Wycombe, UK) and MapMarker D1 labeled 640–1000 (BioVentures, Inc., Murfreesboro, TN, USA) because fragments were expected to be >600 bp. Three parameter sets (Table 1) were used to analyze all fragments. The different parameters examined were capillary temperature (60°C for methods 1 and 2 and 50°C for method 3, respectively), denaturation time (120 s for method 1 and 180 s for methods 2 and 3, respectively) and separation time (60 min for methods 1 and 2 and 70 min for method 3, respectively). Fragment data traces were automatically analyzed by using the scheme shown in Table 1. For locus 3232, we accounted for offset values (i.e., difference among actual sizes of PCR fragments and apparent sizes indicated by electrophoresis) when calculating number of repeats in Table 1.

Assessing Stability and Reproducibility of VNTR Loci

All isolates were grouped into 265 sets of serial isolates (2–6 isolates each) and typed at all 22 loci. Primer sequences for all loci were as described previously (*3*,*9*,*20*,*21*). PCR was set up by using BIOTAQ polymerase for amplifying 12 MIRU and 3 ETR loci and Diamond polymerase for the

Table 1. Expected molecular weights of *Mycobacterium tuberculosis* of fragments at each locus, with different numbers of copies, London, UK, 2005–2006*

	Length of expected fragments for each locus, bp			
No. repeats	MIRU 26	ETR-B	VNTR 1982	VNTR 3232†
0	244	121	178	
1	295	174	256	242
2	344	227	334	286
3	393	280	412	330
4	442	333	490	372
5	491	386	568	415
6	540	439	646	458
7	589	492	727	501
8	638	545	802	546
9	687	598	880	587
10	736	651	958	630
11	785		1,038	673
12	834		1,116	716
13	883		1,194	759
14	932			802
15				845
16				888
17				931
18				974
19				1,017
20				1,060

*MIRU, mycobacterial interspersed repetitive unit; ETR, exact tandem repeat; VNTR, variable number tandem repeats.
†No isolates had 0 repeats in locus 3232 in our population.

additional 7 VNTR loci. Capillary electrophoresis was performed by using the parameters described in method 1.

Results

Optimization of Hypervariable Loci

We evaluated factors that potentially affect the reproducibility of hypervariable VNTR loci by using various PCR and capillary and manual electrophoresis separation conditions as described in the Materials and Methods. The ability to correctly amplify different VNTR loci depended on the enzyme used (Table 2); all polymerases efficiently amplified MIRU 26 and ETR-B, as indicated by the presence of PCR fragments on agarose gels and capillary electrophoresis peaks. However, locus VNTR 3232 was amplified effectively only with Bioline Diamond (15/16 strains, 93.8%). Although all polymerases except Bioline BIOTAQ were able to amplify DNA at locus VNTR 1982, longer fragments were amplified more efficiently by QIAGEN and Bioline Diamond polymerases. Therefore, Diamond polymerase was selected for the amplification of additional VNTR loci.

We assessed 3 methods for capillary electrophoresis. For each locus, apparent fragment sizes were plotted against expected fragment sizes for each method (Figure 1).

MIRU 26 fragments sizes were as expected for all allelic variants (except for the variant with 2 repeats) when BIOTAQ and Diamond polymerases were used, but sizes were larger than expected with QIAGEN polymerases. The smaller ETR-B fragments with 1 and 2 repeats all gave expected sizes with methods 1 and 2 but were less than expected with method 3 (where the capillary temperature was decreased). These results did not affect overall interpretation. For the higher number of repeats (4–6 repeats), all polymerases generated fragments that, when analyzed by using method 3, gave apparent sizes lower than expected. In some cases, this result affected the interpretation. The apparent sizes of VNTR 1982 fragments were all similar to the expected values independent of the polymerase used and the method used for capillary electrophoresis.

Serial Isolates

Amplification was performed by using BIOTAQ polymerase for 12 MIRU and 3 ETR loci and Diamond

Table 2. Number of DNA extracts (from n = 16) for which peaks were detected by different conditions for capillary electrophoresis of *Mycobaterium tuberculosis* after amplifying the loci with different polymerases, London, UK, 2005–2006*

		Bioline polymerases‡		QIAGEN polymerases‡	
Locus	Method†	BIOTAQ	Diamond	HotStartTaq	HotStartTaq Plus
MIRU 26	1	16	16	16 (1)	16 (1)
	2	15	16	16 (1)	16 (1)
	3	16	16	16	16
ETR-B	1	16	16	16 (1)	16 (1)
	2	15	16	16 (1)	16 (1)
	3	16	16	16 (2)	16 (2)
VNTR 1982	1	8	13	14	14
	2	9	13	12	14
	3	6	11	12	14
VNTR 3232	1	11	15	13	14
	2	10	15	14	14
	3	11 (3)	15 (7)	13 (6)	13 (4)

*MIRU, mycobacterial interspersed repetitive unit; ETR, exact tandem repeat; VNTR, variable number tandem repeats.
†Refer to Table 1.
‡Numbers in parentheses represent number of extracts whose calculated number of repeats were higher and lower than the expected value on the basis of that produced by the standard procedure (method 1).

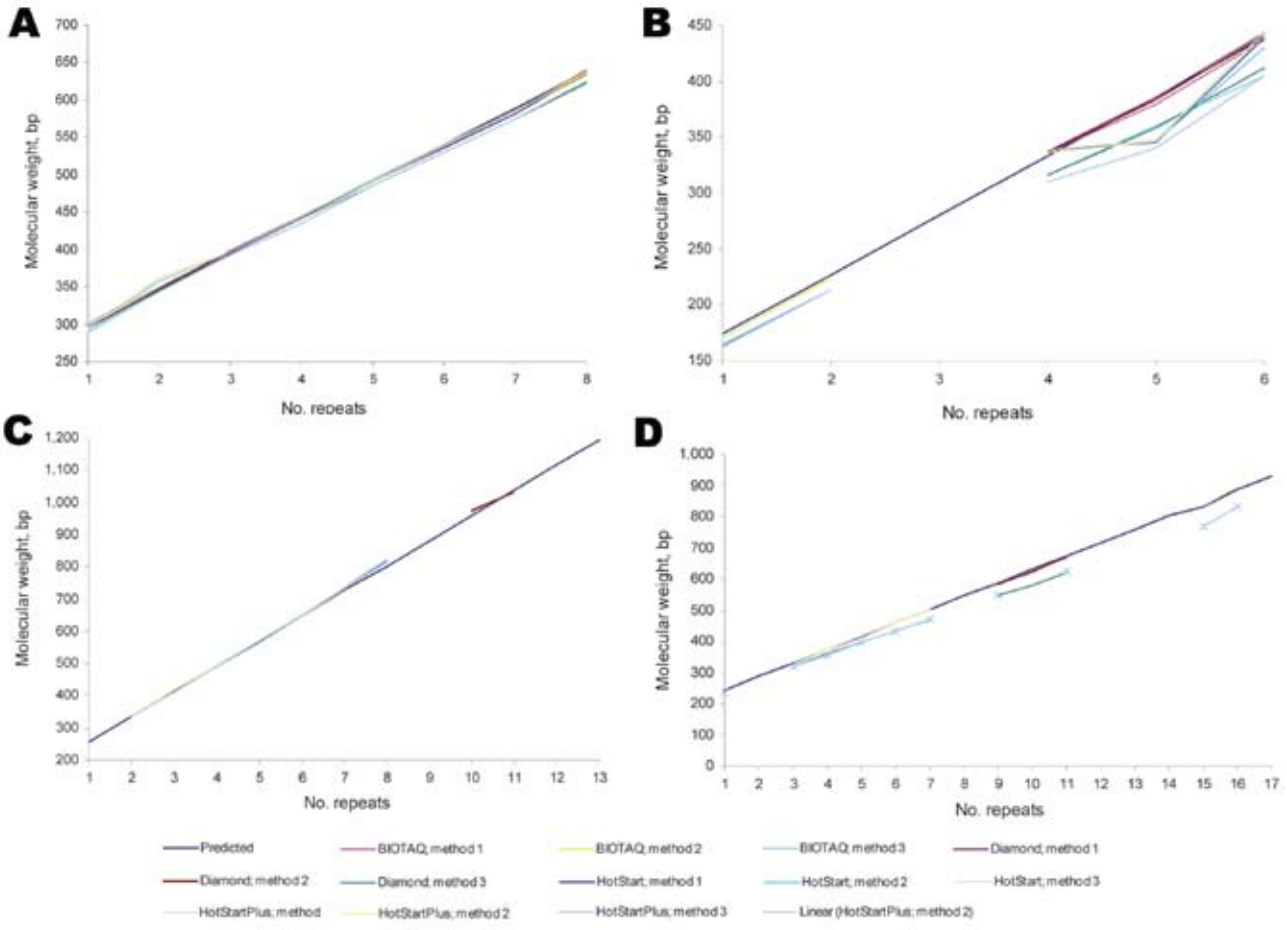


Figure 1. Effect of various enzymes and separation conditions on amplification and detectable molecular weights of PCR fragments for 4 variable number tandem repeat (VNTR) loci. A) Mycobacterial interspersed repetitive unit locus 26; B) locus exact tandem repeat; C) locus 1982; D) locus 3232.

polymerase for 7 VNTR loci with the optimized parameters in method 1. Analysis was blinded. No disagreements occurred in the interpretation of VNTR repeat numbers among isolates in a set. In a proportion of isolates (N = 124), genotyping results were validated by using both capillary electrophoresis and manual electrophoresis for PCR fragment separation, and again, no discrepancies were found between VNTR loci copy numbers in strains isolated from the same patient at different time points (Figure 2).

Population Genotyping in Metropolitan Setting with 2 Panels of VNTR Loci

A total of 2,261 MTBC isolates circulating in London with known spoligotypes were genotyped by using a defined set of 15 loci (12 MIRU and 3 ETR); all known spoligotyping families were represented in the test population (online Technical Appendix, available from www.cdc.gov/EID/content/15/10/1609-Techapp.pdf). Complete 15-loci profiles were obtained for 2,046 strains (90.5% of all strains). Data for the remaining profiles were incomplete for ≥1 locus. Overall PCR failure rate was 1.6%, with the highest number of failures (n = 72) at locus ETR-A and the lowest number of failures (n = 4) at locus ETR-C. When PCR failed, DNA was reextracted from original cultures, and genotyping was attempted again. If the second attempt was unsuccessful, the results for the locus were marked as missing.

Genotyping of MTBC isolates by using 15 MIRU-ETR loci yielded 1,036 unique profiles and 235 clusters containing 2–53 isolates (Table 3). Clustered profiles were shared by 1,225 isolates, giving a clustering rate of 54.2%.

Subsequently, 1,196 (97.6%) of 1,225 isolates (15 MIRU-ETR clustered isolates) were subjected to secondary typing by using VNTR loci 2163b, 2347, 3232, 2163a, 1982, 3336, and 4052. Resolution improved because strains that had been clustered initially were subdivided into new groups: 1,730 isolates now had unique genotyping patterns, and the remaining 502 isolates were grouped into 158 clusters, giving a new, substantially lower, clustering rate of 22.2% (Table 3).

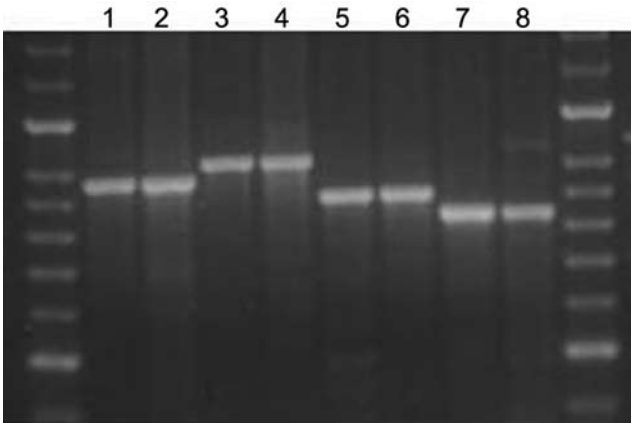


Figure 2. Agarose gel showing the stability of amplified fragments of variable number tandem repeat (VNTR) 3336 from 2 serial isolates isolated from 4 patients. Lane 1, patient A, isolate 1, isolated 2005 Jun 20, 8 copies; lane 2, patient A, isolate 2, isolated 2005 Jul 11, 8 copies; lane 3, patient B, isolate 1, isolated 2005 Jul 8, 9 copies; lane 4, patient B, isolate 2, isolated 2005 Aug 8, 9 copies; lane 5, patient C, isolate 1, isolated 2005 Nov 11, 7 copies; lane 6, patient C, isolate 2, isolated 2005 Nov 15, 7 copies; lane 7, patient D, isolate 1, isolated 2005 May 16, 6 copies; lane 8, patient D, isolate 2, isolated 2005 May 25, 6 copies.

Variability and Discriminative Power of VNTR Loci

The discriminative ability of VNTR loci varied markedly among the 22 VNTR loci and among spoligotyping families (online Technical Appendix) with locus VNTR 3232 showing the greatest variation (HGDI = 0.909 and 19 allelic variants) and loci MIRU 2 and 20, the least (HGDI = 0.134 and 0.196; number of allelic variants 4 and 3, respectively). Twelve loci each had ≥10 allelic variants. MIRU 4 showed moderate discriminative power, and MIRU 10, MIRU 16, MIRU 23, MIRU 26, MIRU 40, ETR-A, ETR-C, and VNTR 2163B, 2163A, 1982, 3232, 3336, and 4052 showed high discriminative power with HGDI values varying from 0.524 to 0.909. None of the 22 loci were monomorphic in the current study. With the exception of VNTR 2347, all loci included in the additional VNTR panel displayed higher variability than the primary panel of 15 MIRU-ETR loci used for UK national typing, which indicates their potential for increasing the power of prospective molecular genotyping.

The discriminative power of VNTR loci also varied among spoligotype families. The mean 15 MIRU-ETR HGDI value for the Beijing family was low (0.163), which indicates that this family is relatively homogeneous, even within the diverse London population settings. Notably, mean 15 MIRU-ETR HGDI values for genetic families within the Euro-American lineage (T, Haarlem, S, X, Latin American–Mediterranean) were generally higher (0.307–0.378) than those for Beijing and Central Asian (CAS) (0.235). Within spoligotype families, the additional 7 VNTR increased variability in all cases, except for *M. bovis*. The highest HGDI were seen in the Latin American–Mediterranean family with locus 2163B; in Beijing, Haarlem, and *M. africanum* with VNTR 3232; in East African–Indian with VNTR 2163A; in X with VNTR 1982; in T with VNTR 3336; and in CAS with VNTR 4052. Within the East African–Indian family, the hypervariable loci VNTR 3232 varied little, with 93.7% isolates having a single copy. A small proportion of strains (Table 4) analyzed by using more discriminative loci, including VNTR 3232, 1982, 2163A, and 3336, generated PCR products that were too large for automated analysis but were resolved manually.

Discussion

Polymorphisms in rapidly evolving repetitive sequences, such as minisatellite VNTR, are a valuable tool for prospective epidemiologic analyses and provide a high degree of discrimination in situations in which few a priori epidemiologic data are available. In this population-based study, we genotyped 2,261 individual MTBC isolates obtained from patients residing in London by using 22 VNTR-MIRU loci.

Conflicting views on the use of hypervariable loci for typing have been reported, even when loci such as VNTR 3232 have been shown to have high discriminatory power (*3*,*5*,*14*). Some studies have demonstrated difficulty in amplification of multiple alleles, absence of PCR amplification products, varying data interpretation, and lack of reproducibility among laboratories (*7*). Similar problems were found with another potentially valuable hypervariable locus, VNTR 1982 (*5*,*7*). Therefore, we believed that by identifying the conditions that provided good, reproducible discrimination, we would be able to define the optimal conditions that would enable molecular epidemiologists to use VNTR 1982 and 3232. We addressed variability and reproducibility for these 2 loci using MIRU 26 and ETR-B as

Table 3. Discriminatory power of VNTR typing used in the study in establishing true minimum cluster size as marker of real transmission rate*

Genotyping method	No. distinct profiles (variety of types)	No. clusters	Size of clusters, no. isolates	Clustering rate, % (n/N)	Recent transmission rate, % ([*n* – *c*]/*N*)	No. unique isolates
MIRU15 (n = 2261)	1,271	235	2–53	54.2	44.0	1,036
MIRU15 + Spoligotyping	1,619	196	2–48	37.1	29.0	1,423
MIRU15 + VNTR7	1,888	158	2–35	22.2	17.0	1,730

*MIRU, mycobacterial interspersed repetitive unit; VNTR, variable number tandem repeats; n, no. clustered cases; N, total no. of strains; c, no. of clusters.

Table 4. Allelic variants of additional hypervariable VNTR loci that cannot be resolved with the CEQ automated sequencer*†

Locus	Maximum no. repeats suitable for automated analysis	Fragment size, bp	Proportion of strains with allelic variants beyond the automated system resolution, %
3232	15	830	4.1
1982	9	880	11.9
2163A	11	876	10.8
3336	11	875	21.1

*Beckman Coulter, Fullerton, CA, USA.
†VNTR, variable number tandem repeats.

controls that give stable comparable results in both agarose gel and capillary electrophoresis and have been used previously in a multilaboratory comparative study (*7*).

In all cases, identical data were produced for MIRU 26 and ETR-B irrespective of the DNA polymerase used. Amplification of VNTR 1982 and 3232 varied with different DNA polymerases, particularly when expected fragments were long.

The differing performances of polymerases for amplifying different loci can be explained by their varying properties. BIOTAQ polymerase is a basic Taq that can be used for a wide range of templates, whereas Diamond polymerase has been modified by a point mutation at the active site of the enzyme, enabling it to read through regions of secondary structure, microsatellites, and guanine cytosine–rich templates, such as those found in the *M. tuberculosis* genome. The QIAGEN polymerases are chemically modified polymerases with a high specificity similar to that of Diamond polymerase; thus they showed similar capabilities in amplifying VNTR 1982 and 3232. In addition, the buffer used with the QIAGEN polymerases is designed to increase the specificity of primer binding, making these polymerases suitable for dealing with complex genomic DNA.

Conditions that affect the denaturation of PCR products, and therefore their linearity before fragment sizing by electrophoresis, would be expected to influence apparent sizes of PCR fragments and copy number enumeration. We investigated the influence of DNA denaturation time and capillary separation temperature. As expected, we found that lowering the separation rate increased the discrimination of fragments >1,000 bp.

A marked difference was observed when the capillary temperature was decreased (method 3), which was independent of the polymerase used and locus investigated and demonstrated that separation conditions are critical for the correct interpretation of the VNTR typing results. In method 3, apparent fragment sizes were smaller and offset values were markedly larger, to the point that in some cases the calculated copy number was different from that expected.

Taking all the data together, we used BIOTAQ for amplifying MIRU and ETR loci, and Diamond polymerase for amplifying the extra 7 hypervariable VNTR loci, using the separation conditions detailed in method 1. We also demonstrated the reproducibility and stability of the extra 7 VNTR loci by comparing 22 MIRU-VNTR profiles from serial isolates. The resulting profiles of serial isolates from the same patients were identical, indicating that the conditions used for fragment amplification, detection, and analysis were ideal for typing of these loci and that these loci could be used for routine genotyping.

Clustering rates seen by using 15 MIRU-ETR loci far exceeded those previously reported when IS*6110* RFLP was used in a London population study (*22*,*23*). We concluded that 15-MIRU-ETR genotyping was insufficiently discriminative and was producing so-called false clustering. This view was supported by the spoligotyping results in which 38 (16%) of 235 isolates of 15 MIRU-ETR clusters contained isolates that belonged to ≥2 spoligo families (Table 3).

Applying all 22 loci gave the lowest clustering rate (22.2%) in MTBC strains obtained over 1 year from a single metropolitan setting (London), a rate almost identical to the proportion established in previous studies conducted in London in 1993 and 1995–1997 (*22*,*23*) and similar to previously reported rates in population-based studies in low- to-middle TB incidence settings where RFLP and PCR-based genotyping methods were used (*11*,*24–26*). These findings suggest, from the public health viewpoint, that TB transmission in London has remained stable over the past decade. Our study provides strong evidence that PCR-based methods, especially VNTR-MIRU, can replace IS*6110* RFLP typing for prospective analysis and that 12 MIRU (*27*), and 15 MIRU-ETR loci panels alone are insufficiently discriminating for evaluation of TB transmission.

The recently proposed VNTR panel (*3*,*5*,*7*,*11*) provides similar degrees of discrimination (comparable to that achieved by IS*6110* RFLP), although discrimination of individual VNTR loci is not equal for different MTBC genetic families (*13*). Inclusion of highly polymorphic VNTR loci effectively differentiates strains within highly conserved groups and is vital for prospective genotyping. Our study demonstrated that even in settings of low TB incidence and relatively low TB transmission rates, TB families, such as Beijing and CAS, remain more conserved than others, and hypervariable loci (e.g., VNTR 3232, 2163A, 4052) provide much higher discrimination than MIRU and ETR loci either alone or in combination.

Our current results agree with the preliminary results of our earlier studies about the applicability of hypervariable VNTR loci (VNTR 3232, VNTR 3336; VNTR 2163a,

and VNTR1982, in particular) and recent reports (*28*–*30*) demonstrating their effectiveness for discrimination among Beijing strains. This agreement suggests that these loci are discriminating and reproducible, especially where Beijing strains are dominant (e.g., China, Russia, Baltic countries) (*28*) and should be included in standardized VNTR panels. They can be used successfully at multiple laboratories with consistent results, provided the conditions for proposed reaction and PCR fragment separation are adhered to and specific DNA polymerases are used.

Acknowledgments

We thank the reference staff at the Health Protection Agency National Mycobacterium Reference Laboratory for providing the DNA extracts used in this study and the research staff for their assistance with the VNTR typing of all of the isolates.

This research was funded through the UK Department of Health grant "Genotyping of *Mycobacterium tuberculosis* in London."

Ms Velji is a PhD student at the UK Health Protection Agency Mycobacterium Reference Laboratory, Clinical TB and HIV Group, Barts and The London School of Medicine, Queen Mary, University of London, UK. Her research interests are molecular microbiology and respiratory infections, especially TB.

References

1. World Health Organization (WHO). Global tuberculosis control: surveillance, planning, financing. WHO report 2007. Geneva: The Organization; 2007 [cited 2009 Sep 7]. Available from www.who.int/tb/publications/global_report/2007/en
2. Anderson SR, Maguire H, Carless J. Tuberculosis in London: a decade and a half of no decline [corrected]. Thorax. 2007;62:162–7. DOI: 10.1136/thx.2006.058313
3. Gopaul KK, Brown TJ, Gibson AL, Yates MD, Drobniewski FA. Progression toward an improved DNA amplification-based typing technique in the study of *Mycobacterium tuberculosis* epidemiology. J Clin Microbiol. 2006;44:2492–8. DOI: 10.1128/JCM.01428-05
4. Kam KM, Yip CW, Tse LW, Leung KL, Wong KL, Ko WM, et al. Optimization of variable number tandem repeat typing set for differentiating *Mycobacterium tuberculosis* strains in the Beijing family. FEMS Microbiol Lett. 2006;256:258–65. DOI: 10.1111/j.1574-6968.2006.00126.x
5. Nikolayevskyy V, Gopaul K, Balabanova Y, Brown T, Fedorin I, Drobniewski F. Differentiation of tuberculosis strains in a population with mainly Beijing-family strains. Emerg Infect Dis. 2006;12:1406–13.
6. Supply P, Lesjean S, Savine E, Kremer K, van Soolingen D, Locht C. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. J Clin Microbiol. 2001;39:3563–71. DOI: 10.1128/JCM.39.10.3563-3571.2001
7. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rusch-Gerdes S, Willery E, et al. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. J Clin Microbiol. 2006;44:4498–510. DOI: 10.1128/JCM.01392-06
8. Roring S, Scott AN, Glyn Hewinson R, Neill SD, Skuce RA. Evaluation of variable number tandem repeat (VNTR) loci in molecular typing of *Mycobacterium bovis* isolates from Ireland. Vet Microbiol. 2004;101:65–73. DOI: 10.1016/j.vetmic.2004.02.013
9. Kwara A, Schiro R, Cowan LS, Hyslop NE, Wiser MF, Roahen Harrison S, et al. Evaluation of the epidemiologic utility of secondary typing methods for differentiation of *Mycobacterium tuberculosis* isolates. J Clin Microbiol. 2003;41:2683–5. DOI: 10.1128/JCM.41.6.2683-2685.2003
10. van Deutekom H, Supply P, de Haas PE, Willery E, Hoijng SP, Locht C, et al. Molecular typing of *Mycobacterium tuberculosis* by mycobacterial interspersed repetitive unit–variable-number tandem repeat analysis, a more accurate method for identifying epidemiological links between patients with tuberculosis. J Clin Microbiol. 2005;43:4473–9. DOI: 10.1128/JCM.43.9.4473-4479.2005
11. Oelemann MC, Diel R, Vatin V, Haas W, Rusch-Gerdes S, Locht C, et al. Assessment of an optimized mycobacterial interspersed repetitive-unit–variable-number tandem-repeat typing system combined with spoligotyping for population-based molecular epidemiology studies of tuberculosis. J Clin Microbiol. 2007;45:691–7. DOI: 10.1128/JCM.01393-06
12. Yokoyama E, Kishida K, Uchimura M, Ichinohe S. Improved differentiation of *Mycobacterium tuberculosis* strains, including many Beijing genotype strains, using a new combination of variable number of tandem repeats loci. Infect Genet Evol. 2007;7:499–508. DOI: 10.1016/j.meegid.2007.02.006
13. Arnold C, Thorne N, Underwood A, Baster K, Gharbia S. Evolution of short sequence repeats in *Mycobacterium tuberculosis*. FEMS Microbiol Lett. 2006;256:340–6. DOI: 10.1111/j.1574-6968.2006.00142.x
14. Roring S, Scott A, Brittain D, Walker I, Hewinson G, Neill S, et al. Development of variable-number tandem repeat typing of *Mycobacterium bovis*: comparison of results with those obtained by using existing exact tandem repeats and spoligotyping. J Clin Microbiol. 2002;40:2126–33. DOI: 10.1128/JCM.40.6.2126-2133.2002
15. Kremer K, Arnold C, Cataldi A, Gutierrez MC, Haas WH, Panaiotov S, et al. Discriminatory power and reproducibility of novel DNA typing methods for *Mycobacterium tuberculosis* complex strains. J Clin Microbiol. 2005;43:5628–38. DOI: 10.1128/JCM.43.11.5628-5638.2005
16. Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, Al-Hajoj SA, et al. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiol. 2006;6:23. DOI: 10.1186/1471-2180-6-23
17. Vitol I, Driscoll J, Kreiswirth B, Kurepina N, Bennett KP. Identifying *Mycobacterium tuberculosis* complex strain families using spoligotypes. Infect Genet Evol. 2006;6:491–504. DOI: 10.1016/j.meegid.2006.03.003
18. Hunter PR, Gaston MA. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol. 1988;26:2465–6.
19. Sola C, Filliol I, Legrand E, Lesjean S, Locht C, Supply P, et al. Genotyping of the *Mycobacterium tuberculosis* complex using MIRUs: association with VNTR and spoligotyping for molecular epidemiology and evolutionary genetics. Infect Genet Evol. 2003;3:125–33. DOI: 10.1016/S1567-1348(03)00011-X
20. Warren RM, Victor TC, Streicher EM, Richardson M, van der Spuy GD, Johnson R, et al. Clonal expansion of a globally disseminated lineage of *Mycobacterium tuberculosis* with low IS*6110* copy numbers. J Clin Microbiol. 2004;42:5774–82. DOI: 10.1128/JCM.42.12.5774-5782.2004
21. Frothingham R, Meeker-O'Connell WA. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. Microbiology. 1998;144:1189–96. DOI: 10.1099/00221287-144-5-1189

22. Maguire H, Dale JW, McHugh TD, Butcher PD, Gillespie SH, Costetsos A, et al. Molecular epidemiology of tuberculosis in London 1995–7 showing low rate of active transmission. Thorax. 2002;57:617–22. DOI: 10.1136/thorax.57.7.617
23. Hayward AC, Goss S, Drobniewski F, Saunders N, Shaw RJ, Goyal M, et al. The molecular epidemiology of tuberculosis in inner London. Epidemiol Infect. 2002;128:175–84. DOI: 10.1017/S0950268801006690
24. Small PM, Hopewell PC, Singh SP, Paz A, Parsonnet J, Ruston DC, et al. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. N Engl J Med. 1994;330:1703–9. DOI: 10.1056/NEJM199406163302402
25. Allix-Beguec C, Fauville-Dufaux M, Supply P. Three-year population-based evaluation of standardized mycobacterial interspersed repetitive-unit–variable-number tandem-repeat typing of *Mycobacterium tuberculosis*. J Clin Microbiol. 2008;46:1398–406. DOI: 10.1128/JCM.02089-07
26. Durmaz R, Zozio T, Gunal S, Allix C, Fauville-Dufaux M, Rastogi N. Population-based molecular epidemiological study of tuberculosis in Malatya, Turkey. J Clin Microbiol. 2007;45:4027–35. DOI: 10.1128/JCM.01308-07
27. Supply P, Lesjean S, Savine E, Kremer K, van Soolingen D, Locht C. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. J Clin Microbiol. 2001;39:3563–71. DOI: 10.1128/JCM.39.10.3563-3571.2001
28. Drobniewski F, Balabanova Y, Nikolayevsky V, Ruddy M, Kuznetzov S, Zakharova S, et al. Drug-resistant tuberculosis, clinical virulence, and the dominance of the Beijing strain family in Russia. JAMA. 2005;293:2726–31. DOI: 10.1001/jama.293.22.2726
29. Wada T, Maeda S, Hase A, Kobayashi K. Evaluation of variable numbers of tandem repeat as molecular epidemiological markers of *Mycobacterium tuberculosis* in Japan. J Med Microbiol. 2007;56:1052–7. DOI: 10.1099/jmm.0.46990-0
30. Iwamoto T, Yoshida S, Suzuki K, Tomita M, Fujiyama R, Tanaka N, et al. Hypervariable loci that enhance the discriminatory ability of newly proposed 15-loci and 24-loci variable-number tandem repeat typing method on *Mycobacterium tuberculosis* strains predominated by the Beijing family. FEMS Microbiol Lett. 2007;270:67–74. DOI: 10.1111/j.1574-6968.2007.00658.x

Address for correspondence: Francis Drobniewski, UK Health Protection Agency Mycobacterium Reference Unit, Clinical TB and HIV Research Group, Queen Mary, University of London, 2 Newark St, E1 2AT, London, UK; email: f.drobniewski@qmul.ac.uk