The Molecular Epidemiology of *Mycobacterium tuberculosis*: Host and Bacterial factors perpetuating the epidemic.

Madeleine Hanekom

Dissertation presented for the degree of Doctor of Philosophy
at
Stellenbosch University

Promotor: Prof RM Warren
Co-promotor: Prof PD van Helden
Co-promotor: Prof TC Victor

August 2009

Declaration

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or in part, submitted it at any university for a degree.

_____
Signature

_____
Name in full

_____/_____/_____
Date

Overview

This study describes the molecular epidemiology of *Mycobacterium tuberculosis* strains with the Beijing genotype. This genotype has received clinical prominence due to its global distribution and the hypothesis that these strains have acquired the ability to evade the protective effect of BCG vaccination, spread more readily, acquire drug resistance and cause severe forms of disease. Molecular biological techniques were used in a series of studies to elucidate the genetic evolutionary mechanisms underlying the success of this genotype in Cape Town, South Africa.

Using a collection of 40 different markers it was possible for the first time to construct a phylogenetic history of Beijing genotype strains. This phylogeny was characterized by the consecutive evolution of 7 sublineages. Analysis of epidemiological data in relation to these sublineages showed an association between more recently evolved Beijing strains and an increased ability to transmit and cause disease. From these findings it was hypothesized that the pathogenic characteristics of the Beijing genotype were not conserved but rather that strains representative of the different sublineages had evolved unique properties. In order to determine whether these so-called unique properties were associated with either the host population or the genetic background of strains from sublineage 7, a meta-analysis of published Mycobacterial Interspersed Repetitive-Unit (MIRU) typing data (East Asia) was compared with MIRU typing data from the South African strains in the context of their phylogenetic histories. This study showed that Beijing genotype strains in South Africa originated in East Asia following their introduction during the early 18^{th} century. A significant association was observed between the frequency of occurrence of strains from defined Beijing sublineages and the human population from whom they were cultured ($p < 0.0001$). Based on these findings it was proposed that either the host population (South African) had selected for a particular Beijing sublineage (i.e. sublineage 7) or that strains from that sublineage had adapted to be more successful in the South African population.

In a subsequent study, using the methodology developed in the above studies, it was shown that strains from the ancestrally positioned lineage (termed “atypical” Beijing

genotype) were over-represented in drug resistant isolates in the Eastern Cape region. This contradicts current dogma which suggests that "atypical" Beijing genotype strains are attenuated in their ability to transmit. However, this phenomenon may be ameliorated in immune-compromised patients as review of the clinical records showed that transmission was associated with HIV co-infection. These findings highlight the need to improve tuberculosis control in vulnerable populations as strains which would normally not contribute significantly to the epidemic now become a cause for concern especially if they are associated with drug resistance.

To improve our understanding of the evolution of the Beijing genotype, the genomic stability of an additional 27 polymorphic markers were analysed. These markers have recently been proposed as the new standard in molecular epidemiological studies and were based on MIRU-Variable Number Tandem-Repeats (VNTR) sequences. Superimposition of the MIRU-VNTR data onto the phylogenetic tree showed excellent concordance thereby demonstrating that these alleles were largely stable over time. It is currently not known how the alleles that do change could influence pathogenicity. The results of this study also demonstrated discordance between strains defined by IS*6110* DNA fingerprinting and those defined by MIRU-VNTR typing thereby demonstrating that these markers evolve independently and at different rates. Furthermore, the MIRU-VNTR typing method was unable to predict transmission of drug resistant strains which contradict previous reports from low incidence settings. This has significant implications for the use of this typing method in high incidence settings.

Using an improved PCR-based method it was possible for the first time, to identify the 5 most prominent phylogenetic lineages in primary cultures of adult tuberculosis patients resident in a high HIV/TB co-infection setting. The results of this study showed that 15% of the study population was infected with two or more strains and Beijing genotype strains were over-represented in these mixed infections. Furthermore, drug susceptibility tests showed that one patient was co-infected with both a drug sensitive and a drug resistant strain. Since mixed infections have been implicated in treatment failure, these findings demonstrate the epidemiological importance of detecting mixed infections in vulnerable populations. This PCR-based

method was further applied to cultures of paediatric tuberculosis patients to classify strains which spoligotyping was unable to define. The result of this study showed three mixed infections which otherwise would have been missed.

In order to determine whether clinical disease presentation of patients infected with strains of the Beijing genotype were different from that of patients infected with non-Beijing genotype strains, clinical and demographic data of these two groups were analysed. This study showed that patients infected with strains of the Beijing genotype were highly infectious as defined by the increased bacterial load in sputum specimens. However, this finding could not be validated by lung pathology according to chest radiographs of infected patients.

Oorsig

Hierdie studie beskryf die molekulêre epidemiologie van *Mycobacterium tuberculosis* rasse met die Beijing genotipe. Hierdie genotipe is van groot kliniese belang weens hul globale verspreiding en die hipotese dat hierdie rasse die vermoë ontwikkel het om die beskermende effek van BCG vaksinasie te vermy, om meer geredelik te versprei, middelweerstandigheid te ontwikkel en erger vorms van siekte te veroorsaak. Molekulêre biologiese tegnieke is gebruik in 'n reeks studies om die genetiese evolusionêre meganismes onderliggend tot die sukses van hierdie genotipe in Kaapstad, Suid-Afrika te verklaar.

Deur 'n versameling van 40 verskillende merkers te gebruik, was dit moontlik om vir die eerste keer 'n filogenetiese stamboom van die Beijing ras genotipe te skep. Hierdie filogenie word gekenmerk deur die opeenvolgende evolusie van 7 ras sublyne. Met die analise van epidemiologiese data in verhouding tot hierdie ras sublyne, is 'n assosiasie tussen die mees onlangs ontwikkelde Beijing rasse en die verhoogde vermoë om te versprei en siekte te veroorsaak, getoon. Vanweë hierdie bevindinge, is 'n hipotese daargestel dat die patogeniese kenmerke van die Beijing genotipe nie in alle raslyne voorkom nie, maar eerder dat verteenwoordigende rasse van die verskillende sublyne unieke eienskappe deur evolusie ontwikkel het. 'n Meta-analise van gepubliseerde MIRU tipering data van Oos-Asië is vergelyk met MIRU tipering data van Suid-Afrikaanse rasse in die konteks van hul filogenetiese geskiedenis om te bepaal watter van hierdie sogenoemde unieke eienskappe geassosieer is met die gasheerpopulasie en watter eienskappe geassosieer is met die genetiese agtergrond van die sublyn 7 rasse. Hierdie studie het getoon dat die Beijing ras genotipe van Suid-Afrika hul oorsprong gekry het van Oos-Asië en vir die eerste keer waargeneem is in die vroeë 18de eeu. 'n Betekenisvolle assosiasie is waargeneem tussen die frekwensie waarteen die rasse van 'n bepaalde Beijing sublyn voorkom en die menslike populasie van wie hulle geïsoleer is ($p < 0.0001$). Gebaseer op hierdie bevindinge is dit voorgestel dat die menslike populasie (Suid-Afrikaners) vir 'n spesifieke Beijing sublyn geselekteer het (bv. Sublyn 7) of dat rasse van hierdie sublyn aangepas het om meer suksesvol te wees in die Suid-Afrikaanse populasie.

In ‘n daaropvolgende studie is, deur gebruik te maak van die metodiek wat ontwikkel is vir die bogenoemde studies, getoon dat die voorouerlike sublyn (bekend as die“atipiese” Beijing genotipe) die mees verteenwoordigende sublyn was onder middelweerstandige isolate van die Oos-Kaap gebied. Dit is teenstrydig met die bestaande dogma wat bepaal dat die “atipiese” Beijing genotipe rasse hulle vermoë om te versprei verloor het. Hierdie verskynsel kan egter versterk word in immuun inkompetente pasiënte aangesien hersiening van die kliniese rekords aangedui het dat verspreiding geassosieer was met HIV ko-infeksie. Hierdie bevindinge bring die behoefte om TB beheer in vatbare populasies te verbeter, na vore, omrede rasse wat gewoonlik `n onbetekenisvolle bydrae tot die epidemie lewer, nou ‘n rede vir kommer is veral as hulle met middelweerstandigheid geassosieer is.

Om ons insig rakende die evolusie van die Beijing genotipe te verbeter, is die genomiese stabiliteit van ‘n addisionele 27 polimorfiese merkers geanaliseer. Daar is onlangs voorgestel dat hierdie merkers, wat gebaseer is op MIRU-VNTR volgordes, die nuwe standaard vir molekulêre studies is. Die MIRU-VNTR data is op die filogenetiese boom geplaas en het uitstekende ooreenstemming getoon wat die allele se stabiliteit oor tyd gedemonstreer het. Dit is tans nie duidelik hoe van die allele wat wel verander, die patogenisiteit beïnvloed nie. Die resultate van die studie wys ook onenigheid tussen rasse wat deur IS*6110* DNA tipering gedefinieer is en dié wat deur MIRU-VNTR tipering gedefinieer is. Dit impliseer dus dat die evolusie van merkers onafhanklik van mekaar plaasvind en teen verskillende tempos. Verder was die MIRU-VNTR tipering metode nie in staat om verspreiding van middelweerstandige rasse te voorspel nie, wat teenstrydig is met vorige verslae waar lae insidensie omgewings bestudeer is. Dit het noemenswaardige implikasies vir die gebruik van hierdie tipering metode in hoë insidensie omgewings.

‘n Verbeterde PKR-gebaseerde metode is vir die eerste keer gebruik om die 5 mees prominente filogenetiese sublyne in primêre kulture van volwasse tuberkulose pasiënte van ‘n hoë MIV/TB ko-infeksie omgewing, te identifiseer. Die resultate van hierdie studie het gewys dat 15% van die studiepopulasie geïnfekteer is met twee of meer rasse en dat die Beijing genotipe ras die meeste voorgekom het in gemengde infeksies. Verder het middelweerstandige toetse gewys dat een pasiënt geïnfekteer

was met beide ‘n middelsensitiewe en ‘n middelweerstandige ras. Gemengde infeksies is al vantevore gekoppel aan onsuksesvolle behandeling en dus demonstreer hierdie bevindinge die epidemiologiese belang van die opsporing van gemengde infeksies in vatbare populasies. Hierdie PKR-gebaseerde metode is verder gebruik om rasse wat voorkom in kulture van pediatriese pasiënte, wat spoligotipering nie kon klassifiseer nie, te klassifiseer. Die resultate het drie gemengde infeksies gewys wat sonder die PKR-gebaseerde metode, nie geïdentifiseer sou gewees het.

Om te bepaal of die kliniese beeld van pasiënte wat geïnfekteer is met rasse van die Beijing genotipe verskil van dié van pasiënte wat geïnfekteer is met rasse van die nie-Beijing genotipe, is die kliniese en demografiese data van die twee groepe pasiënte geanaliseer. Hierdie studie wys dat pasiënte wat geïnfekteer is met rasse van die Beijing genotipe hoogs aansteeklik is (gedefinieer op grond van hoë bakteriële lading in sputum monsters). Hierdie bevindinge kon egter nie met behulp van long patologie op borskas X-strale bevestig word nie.

Acknowledgements

I would like to express my sincere thanks and appreciation to the following people, without whom this thesis and the studies comprising it would not have been possible.

Firstly, I would like to thank my promotor Prof. Rob Warren without whom my achievements thus far would not have been possible. He has thought me all I know about science and he is my greatest mentor. His knowledge, ambition and dedication for his work have inspired me with my own work/research. I would like to thank him for his help, encouragement, motivation, guidance and exciting food recipes.

Secondly, I would like to thank my co-promotor Prof. Paul van Helden for giving me the opportunity to be a part of the Division of Molecular Biology and Human Genetics and to study for this degree. He has always shown interest in my scientific and clinical research.

Thirdly, I would like to thank my co-promotor Prof. Tommie Victor for his supervision and advice during the course of these studies.

Thanks also to all my fellow lab-workers that have contributed to my learning process. They have created a great scientific environment for me to work in and I appreciate their support, advice, many discussions and the opportunity to learn from their variety of work.

Thanks to all my fellow colleagues at Task Applied Science for their support, patience and understanding.

Thanks to Martin Kidd for statistical advice, all the nurses and analysts for the collection and processing of clinical data, for the Department of Health, City of Cape Town for providing facilities to conduct our studies in the epidemiological field sites and also importantly the residents of the epidemiological field sites for their co-operation and giving us the chance to improve TB healthcare.

To the various funders of these projects: Harry Crossley foundation, the European Commission 6th framework programme for research and technological development, TB in the 21st Century Consortium and the National Research Foundation.

Finally, I would like to thank my family and friends for their wonderful support and encouragement.

Table of Contents

Declaration		ii
Overview		iii - v
Oorsig		vi - viii
Acknowledgments		ix
Table of contents		x - xi
Preface		xii - xiv
Chapter 1	Phenotypic and Genotypic characterisation of the *Mycobacterium tuberculosis* Beijing genotype.	1 - 55
Chapter 2	A recently evolved sublineage of the *Mycobacterium tuberculosis* Beijing strain family is associated with an increased ability to spread and cause disease. *J. Clin. Microbiol. 2007; 45:1483-1490*	56 - 78
	Spread of a low-fitness drug-resistant *Mycobacterium tuberculosis* strain in a setting of high human immunodeficiency virus prevalence. *J. Clin. Microbiol. 2008; 46:1514-1516*	79 - 87
Chapter 3	Evidence that the spread of *Mycobacterium tuberculosis* strains with the Beijing genotype is human population dependent. *J. Clin. Microbiol. 2007; 45:2263-2266*	88 - 99
Chapter 4	Discordance between Mycobacterial Interspersed Repetitive-Unit–Variable-Number Tandem-Repeat Typing and IS*6110* Restriction Fragment Length Polymorphism Genotyping for analysis of *Mycobacterium tuberculosis* Beijing strains in a setting of high incidence of tuberculosis. *J. Clin. Microbiol. 2008; 46:3338-3345*	100 - 146

Chapter 5	Mixed *Mycobacterium tuberculosis* infection in patients resident in a high HIV prevalence setting. (*To be submitted*)	147 - 171
	Beijing and Haarlem genotypes are overrepresented among children with drug-resistant tuberculosis in the Western Cape Province of South Africa. *J. Clin. Microbiol. 2006; 44:3539-3543*	172 - 189
Chapter 6	Increased bacterial load is associated with strains of the *Mycobacterium tuberculosis* Beijing genotype. (*To be submitted*)	190 - 204
Conclusion		205 - 208
Candidate's Contributions		209 - 210
Additional Publication		211
Supplemental Tables		212 - 235

Preface

Prior to the advent of molecular epidemiological strain typing tools it was thought that *Mycobacterium tuberculosis* was a ubiquitous bacillus responsible for the global tuberculosis epidemic. This view has changed with the development of IS*6110* DNA fingerprinting which showed that the tuberculosis epidemic was composed of numerous different *Mycobacterium tuberculosis* strains. Analysis of molecular epidemiological data subsequently showed the over-representation of certain strains and that these strains were often associated with major outbreaks. This led to the hypothesis that these strains have evolved unique properties of hyper-transmissibility and increased virulence which could manifest with unique clinical features. Beijing strains represent such an aggressively emerging genotype. This genotype has been extensively studied and the research findings are reviewed in chapter 1. Subsequent chapters describe the phenotypic and genotypic characteristics of the Beijing genotype strains circulating in Cape Town, South Africa, using molecular genotyping methods in association with clinical and demographic data.

In chapter 2, 40 genetic loci were analysed in a collection of Beijing genotype strains in order to reconstruct their evolutionary history and to determine whether pathogenic characteristics were associated with defined sublineages. This study described 7 independently evolving sublineages of which the most recently evolved sublineage was associated with increased pathogenicity. Using the methodology developed it was possible to show a different distribution of strains among the different sublineages in a different host population. Rarely observed strains, known not to transmit, were over-represented compared to frequently observed strains. This was explained by different host immune factors between the different host populations.

In chapter 3 the Mycobacterial Interspersed Repetitive-Unit (MIRU) typing method was used to show that the frequency of occurrence of strains from a defined Beijing sublineage was associated with the human host population from whom they were isolated. This suggested that either, 1) strains of a defined sublineage have evolved a higher level of pathogenicity in response to the immunity presented by a novel host

population or 2) that a certain host population was more vulnerable to strains from that sublineage.

Chapter 4 describes the evolutionary characteristics of a further 27 MIRU-Variable-Number Tandem-Repeats (VNTR) loci in strains of the Beijing genotype. This study showed that these loci were evolving at a rate slower than IS*6110* and therefore were less informative as molecular epidemiological markers. However, we note that the evolution of these MIRU-VNTR loci allowed for the accurate grouping of Beijing genotype strains into their respective phylogenetic lineages.

In chapter 5, clinical isolates were subjected to a PCR-based method to classify them into one of the 5 most prominent phylogenetic lineages of *Mycobacterium tuberculosis*. This methodology allowed for the quantification of the extent of mixed infections and thereby showed that strains of the Beijing genotype were more frequently associated with mixed infections than the other genotypes tested in this study. Again, this supports the notion that strains of the Beijing genotype have evolved unique properties which either allowed superinfection (vulnerability of a distinct host population to a defined Beijing sublineage) or strains from a defined Beijing sublineage have selected for a distinct host population. This PCR-based genotyping method was instrumental in grouping strains cultured from paediatric patients, thereby allowing the demonstration that the Beijing and Haarlem genotypes were associated with drug resistance.

In chapter 6, clinical and demographic data of tuberculosis patients infected with Beijing and non-Beijing genotype strains respectively were analysed to determine whether differences in clinical presentation in relationship to Beijing and non-Beijing genotype strains circulating in Cape Town, South Africa existed. In this study, patients with positive-smear disease were more often infected with strains of the Beijing genotype.

This series of studies have substantially enhanced our understanding of the influence of genetic strain variation on disease outcome. These findings could help future drug,

vaccine and diagnostic development in an attempt to limit tuberculosis transmission and the perpetuation of the TB epidemic.

Chapter 1

Phenotypic and Genotypic characterisation of the *Mycobacterium tuberculosis* Beijing genotype

Hanekom, M., Victor, T. C., van Helden, P. D., and Warren R. M.

Table of contents

1.1 The tuberculosis epidemic	3
1.2 Characteristics of the tubercle bacillus	3 - 4
1.3 Using molecular genotyping methods to identify strains	4 - 7
1.4 The genetic diversity among strains	7 - 8
1.5 Distribution and dominance of strain genotypes	8 - 12
1.6 Beijing strain genotype	12 - 13
2.1 Epidemiology of the Beijing strain genotype	14 - 16
2.2 *Mycobacterium bovis* bacilli Calmette-Guérin (BCG) vaccine	16 - 19
2.3 Drug resistance	19 - 21
2.4 Virulence	22 - 32
2.5 Clinical presentation	32 - 34
3.1 Conclusion	34 - 35
Reference List	35 - 54

1.1 The tuberculosis epidemic

In 1882, Robert Koch announced the discovery that *Mycobacterium tuberculosis (M. tuberculosis)* is the principal causative agent of human tuberculosis (TB). *M. tuberculosis* is a member of a closely related group of organisms known as the *M. tuberculosis* complex (MTBC); other species within the MTBC (e.g. *M. microtti*, *M. caprae* and *M. pinnipedi*) primarily known to cause disease in animals are also known to cause disease in humans. After Koch's discovery, research was directed towards the development of anti-TB drugs and vaccines to eliminate TB. However, despite any advances, 126 years later more people die from TB than ever before and in 1993 the World Health Organisation (WHO) declared TB a public health emergency.

It is important to note that a third of the world's population is estimated to be infected with *M. tuberculosis* and up to 9 million TB cases are diagnosed annually, resulting in over three million deaths every year (1). This makes TB the infectious disease with the highest adult mortality rate. TB accounts for more than a quarter of all preventable adult deaths in developing countries and for a third of HIV/AIDS related deaths. Individuals with undiagnosed or untreated TB are thought to infect 10 to 15 of their contacts each year of which 5 to 10 percent will develop disease in the following 2 years, thereby perpetuating the epidemic (2). Factors that probably influence the worldwide prevalence are the escalating numbers of HIV/AIDS cases, emergence of drug-resistant TB (especially multidrug-resistance (MDR) and extensively drug-resistance (XDR)), the increase in population mobility and failure to implement effective TB control. In order to combat the global TB epidemic, the WHO has been focusing on a strategy of political commitment, case detection, directly observed treatment, an uninterrupted drug supply and standardized reporting (so called Direct Observed Short-Course Therapy System (DOTS)) which is currently implemented in 182 countries worldwide (3).

1.2 Characteristics of the tubercle bacillus

The tubercle bacillus is a highly successful pathogen. It has the ability to persist in a dormant state within the host for extended periods, despite the presence of host

immune cells specialised to kill intracellular bacteria (termed latent infection), to manifest disease in a variety of ways (duration of disease, severity and anatomic distribution (from self-limiting pulmonary infection to extrapulmonary infection and disseminated disease)) and to manifest disease in different human host populations (some geographical areas are more affected than others). However, despite ongoing research, the mechanisms governing pathogenicity and the factors influencing the degree of disease variability remain largely unknown. Several host factors (i.e. malnutrition, alcoholism, homelessness, overcrowding, genetic predisposition (i.e. polymorphism in natural resistance-associated macrophage proteins (NRAMPs), Toll-like receptors and vitamin D receptor) and occupational lung disease) and environmental factors (i.e. exposure to environmental mycobacteria) have been suggested. However, there is substantial evidence to suggest that bacterial factors i.e. genetic variation in the mycobacterium, also contribute to the variability of disease presentation, frequency of transmission and treatment outcome (4 567). For example, a more recent animal study, which infected guinea pigs with both laboratory strains (H37Rv and Erdman) and clinical isolates, found that guinea pigs infected with laboratory strains of *M. tuberculosis,* showed less lung parenchymal lesions than guinea pigs infected with various clinical isolates (8). Clinical isolates showed rapid and progressive disease with extensive pulmonary and extrapulmonary lesion necrosis including lymph node lesion cavitations, a rare event in animals infected with laboratory strains of *M. tuberculosis* (8).

1.3 Using molecular genotyping methods to identify strains

Members of the MTBC are defined by unique genetic characteristics, although they show highly conserved genomes with 99.9% of the DNA sequence being shared. The use of molecular genotyping methods has made it possible to identify and classify members of the MTBC according to these unique genetic characteristics.

In 1992, DNA repeat sequences (Polymorphic GC-rich sequence (PGRS)) marked the advent of molecular epidemiology, since this genetic marker was able to discriminate among TB isolates causing disease in different patients (9). Thereafter, the method of IS*6110* Restriction Fragment Length Polymorphism (RFLP) was developed (10).

This method is based on the detection of the insertion sequence IS*6110*; an insertion element containing 1,355-base pairs (bp) integrated into different sites in the *M. tuberculosis* genome. These "jumping" elements have been shown to be involved in gene disruption or gene regulation and, along with their positional polymorphism, can either be deleted from or replicated in different positions in the genome leading to highly variable IS*6110* copy numbers (ranging between 0 and 25 copies). The internationally standardised method of IS*6110* RFLP analysis requires digestion of chromosomal DNA with the restriction enzyme *Pvu*II, separation of the restricted DNA fragments by electrophoresis in an agarose gel, Southern blotting, hybridisation with a labelled DNA probe matching the sequence of IS*6110* and detection of the labelled probes by autoradiography. The resulting IS*6110* RFLP (banding pattern) is characteristic of a specific *M. tuberculosis* strain. Studies have shown IS*6110* to evolve fast enough to differentiate distant epidemiological events and yet slow enough to accurately define recent epidemiological events (11, 12). These properties make IS*6110* RFLP a useful marker for strain genotyping. However, this method has a number of limitations that hinders its use: it requires large amounts of good quality DNA which requires time consuming culturing, it cannot accurately differentiate low copy number strains, it is labour intensive, expensive and inter-laboratory comparisons are complicated due to the complexity of standardising methodology and interpretation of the fingerprints obtained. Nevertheless, IS*6110* RFLP fingerprinting has provided a foundation for much of our current understanding of the transmission dynamics of *M. tuberculosis* in different study settings (13), as well as the ability to distinguish between endogenous reactivation and exogenous re-infection in recurrent TB cases (14), identification of outbreaks, confirmation of laboratory cross-contamination, to identify where transmission occurs (i.e. household) and helped to define the global population structure of *M. tuberculosis* strains in different geographical settings.

In order to circumvent the limitations associated with IS*6110* RFLP, two PCR-based molecular typing methods, have been developed, namely spoligotyping (15) and mycobacterial interspersed repetitive-unit–variable-number tandem-repeat (MIRU-VNTR) typing (16, 17). Spoligotyping is a PCR-based method that determines the structure of the direct repeat (DR) locus in the genome of *M. tuberculosis*. The DR

locus contains direct repeats of 36-bp which are interspersed with unique non-repetitive spacer sequences varying from 35 to 41 bps in size; together termed the direct variable repeat (DVR) (18). Variability in the DR locus most likely occurs through homologous recombination between neighbouring or distant direct repeats, IS-mediated transposition and single nucleotide polymorphisms (SNPs). The presence or absence of DVRs is determined by amplification of the DVRs and subsequent hybridisation to a membrane containing 43 different DVR sequences. *M. tuberculosis* strains are defined by the absence or presence of these 43 DVRs. Spoligotyping is faster and simpler than IS*6110* RFLP fingerprinting, requires small amounts of DNA and can be performed on crude DNA extracted from clinical samples, thus avoiding the time intensive work associated with the slow growth of these bacteria and is able to discriminate among isolates with fewer than 5 IS*6110* copy numbers. Nevertheless, this method requires a special apparatus and reagents for hybridisation and signal detection, has a lower overall discriminatory power than IS*6110* RFLP and is incapable of accurately defining transmission events leading to an over-estimation of transmission events. A further limitation of spoligotyping is the observation that the DR locus may undergo convergent evolution, leading to epidemiologically unrelated strains sharing identical spoligotype patterns. However, despite these limitations, spoligotyping provides an overview of the population structure of *M. tuberculosis* strains in different geographical settings, since strains which have evolved from a common ancestor can be grouped into evolutionary lineages according to unique spoligotype signatures (19). An international database of spoligotypes of *M. tuberculosis* has been created, termed SpolD. This database contains spoligotype data from numerous settings throughout the world and thereby serves as an extremely important reference that consistently documents the global population structure (20, 21). MIRU-VNTR typing is based on the characterisation of different tandem DNA repeats scattered in various intergenic regions (loci) in the mycobacterial genome. These repeat sequences can either expand or contract due to changes in the number of repeats at each genomic locus. This is then converted into a numerical code which can be used to define a strain. Since many independent loci are assessed, it has been suggested that this method is more appropriate for phylogenetic analysis, however, the resulting trees are not always concordant with SNP-based phylogenetic studies (22). Unlike IS*6110* RFLP, this method can be automated

making this genotyping technique less labour intensive than IS*6110* RFLP with more reproducible results and straightforward inter-laboratory comparison of data. Various studies using MIRU-VNTR-based typing have shown that this technique has a discriminatory power similar to that of IS*6110* RFLP fingerprinting which could further be enhanced when combined with spoligotyping.

1.4 The genetic diversity among strains

Initial phylogenetic analysis of *M. tuberculosis* grouped strains into 3 principal genetic groups (PGG) according to 2 non-synonomous single nucleotide polymorphisms (nsSNPS): codon 463 of the *kat*G gene and codon 95 of the *gyr*A gene (23, 24). Epidemiological analysis of strains within these 3 PGGs suggested that PGG1 strains were more pathogenic as measured by their frequency of transmission. These PGGs were preserved when additional nsSNPs were used to further refine the phylogeny of *M. tuberculosis*. Filliol and colleagues suggested that *M. tuberculosis* strains could be divided into 6 phylogenetic distinct groups, of which 2 groups could be further subdivided into 5 subgroups, resulting in a combination of 9 SNP cluster groups (SCGs) (22). In contrast, Gutacker and colleagues identified 8 major groups of *M. tuberculosis* without subgroups, as well as a three-branch tree rather than a four-branch tree as obtained by Filliol *et al* (25). Both studies showed all of the *M. tuberculosis* SCGs to be distinct and deeply branched. When traditional genotyping methods (spoligotyping and MIRU typing) were superimposed on these phylogenies there was a large degree of concordance with spoligotyping with the exception of some families found in PGG2. MIRU typing, however, appeared to be the least successful method to assign isolates to their respective SCGs (22). Accordingly, the traditional genotyping methods cannot be exploited to study the global phylogeny of *M. tuberculosis*. Tsolaki and Hirsch proposed the use of long sequence polymorphisms (LSPs) to define the phylogeny of *M. tuberculosis* based on the understanding that homologous recombination between strains was not observed in *M. tuberculosis* and thus the analysis of genomic deletions would provide a robust phylogeny given the absence of parallel evolution (homoplasy) (26). Using this technique, Gagneux *et al* (2006) defined the global phylogeny of *M. tuberculosis* in which strains were grouped into 6 six major lineages: Indo-Oceanic, East Asian, East African-Indian, Euro-American, West African I and II (27). The authors also noted

that the major lineages displayed a strong degree of geographic confinement indicative of host-pathogen compatibility. This notion was further supported by Reed *et al* (2009) who demonstrated a relationship between bacterial lineage and patient origin (28). However, this concept was at first controversial and limited in (geographical) scope. We therefore explored this in order to validate the concept in an entirely independent population. Our study conducted in South Africa used genetic information to reconstruct the evolutionary history of strains of the Beijing genotype found in this study setting; we found that the Beijing strain genotype could be divided into 7 sublineages (29). In a follow-up study we compared the frequency of occurrence of different bacterial sublineages circulating in Cape Town, South Africa and East-Asia. We found a significant association between the frequency of occurrence of strains of a defined bacterial sublineage and the human population they were isolated from (30). The success (frequency and/or association with more severe disease presentation) of certain strain genotypes in defined host populations has prompted the hypothesis that these genotypes have evolved unique properties enabling them to spread and cause disease more readily. However, the relationship between genomic evolution and changes in pathogenicity remains to be described.

1.5 Distribution and dominance of strain genotypes

The fourth international spoligotyping database, SpolDB4, was constructed to determine the global population structure, transmission and evolution of the MTBC (31). This database represents 39 295 isolates collected from patients resident in 141 countries. Isolates classified according to spoligotyping, were grouped into clades or lineages based on defined spoligotype signatures which were recognised using specialised computerised software. In total, strains representative of 62 evolutionary lineages have been identified and it is evident that these lineages are not evenly distributed. *M. tuberculosis* strains from the Beijing evolutionary lineage were found to be present in the largest number of countries (13% of global isolates) (Table 1), suggesting that this evolutionary lineage may have evolved unique properties which has allowed global spread (clonal expansion). This differs from many other *M. tuberculosis* lineages, which are restricted to a defined geographical location (usually

the country of origin i.e. East-African Indian (EAI), X-family, Central Asian strain 1 (CAS1) -Delhi) (Table 2).

Table 1: Total number of clinical isolates collected from different countries with their respective number of Beijing genotype strains isolated according to SpolDB4 (31)

Country of isolation of clinical isolates	Total number of clinical isolates collected	Total number of Beijing genotype strains isolated (%)
AFRICA	1 491	137 (9.2)
Guinea-Bissau	217	1
Kenya	71	6
Libya	54	3
Morocco	127	1
Mozambique	28	2
Malawi	122	1
Senegal	69	8
Sierra Leone	3	4
South Africa	564	106
Zimbabwe	247	5
CENTRAL AMERICA	1 232	34 (2.8)
Cuba	239	27
Guadeloupe	232	1
Haiti	375	3
Mexico	386	3
EUROPE	11 393	813 (7.1)
Austria	1 456	25
Belgium	517	14
Czech Republic	393	14
Germany	574	19
Denmark	281	18
Estonia	115	50
Finland	347	4
France	2 265	40
United Kingdom	1523	11
Italy	934	26
Latvia	138	76
Netherlands	973	25
Poland	227	9
Portugal	336	4
Russia	986	446

Sweden	328	32
FAR-EAST ASIA	2 600	1 120 (43.1)
China	145	46
Indonesia	344	153
Japan	138	101
Korea	4	2
Myanmar	20	11
Mongolia	19	10
Malaysia	598	250
Philippines	237	12
Singapore	4	2
Thailand	302	135
Vietnam	789	398
MIDDLE-EAST-CENTRAL ASIA	2 422	441 (18.2)
Armenia	119	51
Azerbaijan	71	53
Bangladesh	676	123
Georgia	272	65
India	483	35
Iran	110	10
Israel	15	15
Kazakhstan	55	38
Madagascar	395	22
Mauritius	21	10
Pakistan	90	5
Reunion Island	16	8
Saudi Arabia	99	6
NORTH AMERICA	9 149	1 380 (15.1)
Canada	266	4
USA unspecified	1 690	
USA, Alabama	3	1 376
USA, New York	5 948	
USA, Texas	1 242	
OCEANIA	187	46 (24.6)
Australia	36	42
New Zealand	151	4
SOUTH AMERICA	2 471	16 (0.65)

Argentina	1 150	3
Brazil	842	2
French Guiana	375	6
Peru	96	4
Suriname	8	1

Table 2: Distribution of strains other than the *M. tuberculosis* Beijing genotype in different countries (31)

Country of isolation of clinical isolates	Haarlem	LAM	T family	EAI	CAS1 - Delhi	X family
Africa	Most frequent	Most frequent	Most frequent	-	Less frequent	-
Central America	25%	Most frequent	Most frequent	-	Less frequent	11.9%
Europe	25%	Most frequent	Most frequent	-	Less frequent	-
Far-East Asia	-	-	-	33.8%	Less frequent	-
Middle-East-Central Asia	-	-	-	24.3%	75%	-
North America	-	-	-	-	Less frequent	21.5%
Oceania	-	-	-	22.9%	Less frequent	-
South America	Most frequent	50%	Most frequent	-	-	-

Abbreviations: LAM - Latino- American and Mediterranean; EAI - East-African-Indian; CAS1-Delhi – Central Asian strain 1 – Delhi

1.6 Beijing strain genotype

The Beijing strain genotype was first identified by van Soolingen *et al* (1995) after analysing *M. tuberculosis* isolates from TB patients resident in the People's Republic of China and Mongolia (32). A comparison between IS*6110* RFLP fingerprints from the East Asian region showed that strains with this genotype were more frequently observed in East-Asia compared to more distant regions, suggesting that these strains

may have originated and spread from the Beijing area to other regions and thus the naming of the evolutionary lineage as Beijing. Van Soolingen and colleagues hypothesised that the success of this genotype was due to their ability to avoid the protective effect of *Mycobacterium bovis* bacilli Calmette-Guérin (BCG) vaccination. Furthermore, they suggested that the Beijing strains represented an aggressively emerging genotype that has successfully and effectively disseminated from its country of origin to neighbouring and distant regions (32). Subsequently, numerous studies have focused on the analysis of strains representative of the Beijing genotype. According to SNP-based analysis, Beijing genotype strains are members of PGG1 (23), Spoligotype S00034 or ST1 (15), Lineage 1 (33) and Cluster II (25) while according to LSP-based analysis they form part of the East-Asian lineage (27). These characteristics have been used to identify members of the Beijing genotype with the view to discovering unique characteristics associated with their global success. This will be the focus of this review from this point.

Articles were collected with reference to studies that were applicable to the M.tuberculosis Beijing strain genotype. The articles were selected through searches of Pubmed, searching the Internet and checking references in the respective Beijing strain genotype articles (the reference list was investigated for further appropriate studies). The full text of a minor amount of articles was not available (mostly due to articles being published in a foreign language) and therefore only the abstract could be reviewed and were only included if it contained sufficient information. The full text of one article and specific results that were not included in another article, were directly requested from the respective authors. One article was not available on Pubmed and was manually searched from a medical library. The terms "Beijing strain genotype and Mycobacterium tuberculosis" and "HN878" (a clinical isolate with the Beijing genotype) were entered into the search fields of Pubmed and the Internet (using Google). The articles selected included different study designs (population-based cross-sectional studies, reviews, prospective and retrospective studies, pilot studies, descriptive studies, case-control studies and surveillance studies), study settings, experimental models and ranged from the year 1995 when the Beijing strain family was first identified up till the most recent published articles (2009). Articles were included that contained information about any of the different pathogenic characteristics (i.e. drug resistance, host immune response) that have

been hypothesised to be unique to the Beijing strain genotype and could explain the reason for the success of this genotype. Therefore, articles that reported characteristics i.e. drug resistance or immune response activation for M. tuberculosis in general without referring specifically to the Beijing strain genotype were excluded. The included articles defined the Beijing strain genotype either according to IS6110 RFLP fingerprinting and/or spoligotyping. The term "Beijing strain genotype" used in the review refers to Beijing, Beijing-like, W strains and W-like strains unless otherwise stated. The term "non-Beijing strain" refers to strains other than the Beijing, Beijing-like, W strain and W-like strain genotypes. The initial search in Pubmed revealed a hundred and ninety three articles from the year 1995 to 2009 which included 5 reviews. However, 3 reviews upon search were excluded because they were written in foreign languages. Articles were excluded that did not report the relevant information required for this review (i.e. drug resistance data was analysed during the study period but not reported in the article).

2.1 Epidemiology of the Beijing strain genotype

To date, the most frequently observed strain genotype of *M. tuberculosis* causing disease globally is the Beijing strain genotype. Strains of the Beijing genotype have definitive molecular characteristics (Table 3) which differentiate them from all other *M. tuberculosis* strain genotypes. Although they display a variety of multi-copy (15-26) IS*6110* RFLP patterns, they are genetically related to each other based on several independent genetic markers (Table 3). The Beijing strain genotype can be divided into two groups (which share the characteristic spoligotype pattern consisting of at least 3 of the last 9 spacers) according to the absence or presence of an IS*6110* insertion in the NFT region. Strains with this insertion were termed "typical" as these were often observed in different geographical settings. In addition, these strains show similar IS*6110* RFLP patterns along with mutations in the putative mutator genes. Strains lacking this insertion were termed "atypical" as they were relatively rarely observed and demonstrate more diverse IS*6110* RFLP patterns, with no mutations in the putative mutator genes. The atypical Beijing strains have been shown to position near the root of this monophyletic lineage and thereby are suggested to be the progenitor to the typical Beijing strains. The reason for the difference in the frequency of the atypical and typical strains remains to be determined but it has been

suggested to reflect differences in pathogenicity (presumably caused by underlying genetic differences). Using comparative whole-genome hybridization, Tsolaki *et al* (2005) showed that this genotype could be further divided into 4 or 5 subgroups according to LSPs and demonstrated a phylogenetic relationship between the subgroups (34). More recently, the Beijing genotype phylogeny was reconstructed based on a combination of sSNPs, SNPs in mismatch repair genes, IS*6110* insertion points and genomic deletions (regions of difference). This phylogeny was shown to be highly congruent with the "gold standard" phylogenetic tree based on sSNPs. A total of 7 independently evolving Beijing sublineages were identified suggesting that pathogenic characteristics of Beijing genotype strains are not conserved but rather that strains within the Beijing genotype have evolved unique pathogenic characteristics (29).

Table 3: Main molecular characteristics defining the Beijing strain genotype

Unique to the Beijing strain genotype

1. A1 insertion of an IS*6110* element in the origin of replication (intergenic dnaN-dnaN region) (35)
2. Spoligotype S00034 (characterised by deletions of spacers 1-34 and presence of most of the spacers 35-43) (36)
3. Presence of a 3.5-kb *Pvu*II fragment carrying IS*1081* (37)
4. One and two IS*6110* copy(ies) in the NTF chromosomal region for atypical and typical strains respectively (3538)
5. Similar IS*6110* multiband profile (15-26 bands) that show >80% similarity
6. Deletion in region of difference 105 (34)
7. Deletion in Rv0927c and SNP in Rv0927c-pstS3 intergenic spacer (39)
8. Overexpression of DosR regulon with concomitant accumulation of TAG (40)
9. Mutation in Rv2629 (41)

Shared by other strain genotypes

1. Principal genetic group 1 (katG codon 463 CTG (Leu) and gyrA codon 95 ACC (Thr) (23)
2. Intact pks 15/1 gene (42)
3. TLR2 T597C polymorphism (43)

Their success has been attributed to their ability to cause a number of outbreaks in institutional, nosocomial (44, 45) and community settings (46), and their association with MDR (47), indicates their substantial contribution to the TB epidemic. The Beijing strain genotype has also been associated with exogenous re-infection, treatment relapse and treatment failure (48), febrile response during early stages of treatment (6) and HIV (49). In addition, associations with modulation or suppression of the host immune response (inhibition of apoptosis of infected macrophages (50), diminished production of IL-2, IFN-γ, TNF-α and elevated levels of IL-10 and IL-18) (51), decreased expression of certain antigens (heat shock protein of 65-kDa, phosphate transport subunit S and 47-kDa protein) (52), distinct expression of proteins that are associated with virulence (i.e. α-crystallin) (52) and the production of a highly bioactive lipid (a polyketide synthase-derived phenolic glycolipid) (42) have been found. Recently, strains of the Beijing genotype have also been associated with zoonotic transmission (53).

2.2 *Mycobacterium bovis* bacilli Calmette-Guérin (BCG) vaccine

To date, *M. bovis* BCG remains the only agent used to vaccinate children against TB. It is successful in reducing the risk of developing the disseminated form of TB, especially meningitis, during early childhood. However, reports on its protective effect against adult pulmonary TB have been inconsistent and in some cases conflicting. An earlier meta-analysis of studies of the efficacy of BCG concluded that geographic location was the major contributor as this accounted for 41% of the between-study variance (54). Factors associated with varying protective efficacy included prevalence of nontuberculous mycobacterium (NTM), the level of socio-economic development (i.e. population density, capacity for case detection in studies, housing and nutrition), differences in vaccination schedules, differences in climate (which can influence conditions for storage of BCG, exposure to sunlight, and levels of Vitamin D in humans), host genetics and virulence of local endemic strains of *M. tuberculosis* (54 - 56). An alternative view, proposed by Hart and colleagues, suggested that strain differences between BCG preparations were responsible for most of the observed variation in efficacy of the BCG vaccine (55 - 57).

Van Soolingen *et al* (1995) hypothesised that the dissemination of the strains of the Beijing genotype could have been secondary to widespread BCG vaccine use (selective pressure) after their observation that high rates of BCG vaccination was a common factor for all countries in South-East Asia where strains with the Beijing genotype are highly prevalent (32). However, a dissimilar trend in the distribution of Beijing genotype strains was found in South America where BCG vaccination has been implemented since at least 1929 (58). Two possible explanations for the observed trend were reported: (1) the available data on the distribution of strains in these countries are incomplete, therefore it is not known whether or not other new successful strains are already circulating in some communities; and (2) transmission of TB in communities in South American countries may be low compared to South-east Asia, in which case clonal expansion of new strains may take a longer time compared to communities of high transmission (57). Subsequently, numerous studies have attempted to determine whether an association existed between the use of the BCG vaccine and the emergence of the Beijing strain genotype. These findings are summarised in Table 4.

Table 4: Comparison of studies that found an association with the spread of the Beijing strain genotype and BCG vaccination and studies that did not find any associations

BCG vaccination			
Studies that found an **association** with the Beijing strain genotype and the study population investigated		Studies that found **no association** with the Beijing strain genotype and the study population investigated	
Van Soolingen *et al*, 1995 (32)	Human population of China and Mongolia	Anh *et al*, 2000 (59)	Human population of Vietnam
Tuyen *et al*, 2000 (60)	Human population of Vietnam	Van Crevel *et al*, 2001 (6)	Human population of Indonesia
Lopez *et al*, 2003 (61)	mouse model of pulmonary TB	Shi *et al*, 2007 (62)	Human population of Tibet
Castañon-Arreola *et a*l, 2004 (63)	Mouse model of pulmonary TB	Jeon *et al*, 2008 (64)	Mouse model of pulmonary TB
Grode *et al*, 2005 (65)	Mouse model of pulmonary TB		
Tsenova *et al*, 2007 (66)	Rabbit model of TBM		
Kremer *et al*, 2004 (67)	Human population of Vietnam and Hong Kong		

In Vietnam, where BCG vaccination is not uniformly practiced, strains of the Beijing genotype were more commonly associated with vaccinated patients than non-vaccinated patients (60). Subsequently, Kremer *et al* (2004) also showed that BCG vaccinated individuals from Vietnam, Hong Kong and the Netherlands were significantly more often infected with strains of the Beijing genotype than non-Beijing genotype strains (68). However, some studies failed to demonstrate an association between BCG vaccination and the spread of strains with the Beijing genotype in Vietnam, Indonesia, Tibet and various regions of the USA respectively (6, 59, 62). On the other hand, one of these studies did find an association between age and infection and strains of the Beijing genotype suggesting that the prevalence of the Beijing strain genotype rather suggests a cohort effect of BCG vaccination than reduced sensitivity to BCG vaccine-induced immunity of the Beijing strain genotype as postulated by others (59).

Studies conducted in different animal models collectively showed that although BCG conferred a significant protection against Beijing and non-Beijing genotype strains, the protective effect was the lowest for the Beijing strain genotype (61, 63, 65, 66). However, when mice vaccinated with a recombinant BCG (experimental BCG over-expressing the 38-kDa antigen of *M. tuberculosis* or experimental BCG secreting listeriolysin of *Listeria moncytogenesis* which enables the latter to escape from the phagosomes of infected host cells), were challenged with strains of the Beijing genotype, protection was restored (63, 65). In contrast, some animal studies could not find a strain-specific resistance to BCG-induced protective immunity (64). The authors showed that levels of BCG-induced immunity against a classic laboratory strain, 4 Beijing clinical isolates and 4 non-Beijing strains to be generally similar (64). These discrepant results were attributed to the use of different animal TB infection models, *M. tuberculosis* challenge methods, lung pathology analyses and the nature of the strains and their inability to induce protective immunity (63 - 66). Furthermore, the different production methods of BCG can present mycobacterial strain preparations with contrasting immunogenic activities because of different ratios of live and dead organisms, different bacterial concentrations and altered surface composition, which may explain the reason that some results did not correlate with previous findings.

In summary, although it has been hypothesised that strains of the Beijing genotype have acquired mechanisms to circumvent BCG-induced immunity, the association with BCG vaccination and the emergence of strains of the Beijing genotype remains inconclusive. Therefore, the possibility that differences in strains of *M. tuberculosis* might influence the variation in BCG efficacy, cannot completely be ignored.

2.3 Drug resistance

Resistance to one or more of the anti-TB drugs is a major public health concern, which increases TB case numbers and has a severe impact on morbidity and mortality. According to the WHO, the drug resistance epidemic has reached alarming proportions, with more than 490 000 MDR cases and 40 000 XDR cases being diagnosed each year (69).

M. tuberculosis usually develops drug resistance as the result from genomic mutations (in the form of nsSNPs, deletions or insertions) in specific resistance-determining regions of target genes or their promoters (70). Resistance-conferring mutations occur spontaneously at a very low frequency and are selected by inadequate therapy (due to poor patient treatment adherence, unavailability or poor quality of drugs) or sub-therapeutic drug levels (due to drug malabsorption or low drug bioavailability). Drug-resistant cases may then infect close contacts within the community. This is probably exacerbated by prolonged diagnostic delay and failure to ensure optimal treatment (46, 71). Routine culture-based methods for drug susceptibility testing (DST) are slow, expensive and rarely available in resource-limited settings. The fact that it often takes several weeks to produce a result, delays the institution of effective treatment, thereby increasing the risk that drug-resistant bacilli will be transmitted to contacts (71, 72). In addition, patients may receive inappropriate therapy that amplifies resistance and further compromises the chance of a successful treatment outcome (73, 74).

All TB strains can develop drug resistance, however, a number of studies have suggested that the genetic background of the *M. tuberculosis* strain may define how frequently resistance is acquired and/or propagated. Numerous molecular epidemiological studies, done in various geographical settings, have suggested an association between drug resistance and MDR-TB, and the Beijing strain genotype (see appendix C). This association held true in areas where the proportion of the Beijing strain genotype was increasing (epidemic). This prompted the hypothesis that this genotype had the ability to acquire drug resistance more readily than other *M. tuberculosis* genotypes. To address this hypothesis, Rad *et al* (2003) suggested that mutations in putative mutator (*mut*) genes (which encode DNA repair enzymes) may play a role in defining the frequency at which drug resistance is acquired in Beijing genotype strains. Analysis of the *mut* genes in 55 Beijing isolates identified missense alterations in the *mut*T4 (Rv3908), *mut*T2 and *ogt* genes. The authors suggested that these polymorphisms could play a role in the acquisition of drug resistance, since they were unique to strains of the Beijing genotype (75). However these findings are controversial, as drug resistance has been noted in both Beijing and non-Beijing genotype strains without mutations in the putative mutator genes (29, 76).

Subsequent genetic studies have aimed to describe additional mutations in DNA repair genes in order to substantiate the hypothesis of a Beijing mutator phenotype (77, 78). However, the concept of a mutator phenotype was not supported when the frequency of spontaneous mutations was calculated using the Luria-Delbrück fluctuation method (79). The authors found that the rate of spontaneous mutations was the same in Beijing, non-Beijing and laboratory strains.

Comparative studies have suggested that mutations in *rpo*B (confers rifampicin resistance) and *kat*G (confers isoniazid resistance) may define the frequency at which drug resistance is acquired in Beijing genotype strains (80, 81). These studies showed that Beijing genotype strains with an *rpo*B codon 531 and a *kat*G codon 315 mutation occurred more frequently than MDR non-Beijing genotype strains with similar point mutations (80, 81). However, no association with a specific mutation and different *M. tuberculosis* strain genotypes could be found in studies analysing the prevalence of *rpo*B mutations in South-East Asia (82) or *rpo*B and *kat*G mutations in Latvia (83) and England (84). Thus, it has been suggested that the regional differences in the selection of drug resistance-conferring mutations may either be associated with the increased capacity of Beijing genotype strains in different settings to acquire these mutations or simply reflects the efficacy of the TB control program in different countries (i.e. transmission) (81). Alternatively, the association of drug resistance with a specific genotype may be related to the pathogenicity of that genotype, which in turn will be defined by the genetic background. For example, strains with a Beijing genotype may develop resistance at the same frequency as strains with a non-Beijing genotype, although the former spread more efficiently giving an impression that they acquired resistance more readily e.g., MDR-TB outbreak of Beijing genotype strains in the United States and in the Samara region of Russia where two-thirds of TB isolates in prisoners and civilians were infected with the Beijing strain genotype (44). However, this hypothesis contradicts the dogma that the evolution of drug resistance has an associated fitness cost. This may be explained by compensatory mutations which ameliorate the fitness cost associated with resistance (85, 86). Thus, it is possible that Beijing genotype strains are able to accumulate compensatory mutations more readily than non-Beijing genotype strains thereby allowing resistant forms to spread. An alternative explanation may be that the genetic background of the Beijing

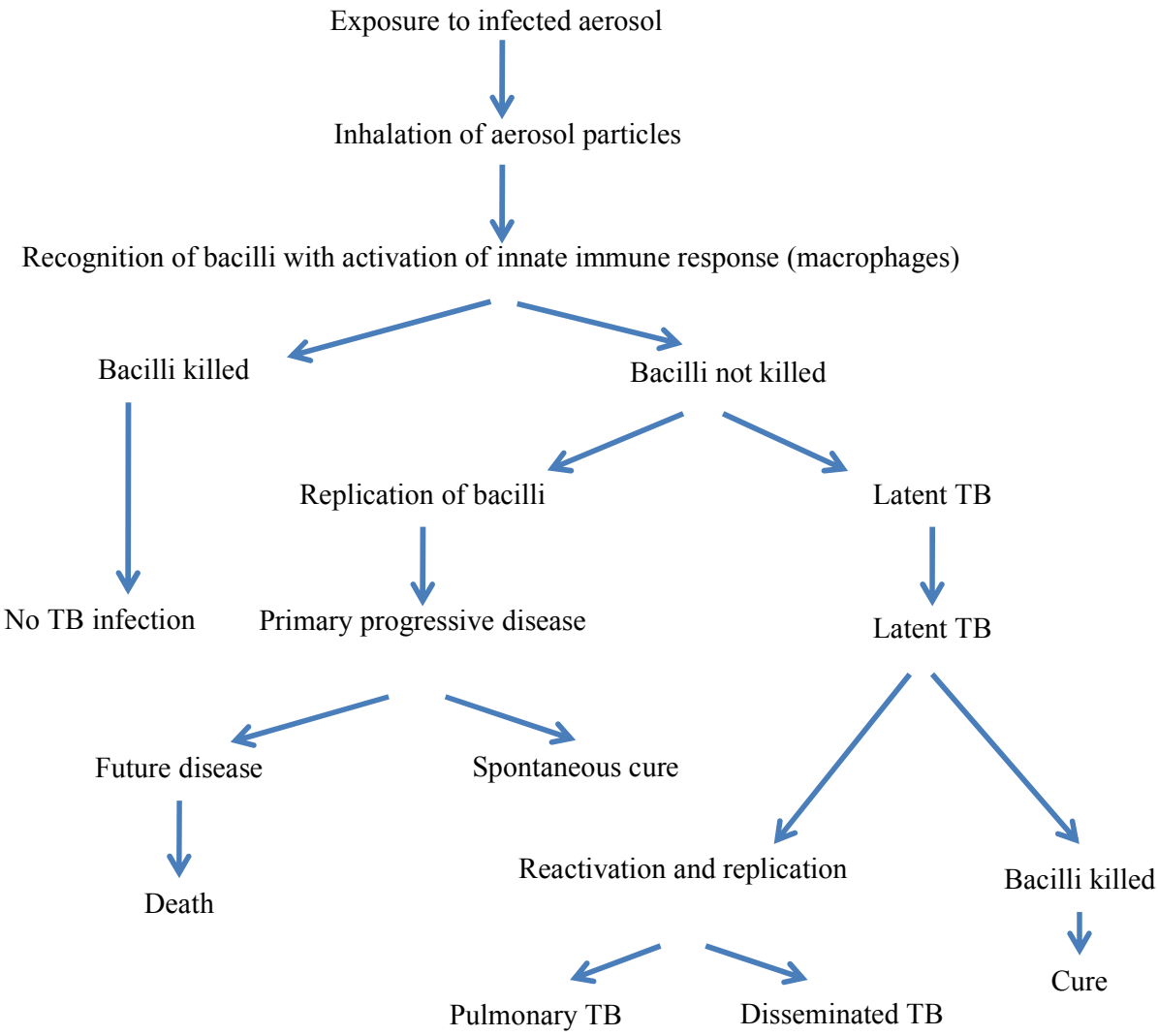
genotype strains confers a high level of fitness when compared to other genotype strains (87). Thus, when the Beijing genotype strains acquire resistance the high intrinsic level of fitness offsets the fitness cost associated with the acquisition of resistance. This was supported by a recent molecular epidemiological study which showed that drug-resistant Beijing genotype strains were able to spread and cause disease more readily than drug-resistant non-Beijing genotype strains (87). An alternative explanation is that the Beijing genotype is dominant in areas where MDR-TB rates are high, as this was the first genotype to develop resistance due to incomplete treatment regimens and thereby has had the greatest opportunity to spread (85, 88, 89). The observations that the Beijing strain genotype is associated with drug resistance have important implications for TB control. Failure to implement rapid diagnostics, improve infection control measures, screen contacts and ensure treatment adherence in these settings will mean that the drug-resistant TB epidemic could continue to increase.

2.4 Virulence

Historically, virulence in mycobacteria has been defined as the ability of the mycobacterium to invade, survive and multiply within host macrophages, the ability to induce a granulomatous inflammatory response and the ability to overcome host defences and persist in the host tissue in a dormant state (even for years). However, virulence in mycobacteria can also be modified by intrinsic factors of the host; avirulent microbes have been noted to become virulent in immune-incompetent hosts (i.e. HIV, steroid therapy). Virulence of mycobacteria has also been quantified in the laboratory; in the majority of mouse model studies, virulence has been measured as the time and inoculum size required to kill the host, number of colony-forming units (CFUs) in the lung homogenates, lung histopathology and the host's defence mechanism (delayed-type hypersensitivity (DTH) reaction). Virulence according to epidemiological studies has been measured as the number of secondary infections caused by an index case (90), the ratio of active to latent infections (91), the propensity to cause cavitatory disease (92) and the ability to disseminate or to cause extrapulmonary infection (93).

After inhalation of viable *M. tuberculosis* bacilli by an immune-competent mycobacterium-naïve host, the innate immune response of the host is activated with the aim of killing the bacilli. However, this is not true for all infected hosts and different scenarios exist (Figure 1): 1) in some infected hosts the bacilli will be killed while in others the bacilli will either persist in a latent form or multiply leading to active disease, 2) in some infected hosts the bacilli will actively multiply in the primary infection area while in others, the bacilli will spread to secondary infection areas and multiply, 3) in vaccinated hosts, some will develop active disease and others latent disease and 4) in hosts with latent disease, some will never develop active disease while others will develop active disease when the host's immune system is compromised.

Figure 1: The response after exposure to an intracellular pathogen



The reasons for the observed discrepancy are not well understood however, there is substantial evidence that host genetics could influence the susceptibility to develop TB disease (94). Furthermore, studies have shown that *M. tuberculosis* has developed intricate mechanisms to suppress the protective host immune response thereby enhancing their ability to survive, multiply inside human macrophages and subsequently to cause active disease (95, 96). However, these intricate mechanisms have not been identified in all *M. tuberculosis* strains, suggesting that different strains exhibit dissimilar pathogenesis leading to different disease outcomes (97).

Much work has been done evaluating *M. tuberculosis* strain-related differences to try to disclose the mechanisms by which strains may suppress the immune response leading to immunopathology. Thus far, various studies have identified a number of possible so-called virulence factors that may play a significant role (42, 50). However, it is not yet clear how these virulence factors enable mycobacteria to grow within its host, to withstand within-host and between-host environmental stresses, to disseminate and to infect a new host.

Molecular epidemiological studies have shown that certain strain genotypes are over-represented, suggesting either host-pathogen compatibility or that the genotype has evolved unique properties enhancing pathogenicity and virulence. To date, most studies have focussed on the Beijing strain genotype as it has been associated with numerous outbreaks worldwide (47) and thus, it has been suggested that the Beijing genotype strain has evolved various mechanisms which enable evasion of the protective host immune response and disease progression. To test this hypothesis, many *in vivo* and *in vitro* TB infection studies have been conducted to analyse the role played by different *M. tuberculosis* strain genotypes in the development of TB in the host. Time to death and organ bacterial load are the most widely used measures, with histopathology and immune parameters studied mainly in the explanation of potential mechanisms underlying the observed virulence.

Earlier studies using infection models in human macrophages found that Beijing genotype strains grew significantly faster than non-Beijing genotype strains. In this study, the Beijing genotype strains caused 25% of TB cases in Los Angeles while the non-Beijing genotype strain was isolated from a diseased patient who had positive sputum for acid-fast bacilli and had contact with many persons at homeless shelters but did not generate secondary cases. Based on these findings, the authors suggested that the Beijing genotype strains had an enhanced ability to avoid host defences (95). Subsequent studies confirmed a higher growth rate associated with strains of the Beijing genotype (98, 99). To determine whether rapid growth in macrophages was characteristic of all Beijing genotype strains, Theus *et al* (2007) infected macrophages with various Beijing genotype strains and found a series of growth phenotypes: three of the Beijing strains grew significantly more slowly than the other Beijing strains,

with Beijing strain 210 growing the fastest. They concluded that rapid growth in macrophages is not a common characteristic of all Beijing genotype strains and that few Beijing genotype strains are as virulent as Beijing strain 210 (100).

In an attempt to determine the mechanism underlying growth rate, numerous *in vitro* TB infection studies have analysed the ability of Beijing and non-Beijing genotype strains to differentially induce cytokines (Table 5).

Table 5: Differential cytokine responses to infection with different *M. tuberculosis* genotype strains

Publications	Beijing strain genotype	Non-Beijing strain genotype
Zhang *et al*,1999 (95)	TNF-α, IL-6, IL-10 and IL-12 equally induced	
Engele *et al*, 2002 (101)	↑TNF-α	↓ TNF-α
Manca *et al*, 2004 (102)	↓ TNF-α, IL-12 ↑IL-4, IL-13	↑TNF-α ↓IL-4, IL-13
Chacón-Salinas *et al*, 2005 (97)	↑TNF-α, IL-12, IL-1ß ↓IL-10	↓TNF-α, IL-12, IL-1ß ↑IL-10
Wong *et al*, 2007 (103)	Not determined	↓ TNF-α
Rocha-Ramirez *et al*, 2007 (43)	↑TNF-α, IL-10	↓ TNF-α, IL-10

From Table 5 it is evident that the respective studies produced different and often conflicting results. A possible explanation for the discordant results reported by Manca *et al* (2004) could be related to the genetic background of the strains used, as it has been shown that different Beijing genotype strains can induce different host immune responses *in vivo* (102). The diverse findings of Rocha-Ramirez *et al* (2007) and Chacon-Salinas *et al* (2005) may be explained by the use of lipid fractions and whole bacteria respectively to infect macrophages (43, 97). Interestingly, the strains that Wong *et al* (2007) defined as hypervirulent were non-Beijing genotype strains isolated from patients with a more severe form of TB infection however, these strains

induced similar cytokine levels as non-Beijing genotype strains used in other studies (103). When macrophages were infected with Beijing and non-Beijing genotype strains with similar drug resistance profiles (MDR), the highest death of macrophages were observed with infection of strains with the Beijing genotype (96). The outcome of infection with Beijing and non-Beijing genotype strains on programmatic cell death of macrophages has also been investigated since induction of macrophage apoptosis accompanies reduced bacterial counts therefore, containing the disease by the host's defence mechanism. The authors found that strains with the Beijing genotype induced a lower extent of macrophage apoptosis, permitting higher bacillary growth. This finding was based on lower expression of IFN-γ, TNF-α and iNOS levels by Beijing genotype strains which subsequently reduces NO production and apoptosis (50).

In an attempt to address the influence of different *M. tuberculosis* strain genotypes on host immune response, a number of different measures of virulence (survival, lung pathology, bacterial load, DTH response) have also been studied in animal TB infection models (mice, rabbits, guinea pigs). Comparative survival studies in mice conducted by Manca *et al* (2001) revealed that strains of the Beijing genotype are of unusually high virulence as mice died more rapidly than mice infected with non-Beijing genotype strains. This was attributed to the decreased ability of spleen and draining lymph node cells of mice to proliferate and produce IFN-γ in response to *M. tuberculosis* antigens (105). This was supported by follow-up studies which found that strains of the Beijing genotype induced the highest mortality in mice compared to mice infected with non-Beijing genotype strains (61, 106). In an extensive study by Dormans *et al* (2004), the survival, lung pathology, bacterial load and DTH responses of BALB/c mice after intratracheal infection with 19 different MTBC strains (including 3 Beijing strains) of 11 major strain genotypes were evaluated. They found that 2 of the 3 Beijing genotype strains were highly virulent indicated by short survival time associated with high CFU counts, while one Beijing genotype strain was less virulent, therefore suggesting that not all Beijing genotype strains are hypervirulent (104).

A study conducted in a guinea pig model used the extent of extrapulmonary lesions and necrosis and time of death as an indicator for virulence and found similar results for Beijing and non-Beijing genotype strains (107).

In a study conducted by Barczak *et al* (2005), mice were simultaneously infected with different concentrations of both Beijing and non-Beijing genotype strains and observed for mycobacterial growth kinetics, strain proportions during infection and mouse survival (108). They revealed that Beijing genotype strains were the most dominant virulent phenotype even when they only made up 4% of the mixed infection. In fact, the simultaneous presence of a non-Beijing and Beijing genotype strain in infected lungs of mice did not appear to influence the outcome of infection in mice. This was explained by the inefficient activation of infected macrophages by Beijing genotype strains rather than inherent differences of strains to overcome macrophage defence mechanisms. Similarly, Tsenova *et al* (2005) showed in tuberculosis meningitis (TBM) rabbit models that the innate immune response to infection with one strain did not appear to affect the innate immune response to the other strain during co-infection. This was based on the observation that macrophages infected with Beijing genotype strains breached the blood brain barrier and seeded to the lungs more efficiently than macrophages infected with non-Beijing genotype strains (109).

In contrast to macrophage infection studies, *in vivo* infection in mouse models showed that strains of the Beijing genotype failed to induce a Th1-type host immune response with subsequent failure to contain Beijing genotype associated disease. They observed a decreased induction of IL-12 and IFN-γ mRNA and increased early production of INF-α mRNA in Beijing-infected mouse lungs (51, 102, 105). It has been suggested that *M. tuberculosis* strains that induce higher levels of type 1 IFN may be more virulent because the development of the IL-12 dependent Th1-type response is reduced (105, 110). However, contradictory results in animal models have also been observed. Sun *et al* (2006) found the Th1-type immune responses to be similar in patients infected with Beijing and non-Beijing genotype strains but strains with a non-Beijing genotype activated more Th2-type immune responses compared to Beijing genotype strains as evidenced by increased IL-4 expression

(111). Some studies found that strains of the Beijing genotype induced a transient increase in TNF-α and iNOS mRNA expression compared to non-Beijing genotype strains that on the other hand induced a high and sustained TNF-α and iNOS mRNA expression (61, 112). This was further supported by Ordway *et al* (2007) who found evidence that initially within the first 2 weeks after infection strains of the Beijing genotype induced a strong Th1-type immune response. This was characterised by a significant IFN-γ and TNF-α production which subsequently declined rapidly. The subsequent reduction in Th1-type immunity was further associated with the rapid emergence of a $CD4^{+}CD25^{+}FoxP3^{+}CD223^{+}IL\text{-}10^{+}$ regulatory T cell population. However, with non-Beijing strain infections, the presence of T regulatory cells was not associated with the inhibition of the early acquired immune response by reduced IFN-γ levels (113).

It has been hypothesised that *M. tuberculosis* has evolved a number of mechanisms to reduce recognition by the host innate immune system and to reduce a tendency to stimulate cytokine responses:

Phenolic glycolipid (PGL): Studies of aerosol infection in mice have observed an increase in time to death between animals infected with a Beijing genotype strain compared to strains with the non-Beijing genotype. The increased virulence of the Beijing strain genotype was associated with the production of a polyketide synthase derived phenolic glycolipid, which dose-dependently reduces the production of Th1-type cytokines (TNF-α, IL-6 and IL-12) by macrophages (42, 51, 102, 105, 109). A 7-bp deletion of the polyketide synthase gene cluster pks1-15 which gives rise to a frameshift mutation has been associated with the inability of non-Beijing genotype strains to synthesise PGL (114). Growth of wild-type Beijing strains and Beijing strains with the mutant pks1-15 in mouse lungs and spleens were similar as well as bacterial titres at time of death of mice. However, median survival times of mutant - infected mice were 90 days longer than those of mice infected with the wild-type Beijing strains (51). Therefore inactivation of pks1-15 inhibited production of Beijing-specific lipids as well as reduced virulence to that of non-Beijing strains. On the other hand, when human peripheral blood monocytes were infected with a non-Beijing strain transformed with the functional pks1-15 gene, growth was similar in the

mutant non-Beijing strain compared to the wild-type non-Beijing strain suggesting that PGL did not influence growth of *M. tuberculosis* (115). To determine whether the pks1-15 gene were responsible for successful transmission of a *M. tuberculosis* strain in California, USA, differences in the pks1-15 locus sequence between the outbreak strain and unique strains were compared. They found that the outbreak strain had the same frameshift mutation in the pks1-15 locus as found in non-Beijing strains while the unique strains shared the same pks1-15 sequence found in Beijing strains. Therefore, pks1-15 gene could not explain the outbreak of this strain in the study setting (116). It is possible that other molecular mechanisms could be responsible for their success as suggested by some authors (40, 115). They showed that some Beijing strains that possess the intact pks1-15 gene were deficient in PGL synthesis indicating that not all members of this genotype are producers of this lipid (40). However, regardless of the strain's ability to produce PGL, all clinical isolates have been reported to be virulent and have likely adapted mechanisms to survive in their respective niches (115).

Genotyping studies showed that the pks1-15 gene was intact in all Beijing genotype strains tested, but mutated in non-Beijing genotype strains (42, 114). This was further supported by subsequent studies that found the intact pks1-15 gene in almost all of the Beijing genotype strains (34, 117, 118). However, the intact region was also found in PGG 1 non-Beijing strains isolated from patients with severe manifestations of TB such as TBM, as well as from those with lung disease (117, 118). This suggests that although PGL are involved in the hypervirulence of the PGL-producing strains, they are not a unique characteristic of the Beijing strain genotype. However, according to Sinisimer *et al*, (2008) although the lipid's cytokine-modulatory function is active in non-Beijing genotype strains, this does not necessarily confer hypervirulence in non-Beijing genotype strains (115). When a laboratory non-Beijing strain was transformed with a functional pks1-15 gene, it elicited lesser amounts of particular cytokines than its parental strain. However, the cytokines differentially induced were not the same as previously observed to be differentially elicited by a Beijing strain and its PGL mutant Beijing strain (42). The mutant Beijing strain induced higher levels of IL-12 compared to the wild-type Beijing strain while the mutant and wild-type non-Beijing strains induced similar levels of IL-12. Hence, the hypervirulence of

the Beijing genotype strain was suggested to be attributed to its failure to induce a strong Th1-type immune response including inducing lower levels of IL-12. Thus, it is possible that by affecting the ability of Beijing genotype strains to induce IL-12, PGL induces a hypervirulence phenotype in infected animal models. This suggests that the ability of PGL to modulate the host cytokine response is likely dependent on the background of the strain in which it is produced. In addition, the recombinant non-Beijing strain was not more virulent in the mouse or rabbit model of infection, as judged by bacillary growth in lungs, dissemination of bacilli to other tissues, gross pathology and survival. In fact, the recombinant non-Beijing strain was even less virulent. In contrast, the virulence of Beijing strains in rabbits were increased compared to the PGL-deficient mutant Beijing strain, further supporting the hypothesis that the effect of PGL on virulence in different animal models may be strain dependent (109). Collective data therefore conclude that, while PGL is an important virulence factor for the Beijing genotype strain only, it probably acts synergistically with other protein and/or lipid virulence factors, also present in other *M. tuberculosis* strains (115).

Antigen expression: Differences in antigen expression levels between Beijing and non-Beijing genotype strains have been proposed as a mechanism which may assist strains of the Beijing genotype to minimise recognition by the host immune response, thereby facilitating the increased prevalence of the Beijing genotype strains (52, 119). Strains of the Beijing genotype were found to have increased expression of α-crystallin (*M.tuberculosis* virulence factor (120)) and decreased expression of Hsp65, PstS1 (B- and T-cell stimulant), and the 47kDa protein compared to the non-Beijing genotype strains which may allow strains of the Beijing genotype to evade the host immune response (121). In accordance, Rindi *et al* (2007) found the expression of the ppe44 gene to be significantly higher in Beijing strains than non-Beijing strains and concluded that differential expression of ppe genes could influence the host's immune responsiveness (119).

Sigma factor, Sig A and the eis gene: Wu *et al* (2004) proposed that the upregulation of the principal sigma factor, SigA in Beijing genotype strains, enhances growth in macrophages and in the lungs of mice during the initial phase of infection (122).

They found that strains with the Beijing genotype had higher sigA mRNA levels and higher intracellular growth rates compared to non-Beijing genotype strains. Furthermore, SigA was also upregulated in Beijing strains during growth in macrophages compared with growth in broth. In contrast, sigA mRNA levels in non-Beijing strains did not change under these conditions (122). Follow-up studies demonstrated that the expression of *eis* (elicits higher levels of IL-10 than TNF-α secretion therefore manipulating the host TNF-α/IL-10 axis to its advantage by inducing a state of localised immunosuppression in the host) was enhanced 12-fold in the Beijing strains compared to the non-Beijing strains when grown in a human monocyte cell line, whereas the expression was the same when the strains were grown *in vitro* (123). They also showed that SigA binds to the promoter region of *eis* and is responsible for the upregulation of *eis* expression in Beijing genotype strains. They concluded that *eis* may contribute to the enhanced growth of the Beijing genotype strains in human macrophages. This was further supported by a subsequent study where increased IL-10 and decreased TNF-α production driven by *eis*, and possibly other *M. tuberculosis* antigens, was proposed to possibly contribute to the intracellular survival of the pathogen by suppressing host cell apoptosis during mycobacterial infection (124).

Dormancy regulon genes (dosR, Rv3130c, hspX, fdxA and narX): Reed and colleagues have demonstrated that strains of the Beijing genotype have a basal level of transcription for several dormancy regulon genes (known to be involved in the adaptive response of *M. tuberculosis* to hypoxia and reactive nitrogen intermediates) that is up to 50-fold higher than that of non-Beijing genotype strains (40). Similarly, the level of the Rv3130c transcript was 10-fold higher in Beijing than non-Beijing genotype strains under standard growth conditions (Rv3130c which encodes a Triacylglycerol (TAG) synthase whose activity is induced in response to hypoxia and NO treatment). This suggests that the Beijing genotype strain has the ability to accumulate fatty acids in the form of TAG which is a readily available source of energy during anaerobic conditions (i.e. latency). The constitutive accumulation of TAG by strains of the Beijing genotype supports the hypothesis that these strains are more adept than strains of the non-Beijing genotype at altering their metabolism to that required for survival within host tissues. While the role of the DosR regulon and

TAG in human TB remains to be explored, the study by Reed *et al* (2007) represents the description of a series of phenomena that occur uniquely within members of a single *M. tuberculosis* genotype (40).

Toll-like receptor 2 (TLR2) gene: Caws *et al* (2008) recently identified a relationship between bacterial and host genotype in TB for the first time. They observed a polymorphism in the TLR2 gene and an association with disease caused by the Beijing genotype strain (7). Therefore, they hypothesised that polymorphisms in genes responsible for the innate immune response to infection may influence the host's response to infection and may result in increased susceptibility to disease from some bacterial lineages but not others. However, the Beijing genotype's own ability to subvert the host innate immune response and subsequently make the host more susceptible to TB could not be excluded.

2.5 Clinical presentation

To determine whether a correlation existed between mycobacterial genotype and clinical presentation and outcome (phenotype), many studies have integrated molecular data (Beijing and non-Beijing genotype strains) with clinical information (chest X-ray (CXR) features, symptoms and signs, age, smear grading, treatment category, disease site and co-morbid diseases). Table 6 summarises these findings. From Table 6 it is evident that a high degree of variability exists between the different studies which may be due to the host population in which the study was performed (including HIV status), the power of the different studies and diagnostic delay (masking extent of disease). Thus it is not possible to conclusively demonstrate a universal association between the Beijing strain genotype and clinical presentation.

Table 6: Correlation between phenotype and the Beijing genotype

Phenotype	**Association** with the Beijing genotype	**No association** with the Beijing genotype
Pre-treatment CXR findings	Van Rie *et al*, 1999 (46) Drobniewski *et al*, 2005 (125)	Borgdorff *et al*, 2004 (126) Van Crevel *et al*, 2001 (6) Thwaites *et al*, 2008 (127)

		Feng *et al*, 2008 (128) Kong *et al*, 2007 (93) Sun *et al*, 2006 (110)
Progression to disease	De Jong *et al*, 2008 (129) Thwaites *et al*, 2008 (127) Sapozhnikova *et al*, 2003 (130)	-
Febrile response during early stages of treatment	Van Crevel *et al*, 2001 (6)	-
Symptoms: Night sweats	-	Drobniewski *et al*, 2005 (125) Sun *et al,* 2006 (110)
Duration of disease symptoms prior to TB diagnosis	Thwaites *et al*, 2008 (127) Sun *et al*, 2006 (131)	Van Crevel *et al*, 2001 (6)
Younger age (<65)	Thwaites *et al*, 2008 (127) Anh *et al*, 2000 (59) Lillebaek *et al*, 2003 (132) Van Rie *et al*, 1999 (46) Kubica *et al*, 2004 (133) Borgdorff *et al*, 2003 (134) Moss *et al*, 1997 (135) Caminero *et al*, 2001 (5) Jou *et al*, 2005 (136) Drobniewski *et al*, 2005 (125) Li *et al*, 2005 (137) Kong *et al*, 2007 (93)	Van Crevel *et al*, 2001 (6) Borgdorff *et al*, 2004 (126) Sun *et al*, 2006 (111) Glynn *et al*, 2002 (47) Dale *et al*, 2005 (138) Tracevska *et al*, 2003 (83) Chan *et al*, 2001 (139) Bifani *et al*, 1999 (36) Prodinger *et al*, 2001 (140) Palittapongarnpim *et al*, 1997 (141) Krüüner *et al*, 2001 (142)
Previous TB treatment	Feng *et al*, 2008 (128) Drobniewski *et al*,2005 (125)	Borgdorff *et al*, 2004 (126) Toungoussova *et al*, 2002 (143) Lillebaek *et al*, 2003 (132) Glynn *et al*, 2005 (144)
Extrapulmonary disease	Kong *et al*, 2007 (93)	Feng *et al*, 2008 (128) Borgdorff *et al*, 2004 (126)

		Caws *et al*, 2008 (7) Lillebaek *et al*, 2003 (132) Dale *et al*, 2005 (138) Glynn *et al*, 2005 (144) Hasan *et al*, 2006 (145) Nicol *et al*, 2005 (146)
Smear positivity	-	Feng *et al*, 2008 (128) Borgdorff *et al*, 2004 (126) Glynn *et al*, 2005 (144) Van Crevel *et al*, 2001 (6) Kong *et al*, 2007 (93)
Treatment failure and disease relapse	Lan *et al*, 2003 (48) Sun *et al*, 2006 (111)	Feng *et al*, 2008 (128)
Clustering	Cox *et al*, 2005 (147)	Kong *et al*, 2007 (93)
Previous imprisonment	Drobniewski *et al*, 2005 (125)	Cox *et al*, 2005 (147)
HIV status	Caws *et al*, 2006 (49)	Drobniewski *et al*, 2005 (125) Palittapongarnpim *et al*, 1997 (141)

3.1 Conclusion

A prominent theme emerging from studies in the present review is that the genetic background of *M. tuberculosis* strains may influence the risk of disease progression. Based on the data accumulated it is clear that virulence of the Beijing strain genotype is not determined by a single characteristic but is the result of a unique, intricate, dynamic interaction between host and the pathogen that is encoded by the pathogen's genome. The association between genetic diversity and differential induction of the host immune response, drug resistance patterns and clinical or pathogenic features emphasizes the importance of continued investigation into the genetic mechanisms underlying the phenotypic and genotypic characteristics of different *M. tuberculosis* strains. Advances in knowledge regarding the Beijing strain genotype will enhance

our understanding of the contribution of pathogenic factors which define disease progression following infection. This in turn may enable the development and assessment of new interventions. In future, these findings could help to direct drug and vaccine development towards targeting critical virulence factors and could also enable the development of new diagnostic tests to identify persons at risk for developing active disease. The latter should in turn guide TB control programs towards high risk patients who are in need of preventative therapy or prompt diagnosis thereby limiting transmission and the perpetuation of the TB epidemic.

Reference List

1. **WHO** REPORT 2009, Global Tuberculosis Control.

2. **WHO**. Fact sheet No104, (Revised March 2007).

3. **THE STOP TB STRATEGY**: Building on and enhancing DOTS to meet the the TB-related Millennium Development Goals. WHO/HTM/TB/2006.368.

4. **Valway, S.E., Sanchez, M.P., Shinnick, T.F., Orme, I., Agerton, T., Hoy, D. et al**. An outbreak involving extensive transmission of a virulent strain of *Mycobacterium tuberculosis*. *N. Engl. J. Med*. 1998, **338**:633-639.

5. **Caminero, J.A., Pena, M.J., Campos-Herrero, M.I., Rodríguez, J.C., García, I., Cabrera, P. et al.** Epidemiological evidence of the spread of a *Mycobacterium tuberculosis* strain of the Beijing genotype on Gran Canaria Island. *Am. J. Respir.Crit. Care Med.* 2001, **164**:1165-1170.

6. **van Crevel, R., Nelwan, R.H., de Lenne, W., Veeraragu, Y., van der Zanden, A.G., Amin, Z. et al.** *Mycobacterium tuberculosis* Beijing genotype strains associated with febrile response to treatment. *Emerging Infect. Dis*. 2001, 7:880-883.

7. **Caws, M., Thwaites, G., Dunstan, S., Hawn, T.R., Lan, N.T.N., Thuong, N.T.T. et al**. The influence of host and bacterial genotype on the development of disseminated disease with *Mycobacterium tuberculosis*. *PLoS Pathog*. 2008, **4:**e1000034.

8. **Palanisamy, G.S., DuTeau, N., Eisenach, K.D., Cave, D.M., Theus, S.A., Kreiswirth, B.N. et al.** Clinical strains of *Mycobacterium tuberculosis* display a wide range of virulence in guinea pigs. *Tuberculosis* (Edinb) 2009, **89**:203-209.

9. **Ross, B.C., Raios, K., Jackson, K. & Dwyer, B.** Molecular cloning of a highly repeated DNA element from *Mycobacterium tuberculosis* and its use as an epidemiological tool. *J. Clin. Microbiol.* 1992, **30**:942-946.

10. **van Embden, J.D., Cave, M.D., Crawford, J.T., Dale, J.W., Eisenach, K.D., Gicquel, B. et al**. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J. Clin. Microbiol.* 1993, **31:**406-409.

11. **Hermans, P.W., van Soolingen, D., Dale, J.W., Schuitema, A.R., McAdam, R.A., Catty, D. et al**. Insertion element IS986 from *Mycobacterium tuberculosis:* a useful tool for diagnosis and epidemiology of tuberculosis. *J. Clin. Microbiol*. 1990, **28**:2051-2058.

12. **van Soolingen, D., Hermans, P.W., de Haas, P.E., Soll, D.R. & van Embden, J.D.** 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*. 1991, **29**:2578- 2586.

13. **Behr, M.A., Warren, S.A., Salamon, H., Hopewell, P.C., Ponce de Leon, A., Daley, C.L. et al**. Transmission of *Mycobacterium tuberculosis* from patients smear-negative for acid-fast bacilli. *Lancet* 1999, **353**:444-449.

14. **van Rie, A., Warren, R., Richardson, M., Victor, T.C., Gie, R.P., Enarson, D.A. et al.** Exogenous reinfection as a cause of recurrent tuberculosis after curative treatment. *N. Engl. J. Med*. 1999, **341**:1174-1179.

15. **Kamerbeek, J., Schouls, L., Kolk, A., van Agterveld, M., van Soolingen, D., Kuijper, S. et al.** Simultaneous detection and strain differentiation of

Mycobacterium tuberculosis for diagnosis and epidemiology. *J. Clin. Microbiol*. 1997, **35**:907-914.

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

17. **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**:3563-3571.

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

19. **Streicher, E.M., Victor, T.C., van der Spuy, G., Sola, C., Rastogi, N., van Helden, P.D. et al.** Spoligotype signatures in the *Mycobacterium tuberculosis* complex. *J. Clin. Microbiol*. 2007, **45**:237-240.

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

21. **Filliol, I., Driscoll, J.R., Van Soolingen, D., Kreiswirth, B.N., Kremer, K., Valétudie, G. et al.** Global distribution of *Mycobacterium tuberculosis* spoligotypes. *Emerging Infect. Dis*. 2002, **8**:1347-1349.

22. **Filliol, I., Motiwala, A.S., Cavatore, M., Qi, W., Hazbón, M.H., Bobadilla del Valle, M. et al.** Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and

recommendations for a minimal standard SNP set. *J. Bacteriol*. 2006, **188**:759-772.

23. **Sreevatsan, S., Pan, X., Stockbauer, K.E., Connell, N.D., Kreiswirth, B.N., Whittam, T.S. et al**. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc. Natl. Acad. Sci.* U.S.A. 1997, **94**:9869-9874.

24. **Rhee, J.T., Piatek, A.S., Small, P.M., Harris, L.M., Chaparro, S.V., Kramer, F.R. et al.** Molecular epidemiologic evaluation of transmissibility and virulence of *Mycobacterium tuberculosis*. *J. Clin. Microbiol*. 1999, **37**:1764-1770.

25. **Gutacker, M.M., Mathema, B., Soini, H., Shashkina, E., Kreiswirth, B.N., Graviss, E.A. et al**. Single-nucleotide polymorphism-based population genetic analysis of *Mycobacterium tuberculosis* strains from 4 geographic sites. *J. Infect. Dis*. 2006, **193**:121-128.

26. **Tsolaki, A.G., Hirsh, A.E., DeRiemer, K., Enciso, J.A., Wong, M.Z., Hannan, M. et al.** Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. *Proc. Natl. Acad. Sci.* U.S.A. 2004, **101**:4865-4870.

27. **Gagneux, S., DeRiemer, K., Van, T., Kato-Maeda, M., de Jong, B.C., Narayanan, S. et al.** Variable host-pathogen compatibility in *Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. U.S.A*. 2006, **103**:2869-2873.

28. **Reed, M.B., Pichler, V.K., McIntosh, F., Mattia, A., Fallow, A., Masala, S. et al.** Major *Mycobacterium tuberculosis* lineages associate with patient country of origin. *J. Clin. Microbiol*. 2009, **47**:1119-1128.

29. **Hanekom, M., van der Spuy, G.D., Streicher, E., Ndabambi, S.L., McEvoy, C.R.E., Kidd, M. et al.** A recently evolved sublineage of the *Mycobacterium tuberculosis* Beijing strain family is associated with an increased ability to spread and cause disease. *J. Clin. Microbiol*. 2007, **45**:1483-1490.

30. **Hanekom, M., van der Spuy, G.D., Gey van Pittius, N.C., McEvoy, C.R.E., Ndabambi, S.L., Victor, T.C. et al**. Evidence that the spread of *Mycobacterium tuberculosis* strains with the Beijing genotype is human population dependent. *J. Clin. Microbiol.* 2007, **45**:2263-2266.

31. **Brudey, K., Driscoll, J.R., Rigouts, L., Prodinger, W.M., Gori, A., Al-Hajoj, S.A. 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.

32. **van Soolingen, D., Qian, L., de Haas, P.E., Douglas, J.T., Traore, H., Portaels, F. et al.** Predominance of a single genotype of Mycobacterium tuberculosis in countries of East Asia. *J. Clin. Microbiol*. 1995, **33**:3234-3238.

33. **Baker, L., Brown, T., Maiden, M.C. & Drobniewski, F**. Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. *Emerging Infect. Dis*. 2004, **10**:1568-1577.

34. **Tsolaki, A.G., Gagneux, S., Pym, A.S., Goguet de la Salmoniere, Y.L., Kreiswirth, B.N., Van Soolingen, D. et al.** Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis. J. Clin. Microbiol.* 2005, **43**:3185-3191.

35. **Kurepina, N.E., Sreevatsan, S., Plikaytis, B.B., Bifani, P.J., Connell, N.D., Donnelly, R.J. et al.** Characterization of the phylogenetic distribution and chromosomal insertion sites of five IS6110 elements in *Mycobacterium tuberculosis*: non-random integration in the dnaA-dnaN region. *Tuber. Lung Dis*. 1998, **79**:31-42.

36. **Bifani PJ, Mathema B, Liu Z, et al** Identification of a W variant outbreak of *Mycobacterium tuberculosis* via population-based molecular epidemiology. *JAMA* 1999, **282**;2321-2327.

37. **van Soolingen, D., de Haas, P.E., Blumenthal, R.M., Kremer, K., Sluijter, M., Pijnenburg, J.E. et al.** Host-mediated modification of PvuII restriction in *Mycobacterium tuberculosis*. *J. Bacteriol.* 1996, **178**:78-84.

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

39. **Jiang, X., Lu, C., Gao, F., Wang, F., Zhang, W., Portugal, I. et al.** A rapid and simple method for identifying *Mycobacterium tuberculosis* W-Beijing strains based on detection of a unique mutation in Rv0927c by PCR-SSCP. *Microbes Infect*. 2009, **11**:419-423.

40. **Reed, M.B., Gagneux, S., Deriemer, K., Small, P.M. & Barry, C.E.3rd**. The W-Beijing lineage of *Mycobacterium tuberculosis* overproduces triglycerides and has the DosR dormancy regulon constitutively upregulated. *J. Bacteriol*. 2007, **189**:2583-2589.

41. **Homolka, S., Köser, C., Archer, J., Rüsch-Gerdes, S. & Niemann, S.** Single-nucleotide polymorphisms in Rv2629 are specific for *Mycobacterium tuberculosis* genotypes Beijing and Ghana but not associated with rifampin resistance. *J. Clin. Microbiol*. 2009, **47**:223-226.

42. **Reed, M.B., Domenech, P., Manca, C., Su, H., Barczak, A.K., Kreiswirth, B.N. et al**. A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. *Nature* 2004, **431**:84-87.

43. **Rocha-Ramírez, L.M., Estrada-García, I., López-Marín, L.M., Segura-Salinas, E., Méndez-Aragón, P., Van Soolingen, D. et al.** *Mycobacterium tuberculosis* lipids regulate cytokines, TLR-2/4 and MHC class II expression in human macrophages. *Tuberculosis* (Edinb) 2008, **88**:212-220.

44. **Frieden, T.R., Sherman, L.F., Maw, K.L., Fujiwara, P.I., Crawford, J.T., Nivin, B. et al.** A multi-institutional outbreak of highly drug-resistant

tuberculosis: epidemiology and clinical outcomes. *JAMA* 1996, **276**:1229-1235.

45. **Agerton, T.B., Valway, S.E., Blinkhorn, R.J., Shilkret, K.L., Reves, R., Schluter, W.W. et al**. Spread of strain W, a highly drug-resistant strain of *Mycobacterium tuberculosis*, across the United States. *Clin. Infect. Dis*. 1999, **29**:85-92; discussion 93-95.

46. v**an Rie, A., Warren, R.M., Beyers, N., Gie, R.P., Classen, C.N., Richardson, M. et al.** Transmission of a multidrug-resistant *Mycobacterium tuberculosis* strain resembling "strain W" among noninstitutionalized, human immunodeficiency virus-seronegative patients. *J. Infect. Dis*. 1999, **180**:1608-1615.

47. **Glynn, J.R., Whiteley, J., Bifani, P.J., Kremer, K. & van Soolingen, D.** Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: a systematic review. *Emerging Infect. Dis*. 2002, **8**:843-849.

48. **Lan, N.T.N., Lien, H.T.K., Tung, L.B., Borgdorff, M.W., Kremer, K. & van Soolingen, D**. *Mycobacterium tuberculosis* Beijing genotype and risk for treatment failure and relapse, Vietnam. *Emerging Infect. Dis*. 2003, **9**:1633-1635.

49. **Caws, M., Thwaites, G., Stepniewska, K., Nguyen, T.N.L., Nguyen, T.H.D., Nguyen, T.P. et al.** Beijing genotype of *Mycobacterium tuberculosis* is significantly associated with human immunodeficiency virus infection and multidrug resistance in cases of tuberculous meningitis. *J. Clin. Microbiol*. 2006, **44**:3934-3939.

50. **Ríos-Barrera, V.A., Campos-Peña, V., Aguilar-León, D., Lascurain, L.R., Meraz-Ríos, M.A., Moreno, J. et al.** Macrophage and T lymphocyte apoptosis during experimental pulmonary tuberculosis: their relationship to mycobacterial virulence. *Eur. J. Immunol*. 2006, **36**:345-353.

51. **Manca, C., Tsenova, L., Freeman, S., Barczak, A.K., Tovey, M., Murray, P.J. et al.** Hypervirulent *Mycobacterium tuberculosis* W/Beijing strains

upregulate type I IFNs and increase expression of negative regulators of the Jak-Stat pathway. *J. Interferon Cytokine Res.* 2005, **25**:694-701.

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

53. **Parsons SD, Gous TA, Warren RM, van Helden PD.** Pulmonary *Mycobacterium tuberculosis* (Beijing strain) infection in a stray dog. *J S Afr Vet Assoc.* 2008 Jun, **79**(2):95-98.

54. **Colditz, G.A., Brewer, T.F., Berkey, C.S., Wilson, M.E., Burdick, E., Fineberg, H.V. et al**. Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature. *JAMA* 1994, **271**:698-702.

55. **Fine, P.E**. Bacille Calmette-Guérin vaccines: a rough guide. *Clin. Infect. Dis*. 1995, **20**:11-14.

56. **Comstock, G.W**. Efficacy of BCG vaccine. *JAMA* 1994, **272**:766.

57. **Hart, P.D**. Efficacy and applicability of mass B.C.G. vaccination in tuberculosis control. *Br Med J* 1967 Mar 11, **1**(5540):587-592.

58. **Abebe, F. & Bjune, G**. The emergence of Beijing family genotypes of *Mycobacterium tuberculosis* and low-level protection by bacille Calmette-Guérin (BCG) vaccines: is there a link? *Clin. Exp. Immunol*. 2006, **145**:389-397.

59. **Ahn, H.C., Yang, J.H., Lee, H.B., Rhee, Y.K. & Lee, Y.C.** Effect of combined therapy of oral anti-tubercular agents on theophylline pharmacokinetics. *Int. J. Tuberc. Lung Dis*. 2000, **4**:784-787.

60. **Le TK, Bach KH, Ho ML, Le NV, Nguyen TN, Chevrier D, Guesdon JL**. Molecular fingerprinting of *Mycobacterium tuberculosis* strains isolated in Vietnam using IS 6110 as probe. *Tuber Lung Dis*. 2000, **80**(2):75-83.

61. **López, B., Aguilar, D., Orozco, H., Burger, M., Espitia, C., Ritacco, V. et al.** A marked difference in pathogenesis and immune response induced by different *Mycobacterium tuberculosis* genotypes. *Clin. Exp. Immunol*. 2003, **133**:30-37.

62. **Shi, C., Wang, X., Zhang, H., Xu, Z., Li, Y. & Yuan, L**. Immune responses and protective efficacy induced by 85B antigen and early secreted antigenic target-6 kDa antigen fusion protein secreted by recombinant bacille Calmette-Guérin. *Acta Biochim. Biophys. Sin.* (Shanghai) 2007, **39**:290-296.

63. **Castañon-Arreola, M., López-Vidal, Y., Espitia-Pinzón, C. & Hernández-Pando, R.** A new vaccine against tuberculosis shows greater protection in a mouse model with progressive pulmonary tuberculosis. *Tuberculosis* (Edinb) 2005, **85**:115-126.

64. **Jeon, B.Y., Derrick, S.C., Lim, J., Kolibab, K., Dheenadhayalan, V., Yang, A.L. et al.** *Mycobacterium bovis* BCG immunization induces protective immunity against nine different *Mycobacterium tuberculosis* strains in mice. *Infect. Immun*. 2008, **76**:5173-5180.

65. **Grode, L., Seiler, P., Baumann, S., Hess, J., Brinkmann, V., Nasser Eddine, A. et al.** Increased vaccine efficacy against tuberculosis of recombinant *Mycobacterium bovis* bacille Calmette-Guérin mutants that secrete listeriolysin. *J. Clin. Invest*. 2005, **115**:2472-2479.

66. **Tsenova, L., Harbacheuski, R., Sung, N., Ellison, E., Fallows, D. & Kaplan, G.** BCG vaccination confers poor protection against *Mycobacterium tuberculosis* HN878-induced central nervous system disease. *Vaccine* 2007, **25**:5126-5132.

67. **Kremer, K., Glynn, J.R., Lillebaek, T., Niemann, S., Kurepina, N.E., Kreiswirth, B.N. et al.** Definition of the Beijing/W lineage of *Mycobacterium tuberculosis* on the basis of genetic markers. *J. Clin. Microbiol*. 2004, **42**:4040-4049.

68. **Kremer, K., van-der-Werf, M.J., Au, B.K.Y., Anh, D.D., Kam, K.M., van-Doorn, H.R. et al.** Vaccine-induced immunity circumvented by typical *Mycobacterium tuberculosis* Beijing strains. *Emerging Infect. Dis*. 2009, **15**:335-339.

69. TUBERCULOSIS MDR-TB & XDR-TB. THE 2008 REPORT. **WHO** Feb 2008.

70. **Ramaswamy, S. & Musser, J.M**. Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 update. *Tuber. Lung Dis*. 1998, **79**:3-29.

71. **Victor, T.C., Streicher, E.M., Kewley, C., Jordaan, A.M., van der Spuy, G.D., Bosman, M. et al.** Spread of an emerging *Mycobacterium tuberculosis* drug-resistant strain in the western Cape of South Africa. *Int. J. Tuberc. Lung Dis*. 2007, **11**:195-201.

72. **Johnson, R., Streicher, E.M., Louw, G.E., Warren, R.M., van Helden, P.D. & Victor, T.C.** Drug resistance in *Mycobacterium tuberculosis*. *Curr Issues Mol Biol* 2006, **8**:97-111.

73. **Pillay, M and Sturm, W.A**. Evolution of the Extensively Drug□Resistant F15/LAM4/KZN Strain of *Mycobacterium tuberculosis* in KwaZulu□Natal, South Africa. Clin.Infect.Dis. 2007, **45**:1409–1414.

74. **van Helden, P.D., Donald, P.R., Victor, T.C., Schaaf, H.S., Hoal, E.G., Walzl, G. et al.** Antimicrobial resistance in tuberculosis: an international perspective. *Expert Rev Anti Infect Ther* 2006, **4**:759-766.

75. **Ebrahimi-Rad, M., Bifani, P., Martin, C., Kremer, K., Samper, S., Rauzier, J. et al**. Mutations in putative mutator genes of *Mycobacterium tuberculosis* strains of the W-Beijing family. *Emerging Infect. Dis*. 2003, **9**:838-845.

76. **Lari, N., Rindi, L., Bonanni, D., Tortoli, E. & Garzelli, C**. Mutations in mutT genes of *Mycobacterium tuberculosis* isolates of Beijing genotype. *J. Med. Microbiol*. 2006, **55**:599-603.

77. **Nouvel, L.X., Dos Vultos, T., Kassa-Kelembho, E., Rauzier, J. & Gicquel, B.** A non-sense mutation in the putative anti-mutator gene ada/alkA of *Mycobacterium tuberculosis* and *M. bovis* isolates suggests convergent evolution. *BMC Microbiol*. 2007, **7**:39.

78. **Dos Vultos, T., Mestre, O., Tonjum, T. & Gicquel, B.** DNA repair in *Mycobacterium tuberculosis* revisited. *FEMS Microbiol. Rev*. 2009, **33**:471-487.

79. **Werngren, J. & Hoffner, S.E.** Drug-susceptible *Mycobacterium tuberculosis* Beijing genotype does not develop mutation-conferred resistance to rifampin at an elevated rate. *J. Clin. Microbiol*. 2003, **41**:1520-1524.

80. **Hillemann, D., Kubica, T., Agzamova, R., Venera, B., Rüsch-Gerdes, S. & Niemann, S**. Rifampicin and isoniazid resistance mutations in *Mycobacterium tuberculosis* strains isolated from patients in Kazakhstan. *Int. J. Tuberc. Lung Dis*. 2005, **9**:1161-1167.

81. **Mokrousov I, Otten T, Vyazovaya A, Limeschenko E, Filipenko ML, Sola C, Rastogi N, Steklova L, Vyshnevskiy B, Narvskaya O**. PCR-based methodology for detecting multidrug-resistant strains of *Mycobacterium tuberculosis* Beijing family circulating in Russia. *Eur J Clin Microbiol Infect Dis*. 2003, **22**(6):342-348.

82. **Qian, L., Abe, C., Lin, T., Yu, M., Cho, S., Wang, S. et al**. *rpo*B genotypes of *Mycobacterium tuberculosis* Beijing family isolates from East Asian countries. *J. Clin. Microbiol*. 2002, **40**:1091-1094.

83. **Tracevska, T., Jansone, I., Baumanis, V., Marga, O. & Lillebaek, T**. Prevalence of Beijing genotype in Latvian multidrug-resistant *Mycobacterium tuberculosis* isolates. *Int. J. Tuberc. Lung Dis*. 2003, **7**:1097-1103.

84. **Baker, L.V., Brown, T.J., Maxwell, O., Gibson, A.L., Fang, Z., Yates, M.D. et al.** Molecular analysis of isoniazid-resistant *Mycobacterium tuberculosis* isolates from England and Wales reveals the phylogenetic

significance of the ahpC-46A polymorphism. *Antimicrob. Agents Chemother*. 2005, **49**:1455-1464.

85. **Gagneux, S**. Fitness cost of drug resistance in *Mycobacterium tuberculosis*. *Clin. Microbiol. Infect*. 2009, Jan; 15 Suppl **1**:66-68. Review.

86. **Gagneux, S., Long, C.D., Small, P.M., Van, T., Schoolnik, G.K. & Bohannan, B.J.M**. The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis. Science* 2006, **312**:1944-1946.

87. **van der Spuy, G.D., Kremer, K., Ndabambi, S.L., Beyers, N., Dunbar, R., Marais, B.J. et al.** Changing *Mycobacterium tuberculosis* population highlights clade-specific pathogenic characteristics. *Tuberculosis* (Edinb) 2009, **89**:120-125.

88. **Nettleman, M.D**. Multidrug-resistant tuberculosis: news from the front. *JAMA* 2005, **293**:2788-2790.

89. **Gagneux, S., Burgos, M.V., DeRiemer, K., Encisco, A., Muñoz, S., Hopewell, P.C. et al.** Impact of bacterial genetics on the transmission of isoniazid-resistant *Mycobacterium tuberculosis*. *PLoS Pathog*. 2006, **2**:e61.

90. **Theus, S.A., Cave, M.D., Eisenach, K., Walrath, J., Lee, H., Mackay, W. et al**. Differences in the growth of paired Ugandan isolates of *Mycobacterium tuberculosis* within human mononuclear phagocytes correlate with epidemiological evidence of strain virulence. *Infect. Immun.* 2006, **74**:6865-6876.

91. **Newton, S.M., Smith, R.J., Wilkinson, K.A., Nicol, M.P., Garton, N.J., Staples, K.J. et al**. A deletion defining a common Asian lineage of *Mycobacterium tuberculosis* associates with immune subversion. *Proc. Natl. Acad. Sci*. U.S.A. 2006, **103**:15594-15598.

92. **Kato-Maeda, M., Rhee, J.T., Gingeras, T.R., Salamon, H., Drenkow, J., Smittipat, N. et al**. Comparing genomes within the species *Mycobacterium tuberculosis*. *Genome Res.* 2001, **11:**547-554.

93. **Kong, Y., Cave, M.D., Zhang, L., Foxman, B., Marrs, C.F., Bates, J.H. et al.** Association between *Mycobacterium tuberculosis* Beijing/W lineage strain infection and extrathoracic tuberculosis: Insights from epidemiologic and clinical characterization of the three principal genetic groups of *M. tuberculosis* clinical isolates. *J. Clin. Microbiol*. 2007, **45**:409-414.

94. **Lombard, Z., A. E. Brune, E. G. Hoal, C. Babb, P. D. van Helden, J. T. Epplen, and L. Bornman** HLA class II disease associations in southern Africa. *Tissue Antigens* 2006, **67**:97-110.

95. **Zhang, M., Gong, J., Yang, Z., Samten, B., Cave, M.D. & Barnes, P.F**. Enhanced capacity of a widespread strain of *Mycobacterium tuberculosis* to grow in human macrophages. *J. Infect. Dis*. 1999, **179**:1213-1217.

96. **Manicheva, O.A., Lasunskaia, E.B., Zhuravlev, V.I., Otten, T.F., Barnaulov, A.O., Mokrousov, I.V. et al.** Drug sensitivity in Mycobacterium tuberculosis versus its viability, cytotoxicity, genotype, and the course of the process in patients with pulmonary tuberculosis. *Probl Tuberk Bolezn Legk*, 2008, 18-22.

97. **Chacón-Salinas, R., Serafín-López, J., Ramos-Payán, R., Méndez-Aragón, P., Hernández-Pando, R., Van Soolingen, D. et al**. Differential pattern of cytokine expression by macrophages infected in vitro with different *Mycobacterium tuberculosis* genotypes. *Clin. Exp. Immunol*. 2005, **140**:443-449.

98. **Li Q, Whalen CC, Albert JM, Larkin R, Zukowski L, Cave MD, Silver RF.** Differences in rate and variability of intracellular growth of a panel of *Mycobacterium tuberculosis* clinical isolates within a human monocyte model. *Infect. Immun*. 2002, **70**(11):6489-6493.

99. **Theus, S.A., Cave, M.D. & Eisenach, K.D.** Activated THP-1 cells: an attractive model for the assessment of intracellular growth rates of *Mycobacterium tuberculosis* isolates. *Infect. Immun*. 2004, **72**:1169-1173.

100. **Theus, S., Eisenach, K., Fomukong, N., Silver, R.F. & Cave, M.D**. Beijing family *Mycobacterium tuberculosis* strains differ in their intracellular growth in THP-1 macrophages. *Int. J. Tuberc. Lung Dis*. 2007, **11**:1087-1093.

101. **Engele, M., Stössel, E., Castiglione, K., Schwerdtner, N., Wagner, M., Bölcskei, P. et al**. Induction of TNF in human alveolar macrophages as a potential evasion mechanism of virulent *Mycobacterium tuberculosis. J. Immunol.* 2002, **168**:1328-1337.

102. **Manca, C., Reed, M.B., Freeman, S., Mathema, B., Kreiswirth, B., Barry, C.E. 3rd et al**. Differential monocyte activation underlies strain-specific *Mycobacterium tuberculosis* pathogenesis. *Infect. Immun*. 2004, **72**:5511-5514.

103. **Wong, K.C., Leong, W.M., Law, H.K.W., Ip, K.F., Lam, J.T.H., Yuen, K.Y. et al**. Molecular characterization of clinical isolates of *Mycobacterium tuberculosis* and their association with phenotypic virulence in human macrophages. *Clin. Vaccine Immunol*. 2007, **14**:1279-1284.

104. **Dormans, J., Burger, M., Aguilar, D., Hernandez-Pando, R., Kremer, K., Roholl, P. et al.** Correlation of virulence, lung pathology, bacterial load and delayed type hypersensitivity responses after infection with different *Mycobacterium tuberculosis* genotypes in a BALB/c mouse model. *Clin. Exp. Immunol.* 2004, **137**:460-468.

105. **Manca, C., Tsenova, L., Bergtold, A., Freeman, S., Tovey, M., Musser, J.M. et al.** Virulence of a *Mycobacterium tuberculosis* clinical isolate in mice is determined by failure to induce Th1 type immunity and is associated with induction of IFN-alpha/beta. *Proc. Natl. Acad. Sci.* U.S.A. 2001, **98**:5752-5757.

106. **Roberts, T., Beyers, N., Aguirre, A. & Walzl, G**. Immunosuppression during active tuberculosis is characterized by decreased interferon-gamma production and CD25 expression with elevated forkhead box P3, transforming growth factor- beta , and interleukin-4 mRNA levels. *J. Infect. Dis.* 2007, **195**:870-878.

107. **Palanisamy, G.S., Smith, E.E., Shanley, C.A., Ordway, D.J., Orme, I.M. & Basaraba, R.J.** Disseminated disease severity as a measure of virulence of *Mycobacterium tuberculosis* in the guinea pig model. *Tuberculosis* (Edinb) 2008, **88**:295-306.

108. **Barczak, A.K., Domenech, P., Boshoff, H.I.M., Reed, M.B., Manca, C., Kaplan, G. et al.** In vivo phenotypic dominance in mouse mixed infections with *Mycobacterium tuberculosis* clinical isolates. *J. Infect. Dis*. 2005, **192**:600-606.

109. **Tsenova, L., Ellison, E., Harbacheuski, R., Moreira, A.L., Kurepina, N., Reed, M.B. et al.** Virulence of selected *Mycobacterium tuberculosis* clinical isolates in the rabbit model of meningitis is dependent on phenolic glycolipid produced by the bacilli. *J. Infect. Dis*. 2005, **192**:98-106.

110. **Ehrt, S. & Schnappinger, D**. *Mycobacterium tuberculosis* virulence: lipids inside and out. *Nat. Med*. 2007, **13**:284-285.

111. **Sun, Y., Lee, A.S.G., Wong, S. & Paton, N.I.** Association of *Mycobacterium tuberculosis* Beijing genotype with tuberculosis relapse in Singapore. *Epidemiol. Infect*. 2006, **134**:329-332.

112. **McShane, H**. Susceptibility to tuberculosis--the importance of the pathogen as well as the host. *Clin. Exp. Immunol*. 2003, **133:**20-21.

113. **Ordway, D., Henao-Tamayo, M., Harton, M., Palanisamy, G., Troudt, J., Shanley, C. et al.** The hypervirulent *Mycobacterium tuberculosis* strain HN878 induces a potent TH1 response followed by rapid down-regulation. *J. Immunol*. 2007, **179**:522-531.

114. **Constant, P., Perez, E., Malaga, W., Lanéelle, M., Saurel, O., Daffé, M. et al.** Role of the pks15/1 gene in the biosynthesis of phenolglycolipids in the *Mycobacterium tuberculosis* complex. Evidence that all strains synthesize glycosylated p-hydroxybenzoic methyl esters and that strains devoid of phenolglycolipids harbor a frameshift mutation in the pks15/1 gene. *J. Biol. Chem*. 2002, **277**:38148-38158.

115. **Sinsimer, D., Huet, G., Manca, C., Tsenova, L., Koo, M., Kurepina, N. et al**. The phenolic glycolipid of *Mycobacterium tuberculosis* differentially modulates the early host cytokine response but does not in itself confer hypervirulence. *Infect. Immun*. 2008, **76**:3027-3036.

116. **Cantrell, S.A., Pascopella, L., Flood, J., Crane, C.M., Kendall, L.V. & Riley, L.W.** Community-wide transmission of a strain of *Mycobacterium tuberculosis* that causes reduced lung pathology in mice. *J. Med. Microbiol*. 2008, **57**:21-27.

117. **Chaiprasert, A., Yorsangsukkamol, J., Prammananan, T., Palittapongarnpim, P., Leechawengwong, M. & Dhiraputra, C**. Intact pks15/1 in non-W-Beijing *Mycobacterium tuberculosis* isolates. *Emerging Infect. Dis*. 2006, **12**:772-774.

118. **Martinez-Gamboa, A., Ponce-de-Leon, A., Galindo-Fraga, A., Bobadilla-del-Valle, M., Kato-Maeda, M., Robertson, B.D. et al**. Molecular analysis of *Mycobacterium tuberculosis* strains with an intact pks15/1 gene in a rural community of Mexico. *Arch. Med. Res.* 2008, **39**:809-814.

119. **Rindi, L., Peroni, I., Lari, N., Bonanni, D., Tortoli, E. & Garzelli, C.** Variation of the expression of *Mycobacterium tuberculosis* ppe44 gene among clinical isolates. FEMS *Immunol. Med. Microbiol*. 2007, **51**:381-387.

120. **Sherman, D.R., Voskuil, M., Schnappinger, D., Liao, R., Harrell, M.I. & Schoolnik, G.K**. Regulation of the *Mycobacterium tuberculosis* hypoxic response gene encoding alpha -crystallin. *Proc. Natl. Acad. Sci*. U.S.A. 2001, **98**:7534-7539.

121. **Stewart, G.R., Snewin, V.A., Walzl, G., Hussell, T., Tormay, P., O'Gaora, P. et al**. Overexpression of heat-shock proteins reduces survival of *Mycobacterium tuberculosis* in the chronic phase of infection. *Nat. Med*. 2001, **7**:732-737.

122. **Wu, S., Howard, S. T., Lakey, D. L., Kipnis, A., Samten, B., Safi, H., Gruppo, V., Wizel, B., Shams, H. & et. al.** The principal sigma factor sigA

mediates enhanced growth of *Mycobacterium tuberculosis in vivo*. *Mol Microbiol* 2004, **51**:1551–1562.

123. **Wu, S., Barnes, P.F., Samten, B., Pang, X., Rodrigue, S., Ghanny, S. et al.** Activation of the eis gene in a W-Beijing strain of *Mycobacterium tuberculosis* correlates with increased SigA levels and enhanced intracellular growth. *Microbiology* 2009, **155**:1272-1281.

124. **Samuel, L.P., Song, C., Wei, J., Roberts, E.A., Dahl, J.L., Barry, C.E. 3rd et al.** Expression, production and release of the Eis protein by *Mycobacterium tuberculosis* during infection of macrophages and its effect on cytokine secretion. *Microbiology* 2007, **153**:529-540.

125. **Drobniewski F, Balabanova Y, Nikolayevsky V, Ruddy M, Kuznetzov S, Zakharova S, Melentyev A, Fedorin I**. Drug-resistant tuberculosis, clinical virulence, and the dominance of the Beijing strain family in Russia. *JAMA* 2005, **22**:2726-2731.

126. **Borgdorff, M.W., Van Deutekom, H., De Haas, P.E.W., Kremer, K. & Van Soolingen, D**. *Mycobacterium tuberculosis*, Beijing genotype strains not associated with radiological presentation of pulmonary tuberculosis. *Tuberculosis* (Edinb) 2004, **84**:337-340.

127. **Thwaites, G., Caws, M., Chau, T.T.H., D'Sa, A., Lan, N.T.N., Huyen, M.N.T. et al.** Relationship between *Mycobacterium tuberculosis* genotype and the clinical phenotype of pulmonary and meningeal tuberculosis. *J. Clin. Microbiol*. 2008, **46**:1363-1368.

128. **Feng, J., Su, W., Tsai, C. & Chang, S**. Clinical impact of *mycobacterium tuberculosis* W-Beijing genotype strain infection on aged patients in Taiwan. *J. Clin. Microbiol*. 2008, **46**:3127-3129.

129. **de Jong, B.C., Hill, P.C., Aiken, A., Awine, T., Antonio, M., Adetifa, I.M. et al**. Progression to active tuberculosis, but not transmission, varies by *Mycobacterium tuberculosis* lineage in The Gambia. *J. Infect. Dis*. 2008, **198**:1037-1043.

130. **Sapozhnikova, N.V., Skvortsova, L.A., Pavlova, M.V., Vishnevskiĭ, B.I., Otten, T.F., Vasil'eva, S.N. et al**. Pulmonary tuberculosis caused by *Mycobacterium tuberculosis* of different genotypes. *Probl Tuberk Bolezn Legk*, 2003, 13-15.

131. **Sun, Y., Lim, T.K., Ong, A.K.Y., Ho, B.C.H., Seah, G.T. & Paton, N.I.** Tuberculosis associated with *Mycobacterium tuberculosis* Beijing and non-Beijing genotypes: a clinical and immunological comparison. *BMC Infect. Dis*. 2006, **6**:105.

132. **Lillebaek, T., Andersen, A.B., Dirksen, A., Glynn, J.R. & Kremer, K**. *Mycobacterium tuberculosis* Beijing genotype. *Emerging Infect. Dis*. 2003, **9**:1553-1557.

133. **Kubica, T., Rüsch-Gerdes, S. & Niemann, S**. The Beijing genotype is emerging among multidrug-resistant *Mycobacterium tuberculosis* strains from Germany. *Int. J. Tuberc. Lung Dis*. 2004, **8**:1107-1113.

134. **Borgdorff, M.W., de Haas, P., Kremer, K. & van Soolingen, D.** *Mycobacterium tuberculosis* Beijing genotype, the Netherlands. *Emerging Infect. Dis*. 2003, **9**:1310- 1313.

135. **Moss AR, Alland D, Telzak E, Hewlett D Jr, Sharp V, Chiliade P, LaBombardi V, Kabus D, Hanna B, Palumbo L, Brudney K, Weltman A, Stoeckle K, Chirgwin K, Simberkoff M, Moghazeh S, Eisner W, Lutfey M, Kreiswirth B**. A city-wide outbreak of a multiple-drug-resistant strain of *Mycobacterium tuberculosis* in New York. *Int J Tuberc Lung Dis*. 1997, **2**:115-121.

136. **Jou, R., Chiang, C. & Huang, W**. Distribution of the Beijing family genotypes of *Mycobacterium tuberculosis* in Taiwan. *J. Clin. Microbiol*. 2005, **43**:95-100.137.**Li, W.M., Wang, S.M., Li, C.Y., Liu, Y.H., Shen, G.M., Zhang, X.X. et al.** Molecular epidemiology of *Mycobacterium tuberculosis* in China: a nationwide random survey in 2000. *Int. J. Tuberc. Lung Dis*. 2005, **9**:1314-1319.

138. **Dale, J.W., Bothamley, G.H., Drobniewski, F., Gillespie, S.H., McHugh, T.D. & Pitman, R**. Origins and properties of *Mycobacterium tuberculosis* isolates in London. *J. Med. Microbiol.* 2005, **54**:575-582.

139. **Chan, M.Y., Borgdorff, M., Yip, C.W., de Haas, P.E., Wong, W.S., Kam, K.M. et al**. Seventy percent of the *Mycobacterium tuberculosis* isolates in Hong Kong represent the Beijing genotype. *Epidemiol. Infect*. 2001, **127**:169-171.

140. **Prodinger, W.M., Bunyaratvej, P., Prachaktam, R. & Pavlic, M**. *Mycobacterium tuberculosis* isolates of Beijing genotype in Thailand. *Emerging Infect. Dis*. 2001, **7**:483-484.

141. **Palittapongarnpim P, Luangsook P, Tansuphaswadikul S, Chuchottaworn C, Prachaktam R, Sathapatayavongs B** Restriction fragment length polymorphism study of *Mycobacterium tuberculosis* in Thailand using IS6110 as probe. *Int J Tuberc Lung Dis*. 1997, **4**:370-376.

142. **Krüüner, A., Hoffner, S.E., Sillastu, H., Danilovits, M., Levina, K., Svenson, S.B. et al**. Spread of drug-resistant pulmonary tuberculosis in Estonia. *J. Clin. Microbiol.* 2001, **39**:3339-3345.

143. **Toungoussova, O.S., Sandven, P., Mariandyshev, A.O., Nizovtseva, N.I., Bjune, G. & Caugant, D.A**. Spread of drug-resistant *Mycobacterium tuberculosis* strains of the Beijing genotype in the Archangel Oblast, Russia. *J. Clin. Microbiol*. 2002, **40**:1930-1937.

144. **Glynn, J.R., Crampin, A.C., Traore, H., Yates, M.D., Mwaungulu, F.D., Ngwira, B.M. et al**. *Mycobacterium tuberculosis* Beijing genotype, northern Malawi. *Emerging Infect. Dis*. 2005, **11**:150-153.

145. **Hasan Z, Tanveer M, Kanji A, Hasan Q, Ghebremichael S, Hasan R** Spoligotyping of *Mycobacterium tuberculosis* isolates from Pakistan reveals predominance of Central Asian Strain 1 and Beijing isolates.. *J Clin Microbiol*. 2006, **5**:1763-1768.

146. **Nicol, M.P., Sola, C., February, B., Rastogi, N., Steyn, L. & Wilkinson, R.J.** Distribution of strain families of *Mycobacterium tuberculosis* causing pulmonary and extrapulmonary disease in hospitalized children in Cape Town, South Africa. *J. Clin. Microbiol*. 2005, **43**:5779-5781.

147. **Cox HS, Kubica T, Doshetov D, Kebede Y, Rüsch-Gerdess S, Niemann S** The Beijing genotype and drug resistant tuberculosis in the Aral Sea region of Central Asia. *Respir Res*. 2005, **6**:134.

Chapter 2

A recently evolved sublineage of the *Mycobacterium tuberculosis* Beijing strain family is associated with an increased ability to spread and cause disease

Hanekom, M., van der Spuy, G. D., Streicher, E. M., Ndabambi, S. L., McEvoy, C. R. E., Kidd, M., Beyers, N., Victor, T. C., van Helden, P. D. and Warren R.M.

J.Clin.Microbiol. (2007). 45:1483-1490

Abstract

This study aimed to reconstruct the evolutionary history of Beijing strains of *Mycobacterium tuberculosis* and to test the hypothesis that evolution has influenced the ability of the Beijing strains within the different Beijing sublineages to spread and cause disease. A PCR-based method was used to analyze the genome structure of 40 different loci in 325 Beijing isolates collected from new and retreatment tuberculosis patients from an urban setting and 270 Beijing isolates collected from high-risk tuberculosis patients from a rural setting in the Western Cape, South Africa. The resulting data were subjected to phylogenetic analysis using the neighbor joining algorithm. Phylogenetic reconstructions were highly congruent with the "gold standard" phylogenetic tree based on synonymous single-nucleotide polymorphisms, thereby allowing a prediction of the order in which the evolutionary events had occurred. A total of seven independently evolving Beijing sublineages were identified. Analysis of epidemiological data in relation to the Beijing sublineage suggested an association between recent evolutionary change and frequency of occurrence in an urban population ($p < 0.001$) as well as in the rural population ($p < 0.001$). This concept was further supported by an association between more recently evolved Beijing strains and an increased ability to transmit and to cause disease (odds ratio, 5.82; 95% confidence interval, 3.13 to 10.82 [$p < 0.001$]). An association between Beijing sublineage and demographic and clinical parameters and drug resistance could not be demonstrated. From these data, we suggest that the pathogenic characteristics of Beijing strains are not conserved but rather that strains within individual lineages have evolved unique pathogenic characteristics.

Introduction

From molecular epidemiological data, it has been hypothesized that Beijing strains of *Mycobacterium tuberculosis* have evolved unique properties (10), including the abilities to evade the protective effect of *Mycobacterium bovis* BCG vaccination (23), to disseminate more efficiently than non-Beijing strains (4), and to acquire drug resistance more frequently as the consequence of single-nucleotide polymorphisms (SNPs) in the mismatch repair genes (15). In some studies, disease caused by a Beijing strain has been associated with treatment relapse (16), while in

another study a febrile response was noted during the earlier stages of treatment of tuberculosis caused by a Beijing strain (21). However, studies of chest radiography data have not been able to consistently demonstrate differences between the pathogenic characteristics of Beijing and non-Beijing strains (5, 7, 17). It is currently not known whether these results can be explained by different pathogenic characteristics among different Beijing strains, as a recent study has shown that extrathoracic tuberculosis was associated with only a subset of Beijing strains (12).

M. tuberculosis strains with a Beijing genotype are members of principal genetic group 1 and are characterized by a distinct spoligotype pattern, closely related IS*6110* DNA fingerprints, identical variable-number tandem repeat sequences, and an IS*6110* insertion in the NTF 1 region (Rv0001:Rv0002) (10). Comparative genomics have shown that Beijing strains have evolved through mechanisms including IS*6110* transposition (2), SNPs in the mismatch repair genes (15), deletion of chromosomal domains (regions of difference [RDs]) (19), and synonymous SNP(sSNPs) (9). An evolutionary scenario based on mutations in the mismatch repair genes *mutT2* and *mutT4* and *ogt* genes has suggested the absence of these mutations in the ancestral Beijing strains (15). Subsequent mutations in the *mutT4* gene (codon 48) and *ogt* gene (codon 37) or *ogt* and *mutT2* genes (codons 12 and 58, respectively) generated two independently evolving subgroups (15). Mapping of the RD domains showed that the Beijing strain family is a monophyletic clade, with four distinct evolutionary branches defined by the RD105, RD181, RD150, and RD142 deletions, respectively (19). Phylogenetic analysis according to sSNPs supported the monophyletic clade structure and suggested 11 Beijing sequence types (STs) (9). However, the order in which all of the above evolutionary events occurred remained to be determined.

In this study we aimed to reconstruct the evolutionary history of Beijing strains from the Western Cape, South Africa, through the analysis of the genomic structure of 40 different loci. This phylogeny was used to test the hypothesis that evolution has influenced the ability of the Beijing strains within the different Beijing sublineages to spread and cause disease.

Material and methods

Urban setting

As part of an ongoing molecular epidemiological study, an attempt was made to collect sputum samples at diagnosis from all new and retreatment tuberculosis cases who attended primary health care clinics and who were resident in an epidemiological field site in Cape Town, Western Cape, South Africa (3), during the period January 1993 to December 2004. Sputum smear microscopy (fluorescent staining) and/or culture (BACTEC 460, MGIT 960, or Löwenstein-Jensen medium) was performed by the National Health Laboratory Service or the Stellenbosch University laboratories to confirm the presence of *M. tuberculosis*. Clinical and demographic data including previous history of tuberculosis, gender and age, smear positivity, and drug susceptibility test results (if requested) were recorded in a database. Chest radiography was not done routinely and therefore was not included in this study. A case was considered as smear positive if any sputum sample, collected within the first 2 months after diagnosis, was found to be smear-positive. Human immunodeficiency virus (HIV) testing was not routinely done in the initial years of this study although a recent survey of 366 new adult smear-positive tuberculosis cases (2000 to 2002) in this epidemiological field site showed that 10% were HIV positive.

Rural setting

Sputum isolates were routinely collected at diagnosis from all high-risk tuberculosis patients (patients whose sputum failed to convert after the intensive phase of therapy or patients who had previously been treated with antituberculosis drugs) who were attending primary health care clinics and who were resident in the Boland, Overberg, Karoo, and Southern Cape region of the Western Cape, South Africa, during the period February 2003 to August 2004. This region is located 150 to 500 km northeast of the epidemiological field site in Cape Town. Sputum smear microscopy (fluorescent staining) and/or culture (BACTEC 460, MGIT 960, or Löwenstein-Jensen medium) was performed by the National Health Laboratory Service or the Stellenbosch University laboratories to confirm the presence of *M. tuberculosis*. Clinical and demographic data including previous history of

tuberculosis, gender and age, smear positivity, and drug susceptibility test results were recorded in a database. A survey of 384 tuberculosis retreatment patients (2003 to 2004) in this study setting showed that 13% were HIV coinfected. This study setting was included to enable a comparison of the *M. tuberculosis* strain population structures in two independent communities. This study was approved by the Ethics Committee (Institutional Review Board) of Stellenbosch University.

Drug susceptibility testing

Drug susceptibility testing was done by the National Health Laboratory Service, using the indirect proportion method on Middlebrook medium containing critical concentrations of 0.2 µg/ml isoniazid and 30 µg/ml rifampin. In this study drug resistance was defined as resistance to either isoniazid or rifampin or a combination of isoniazid and rifampin (multidrug-resistant [MDR] tuberculosis). Drug susceptibility testing for other resistance markers was not routinely done.

Genotypic classification

Sputum isolates from patients resident in the urban epidemiological field site were subcultured on Löwenstein-Jensen medium and genotyped by IS*6110* DNA fingerprinting (22) and spoligotyping according to internationally standardized methods (11). Sputum isolates from patients resident in the rural setting were cultured on BACTEC 12B medium, and a boiled aliquot was genotyped by spoligotyping according to internationally standardized methods (11). Subcultures of the isolates from the rural setting were not available for IS*6110* DNA fingerprinting. Isolates were classified as members of the Beijing genotype if they had the characteristic Beijing spoligotype (10). Only the first *M. tuberculosis* isolate from each case with a Beijing strain was included for subsequent genetic analysis.

Genomic comparison

Each Beijing strain from each patient was subjected to PCR amplification in a reaction mixture containing 0.2 µg DNA template, 5 µl Q buffer, 2.5 µl 10x buffer, 2 µl 25 mM $MgCl_2$, 4 µl 10 mM deoxynucleoside triphosphates, 1 µl of each primer (50 pmol/µl) (Table 1), and 0.125 µl HotStarTaq DNA polymerase (QIAGEN,

Germany) and made up to 25 µl with H_2O. Primers 1 to 22 (Table 1) were used in combination with the universal forward primer and the internal control primers to determine the presence or absence of IS*6110* elements at specific chromosomal loci (2). Primer sets 23 to 26 (Table 1) was used to determine the presence or absence of SNPs in the *mutT2* (codon 58), *mutT4* (codon 48), and *ogt* (codons 12 and 37) genes (15) by the amplification refractory mutation system method. Primer sets 27 to 30 (Table 1) was used to determine whether RDs (RD105, RD181, RD150, and RD142) (19) were present or absent. Primer sets 31 to 40 (Table 1) was used to determine the presence or absence of sSNPs at chromosomal positions 797736, 909166, 1477596, 1548149, 1692069, 1892017, 2376135, 2532616, 2825581, and 4137829 (9) relative to the H37Rv whole-genome sequence (6) by the amplification refractory mutation system method. Amplification was initiated by incubation at 95°C for 15 min, followed by 35 to 45 cycles at 94°C for 1 min, annealing temperature (Table 1) for 1 min, and 72°C for 1 min. After the last cycle, the samples were incubated at 72°C for 10 min. PCR amplification products were electrophoretically fractionated in 3.0% agarose in 1x Tris-buffered EDTA, pH 8.3, at 6 V/cm for 4 h and visualized by staining with ethidium bromide. The existence of a mutational event was determined by the presence or absence and the size of the respective PCR product.

Table 1: Primer sequences

Primer Set	Primer Name [e]	Sequence (5' to 3')	T_m (°C) [f]
	Internal control	TCC CAG TGA CGT TGC CTT C	62
	Internal control	GAG CAG CAG TGG AAT TTC GC	62
	Universal Forward	TTC AAC CAT CGC CGC CTC TAC	62
1	IS21-1(1262963)[a]	GTC GCC GGA GTT GAA GAA GCT GAA CC	62
2	IS21-2 (absent)[a]	CGA TCA ATG TTC CGC CTA ATT GAA CC	62
3	IS21-3 (3797825)[a]	AAT GCA GAC GAC GCG ACG ATT GAA CC	62

4	IS21-4 (1592)[a]	CAG CGA CAC TCA CAG CCA ATT GAA CC	62
5	IS21-5 (1543972)[a]	CTG CAC CGC GCG CAA CGA GGT GAA CC	62
6	IS21-6 (2263627)[a]	TCC AGG CAC CAG CAT CAA GGT GAA CC	62
7	IS21-7 (2038898)[a]	CGT TTG TGG GTG TCC GGT ATT GAA CC	62
8	IS21-8 (2367677)[a]	TGT CGC CAG TTA CGC ACG AGT GAA CC	62
9	IS21-9 (3549199)[a]	CGC GGA GCC GTC GGC CGC GGT GAA CC	62
10	IS21-10 (888990)[a]	TGA GGG GGT GTT GAG GTT GGT GAA CC	62
11	IS21-11 (3379025)[a]	GGG TTT TAA AAA AGT CGC TGT GAA CC	62
12	IS21-12 (3127932)[a]	GAT GGC ACG GCC GAC CTG AAT GAA CC	62
13	IS21-13 (3547342)[a]	GAC CGC GGT TGG GTG GAC ATT GAA CC	62
14	IS21-14 (3378553)[a]	CCC GCG CCT GCG CAA TTG GCT GAA CC	62
15	IS21-15 (3844681) [a]	GCG ATG TGG ATT GTG TCG GGT GAA CC	62
16	IS21-16 (1986638)[a]	ATC TGT TCA TCT CCG ACC TGT GAA CC	62
17	IS21-17 (absent)[a]	CGC GCG GTT CGA ATA CGG TGT GAA CCC	62
18	IS21-18 (2163649)[a]	TAC CAG CGA CGT TAA ACG GAT GAA CC	62
19	IS21-19 (3493910)[a]	GAA CCC CGG CTC CGC CTC GAT GAA CC	62
20	IS21-20 (1657015)[a]	GAG GCT CCT TTC GAC CGC GTT GAA CC	62
21	IS21-21 (2366892)[a]	GCG TGA TGT GCA CCA TAG TGT GAA CC	62
22	IS21-22 (2634022) [a]	GGT GCG GGG TCG GGG CCG TTT GAA CC	62
23	*mut*T2(58)[b] wt F	AGA GCT CGC CGA AGA ACC GG	70
23	*mut*T2(58)[b] mut F	AGA GCT CGC CGA AGA ACC GC	70
23	*mut*T2(58[b]) R	AAG CAG ATG CAC GCG ATA GG	70

24	*mut*T4(48)[b] F	AGC CGA GAA TCA CAT GGA CG	68
24	*mut*T4(48)[b] wt R	CGA GGT GAG CGG GAT CG	68
24	*mut*T4(48)[b] mut R	CGA GGT GAG CGG GAT CC	68
25	*ogt*(12)[b] F	CCG CAG GAG AAG ATC GCA T	68
25	*ogt*(12)[b] wt R	GCC CGG CCA GGG TTA ATA GC	68
25	*ogt*(12)[b] mut R	GCC CGG CCA GGG TTA ATA GT	68
26	*ogt*(37)[b] F	CCA TCG GGC CAT TAA CCC T	66
26	*ogt*(37)[b] wt R	TCG GGT GTC CAG TGT GCG C	66
26	*ogt*(37)[b] mut R	TCG GGT GTC CAG TGT GCG A	66
27	RD105[c] F	ACA GCG CGG GTC ATA TCA C	62
27	RD105[c] int	GCA ACA CCC GCT TGT CTT TG	62
27	RD105[c] R	AAC CAG CTC CTC GAC GCT ATC	62
28	RD142[c] F	CCG GTG GTA CGG GTA TTT CC	62
28	RD142[c] int	GCT CGA GCA TGA TCA GCA AAG	62
28	RD142[c] R	TAG CAC CAG TAC CGG ATG TCC	62
29	RD150[c] F	AGT GCT GGC AAT AGC GGT TG	62
29	RD150[c] int	CAC CGG CAC TTA CCA TCT CG	62
29	RD150[c] R	CCA GCA CTT GTT GCA ACT TCG	62
30	RD181[c] F	AAA TCC GCC CAT ACC CGT C	62
30	RD181[c] R	AGC TTC GAC TGG CCA TAG GC	62
31	797736[d] wt F	CAG CTC ATC TGT TGA TGT TC	66
31	797736[d] mut F	CAG CTC ATC TGT TGA TGT TT	66

31	797736[d] R	ACG CCA TAA GCA CCT TCA CAC	66
32	909166[d] wt F	CAC GTT TGA CCG GAT CCC GC	69
32	909166[d] mut F	CAC GTT TGA CCG GAT CCC GT	69
32	909166[d] R	CGT CCC AAG GAG CAG TCA AG	69
33	1477596[d] wt F	GCC CGG CCA GGG TTA ATA GC	69
33	1477596[d] mut F	GCC CGG CCA GGG TTA ATA GT	69
33	1477596[d] R	CCG GTA AAC CAA TGG CCA C	69
34	1548149[d] F	CGA CGT GAC ATG GCT GGA TT	68
34	1548149[d] wt R	CCG ACA ACG TCA GCG ATA CC	68
34	1548149[d] mut R	CCG ACA ACG TCA GCG ATA CT	68
35	1692069[d] F	CGA CCC GTG ATC GCA TGT A	69
35	1692069[d] wt R	CAC CGA GCT CAC CGG CCT TT	69
35	1692069[d] mut R	CAC CGA GCT CAC CGG CCT TC	69
36	1892017[d] wt F	CCA CGT TTC TTG ATG CCT AT	66
36	1892017[d] mut F	CCA CGT TTC TTG ATG CCT AC	66
36	1892017[d] R	TAT CGA GGC CGA CGA AAG G	66
37	2376135[d] wt F	GTT GAT GTA TAT CGC GGA CA	66
37	2376135[d] mut F	GTT GAT GTA TAT CGC GGA CG	66
37	2376135[d] R	GCC GCC GAA TTA GAA CAG C	66
38	2532616[d] F	CTG CTT CGA CAC CTT TAA CGC	72
38	2532616[d] wt R	ATG CCC AAC GCC GCA GTG GGC	72
38	2532616[d] mut R	ATG CCC AAC GCC GCA GTG GGT	72

39	2825581[d] wt F	TGA CGG TCG GAT TCT TGG GT	68
39	2825581[d] mut F	TGA CGG TCG GAT TCT TGG GG	68
39	2825581[d] R	AGC CCA CGA GAT ACT GAG CG	68
40	4137829[d] wt F	AGA TGG CCT ACC GGA TCA CC	69
40	4137829[d] mut F	AGA TGG CCT ACC GGA TCA CT	69
40	4137829[d] R	GAC GCA GTC GCA ACA GTT CAC	69

[a] chromosomal position of IS*6110* insertion sites (2).

[b] codon positions of SNPs in the mismatch repair genes (15)

[c] regions of difference (RD) (19)

[d] chromosomal positions of sSNPs (9)

[e] F – forward primer; R – reverse primer; int – internal primer; wt – wild-type and mut - mutant

[f] T_m, annealing temperature

Phylogenetic analysis

The evolutionary state(s) for the SNPs in the mismatch repair genes and the 11 Beijing ST sSNPs was assigned according to the DNA sequence of H37Rv (6). IS*6110* insertion sites were assigned according to the absence or presence of an IS*6110* element at a specific chromosomal site (IS*6110* insertions at positions 3547342 and 3378553 were excluded from the final analysis as they showed a high degree of homoplasy). The RDs were assigned according to their absence or presence. The complete set of evolutionary states for the different mutational events or subsets thereof was subjected to phylogenetic analysis using the neighbor joining algorithm (PAUP* 4.0; Phylogenetic Analysis Using Parsimony [*Other Methods], version 4b10; Sinauer Associates, Sunderland, MA). Bootstrapping was performed to establish a degree of statistical support for branch nodes (8). A consensus tree rooted to *M. bovis* was generated using the program Contree (PAUP* 4.0). Only branches which occurred in >50% of the bootstrap trees were included in the final tree. Each branch was defined as a Beijing sublineage.

Statistical analysis

The *z* test for proportions was used for testing frequency of occurrence of cases in the respective Beijing sublineages. The Fisher exact test was used to determine the association between Beijing sublineage and gender, new versus retreatment cases, smear positivity within 2 months of diagnosis, drug resistance (for patients resident in either the urban or the rural setting), and IS*6110* clustering (urban setting only). In all cases, a *P* value of 0.05 was used as the cutoff level for significance.

Results

Genotyping showed that 325/1,525 patients (21%) from the urban epidemiological field site in Cape Town and 270/904 patients (30%) from the rural setting in the Boland, Overberg, Karoo, and Southern Cape regions of the Western Cape had tuberculosis with an *M. tuberculosis* Beijing strain. Drug susceptibility testing showed that of the patients with Beijing strains, 50 (15%) from the urban setting and 63 (23%) from the rural setting had drug-resistant tuberculosis.

Phylogenetic analysis according to the neighbor joining algorithm grouped the Beijing strains from the two settings into four STs, ST11, ST19, ST22, and a novel ST (Beijing sublineages 6 and 7) (Figure 1, left). The evolutionary history of these STs was further resolved according to SNPs in the mismatch repair genes, mutations induced by IS*6110* insertion, and the structure of the RDs (Figure 1, right). This phylogenetic tree was statistically robust, and the predicted evolutionary order was congruent with the tree based only on sSNPs (compare left and right sides of Figure 1). This phylogeny demonstrates that the Beijing genotype is a monophyletic clade, defined by the RD105 deletion; IS*6110* insertions at chromosome positions 1592, 1986638, 3127932, 3549199, and 3844681; and an sSNP at position 1548149 (Figure 1, right). Subsequent evolution divided the Beijing strains into seven Beijing sublineages (Figure 1, right) according to RDs (19), sSNPs (9), SNPs in the mismatch repair genes (15), and IS*6110* insertions (2). The evolutionary order suggests that the RD181 deletion and the *mutT4* (codon 48) SNP occurred before ST19, while the SNPs in the *mutT2* and *ogt* mismatch repair genes occurred after ST19 and before ST10. Beijing sublineages 3 to 6 were distinguished by a

combination of specific IS*6110* insertions and/or sSNPs, while Beijing sublineage 7 was characterized by the RD150 deletion. The overall topology of the tree suggests divergent evolution, and this is supported by the increasing complexity in the IS*6110* banding patterns between Beijing sublineage 1 (closest to the ancestor) and Beijing sublineage 7 (most distant from the ancestor) isolates from the urban setting (Figure 2).

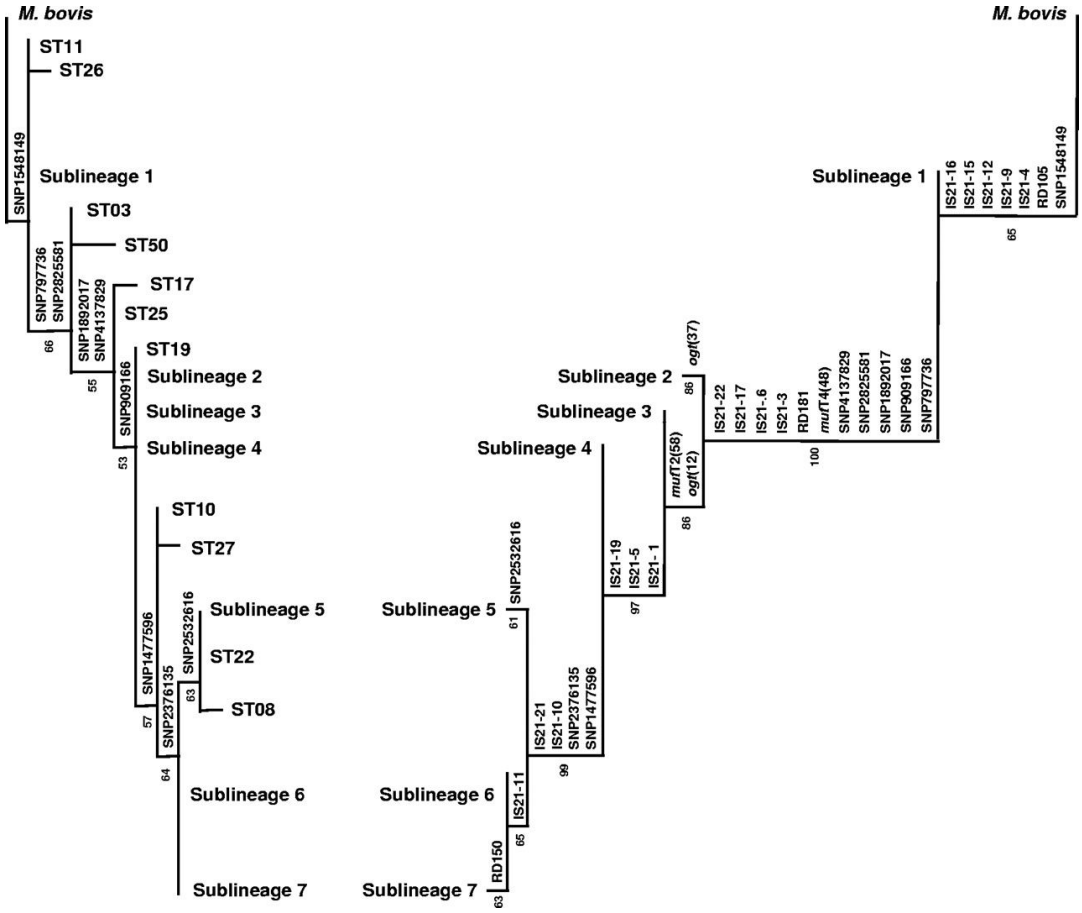


Figure 1: The evolutionary history of *M. tuberculosis* Beijing strains from an urban and a rural setting in the Western Cape, South Africa. (Left) Phylogenetic reconstruction based on sSNPs (9) using the bootstrapping and neighbor joining algorithms (PAUP* 4.0). The tree was rooted to *M. bovis*. Bootstrap values are given at internal nodes. The chromosomal positions of the predominant sSNPs are given at the nodes where they occur. All branches with a zero length were collapsed. (Right) Phylogenetic reconstruction based on sSNPs (9), SNPs in mismatch repair genes (15), IS*6110* insertion points (2), and RDs (19) using the bootstrapping and neighbor joining algorithms (PAUP* 4.0). The tree was rooted to *M. bovis*. Bootstrap values are given at internal nodes. The respective evolutionary events are indicated at nodes where they occur (the order in which these events occurred is unknown). All branches with a zero length were collapsed. The scale indicates the number of steps per unit length.

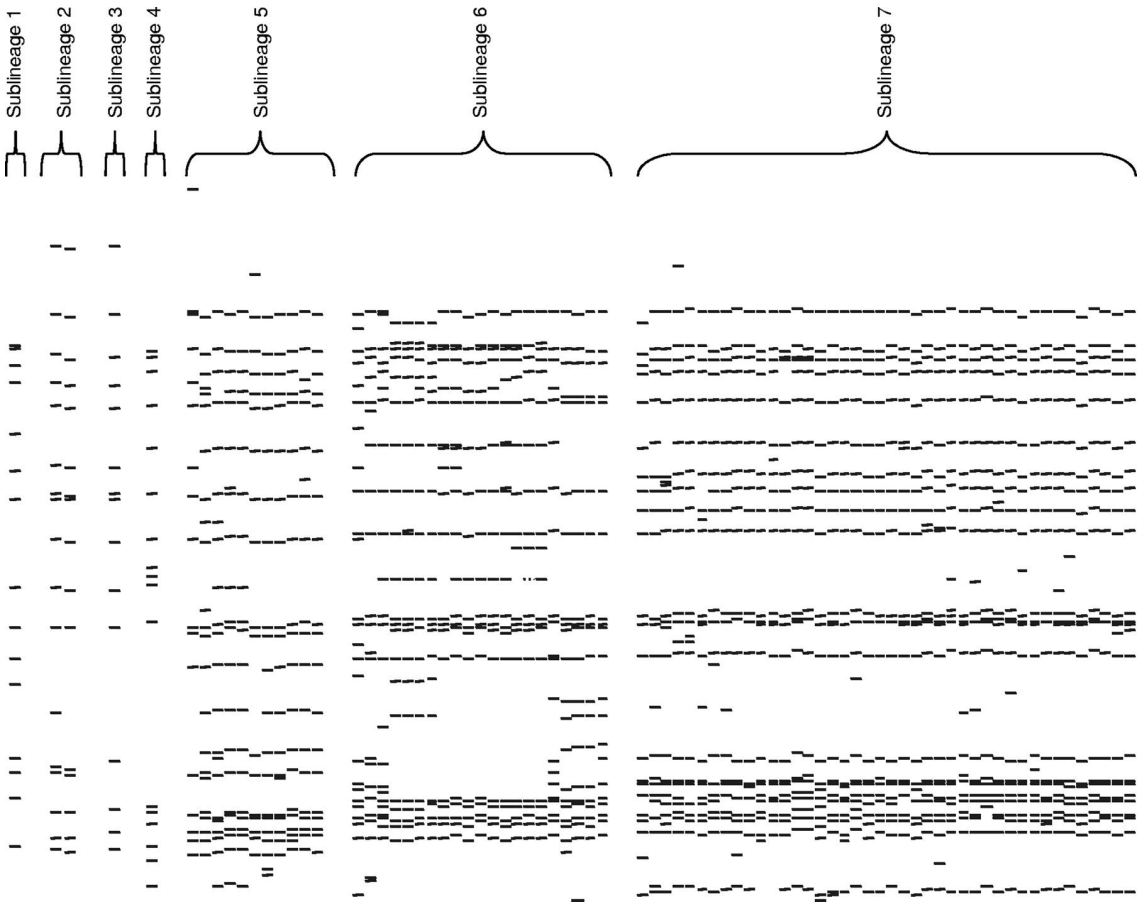


Figure 2: IS*6110* DNA fingerprints of *M. tuberculosis* Beijing strains from the urban setting. *M. tuberculosis* isolates were classified by IS*6110* DNA fingerprinting according to the internationally standardized method (22). The DNA fingerprints from representative isolates are grouped according to their Beijing sublineage. Only a single representative of each IS*6110* DNA fingerprint is shown. Beijing strains with STs ST03, ST08, ST10, ST17, ST50, ST25, ST26, and ST27; strains with the RD142 deletion; and strains with IS*6110* insertions at positions 2038898, 2163649, and 2367677 were absent from this study setting although Beijing strains with these evolutionary events have been described elsewhere (2, 9, 15, 19). In order to determine whether the evolutionary process had altered the ability of drug-sensitive Beijing strains to spread and cause disease, the frequency of occurrence of strains from the more recently evolved Beijing sublineage 7 was compared to that of strains belonging to the more distantly evolved Beijing sublineages 2 to 6.

Table 2: Clinical, demographic, and molecular epidemiological data from patients resident in the urban epidemiological field site in Cape Town, South Africa

	Number of drug sensitive cases (%)							Number of drug resistant cases (%)						
Beijing types	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Total number of	0	4	2	1	7	34	227	1	0	0	0	33	8	8
Male	-	2 (50)	2 (100)	-	4 (57)	22	144 (63)	1 (100)	-	-	-	15 (45)	6 (75)	3 (37)
Average Age (years)	-	38.5	35.3	22	36.9	34.5	35	25.7	-	-	-	36.3	36.7	35.9
New cases	-	2 (50)	1 (50)	1 (100)	3 (43)	23	136 (60)	-	-	-	-	15 (45)	2 (25)	7 (87)
Smear negative	-	-	-	-	2 (29)	7 (20)	56 (24)	-	-	-	-	9 (27)	3 (38)	3 (38)
Smear positive	-	3 (75)	1 (50)	1 (100)	4 (57)	24	156 (70)	1 (100)	-	-	-	22 (67)	5 (62)	5 (62)
Smear unknown	-	1 (25)	1 (50)	-	1 (14)	3 (9)	15 (6)	-	-	-	-	2 (6)	-	-
INH mono-resistant	-	-	-	-	-	-	-	-	-	-	-	-	7 (87)	2 (25)
RIF mono-resistant	-	-	-	-	-	-	-	1 (100)	-	-	-	-	1 (13)	-
MDR	-	-	-	-	-	-	-	-	-	-	-	33 (100)	-	6 (75)
Number of strains	-	2	1	1	6	18	44	1	-	-	-	6	5	5
Number of unique	-	-	-	1	5	12	31	1	-	-	-	4	4	4
No of clusters	-	2	1	-	1	6	13	-	-	-	-	2	1	1
Cases in clusters (%)	-	4 (100)	2 (100)	-	2 (29)	22	196 (86)	-	-	-	-	29 (88)	4 (50)	4 (50)

Table 3: Clinical and demographic data from patients resident in the rural setting of the Boland, Overberg, Karoo, and Southern Cape region of the Western Cape, South Africa

	Number of drug sensitive cases (%)							Number of drug resistant cases (%)						
Beijing Types	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Total number of patients	0	0	0	1	7	68	131	3	2	3	1	5	30	19
Male	-	-	-	-	4 (57)	42 (62)	72 (55)	2 (67)	-	1 (33)	1 (100)	3 (60)	20 (67)	14 (74)
Average Age (years)	-	-	-	29	36.3	27.7	34.5	29.7	23.5	31	39	34.4	32.1	29.3
New case	-	-	-	1 (100)	-	22 (32)	45 (34)	na	na	na	na	na	na	na
Smear negative	-	-	-	1 (100)	2 (29)	37 (55)	58 (44)	2 (67)	1 (50)	1 (33)	-	2 (40)	13 (43)	9 (47)
Smear positive	-	-	-	-	5 (71)	24 (35)	54 (41)	-	-	1 (33)	-	1 (20)	12 (40)	8 (42)
Smear unknown	-	-	-	-	-	7 (10)	19 (15)	1 (33)	1 (50)	1 (33)	-	2 (40)	5 (17)	2 (11)
INH mono-resistant								-	-	1 (33)	-	-	9 (30)	6 (32)
RIF mono-resistant								-	-	1 (33)	-	-	1 (3)	1 (5)
MDR								3 (100)	2 (100)	1 (33)	1 (100)	5 (100)	20 (67)	12 (63)

This analysis showed an overabundance of Beijing sublineage 7 strains in the urban population (Table 2) ($p < 0.001$ from the z test for the hypothesis that proportion of sublineage 7 cases = 0.14) as well as in the rural population (Table 3) ($p < 0.001$ from the z test for the hypothesis that proportion of sublineage 7 cases = 0.14). In addition, strain clustering according to IS*6110* DNA fingerprinting in the urban community was strongly associated with the Beijing sublineage 7 (Table 2) (odds ratio, 5.82; 95% confidence interval, 3.13 to 10.82 [$p < 0.001$]) (A similar analysis could not be done for isolates from the rural setting as cultures were not available for DNA fingerprinting). Together, this suggests that strains from sublineage 7 have evolved a phenotype which allows them to transmit and cause disease more frequently than strains from sublineages 2 to 6.

To determine whether the frequency of occurrence of drug-susceptible strains could be explained by either a demographic or a clinical phenomenon, the demographic and clinical data from patients (in both settings) were compared according to their strains' Beijing sublineage. No association could be shown between Beijing sublineage and age, gender, previous history of tuberculosis, and smear positivity within the first 2 months following diagnosis.

To determine whether evolution has altered the ability of the Beijing sublineages to acquire drug resistance, phenotypic drug susceptibility data were compared according to the respective strains' Beijing sublineage. The results from the 50 patients resident in the urban setting showed no association between the Beijing sublineage and the number of strains with drug resistance (Table 2). However, a high level of transmission of resistant strains was seen in Beijing sublineage 5, which was associated with MDR tuberculosis (Table 2). Similarly, the high number of drug-resistant tuberculosis cases seen with Beijing sublineages 6 and 7 in the rural setting was associated with MDR tuberculosis (Table 3).

In this study we have used genetic information to predict evolutionary relationships between strains within the Beijing evolutionary lineage in order to test the hypothesis that the evolutionary process has altered the ability of these strains to spread and cause disease. According to the phylogenetic prediction we demonstrate

that the Beijing strain family is a monophyletic clade which could be divided into seven sublineages. The predicted evolutionary order was congruent with the "gold standard" tree based on sSNPs (9). The observation that many of the evolutionary events have also been described for Beijing isolates from different geographical regions suggests that these evolutionary events occurred in the distant past and that the resulting progeny (including both distantly and more recently evolved clones) were subsequently imported into South Africa (13).

Discussion

In this study, cases with a sublineage 7 strain were over-represented, suggesting that the evolutionary process has had a positive influence on the strain's ability to spread and cause disease. This concept was further supported by an association between sublineage 7 and transmissibility. Similarly, in a previous study, a significant association was observed between extrathoracic tuberculosis and a subset of Beijing strains characterized by either RD142 or RD150 deletion (12). Together these data support the hypothesis that phenotypic differences may exist within the Beijing strain family (7, 19). Molecular epidemiological studies also support the notion that the more recently evolved strains (termed typical Beijing strains) are adapted to spread and cause disease, given their frequency of occurrence, in comparison to distantly evolved strains (termed atypical Beijing strains) (14, 18). However, factors including social behavior which could lead to the spread of sublineage 7 strains were not investigated in this study. An alternative explanation for the high frequency of occurrence of the Beijing sublineage 7 strains in South Africa could be a strong founder effect. However, comparison between Beijing genotypes from South Africa and East Asia showed that at least nine different Beijing founder strains were introduced into South Africa (data not shown). The spectrum of founder strains is likely to reflect that in East Asia at the time of importation into South Africa since the East Asia immigrants were derived from many different locations throughout the region. Thus, it is probable that the subsequent spread of sublineage 7 in South Africa represents evolutionary selection of an individual sublineage rather than a founder effect.

The observation of a changing epidemiological picture concurrent with the accumulation of evolutionary events cannot be linked to the sSNPs analyzed as these events are thought to be largely neutral (9). Conversely, mutation in the mismatch repair genes (15), IS*6110* integration (2), and chromosomal deletions (19) may influence the strains' ability to spread and cause disease by altering or eliminating gene function. Mutation in the mismatch repair genes has been hypothesized to enhance the ability of the Beijing strains to acquire drug resistance (15). IS*6110* insertion is thought to influence the characteristics of the Beijing strains by altering the function of genes by either causing knockouts or disrupting or enhancing promoter activity (2). The accumulation of chromosomal deletions has been suggested to contribute to an altered virulence in the Beijing strain family (19), and in a previous study, the analysis of 100 *M. tuberculosis* strains suggested an association between deletion and an altered phenotype (20). In our study the most pathogenic strains, as measured by their frequency of occurrence, were associated with the accumulation of deletions, including RD150. The true extent of chromosomal evolution within the Beijing strain family almost certainly extends beyond the evolutionary events tested in this study, and the influence of such events on the phenotype of different strains remains to be determined.

Our study design differs from previous studies as we analyzed intrastrain family phenotypic-genotypic correlations compared to interstrain family phenotypic-genotypic correlations (5, 7, 16, 17, 21). In our study, the analysis of patient demographic and clinical data failed to demonstrate an association with any of the Beijing strain sublineages. This implies that the appearance of secondary cases occurs independently of the proportion of infectious (smear-positive) cases in the different sublineages, thereby suggesting that the characteristics of the strains from the different sublineages could be the key factors influencing the ability to spread and cause disease.

We noted the presence of drug resistance in all of the Beijing sublineages and noted that the frequency of occurrence of the drug resistance phenotype was not associated with the most recently evolved Beijing sublineage. The absence of spread of drug resistance in the most recently evolved Beijing sublineage could be explained by the

fitness cost as a result of the acquisition of mutations in the genes conferring drug resistance (1). The high number of resistant cases in sublineage 5 (urban setting) and sublineages 6 and 7 (rural settings) was associated with MDR tuberculosis and clearly indicates that MDR tuberculosis can spread within communities if not managed and treated appropriately.

In summary, we suggest that the pathogenic characteristics of *M. tuberculosis* strains are not conserved within defined strain families but rather that individual lineages within strain families have evolved unique pathogenic characteristics.

Acknowledgments

TB in the 21st Century Consortium, NRF (2054201), and the Harry Crossley Foundation and the European Commission 6th Framework Program on Research Technological Development Demonstration (project no. 037919) are thanked for financially supporting this project. We thank the nurses and data analysts for the collection and processing of clinical data. We thank I. Toms, Department of Health, City of Cape Town, and are indebted to the residents of the epidemiological field site.

Reference List

1. **Andersson, D. I.** 2006. The biological cost of mutational antibiotic resistance: any practical conclusions? Curr.Opin.Microbiol. **9**:461-465.

2. **Beggs, M. L., K. D. Eisenach, and M. D. Cave**. 2000. Mapping of IS*6110* insertion sites in two epidemic strains of mycobacterium tuberculosis. J.Clin.Microbiol. **38**:2923-2928.

3. **Beyers, N., R. P. Gie, H. L. Zietsman, M. Kunneke, J. Hauman, M. Tatley, and P. R. Donald**. 1996. The use of a geographical information system (GIS) to evaluate the distribution of tuberculosis in a high-incidence community. S.Afr.Med.J. **86**:40-1, 44.

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

5. **Borgdorff, M. W., H. van Deutekom, P. E. de Haas, K. Kremer, and D. van Soolingen**. 2004. *Mycobacterium tuberculosis*, Beijing genotype strains not associated with radiological presentation of pulmonary tuberculosis. Tuberculosis.(Edinb.) **84**:337-340.

6. **Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, III, 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, B. G. Barrell, and a. et**. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature **393**:537-544.

7. **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**:2726-2731.

8. **Felsenstein, J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**:783-793.

9. **Filliol, I., A. S. Motiwala, M. Cavatore, W. Qi, M. H. Hazbon, d. Bobadilla, V, J. Fyfe, L. Garcia-Garcia, N. Rastogi, C. Sola, T. Zozio, M. I. Guerrero, C. I. Leon, J. Crabtree, S. Angiuoli, K. D. Eisenach, R. Durmaz, M. L. Joloba, A. Rendon, J. Sifuentes-Osornio, d. L. Ponce, M. D. Cave, R. Fleischmann, T. S. Whittam, and D. Alland**. 2006. Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. J.Bacteriol. **188**:759-772.

10. **Glynn, J. R., J. Whiteley, P. J. Bifani, K. Kremer, and D. van Soolingen**. 2002. Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: a systematic review. Emerg.Infect.Dis. **8**:843-849.

11. **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**:907-914.

12. **Kong, Y., M. Cave, L. Zhang, B. Foxman, C. Marrs, J. Bates, and Z. Yang**. 2006. A Population-Based Study of Deletions in Five Different Genomic Regions of *Mycobacterium tuberculosis* and Their Possible Clinical Relevance. J.Clin.Microbiol.

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

14. **Mokrousov, I., O. Narvskaya, T. Otten, A. Vyazovaya, E. Limeschenko, L. Steklova, and B. Vyshnevskyi**. 2002. Phylogenetic reconstruction within *Mycobacterium tuberculosis* Beijing genotype in northwestern Russia. Res.Microbiol. **153**:629-637.

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

16. **Sun, Y. J., A. S. Lee, S. Y. Wong, and N. I. Paton**. 2006. Association of *Mycobacterium tuberculosis* Beijing genotype with tuberculosis relapse in Singapore. Epidemiol.Infect. **134**:329-332.

17. **Sun, Y. J., T. K. Lim, A. K. Ong, B. C. Ho, G. T. Seah, and N. I. Paton**. 2006. Tuberculosis associated with *Mycobacterium tuberculosis* Beijing and non-

Beijing genotypes: A clinical and immunological comparison. BMC.Infect.Dis. **6**:105.

18. **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**:1930-1937.

19. **Tsolaki, A. G., S. Gagneux, A. S. Pym, Goguet de la Salmoniere YO, B. N. Kreiswirth, D. van Soolingen, and P. M. Small**. 2005. Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis.* J.Clin.Microbiol. **43**:3185-3191.

20. **Tsolaki, A. G., A. E. Hirsh, K. DeRiemer, J. A. Enciso, M. Z. Wong, M. Hannan, Goguet de la Salmoniere YO, 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**:4865-4870.

21. **van Crevel, R., R. H. Nelwan, W. de Lenne, Y. Veeraragu, A. G. van der Zanden, Z. Amin, J. W. van der Meer, and D. van Soolingen.** 2001. *Mycobacterium tuberculosis* Beijing genotype strains associated with febrile response to treatment. Emerg.Infect.Dis. 7:880-883.

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

23. **van Soolingen, D., L. Qian, P. E. de Haas, J. T. Douglas, H. Traore, F. Portaels, H. Z. Qing, D. Enkhsaikan, P. Nymadawa, and J. D. van Embden**. 1995. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. J.Clin.Microbiol. **33**:3234-3238.

Spread of a low-fitness drug-resistant *Mycobacterium tuberculosis* strain in a setting of high human immunodeficiency virus prevalence

Strauss, O. J., Warren, R. M., Jordaan, A., Streicher, E. M., Hanekom, M., Falmer, A. A., Albert, H., Trollip, A., Hoosain, E., van Helden, P. D., Victor, T.C.

J.Clin.Microbiol. (2008). 46:1514-1516

Abstract

The fitness cost associated with the evolution of resistance to rifampin in *Mycobacterium tuberculosis* may be different in clinical isolates compared to *in vitro*-generated mutants. An atypical Beijing strain (attenuated phenotype) demonstrated the ability to spread despite acquiring resistance to rifampin. Transmission was linked to human immunodeficiency virus co-infection ($p = 0.029$), raising concern for the spread of drug resistance in vulnerable populations.

Text

The fitness of *Mycobacterium tuberculosis* strains circulating in a community is possibly the driving force perpetuating the tuberculosis epidemic. This is particularly true for the spread of drug resistance, since it has been suggested that the evolution of drug resistance has a fitness cost resulting in the overall attenuation of the pathogen (1). This phenomenon has been demonstrated during the spontaneous *in vitro* evolution of resistance to rifampin, where a direct correlation was observed between the frequency at which a specific non-synonymous single nucleotide polymorphism (nsSNP) occurred and the fitness of the mutant clone (4). That study showed that rarely observed nsSNPs had a high fitness cost, whereas frequently observed nsSNPs had a low fitness cost. These results correlated well with the frequency of nsSNPs observed in rifampin-resistant clinical isolates (4).

Molecular epidemiological studies have suggested that certain *M. tuberculosis* genotypes are more successful than others (3, 5). Accordingly, it has been hypothesized that the Beijing genotype is a high fitness genotype, possibly as a result of the evolution of unique properties (2, 11, 16). Strains with the Beijing genotype can be broadly grouped as typical or atypical according to the presence or absence, respectively, of an IS*6110* insertion in the NTF region of the *M. tuberculosis* genome (9). Phylogenetic analysis has provided evidence that these two genotypes are derived from a common progenitor (6); however, they demonstrate vastly different epidemiological characteristics, since strains with the atypical Beijing genotype are only rarely observed (9, 14). This has prompted speculation that atypical Beijing strains have lower fitness than typical Beijing strains. We hypothesize that rifampin resistance causing nsSNPs that have a fitness cost would only be rarely observed in

M. tuberculosis strains with an atypical Beijing genotype unless epidemiological factors favouring their spread were present.

To test this hypothesis, sputum specimens were collected from retreatment cases attending health care clinics or referral hospitals in two provinces in South Africa. Each specimen was subjected to routine culture-based drug susceptibility testing for isoniazid and rifampin. During the period from January 2001 to October 2004, 312 rifampin-resistant cases were identified from the Western Cape (WC) region (12), while during the period from September 2003 to May 2004, 117 rifampin resistant cases were identified from the Eastern Cape (EC) region. DNA sequence analysis of the rifampin resistance determining region (RRDR) of the *rpoB* gene (13) showed that >90% of rifampin-resistant isolates had an nsSNP in the RRDR region. Of the 30 nsSNPs identified, 25 nsSNPs appeared at a frequency consistent with frequencies reported in in vitro-generated rifampin-resistant mutants (4, 7, 10) (Table 1). The frequency of appearance of the remaining five nsSNPs was discordant compared to the *in vitro*-generated rifampin resistant mutants (Table 1). The nsSNPs at codons 516 (GAC→GTC and GAC→TAC) and 533 (CTG→CCG) were significantly under-represented in the *in vitro*-generated rifampin resistant mutants, while nsSNPs at codon 522 (TCG→TTG) and codon 526 (CAC→CGC) were significantly over-represented in the *in vitro*-generated rifampin-resistant mutants (Table 1). This suggests that the nsSNPs at codons 516 (GAC→GTC and GAC→TAC) and 533 (CTG→CCG) had a lower fitness cost in clinical isolates compared to in vitro-generated rifampin resistant mutants. Conversely, the nsSNPs at codons 522 (TCG→TTG) and 526 (CAC→CGC) appear to have a high fitness cost in clinical isolates.

Table 1: Distribution of nsSNPs conferring resistance to rifampin in clinical isolates from the WC and EC regions of South Africa[a]

Codon	WT		Mut	*In vitro* RIF^R studies combined[a] (n=368)	Western Cape RIF^R				Eastern Cape RIF^R			
					All (n=312)	Non-Beijing (n=196)	Beijing (n=116)		All (n=117)	Non-Beijing (n=58)	Beijing (n=59)	
							Typical (n=101)	Atypical (n=15)			Typical (n=14)	Atypical (n=45)
490	CAG	→	CAA		1	1						
511	CTG	→	CCG		3		2	1	3		3	
512	CAA	→	AAA		3	3						
513	CAA	→	AAA	2	4	4						
	CAA	→	CCA		1			1				
	CAA	→	GAA	4								
	CAA	→	CTA	2								
516	GAC	→	GTC	3[b]	15	7	5	3	29	1	1	27
	GAC	→	TAC		15	13	1	1	6	2	1	3
	GAC	→	TTC		1	1						
	GAC	→	GGC		1		1					
519	AAC	→	AAG	1								
522	TCG	→	TTG	46[c]								
526	CAC	→	AAC		3	1	1	1	1	1		
	CAC	→	CTC	1	5	4	1		1		1	
	CAC	→	TAC	58	19	14	5		6	3	1	2
	CAC	→	GAC	18	8	7	1					
	CAC	→	GCC		2	2						
	CAC	→	TGC		1	1						
	CAC	→	CGC	57[c]	1	1						
	CAC	→	CCC	4								
529	CGA	→	CCA	1								

	CGA	→	CAA	1								
	CGA	→	GGA	1								
531	TCG	→	TGG	17	8	2	3	3	2	2		
	TCG	→	TTG	132	169	110	54	5	58	42	5	11
	TCG	→	CAG		3		3					
	TCG	→	TTC		1		1					
533	CTG	→	CCG	1[b]	15	5	10					
	CTG	→	GAG		2		2					
Multiple nsSNPs				2	1	1						
Insertions				1	1	1						
Deletions				15[d]	2	2						
nsSNPs absent from RRDR				1	27	16	11		11	7	2	2

[a] Combined data from (4,7,10)

[b] Under-represented in the *in vitro*-generated rifampicin-resistant mutants [z-test ($p < 0.001$)]

[c] Over-represented in the *in vitro*-generated rifampicin-resistant mutants [z-test ($p < 0.001$)]

[d] Eleven different deletion events (7,10)

WT = wild type codon sequence, Mut - mutant codon sequence, RIF^R - rifampicin-resistant, RRDR - rifampicin-resistance determining region

In order to determine whether a relationship existed between strain genotype and the nsSNPs conferring rifampin resistance, the isolates from this study were classified as either Beijing or non-Beijing by spoligotyping (8). The results showed that 116 (37%) of the rifampin-resistant cases from the WC region and 59 (50%) of the rifampin-resistant cases from the EC region were infected with a Beijing genotype strain (Table 1). Subclassification of the Beijing isolates as either typical or atypical (6) showed that the population structure of rifampin resistant Beijing strains was significantly different in the two study settings (Fisher exact test odds ratio = 21.6; 95% confidence interval = 9.6 to 48.6, $p < 0.0001$) (Table 1). The nsSNP at codon 516 (GAC→GTC) was associated with the atypical Beijing genotype from the EC (Fisher exact test odds ratio = 45; 95% confidence interval = 3.8 to 525, $p < 0.0008$), while the nsSNP at codon 533 was mostly found in isolates with the typical Beijing genotype from the WC (Table 1).

IS*6110* DNA fingerprinting (15) showed that isolates from the WC region with the typical Beijing genotype and either an nsSNP at codon 516 (GAC→GTC) or codon 533 (CTG→CCG) were not clustered (5 of 5 banding patterns and 9 of 10 banding patterns, respectively), thereby suggesting that these nsSNPs had evolved independently and that the resulting clones were not transmitted. In contrast, the isolates from the EC region with the atypical Beijing genotype and an nsSNP at codon 516 (GAC→GTC) were clustered and also shared the rare -17 *inhA* promoter mutation (G→T) (data not shown), suggesting ongoing transmission. These isolates clustered with the atypical Beijing strains from the WC which had an nsSNP at codon 516 (GAC→GTC) and the -17 *inhA* promoter mutation, suggesting interprovincial spread. This finding was contrary to previous reports, which suggested that atypical Beijing strains are attenuated in their ability to transmit (9, 14), while the mutation at codon 516 (GAC→GTC) would have been expected to further compromise the ability of these strains to transmit unless compensatory mutations were present or the epidemiological context allowed transmission to occur. Analysis of the host population in the EC region showed human immunodeficiency virus (HIV) coinfection to be a risk factor for the spread of the atypical Beijing strains (z-test for the hypothesis that a proportion of HIV-positive cases = 0.42, $p < 0.029$). In contrast, the frequency of atypical Beijing strains was low in the WC region, which in turn has

a low incidence of HIV-tuberculosis coinfection (6). This raises concern for the spread of all drug-resistant strains in vulnerable populations.

In summary, greater vigilance is required to contain the drug-resistant tuberculosis epidemic in high-HIV-prevalence settings. This can be achieved by the development and implementation of rapid diagnostics, the provision of appropriate therapy, ensuring treatment adherence, and intensified screening of contacts. However, in order for diagnosis and treatment to be effective, it is essential that communities are educated to improve health-seeking behavior.

Acknowledgments

We thank the South African National Research Foundation (grants GUN 2054278 and DST/NRF Centre of Excellence in Biomedical Tuberculosis Research), the Wellcome Trust (grant 072402/Z/03/Z), the NIH (grant R21 A155800-01), the IAEA (grant SAF6008), European Commission 6th framework program (project no. 037919), and the Harry Crossly Foundation for financial support. We thank Lindiwe Kani for reviewing the clinical files.

Reference List

1. **Andersson, D. I.** 2006. The biological cost of mutational antibiotic resistance: any practical conclusions? Curr.Opin.Microbiol. **9**:461-465.

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

3. **European Concerted Action on New Generation Genetic Markers and Techniques for the Epidemiology and Control of Tuberculosis**. 2006. Beijing/W genotype *Mycobacterium tuberculosis* and drug resistance. Emerg.Infect.Dis. **12**:736-743.

4. **Gagneux, S., C. D. Long, P. M. Small, T. Van, G. K. Schoolnik, and B. J. Bohannan**. 2006. The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. Science **312**:1944-1946.

5. **Glynn, J. R., J. Whiteley, P. J. Bifani, K. Kremer, and D. van Soolingen**. 2002. Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: a systematic review. Emerg.Infect.Dis. **8**:843-849.

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

7. **Huitric, E., J. Werngren, P. Jureen, and S. Hoffner**. 2006. Resistance levels and rpoB gene mutations among in vitro-selected rifampin-resistant *Mycobacterium tuberculosis* mutants. Antimicrob.Agents Chemother. **50**:2860-2862.

8. **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**:907-914.

9. **Mokrousov, I., O. Narvskaya, T. Otten, A. Vyazovaya, E. Limeschenko, L. Steklova, and B. Vyshnevskyi**. 2002. Phylogenetic reconstruction within *Mycobacterium tuberculosis* Beijing genotype in northwestern Russia. Res.Microbiol. **153**:629-637.

10. **Morlock, G. P., B. B. Plikaytis, and J. T. Crawford**. 2000. Characterization of spontaneous, In vitro-selected, rifampin-resistant mutants of *Mycobacterium tuberculosis* strain H37Rv. Antimicrob.Agents Chemother. **44**:3298-3301.

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

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

13. **Telenti, A., P. Imboden, F. Marchesi, D. Lowrie, S. Cole, M. J. Colston, L. Matter, K. Schopfer, and T. Bodmer**. 1993. Detection of rifampicin-resistance mutations in *Mycobacterium tuberculosis*. Lancet **341**:647-650.

14. **Toungoussova, O. S., A. Mariandyshev, G. Bjune, P. Sandven, and D. A. Caugant**. 2003. Molecular epidemiology and drug resistance of *Mycobacterium tuberculosis* isolates in the Archangel prison in Russia: predominance of the W-Beijing clone family. Clin.Infect.Dis. **37**:665-672.

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

16. **van Soolingen, D., L. Qian, P. E. de Haas, J. T. Douglas, H. Traore, F. Portaels, H. Z. Qing, D. Enkhsaikan, P. Nymadawa, and J. D. van Embden**. 1995. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. J.Clin.Microbiol. **33**:3234-3238.

Chapter 3

Evidence that the spread of *Mycobacterium tuberculosis* strains with the Beijing genotype is human population dependent

Hanekom, M., van der Spuy, G. D., Gey van Pittius, N. C., McEvoy, C. R. E., Ndabambi, S. L., Victor, T. C., Hoal, E. G., van Helden, P. D. and Warren, R. M.

J. Clin. Microbiol. (2007). 45:2263-2266

Abstract

This study describes a comparative analysis of the Beijing mycobacterial interspersed repetitive unit types of *Mycobacterium tuberculosis* isolates from Cape Town, South Africa, and East Asia. The results show a significant association between the frequency of occurrence of strains from defined Beijing sublineages and the human population from whom they were cultured ($p < 0.0001$).

Text

Mycobacterium tuberculosis strains with the Beijing genotype have been shown to be globally widespread and are particularly prevalent in East Asia, where over 80% of strains from the Beijing, China region are of this genotype (5). It has been hypothesized that Beijing strains have evolved unique properties, including the ability to evade the protective effect of *Mycobacterium bovis* BCG vaccination (19) and the ability to spread more efficiently than non-Beijing strains (2). However, the clinical presentations of patients with tuberculosis caused by a Beijing strain were found to vary between different geographical settings (3-5,16). Currently, it is not known whether the observed variability in clinical presentation is a function of the Beijing strain population found in particular geographical settings, is a function of the genetic composition of the human population, or is a function of a combination of these two variables.

This study aimed to test the hypothesis that host-pathogen compatibility determined the Beijing strain population structure in different host populations in different geographical settings. Cultures of *M. tuberculosis* isolates from patients of mixed ancestry (14) who were resident in Cape Town, South Africa (20) were classified as being of the Beijing genotype by spoligotyping (10). The Beijing strains were assigned to phylogenetic sublineages as described previously (8) and were genotyped by mycobacterial interspersed repetitive-unit (MIRU) typing (17).

During the study period (January 1993 to December 2004), 25 MIRU types were identified among 321 tuberculosis cases infected with a Beijing strain (Table 1). A comparison of the MIRU type data for the Beijing strains from Cape Town and previously published MIRU type data for the Beijing strains from East Asia (1, 9, 12,

13, 15, 18) showed that nine of the Beijing MIRU types (types MT01, MT08, MT11, MT18, MT19, MT21, MT28, MT33, and MT54) were shared between these geographical settings (Table 1). This suggests that the nine shared Beijing MIRU types represent founder strains that were introduced into Cape Town from East Asia, as the latter is thought to be the evolutionary origin of strains with a Beijing genotype (5, 7, 12). The definition of founder MIRU types was supported by their disproportionately high number (n = 267) compared to the number with non-founder MIRU types (n = 54) in tuberculosis patients from Cape Town (z test for the hypothesis that the proportion of types were founder MIRU types = 0.5; $p < 0.001$).

Superimposition of the Beijing MIRU type data onto the previously described phylogenetic tree of the Beijing strain family (8) provided a framework that could be used to predict the evolutionary order in which the 25 Beijing MIRU types had evolved (Figure 1). From this prediction, the Beijing MIRU types could be partitioned into seven Beijing sublineages. The number of founder Beijing MIRU types was variable among the different Beijing sublineages (Figure 1). Twenty-four of the Beijing MIRU types were unique to their respective sublineages, while the remaining Beijing MIRU type (MT11) was shared by three different sublineages (sublineages 2, 3, and 6) (Figure 1 and Table 1), suggesting that MT11 was an ancestral Beijing MIRU type (12).

Table 1: Geographical distribution of Beijing MIRU types from Asia and South Africa

MIRU type	Beijing sublineage[b]	Number of copies in polymorphic MIRU loci												Number of strains (%)						
		2	4	10	16	20	23	24	26	27	31	39	40	RUS[c,d]	CHN[d]	HK[e]	VNM[d]	SGP[f]	BGD[g]	CT-SA
MT01[a]	4	2	2	3	3	2	5	1	7	3	4	3	3	1 (2.2)	5 (3.8)	7 (3.3)			2 (16.7)	1 (0.3)
MT02	n/a	2	2	3	3	2	5	1	5	3	5	3	3	27 (60.0)	2 (1.5)	4 (1.9)		2 (3.6)	7 (58.3)	
MT04	n/a	2	2	3	2	2	5	1	5	3	5	3	3	1 (2.2)			1 (2.7)			
MT05	n/a	2	2	3	3	2	5	1	4	3	5	3	3	1 (2.2)						
MT07	n/a	2	2	3	3	2	5	1	5	3	6	3	3	2 (4.4)						
MT08[a]	6	2	2	3	2	2	5	1	7	3	5	3	3	1 (2.2)	2 (1.5)	3 (1.4)				2 (0.6)
MT09	n/a	2	2	3	3	2	5	1	7	3	5	3	1	1 (2.2)			1 (2.7)			
MT11[a]	2 3 6	2	2	3	3	2	5	1	7	3	5	3	3	10 (22.2)	42 (32.3)	77 (36.5)	10 (27.0)	39 (69.6)		3 (0.9) 2 (0.6) 8 (2.5)
MT12	n/a	2	2	1	3	2	5	1	7	3	5	3	3	1 (2.2)	1 (0.8)					
MT13	n/a	2	2	3	3	2	6	1	7	1	5	3	1			8 (3.8)	1 (2.7)			
MT14	n/a	2	2	3	3	2	5	1	7	3	6	3	3		2 (1.5)	3 (1.4)	1 (2.7)	2 (3.6)		
MT16	n/a	2	2	3	3	2	5	1	7	3	5	3	2		5 (3.8)	12 (5.7)		5 (8.9)		
MT17	n/a	2	2	3	4	2	5	1	7	3	5	3	3		1 (0.8)			1 (1.8)		
MT18[a]	6	2	2	3	3	2	5	1	7	3	5	4	3		2 (1.5)	5 (2.4)		2 (3.6)		7 (2.2)
MT19[a]	5	2	0	3	3	2	5	1	7	3	5	3	3			2 (0.9)	1 (2.7)			39 (12.1)
MT20	7	2	2	2	3	2	5	1	7	4	5	4	3							19 (5.9)
MT21[a]	6	2	2	3	3	2	5	1	8	3	5	3	3		4 (3.1)	6 (2.8)	2 (5.7)	4 (7.1)		13 (4.0)
MT25	7	2	2	2	3	2	5	1	7	3	3	4	3							18 (5.6)
MT26	6	2	2	3	3	2	5	1	7	3	5	4	4							1 (0.3)
MT27	7	2	2	2	3	2	5	1	3	3	5	4	3							1 (0.3)
MT28[a]	7	2	2	2	3	2	5	1	7	3	5	4	3		16 (12.3)	23 (10.9)	4 (10.8)			189 (58.9)
MT29	6	2	2	3	3	2	5	1	7	4	5	3	3							1 (0.3)
MT33[a]	6	n/a	2	3	3	n/a	n/a	n/a	6	3	5	3	3		11 (8.5)	4 (1.9)	12 (32.4)			3 (0.9)
MT37	n/a	n/a	2	3	3	n/a	n/a	n/a	2	3	5	3	3		2 (1.5)				1 (8.3)	

MT43	n/a	n/a	2	3	3	n/a	n/a	n/a	6	3	4	3	3					2 (16.7)	
MT44	n/a	n/a	2	2	3	n/a	n/a	n/a	7	3	5	3	3	6 (4.6)	9 (4.3)	1 (2.7)			
MT47	n/a	n/a	2	3	3	n/a	n/a	n/a	7	1	5	3	1	3 (2.3)	6 (2.8)	3 (8.1)			
MT48	n/a	n/a	2	3	3	n/a	n/a	n/a	4	3	4	3	3	2 (1.5)					
MT49	n/a	n/a	2	7	2	n/a	n/a	n/a	7	3	4	3	3	3 (2.3)					
MT50	n/a	n/a	2	5	2	n/a	n/a	n/a	7	3	4	3	3	2 (1.5)					
MT51	n/a	n/a	2	3	2	n/a	n/a	n/a	5	1	5	3	3	2 (1.5)					
MT52	n/a	n/a	2	3	3	n/a	n/a	n/a	6	3	5	3	2	2 (1.5)					
MT53	n/a	n/a	2	3	3	n/a	n/a	n/a	7	3	5	1	3	2 (1.5)					
MT54[a]	6	n/a	2	3	3	n/a	n/a	n/a	7	3	5	2	3	7 (5.4)	4 (1.9)				1 (0.3)
MT55	n/a	n/a	2	2	3	n/a	n/a	n/a	5	3	3	4	3	2 (1.5)					
MT56	n/a	n/a	2	2	3	n/a	n/a	n/a	9	3	5	4	3	2 (1.5)					
MT57	n/a	n/a	2	2	3	n/a	n/a	n/a	7	3	5	2	3	2 (1.5)					
MTSing76	n/a	2	2	6	2	2	5	1	7	3	4	3	3				1 (1.8)		
MTZAF1	1	2	2	3	3	2	5	1	6	3	5	3	1						1 (0.3)
MTZAF2	5	2	3	3	3	2	5	1	7	3	5	3	3						1 (0.3)
MTZAF3	6	2	2	3	3	2	5	1	7	2	5	3	3						3 (0.9)
MTZAF4	6	2	2	3	3	2	5	1	5	3	5	3	3						1 (0.3)
MTZAF5	6	2	2	4	3	2	6	1	5	3	3	2	3						1 (0.3)
MTZAF6	7	2	2	2	3	2	5	1	7	4	4	4	3						1 (0.3)
MTZAF7	7	2	2	2	3	2	5	1	8	4	5	4	3						1 (0.3)
MTZAF8	7	2	2	2	3	2	5	1	9	3	5	3	3						1 (0.3)
MTZAF9	7	2	2	2	3	2	5	1	7	3	4	4	3						1 (0.3)
MTZAF10	7	2	2	2	3	2	5	1	5	3	5	4	3						1 (0.3)
MTZAF11	7	2	2	4	3	2	5	1	5	3	3	2	3						1 (0.3)
MTHK1	n/a	2	2	3	3	2	5	1	7	3	5	3	4		6 (2.8)				
MTHK2	n/a	2	2	3	3	2	5	1	7	3	5	5	3		5 (2.4)				
MTHK3	n/a	2	2	3	4	2	5	1	7	3	5	3	2		5 (2.4)				
MTHK4	n/a	2	1	3	3	2	5	1	7	3	5	3	3		4 (1.9)				
MTHK5	n/a	2	2	3	3	2	5	1	6	3	5	3	4		3 (1.4)				
MTHK6	n/a	2	0	3	3	2	5	1	7	3	5	3	4		3 (1.4)				
MTHK7	n/a	2	2	3	3	2	5	1	7	3	5	4	2		2 (0.9)				
MTHK8	n/a	2	2	3	3	2	5	1	8	3	5	5	3		2 (0.9)				
MTHK9	n/a	2	2	2	3	2	5	1	7	3	5	3	2		2 (0.9)				
MTHK10	n/a	2	2	3	3	2	5	1	6	3	7	3	3		2 (0.9)				
MTHK11	n/a	2	2	3	3	2	2	1	7	4	4	3	3		2 (0.9)				
MTHK12	n/a	2	2	2	3	2	5	1	7	3	5	4	4		2 (0.9)				

RUS – Russia; CHN – China; HK – Hong Kong; VNM – Vietnam; SGP – Singapore; BGD – Bangladesh; CT-SA – Cape Town South Africa.

n/a – not available

[a] founder MIRU types

[b] according to (8)

[c] according to (13)

[d] according to (12)

[e] according to (9)

[f] according to (15)

[g] according to (1)

Figure 1

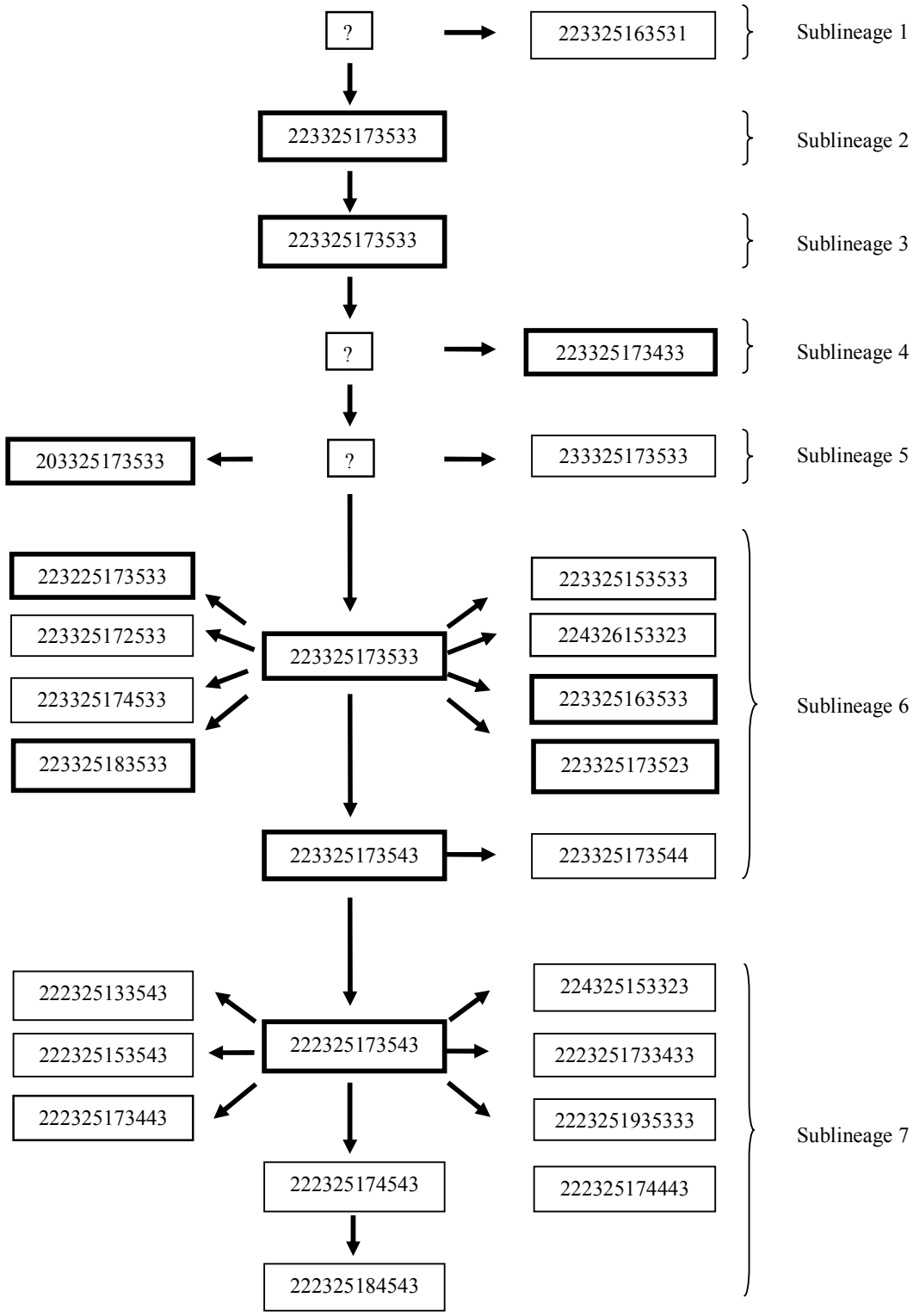


Figure 1: Evolutionary scenario of Beijing MIRU types according to Beijing sublineage. Beijing MIRU types were grouped according to their respective Beijing sublineages (8), and the most

parsimonious evolutionary order was proposed. The Beijing MIRU types are indicated within each box. Founder Beijing MIRU types are indicated by bold boxes. Unknown Beijing MIRU types are indicated by "?."

To determine the propensity of Beijing strains from the different sublineages to spread in the human population in Cape Town, the number of cases in circulation within each sublineage was compared to the number of founder strains for that sublineage (Table 1). The number of representatives of these founder strains was shown to be overrepresented in sublineage 7 (n = 233 cases from one founder strain) compared with the numbers in sublineages 1 to 6 (n = 88 cases from eight founder strains) (z test for the hypothesis that the proportion of sublineage 7 cases = 0.11; $p < 0.001$). In comparison, founder Beijing MIRU types MT01, MT08, MT11, MT18, MT19, MT21, MT33, and MT54 were overrepresented in the human population in East Asia (China, 73/130; Hong Kong, 108/211; Vietnam, 25/37; and Singapore, 45/56) compared to their representation in Cape Town, South Africa (79/321) (Fisher's exact test odds ratio [OR], 4.20; 95% confidence interval [CI], 3.06 to 5.77; $p < 0.0001$).

A significant association between the frequency of occurrence of strains of defined Beijing sublineages and the human population from whom they were isolated was observed (for sublineages 1 to 6, n = 88 for Cape Town and n = 253 for East Asia; for sublineage 7, n = 233 for Cape Town and n = 43 for East Asia; Fisher's exact test OR, 15.58; 95% CI, 10.38 to 23.38; $p < 0.0001$).

It is unlikely that these findings can be explained by multiple importations of founder strains of sublineage 7 in preference to founder strains of sublineages 1 to 6, given that immigrants to South Africa came from many different geographical regions in East Asia and that sublineage 7 founder strains are less frequently observed in East Asia. Accordingly, we propose that the situation in Cape Town represents an approximation to a common starting point for all the founder strains introduced, with those best adapted to the local population spreading most efficiently. This could be due to the innate characteristics of the strains within defined Beijing sublineages or the local host population. Susceptibility to *M. tuberculosis* per se has frequently been associated with the HLA genotype (11), and HLA allele frequencies are known to differ widely between human populations with different histories, with certain alleles

totally absent in some populations. Our conclusion differs from that of Gagneux et al. (6), as we demonstrate that strains from a defined sublineage (a subset of strains from an evolutionary lineage) may have been selected by a human population in a defined geographical setting.

In summary, the global success of the Beijing lineage may reflect either the selection of defined sublineages in different geographical settings by distinct human populations or the adaptation of strains in a defined sublineage to spread more readily in a distinct human population. We acknowledge that these contrasting conclusions cannot be easily distinguished with the available data. However, the emergence of a sublineage of Beijing strains with increased pathogenicity may have important implications for the Tuberculosis Control Program. Early diagnosis and contact tracing will be essential to curb the spread of these strains. Furthermore, it will be important to ensure that future vaccines protect against these strains.

Acknowledgments

TB in the 21st Century Consortium, the Harry Crossley Foundation and the European Commission 6th framework programme on research technological development demonstration (project no. 037919) are thanked for financially supporting this project. Prof M. Kidd is thanked for statistical advice.

Reference List

1. **Banu, S., S. V. Gordon, S. Palmer, M. R. Islam, S. Ahmed, K. M. Alam, S. T. Cole, and R. Brosch**. 2004. Genotypic analysis of *Mycobacterium tuberculosis* in Bangladesh and prevalence of the Beijing strain. J.Clin.Microbiol. **42**:674-682.

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

3. **Borgdorff, M. W., H. van Deutekom, P. E. de Haas, K. Kremer, and D. van Soolingen**. 2004. *Mycobacterium tuberculosis*, Beijing genotype strains not associated with radiological presentation of pulmonary tuberculosis. Tuberculosis.(Edinb.) **84**:337-340.

4. **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**:2726-2731.

5. **European Concerted Action on New Generation Genetic Markers and Techniques for the Epidemiology and Control of Tuberculosis.** 2006. Beijing/W genotype *Mycobacterium tuberculosis* and drug resistance. Emerg.Infect.Dis. 12:736-743.

6. **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 .

7. **Glynn, J. R., J. Whiteley, P. J. Bifani, K. Kremer, and D. van Soolingen**. 2002. Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: a systematic review. Emerg.Infect.Dis. **8**:843-849.

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

9. **Kam, K. M., C. W. Yip, L. W. Tse, K. L. Wong, T. K. Lam, K. Kremer, B. K. Au, and D. van Soolingen**. 2005. Utility of mycobacterial interspersed repetitive unit typing for differentiating multidrug-resistant *Mycobacterium tuberculosis* isolates of the Beijing family. J.Clin.Microbiol. **43**:306-313.

10. **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**:907-914.

11. **Lombard, Z., A. E. Brune, E. G. Hoal, C. Babb, P. D. van Helden, J. T. Epplen, and L. Bornman**. 2006. HLA class II disease associations in southern Africa. Tissue Antigens **67**:97-110.

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

13. **Mokrousov, I., O. Narvskaya, E. Limeschenko, A. Vyazovaya, T. Otten, and B. Vyshnevskiy**. 2004. Analysis of the allelic diversity of the mycobacterial interspersed repetitive units in *Mycobacterium tuberculosis* strains of the Beijing family: practical implications and evolutionary considerations. J.Clin.Microbiol. **42**:2438-2444.

14. **Nurse, G. T., J. S. Weiner, and T. Jenkins**. 1985. The Peoples of Southern Africa and their Affinities, p. 410. *In* Clarendon Press, Oxford (United Kingdom).

15. **Sun, Y. J., R. Bellamy, A. S. Lee, S. T. Ng, S. Ravindran, S. Y. Wong, C. Locht, P. Supply, and N. I. Paton**. 2004. Use of mycobacterial interspersed repetitive unit-variable-number tandem repeat typing to examine genetic diversity of *Mycobacterium tuberculosis* in Singapore. J.Clin.Microbiol. **42**:1986-1993.

16. **Sun, Y. J., T. K. Lim, A. K. Ong, B. C. Ho, G. T. Seah, and N. I. Paton**. 2006. Tuberculosis associated with *Mycobacterium tuberculosis* Beijing and non-Beijing genotypes: A clinical and immunological comparison. BMC.Infect.Dis. **6**:105.

17. **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**:3563-3571.

18. **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**:529-538.

19. **van Soolingen, D., L. Qian, P. E. de Haas, J. T. Douglas, H. Traore, F. Portaels, H. Z. Qing, D. Enkhsaikan, P. Nymadawa, and J. D. van Embden**. 1995. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. J.Clin.Microbiol. **33**:3234-3238.

20. **Verver, S., R. M. Warren, Z. Munch, E. Vynnycky, P. D. van Helden, M. Richardson, G. D. van der Spuy, D. A. Enarson, M. W. Borgdorff, M. A. Behr, and N. Beyers**. 2004. Transmission of tuberculosis in a high incidence urban community in South Africa. Int.J.Epidemiol. **33**:351-357.

Chapter 4

Discordance between Mycobacterial Interspersed Repetitive-Unit–Variable-Number Tandem-Repeat Typing and IS*6110* Restriction Fragment Length Polymorphism Genotyping for analysis of *Mycobacterium tuberculosis* Beijing strains in a setting of high incidence of tuberculosis.

Hanekom, M.,‡ van der Spuy, G. D.,‡ Gey van Pittius, N. C., McEvoy, C. R. E., Hoek, K. G. P., Ndabambi, S. L., Jordaan, A. M., Victor, T. C., van Helden, P. D., and Warren, R. M.

‡ Joint first authorship

J. Clin. Microbiol. (2008). 46:3338-3345

Abstract

IS*6110* restriction fragment length polymorphism (RFLP) genotyping is the most widely used genotyping method to study the epidemiology of *Mycobacterium tuberculosis*. However, due to the complexity of the IS*6110* RFLP genotyping technique, and the interpretation of RFLP data, mycobacterial interspersed repetitive-unit–variable-number tandem-repeat (MIRU-VNTR) genotyping has been proposed as the new genotyping standard. This study aimed to determine the discriminatory power of different MIRU-VNTR locus combinations relative to IS*6110* RFLP genotyping, using a collection of Beijing genotype *M. tuberculosis* strains with a well-established phylogenetic history. Clustering, diversity index, clustering concordance, concordance among unique genotypes, and divergent and convergent evolution were calculated for seven combinations of 27 different MIRU-VNTR loci and compared to IS*6110* RFLP results. Our results confirmed previous findings that MIRU-VNTR genotyping can be used to estimate the extent of recent or ongoing transmission. However, molecular epidemiological linking of cases varied ignificantly depending on the genotyping method used. We conclude that IS*6110* RFLP and MIRU-VNTR loci evolve independently and at different rates, which leads to discordance between transmission chains predicted by the respective genotyping methods. Concordance between the two genotyping methods could be improved by the inclusion of genetic distance (GD) into the clustering formulae for some of the MIRU-VNTR loci combinations. In summary, our findings differ from previous reports, which may be explained by the fact that in settings of low tuberculosis incidence, the genetic distance between epidemiologically unrelated isolates was sufficient to define a strain using either marker, whereas in settings of high incidence, continuous evolution and persistence of strains revealed the weaknesses inherent to these markers.

Introduction

Over the past 2 decades, molecular genotyping methods have enhanced our understanding of the epidemiology of tuberculosis (TB) in numerous geographical settings. These methods have enabled geo-temporal tracking of *Mycobacterium tuberculosis* strains with the view to identifying source cases responsible for TB outbreaks (3), tracking of recent and ongoing disease transmission (31),

distinguishing between reinfection and relapse (28), evaluating the effectiveness of direct observed therapy short-course-based TB control programs (5,16), and identifying global genetic lineages (7). Ideally, molecular genotyping tools should be inexpensive, highly discriminative, deliver rapid results, be straightforward to perform, and produce easily interpretable results that allow for accurate interlaboratory comparisons (universally comparable databases).

Three genotyping methods are currently widely used in molecular epidemiological studies of TB: IS*6110* restriction fragment length polymorphism (RFLP) genotyping (27), spoligotyping (14), and mycobacterial interspersed repetitive-unit–variable-number tandem-repeat (MIRU-VNTR) genotyping (21, 22). Currently, IS*6110* RFLP genotyping is the most widely used genotyping method (27). However, this method is time-consuming, laborious, and complex. Furthermore, differences in application can make inter-laboratory comparisons difficult, and the data generated may have limitations (i.e., comparison of strains with high versus low IS*6110* copy numbers). More recently, the validity of the calculation of IS*6110* RFLP clustering, as a surrogate for transmission, has been questioned, as the IS*6110* banding pattern may change during transmission (33, 35). A nearest genetic distance model has been evaluated to incorporate IS*6110* banding changes into the calculation of ongoing transmission (24). The term "cluster" has also been questioned in studies which have compared contact tracing data with IS*6110* RFLP data (4, 26). In response, numerous studies have been conducted to try to identify alternative methods that have the ability to accurately describe epidemiological events in different settings at a similar discriminatory level to that of IS*6110* RFLP genotyping. One of the most promising methods is MIRU-VNTR genotyping, a PCR-based method for detecting the number of tandem repeats at a given genetic locus. Supply *et al*. (21) defined a set of 15 MIRU-VNTR loci for molecular epidemiological investigations and a set of 24 MIRU-VNTR loci for phylogenetic analysis of *M. tuberculosis* strains worldwide. In support of this, another study concluded that this "real-time" MIRU-VNTR genotyping approach was highly applicable for population-based studies (18). This view was reinforced by a study conducted in the Brussels region, where the authors concluded that a standardized MIRU-VNTR genotyping method could be a new

reference for epidemiological and phylogenetic screening of *M. tuberculosis* strains (2).

A study from Japan (10) investigated the differentiation power of the proposed 15- and 24-loci MIRU-VNTR genotyping methods for strains with the Beijing genotype and concluded that the analyses of these loci were of limited use for discriminating strains of this genotype. In their study they showed that VNTR loci 3820, 3232, and 4120 were highly polymorphic in Beijing genotype strains and thus proposed the use of these loci to enhance the discriminatory power of the proposed 15-loci MIRU-VNTR genotyping method. However, other studies have excluded these loci due to difficulties associated with the reproducibility of PCR amplification (15, 21, 36). Subsequently, a study in Hong Kong, which also examined strains of the Beijing genotype, showed that a different combination of 12 VNTR and QUB (Queen's University of Belfast) loci gave a Hunter-Gaston discriminatory index value which was almost equal to that obtained in IS*6110* RFLP genotyping (12, 13). However, this was refuted by a more recent study from China which suggested that MIRU-VNTR genotyping may over-estimate transmission in isolates with the Beijing genotype (11). Collectively, these findings suggest that the selection of MIRU-VNTR loci for optimal differentiation of *M. tuberculosis* requires further validation in different geographical settings. To date, the performance of the MIRU-VNTR genotyping method has not been evaluated in an epidemic setting, nor has it been tested within the context of a robust *M. tuberculosis* phylogeny.

In this study the discriminatory power of different MIRUVNTR locus combinations was determined as previously described (8, 10, 21, 22) and compared to the IS*6110* RFLP genotyping method by using a collection of Beijing genotype *M. tuberculosis* strains with a well-established phylogenetic history (9). The results are discussed in the context of concordance between the different genotyping methods in their abilities to define a strain and to accurately describe the epidemiology of TB in a high-incidence setting.

Materials and methods

Study population

Sputum samples were collected during the period from January 1993 to December 2004 from new and retreatment TB patients who were resident and attending health care clinics in an epidemiological field site in Cape Town, South Africa (31). This study is part of a larger, long-term molecular epidemiological project which has been approved by the ethics committee of Stellenbosch University.

IS6110 RFLP genotyping

M. tuberculosis isolates were cultured on MGIT (Becton-Dickinson) or Löwenstein-Jensen medium, and DNA was extracted as previously described (32). Each isolate was classified by IS*6110* RFLP genotyping (27) and spoligotyping (14) using internationally standardized protocols. IS*6110* RFLP patterns were analyzed using Gelcompar II (Applied-Maths, Sint-Martens-latem, Belgium) with tolerance settings allowing a 5% shift in lane position and a 0.6% variation in individual band position to compensate for minor technical errors. Isolates were assigned as members of the Beijing genotype if they had the characteristic Beijing spoligotype (30). Only the first *M. tuberculosis* isolate from each case was included for subsequent analysis. Each Beijing isolate was grouped into one of seven phylogenetic sublineages according to 40 different genetic markers, as previously described (9).

DNA sequencing

The DNA sequence of the *katG*, *rpoB*, *embB*, and *rrs* genes of isolates classified as members of the Beijing sublineage 5 were determined as previously described (19, 25).

MIRU-VNTR typing

Twenty-seven MIRU-VNTR loci were amplified by PCR as described previously (8, 10, 21, 22). The number of repeats at each genomic locus was calculated according to the electrophoretic mobility of the corresponding PCR product (23). Alleles were assigned numerical values according to the number of repeats present in that genomic

locus. Isolates were genotypically classified according to seven different MIRU-VNTR locus combinations (Table 1).

Table 1: MIRU-VNTR locus combinations

	MIRU-VNTR Locus Combinations						
	12-MIRU[a]	12-MIRU + ETR A, B, C[b]	12-MIRU + hyper-variable loci[e]	15-MIRU-VNTR[c]	15-MIRU-VNTR + hyper-variable loci[d]	24-MIRU-VNTR[c]	24-MIRU-VNTR + hyper-variable loci[e]
MIRU02	●	●	●			●	●
MIRU04	●	●	●	●	●	●	●
MIRU10	●	●	●	●	●	●	●
MIRU16	●	●	●	●	●	●	●
MIRU20	●	●	●			●	●
MIRU23	●	●	●			●	●
MIRU24	●	●	●			●	●
MIRU26	●	●	●	●	●	●	●
MIRU27	●	●	●			●	●
MIRU31	●	●	●	●	●	●	●
MIRU39	●	●	●			●	●
MIRU40	●	●	●	●	●	●	●
VNTR1955				●	●	●	●
VNTR2165/ETR-A		●		●	●	●	●
QUB11b				●	●	●	●
QUB26b				●	●	●	●
VNTR0424				●	●	●	●
VNTR2401				●	●	●	●
VNTR4156				●	●	●	●
VNTR3690				●	●	●	●
ETR-C		●		●	●	●	●
VNTR2347						●	●
ETR-B		●				●	●

Mtub 34			•	•
QUB3232	•	•		•
VNTR3820	•	•		•
VNTR4120	•	•		•

• locus included

[a] according to (22)

[b] according to (7)

[c] according to (21)

[d] according to (9)

[e] this study

Analytical calculations

(i) Estimation of clustering: A cluster (representing either recent or ongoing transmission or a <2-year interval) was defined as a series of isolates having the same genotype (IS*6110* RFLP or MIRU-VNTR), while isolates with unique IS*6110* RFLP or MIRU-VNTR genotypes were considered to represent reactivation or influx of disease into the study community (20). Secondary analyses which incorporated the concept of evolution during transmission were done using data sets (genotypes according to IS*6110* RFLP or a particular MIRU-VNTR locus combination) in which isolates separated by a single evolutionary event were combined into transmission chains with a genetic distance of 1 (24).

(ii) Estimation of genetic diversity: The genetic diversity for each individual MIRU-VNTR locus, each of the seven MIRU-VNTR locus combinations (Table 1), and the IS*6110* RFLP fingerprints was calculated as $h = 1 - \sum xi^2[n/(n-1)]$, where x_i is the frequency of the ith allele at the locus, n is the number of isolates in the sample, and the term $n/(n - 1)$ is a correction for bias in small samples (17).

(iii) Estimation of matching and mismatching concordance: Concordance between the IS*6110* RFLP genotypes and the respective MIRU-VNTR genotypes was calculated as follows: each isolate was paired with every other isolate in the data set, and their genotypes (IS*6110* RFLP and MIRU-VNTR) were scored as either a match

(identical) or mismatch (nonidentical). Matching concordance between the respective genotyping methods was calculated according to the number of paired isolates having a match for both of the methods as a proportion of the total number of pairs having matching IS*6110* RFLP genotypes. This is a measure of agreement between two methods as to whether any two isolates form part of the same transmission chain. Mismatching concordance was calculated as the number of paired isolates having nonmatching genotypes for both of the methods as a proportion of the total number of pairs having nonmatching IS*6110* RFLP genotypes. This is a measure of agreement between two methods for any two isolates that do not form part of the same transmission chain.

(iv) Estimation of concordance among unique genotypes: Concordance between uniquely occurring IS*6110* RFLP genotypes and the MIRU-VNTR genotypes was calculated as the proportion of isolates having unique IS*6110* RFLP genotypes that also had unique MIRU-VNTR genotypes.

(v) Estimation of number of convergent events: Convergent evolution was identified by drawing connecting lines between each IS*6110* RFLP genotype and each MIRU-VNTR genotype for which isolates were found to have that particular genotype combination (Figure 1). Convergent evolution was defined, conservatively, as the existence of isolates representing each of the four possible combinations of two IS*6110* RFLP genotypes (e.g., IS1 and IS2) and two MIRU-VNTR genotypes (e.g., M1 and M2) (Figure 1). This scenario would only be possible if one of the MIRU-VNTR genotypes had evolved more than once, assuming that the chance of IS*6110* RFLP genotype convergence was significantly lower than that of MIRU-VNTR genotype convergence. The validity of this method was confirmed by plotting the IS*6110* RFLP genotypes onto a phylogenetic tree constructed using the MIRU-VNTR data in combination with a neighbor-joining algorithm (data not shown) (34).

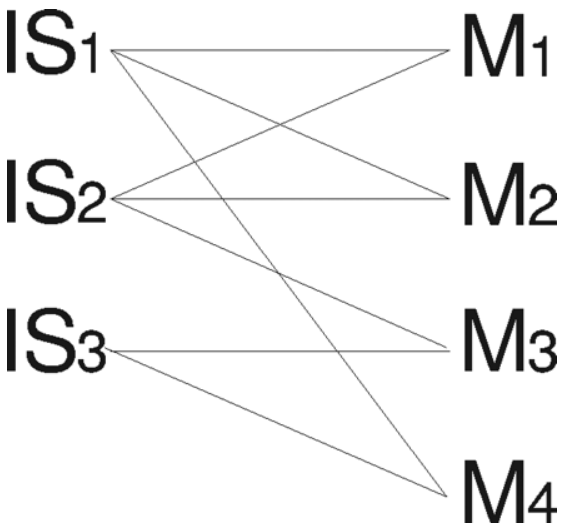


Figure 1: An example of MIRU-VNTR (M_x) and IS*6110* RFLP (IS_x) genotypes. The connecting lines represent the MIRU-VNTR and IS*6110* RFLP genotype combinations observed in *M. tuberculosis* isolates in the study setting. M_1 and M_2 are both linked to IS_1 and IS_2 and therefore represent a convergent event. Neither M_3 nor M_4 share common connections to more than one IS_x with any other M_x. Their connecting lines therefore indicate simple, linear evolution.

(vi) Estimation of number of divergent events: A divergent evolutionary event was scored for each MIRU-VNTR genotype which existed in combination with only one IS*6110* RFLP genotype and where this IS*6110* RFLP genotype was found in combination with more than one MIRU-VNTR genotype (Figure 2). This implies that the MIRU-VNTR genotype arose subsequent to the IS*6110* RFLP genotype. A divergent event was also added for each convergent event, since a convergent event implies a prior divergent event.

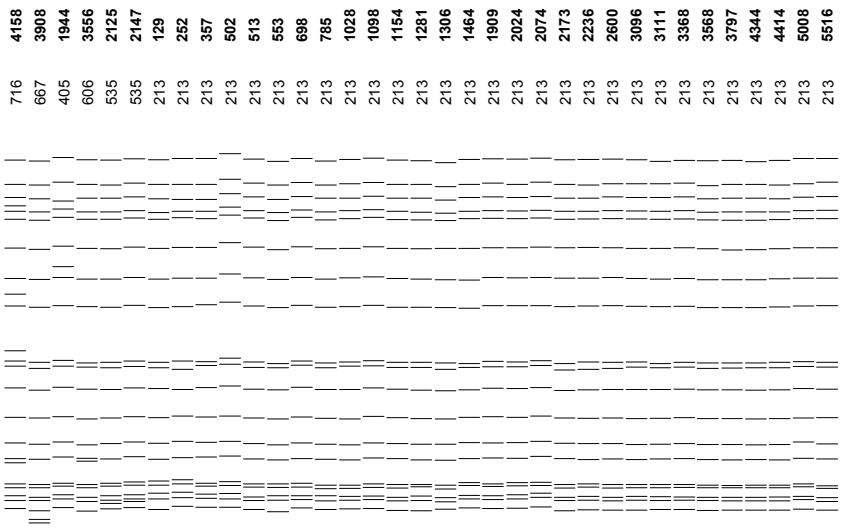


Figure 2: IS*6110* RFLP banding patterns of Beijing sublineage 5 isolates sharing identical *katG*, *rpoB*, *embB*, and *rrs* gene mutations. The isolate numbers are indicated in bold, while the IS*6110*-RFLP cluster numbers are indicated in standard text (these numbers correspond to the numbers given in data set S2 of the supplemental material).

(vii) Sensitivity and specificity calculations: The sensitivity and specificity (and positive and negative predictive values) of the IS*6110* RFLP and respective MIRU-VNTR genotyping methods were calculated using GraphPad Prism 5 software (La Jolla, CA) based on their ability to correctly identify an independently genotyped drug-resistant cluster.

Results

IS*6110* RFLP genotyping identified 74 different strains among the 321 isolates with the Beijing spoligotype collected over a 12-year period (Table 2). Of these strains, 272 were grouped into 25 clusters (containing between 2 and 100 isolates), and 49 were unique strains. The overall percent clustering was calculated to be 84.7% using the n/T formula (1). Each isolate was subsequently genotyped with 27 MIRU-VNTR loci and analyzed according to seven different MIRU-VNTR locus combinations (Table 1 and Supplemental data). The performance of these locus combinations, in relation to the IS*6110* RFLP genotyping method, was determined either over a 12-year period (Table 2) or over six consecutive 2-year periods (Table 3). In both

analyses the traditional 12-MIRU loci genotyping method underestimated the number of genotypes (strains) identified and thereby overestimated the percentage of clustering (Tables 2 and 3). The inclusion of exact tandem repeat (ETR) alleles A, B, and C to the 12-MIRU loci set did not significantly improve the number of strains detected or the estimate of clustering (Tables 2 and 3). Analysis of the isolates using the newly proposed 15- and 24-MIRU-VNTR locus combinations increased the number of strains identified; however, the discriminatory power of these locus combinations remained lower than that observed using IS*6110* RFLP genotyping (Tables 2 and 3). Consequently, these locus combinations over-estimated clustering. The addition of the VNTR loci 3232, 4120, and 3820 to the 12-, 15-, and 24-MIRU-VNTR locus combinations increased the number of strains detected and thereby produced clustering estimates similar to or slightly lower than that of IS*6110* RFLP genotyping (Tables 2 and 3). This implies that some MIRU-VNTR locus combinations could be selected as epidemiological markers to estimate the extent of both recent (<2-year interval) and ongoing (unrestricted interval) transmission in settings with a high incidence of strains with the Beijing genotype.

To determine whether a correlation existed between the definitions of a strain according to IS*6110* RFLP or MIRU-VNTR genotyping methods, the respective genotypes were compared. From the results shown in Table 2 it is evident that a strain classified as a cluster according to IS*6110* RFLP genotyping may in some instances be classified as unique according to the different MIRU-VNTR locus combinations, or vice versa. Using a pair-wise analysis, we estimated the degree of matching concordance between the IS*6110* RFLP and MIRU-VNTR genotyping methods to range between 39% and 68% depending on the locus combinations used (Tables 2 and 3). The inclusion of additional MIRU-VNTR loci decreased the degree of matching concordance, as a result of an increased rate of divergence caused by more rapid evolution, with the hypervariable loci having the greatest effect. Conversely, the inclusion of additional loci increased the degree of mismatching concordance, as well as concordance between strains identified as having unique genotypes according to both genotyping methods (IS*6110* RFLP and MIRU-VNTR). A consequence of more rapid evolution was the increased risk of convergent evolutionary events (Table 2).

To determine whether concordance between the respective genotyping methods could be improved, the analysis was repeated to allow for a genetic distance of 1, i.e., evolution of single MIRU-VNTR loci or single-band changes in the IS*6110* pattern within the definition of a cluster. The results showed that the inclusion of genetic distance had a significant influence on the MIRU-VNTR definition of a cluster, collapsing many of the genotypes (Table 2 and Supplemental data 1). This was less pronounced for IS*6110* RFLP analysis (Table 2 and Supplemental data 1). Matching concordance was improved by allowing for evolution of the MIRU-VNTR genotypes; however, mismatching concordance was concomitantly reduced for genotypes based on the 12-MIRU loci combinations. This may be explained by the loss of discriminatory power as a result of the collapsing of genotypes, which is associated with a low rate of evolution. In contrast, mismatching concordance was improved for 15- and 24-MIRU-VNTR combinations due to the higher evolutionary rates of these markers. However, the concordance among unique genotypes remained low (Table 2).

To establish which of the genotyping methods provided the most accurate description of ongoing transmission in the study setting, the largest group of drug-resistant isolates (found within sublineage 5) was selected, based on identical mutations conferring resistance to isoniazid, rifampin, ethambutol, and streptomycin (see Supplemental data 2). These isolates represent the continuing spread of a previously described multidrug-resistant TB outbreak (29). A total of 35 isolates were identified with the *katG*315 AGC to ACC, *rpoB*531 TCG to TTG, *embB*306 ATG to ATA, and *rrs*513 CAG to CCG mutations, forming a single drug resistance-based cluster (Fig. 2). The sensitivity, specificity, and positive and negative predictive values related to the abilities of the different markers to identify the drug-resistant cluster are given in Table 4. While the sensitivities of all the markers were high, with some of those based on MIRU-VNTR loci outperforming IS*6110* RFLP, the specificity of all MIRU-VNTR markers was substantially lower than that of IS*6110* RFLP. The inclusion of genetic distance (single events) within the definition of a cluster appeared to improve the sensitivity of most of the markers but concomitantly decreased the specificity of the MIRU-VNTR markers. The specificity of IS*6110* was not affected by the inclusion of genetic distance. Positive predictive values were not significantly

affected by allowing for evolution of the markers; however, with the exceptions of IS*6110*, which increased, and the 24 MIRU and three hypervariable loci, which remained unchanged, the negative predictive values for all markers were reduced to zero.

To determine whether MIRU-VNTR genotyping could be used as a method to phylogenetically group strains with the Beijing genotype, the correlation between MIRU-VNTR genotype and Beijing sublineage was quantified. As sublineages 3 and 4 and sublineages 5 and 6 were distinguished solely on the basis of IS*6110* in our data set, these two pairs of sublineages were combined for the purposes of this analysis. Table 2 shows that the respective MIRU-VNTR locus combinations correctly grouped >96% of the isolates according to their sublineage designation, in comparison to 100% with IS*6110* RFLP genotyping. The incorporation of genetic distance reduced the ability of genotyping methods based on the 12-MIRU locus combinations to correctly group isolates (Table 2).

Table 2: Comparison between molecular epidemiological data generated over a 12-year interval using IS*6110* RFLP and MIRU-VNTR genotyping methods

			MIRU-VNTR Locus Combinations						
		IS*6110*-RFLP	12-MIRU[a]	12-MIRU + ETR A, B, C[b]	12-MIRU + hyper-variable loci[e]	15-MIRU-VNTR[c]	15-MIRU-VNTR + hyper-variable loci[d]	24-MIRU-VNTR[c]	24-MIRU-VNTR + hyper-variable loci[e]
GD = 0	Genotypes (n)	74	27	39	67	47	83	57	91
	Clustering (%)	84.7	95.0	91.2	84.1	89.4	78.8	87.2	77.3
	Unique IS*6110* genotypes (n = 49) with unique MIRU-VNTR genotypes (n)	N/A	6	8	14	12	19	12	19
	Unique IS*6110* genotypes (n = 49) with clustered MIRU-VNTR genotypes (n)	N/A	43	41	35	37	30	37	30
	Clustered IS*6110* genotype (n = 272) with unique MIRU-VNTR genotypes (n)	N/A	10	18	38	22	49	28	54
	Clustered IS*6110* genotype (n = 272) with clustered MIRU-VNTR genotypes (n)	N/A	262	254	234	250	223	244	218
	Pair-wise matching concordance (%)	100	68	60	53	67	51	52	39
	Pair-wise mismatching concordance (%)	N/A	69	72	71	68	71	73	81
	Concordance between unique strains (%)	100	20.4	22.4	28.6	34.7	40.8	40.8	42.9
	Converged genotypes (n)	N/A	3	5	5	3	5	7	9
	Diverged genotypes (n)	N/A	14	26	47	27	58	38	68
	Diversity index[f]	0.85	0.63	0.67	0.7	0.63	0.7	0.72	0.78

GD = 1	**Genotypes (n)**	40	3	4	11	11	20	15	27
	Pair-wise matching concordance (%)	100	99	99	96	99	96	99	96
	Pair-wise mismatching concordance (%)	N/A	2	3	8	80	82	83	88
	Concordance between unique strains (%)	100	0	4	4	15	30	22	41
	Beijing sublineage discrimination GD = 0 (%)	100	96	96	99	99	99	99	99
	Beijing sublineage discrimination GD = 1 (%)	99	1	1	3	99	99	99	99

[a] according to (22),

[b] according to (7),

[c] according to (21),

[d] according to (9),

[e] this study,

[f] according to (15)

N/A: not applicable

Table 3: Comparison between molecular epidemiological data generated over six consecutive 2-year intervals by IS*6110* RFLP and MIRU-VNTR genotyping methods

	MIRU-VNTR Locus Combinations							
	IS*6110*-RFLP	12-MIRU[a]	12-MIRU + ETR A, B, C[b]	12-MIRU + hyper-variable loci[e]	15-MIRU-VNTR[c]	15-MIRU-VNTR + hyper-variable loci[d]	24-MIRU-VNTR[c]	24-MIRU-VNTR + hyper-variable loci[e]
Average No strains (range)	17.9 (7-20)	9.9 (5-14)	10.9 (7-14)	15.4 (8-22)	12.0 (5-17)	16.6 (7-23)	13.6 (7-19)	18.1 (9-25)
Average No clusters (range)	6.1 (4-10)	4.6 (2-7)	4.9 (3-7)	4.0 (2-6)	3.7 (2-5)	3.6 (2-6)	4.6 (3-6)	4.3 (3-7)
Average % clustering (range)	74.5 (55.4-86.7)	85.1 (60.0-93.1)	83.1 (60.0-91.7)	73.1 (60.0-77.8)	80.3 (66.7-93.1)	70.7 (57.6-76.4)	77.8 (60.0-89.7)	67.7 (54.5-75.0)
Average % pair-wise matching concordance (range)	100	69.7 (60-80)	63.7 (44-72)	52.7 (40-58)	68.2 (45-78)	51.0 (22-63)	55.0 (44-69)	40.2 (22-51)

Average % pair-wise mis-matching concordance (range)	N/A	92.8 (72-99)	95.0 (84-99)	94.0 (77-99)	92.3 (70-99)	94.0 (76-100)	95.2 (84-99)	95.0 (84-99)
Average % concordance between unique strains (range)	100	19.1 (6.1-60.0)	25.7 (15.6-60.0)	42.2 (27.0-66.7)	31.3 (16.7-60.0)	50.9 (32.0-66.7)	35.4 (24.0-60.0)	51.3 (32.0-66.7)

[a] according to (22)

[b] according to (7)

[c] according to (21)

[d] according to (9)

[e] this study

N/A: not applicable

Table 4: Sensitivity, specificity, PPV, and NPV values for IS*6110* RFLP and MIRU-VNTR genotyping methods based on correct identification of an independently genotyped drug-resistant cluster characterized by unique mutations in the *katG*, *rpoB*, *embB*, and *rrs* genes.

			MIRU-VNTR Locus Combinations						
		IS*6110*-RFLP	12-MIRU[a]	12-MIRU + ETR A, B, C[b]	12-MIRU + hyper-variable loci[e]	15-MIRU-VNTR[c]	15-MIRU-VNTR + hyper-variable loci[d]	24-MIRU-VNTR[e]	24-MIRU-VNTR + hyper-variable loci[e]
GD = 0	Sensitivity	0.83 (0.66 - 0.93)	1.00 (0.90 - 1.00)	0.97 (0.85 - 1.00)	0.80 (0.63 - 0.92)	0.91 (0.77 - 0.98)	0.74 (0.57 - 0.88)	0.91 (0.77 - 0.98)	0.74 (0.57 - 0.88)
	Specificity	1.00 (0.48 - 1.00)	0.20 (0.01 - 0.72)	0.20 (0.01 - 0.72)	0.20 (0.01 - 0.72)	0.40 (0.05 - 0.85)	0.40 (0.05 - 0.85)	0.40 (0.05 - 0.85)	0.40 (0.05 - 0.85)
	Positive Predictive Value	1.00 (0.88 - 1.00)	0.90 (0.76 - 0.97)	0.89 (0.75 - 0.97)	0.88 (0.71 - 0.96)	0.91 (0.77 - 0.98)	0.90 (0.73 - 0.98)	0.91 (0.77 - 0.98)	0.90 (0.73 - 0.98)
	Negative Predictive Value	0.45 (0.17 - 0.77)	1.00 (0.03 - 1.00)	0.50 (0.01 - 0.99)	0.13 (0.00 - 0.53)	0.40 (0.05 - 0.85)	0.18 (0.02 - 0.52)	0.400 (0.05 - 0.85)	0.18 (0.02 - 0.52
GD = 1	Sensitivity	0.94 (0.81 - 0.99)	1.00 (0.72 – 1.00)	1.00 (0.72 – 1.00)	0.97 (0.85 - 1.00)	1.00 (0.72 – 1.00)	0.94 (0.81 - 0.99)	1.00 (0.72 – 1.00)	0.74 (0.57 - 0.88)
	Specificity	1.00 (0.48 - 1.00)	0.00 (0.00 - 0.54)	0.00 (0.00 - 0.54)	0.00 (0.00 - 0.52)	0.00 (0.00 - 0.537)	0.00 (0.00 - 0.52)	0.00 (0.00 - 0.54)	0.40 (0.05 - 0.85)
	Positive Predictive Value	1.00 (0.89 - 1.00)	0.88 (0.72 – 0.95)	0.88 (0.72 – 0.95)	0.87 (0.73 - 0.96)	0.88 (0.72 – 0.95)	0.87 (0.72 - 0.96)	0.88 (0.72 – 0.95)	0.90 (0.73 - 0.98)
	Negative Predictive Value	0.71 (0.29 - 0.960)	ND	ND	0.00 (0.00 - 0.98)	ND	0.00 (0.00 - 0.84)	ND	0.18 (0.02 - 0.52)

Sensitivity, specificity and predictive values are given with their 95% confidence intervals in brackets

[a] according to (22)

[b] according to (7)

[c] according to (21)

[d] according to (9)

[e] this study

ND: not determinable

Discussion

IS*6110* RFLP genotyping is the most widely used genotyping method for investigating and understanding the epidemiology of *M. tuberculosis* (27). However, studies comparing IS*6110* RFLP molecular epidemiological and contact tracing data have questioned the validity of the definition of transmission (4, 26). In order to address these concerns MIRU-VNTR genotyping using either 15- or 24-MIRU-VNTR loci combinations have been extensively evaluated as the new genotyping standard for molecular epidemiological studies of *M. tuberculosis* (21). Concordance between MIRU-VNTR genotyping and contact tracing data was found to be superior to that of IS*6110* RFLP in settings of low incidence (2, 18). However, these MIRU-VNTR locus combinations have not been fully tested in geographical regions of TB endemicity or within a robust *M. tuberculosis* phylogeny. Our results confirm previous findings (2, 10, 18, 21) which suggested that MIRU-VNTR genotyping, using carefully selected locus combinations, could be used to estimate the extent of recent or ongoing transmission. The inclusion of the three hypervariable loci improved the discriminatory power of the MIRU-VNTR genotyping method in this Beijing lineage, thereby supporting a previous suggestion for their inclusion (10). However, the use of these loci needs further evaluation in other evolutionary lineages, as difficulties associated with amplification reproducibility have been reported (15, 21, 36).

We conclude that the PCR-based MIRU-VNTR genotyping method could be applied as an epidemiological tool to measure the performance of a TB control program over time in a defined geographical setting. However, the observed concordance in the estimate of recent and ongoing transmission when using the IS*6110* RFLP or MIRU-VNTR genotyping methods was only coincidental. A subsequent analysis of the MIRU-VNTR data, in comparison to the IS*6110* RFLP genotyping data, revealed that the classification of a strain according to its genotype differed significantly depending on the genotyping method used. Accordingly, our study showed that the degree of matching and mismatching concordance as well as concordance among unique strains was low. This led to discordance between the transmission chains predicted by the respective genotyping methods. Matching concordance increased when genetic distance was incorporated into the clustering calculation for all of the MIRU-VNTR

combinations. However, this effect was offset in the case of 12-MIRU-based markers by the concomitant reduction in mismatching concordance, which was not the case for the 15- and 24-MIRU-VNTR combinations. From this, it is apparent that the additional loci included in the 15- and 24-MIRU-VNTR combinations (with or without the addition of the hypervariable loci) improved the overall concordance of MIRU-VNTR with respect to IS*6110* RFLP. This may be due to these loci being inherently less stable and therefore more informative. However, a caveat to the inclusion of genetic distance in the clustering formula is that epidemiologically unrelated cases may be incorrectly linked within a transmission chain.

Our analysis of the drug-resistant cluster to elucidate which of the genotyping methods provided the most accurate reflection of the epidemiology highlighted shortcomings of both the IS*6110* RFLP and MIRU-VNTR genotyping methods. This analysis supported a previous study which demonstrated that ongoing transmission was characterized by the evolution of variant IS*6110* RFLP genotypes while simultaneously preserving existing genotypes (33). A similar observation was found when using the different MIRU-VNTR locus combinations. This could be explained by the fact that the evolution of different loci could take place both convergently and divergently. Together, these results substantiate previous findings which have suggested that the definition of ongoing transmission according to IS*6110* RFLP or MIRU-VNTR genotyping should include closely related genotypes (18, 24, 35). However, when allowing for single MIRU-VNTR changes within the definition of a cluster, the MIRU-VNTR genotyping method collapsed many of the sublineage 5 isolates into a limited number of clusters. As a result, most of the isolates were grouped as resistant, giving the method a high sensitivity, but in doing so, compromising specificity. In contrast, the identification of isolates within the drug-resistant cluster was largely retained by IS*6110* RFLP analysis despite the inclusion of genetic distance. This suggests that IS*6110* RFLP analysis in combination with genetic distance provides a more accurate reflection of ongoing transmission of this multidrug-resistant TB outbreak in this setting. This finding is important for the interpretation of molecular epidemiological data in settings where contact tracing is extremely difficult. However, we acknowledge that the concordance between IS*6110*

RFLP findings and transmission needs further investigation in different settings and in *M. tuberculosis* strains with different genetic backgrounds.

Our results differ from previous studies (2, 18), which demonstrated a close correlation between IS*6110* RFLP and MIRU-VNTR genotyping. These studies were conducted in settings in Western Europe with a low incidence of TB and where the TB epidemic is primarily driven by reactivation and immigration (6). In these settings, efficient TB control programs would largely prevent recent and ongoing transmission and the subsequent generation of closely related clonal variants. Thus, genetic diversity is predicted to be preserved. In most instances, this would imply that the strains cultured from TB cases would be genetically distantly related and thus would not share either IS*6110* RFLP banding patterns or MIRU-VNTR genotypes. Accordingly, MIRU-VNTR genotyping would discriminate strains at a level similar to that of IS*6110* RFLP genotyping. In contrast, our setting of high TB incidence has promoted the evolution of a large number of genetically closely related strains which are maintained within the host population. The genetic distance between these strains is often of such a nature that strains either have identical IS*6110* RFLP genotypes and variant MIRU-VNTR genotypes or vice versa. Accordingly, we hypothesize that the degree of discordance between IS*6110* RFLP and MIRU-VNTR genotyping is dependent on the genetic distance between isolates. This is supported by the observation that distantly related isolates from the different Beijing sublineages have evolved distinct IS*6110* RFLP and MIRU-VNTR genotypes.

In summary, we conclude that both IS*6110* RFLP and MIRU-VNTR genotyping methods have limitations in defining chains of transmission of Beijing genotype *M. tuberculosis* strains in this setting of high incidence.

Acknowledgments

The Harry Crossley Foundation and the European Commission 6th Framework Program on Research Technological Development Demonstration (project no. 037919) are thanked for financially supporting this project.

Reference List

1. **Alland, D., G. E. Kalkut, A. R. Moss, R. A. McAdam, J. A. Hahn, W. Bosworth, E. Drucker, and B. R. Bloom**. 1994. Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. N.Engl.J.Med. **330**:1710-1716.

2. **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**:452-457.

3. **Cowan, L. S., L. Diem, T. Monson, P. Wand, D. Temporado, T. V. Oemig, and J. T. Crawford**. 2005. Evaluation of a two-step approach for large-scale, prospective genotyping of *Mycobacterium tuberculosis* isolates in the United States. J.Clin.Microbiol. **43**:688-695.

4. **Cruz-Ferro, E. and E. Fernandez-Nogueira**. 2007. Epidemiology of tuberculosis in Galicia, Spain, 1996-2005. Int.J.Tuberc.Lung Dis. **11**:1073-1079.

5. **Dahle, U. R., V. Eldholm, B. A. Winje, T. Mannsaker, and E. Heldal**. 2007. Impact of immigration on the molecular epidemiology of *Mycobacterium tuberculosis* in a low-incidence country. Am.J.Respir.Crit Care Med. **176**:930-935.

6. **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**:2869-2873.

7. **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**:8207-8213.

8. **Hanekom, M., G. D. van der Spuy, E. Streicher, S. L. Ndabambi, C. R. McEvoy, M. Kidd, N. Beyers, T. C. Victor, P. D. van Helden, and R. M.**

Warren. 2007. A recently evolved sublineage of the *Mycobacterium tuberculosis* Beijing strain family is associated with an increased ability to spread and cause disease. J.Clin.Microbiol. **45**:1483-1490.

9. **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**:67-74.

10. **Jiao, W. W., I. Mokrousov, G. Z. Sun, Y. J. Guo, A. Vyazovaya, O. Narvskaya, and A. D. Shen**. 2008. Evaluation of new variable-number tandem-repeat systems for typing *Mycobacterium tuberculosis* with Beijing genotype isolates from Beijing, China. J.Clin.Microbiol. **46**:1045-1049.

11. **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**:258-265.

12. **Kam, K. M., C. W. Yip, L. W. Tse, K. L. Wong, T. K. Lam, K. Kremer, B. K. Au, and D. van Soolingen**. 2005. Utility of mycobacterial interspersed repetitive unit typing for differentiating multidrug-resistant *Mycobacterium tuberculosis* isolates of the Beijing family. J.Clin.Microbiol. **43**:306-313.

13. **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**:907-914.

14. **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**:5628-5638.

15. **lix-Beguec, C., M. Fauville-Dufaux, and P. Supply**. 2008. Three-year population-based evaluation of standardized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing of *Mycobacterium tuberculosis*. J.Clin.Microbiol. **46**:1398-1406.

16. **Lopez-Calleja, A. I., M. A. Lezcano, M. A. Vitoria, M. J. Iglesias, A. Cebollada, C. Lafoz, P. Gavin, L. Aristimuno, M. J. Revillo, C. Martin, and S. Samper**. 2007. Genotyping of *Mycobacterium tuberculosis* over two periods: a changing scenario for tuberculosis transmission. Int.J.Tuberc.Lung Dis. **11**:1080-1086.

17. **Nei, M.** 1978. Estimation of Average Heterozygosity and Genetic Distance from a Small Number of Individuals. Genetics **89**:583-590.

18. **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**:691-697.

19. **Pretorius, G. S., P. D. van Helden, F. Sirgel, K. D. Eisenach, and T. C. Victor**. 1995. Mutations in *katG* gene sequences in isoniazid-resistant clinical isolates of *Mycobacterium tuberculosis* are rare. Antimicrob.Agents Chemother. **39**:2276-2281.

20. **Small, P. M., P. C. Hopewell, S. P. Singh, A. Paz, J. Parsonnet, D. C. Ruston, G. F. Schecter, C. L. Daley, and G. K. Schoolnik**. 1994. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. N.Engl.J.Med. **330**:1703-1709.

21. **Supply, P., C. Allix, S. Lesjean, M. Cardoso-Oelemann, S. Rusch-Gerdes, E. Willery, E. Savine, H. P. de, D. H. van, 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 S. D. van**. 2006. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. J.Clin.Microbiol. **44**:4498-4510.

22. **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**:3563-3571.

23. **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**:762-771.

24. **van der Spuy, G. D., R. M. Warren, M. Richardson, N. Beyers, M. A. Behr, and P. D. van Helden**. 2003. Use of genetic distance as a measure of ongoing transmission of *Mycobacterium tuberculosis*. J.Clin.Microbiol. **41**:5640-5644.

25. **van der Zanden, A. G., E. M. Te Koppele-Vije, B. N. Vijaya, D. van Soolingen, and L. M. Schouls**. 2003. Use of DNA extracts from Ziehl-Neelsen-stained slides for molecular detection of rifampin resistance and spoligotyping of *Mycobacterium tuberculosis*. J.Clin.Microbiol. **41**:1101-1108.

26. **van Deutekom H., P. Supply, P. E. de Haas, E. Willery, S. P. Hoijng, C. Locht, R. A. Coutinho, and S. D. van**. 2005. 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. **43**:4473-4479.

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

28. **van Rie, A., R. Warren, M. Richardson, T. C. Victor, R. P. Gie, D. A. Enarson, N. Beyers, and P. D. van Helden**. 1999. Exogenous reinfection as a cause of recurrent tuberculosis after curative treatment. N.Engl.J.Med. **341**:1174-1179.

29. **van Rie, A., R. M. Warren, N. Beyers, R. P. Gie, C. N. Classen, M. Richardson, S. L. Sampson, T. C. Victor, and P. D. van Helden**. 1999. Transmission of a multidrug-resistant *Mycobacterium tuberculosis* strain resembling "strain W"

among noninstitutionalized, human immunodeficiency virus-seronegative patients. J.Infect.Dis. **180**:1608-1615.

30. **van Soolingen, D., L. Qian, P. E. de Haas, J. T. Douglas, H. Traore, F. Portaels, H. Z. Qing, D. Enkhsaikan, P. Nymadawa, and J. D. van Embden**. 1995. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. J.Clin.Microbiol. **33**:3234-3238.

31. **Verver, S., R. M. Warren, Z. Munch, E. Vynnycky, P. D. van Helden, M. Richardson, G. D. van der Spuy, D. A. Enarson, M. W. Borgdorff, M. A. Behr, and N. Beyers**. 2004. Transmission of tuberculosis in a high incidence urban community in South Africa. Int.J.Epidemiol. **33**:351-357.

32. **Warren, R., M. de Kock, E. Engelke, R. Myburgh, v. P. Gey, T. Victor, and P. van Helden**. 2006. Safe *Mycobacterium tuberculosis* DNA extraction method that does not compromise integrity. J.Clin.Microbiol. **44**:254-256.

33. **Warren, R. M., G. D. van der Spuy, M. Richardson, N. Beyers, C. Booysen, M. A. Behr, and P. D. van Helden**. 2002. Evolution of the IS*6110*-based restriction fragment length polymorphism pattern during the transmission of *Mycobacterium tuberculosis*. Journal of Clinical Microbiology **40**:1277-1282.

34. **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**:5774-5782.

35. **Yeh, R. W., d. L. Ponce, 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**:1107-1111.

36. **Yokoyama, E., K. Kishida, M. Uchimura, and S. Ichinohe**. 2007. 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. **7**:499-508.

Table S1. Cross-tabulation of marker genotypes

The tables below show cross-tabulations of the numbers of cases classified into transmission chains according to IS*6110* and each of the MIRU-VNTR combinations, allowing for evolution by single events in either marker.

IS*6110*-RFLP genotypes GD=1	12-MIRU genotypes (GD=1)			
	1	2	3	Total
1	219	1		220
2	1			1
3	1			1
4	1			1
5	4			4
6	33			33
7	1			1
8	9			9
9	8			8
10	1			1
11	1			1
12	1			1
13	1			1
14	1			1
15	2			2
16	1			1
17	2			2
18	1			1
19	3			3
20	1			1
21	4		1	5
22	1			1
23	1			1
24	1			1
25	1			1
26	1			1
27	2			2
28	2			2
29	1			1
30	1			1
31	2			2
32	1			1
33	2			2
34	1			1
35	1			1
36	1			1
37	1			1
38	1			1
39	1			1
40	1			1
Total	319	1	1	321

IS*6110*-RFLP genotypes GD=1	12-MIRU + ETR A, B, C (GD=1)				Total
	1	2	3	4	
1	219		1		220
2	1				1
3	1				1
4	1				1
5	4				4
6	33				33
7	1				1
8	9				9
9	8				8
10	1				1
11	1				1
12	1				1
13	1				1
14	1				1
15	2				2
16	1				1
17	2				2
18	1				1
19	3				3
20	1				1
21	4			1	5
22	1				1
23	1				1
24	1				1
25	1				1
26	1				1
27	2				2
28	2				2
29	1				1
30		1			1
31	2				2
32	1				1
33	2				2
34	1				1
35	1				1
36	1				1
37	1				1
38	1				1
39	1				1
40	1				1
Total	318	1	1	1	321

IS*6110*-RFLP genotypes GD=1	12-MIRU + hyper-variable loci (GD=1)											
	1	2	3	4	5	6	7	8	9	10	11	Total
1	216		1	1	1					1		220
2	1											1
3	1											1
4	1											1
5	4											4
6	32	1										33
7	1											1
8	9											9
9	8											8
10	1											1
11	1											1
12	1											1
13	1											1
14	1											1
15	2											2
16	1											1
17	2											2
18	1											1
19	2								1			3
20	1											1
21	4										1	5
22	1											1
23	1											1
24	1											1
25	1											1
26	1											1
27	2											2
28	2											2
29	1											1
30								1				1
31						1	1					2
32	1											1
33	2											2
34	1											1
35	1											1
36	1											1
37	1											1
38	1											1
39	1											1
40	1											1
Total	311	1	1	1	1	1	1	1	1	1	1	321

IS*6110*-RFLP genotypes GD=1	15-MIRU-VNTR (GD-1)											Total
	1	2	3	4	5	6	7	8	9	10	11	
1		219								1		220
2		1										1
3	1											1
4	1											1
5		4										4
6	33											33
7	1											1
8	9											9
9	8											8
10	1											1
11	1											1
12	1											1
13	1											1
14	1											1
15	2											2
16	1											1
17	2											2
18	1											1
19	3											3
20	1											1
21	4										1	5
22									1			1
23		1										1
24	1											1
25					1							1
26						1						1
27							2					2
28								2				2
29		1										1
30				1								1
31			2									2
32		1										1
33		2										2
34	1											1
35	1											1
36		1										1
37	1											1
38		1										1
39		1										1
40	1											1
Total	77	232	2	1	1	1	2	2	1	1	1	321

IS*6110*-RFLP genotypes GD=1	15-MIRU-VNTR + hyper-variable loci (GD=1)																				Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1				216	1	1	1												1		220
2				1																	1
3																1					1
4															1						1
5				4																	4
6	32	1																			33
7	1																				1
8	9																				9
9	8																				8
10	1																				1
11	1																				1
12	1																				1
13	1																				1
14	1																				1
15	2																				2
16	1																				1
17	2																				2
18	1																				1
19	2													1							3
20	1																				1
21	4																			1	5
22																		1			1
23				1																	1
24	1																				1
25										1											1
26											1										1
27												2									2
28													2								2
29				1																	1
30									1												1
31								2													2
32				1																	1
33				2																	2
34	1																				1
35			1																		1
36				1																	1
37																	1				1
38				1																	1
39				1																	1
40	1																				1
Total	71	1	1	229	1	1	1	2	1	1	1	2	2	1	1	1	1	1	1	1	321

IS*6110*-RFLP genotypes GD=1	24-MIRU-VNTR (GD=1)															Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1		219												1		220
2		1														1
3	1															1
4	1															1
5		4														4
6	33															33
7	1															1
8	9															9
9	8															8
10								1								1
11	1															1
12	1															1
13	1															1
14							1									1
15								2								2
16	1															1
17	2															2
18	1															1
19	3															3
20								1								1
21									4						1	5
22													1			1
23		1														1
24	1															1
25						1										1
26										1						1
27											2					2
28												2				2
29		1														1
30				1												1
31			2													2
32		1														1
33		2														2
34	1															1
35	1															1
36		1														1
37	1															1
38		1														1
39		1														1
40					1											1
Total	67	232	2	1	1	1	1	4	4	1	2	2	1	1	1	321

IS*6110*-RFLP genotypes GD=1	24-MIRU-VNTR + hyper-variable loci (GD=1)																											Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	
1				216	1	1	1																			1		220
2				1																								1
3																						1						1
4																					1							1
5				4																								4
6	32	1																										33
7	1																											1
8									9																			9
9										8																		8
10																1												1
11	1																											1
12	1																											1
13	1																											1
14															1													1
15																2												2
16	1																											1
17	2																											2
18	1																											1
19	2												1															3
20																1												1
21																	4										1	5
22																								1				1
23				1																								1
24																									1			1
25														1														1
26																		1										1
27																			2									2
28																				2								2
29				1																								1
30											1																	1
31								2																				2
32				1																								1
33				2																								2
34	1																											1
35			1																									1
36				1																								1
37																							1					1
38				1																								1
39				1																								1
40												1																1
Total	43	1	1	229	1	1	1	2	9	8	1	1	1	1	1	4	4	1	2	2	1	1	1	1	1	1	1	321

Table S2. Sublineage 5 drug-resistance mutations and MIRU-VNTR repeat numbers

								Number of Repeats (hexadecimal Notation)																														
Isolate No	IS6110 RFLP	Inh	Rif	*katG* mutation	*rpoB* mutation	*embB*	*rrs*	SubLineage	MIRU02	MIRU04	MIRU10	MIRU16	MIRU20	MIRU23	MIRU24	MIRU26	MIRU27	MIRU31	MIRU39	MIRU40	VNTR1955	VNTR2165	QUB11b	QUB26b	VNTR0424	VNTR2401	VNTR4156	VNTR3690	ETR-C	VNTR2347	ETR-B	Mtub34	QUB3232	VNTR3820	VNTR4120
1945	551	S	R					1	2	2	3	3	2	5	1	6	3	5	3	1	4	5	5	9	4	4	4	3	4	3	2	3	D	A	9
2701	515	NT	NT					2	2	2	3	3	2	5	1	7	3	5	3	3	4	5	6	9	4	4	6	3	3	3	2	3	D	C	9
2704	515	NT	NT					2	2	2	3	3	2	5	1	7	3	5	3	3	4	5	6	9	4	4	6	3	3	3	2	3	D	C	9
4690	660	S	S					2	2	2	3	3	2	5	1	7	3	5	3	3	4	5	6	9	4	4	6	3	3	3	2	3	D	C	A
1453	907	NT	NT					3	2	2	3	3	2	5	1	7	3	5	3	3	4	5	6	9	4	4	6	3	3	3	2	3	D	C	9
4510	907	NT	NT					3	2	2	3	3	2	5	1	7	3	5	3	3	4	5	6	9	4	4	5	3	3	3	2	3	D	C	9
1176	484	NT	NT					4	2	2	3	3	2	5	1	7	3	4	3	3	6	5	5	7	4	4	4	3	3	3	2	3	F	D	B
129	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
252	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
357	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
502	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	8	D	2
513	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
553	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
698	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
785	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
1028	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
1098	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
1154	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
1281	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
1306	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	3	D	A
1464	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
1909	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
2024	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
2074	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	1	3	2	3	E	D	A

Isolate No	IS6110 RFLP	Inh	Rif	*katG* mutation	*rpoB* mutation	*embB*	*rrs*	Number of Repeats (hexadecimal Notation)																												
								SubLineage	MIRU02	MIRU04	MIRU10	MIRU16	MIRU20	MIRU23	MIRU24	MIRU26	MIRU27	MIRU31	MIRU39	MIRU40	VNTR1955	VNTR2165	QUB11b	QUB26b	VNTR0424	VNTR2401	VNTR4156	VNTR3690	ETR-C	VNTR2347	ETR-B	Mtub34	QUB3232	VNTR3820	VNTR4120
2173	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	3	D	A
2236	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	9
2600	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	3	D	A
3096	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
3111	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
3368	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
3568	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
3797	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
4344	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
4414	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	2	3	3	3	3	2	3	E	D	A
5008	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
5516	213	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
953	218	S	S	WT	WT	not tested	not tested	5	2	3	3	3	2	5	1	7	3	5	3	3	5	5	3	8	4	4	3	3	3	3	2	3	C	D	A
1646	240	S	S	WT	WT	not tested	not tested	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
819	294	S	S	WT	WT	not tested	not tested	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
1149	316	S	R	WT	526 (CAC to TAC)	not tested	not tested	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	3	8	4	4	3	3	3	3	2	3	E	D	A
1944	405	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	2	3	3	3	3	2	3	C	3	A
2125	535	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
2147	535	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
3556	606	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
3908	667	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
4158	716	R	R	315 (AGC to ACC)	531 (TCG to TTG)	306 (ATG to ATA)	513 (CAG to CCG)	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	B	3	7
4570	810	NT	NT	WT	WT	not tested	not tested	5	2	0	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	D	A
1125	205	S	S					6	2	2	3	3	2	5	1	7	3	5	3	3	5	5	5	6	4	4	3	3	3	3	2	3	D	D	3
407	206	S	S					6	2	2	3	3	2	5	1	7	3	5	3	3	5	5	5	6	4	4	3	3	3	3	2	3	D	D	3
261	219	NT	NT					6	2	2	3	3	2	5	1	6	3	5	3	3	5	4	5	8	4	4	3	3	3	3	2	3	D	D	A

Isolate No	IS6110 RFLP	Inh	Rif	*katG* mutation	*rpoB* mutation	*embB*	*rrs*	Number of Repeats (hexadecimal Notation)																												
								SubLineage	MIRU02	MIRU04	MIRU10	MIRU16	MIRU20	MIRU23	MIRU24	MIRU26	MIRU27	MIRU31	MIRU39	MIRU40	VNTR1955	VNTR2165	QUB11b	QUB26b	VNTR0424	VNTR2401	VNTR4156	VNTR3690	ETR-C	VNTR2347	ETR-B	Mtub34	QUB3232	VNTR3820	VNTR4120
354	219	S	S					6	2	2	3	3	2	5	1	8	3	5	3	3	5	5	5	8	4	4	3	3	3	3	2	3	D	D	A
1307	219	NT	NT					6	2	2	3	3	2	5	1	6	3	5	3	3	5	5	5	8	4	4	3	3	3	3	2	3	D	D	A
1641	219	NT	NT					6	2	2	3	3	2	5	1	6	3	5	3	3	5	5	5	8	4	4	3	3	3	3	2	3	D	D	A
5573	219	S	S					6	2	2	3	3	2	5	1	5	3	5	3	3	5	5	5	8	4	4	3	3	3	3	2	3	C	D	A
954	220	S	S					6	2	2	3	3	2	5	1	8	3	5	3	3	5	4	5	8	4	4	3	3	3	3	2	3	D	D	A
1716	220	S	R					6	2	2	3	3	2	5	1	8	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	D	D	A
1973	220	NT	NT					6	2	2	3	3	2	5	1	8	3	5	3	3	5	5	5	8	4	4	3	3	3	3	2	3	D	D	A
2332	220	NT	NT					6	2	2	3	3	2	5	1	8	3	5	3	3	5	5	5	8	4	2	3	3	3	3	2	3	D	C	A
3166	220	R	S					6	2	2	3	3	2	5	1	8	3	5	3	3	5	5	5	8	4	4	3	3	3	3	2	3	D	D	A
4036	220	R	S					6	2	2	3	3	2	5	1	8	3	5	3	3	5	4	5	8	4	4	3	3	3	3	2	3	C	D	A
4104	220	NT	NT					6	2	2	3	3	2	5	1	8	3	5	3	3	5	5	5	8	4	4	3	3	3	3	2	3	C	D	A
4473	220	R	NT					6	2	2	3	3	2	5	1	8	3	5	3	3	5	5	5	8	4	4	3	3	3	3	2	3	D	D	A
4493	220	NT	NT					6	2	2	3	3	2	5	1	8	3	5	3	3	5	5	5	8	4	4	3	3	3	3	2	3	C	D	A
4576	220	R	S					6	2	2	3	3	2	5	1	8	3	5	3	3	5	5	5	8	4	4	3	3	3	3	2	3	C	D	2
4880	220	NT	NT					6	2	2	3	3	2	5	1	8	3	5	3	3	5	5	5	8	4	4	3	3	3	3	2	3	C	D	A
1189	223	R	S					6	2	2	3	3	2	5	1	7	3	5	4	4	5	5	1	8	4	4	3	3	3	1	2	3	D	D	A
318	307	S	S					6	2	2	3	3	2	5	1	7	3	5	4	3	5	5	1	8	4	4	3	3	3	1	2	3	D	D	A
1507	307	NT	NT					6	2	2	3	3	2	5	1	7	3	5	4	3	5	5	1	8	4	4	3	3	3	1	2	3	D	D	A
204	319	S	S					6	2	2	3	3	2	5	1	7	2	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	E	3	9
919	319	S	S					6	2	2	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	3	D	A
2387	319	S	S					6	2	2	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	2	3	3	3	3	2	3	D	D	9
3514	319	NT	NT					6	2	2	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	D	3	9
700	322	NT	NT					6	2	2	3	3	2	5	1	8	3	5	3	3	5	5	5	8	4	4	3	3	3	3	2	3	D	D	B
1844	323	R	NT					6	2	2	3	3	2	5	1	7	3	5	4	3	5	5	1	8	4	4	3	3	3	1	2	3	D	D	A
3027	323	NT	NT					6	2	2	3	3	2	5	1	7	3	5	4	3	5	4	1	8	4	4	3	4	3	1	2	3	B	D	A
4456	323	S	S					6	2	2	3	3	2	5	1	7	3	5	4	3	5	5	1	8	4	4	3	3	3	1	2	3	C	D	A

Isolate No	IS6110 RFLP	Inh	Rif	*katG* mutation	*rpoB* mutation	*embB*	*rrs*	Number of Repeats (hexadecimal Notation)																												
								SubLineage	MIRU02	MIRU04	MIRU10	MIRU16	MIRU20	MIRU23	MIRU24	MIRU26	MIRU27	MIRU31	MIRU39	MIRU40	VNTR1955	VNTR2165	QUB11b	QUB26b	VNTR0424	VNTR2401	VNTR4156	VNTR3690	ETR-C	VNTR2347	ETR-B	Mtub34	QUB3232	VNTR3820	VNTR4120
1696	343	NT	NT					6	2	2	3	3	2	5	1	7	3	5	4	3	5	5	1	8	4	4	3	3	3	1	2	3	D	D	A
1823	345	NT	NT					6	2	2	3	3	2	5	1	7	3	5	3	3	8	5	5	8	4	4	3	3	3	3	2	3	D	D	9
1892	402	NT	NT					6	2	2	3	3	2	5	1	7	4	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	B	D	A
3255	618	NT	NT					6	2	2	4	3	2	6	1	5	3	3	2	3	4	5	3	7	4	4	3	1	3	3	2	1	3	8	2
3361	618	S	S					6	2	2	3	3	2	5	1	7	3	5	2	3	5	5	1	8	4	4	3	3	3	1	2	3	D	D	A
3654	618	NT	NT					6	2	2	3	3	2	5	1	7	3	5	4	3	5	4	1	8	4	4	3	3	3	1	2	3	D	D	A
3257	650	NT	NT					6	2	2	3	2	2	5	1	7	3	5	3	3	5	3	3	8	4	4	3	3	3	3	2	3	C	D	8
4126	650	NT	NT					6	2	2	3	2	2	5	1	7	3	5	3	3	5	5	3	8	4	4	3	3	3	3	2	3	D	D	8
3518	669	R	S					6	2	2	3	3	2	5	1	7	2	5	3	3	5	4	4	8	4	4	3	3	3	3	2	3	3	8	9
3954	670	NT	NT					6	2	2	3	3	2	5	1	7	3	5	3	3	5	5	4	8	4	4	3	3	3	3	2	3	D	D	9
3047	823	NT	NT					6	2	2	3	3	2	5	1	7	3	5	3	3	5	5	4	6	4	4	3	3	3	3	2	3	A	D	3
5610	985	NT	NT					6	2	2	3	3	2	5	1	7	2	5	2	3	5	5	4	7	4	4	3	3	3	3	2	3	D	C	A
649	2	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
600	187	S	S					7	2	2	2	3	2	5	1	7	3	3	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	9
1639	207	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
2189	207	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	3	5	6	4	4	3	4	5	3	2	3	B	D	A
4048	207	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
137	208	NT	NT					7	2	2	2	3	2	5	1	7	4	5	4	3	5	5	5	6	4	4	3	4	5	3	3	3	B	D	A
151	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
313	208	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
439	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
638	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	3	3	B	D	A
650	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	3	3	B	D	A
714	208	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	6
717	208	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
764	208	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A

Isolate No	IS6110 RFLP	Inh	Rif	*katG* mutation	*rpoB* mutation	*embB*	*rrs*	Number of Repeats (hexadecimal Notation)																												
								SubLineage	MIRU02	MIRU04	MIRU10	MIRU16	MIRU20	MIRU23	MIRU24	MIRU26	MIRU27	MIRU31	MIRU39	MIRU40	VNTR1955	VNTR2165	QUB11b	QUB26b	VNTR0424	VNTR2401	VNTR4156	VNTR3690	ETR-C	VNTR2347	ETR-B	Mtub34	QUB3232	VNTR3820	VNTR4120
767	208	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
768	208	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
815	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	C	A
823	208	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
993	208	S	S					7	2	2	2	3	2	5	1	7	4	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1042	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1058	208	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1130	208	S	S					7	2	2	2	3	2	5	1	7	4	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1232	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1293	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1376	208	NT	NT					7	2	2	2	3	2	5	1	7	4	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1382	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	N/A	N/A
1411	208	NT	NT					7	2	2	2	3	2	5	1	7	4	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1432	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1524	208	NT	NT					7	2	2	2	3	2	5	1	7	4	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1537	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	5	4	3	4	5	3	2	3	B	D	A
1609	208	NT	NT					7	2	2	2	3	2	5	1	3	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1620	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	8
1763	208	S	S					7	2	2	2	3	2	5	1	7	4	4	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1779	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	2
1818	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
2011	208	S	S					7	2	2	2	3	2	5	1	7	4	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
2112	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
2229	208	NT	NT					7	2	2	2	3	2	5	1	7	3	3	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
2256	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	2	D	A
2377	208	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	C	A

Isolate No	IS6110 RFLP	Inh	Rif	*katG* mutation	*rpoB* mutation	*embB*	*rrs*	Number of Repeats (hexadecimal Notation)																												
								SubLineage	MIRU02	MIRU04	MIRU10	MIRU16	MIRU20	MIRU23	MIRU24	MIRU26	MIRU27	MIRU31	MIRU39	MIRU40	VNTR1955	VNTR2165	QUB11b	QUB26b	VNTR0424	VNTR2401	VNTR4156	VNTR3690	ETR-C	VNTR2347	ETR-B	Mtub34	QUB3232	VNTR3820	VNTR4120
2890	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3222	208	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	4	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3259	208	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3305	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3327	208	NT	NT					7	2	2	2	3	2	5	1	7	4	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3339	208	R	S					7	2	2	2	3	2	5	1	7	4	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	B
3496	208	NT	NT					7	2	2	2	3	2	5	1	7	4	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	1	D	A
3641	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3646	208	NT	NT					7	2	2	2	3	2	5	1	7	4	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3688	208	S	S					7	2	2	2	3	2	5	1	8	4	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3835	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4146	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4153	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4174	208	NT	NT					7	2	2	2	3	2	5	1	7	4	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4225	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	9
4290	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4329	208	NT	NT					7	2	2	2	3	2	5	1	7	4	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4424	208	NT	NT					7	2	2	2	3	2	5	1	7	3	3	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4449	208	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4551	208	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
5016	208	S	S					7	2	2	2	3	2	5	1	7	3	5	3	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
5187	208	S	S					7	2	2	2	3	2	5	1	9	3	5	3	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
5357	208	S	S					7	2	2	2	3	2	5	1	7	4	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
212	209	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	3	3	B	D	A
215	209	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	3	3	B	D	A
284	209	NT	NT					7	2	2	2	3	2	5	1	7	3	3	4	3	5	5	5	6	4	4	3	4	5	3	3	3	B	D	A

Isolate No	IS6110 RFLP	Inh	Rif	*katG* mutation	*rpoB* mutation	*embB*	*rrs*	Number of Repeats (hexadecimal Notation)																												
								SubLineage	MIRU02	MIRU04	MIRU10	MIRU16	MIRU20	MIRU23	MIRU24	MIRU26	MIRU27	MIRU31	MIRU39	MIRU40	VNTR1955	VNTR2165	QUB11b	QUB26b	VNTR0424	VNTR2401	VNTR4156	VNTR3690	ETR-C	VNTR2347	ETR-B	Mtub34	QUB3232	VNTR3820	VNTR4120
469	209	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
495	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	3	3	B	D	A
599	209	R	R					7	2	2	2	3	2	5	1	7	3	3	4	3	5	5	5	6	4	4	3	4	5	3	2	3	4	D	A
605	209	S	S					7	2	2	2	3	2	5	1	7	3	3	4	3	5	5	5	6	4	4	3	4	5	3	3	3	B	D	A
618	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
634	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	3	3	B	D	A
687	209	R	R					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
715	209	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
735	209	S	S					7	2	2	2	3	2	5	1	7	3	3	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
759	209	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
795	209	R	R					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
972	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	2	4	5	3	2	3	B	D	A
1151	209	S	S					7	2	2	2	3	2	5	1	7	3	3	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1158	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1170	209	S	S					7	2	2	2	3	2	5	1	7	3	3	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	C	9
1182	209	R	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1259	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1267	209	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1279	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1326	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1360	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1383	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1421	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1427	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1547	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	9
1555	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A

Isolate No	IS6110 RFLP	Inh	Rif	*katG* mutation	*rpoB* mutation	*embB*	*rrs*	Number of Repeats (hexadecimal Notation)																												
								SubLineage	MIRU02	MIRU04	MIRU10	MIRU16	MIRU20	MIRU23	MIRU24	MIRU26	MIRU27	MIRU31	MIRU39	MIRU40	VNTR1955	VNTR2165	QUB11b	QUB26b	VNTR0424	VNTR2401	VNTR4156	VNTR3690	ETR-C	VNTR2347	ETR-B	Mtub34	QUB3232	VNTR3820	VNTR4120
1575	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1588	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1736	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1777	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1780	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	3	D	2
1852	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1871	209	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1924	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1948	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1949	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	8
2076	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
2087	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	1	3	2	3	B	D	A
2206	209	NT	NT					7	2	2	2	3	2	5	1	7	3	3	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
2221	209	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	4	5	5	6	4	4	3	4	5	3	2	3	B	D	A
2307	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	A	D	A
2330	209	NT	NT					7	2	2	2	3	2	5	1	7	3	3	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
2391	209	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
2412	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	4	5	5	6	4	4	3	4	5	3	2	3	B	D	A
2413	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
2463	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
2598	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
2891	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	4	5	5	6	4	4	3	4	5	3	2	3	B	D	A
2903	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	4	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3005	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3033	209	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3039	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A

Isolate No	IS6110 RFLP	Inh	Rif	*katG* mutation	*rpoB* mutation	*embB*	*rrs*	Number of Repeats (hexadecimal Notation)																												
								SubLineage	MIRU02	MIRU04	MIRU10	MIRU16	MIRU20	MIRU23	MIRU24	MIRU26	MIRU27	MIRU31	MIRU39	MIRU40	VNTR1955	VNTR2165	QUB11b	QUB26b	VNTR0424	VNTR2401	VNTR4156	VNTR3690	ETR-C	VNTR2347	ETR-B	Mtub34	QUB3232	VNTR3820	VNTR4120
3052	209	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3188	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3207	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3314	209	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	2	2	3	B	D	A
3324	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3369	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3414	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3416	209	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3506	209	NT	NT					7	2	2	2	3	2	5	1	7	3	3	4	3	5	5	5	6	4	4	3	4	5	3	2	3	3	D	7
3582	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	3	D	8
3589	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3594	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	4	D	A
3687	209	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3752	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3847	209	NT	NT					7	2	2	2	3	2	5	1	7	3	3	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3885	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	3	4	3	4	5	3	2	3	B	D	A
3973	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4002	209	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4004	209	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4122	209	NT	NT					7	2	2	2	3	2	5	1	7	3	3	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4151	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	F	A
4186	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4187	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4195	209	NT	NT					7	2	2	2	3	2	5	1	7	3	3	4	3	5	5	5	6	4	4	3	4	5	3	2	3	3	7	2
4220	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4221	209	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A

Isolate No	IS6110 RFLP	Inh	Rif	*katG* mutation	*rpoB* mutation	*embB*	*rrs*	Number of Repeats (hexadecimal Notation)																												
								SubLineage	MIRU02	MIRU04	MIRU10	MIRU16	MIRU20	MIRU23	MIRU24	MIRU26	MIRU27	MIRU31	MIRU39	MIRU40	VNTR1955	VNTR2165	QUB11b	QUB26b	VNTR0424	VNTR2401	VNTR4156	VNTR3690	ETR-C	VNTR2347	ETR-B	Mtub34	QUB3232	VNTR3820	VNTR4120
4235	209	S	S					7	2	2	2	3	2	5	1	7	3	4	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4271	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4276	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4292	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4376	209	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4395	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4430	209	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4436	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4447	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4459	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4463	209	NT	NT					7	2	2	2	3	2	5	1	7	3	3	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4464	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4477	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4574	209	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4942	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
5218	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	3	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
5240	209	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
5460	209	S	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
5527	209	NT	NT					7	2	2	2	3	2	5	1	5	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
5569	209	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3189	211	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4059	211	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4896	211	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3941	212	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
191	215	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
230	215	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	3	3	B	D	A

Isolate No	IS6110 RFLP	Inh	Rif	*katG* mutation	*rpoB* mutation	*embB*	*rrs*	Number of Repeats (hexadecimal Notation)																														
								SubLineage	MIRU02	MIRU04	MIRU10	MIRU16	MIRU20	MIRU23	MIRU24	MIRU26	MIRU27	MIRU31	MIRU39	MIRU40	VNTR1955	VNTR2165	QUB11b	QUB26b	VNTR0424	VNTR2401	VNTR4156	VNTR3690	ETR-C	VNTR2347	ETR-B	Mtub34	QUB3232	VNTR3820	VNTR4120
590	215	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
791	215	R	R					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	2	5	3
1081	215	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1390	215	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1396	215	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1652	215	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1708	215	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3030	215	S	S					7	2	2	2	3	2	5	1	7	4	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3114	215	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3555	215	NT	NT					7	2	2	2	3	2	5	1	7	4	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4023	215	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4166	215	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4200	215	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4256	215	NT	NT					7	2	2	2	3	2	5	1	7	4	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
2115	216	NT	NT					7	2	2	4	3	2	5	1	5	3	3	2	3	4	5	3	7	4	2	3	1	5	3	2	3	3	6	2
2274	216	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
2542	216	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3915	216	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4097	216	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1389	217	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1492	217	R	R					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1821	217	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3289	217	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1428	255	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1504	255	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	E	D	A
1522	255	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A

Isolate No	IS6110 RFLP	Inh	Rif	*katG* mutation	*rpoB* mutation	*embB*	*rrs*	Number of Repeats (hexadecimal Notation)																												
								SubLineage	MIRU02	MIRU04	MIRU10	MIRU16	MIRU20	MIRU23	MIRU24	MIRU26	MIRU27	MIRU31	MIRU39	MIRU40	VNTR1955	VNTR2165	QUB11b	QUB26b	VNTR0424	VNTR2401	VNTR4156	VNTR3690	ETR-C	VNTR2347	ETR-B	Mtub34	QUB3232	VNTR3820	VNTR4120
1994	255	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
2958	255	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4012	308	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1515	310	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
2315	396	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	4	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3050	411	S	S					7	2	2	2	3	2	5	1	7	4	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
2888	412	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
2895	426	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3066	502	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3181	502	R	R					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
2675	516	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	3	3	2	3	B	D	A
2057	537	S	S					7	2	2	2	3	2	5	1	7	3	3	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
2338	537	NT	NT					7	2	2	2	3	2	5	1	7	3	3	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1015	548	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1463	548	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
793	595	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1813	596	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
1518	598	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4204	656	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3815	657	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3457	661	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	4	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4352	661	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	4	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3584	662	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4006	664	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	4	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3393	665	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3486	665	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A

Isolate No	IS6110 RFLP	Inh	Rif	*katG* mutation	*rpoB* mutation	*embB*	*rrs*	Number of Repeats (hexadecimal Notation)																												
								SubLineage	MIRU02	MIRU04	MIRU10	MIRU16	MIRU20	MIRU23	MIRU24	MIRU26	MIRU27	MIRU31	MIRU39	MIRU40	VNTR1955	VNTR2165	QUB11b	QUB26b	VNTR0424	VNTR2401	VNTR4156	VNTR3690	ETR-C	VNTR2347	ETR-B	MtuB34	QUB3232	VNTR3820	VNTR4120
5189	698	S	S					7	2	2	2	3	2	5	1	7	4	5	4	3	5	5	4	6	4	4	3	4	5	3	2	3	B	D	A
4013	743	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4047	750	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
2383	773	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
3809	785	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
571	851	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
4437	871	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
5144	871	NT	NT					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
5080	878	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	2	4	3	4	5	3	2	3	B	D	A
4457	879	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A
5629	938	S	S					7	2	2	2	3	2	5	1	7	3	5	4	3	5	5	5	6	4	4	3	4	5	3	2	3	B	D	A

R = Resistant

S = Susceptible

NT = Not Tested

Chapter 5

Mixed *Mycobacterium tuberculosis* infection in patients resident in a high HIV prevalence setting

Hanekom, M., Van de Berg, D., Streicher, E. M., Cox, H., McDermid, C., Bosman, M., Gey van Pittius, N. C., Victor, T. C., Kidd, M., van Helden, P. D. and Warren, R. M.

To be submitted
(Formatted as per submission)

Mixed *Mycobacterium tuberculosis* infection in patients resident in a high HIV prevalence setting

Hanekom M[1], Van de Berg D[1], Streicher EM[1], Cox H[2], McDermid C[2], Bosman M[3], Gey van Pittius NC[1], Victor T[1], Kidd M[4], van Helden PD[1], Warren RM[1]*

[1]DST/NRF Centre of Excellence for Biomedical Tuberculosis Research / MRC Centre for Molecular and Cellular Biology, Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, [2]Médecins Sans Frontières, Cape Town, South Africa, [3]Biotec Laboratories, Ltd., c/o National Health Laboratory Service, Cape Town, Western Cape, South Africa, [4]Centre for Statistical Consultation, Stellenbosch University, South Africa.

*Corresponding Author

DST/NRF Centre of Excellence for Biomedical Tuberculosis Research / MRC Centre for Molecular and Cellular Biology, Division of Molecular Biology and Human Genetics, Faculty of Health Sciences, Stellenbosch University, PO Box 19063, Tygerberg, South Africa, 7505
Telephone: 021 938 9073
Fax: 021 938 9476
Email: rw1@sun.ac.za

Running title: Mixed infections

Keywords: *Mycobacterium tuberculosis*, HIV prevalence setting

Abstract

Word count = 332

Background: The frequency and epidemiological significance of mixed *Mycobacterium tuberculosis* infections in patients with pulmonary tuberculosis is still partly unknown.

Aim: First to develop methods to identify the presence or absence of the most prominent phylogenetic lineages in primary cultures. Secondly, using this genotyping data to understand the population structure of mixed *M. tuberculosis* infections in sputum specimens collected from pulmonary tuberculosis cases resident in a high HIV prevalent setting in South Africa.

Method: *M. tuberculosis* cultures (Mycobacterial Growth indicator Tube (MGIT) and/or Löwenstein-Jensen (LJ)), clinical and demographic data from pulmonary tuberculosis patients resident in Khayelitsha, South Africa were collected during the period April to October 2008 as part of a drug surveillance study. Isolates were genotyped using an in-house PCR-based method which identified the most prominent phylogenetic lineages; Beijing, Haarlem, Low Copy Clade, S-family, and LAM.

Results: MGIT and LJ cultures were analysed from 224 patients. Seven cultures failed to amplify on both media respectively. Of the remaining cultures (n=213), 181 (85%) showed infection with a single strain and 32 (15%) showed the presence of more than 1 strain on both MGIT and LJ media. The population structure of mixed *M. tuberculosis* strain genotypes was concordant in 62.5% of the cultures when cultured on either MGIT or LJ media. The remaining 37% showed discordance in either that one or both of the strain genotypes were absent from either the MGIT or LJ media. Twenty-eight percent of the mixed infections were missed by culture on MGIT media, while 9% were missed when using LJ culture media. Strains of the Beijing genotype were strongly associated with mixed infections ($p < 0.01$). Mixed infection was not associated with age and sex, treatment history, smear grading or drug resistance.

Conclusion: Mixed infection was observed in 15% of the cultures in this study population. These may explain heterogeneous clinical responses to anti-tuberculosis treatment or masking

of drug resistance. The growth characteristics of *M. tuberculosis* on different media may influence our ability to detect mixed infections.

Introduction

Treatment failure is a significant concern to tuberculosis (TB) control programs. Programmatically, treatment failure is defined in patients who, while on treatment, remained or became smear-positive again five months or later after commencing treatment or a patient who was initially smear-negative before starting treatment and became smear-positive after the second month of treatment. This is usually thought to result from either undetected primary drug-resistance or acquisition of resistance as a consequence of poor patient treatment adherence. However, with the development of molecular epidemiological tools it has been possible to genotype strains present in a patient's sputum to monitor any changes in the *Mycobacterium tuberculosis* population structure over time. Evidence of mixed infection in a single host at a single point in time, suggesting reinfection before or during disease, was first observed using the phage typing method (1, 2). Subsequently, Yeh *et al* (1999) reported the first case of mixed infection using IS*6110*-based genotyping of *M. tuberculosis* (3). This was followed by several other studies documenting mixed infection (4, -6). However, the resolution of these methods to define the presence of a mixed infection was limited by the relative ratio of the two strains present in the specimen (7). This could in part be overcome by the analysis of individual bacterial colonies from a single specimen (8). Using a highly sensitivity PCR technique, Warren *et al* (2004) was able to show that 19% of the study participants in the given study setting were infected with both Beijing and non-Beijing strain genotypes (9). In response to this finding, it was questioned whether mixed infections were of concern for TB control, since such infections could influence the diagnosis of drug resistance if a patient was infected with both a drug sensitive and a drug resistant strain. To address this concern, a follow up study showed two additional mechanisms which may explain the emergence of drug resistance in a patient. Firstly, the patient may be exogenously reinfected with a drug resistant strain while on treatment for drug susceptible TB (10). Secondly, the patient may have been infected with both a drug susceptible and a drug resistant strain prior to progression towards disease. Subsequent treatment would select for the drug resistant strain leading to drug resistant TB (5, 10).

However, the PCR-based methods used in these studies were able to analyze a small subset of strains only with limited ability to differentiate among the different genotypes. Therefore, the true population structure of mixed infections in these study settings has remained largely unknown. First this study aimed to develop methods to identify the presence or absence of 5

different strain genotypes (Beijing, Latin-American and Mediterranean (LAM), S-family (F28), Low Copy Clade (LCC) and Haarlem) in primary cultures. Secondly, using this genotyping data to quantify the extent of mixed infections in a high incidence setting with a high prevalence of human immunodeficiency virus (HIV) co-infection and to identify risk factors associated with mixed infection.

Methods

Study population

During the period from April to October 2008, sputum specimens were collected from TB suspects resident in Khayelitsha, South Africa, where the incidence of TB is reported to be at least 1500/100 000 per annum and the TB/HIV co-infection rate is 70 % (City of Cape Town, Health Department, Statistics. 2009). All sputum specimens were submitted for culture and drug susceptibility testing to the National Health Laboratory Service. After decontamination, each sputum specimen was cultured on Mycobacterial Growth Indicator Tube (MGIT) liquid medium (Becton-Dickinson) and on Löwenstein-Jensen (LJ) solid medium. Aliquots of the respective cultures were boiled at 100°C to ensure sterilization and release of the *M. tuberculosis* DNA. Drug susceptibility testing results, smear grading and patient demographics including sex, age, and previous history of TB were recorded at diagnosis and stored in a database. This study was approved by the Ethics review board of Stellenbosch University.

Genotyping by molecular techniques

Positive primary cultures were genotyped by spoligotyping using the internationally standardized method (11). Insertion sequence junctions were cloned and sequenced as previously described 10. Primers were designed (table 1) complimentary to the IS*6110* internal sequence and the 3' and 5' chromosomal flanking sequences. *M. tuberculosis* genome sequences were amplified from a panel of *M. tuberculosis* strains representing the different genotypes present in Cape Town, South Africa. Only primer sets that uniquely amplified strains representative of the LAM (12), S, LCC and Haarlem genotypes were selected for subsequent analysis (Table 1). In addition, a primer set spanning the junction site of the Region of Difference (RD) 105 was used to specifically identify the presence of Beijing genotype strains (Table 1) (13).

The sensitivity and specificity of amplification for the different primer sets was determined by PCR amplification of a panel of genetically related and unrelated *M. tuberculosis* strains classified according to IS*6110* RFLP and spoligotyping (9). Briefly, DNA from 40 genetically known related and unrelated *M. tuberculosis* strains, representing 31 distinct strains, were amplified with primer sets 1 to 4 (Figure 1). The sensitivity and specificity of

primer set 1 to 3 was 100% (95%CI 85 – 100%) while the sensitivity of primer set 4 was 75% and specificity 50% (95%CI 85 – 100%) as this primer set recognized *M. tuberculosis* strains in both the LCC and Haarlem genotypes.

To determine the presence of a particular *M. tuberculosis* strain genotype in the selected cultures, DNA was PCR amplified in a total volume of 25 µl, containing 2 µl DNA template, 1 x Q-Buffer, 1 x enzyme buffer, 3.5 mM $MgCl_2$, 1.6 mM dNTP's, 25 pmol of each primer set (Table 1) and 0.25 U HotStar Taq DNA polymerase (Qiagen, Germany). Amplification was initiated by incubation at 95 ˚C for 15 min., followed by 45 cycles at 94 ˚C for 45 sec., 62 - 65 ˚C for 45 sec., and 72 ˚C for 45 sec. After the last cycle, the samples were incubated at 72 ˚C for 10 min. PCR amplification products were electrophoretically fractionated in 2.0 % agarose in 1 x SB pH 8.3 at 3.5V/cm for 4 hours, and visualized by staining with ethidium bromide. The presence of a particular *M. tuberculosis* genotype was defined when the amplified product corresponded to the predicted product size (Table 1).

To minimize the risk of laboratory cross-contamination during the PCR amplification, each procedure (preparation of the PCR reaction mixes, the addition of the DNA, the PCR amplification and the electrophoretic fractionation) was done in physically separated rooms. Negative controls (water) were included to control for reagent contamination.

Statistical Methods

The Kappa coefficient of conformance was used as a measure of conformance between the MGIT and LJ cultures. Comparisons of binary responses were done using cross tabulation and the Chi-square test. Comparisons of ordinal variables (eg. age) between mixed versus single cases were done using one-way ANOVA.

Results

Study population

During the period from April to October 2008, 614 patients enrolled in a drug surveillance study were diagnosed with TB by either routine smear or culture. The average age of these patients was 35.4 years and included 328 (53%) male patients and 223 (36%) female patients (age and gender data was unavailable for 63 patients). Primary MGIT and their corresponding LJ cultures were available from 224 of these cases and formed the cohort for subsequent analysis. The average age of these patients in this cohort was 34.6 years and included 121 (54%) male patients and 88 (39%) female patients (age and sex data was unavailable for 15 patients). Drug susceptibility testing was done on 199 primary cultures of which 119 (60%) were drug-sensitive and 50 (25%) were resistant to any drug. The remaining 30 (15%) cultures were contaminated. The patients represented in this cohort were not statistically different from the patients in the entire cohort according to the demographic measures of age, sex, previous history of tuberculosis and drug susceptibility patterns.

Genotyping by PCR amplification

Spoligotyping identified 188 genotypes cultured from MGIT media and 181 genotypes cultured from LJ media (Table 2). PCR amplification using the five primer sets was applied to determine the presence of Beijing, LAM, S-family, LCC or Haarlem strain genotypes of 244 primary MGIT and their corresponding LJ cultures. Seven MGIT and 7 LJ cultures failed to amplify and were excluded from further analysis. Table 2 shows the genotypes identified by the PCR-genotype method. In total, 181 single infections were identified on both MGIT and LJ media, of which 180 shared the same strain genotype and 1 had a different strain genotype (Table 2 and Figure 2). According to statistical analysis, concordance between the MGIT and LJ cultures was 0.74. Twenty-three mixed infections were identified on MGIT media and 29 on LJ media (Figure 2). The population structure of mixed infections on MGIT media included 14 Beijing and non-Beijing strain genotype combinations and 9 non-Beijing strain genotype combinations (Table 3). In addition, the population structure of mixed infections cultured on LJ media showed the presence of 21 Beijing and non-Beijing strain genotype combinations and 8 non-Beijing genotype combinations (Table 3).

In total, 32 (15%) mixed infections were identified by culture on MGIT and LJ media (Table 3). The population structure of these mixed infections included 21 Beijing and non-Beijing strain genotype combinations and 11 non-Beijing genotype combinations (Table 3). Twenty (62.5%) of the mixed infections showed concordant genotypes on both MGIT and LJ media, although an additional genotype was identified on one of the cultures grown on MGIT media as well as on one of the cultures grown on LJ media (Table 4). Furthermore, 11 of the mixed infections showed that one of the genotypes was present on both media while the other genotype was absent from either the MGIT or LJ media (Table 4). The remaining mixed infection showed dissimilar strain genotypes on both media (Table 4).

Strains of the Beijing ($p < 0.01$), LAM ($p = 0.02$), LCC ($p < 0.01$), Haarlem ($p < 0.01$) and "other strains" ($p < 0.01$) genotypes were more associated with mixed infections than strains of the S-family ($p = 0.28$) genotype. However, strains of the Beijing genotype were significantly more associated with mixed infections of cultures grown on MGIT media ($p < 0.01$) and LJ media ($p < 0.01$) than single infections grown on these 2 media. Also, the Beijing strain genotype was more frequently observed in the total number of mixed infections grown on both MGIT and LJ media ($p < 0.01$). Strains of the Haarlem genotype were more often observed in mixed infections when strains of the LAM and LCC genotypes were absent ($p = 0.03$).

Clinical, demographic and drug susceptibility characteristics of study cohort

The demographics of the 32 patients infected with more than one strain genotype were similar to that of the 181 patients assigned as having single infections (Table 5). No significant association could be demonstrated between mixed infection and age, sex, smear grading or drug-resistance (Table 5). Mono-resistant (resistance to any 1 drug only), sensitive and multidrug- resistant (MDR) (resistance to at least isoniazid and rifampicin) isolates were associated with mixed infection, while poly-resistance (resistance to 2 or more drugs excluding MDR) was not associated with mixed infections. However, the total number of poly-resistant and MDR isolates were too small to report on with any confidence (Table 5).

Discussion

In this study an improved PCR-based method was used to define the population structure of *M. tuberculosis* strain genotypes in primary cultures from patients diagnosed with pulmonary TB who were resident in Khayelitsha, South Africa. Using this method, it was possible to identify members of the Beijing, LAM, S-family, LCC and Haarlem genotypes and to demonstrate that 15% of sputum cultures had more than one genotype present. This was significantly lower than the 57% reported in an epidemiological field site near Cape Town, South Africa using a PCR-based method to identify only Beijing and non-Beijing strain genotypes (9). The difference could either be due to the difference in host population or the sensitivity of the method used to detect mixed infection. The study by Warren *et al* (2004) was conducted in a setting where HIV/TB co-infection was low and 75% of the cases had smear positive disease. In contrast, the proportion of cases with smear positive disease was only 55% in this study and HIV/TB co-infection was high. The lower smear positivity rate implies that they are less infectious thereby decreasing the risk of multiple infections. In addition, it is well established that HIV positive individuals are at a greater risk of progressing to disease following infection (14), thus their risk of reinfection during a short period of latent infection would be reduced. Alternatively, the methodology could have influenced the estimate of mixed infection. We cannot exclude the possibility that culture medium may have disturbed the relative abundance of the different genotypes present due to their different growth characteristics. Furthermore, Warren *et al* (2004) used a non-competitive PCR-based method as compared to the competitive method used in this study which would have influenced the sensitivity for detecting underlying strains. Previously, it was estimated that the competitive method was only able to identify mixed infections at a ratio of 1:150, thus mixed infection with a ratio of greater than 1:150 would be missed. Our estimate of mixed infection was also limited by the fact that the PCR-based method could only detect the presence or absence of a particular genotype and thus it was not possible to determine the extent of mixed infection with different strains from the same genotype.

In our study, we demonstrate for the first time a broader spectrum of genotype strains using a PCR-based method. Previous studies using PCR-based methods could only distinguish between Beijing and non-Beijing genotype strains (9). In addition, studies using IS*6110* DNA fingerprinting to detect mixed infections were low due to the presence of low-intensity bands (7). Furthermore, using this method, we were able to identify up to 4 different *M.*

tuberculosis strain genotypes in one primary culture. Similar findings were reported in previous studies (15, 16) where TB is endemic in these populations. Therefore, this phenomenon could be expected in high TB incidence settings where the infection pressure is high. However, the implication of mixed infections involving more than two strains is unclear. In contrast to our results, a study done in China using MIRU-VNTR typing, found 43% of mixed infections had non-Beijing genotype strains while Beijing genotype strains were relatively seldom found in mixed infections (17). In our study we show that strains of the Beijing genotype in this setting were over-represented in mixed infections of cultures grown on either LJ or MGIT media. This finding may suggest that the Beijing strain genotype found in this setting has special properties allowing it to reinfect patients. A previous study has shown that strains from a defined Beijing sublineage in different geographical settings may have selected for a distinct host population or that that host population was more vulnerable to strains from a defined Beijing sublineage (18). In mixed infections, strains of the Beijing genotype occurred in combination with all other genotypes tested (LAM, LCC, "other strains", S-family, Haarlem) by the PCR-based method. Thus, we hypothesise that the population structure of mixed infections may reflect pathogen-pathogen compatibility. Further studies are needed to determine whether certain genotypes occur together more frequently than other genotypes during a single disease episode, whether certain genotypes influence the growth of other genotypes *in vivo* and whether certain genotypes influence the rate of disease progression.

Comparison of demographic data between patients infected with a single strain genotype and multiple strain genotypes did not identify significant differences. In addition, there was no relationship between mixed infection, treatment history and drug resistance. This differs from the study by Warren *et al* (2004) which showed that mixed infections were more frequently observed in retreatment cases (9). It has been suggested that reinfection in recurrent TB may be explained by the fact that an ongoing TB infection may divert the immune response, thereby increasing the overall susceptibility to reinfection (9). This may also be the case in new TB cases co-infected with HIV where the immune response is suboptimal. Thus, this could explain the similar frequencies of mixed infection in both new and retreatment cases observed in this study.

The population structure of mixed *M. tuberculosis* strain genotypes was concordant in 62.5% of the cultures when cultured on either MGIT or LJ media. The remaining 37% showed discordance in either that one or both of the strain genotypes were absent from either the MGIT or LJ media. Twenty-eight percent of the mixed infections were missed by culture on MGIT media, while 9% were missed when using LJ culture media. Previously, it has been suggested that the culture method could provide a "microbiological bias" and lead to a discrepancy in the growth of mycobacteria and therefore should be considered when interpreting molecular epidemiological data. Our study supports previous reports which have suggested the use of a combination of both liquid and culture media for maximum recovery of *M. tuberculosis* (19, 20).

In summary, this study demonstrates that mixed infections with different *M. tuberculosis* genotypes is an important epidemiological phenomenon in TB patients resident in a setting with one of the highest burdens of HIV and TB co-infection. The clinical significance of mixed infection remains poorly understood although it has been previously shown that mixed infection may mask the detection of drug resistance. In this study, only one patient was infected with both a drug sensitive and a drug resistant strain. Furthermore, strains of the Beijing genotype were significantly associated with mixed infections which further support the hypothesis that strains of the Beijing genotype have evolved selective advantages enabling their increased dissemination. Lastly, the study shows that the growth characteristics of *M. tuberculosis* on different media may influence our ability to detect mixed infections.

Reference List

1. **Mankiewicz, E. & Liivak, M**. 1975. Phage types of *Mycobacterium tuberculosis* in cultures isolated from Eskimo patients. Am. Rev. Respir. Dis. 111:307-312.

2. **Bates, J.H., Stead, W.W. & Rado, T.A**. 1976. Phage type of tubercle bacilli isolated from patients with two or more sites of organ involvement. Am. Rev. Respir. Dis. 114:353-358.

3. **Yeh, R.W., Hopewell, P.C. & Daley, C.L**. 1999. Simultaneous infection with two strains of *Mycobacterium tuberculosis* identified by restriction fragment length polymorphism analysis. Int. J. Tuberc. Lung Dis. 3:537-539.

4. **Chaves, F., Dronda, F., Alonso-Sanz, M. & Noriega, A.R.** 1999. Evidence of exogenous reinfection and mixed infection with more than one strain of *Mycobacterium tuberculosis* among Spanish HIV-infected inmates. AIDS 13:615-620.

5. **Theisen, A., Reichel, C., Rüsch-Gerdes, S., Haas, W.H., Rockstroh, J.K., Spengler, U. et al.** 1995. Mixed-strain infection with a drug-sensitive and multidrug-resistant strain of *Mycobacterium tuberculosis*. Lancet 345:1512.

6. **Niemann, S., Richter, E., Rüsch-Gerdes, S., Schlaak, M. & Greinert, U.** 2000. Double infection with a resistant and a multidrug-resistant strain of *Mycobacterium tuberculosis*. Emerging Infect. Dis. 6:548-551.

7. **Richardson, M., Carroll, N.M., Engelke, E., Van Der Spuy, G.D., Salker, F., Munch, Z. et al**. 2002. Multiple *Mycobacterium tuberculosis* strains in early cultures from patients in a high-incidence community setting. J. Clin. Microbiol. 40:2750-2754.

8. **Braden, C.R., Morlock, G.P., Woodley, C.L., Johnson, K.R., Colombel, A.C., Cave, M.D. et al**. 2001. Simultaneous infection with multiple strains of *Mycobacterium tuberculosis*. Clin. Infect. Dis. 33:e42-47.

9. **Warren, R.M., Victor, T.C., Streicher, E.M., Richardson, M., Beyers, N., Gey van Pittius, N.C. et al**. 2004. Patients with active tuberculosis often have different strains in the same sputum specimen. Am. J. Respir. Crit. Care Med. 169:610-614.

10. **van Rie, A., Victor, T.C., Richardson, M., Johnson, R., van der Spuy, G.D., Murray, E.J. et al**. 2005. Reinfection and mixed infection cause changing *Mycobacterium tuberculosis* drug-resistance patterns. Am. J. Respir. Crit. Care Med. 172:636-642.

11. **Kamerbeek, J., Schouls, L., Kolk, A., van Agterveld, M., van Soolingen, D., Kuijper, S. et al**. 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J. Clin. Microbiol. 35:907-914.

12. **Marais, B.J., Victor, T.C., Hesseling, A.C., Barnard, M., Jordaan, A., Brittle, W. et al.** 2006. Beijing and Haarlem genotypes are overrepresented among children with drug-resistant tuberculosis in the Western Cape Province of South Africa. J. Clin. Microbiol. 44:3539-3543.

13. **Tsolaki, A.G., Gagneux, S., Pym, A.S., Goguet de la Salmoniere, Y.L., Kreiswirth, B.N., Van Soolingen, D. et al.** 2005. Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis*. J. Clin. Microbiol. 43:3185-3191.

14. **Patel, N.R., Swan, K., Li, X., Tachado, S.D. & Koziel, H**. 2009. Impaired *M. tuberculosis*-mediated apoptosis in alveolar macrophages from HIV+ persons: potential role of IL-10 and BCL-3. J. Leukoc. Biol. 86:53-60.

15. **Das, S., Narayanan, S., Hari, L., Mohan, N.S., Somasundaram, S., Selvakumar, N. et al.** 2004. Simultaneous infection with multiple strains of *Mycobacterium tuberculosis* identified by restriction fragment length polymorphism analysis. Int. J. Tuberc. Lung Dis. 8:267-270.

16. **Shamputa, I.C., Jugheli, L., Sadradze, N., Willery, E., Portaels, F., Supply, P. et al.** 2006. Mixed infection and clonal representativeness of a single sputum sample in tuberculosis patients from a penitentiary hospital in Georgia. Respir. Res. 7:99.

17. **Fang, R., Li, X., Li, J., Wu, J., Shen, X., Gui, X. et al.** 2008. Mixed infections of *Mycobacterium tuberculosis* in tuberculosis patients in Shanghai, China. Tuberculosis (Edinb) 88:469-473.

18. **Hanekom, M., van der Spuy, G.D., Gey van Pittius, N.C., McEvoy, C.R.E., Ndabambi, S.L., Victor, T.C. et al**. 2007. Evidence that the spread of *Mycobacterium tuberculosis* strains with the Beijing genotype is human population dependent. J. Clin. Microbiol. 45:2263-2266.

19. **Rishi, S., Sinha, P., Malhotra, B. & Pal, N.** 2007. A comparative study for the detection of Mycobacteria by BACTEC MGIT 960, Lowenstein Jensen media and direct AFB smear examination. Indian J Med Microbiol 25:383-386.

20. **van der Zanden, A.G.M., Rahim, Z., Fedder, G., Vos-van Adrichem, J., Sebens, F.W., Heilmann, F.G.C. et al**. 2007. Multiple *Mycobacterium tuberculosis* infections in an HIV-infected patient. Southeast Asian J. Trop. Med. Public Health 38:704-705.

21. **Brosch, R., Gordon, S.V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K. et al.** 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proc. Natl. Acad. Sci. U.S.A. 99:3684-3689.

Figure 1: Order in which primer sets were used to identify strains of the Beijing, LAM, S-family, LCC and Haarlem genotype respectively

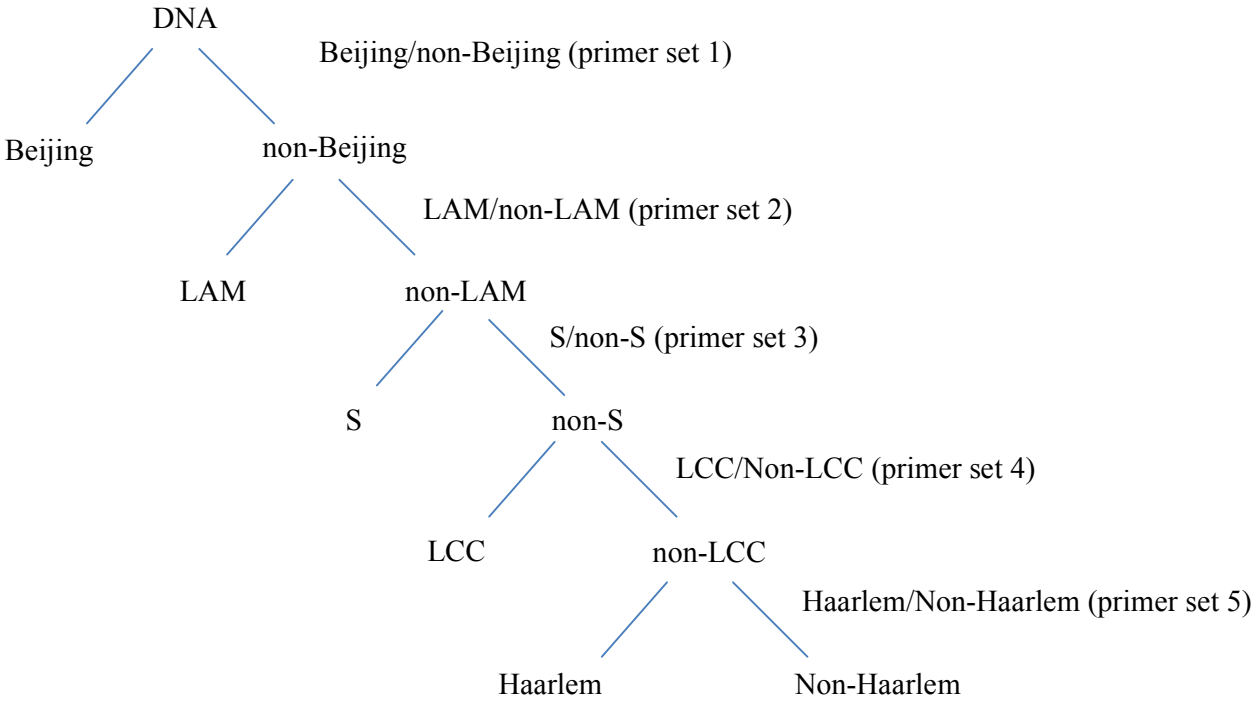


Figure 2: Flow diagram illustrating PCR results of primary cultures grown on MGIT and LJ media

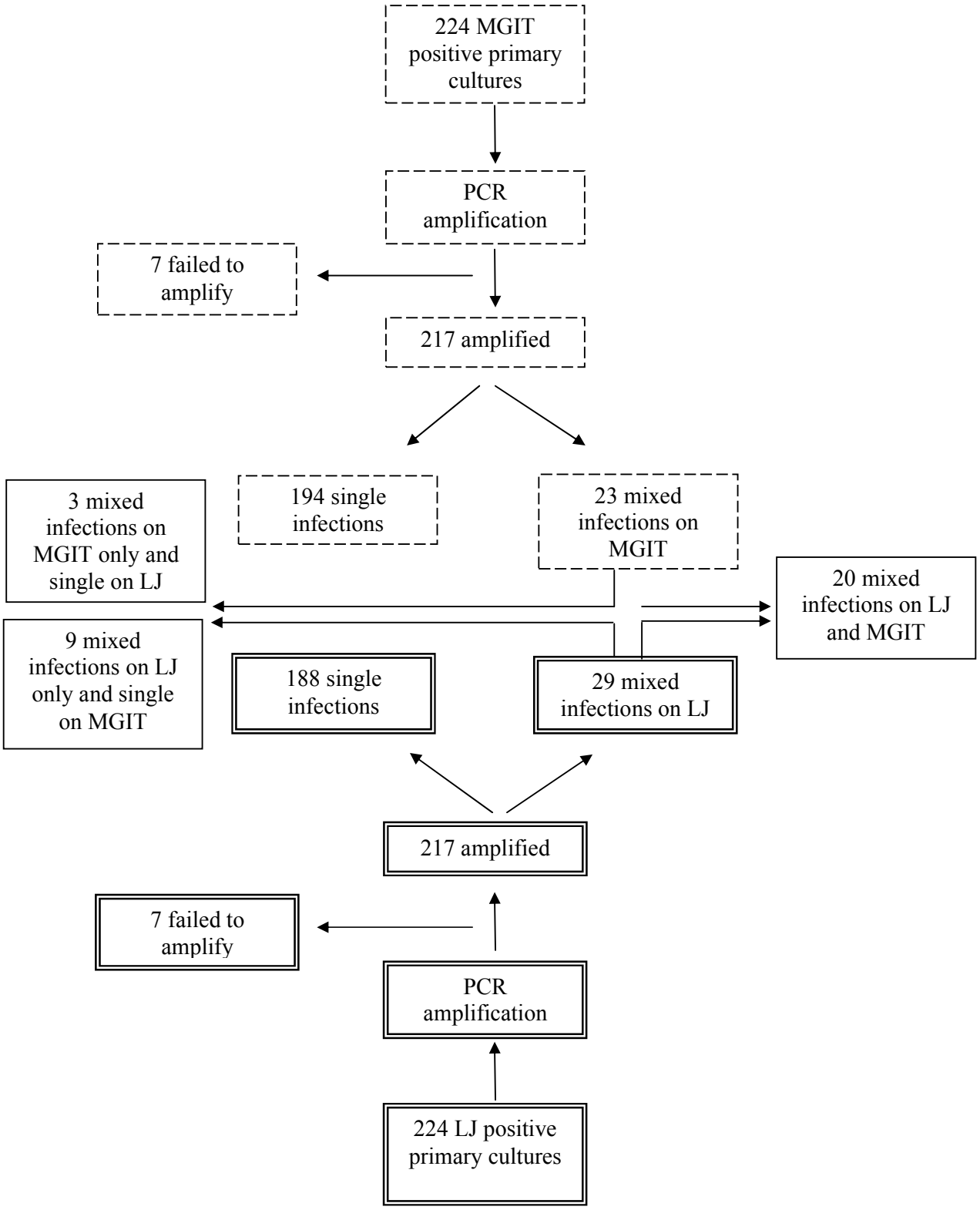


Table 1: Primer sequences used to identify strains of the Beijing, LAM, S-family, LCC and Haarlem genotype respectively

Primer	Primer set	Primer sequence (5' – 3')	Primer position	Chromosomal position[a]	PCR product size (base pairs)
RD105 F		ACA GCG CGG GTC ATA TCA C	79385		
RD105 R	1	AAC CAG CTC CTC GAC GCT ATC	83257	79567 to 83034[c]	Beijing = 405 Non-Beijing = 615
RD105 INT		GCA ACA CCC GCT TGT CTT TG	79999		
Lam F[d]	2	TAG CCC ACC ACC ACA GCT TC	932281	932204	LAM = 205 Non-LAM = 141
Lam R[d]		ACC ACC CTG CCT AAC CAA TTC	932141		
F28/480.9 F	3	GGC GGT GTT AGC GAT TGA A	1889015	1889068	S = 194 Non-S = 236
F28/480.9 R		CTG CGG CAA CAG ATT CCA CTA	1889250		
Rv0403cF	4	GAC AAC GCA TGG ATC GTC C	483170	483298	LCC = 270 Non-LCC = 388
Rv0403cR		TCA CAT CAA CAT GCG CCC	483557		
Rv2336 F	5	GGT GGC GAA AGC TTT AGC C	2610998	2610861	Haarlem = 279 Non-Haarlem = 212
Rv2336 R		TGC GCC AAA CAT GCA GTC	2610787		
Internal primer[b]		TTC AAC CAT CGC CGC CTC TAC			

[a] relative to the H37Rv genome sequence 21

[b] internal primer used for primer set 2 to 4

[c] RD105 region according to (13).

[d] according to (12)

Table 2: Strain genotypes cultured on MGIT and LJ media classified according to spoligotyping and PCR amplification

	MGIT		LJ		MGIT-LJ combinations
Strain genotypes	Spoligotype n = 188 (%)	PCR- genotype n = 217(%)	Spoligotype n = 181 (%)	PCR- genotype n = 217 (%)	PCR- genotype n = 213 (%)
Single infection	n = 182 (96.9)	n = 194 (89.4)	n = 179 (98.9)	n = 188 (86.6)	n = 181[a] (85)
Beijing	63 (33.5)	67 (30.9)	52 (28.7)	56 (25.8)	56 (26.3)
LAM	70 (37.2)	73 (33.6)	78 (43.1)	77 (35.5)	73 (34.3)
S-family	11 (5.9)	12 (5.5)	11 (6.1)	12 (5.5)	12 (5.6)
LCC	16 (8.5)	21 (9.7)	18 (9.9)	21 (9.7)	20 (9.4)
Haarlem	6 (3.2)	3 (1.4)	5 (2.8)	4 (1.8)	3 (1.4)
Other strains	16 (8.5)	18[b] (8.3)	15 (8.3)	18[b] (8.3)	18 (8.5)
Family 14	8		7		
CAS	7		7		
EAI	1		1		
Mixed infection[c]	n = 6 (3.2)	n = 23 (10.6)	n = 2 (1.1)	n = 29 (13.4)	n = 32 (15.0)

Abbreviations: CAS – Central African strain, EAI – East-African-Indian

[a] Total number of single infections add up to 182 because one culture showed the presence of a Beijing strain on MGIT media and the presence of a LAM strain on LJ media

[b] strains other than the strains tested with the 5 primer sets

[c] more than 1 strain genotype was present in the primary culture

Table 3: The population structure of mixed infections cultured on MGIT media and their corresponding LJ media as well as the population structure of the total mixed infections.

MGIT culture	LJ culture	MGIT-LJ combinations
LAM + "other strains"	LAM + "other strains"	LAM + "other strains"
Beijing + LAM	Beijing + LAM	Beijing + LAM
Beijing + LAM + LCC	Beijing + LAM	Beijing + LAM
Beijing + LAM	Beijing + LAM	Beijing + LAM
Beijing	Beijing + LAM	Beijing + LAM
S-family + "other strains"	S-family + "other strains	S-family + "other strains"
Beijing + Haarlem	Beijing + Haarlem	Beijing + Haarlem
Beijing + LAM	Beijing + LAM	Beijing + LAM
Beijing	LAM + "other strains"	LAM + "other strains"
Beijing + LCC	Beijing + LCC	Beijing + LCC
Beijing + LAM	Beijing + LAM	Beijing + LAM
Beijing + Haarlem	Beijing + Haarlem	Beijing + Haarlem
Beijing	Beijing + LAM	Beijing + LAM
Haarlem + "other strains"	Haarlem	Haarlem + "other strains"
LAM	LAM + LCC	LAM + LCC
Beijing	Beijing + LAM + S-family + LCC	Beijing + LAM + S-family + LCC
Beijing	Beijing + LAM	Beijing + LAM
Beijing	Beijing + other	Beijing + other
Beijing + LAM	Beijing + LAM	Beijing + LAM
Beijing + "other strains"	Beijing + LAM + S-family + LCC	Beijing + "other strains"
LCC + "other strains"	LCC + "other strains"	LCC + "other strains"
LAM + Haarlem	LAM + Haarlem	LAM + Haarlem
LCC + "other strains"	LCC + "other strains"	LCC + "other strains"
LAM + "other strains"	LAM	LAM + "other strains"
LAM + "other strains"	LAM	LAM + "other strains"
Beijing + LCC	Beijing + LCC	Beijing + LCC
Beijing + S-family	Beijing + S-family	Beijing + S-family
Beijing	Beijing + LCC	Beijing + LCC
Beijing + LAM	Beijing + LAM	Beijing + LAM
Beijing	Beijing + LAM + LCC	Beijing + LAM + LCC

Beijing + Haarlem LAM + LCC	Beijing + Haarlem LAM + LCC	Beijing + Haarlem LAM + LCC

Table 4: Discrepant PCR amplification results of strain genotypes grown on MGIT media and their corresponding LJ media.

MGIT	LJ
Haarlem + "other strains"	Haarlem
LAM + "other strains"	LAM
LAM + "other strains"	LAM
Beijing	Beijing + LAM
Beijing	LAM + "other strains"
Beijing	Beijing + LAM
LAM	LAM + LCC
Beijing	Beijing + LAM + S-family + LCC
Beijing	Beijing + LAM
Beijing	Beijing + "other strains"
Beijing	Beijing + LCC
Beijing	Beijing + LAM + LCC

Table 5: Demographic and drug susceptibility data of patients infected with one or more *M. tuberculosis* strain genotypes.

Demographic and clinical data	All	Mixed infection (%) (n=32)	Single strain infection (%) (n=181)	*P*-values
Mean age, yr	199	33.1 yr	35 yr	
Sex				
Male	113	16 (51.6)	97 (57.7)	$p = 0.53$
Female	86	15 (48.4)	71 (42.3)	
Treatment history				
New	130	21 (67.7)	109 (67.7)	$p = 0.99$
Retreatment	62	10 (32.3)	52 (32.3)	
Smear grading				
Negative	77	13 (41.9)	64 (38.10)	$p = 0.69$
Positive	122	18 (58.1)	104 (61.9)	
Drug susceptibility				
Sensitive	114	17 (70.8)	97 (70.3)	$p = 0.88$
All resistance	50[a]	7 (29.2)	43 (30.7)	
Mono-resistance	30	6 (25)	24 (17.4)	
Poly-Resistance	12	0	12 (8.7)	$p = 0.22$
MDR	6	1(4.2)	5 (3.6)	

[a] Two extensively drug-resistant (XDR) isolates were excluded from the table.

Beijing and Haarlem genotypes are overrepresented among children with drug-resistant tuberculosis in the Western Cape Province of South Africa

Marais, B. J., Victor, T. C., Hesseling, A. C., Barnard, M., Jordaan, A., Brittle, W., Reuter, H., Beyers, N., van Helden, P. D., Warren, R. M. and Schaaf, H. S.

J. Clin. Microbiol. (2006). 44:3539-3543

Abstract

Drug resistance among children with culture-confirmed tuberculosis (TB) provides an accurate measure of transmitted drug resistance within the community. We describe the genotype diversity in children with culture-confirmed TB and investigate the relationship between genotype and drug resistance. A prospective study was conducted from March 2003 through August 2005 at Tygerberg Children's Hospital, in the Western Cape Province of South Africa. All children (<13 years of age) diagnosed with culture-confirmed TB were included. Genotype analysis and phenotypic drug susceptibility testing were performed on the first culture-positive isolate from each patient. Mutation analysis was performed on all drug-resistant isolates. Spoligotyping was successfully performed on isolates from 391/399 (98%) children diagnosed with culture-confirmed TB. Drug susceptibility testing was also performed on 391 isolates; 49 (12.5%) were resistant to isoniazid, and 20 (5.1%) of these were resistant to both isoniazid and rifampin. Beijing was the most common genotype family, identified in 130/391 (33.2%) cases, followed by LAM in 114/391 (29.2%) cases. The presence of both Beijing and Haarlem genotype families was significantly associated with drug resistance (26/49 [53.1%] versus 113/342 [33.0%]; odds ratio, 1.7; 95% confidence interval, 1.0 to 2.9). The high prevalence of Beijing and LAM in children with culture-confirmed TB reflects considerable transmission of these genotype families within the community. The overrepresentation of Beijing and Haarlem genotype families in children with drug-resistant TB demonstrates their contribution to transmitted drug resistance and their potential importance in the emergent drug-resistant TB epidemic.

Introduction

At a molecular level, the global tuberculosis (TB) epidemic consists of multiple genotype-specific subepidemics. Over time, the host-pathogen interaction, as well as the environment, selects the most "successful" genotypes within particular geographic locations (10). Different *Mycobacterium tuberculosis* (*M. tuberculosis*) genotypes can be identified by variation in certain well-characterized repeat sequences, such as the IS*6110* transposable element and the direct repeat region (33).

Host-related selective forces include variability in the immune response to *M. tuberculosis* infection and the provision of anti-TB chemotherapy to TB patients. The immune response of the host is influenced by genetic predisposition and acquired factors (6, 10), including bacillus Calmette-Guerin (BCG) vaccination, previous *M. tuberculosis* infection, exposure to environmental mycobacteria, and human immunodeficiency virus (HIV) infection. Genotype-specific differences in the host-pathogen interaction are well documented in animal models, where infection with different *M. tuberculosis* genotypes evokes markedly different immune responses (14). In addition, BCG vaccination provides variable genotype-specific protection; infection with strains from the Beijing genotype family in particular has been associated with decreased protection following BCG vaccination in mice (14).

The Beijing genotype family is well recognized as a distinct genetic lineage, and it is genetically highly conserved (11), although large sequence polymorphisms identify four monophyletic subgroups (30). It has a worldwide distribution but predominates in certain geographic areas, particularly in parts of Asia (13, 31) and Russia (7). Based on its association with young age in adult studies from Vietnam and Russia (7), suggesting more recent spread, it has been proposed that Beijing should be regarded as an emerging genotype family (2).

The association between drug resistance and the Beijing genotype is well documented in adults (1, 4, 13, 23, 29, 30). The geographic variability observed in this association (30), the frequent clustering of resistant genotypes, and their successful spread within the Russian prison system (7, 29), suggest recent clonal expansion. This is supported by evidence that some strains of the Beijing genotype family retain fitness despite the acquisition of drug resistance (28). The Haarlem genotype family has also been

associated with drug resistance and rapid clonal expansion (8, 19). The association of these genotype families with drug-resistant outbreaks clearly demonstrates their epidemic potential (19, 21). From a TB control point of view, it is relevant to know whether specific genotype families are overrepresented among drug-resistant cases and in particular if these resistant strains are successfully transmitted within the community.

Children provide a valuable epidemiologic perspective, as the burden of childhood TB reflects the level of epidemiologic control achieved within a particular community (18). Because children usually progress to disease within 12 months of primary infection (15, 16), genotype analysis of isolates collected from children reflects current transmission patterns within the community (35). Children tend to develop pauci-bacillary disease and are therefore unlikely to acquire drug resistance (15, 26). Consequently, drug resistance patterns among children with culture-confirmed TB serve as a good indicator of transmitted (primary) drug resistance within the community (26).

Genotype diversity, and in particular the relation between genotype and drug resistance, are poorly documented in children. The aims of this study were to describe the genotype diversity in children with culture-confirmed TB and to investigate the relationship between genotype and drug resistance.

Materials and methods

Study population and setting

Isolates were prospectively collected from March 2003 through August 2005 at Tygerberg Children's Hospital, a pediatric referral hospital in the Western Cape Province of South Africa. All children (<13 years of age) diagnosed with culture-confirmed tuberculosis were included. The Western Cape Province reported a TB incidence of 678/100,000 in 2003 (3), and the incidence of childhood TB in areas adjacent to Tygerberg Children's Hospital was 407/100,000 in the same period, representing 13.7% of all treated cases (17).

Laboratory procedures

Diagnostic cultures were performed using Middlebrook 7H9 broth-based Mycobacteria Growth Indicator Tubes (MGIT; Becton Dickinson, Sparks, MD). *M. tuberculosis* was confirmed by routine PCR-based speciation of the *M. tuberculosis* complex (5). The first culture-positive isolate from each patient was sent to the regional diagnostic laboratory for phenotypic drug susceptibility testing, using the indirect proportional method. Drug resistance was defined according to international criteria (≥1% bacterial growth on drug-containing media compared to a non-drug-containing control) using critical concentrations of 0.2 µg/ml isoniazid (INH) or 30 µg/ml rifampin (RMP) (26). In-house quality assurance checks were performed after every batch, with external review by the national reference TB laboratory on a quarterly basis.

Genotypic testing for specific mutations in the *katG*, *inhA*, and *rpoB* genes that may confer resistance to INH and/or RMP was performed on all isolates with phenotypic drug resistance. DNA sequencing included the following regions: position 15 in the *inhA* promoter region, codon 315 of the *katG* gene, and codons 526 and 531 of the *rpoB* gene. An equal number of drug-sensitive specimens were tested as controls.

Mycobacterial subcultures made on standard Löwenstein-Jensen slants were used for genotyping. Strict protocols were in place to prevent cross-contamination; subcultures were prepared within a negative-pressure extraction hood, using sterile equipment and containers, by direct aspiration from the liquid growth media. Genotype determination was done using standardized spoligotyping methodology (33). Isolates were assigned to specific genotype families according to their spoligotype signature, using the spoldb3 database (www.pasteur-guadeloupe.fr/tb/spoldb3). The following internationally recognized genotype families were identified: Beijing, LAM (Latin American and Mediterranean family), X (a European clade of IS*6110* low banders), and Haarlem (9), while Family 28 was identified as a "stand alone" genotype family (E. Streicher, personal communication). Genotype families with a low frequency (<10 cases) were combined and categorized as "Other."

Some isolates could not be assigned to an internationally recognized genotype family, and these specimens were further characterized by using a novel PCR-based technique. PCR primers were designed to identify the presence of an IS*6110* insertion

(position 932204 according to the whole H37RV genome sequence), which is unique to all members of the LAM family (25). DNA was PCR amplified in a total volume of 25 µl, containing 1 µl DNA template, 1x Q-Buffer, 1x enzyme buffer, 3.5 mM $MgCl_2$, 1.6 mM deoxynucleoside triphosphates, 25 pmol of each primer (LAM F, TAG CCC ACC ACC ACA GCT TC; LAM R, ACC ACC CTG CCT AAC CAA TTC; Xho1, TTC AAC CAT CGC CGC CTC TAC), and 0.25 U HotStarTaq DNA polymerase (QIAGEN, Germany). Amplification was initiated by incubation at 95°C for 15 min, followed by 40 cycles at 94°C for 1 min, 62°C for 1 min, and 72°C for 1 min. After the last cycle, isolates were incubated at 72°C for 10 min. PCR amplification products were electrophoretically fractionated in 3.0% agarose in 1x Tris-borate-EDTA, pH 8.3, at 3.5 V/cm for 4 h and visualized by staining with ethidium bromide. The presence of a LAM strain was represented by a band of 205 base pairs, while a non-LAM strain was represented by a band of 141 base pairs. The specificity of amplification was determined by amplification of a panel of genetically unrelated and related strains and was shown to be 100% (95% confidence interval [CI], 85 to 100%) compared to the gold standard of IS*6110* DNA fingerprinting. To minimize the risk of laboratory cross-contamination during PCR amplification, each procedure (preparation of the PCR mixes, the addition of the DNA, the PCR amplification, and the electrophoretic fractionation) was conducted in physically separated rooms. Negative controls (water) were included to control for reagent contamination. After performing the LAM-specific PCR test, some isolates could still not be assigned to an internationally recognized genotype family, and these isolates were categorized as "unclassified."

Children were tested for human immunodeficiency virus (HIV) infection, using the Determine HIV1/2 rapid test (Abbot, Tokyo, Japan) as a screening test and an enzyme-linked immunosorbent assay (DNA PCR in children <18 months of age) for confirmation. The study was approved by the Ethics Committee, Faculty of Health Sciences, Stellenbosch University.

Data analysis

Descriptive statistical analysis was done with SPSS, version 13 (SPSS Inc., Chicago, IL) on anonymous unlinked data. Categorical data were analyzed with the χ^2 test,

using Epi-Info version 6.04. We report the relative proportion of different genotype families, comparing drug-susceptible to drug-resistant isolates, and describe the specific mutations associated with resistance.

Results

During the study period, 399 children were diagnosed with culture-confirmed TB. Patient characteristics and drug susceptibility test (DST) results are reflected in Table 1. Spoligotyping and DST analysis were performed on 391 isolates; 8 (2%) isolates developed bacterial contamination or lost viability during drug susceptibility testing. Of the isolates tested, 49 (12.5%) were resistant to INH, of which 20 (5.1%) were resistant to both INH and RMP (multidrug resistant [MDR]). No case of RMP monoresistance was recorded.

Table 1: Patient characteristics and drug susceptibility test results of children diagnosed with culture-confirmed tuberculosis

Patient group and characteristic	Value[a]
All patients with culture-confirmed TB (n = 399)	
Boys (no. of patients)	210 (52.6)
Age (yr)	
Median	2
Range	0-12
HIV testing (no. of patients)	
HIV test performed	297 (74.4)
HIV infected	73 (18.3)
DST result available (no. of patients)	391 (98.0)
All patients with a DST result (*n* = 391)	
Drug susceptible	342 (87.5)
All INH resistance	49 (12.5)
INH monoresistance	29 (7.4)
MDR	20 (5.1)

[a] Values in parentheses are percentages

Table 2 reflects the relative proportion of different genotype families, comparing drug-susceptible to drug-resistant isolates. Beijing was the most prevalent genotype family, identified in 130/391 (33.2%) isolates, followed by LAM, identified in

114/391 (29.2%) isolates. Children infected with either the Beijing or Haarlem strain families were more likely to have drug-resistant disease compared to those infected with other strain families (26/139 [18.7%] versus 23/252 [9.1%]; odds ratio [OR], 2.29; 95% confidence interval [CI], 1.2 to 4.4). The Beijing and Haarlem families combined were also more prevalent in the drug-resistant compared to the drug-susceptible group (26/49 [53.1%] versus 113/342 [33.0%]; OR, 1.7; 95% CI, 1.0 to 2.9).

Table 2:. Mycobacterial genotype and drug resistance in children with culture-confirmed tuberculosis

Genotype family designation[a]	No. (%) of isolates	No. (%) of isolates with DST result available		Odds ratio (95% CI)
		Drug susceptible	Drug resistant	
Beijing	130 (32.6)	105 (30.7)	22 (44.9)	1.5 (0.8-2.6)
LAM	114 (28.6)	99 (28.9)	13 (26.5)	0.9 (0.4-1.8)
X	25 (6.3)	22 (6.4)	3 (6.1)	1.0 (0.2-3.5)
Family 28	25 (6.3)	24 (7.1)	1 (2.0)	0.3 (0.0-1.8)
Haarlem	12 (3.0)	8 (2.3)	4 (8.2)	3.7 (0.8-14.5)
Other	7 (2.0)	7 (2.0)	0	
Unclassified	78 (19.5)	71 (20.8)	5 (10.3)	0.4 (0.1-1.1)
Not done	8 (2.0)	6 (1.8)	1 (2.0)	1.2 (0.0-9.9)
Total	399 (100.0)	342 (100)	49 (100)	

[a] Other, all genotype families with a frequency of less than 10 cases; X, European clade of IS*6110* low banders, identified by spoligotyping (9); unclassified, no internationally recognized genotype family assigned, based on the spoldb3 spoligotype database (9), and negative for the LAM-specific PCR test.

Table 3 reflects specific mutations in the *inhA* promoter region and in the *katG* gene. Resistance mutations were detected in 31/49 (64.6%) isolates with phenotypic INH resistance. The two most common mutations detected were –15 in the *inhA* promoter region (18/49; 36.7%) and 315ACC in the *katG* gene (16/49; 32.7%). Among Beijing isolates, the –15 *inhA* promoter region mutation occurred most frequently (13/22; 59.1%) and was exclusively detected in the Beijing 220 subgroup. The Beijing 220 subgroup is a well-recognized drug-resistant strain in the Western Cape Province (T. C. Victor, personal communication). Mutations found in the MDR isolates included –15 in the *inhA* promoter region (7/20; 35%) as well as 315ACC (10/20; 50%) and 315ACA in the *katG* gene (1/20; 5%). No INH resistance-associated mutation was detected in 2/20 (10%) MDR isolates, using the limited mutation analysis described, and no multiple mutations were detected in the same isolate. No resistance-associated mutations, either for INH or RMP, were identified in any of the drug-sensitive control isolates.

Table 3:. Specific resistance mutations in children with INH-resistant tuberculosis

Genotype family designation[a]	No. of strains with resistance mutation:			No. of wild-type strains	Total no. of strains
	inhA –15	*katG* 315ACA	*katG* 315ACC		
Beijing	13	0	5	4	22
LAM	3	0	4	6	13
X	0	1	1	1	3
Family 28	1	0	0	0	1
Haarlem	0	0	3	1	4
Other	0	0	0	0	0
Unclassified	0	0	3	2	5
Not done	1	0	0	0	1
Total	18	1	16	14	49

[a] Other, all genotype families with a frequency of less than 10 cases; X, European clade of IS*6110* low banders, identified by spoligotyping (9); unclassified, no internationally recognized genotype family assigned, based on the spoldb3 spoligotype database (9), and negative for the LAM-specific PCR test.

Table 4 reflects specific mutations in the *rpoB* gene in children with RMP-resistant TB. Resistance mutations were detected in all isolates with phenotypic RMP resistance. The most common *rpoB* gene mutation was 531TTG (13/20; 65%). This mutation was identified in all (8/8) multidrug-resistant Beijing isolates.

Table 4: Specific *rpoB* mutations in children with MDR[a] tuberculosis

Genotype family designation[b]	No. of strains with resistance mutation:			No. of wild-type strains	Total no. of strains
	526	531 CAG	531 TTG		
Beijing	0	0	8	0	8
LAM	1	1	3	0	5
X	0	0	1	0	1
Family 28	1	0	0	0	1
Haarlem	0	0	2	0	2
Other	0	0	0	0	0
Unclassified	2	0	1	0	3
Total	4	1	15	0	20

[a] MDR, multidrug resistant (resistant to both INH and rifampin); all rifampin-resistant isolates were resistant to isoniazid as well.

[b] Other, all genotype families with a frequency of less than 10 cases; X, European clade of IS*6110* low banders, identified by spoligotyping (9); unclassified, no internationally recognized genotype family assigned, based on the spolDB3 spoligotype database (9), and negative for the LAM-specific PCR test.

Discussion

To our knowledge this is the first report of genotype diversity in relation to phenotypic and genotypic drug resistance in children with TB, and it represents one of the largest cohorts of children with DST results reported to date. As such, the report provides an estimate of genotype-specific transmission patterns at the community level, both for drug-susceptible and drug-resistant organisms. The association found between Beijing and Haarlem genotype families and drug-resistant TB is consistent with previous reports on adults (4, 7, 8, 19, 23, 29).

The predominance of specific mutations such as the –15 *inhA* promoter region mutation and the 531TTG *rpoB* mutation in the Beijing genotype family is an interesting observation. The frequency with which these specific mutations were detected may reflect random mutation and subsequent clonal expansion. Alternatively, certain genotype families may preferentially develop resistance mutations at particular locations in so-called mutation hotspots. Unique polymorphisms in the genes involved in DNA repair (*mut* genes) have been demonstrated in the Beijing genotype family, which may explain the ability of this clonal family to rapidly adapt to selective pressure (24); however, functional correlates of the "mutator phenotype" hypothesis have not been demonstrated. In a comparative RMP resistance study, Beijing genotypes did not acquire *rpoB* mutations more rapidly than the reference H37Rv strain (34). However, the Beijing genotype family is not a homogeneous group and demonstrates a deeper population structure than previously appreciated (30). Therefore, the reported comparable acquisition of RMP resistance may not necessarily apply to the Beijing genotype family as a whole. In addition, the ability to acquire INH resistance may be the most important step in the acquisition of MDR disease. RMP monoresistance is rare. Due to the exceptional early bactericidal activity of INH (20), the development of monoresistance to INH is probably a more relevant event in the acquisition of multidrug resistance. In this regard, the specific –15 *inhA* promoter region mutation, documented in children infected with the particular Beijing 220 subgroup, is of particular interest and warrants further investigation. The substantial pool of INH monoresistant isolates identified may represent a source of future MDR disease (27).

Similar to a recent report from an adjacent study setting in Cape Town, South Africa (22), this study demonstrates that Beijing is the dominant genotype family in children with culture-confirmed TB. An important finding is that the prevalence of the Beijing genotype in children seems to differ from that described in adults from the same geographic area. LAM (which includes Family 11) has been reported as the dominant genotype family among adults in Cape Town (32). This discrepancy may reflect an important "age-shift," similar to adult data from Vietnam (2), indicating emergence of the Beijing genotype family in the Western Cape Province of South Africa. However, the validity of comparisons with adult data is limited, as adult and child data were not collected from the same geographic area or the same period of time. In Vietnam, the prevalence of the Beijing genotype family was found to be associated with the presence of a BCG scar (2). Although this was not significant after

correcting for the cohort effect (2), it lends preliminary support to observations in animal studies that BCG provides less efficient protection against Beijing (14). This may provide a selective advantage to the Beijing genotype family in countries with universal BCG vaccination policies, such as South Africa.

This study has several limitations. Limiting selection bias presents a huge challenge when studying children with culture-confirmed TB, as culture confirmation is rarely achieved at the primary health care level, and hospital-based studies are inherently limited by selection bias. However, a recent report that compared drug resistance patterns between hospital- and community-recruited children from the same cohort found no significant differences between the two groups (26).

When investigating genotype diversity, an important consideration that is often overlooked is the potential selection bias imposed by the variable culturability of different genotypes. If particular genotypes are easier to culture in vitro, particularly in children with pauci-bacillary disease, this may introduce a significant bias. *In vivo* data demonstrate that variation in strain-specific culturability is a reality that requires further exploration (12). Finally, the spoligotype fingerprint of the Beijing genotype family is short and distinct. Therefore, the probability of losing the spoligotype fingerprint of other strains through mutational events is greater. As a result, most of the unclassified strains would represent non-Beijing genotypes. We used a novel

PCR-based test to reduce the number of unclassified isolates, and during analysis, all unclassified strains were regarded as non-Beijing to limit potential bias.

In conclusion, the high prevalence of Beijing and LAM in children with culture-confirmed TB reflects considerable transmission of these genotype families within the community. The overrepresentation of Beijing and Haarlem genotype families in children with drug-resistant TB demonstrate their contribution to transmitted drug resistance and their potential importance in the emergent drug-resistant TB epidemic.

Acknowledgments

We thank the TB in the 21st Century Consortium, an international network supported by the Norwegian Research Council and the Centre for Prevention of Global Infections, University of Oslo. We acknowledge financial support from the Harry and Doris Crossly Foundation and the South African Medical Research Council.

Reference List

1. **Almeida, D., C. Rodriques, T. F. Ashavid, A. Lalvani, Z. F. Udwadia, and A. Mehta.** 2005. High incidence of the Beijing genotype among multi-drug resistant isolates of *Mycobacterium tuberculosis* in a tertiary care centre in Mumbai, India. Clin. Infect. Dis. **40:**881-886.

2. **Anh, D. D., M. W. Borgdorff, L. N. Van, N. T. N. Lan, T. van Gorkom, K. Kremer, and D. van Soolingen.** 2000. Mycobacterium tuberculosis Beijing genotype emerging in Vietnam. Emerg. Infect. Dis. **6:**302-305.

3. **Cape Town TB Control.** Progress report 1997-2003. [Online.] http://www.hst.org.za/publications/618.

4. **Cox, H. S., T. Kubica, D. Doshetov, Y. Kebede, S. Rusch-Gerdess, and S. Niemann.** 2005. The Beijing genotype and drug resistant tuberculosis in the Aral Sea region of Central Asia. Respir. Res. **6:**134.

5. **De Wit, D., L. Steyn, S. Shoemaker, and M. Sogin.** 1990. Direct determination of *Mycobacterium tuberculosis* in clinical specimens by DNA amplification. Clin. Microbiol. **28:**2437-2441.

6. **Dormans, J., M. Burger, D. Aguilar, R. Hernandez-Pando, K. Kremer, P. Roholl, S. M. Arend, and D. van Soolingen.** 2004. Correlation of virulence, lung pathology, bacterial load and delayed type hypersensitivity responses after injection with different *Mycobacterium tuberculosis* genotypes in a BALB/c mouse model. Clin. Exp. Immunol. **137:**460-468.

7. **Drobniewski, F., Y. Balabanova, V. Nikolayevsky, M. Ruddy, S. Kuznetzov, A. Zakharova, A. Melentyev, and I. Fedorin.** 2005. Drug-resistant tuberculosis, clinical virulence and the dominance of the Beijing strain family in Russia. JAMA **293:**2726-2731.

8. **Farnia, P., M. R. Masjedi, M. Misaedi, F. Mohammadi, A. Jallaledin-Ghanavi, V. Vincent, M. Bahadori, and A. A. Velayati.** 11 February 2006, posting date. Prevalence of Haarlem 1 and Beijing types of *Mycobacterium tuberculosis* strains in Iranian and Afghan MDR-TB patients. J. Infect. [Epub ahead of print]

9. **Filiol, 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, E. Kassa-Kelembho, 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.** 2003. Snapshot of moving and expanding clones of *Mycobacterium tuberculosis* and their global distribution assesses by spoligotyping in an international study. J. Clin. Microbiol. **41:**1963-1970.

10. **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. USA **103:**2869-2873.

11. **Glynn, J., J. R. Whiteley, P. J. Bifani, K. Kremer, and D. van Soolingen.** 2002. Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: a systematic review. Emerg. Infect. Dis. **8:**843-849.

12. **Li, Q., C. C. Whalen, J. Albert, R. Larkin, L. Zukowski, M. D. Cave, and R. Silver.** 2002. Differences in rate and variability in intracellular growth of a panel of *Mycobacterium tuberculosis* clinical isolates within a human monocyte model. Infect. Immun. **70:**6489-6493.

13. **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.** 2000. Molecular epidemiology of *Mycobacterium tuberculosis* in China: a nationwide random survey in 2000. Int. J. Tuberc. Lung Dis. **9:**1314-1319.

14. **Lopez, B., D. Aguilar, H. Orozco, M. Burger, C. Espetia, 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:**30-37.

15. **Marais, B. J., R. P. Gie, H. S. Schaaf, N. Beyers, P. R. Donald, and J. Starke.** 2006. Childhood pulmonary tuberculosis-old wisdom and new challenges. Am. J. Resp. Crit. Care Med. **173:**1078-1090.

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

17. **Marais, B. J., A. C. Hesseling, R. P. Gie, H. S. Schaaf, and N. Beyers.** 2006. The burden of childhood tuberculosis and the accuracy of routine surveillance data in a high-burden setting. Int. J. Tuberc. Lung Dis. **10:**259-263.

18. **Marais, B. J., C. C. Obihara, R. W. Warren, H. S. Schaaf, R. P. Gie, and P. R. Donald.** 2005. The burden of childhood tuberculosis: a public health perspective. Int. J. Tuberc. Lung Dis. **9:**1305-1313.

19. **Mardassis, H., A. Namouchi, R. Haltiti, M. Zarrouk, B. Mhenni, A. Karboul, N. Khabouchi, N. C. Gey van Pittius, E. M. Streicher, J. Rauzier, B. Gicquel, and K. Dellagi.** 2005. 2005. Tuberculosis due to resistant Haarlem strain, Tunisia. Emerg. Infect. Dis. **11:**957-961.

20. **Mitchison, D. A.** 2000. Role of individual drugs in the chemotherapy of tuberculosis. Int. J. Tuberc. Lung Dis. **4:**796-806.

21. **Moss, A. R., D. Alland, E. Telzak, D. Hewlett, 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:**115-121.

22. **Nicol, M., C. Sola, B. February, N. Rastogi, L. Steyn, and R. J. Wilkinson.** 2005. Distribution of strain families of *Mycobacterium tuberculosis* causing pulmonary and extra-pulmonary disease in hospitalized children in Cape Town, South Africa. J. Clin. Microbiol. **43:**5779-5781.

23. **Park, Y. K., S. Shin, S. Ryu, S. N. Cho, W. Koh, O. J. Kwon, Y. S. Shim, W. J. Lew, and G. H. Bai.** 2005. Comparison of drug resistance genotypes between Beijing and non-Beijing family strains of *Mycobacterium tuberculosis* in Korea. J. Microbiol. Methods **63:**165-172.

24. **Rad, M. E., P. Bifani, C. Martin, K. Kremer, S. Samper, J. Rauzier, B. Kreiswirth, J. Blazquez, M. Jouan, D. van Soolingen, and B. Gicquel.** 2003.

Mutations in putative mutator genes of Mycobacterium tuberculosis strains of the W-Beijing family. Emerg. Infect. Dis. **9:**838-845.

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

26. **Schaaf, H. S., B. J. Marais, A. C. Hesseling, R. P. Gie, N. Beyers, and P. R. Donald.** 2006. Childhood drug-resistant tuberculosis in the Western Cape Province of South Africa. Acta Paediatrica **15:**523-528.

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

28. **Toungoussova, O. S., D. A. Caugant, P. Sandven, A. O. Mariandyshev, and G. Bjune.** 2004. Impact of drug resistance on fitness of *Mycobacterium tuberculosis* strains of the W-Beijing genotype. FEMS Immunol. Med. Microbiol. **42:**281-290.

29. **Toungoussova, O. S., A. Mariandyshev, G. Bjune, P. Sandven, and D. A. Caugant.** 2003. Molecular epidemiology and drug resistance of *Mycobacterium tuberculosis* isolates in the Archangel prison in Russia: predominance of the W-Beijing clone family. Clin. Infect. Dis. **37:**665-672.

30. **Tsolaki, A. G., S. Gagneux, A. S. Pym, Y. O. Gogeut de la Salmoniere, B. N. Kreiswirth, D. van Soolingen, and P. M. Small.** 2005. Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **43:**3185-3191.

31. **van Soolingen, D., L. Qian, P. E. W. de Haas, J. T. Douglas, H. Traore, F. Portaels, H. Z. Qing, D. Enkhsaikan, P. Nymadawa, and J. D. A. van Embden.** 1995. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. J. Clin. Microbiol. **33:**3234-3238.

32. **Victor, T. C., P. E. de Haas, A. M. Jordaan, G. D. van der Spuy, M. Richardson, D. van Soolingen, P. van Helden, and R. Warren.** 2004. 2004. Molecular characteristics and global spread of *Mycobacterium tuberculosis* with a Western Cape F11 genotype. J. Clin. Microbiol. **42:**769-772.

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

34. **Werngren, J., and S. E. Hoffner.** 2003. Drug-susceptible *Mycobacterium tuberculosis* Beijing genotype does not develop mutation conferred resistance to rifampin at an elevated rate. J. Clin. Microbiol. **41:**1520-1524.

35. **Wootton, S. H., B. E. Gonzalez, R. Pawlak, L. D. Teeter, K. C. Smith, J. M. Musser, J. R. Starke, and E. A. Graviss.** 2005. Epidemiology of pediatric tuberculosis using traditional and molecular techniques; Houston, Texas. Pediatrics **116:**1141-1147.

Chapter 6

Increased bacterial load is associated with strains of the *Mycobacterium tuberculosis* Beijing genotype

Hanekom, M., Claassens, M., Lawrence, K., van der Spuy G. D., Beyers, N, Gie, R., Kidd, M., van Helden, P. D. and Warren, R. M.

To be submitted
(Formatted as per submission)

Increased bacterial load is associated with strains of the *Mycobacterium tuberculosis* Beijing genotype

Hanekom M[1], Claassens M[2], Lawrence K[2], van der Spuy G[1], Beyers N[2], Gie R[2], Kidd M[3], van Helden, PD[1], Warren, RM[1*]

[1]DST/NRF Centre of Excellence for Biomedical Tuberculosis Research / MRC Centre for Molecular and Cellular Biology, Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, [2]Desmond Tutu Tuberculosis Centre, Department of Paediatrics and Child Health, Faculty of Health Sciences, Stellenbosch University, South Africa. [3]Centre for Statistical Consultation, Stellenbosch University, South Africa.

*Corresponding Author

DST/NRF Centre of Excellence for Biomedical Tuberculosis Research / MRC Centre for Molecular and Cellular Biology, Division of Molecular Biology and Human Genetics, Faculty of Health Sciences, Stellenbosch University, PO Box 19063, Tygerberg, South Africa, 7505
Telephone: 021 938 9073
Fax: 021 938 9476
Email: rw1@sun.ac.za

Running title: Beijing genotype
Keywords: *Mycobacterium tuberculosis*, Beijing, smear positive, clinical presentation

Abstract

Word count = 199

Background: Molecular epidemiological and murine infection model studies have suggested that distinct *Mycobacterium tuberculosis* genotypes evolved unique pathogenic characteristics.

Aim: Clinical and demographic parameters of new tuberculosis (TB) patients infected with Beijing and non-Beijing genotype strains were compared to identify genotype-specific pathogenic characteristics.

Methods: Clinical and demographic data were collected from new pulmonary tuberculosis patients that were prospectively recruited at 5 health care clinics in Cape Town, South Africa over a period of 24 months as part of a longitudinal cohort study. Chest radiographs were read, scored and recorded in a database. *M. tuberculosis* strains were genotyped according to IS*6110* DNA fingerprinting and spoligotyping.

Results: Patients infected with a Beijing genotype strain were more likely to have a higher bacillary load as measured by their sputum smear gradation (p = 0.008). Review of the chest radiographs showed a weak association between pleural TB (pleural fluid and pleural fibrosis) and the non-Beijing genotype (p = 0.03). All of the other clinical and demographic parameters analyzed were not associated with a defined *M. tuberculosis* genotype.

Conclusion: This study showed that Beijing genotype strains were associated with an increased bacterial load in sputum samples thereby enhancing the transmissibility of this genotype.

Introduction

Molecular epidemiological studies have shown that the tuberculosis (TB) epidemic is composed of subsets of smaller sub-epidemics which consist of various genetically related and unrelated *Mycobacterium tuberculosis* genotypes. Subsequent studies

have found that the Beijing genotype occurs worldwide and has been associated with numerous TB outbreaks (1, 2). In addition, the Beijing strain genotype has been associated with drug resistance (3), the ability to evade the protective effect of the BCG vaccine (4), treatment failure and relapse (5, 6) and to cause more severe forms of disease (7). Based on these findings, many studies attempted to determine whether a correlation exists between the *M. tuberculosis* genotype and clinical presentation. In the murine infection model Beijing genotype strains were more frequently associated with increased growth rate, bacterial load and time to death and severe lung pathology of infected mice (8, 9). These findings were attributed to the ability of the Beijing strain genotype to induce a non-protective Th2-type immune response characterized by increased levels of IL-4 and IgE and decreased levels of IL-2 and IFN-γ (8). In humans, the interaction between *M. tuberculosis* genotypes and the immune response is not well established although certain clinical and demographic parameters (extrapulmonary disease, febrile response, human immunodificiency virus (HIV) status and young age) have been associated with the Beijing genotype (7, 10, 11, 12). Interestingly, the association between the Beijing genotype and transmission has not been linked to the clinical parameter of sputum smear gradation (7, 10, 13, 14). Recent studies from The Gambia and South Africa suggest that the Beijing and non-Beijing genotype strains are equally infectious (15, 16), although, infection with Beijing genotype strains was more likely to progress towards disease, possibly as a result of immune modulation (15, 17).

A limited number of studies have attempted to use chest radiography to demonstrate an association between lung pathology and the genotype of the *M. tuberculosis* strain causing disease (10, 13, 18, 19). However these finding have not been consistent. Studies done in Indonesia (10), the Netherlands (18) and Taiwan (13) found no significant differences between chest radiograph findings and these two groups of patients. In contrast, a subsequent study done in Russia (19) showed that TB caused by a Beijing genotype strain was more commonly associated with advanced radiological abnormalities (multiple lung zones with fibrotic changes and widespread cavitations) than strains with a non-Beijing genotype.

In this study we compared clinical and demographic data of newly diagnosed TB patients infected with Beijing and non-Beijing genotype strains in order to determine whether a correlation existed between clinical presentation and the genotype of the *M. tuberculosis* strain causing disease.

Methods

Sputum specimens were collected from pulmonary TB patients that were prospectively recruited at 5 health care clinics in Cape Town, South Africa over a period of 24 months as part of a longitudinal cohort study. Patients were excluded for the following reasons: multidrug-resistant (MDR) TB, positive HIV status, co-infection with nontuberculous mycobacteria (NTM), co-morbid diseases or history of previous TB. Sputum smear and culture, radiological findings, drug susceptibility test results and patient demographics including gender and age were recorded at diagnosis and stored in a database. Strains were genotyped by IS*6110* DNA fingerprinting (20) and spoligotyping (21) using internationally standardized methods.

Chest radiographs were read and scored using a modified version of the recently described chest radiograph reading and recording system (CRRS) (22). This system provides for the collection of data on the most significant forms of lung disease for which radiology is a useful screening method. Chest radiographs were read by a single reader with clinical experience in chest radiology and who was blinded to the genotypic data. Chest radiographs were evaluated according to film quality (optimal, suboptimal and unreadable) and the presence or absence of abnormalities associated with TB. Abnormalities recorded included parenchymal (large (>1cm) and small (<1cm) opacities, cavities), pleural (pleural fluid/ fibrosis), central structure (hilar and mediastinal lymphadenopathy) and other (bronchiectasis, hyperinflation) abnormalities. Relationships between the occurrences of various chest radiograph findings and Beijing/non-Beijing genotype cases were tested using cross tabulation and the Fisher exact test.

This study was approved by the Ethics review board of Stellenbosch University. Permission for the study was also granted by the Director of Health, City of Cape Town, and the local health committees representing the communities.

Results

Two hundred and seventy-four HIV-uninfected, new pulmonary TB patients were enrolled in the study. Complete clinical and demographic data as well as good quality chest radiographs were available from 241 of these patients. The patients represented in this cohort were not statistically different from the patients in the entire cohort according to the demographic measures of age, gender, smear grading and drug susceptibility patterns. Eighty-five (35%) of the patients were infected with a Beijing genotype strain, while the remaining 156 (65%) patients were infected with a non-Beijing genotype strain. Comparison of clinical and demographic data showed that patients infected with a Beijing genotype strain were more likely to have a higher bacillary load as measured by their sputum smear gradation (p = 0.008; Table 1). Analysis of the chest radiographs showed a weak association between extrapulmonary TB (pleural fluid and pleural fibrosis) and the non-Beijing genotype (p = 0.03; Table 2). All of the other clinical and demographic parameters were not associated with a defined *M. tuberculosis* genotype.

Discussion

This study showed that the clinical and demographic characteristics of patients infected with either a Beijing genotype strain or a non-Beijing genotype strain are very similar with the exception that the bacterial load was significantly higher in patients infected with a Beijing genotype strain. This suggests that these patients are more infectious leading to ongoing transmission in this setting. This supports a recent finding which showed that the proportion of TB cases with a Beijing genotype strain was increasing significantly over time (23, 24). Van der Spuy *et al* (2009) showed a weak association between smear-positive disease and the Beijing genotype but suggested that other factors were required to explain the growth rate in the number of cases with a Beijing genotype strain (23). The association between smear gradation and extent of disease suggests that Beijing genotype strains cause more extensive disease, however, this was not supported by macroscopic lung pathology according to the chest radiograph analysis as no statistical difference could be found for the expected abnormalities (cavitations, consolidation and infiltration) associated with smear positivity (25) and the genotype of the strain causing disease. This can in part

be explained by the fact that the majority of cases included in the study cohort had smear-positive disease and advanced radiological pathology suggestive of diagnostic delay. Both diagnostic delay and the extent of disease may mask subtle pathological differences associated with the genotype of the strain causing disease. Thus, in these patients a relationship between bacterial load and lung pathology may not be evident.

In this study, the non-Beijing genotype was weakly associated with chest radiograph abnormalities consistent with pleural TB as defined by the presence of pleural fluid or pleural fibrosis. This suggests that non-Beijing genotype strains may cause more disseminated disease than strains of the Beijing genotype in the study setting. However, we acknowledge that we were unable to show that these patients had other extrapulmonary manifestations (i.e. skeletal, central nervous system) beyond the pleura since the clinical investigation in this study was limited to pulmonary and pleural TB. In general our findings concurred with previous studies which failed to demonstrate an association between the Beijing genotype and radiological features (7, 13, 26 27). However, our findings differed from the study done by Drobniewski *et al* (2005) who showed that strains of the Beijing genotype were associated with more advanced disease as defined by amongst others fibrotic changes (19). These conflicting findings could be attributed to the population structure of the Beijing genotype strains in the different study settings as well as host population genetics (28, 29). Previously it was suggested that strains from the different Beijing sublineages may have evolved unique pathogenic characteristics (29).

We acknowledge that this study was limited by the fact that only one reader's observation was compared with the genotypic data. However, this is in agreement with a previous study (18) and inter-reader agreement has been shown to be limited (30).

In conclusion, this study supports the notion that the Beijing genotype may influence the clinical presentation of disease in certain host populations. However, we conclude that the resolution of the chest radiograph may be insufficient in patients with prolonged diagnostic delay to identify minor differences in pathology associated with a defined bacterial genotype.

Acknowledgments

We thank the European Commission 6th framework programme (project no. 037919) for financial support. We thank the nurses and data annalists for the collection and processing of clinical data. We thank the Department of Health, City of Cape Town and are indebted to the residents of the epidemiological field site.

Reference List

1. **Glynn, J.R., Whiteley, J., Bifani, P.J., Kremer, K. & van Soolingen, D**. 2002. Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: a systematic review. Emerging Infect. Dis. 8, 843-849.

2. **Bifani PJ, Mathema B, Kurepina NE, Kreiswirth BN**. 2002. Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. Trends Microbiol. 10:45–52.

3. **Agerton, T.B., Valway, S.E., Blinkhorn, R.J., Shilkret, K.L., Reves, R., Schluter, W.W. et al.** 1999. Spread of strain W, a highly drug-resistant strain of *Mycobacterium tuberculosis*, across the United States. Clin. Infect. Dis. 29:85-92; discussion 93-95.

4. **Abebe, F. & Bjune, G.** 2006. The emergence of Beijing family genotypes of *Mycobacterium tuberculosis* and low-level protection by bacille Calmette-Guérin (BCG) vaccines: is there a link? Clin. Exp. Immunol. 145:389-397.

5. **Lan, N.T.N., Lien, H.T.K., Tung, L.B., Borgdorff, M.W., Kremer, K. & van Soolingen, D.** 2003. *Mycobacterium tuberculosis* Beijing genotype and risk for treatment failure and relapse, Vietnam. Emerging Infect. Dis. 9:1633-1635.

6. **Sun, Y., Lee, A.S.G., Wong, S. & Paton, N.I.** 2006. Association of *Mycobacterium tuberculosis* Beijing genotype with tuberculosis relapse in Singapore. Epidemiol. Infect. 134:329-332.

7. **Kong, Y., Cave, M.D., Zhang, L., Foxman, B., Marrs, C.F., Bates, J.H. et al.** 2007. Association between *Mycobacterium tuberculosis* Beijing/W lineage

strain infection and extrathoracic tuberculosis: Insights from epidemiologic and clinical characterization of the three principal genetic groups of M. tuberculosis clinical isolates. J. Clin. Microbiol. 45:409-414.

8. **López, B., Aguilar, D., Orozco, H., Burger, M., Espitia, C., Ritacco, V. et al**. 2003. A marked difference in pathogenesis and immune response induced by different *Mycobacterium tuberculosis* genotypes. Clin. Exp. Immunol. 133:30-37.

9. **Dormans, J., Burger, M., Aguilar, D., Hernandez-Pando, R., Kremer, K., Roholl, P. et al**. 2004. Correlation of virulence, lung pathology, bacterial load and delayed type hypersensitivity responses after infection with different *Mycobacterium tuberculosis* genotypes in a BALB/c mouse model. Clin. Exp. Immunol. 137:460-468.

10. **van Crevel, R., Nelwan, R.H., de Lenne, W., Veeraragu, Y., van der Zanden, A.G., Amin, Z. et al**. 2001. *Mycobacterium tuberculosis* Beijing genotype strains associated with febrile response to treatment. Emerging Infect. Dis. 7:880-883.

11. **Borgdorff, M.W., de Haas, P., Kremer, K. & van Soolingen, D**. 2003. *Mycobacterium tuberculosis* Beijing genotype, the Netherlands. Emerging Infect. Dis. 9:1310-1313.

12. **Caws, M., Thwaites, G., Stepniewska, K., Nguyen, T.N.L., Nguyen, T.H.D., Nguyen, T.P. et al**. 2006. Beijing genotype of *Mycobacterium tuberculosis* is significantly associated with human immunodeficiency virus infection and multidrug resistance in cases of tuberculous meningitis. J. Clin. Microbiol. 44:3934-3939.

13. **Feng, J., Su, W., Tsai, C. & Chang, S.** 2008. Clinical impact of *Mycobacterium tuberculosis* W-Beijing genotype strain infection on aged patients in Taiwan. J. Clin. Microbiol. 46:3127-3129(2008).

14. **Glynn, J.R., Crampin, A.C., Traore, H., Yates, M.D., Mwaungulu, F.D., Ngwira, B.M. et al**. 2005. *Mycobacterium tuberculosis* Beijing genotype, northern Malawi. Emerging Infect. Dis. 11:150-153.

15. **de Jong, B.C., Hill, P.C., Aiken, A., Awine, T., Antonio, M., Adetifa, I.M. et al.** 2008. Progression to active tuberculosis, but not transmission, varies by *Mycobacterium tuberculosis* lineage in The Gambia. J. Infect. Dis. 198:1037-1043.

16. **Marais, B.J., Hesseling, A.C., Schaaf, H.S., Gie, R.P., van Helden, P.D. & Warren, R.M.** 2009. *Mycobacterium tuberculosis* transmission is not related to household genotype in a setting of high endemicity. J. Clin. Microbiol. 47:1338-1343(2009).

17. **de Jong, B.C., Hill, P.C., Brookes, R.H., Gagneux, S., Jeffries, D.J., Otu, J.K. et al.** 2006. *Mycobacterium africanum* elicits an attenuated T cell response to early secreted antigenic target, 6 kDa, in patients with tuberculosis and their household contacts. J. Infect. Dis. 193:1279-1286.

18. **Borgdorff, M.W., Van Deutekom, H., De Haas, P.E.W., Kremer, K. & Van Soolingen, D.** 2004. *Mycobacterium tuberculosis*, Beijing genotype strains not associated with radiological presentation of pulmonary tuberculosis. Tuberculosis (Edinb) 84:337-340.

19. **Drobniewski F, Balabanova Y, Nikolayevsky V, Ruddy M, Kuznetzov S, Zakharova S, Melentyev A, Fedorin I**. 2005. Drug-resistant tuberculosis, clinical virulence, and the dominance of the Beijing strain family in Russia. JAMA 22:2726-2731.

20. **van Embden, J.D., Cave, M.D., Crawford, J.T., Dale, J.W., Eisenach, K.D., Gicquel, B. et al.** 1993. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J. Clin. Microbiol. 31:406-409.

21. **Kamerbeek, J., Schouls, L., Kolk, A., van Agterveld, M., van Soolingen, D., Kuijper, S. et al.** 1997. Simultaneous detection and strain differentiation of

Mycobacterium tuberculosis for diagnosis and epidemiology. J. Clin. Microbiol. 35:907-914.

22. **Den Boon, S., Bateman, E.D., Enarson, D.A., Borgdorff, M.W., Verver, S., Lombard, C.J. et al.** 2005. Development and evaluation of a new chest radiograph reading and recording system for epidemiological surveys of tuberculosis and lung disease. Int. J. Tuberc. Lung Dis. 9:1088-1096.

23. **van der Spuy, G.D., Kremer, K., Ndabambi, S.L., Beyers, N., Dunbar, R., Marais, B.J. et al.** 2009. Changing *Mycobacterium tuberculosis* population highlights clade-specific pathogenic characteristics. Tuberculosis (Edinb) 89:120-125.

24. **Cowley, D., Govender, D., February, B., Wolfe, M., Steyn, L., Evans, J. et al.** 2008. Recent and rapid emergence of W-Beijing strains of *Mycobacterium tuberculosis* in Cape Town, South Africa. Clin. Infect. Dis. 47:1252-1259.

25. **Wang, C., Chen, H., Chong, I., Hwang, J. & Huang, M**. 2008. Predictors for identifying the most infectious pulmonary tuberculosis patient. J. Formos. Med. Assoc. 107:13-20.

26. **Thwaites, G., Caws, M., Chau, T.T.H., D'Sa, A., Lan, N.T.N., Huyen, M.N.T. et al.** 2008. Relationship between *Mycobacterium tuberculosis* genotype and the clinical phenotype of pulmonary and meningeal tuberculosis. J. Clin. Microbiol. 46:1363-1368.

27. **Sun, Y., Lim, T.K., Ong, A.K.Y., Ho, B.C.H., Seah, G.T. & Paton, N.I**. 2006. Tuberculosis associated with *Mycobacterium tuberculosis* Beijing and non-Beijing genotypes: a clinical and immunological comparison. BMC Infect. Dis. 6:105.

28. **Hanekom, M., van der Spuy, G.D., Streicher, E., Ndabambi, S.L., McEvoy, C.R.E., Kidd, M. et al**. 2007. A recently evolved sublineage of the *Mycobacterium tuberculosis* Beijing strain family is associated with an increased ability to spread and cause disease. J. Clin. Microbiol. 45:1483-1490.

29. **Hanekom, M., van der Spuy, G.D., Gey van Pittius, N.C., McEvoy, C.R.E., Ndabambi, S.L., Victor, T.C. et al.** 2007. Evidence that the spread of *Mycobacterium tuberculosis* strains with the Beijing genotype is human population dependent. J. Clin. Microbiol. 45:2263-2266.

30. **Yerushalmy, J.** 1969. The statistical assessment of the variability in observer perception and description of roentgenographic pulmonary shadows. Radiol. Clin. North Am. 7:381-392.

Table 1: Clinical and demographic of patients according to the genotype of the M. tuberculosis strain causing disease.

Characteristics of cohort	Beijing (n=85) (%)	Non-Beijing (n = 156) (%)	*p*-values
Age	33.7	33.1	$p = 0.7$
Gender			
Male	52 (35.6)	94 (64.4)	$p = 0.7$
Female	33 (34.7)	62 (65.3)	
Drug susceptibility			
Sensitive	76 (36)	133 (64)	$p = 0.25$
INH mono-resistance	6 (25)	18 (75)	
Smear gradation			
Low (negative/1+)	39 (29)	98 (71)	$p = 0.008$
High (2+/3+)	46 (45)	56 (55)	
Chest radiograph abnormalities[a]			
Yes	84 (35.0)	156 (65)	$p = 0.35$
No	1 (100)	0	

[a] see Table 2

Table 2: Chest radiograph abnormalities of TB patients infected with either strains of the Beijing genotype or strains of the non-Beijing genotype.

Chest radiograph findings	Beijing (n = 85) (%)	Non-Beijing (n = 156) (%)	*p*-values
Abnormalities consistent with TB			
Yes	84 (35.3)	154 (64.7)	$p = 0.72$
No	1 (33.3)	2 (66.7)	
A. Pulmonary TB			
Parenchymal abnormalities			
Yes	82 (34.6)	155 (65.4)	
No	3 (75.0)	1 (25.0)	$p = 0.13$
• Cavity			
Yes	80 (35.2)	147 (64.8)	$p = 0.58$
No	5 (35.7)	9 (64.3)	
• Small opacities (< 1cm)			
○ Number of lobes involved			
1 lobe	4 (33.3)	8 (66.7)	$p = 0.58$
>1 lobe	81 (35.4)	148 (64.6)	
○ Extent of disease			
unilateral	30 (33.7)	59 (66.3)	$p = 0.40$
bilateral	55 (36.2)	97 (63.8)	
• Large opacities (> 1cm)			
○ Number of lobes involved			
1 lobe	73 (35.1)	135 (64.9)	$p = 0.52$
>1 lobe	12 (36.4)	21 (63.6)	
○ Extent of disease			
unilateral	82 (35.0)	152 (65.0)	$p = 0.47$
bilateral	3 (42.9)	4 (57.1)	
Central structure abnormalities			
Yes	82 (35.7)	148 (64.3)	$p = 0.41$
No	3 (27.3)	8 (72.7)	
• Enlarged hilar lymph nodes			
Yes	6 (28.6)	15 (71.4)	$p = 0.34$
No	79 (35.9)	141 (64.1)	
• Enlarged mediastinal lymph nodes			
Yes	2 (100.0)	0	$p = 0.12$

No	83 (34.7)	156 (65.3)	
Other lung abnormalities			
• Hyperinflation			
Yes	20 (34.5)	38 (65.5)	$p = 0.51$
No	65 (35.5)	118 (64.5)	
• Bronchiectasis			
Yes	9 (27.3)	24 (72.7)	$p = 0.20$
No	76 (36.5)	132 (63.5)	
B. Pleural TB			
Pleural abnormalities			
Yes	54 (32.9)	110 (67.1)	$p = 0.17$
No	31 (40.3)	46 (59.7)	
• Pleural fluid/fibrosis			
Yes	18 (25.7)	52 (74.3)	$p = 0.03$
No	67 (39.2)	104 (60.8)	

Conclusion

Tuberculosis (TB) has been a major concern to those providing health care since the dawn of recorded human history. The specific approach to TB control has been based on prevention and treatment of infection by means of vaccination (primary prevention), decrease the development of active disease in individuals already infected by means of preventative therapy (secondary prevention) and eradication of the bacilli in individuals with active disease by means of curative chemotherapy (active treatment). However, evidence shows that BCG vaccine protection ranges from 0 to 80% and antimicrobial agents have encouraged the evolution of resistant bacteria often resulting in treatment failure. To address these limitations in TB control, the focus of research has been on the development of new vaccines and anti-TB drugs. However, this has been done without a good understanding of the genetic background of the strains of *Mycobacterium tuberculosis* and the population structure of these strains in different geographical settings.

The development of molecular epidemiological tools has allowed the focus of research to turn towards understanding the *M. tuberculosis* genome in order to amongst others, determine the influence of genetic strain variation on disease presentation and outcome (association of different bacterial strains with drug resistance, treatment failure, clustering, and the BCG vaccine).

The Beijing strain genotype has been extensively studied (as discussed in Chapter 1) in different geographical settings because of their worldwide occurrence and association with major TB outbreaks. It has been suggested that strains with the Beijing genotype have the ability to evade the protective effect of the BCG vaccine, to transmit more readily, to cause more severe forms of disease, to elicit a non-protective host immune response and to acquire drug resistance. In Chapter 2, we described the phylogeny of the Beijing lineage and showed that the pathogenic characteristics of the Beijing lineage are not conserved but rather that strains within the different sublineages have evolved unique pathogenic characteristics. The most pathogenic strains, as measured by their frequency of occurrence, were associated with the accumulation of deletions thus, suggesting that evolution has influenced the ability of

strains to transmit and cause disease. Strains with the least genomic changes (termed "atypical Beijing strains) were less frequently found to transmit than the more recently evolved strains (termed "typical" Beijing strains). Analysis of drug resistance data showed that the frequency of occurrence of the drug resistance phenotype was not associated with the most recently evolved sublineage. Interestingly, we showed in another setting that the "atypical" Beijing strains, which according to current dogma are attenuated in their ability to transmit, were over-represented. Transmission was linked to HIV co-infection supporting the concept that less fit strains are able to spread and cause disease in vulnerable host populations causing further concerns for TB control.

In Chapter 3, the characteristics of the Beijing strain genotype i.e. ability to transmit and cause disease were further investigated. We found a strong association between prevalence of a specific sublineage and the host population in which the strains were isolated. From this we concluded that either the strains had evolved unique properties which allowed rapid spread in their respective host populations or that the immunity of the respective host populations was unable to prevent progression to disease and subsequent spread. The findings from the above chapters have important implications for TB control as it is evident that Beijing genotype strains from the most recently evolved sublineage are highly pathogenic in their ability to spread and cause disease. Greater vigilance is needed to prevent the spread of these strains and it is thus suggested that screening of contacts be intensified as well as the implementation of rapid diagnostic tools.

The evolution of the Beijing genotype was further explored in Chapter 4 using Mycobacterial Interspersed Repetitive-Unit-Variable-Number Tandem-Repeat which is based on a PCR amplification of a combination of minisatellite-like variable-number repeat regions scattered around the bacterial genome. The ability of these markers to identify transmission of Beijing genotype strains in Cape Town, South Africa was similar to that of IS*6110* DNA fingerprinting. However, comparison of the MIRU-VNTR loci combinations and IS*6110* DNA fingerprints showed discordance between the two methods. This implies that these chromosomal markers evolve independently and at different rates. Their use as molecular epidemiological

tools is defined by the incidence of TB in the study setting and the role of immigration in introducing new strains into the study population.

Treatment failure, a significant concern for TB control, has been associated with strains of the Beijing genotype as well as with mixed infections where a drug-resistant and a drug-sensitive strain are present in a single primary culture of a TB case. In Chapter 5, using an improved PCR-based method, we were able to identify, for the first time, 5 prominent strain genotypes (Beijing, LAM, S-family, LCC and Haarlem) in primary cultures and quantify the extent of mixed infections in a high incidence setting with a high prevalence of HIV co-infection. In this study we found a low incidence of mixed infections, however, one patient presented with both a drug-sensitive and a drug-resistant strain underscoring the importance of detecting mixed infections in order to assure successful treatment and limiting TB transmission. Furthermore, we found that strains of the Beijing genotype were associated with mixed infections, which further supports the hypothesis that strains of the Beijing genotype have evolved selective advantages enabling their increased dissemination.

Genotype-specific characteristics were further explored in chapter 6 by comparing clinical and demographic data of patients infected with Beijing and non-Beijing genotype strains. We found that strains of the Beijing genotype were associated with increased bacillary load in sputum of TB cases supporting the hypothesis that these strains transmit more readily. However, this association could not be confirmed by lung pathology according to chest radiographs. This contradicts the hypothesis that the Beijing genotype cause more severe forms of disease. However, the severity of disease may be masked by diagnostic delay (either due to prolonged patient seeking behaviour or time consuming diagnostic tools) as most of these patients were highly infectious (smear-positive) and had extensive disease according to chest radiography. This demonstrates the need to educate the human population on the signs and symptoms of TB to ensure that suspects seek medical advice earlier.

Collectively, these studies show the importance of improving TB control and enhancing our understanding of the influence of different *M. tuberculosis* genotypes on clinical presentation. The identification of unique pathogenic characteristics

associated with the Beijing genotype could be useful for future vaccine and drug development. Our findings show that these Beijing genotype strains could be adapted to a certain host population and thus, future vaccines should target bacterial populations that are prevalent in a specific geographical setting (individualized rather than generalized preventative therapy). In addition, new generation drugs should target virulence factors in different *M. tuberculosis* strains.

Candidate's Contributions

Chapter 1 Phenotypic and Genotypic characterisation of the *Mycobacterium tuberculosis* Beijing genotype

- Primary author

Chapter 2 A recently evolved sublineage of the *Mycobacterium tuberculosis* Beijing strain family is associated with an increased ability to spread and cause disease

- First author
- Planning of study
- Development of molecular tools and laboratory work
- Construction of database
- Analysis and interpretation of data
- Writing of manuscript

Chapter 3 Spread of a low-fitness drug-resistant *Mycobacterium tuberculosis* strain in a setting of high human immunodeficiency virus prevalence

- Co-author
- Development of molecular tools and laboratory work

Chapter 4 Evidence that the spread of *Mycobacterium tuberculosis* strains with the Beijing genotype is human population dependent

- First author
- Planning of study
- Development of molecular tools and laboratory work
- Construction of database
- Analysis and interpretation of data
- Writing of manuscript

Chapter 5 Discordance between Mycobacterial Interspersed Repetitive-Unit–Variable-Number Tandem-Repeat Typing and IS*6110* Restriction Fragment Length Polymorphism Genotyping for analysis of *Mycobacterium tuberculosis* Beijing strains in a setting of high incidence of tuberculosis

- Joint first author
- Planning of study
- Development of molecular tools and laboratory work
- Construction of database
- Analysis and interpretation of data

- Writing of manuscript

Chapter 6 Mixed *Mycobacterium tuberculosis* infection in patients isolates resident in a high HIV prevalence setting

- First author
- Planning of study
- Development of molecular tools
- Analysis and interpretation of data
- Writing of manuscript

Chapter 7 Radiological characterisation of tuberculosis cases infected with either Beijing or non-Beijing genotype *Mycobacterium tuberculosis* strains \

- First author
- Conseptualisation and Planning of study
- Reading chest-radiographs
- Analysis and interpretation of data
- Writing of manuscript

Additional Publication

In addition to the papers which form part of this thesis, I have also contributed to the following publication listed below. I was involved with the development of the molecular tools, the laboratory work and writing and editing of the manuscript.

Warren R M, Gey van Pittius N C, Barnard M, Hesseling A, Engelke E, de Kock M, Gutierrez M C, Chege G K, Victor T C, Hoal E G and van Helden P D. *(2006)*. Differentiation of *Mycobacterium tuberculosis* complex by PCR amplification of genomic regions of difference. *J Clin Microbiol 10(7):818-822*

Supplemental Tables

Table 1

	PUBLICATION	YEAR	POPULATION STUDIED	Proportion Beijing strains to total Mtb population
AFRICA				
Bangui	Nouvel *et al*, 2006	1993-2001	Studied 53 MDR *M. tuberculosis* strains isolated from different patients at the Bangui Pasteur institute to look for diversity in a series of putative anti-mutator genes; 230 non-MDR strains were used as control group isolated from other regions	0
Burkina Faso	Godreuil *et al*, 2007	2001	120 isolates were collected from culture positive TB cases; 35.8% were HIV positive	0
Kenya (Nairobi)	Githui *et al*, 2003	Not available	73 isolates were cultured from suspected TB cases referred to the Centre for Respiratory Diseases Research, Kenya Medical Research Institute, Tuberculosis Reference and Research laboratory from public and private health care centers	6/73 (8%)
Madagascar	Ferdinand *et al*, 2005	1994-2000	333 TB strains from 4 different studies were arbitrarily selected to determine the genetic diversity of strains found in Madagascar; represented a significant proportion of the 1497 strains studied in the country between 1994 and 2000	4%
Morocco (Casablanca)	Tazi *et al*, 2007	1997-1998	150 isolates were collected from PTB cases resident in 6 different districts of Casablanca; represents over a third of isolates collected in this period	Not available

Malawi (Karonga District)	Glynn *et al*, 2005	1995-2003	1029 isolates were collected from new and retreatment culture-positive TB cases identified at hospital and peripheral clinics as part of a longstanding epidemiologic study of TB and HIV in a rural area	44/1029 (4%)
Senegal	Niang *et al*, 1999	1994-1995	69 and 20 *M. tuberculosis* strains isolated from Senegalese and Ivory Coast patients respectively to investigate the degree of heterogeneity of the DR region; majority of isolates from non-hospitalised patients	8/89 (9%)
Sierra Leone	Homolka *et al*, 2008	2003-2004	97 strains isolated from smear positive cases registered for retreatment in western area and Kenema districts	4/97 (4%)
Tanzania (Dar es Salaam)	Eldholm *et al*, 2006	Oct and Nov 2005	147 isolates were collected in Dar es Salaam	7/145 (5%)
Tanzania	Kibiki *et al*, 2007	Apr-Sept 2005	130 culture-positive samples were collected from HIV positive and negative patients from the majority of Chagga and Masai tribes	7/130 (5%)
Uganda (Kampala)	Asiimwe *et al*, 2008	Feb 2006-Nov 2006	Cross-sectional study; 344 isolates from new, smear-positive adult peri-urban patient population from Rubaga division, Kampala attending 4 main division TB clinics	4/344 (1.2%)
Uganda (Kampala)	Niemann *et al*, 2002	1995-1997	234 strains were isolated from sputum samples which were obtained from newly diagnosed smear-positive PTB patients from the largest clinic in Uganda	3/234 (1.3%)
South Africa (Cape Town)	Richardson *et al*, 2002	1993-1998	813 isolates were collected from culture-positive TB cases	137/813(17%)
South Africa (Cape Town)	Van Rie *et al*, 2001	Apr 1992-March 1997	61 drug R TB cases resident in 2 neighbouring communities were selected	15/61 (25%)
South Africa (Cape Town)	Van Rie *et al*, 1999	Jan 1993-March 1997	606 isolates were collected at the primary health care clinics of 2 neighbouring communities studied from new and retreatment cases	17/606 (3%)
South Africa (Western and Eastern Cape)	Strauss *et al*, 2007	2001-2004	312 Rif resistant isolates were collected from Cape Town and 117 from Port Elizabeth from retreatment PTB patients with and	175/429 (41%)

			without HIV co-infection respectively, attending health care clinics or referral hospitals	
South Africa (BOKS)	Streicher *et al*, 2004	2001-2002	328 drug R isolates were collected from 72 different clinics in the Boland-Overberg and Southern Cape-Karoo regions of the Western Cape Province	91/328 (28%)
Tunisia	Namouchi *et al*, 2008	2001-2005	378 isolates were collected from 3 northern representative regions of Tunisia	0
Zimbabwe	Easterbrook *et al*, 2004	May 1997-Oct 1997	214 isolates were collected from new and retreatment smear-positive adult PTB patients presenting to the main referral hospital for TB and other infectious diseases in Harare (represents half of *M. tuberculosis* infected population). 74% of TB patients also had HIV	4/214(2%)
CENTRAL AMERICA				
Cuba (Havana)	Diaz *et al*, 1998	1994-1995	160 isolates from new and retreatment cases analysed were obtained from the National Reference Laboratory for Tuberculosis (37% of TB cases proven by culture during the study period)	23/160 (14%)
EUROPE				
Austria	Stauffer *et al* 2000	1995-1998	Drug resistance patterns of 3559 M. tuberculosis strains were evaluated; results obtained from all laboratories (9) performing drug resistant testing in Austria	Not available (only mentions 2 Beijing strains with MDR; ?Beijing=drug sensitive or drug resistant)
Azerbaijan	Pfyffer *et al*, 2001	1995-1996	65 smear-positive PTB male prisoners from Azerbaijan were included in the study; on diagnosis patients were transferred to the Central Penitentiary Hospital in Baku	46/65 (71%)
Belgium	Allix-Béguec *et al*, 2008	2003-2004	530 isolates were collected from culture-confirmed new and retreatment TB cases; represents 89% of all bacteriologically	5/530 (0.9%)

			confirmed cases	
Belgrade	Vuković *et al*, 2003	Apr-Sept 1998 May-Oct 1999	176 randomly chosen patients with permanent residence in Belgrade who were diagnosed with culture proven PTB; representative of 48.4% of total cases reported however representative of entire population of TB patients	0
Bulgaria	Valcheva *et al*, 2008	Jan 2005-June 2007	133 isolates collected from newly diagnosed adult PTB patients resident in different regions in Bulgaria	0
Bulgaria	Valcheva *et al*, 2008	2004-2006	113 isolates collected from epidemiologically unlinked, HIV negative adult PTB patients who were admitted to TB hospitals in different regions in Bulgaria to study the population structure of *M. tuberculosis* in Bulgaria	0
Germany	Diel *et al*, 2002	1997-1999	423 isolates collected from culture-confirmed PTB and non-PTB patients that were reported to each of the 7 district public health departments of Hamburg; prospective, population-based study	15/423 (4%)
Germany	Hilleman *et al*, 2005	2001	Resistance mutations of 103 MDR strains isolated in Germany were investigated (history of previous treatment unknown) (represent >90% of all MDR cases reported in 2001) and 10 drug sensitive isolates were used as controls	62/103 (60%)
Germany	Kubica *et al*, 2004	1995-2001	451consecutive MDR-TB strains were obtained from patients living in Germany. Strains analysed were a subset of 1511 drug resistant isolates partly used for previous studies	175/451 (39%)
Estonia	Kruuner *et al*, 2001	1994	299 (only 209 genotyped) smear positive cases diagnosed and treated at 2 hospitals; represents 58% of new cases	61/209 (29%)
Estonia	Werngren *et al*, 2003	Not available	13 drug sensitive isolates were included in this study to determine whether drug sensitive Beijing isolates *in vitro* generated a larger number of Rif resistant mutants and therefore are more prone to acquire drug resistance than non-Beijing isolates	6/13 (46%)
Finland	Puustinen *et al*, 2003	2000-2001	380 *M. tuberculosis* isolates from TB patients living in Finland;	4/380 (1%)

			represents 90% of new cases	
Finland	Martilla *et al*, 2008	1995-2004	181 drug resistant culture-positive isolates from new TB cases were collected and genotyped	16/181 (9%)
United Kingdom (Scotland)	Fang *et al*, 1999	1990-1997	715 *M. tuberculosis* isolates received by the Scottish Mycobaterium Reference Laboratory in Edingburgh were analysed	2/715 (0.3%)
United-Kingdom (London)	Goyal *et al*, 1997	Not available	isolates were collected from 167 available patients with culture-confirmed TB presenting to 3 large hospitals in Northwest London; RFLP and spoligotyping's abilities were compared to identify clusters	7/167 (4.2%)
Italy (Tuscany)	Lari *et al*, 2006	2002-2004	30 Beijing, 24 non-Beijing MDR and 28 non-Beijing drug sensitive isolates were collected randomly to evaluate the contribution of *mut*T gene mutations to drug resistance	Not applicable
Italy (Tuscany)	Lari *et al*, 2007	2002-2004	829 *M. tuberculosis* isolates from TB patients living in Tuscany and admitted to 10 major community hospitals	54/829 (7%)
Italy (Tuscany)	Rindi *et al*, 2008	2002-2005	1079 isolates were collected from TB patients hospitalized in Tuscany, Italy	73/1079 (6.8%)
Italy (Trieste district)	Dolzani *et al*, 2004	1999-2000	51 consecutive isolates of *M. tuberculosis* were collected; they represent virtually all isolates recovered from patients during 1999 and 2000	2/51 (4%) belong to PGG1
Mediterranean islands	Sola *et al*, 2005	1994-2001	192 isolates were collected from patients living in Sardinia and Sicily	2/192 (1%)
Latvia	Tracevska *et al*, 2003	1999-2001	109 isolates were randomly collected from new and retreatment MDR cases to determine the prevalence of the Beijing genotype in Latvian MDR isolates	63/109 (58%)
Poland	Sajduda *et al*, 2004	2000	251 isolates were collected from new and retreatment patients as part of a national drug resistance survey	7/251 (3%)
Portugal	Perdigão *et al*, 2008	Not available	58 MDR isolates were collected to determine drug resistance profile, genotyping according to MIRU-VNTR and analysing drug	Not available

			conferring mutations	
Portugal	Portugal *et al*, 2008	2001-2002	290 isolates were randomly collected from patients in Lisbon hospitals to determine genetic diversity of *M. tuberculosis* strains	Not available
Russia	Mokrousov *et al*, 2003	1996-2002	354 Isolates obtained from newly diagnosed or recurrent TB epidemiologically unlinked adult patients living in different regions of Russian Federation and who were admitted to hospitals. (cross-sectional/population-based studies)	198/354 (56%)
Russia (Archangel Oblast)	Toungoussova *et al*, 2002	1998-1999	119 isolates from new and retreatment PTB patients from Archangel Oblast	53/119 (44.5%)
Russia (Archangel prison)	Toungoussova *et al*, 2003	2001	114 isolates were collected from new and retreatment male prisoners	87/114 (76%)
Russia (Archangel Oblast)	Toungoussova *et al*, 2005	1998-2005	77 MDR isolates from new and retreatment patients from community and prisons of the Archangel Oblast were collected to test their R against capreomycin, kanamycin, ethionamide and ofloxacicin	63/77 (82%)
Russia (Tula Region)	Ignatova *et al*, 2006	2001-2002	87 isolates were collected from culture-positive new and retreatment male PTB patients diagnosed in the Ozerki prison	38/87 (44%)
Russia (Samara region)	Drobniewski *et al*, 2005	2001-2002	880 isolates from new and retreatment cases were collected as part of a cross-sectional population-based molecular epidemiological study of all civilian and penitentiary TB facilities	586/880 (67%)
Russia (Samara region)	Balabonova *et al*, 2006	Not available	Cross-sectional population molecular epidemiological study that showed that the Beijing genotype prevailed among the 880 obtained isolates	586/880 (67%)
Russia (St. Petersburg)	Narvskaya *et al*, 2002	Aug 1998-May 1999	35 *M. tuberculosis* isolates from 19 patients were collected to define a nosocomial outbreak of MDR TB	7/19 (37%)
Russia	Mokrousov *et al*, 2002	1996-2001	261 isolates (204 INH R and 57 INH S) were collected from new and retreatment hospitalised cases resident in St. Petersburg,	123/261 (47%)

			Leningrad Oblast, Novgorod and Pskov respectively	
Russia (Serpukhov prison)	Shemyakin *et al*, 2004	2001	130 isolates (17=drug sensitive and 113=drug resistant) from HIV (-) male patients who were previously treated for TB, were studied	55/113 (49%)
Russia (Archangel Oblast)	Toungoussova *et al*, 2004	1998-2001	55 *M. tuberculosis* isolates were selected from a collection of clinical *M. tuberculosis* isolates from patients with PTB	24/55 (44%)
Russia (Ukraine)	Nikolayevskyy *et al*, 2006	Jun-Jul 2003 and 2004 (Odessa);2003 (Nikolaev)	225 isolates collected from clinically and radiological confirmed new and chronic TB patients attending 2 central regional TB dispensaries in Odessa and Nikolaev; all INH and/or Rif resistant	89/225 (40%)
Russia	Mokrousov *et al*, 2002	May 1997-Sept 1999	38 MDR strains were isolated in St. Petersburg, Leningrad Oblast and Novgorod from retreatment patients hospitalised in specialised TB hospital most of whom had been subjected to surgery	26/38 (68%)
Russia (Kaliningrad)	Mokrousov *et al*, 2008	March-June 2006	90 isolates were randomly selected collected from new and retreatment PTB epidemiologically unlinked patients who were permanent residents in Kaliningrad region	41/90 (46%)
Russia	Narvskaia *et al*, 2002	Not available	67 isolates collected from different specimens of 24 patients operated on for chronic progressive PTB living in north-western Russia	36/67 (53%)
Russia (Ural region)	Kovalev *et al*, 2005	Oct 2001-June 2002	92 isolates were collected from new and retreatment PTB patients	50/92 (54%)
Russia (Moscow)	Lipin *et al*, 2007	1998-2001	217 MDR isolates were obtained from new and retreatment PTB patients resident in Tula and Serpukhov District	92/217 (42%)
Sweden	Ghebremichael *et al*, 2008	1994-2005	400 Drug resistant isolates obtained from all 6 laboratories; represents 8,5% of all culture-confirmed TB patients in Sweden	48/400 (12%)
Sweden	Brudey *et al*, 2004	1999-2002	220 isolates cultured from pathological samples in Göteberg; only 191 were genotyped and 213 had DST done	11/191 (6%)

Spain	Samper *et al*, 2005	1998-2000	180 MDR-TB isolates collected from 118 mycobacteriology laboratories located within different hospitals in Spain (isolates represent two-thirds of all Spanish MDR-TB cases); new and retreatment cases included	11/180 (6%)
Spain (Gran Canaria Island)	Pena *et al*, 2003	1993-1996	409 isolates collected from culture-confirmed TB patients in either 2 of the referral hospitals on the Gran Canaria Island.	75/409 (18%)
Spain (Gran Canaria Island)	Caminero *et al*, 2001	1993-1996	566 pts diagnosed with culture-confirmed TB (78.5% of cases) were evaluated	75/566 (13%)
Spain (Elche)	Garcia-Pachon *et al*, 2007	1993-2005	A search for Beijing family strains was performed among 332 *M. tuberculosis* isolates obtained from TB cases (represents 73% of isolations during this period)	1/332 (0.3%)
Spain (Madrid)	De Viedma *et al*, 2006	2001-2002	510 MTB isolates obtained from patients from 2 tertiary teaching hospitals in Madrid; representative of 30% of all TB cases	8/510 (1.6%)
Turkey	Durmaz *et al*, 2007	1994-2005	200 drug resistant isolates were collected in four different regions in Turkey (Marmara, Mediterranean, Aegean, East Anatolia) (regions represent 2/3rd of all TB cases in this country)	1/200 (0.5%)
Turkey	Gencer *et al*, 2005	2001-2002	381 isolates were collected from 6 regional TB laboratories (Ankare, Antalya, Kayseri, Trabzan, Samsun and Van)	7/381 (2%)
Turkey (Istanbul)	Koksalan *et al*, 2006	March 2002-Sept 2004	4069 isolates were collected as part of a prevalence study	46/4069 (1%)
Turkey (Izmir)	Bicmen *et al*, 2007	Not available	Description of the molecular characterisation of 56 *M. tuberculosis* isolates recovered from patients in Izmir by 3 different molecular methods	1/56 (1.8%)
FAR-EAST ASIA				
China (Beijing)	Van Soolingen *et al*, 1995	1992-1994	49, 20, 2594 and 934 isolates were collected from Beijing, Mongolia, The Netherlands and 16 other countries respectively from both hospitalised and non-hospitalised TB patients	42/49 (86%) 10/20 (50%) 82/2594 (3%) 32/934 (8%)

China (Beijing)	Mokrousov *et al*, 2006	2002-2005	123 isolates randomly selected from adult PTB pts admitted to Beijing Chest Hospital; epidemiologically unlinked; Beijing isolates further subdivided into old and modern sublineages on basis of NTF locus analysis	113/123 (92%)
China (Beijing)	Wei-Wei *et al*, 2007	2002-2005	121 isolates were randomly selected among those isolated in the Beijing Chest Hospital (these subsets did not reflect the real state of resistant strains in the local M. tuberculosis population).	111/121 (92%)
China (Beijing, Shanghai, Tianjin)	Li *et al*, 2005	2000	408 isolates were randomly collected from PTB patients enrolled in a nationwide random survey	265/408 (65%)
China (Hong Kong)	Chan *et al*, 2001	1998-1999	500 isolates from new TB patients were selected at random from samples received at the TB Reference Laboratory in Hong Kong	337/500 (67%)
China (Tianjin)	Chai *et al*, 2007	Aug-Dec 2005	96 isolates were collected in Tianjin Haihe Hospital	88/96 (92%)
China (Tibet)	Shi *et al*, 2007	Not available	217 isolates were collected from 6 regions of Tibet	190/217 (88%)
Indonesia (Jakarta)	Van Crevel *et al*, 2001	Dec 1998-Mrch 1999	92 drug resistant isolates were obtained from new and retreatment patients enrolled in an outpatient clinic in Jakarta	31 (34%)
Indonesia (West Java and Timor)	Parwati *et al*, 2008	2001-2006	813 and 84 isolates were prospectively collected from TB patients in West Java (Jakarta and Bandung) and Timor respectively at 2 outpatient clinics and 2 hospitals (EPTB cases were included) (not representative of Indonesia as a whole)	280/897 (31%)
Japan	Millet *et al*, 2007	Jul 2003-Feb 2005	101 isolates were collected from patients at the University Hospital, Okinawa prefecture, Ryukyu Islands (representative of 1 out of 2 cases in setting)	72/101 (71%)
Korea	Park *et al*, 2005	1998-2001	241 drug sensitive and 502 drug resistant strains were selected for this study after DST among clinical isolates referred to Korean Institute of Tuberculosis in Seoul, Korea	569/743 (77%)
Malaysia	Dale *et al*, 1999	1993-1994	439 random samples of *M. tuberculosis* complex isolates from all parts of Malaysia were provided by the Institute for Respiratory	83/439 (19%)

			Medicine	
Singapore	Sun *et al*, 2007	Aug 1994-Dec 1996	Population-based study where 598 isolates were collected from the Central Tuberculosis Laboratory, Singapore General Hospital	338/598 (57%)
Thailand (Bangkok)	Prodinger *et al*, 2001	1999-2000	204 drug sensitive isolates were obtained from TB patients at a hospital in Bangkok that treats 625 TB patients annually	90/204 (44%)
Taiwan (Sioulin township)	Chuang *et al*, 2008	2003-2004	138 isolates collected from culture-confirmed PTB and non-PTB cases resident in an aboriginal township (Sioulin Township) from Taiwan; new and retreatment cases included	53/138 (38%)
Taiwan	Jou *et al*, 2005	2002	421 isolates were collected from 5 general hospitals located in northern, eastern, western and central regions of Taiwan; included new and retreatment cases, pulmonary and extra-pulmonary TB	187/421 (44%)
Vietnam (Hanoi and Ho Chi Minh City)	Tuyen *et al*, 2000	1994-1996	168 isolates were collected from new TB cases; 67 from the north, 66 HIV negative from the south and 35 HIV positive from the south of Vietnam (representatives more of the health status of towns than rural areas)	65/168 (38%)
Vietnam (Hanoi and Ho Chi Minh City)	Anh *et al*, 2000	1998-April 1999	563 isolates were collected from new TB patients whose age and BCG status were known, from 2 centers in Hanoi and Ho Chi Minh City	301/563 (54%)
Vietnam (Ho Chi Minh City)	Caws *et al*, 2006	2001-2003	222 isolates collected from HIV negative and positive TBM adult patients enrolled as part of a randomized control trial and typing study on bacterial genetics in extrapulmonary TB respectively.	113/222 (51%)
China, Japan, Korea and Taiwan	Qian *et al*, 2002	Not available	66 Rifampin-resistant MDR TB isolates were collected from 4 different regions to analyse the 81-bp region of the *rpo*B gene.	50/66 (76%)
Indonesia, Malaysia, the Philippines, Vietnam	Ang *et al*, 2008	1996-2006	103 MDR isolates were collected from government state-owned hospitals and private clinics	32/103 (31.1%)
MIDDLE-EAST-CENTRAL ASIA				
Bangladesh	Shamputa *et al*, 2004	Sept 2001-Feb	97 smear positive TB patients were enrolled from 3 hospitals in the	7/97 (7%)

		2003	greater Mymensingh District of Bangladesh; all except 1 were retreatment cases	
Bangladesh	Rahim *et al*, 2007	Jul 2001-Nov 2003	224 strains were isolated from non-hospitalised TB patients of selected urban (Dhaka) and rural (Matlab) areas of Bangladesh	34/224 (15.2%)
Bangladesh	Storla *et al*, 2006	Not available	111 isolates from new PTB patients living in rural Sunamganj district were collected	3/111 (2.7%)
India (Delhi)	Singh *et al*, 2004	Jan 2001-Mrch 2001	105 isolates from culture-pos new and retreatment TB patients enrolled from district TB and primary health centre and tertiary-care referral hospital in the beginning of a multicentre study; representative of the Delhi community	9/105 (8.6%)
India (7 regions)	Singh *et al*, 2007	Jan 2001-Sept 2003	540 isolates were collected from new, smear-positive PTB cases resident in Delhi, Lucknow, Pune, Vellore, Nagpur, Chennai and Trivandrum.	23/540 (4.3%)
India (Kanpur)	Sharma *et al*, 2008	May 2004-Feb 2005	97 isolates obtained from new and retreatment TB patients coming at 3 TB units at Ghatampur, Bidhnoo and Bilhaur of Kanpur Nagar rural district; part of a drug surveillance study (represents 80% of total isolates from Kanpur Nagar district)	4/97 (4%)
India	Gutierrez *et al*, 2006	1997-2002	91 isolates collected from adult PTB cases resident in 12 different regions spread across the country, including northern, central and southern India (not representative of all isolates, however provide diversity present in India)	9/91 (10%)
India	Suresh *et al*, 2006	2001-2003	As part of a drug resistant study in 3 cities in North India (Delhi, Lucknow, Pune) and 2 cities in South India (Chennai and Trivandrum), 149 isolates phenotypically resistant to RIF and INH with or without resistance to any other drugs, were collected from smear-positive new and retreatment PTB cases	16/149 (10.7%)
India (Mumbai)	Almeida *et al*, 2005	Jan 2001-July 2002	196 isolates with available patient data were analysed collected from cases visiting a tertiary care center (referral bias towards	45/196 (23%)

			nonresponding cases)	
India (Bombay)	Mistry *et al*, 2002	Not available	Describe a cross-sectional analysis of a panel of 65 MDR TB isolates	2/65 (3%)
Iran and Afghanistan	Ramazanzadeh *et al*, 2006	2003-2004	345 patients were enrolled with TB who presented to the National Research Institute of Tuberculosis and Lung Disease; 258 were Iranian and 87 Afghan patients	9.4%
Iran	Doroudchi *et al*, 2000	Not available	To evaluate the usefulness of IS*6110*-RFLP and spoligotyping in the epidemiology of TB in Iran, 48 isolates were analysed	5/48 (10%)
Iran (Tehran)	Farnia *et al*, 2004	2001	129 isolates were collected from new TB cases (89 retreatment cases were excluded) (59% of total MTB population; all were incidents in Tehran during last 10 years	1/129 (0.8%)
Kazakhstan	Kubica *et al*, 2005	2001	150 isolates from new and retreatment drug resistant PTB cases were collected as part of drug surveillance study (8 isolates were dual infections)	100/142 (70%)
Kazakhstan	Hillemann *et al*, 2005	2001	179 drug resistant strains from smear-positive patients representing 100% of the drug resistant strains from 9 of the 14 Kazakhstan oblasts and 42 randomly selected drug sensitive controls were analysed during 2004. Of these 142 and 10 drug resistant (new and retreatment) and drug sensitive isolates were selected respectively for this study	95/152 (63%)
Sri Lanka	Rajapaksa *et al*, 2008	July 2005-Dec 2006	100 isolates were collected from new and retreatment TB cases at the Microbiology Laboratory at University of Colombo; not representative of whole country	14/100 (14%)
Pakistan	Hasan *et al*, 2006	Jan 2003-Jun 2004	314 isolates were collected from The Aga Khan University Hospital which represented pulmonary and extra-pulmonary samples from different geographical locations across the country	18/314 (6%)
Saudi Arabia	Al-Hajoj *et al*, 2007	2002-2005	1505 isolates were collected from 7 different regions in Saudi Arabia	67/1505 (4%)

Uzbekistan and Turkmenistan	Cox *et al*, 2005	Jul 2001-Jan 2002	382 isolates were collected in a cross-sectional survey from new and retreatment smear-positive PTB patients from 4 districts in Dashoguz Velayat in Turkmenistan and in 4 districts in the Autonomous Republic of Karakalpakstan in Uzbekistan	190/382 (50%)
Uzbekistan and Turkmenistan	Cox *et al*, 2007	Jul 2001-Jan 2002	382 isolates were collected in a cross-sectional survey from new and retreatment smear-positive PTB patients from 4 districts in Dashoguz Velayat in Turkmenistan and in 4 districts in the Autonomous Republic of Karakalpakstan in Uzbekistan	190/382 (50%)
NORTH AMERICA				
USA (NYC)	Frieden *et al*, 1996	Jan 1990-Aug 1993	267 isolates were collected from MDR TB patients infected with Beijing W strain of *M. tuberculosis* resident in all boroughs, were cared for at 41 different hospitals and were hospitalized	100%
USA (NYC)	Munsiff *et al*, 2003	1993-1999	567 patients were diagnosed with MDR-TB; 28% were found to have an isolate that was identified as either strain W or one of its variants; patients resided in all 5 NYC boroughs; Diagnosis at 56 medical and 2 correctional facilities	161/567 (28%)
USA (New-York)	Moss *et al*, 1997	1992	530 stored MDR cultures from the 1992 outbreak in NYC hospitals of new, mostly HIV positive TB cases and cultures referred for second-line DST at the Department of Health respectively, were genotyped	112/530
USA (Texas)	Soini *et al*, 2000	Not available	1429 isolates were collected as part of an ongoing, population-based TB epidemiology study in Houston, Texas	376/1429 (26%)
USA (New Jersey)	Bifani *et al*, 1999	Jan 1996-Sep 1998	1207 isolates of culture positive TB cases reported to the NJDHSS were included and genotyped as part of the National Tuberculosis Genotyping and Surveillance Network, Centers for Disease control and prevention. 68 belonged to the W family which were further analysed	68/1207 (6%)
USA	Agerton *et al*, 1999	1992-1997	104 549 TB cases reported to CDC from 9 states of the US (excl NYC and NY state)(California, Colorado, Georgia, Florida,	23/104 549 (0.02%)

			Maryland, New Jersey, Ohio, Pennsylvania and South Carolina) and Puerto Rico; of these 23 met case definition of strain W TB	
OCEANIA				
Australia (Papua New Guinea)	Gilpin *et al*, 2008	2000-2006	60 patients were diagnosed with TB from the Western Province of PNG)	Not available
SOUTH AMERICA				
Argentina (Beunos Aires)	Palmero *et al*, 2003	1991-2000	79 Initial (primary) MDR TB in HIV-negative patients treated at a referral hospital was examined by using molecular clustering	Not available
Brazil	Borsuk *et al*, 2005	1997-2001	170 isolates collected from patients attending the Health Center in Pelotas and or Rio Grande University Hospital, Rio Grande do Sul; represents 20% of all TB cases in this region.	0
Brazil (Rio de Janeiro)	Lazzarini *et al*, 2007	Jan 2002-Aug 2003	404 cultures from a repository of mycobacterial cultures were randomly selected from 1202 positive cultures from patients resident in Rio de Janeiro, Brazil.	2/404 (0.5%)
Brazil (Sao Paulo)	Telles *et al*, 2005	2000-2002	420 isolates were collected from new and retreatment cases examined at 4 health care centers located in Zona Norta, a urban community of Sao Paulo	4/420 (0.95%)
French Guiana and French West Indies	Brudey *et al*, 2006	1994-2003	718 Isolates were collected from culture positive cases resident in French Guiana, Guadeloupe and Martinique; HIV=25%	6/718 (0.8%) (only found in French Guiana)
Peru	Escalante *et al*, 1998	1995-1996	29 isolates cultured from PTB patients and suspected to be infected with drug resistnat organisms	2/29 (6.9%)
Paraguay	Candia *et al*, 2007	2003	Cross-sectional study; 220 strains isolated from new and retreatment PTB pts during the national survey of drug resistance carried out in Paraguay; represents 77% of strains composing the drug resistance survey.	1/220 (0.5%)
Venezuela	Artistimuño *et al*, 2006	1998-1999	556 isolates were collected from new and retreatment smear-positive PTB patients enrolled in a drug survey in 23 provinces in	0

			Venezuela	
Venezuela	Artistimuño *et al*, 2007	2000-2004	36 MDR isolates were obtained from smear-positive TB cases diagnosed by the Venezuelan National TB Program	2 (6%)
Colombia (Buenaventura0	Laserson *et al*, 2000	1997-1998	24 and 40 isolates from treatment failure patients identified through passive and active surveillance respectively at 9 local health posts and 1 central hospital and 73 isolates from new TB cases were analysed	11/111 (10%)
Peru, Argentina, Brazil, Paraguay, Chile, Colombia, Ecuador	Ritacco *et al*, 2008	1997-2003	1202 isolates were collected from smear positive PTB cases from 7 countries of South America as part of surveys for evaluation of drug resistant or patients with laboratory-confirmed diagnosis of MDR-TB	19/1202 (1.6%)
Argentina and Colombia	Olano *et al*, 2007	2003-2004	92 and 25 isolates were collected from Argentina and Colombia respectively to test the association between mutations in specific DNA repair genes and MDR.	2/117 (1.7%)
OTHER				
Russia, Europe and East Asia	Mokrousov *et al*, 2006	2002-2005	866 isolates were collected representing Russia, China, Vietnam and Bulgaria where the Beijing genotype is endemic. The collections from first 3 countries were convenience samples whereas those from Bulgaria were randomly selected	408/866 (47%)
Texas and Mexico	Quitugua *et al*, 2002	1992-2000	Total of 919 drug resistant isolates from 606 and 313 patients residing in Texas and Mexico respectively were collected	145/919 (16%)

Table 2

COUNTRY	PUBLICATION	Incidence of all TB cases – WHO, 2006 (rate/100 000)	HIV prevalence in incident TB cases (%)	Association between Beijing and drug drug resistance	COMMENTS
AFRICA					
Kenya (Nairobi)	Githui *et al*, 2003	384	52	no	3% of Beijing strains were drug resistant of which all were MDR versus 42% of non-Beijing strains were drug resistant of which 40% had MDR; CAS family associated with drug resistance in this study
Malawi (Karonga District)	Glynn *et al*, 2005	377	70	no	All Beijing strains were drug sensitive
Senegal	Niang *et al*, 1999	270	3	not available	No difference could be observed between the profiles of sensitive and drug resistant TB strains
Sierra Leone	Homolka *et al*, 2008	517	5	yes	All 4 isolates with Beijing genotype were drug resistant (3=MDR)
Tanzania	Kibiki *et al*, 2007	312	18	no	All Beijing isolates were drug sensitive
Uganda (Kampala)	Asiimwe *et al*, 2008	355	16	no	3 out of 4 is drug sensitive
South Africa (Cape Town)	Van Rie *et al*, 2001	940	44	yes	All drug resistant Beijing strains were MDR whereas only 50% of drug resistant non-Beijing strains were MDR
South Africa (Cape Town)	Van Rie *et al*, 1999	940	44	yes	17/21 MDR cases identified were infected with a Beijing genotype
South Africa (Western and Eastern Cape)	Strauss *et al*, 2007			yes	nsSNPs at codon 516 (GAC-GTC) was associated with atypical Beijing strains from the Western Cape and nsSNPs at codon 533 was associated with typical Beijing strains from the Eastern

					Cape
South Africa (BOKS)	Streicher *et al*, 2004			yes	Beijing and LCC were most prevalent drug resistant isolates
Zimbabwe	Easterbrook *et al*, 2004	557	43	no	Only 4 Beijing isolates were found in study setting; all were drug sensitive
CENTRAL AMERICA					
Cuba	Diaz *et al*, 1998	9	0	yes	11/20 drug resistant cases were infected with a Beijing genotype
EUROPE					
Austria	Stauffer *et al*, 2000	13	5	no	15/17 MDR isolates had unique spoligotype patterns; 2/17 were Beijing
Azerbaijan	Pfyffer *et al*, 2001	77	1	yes	63% of drug resistant cases were infected with Beijing strain of which 68% were MDR versus 20 and 46% respectively for cases infected with non-Beijing strain; MDR was associated with the Beijing genotype
Germany	Diel *et al*, 2002	6	2	no	Majority of isolates analysed (401/423) were drug sensitive; remaining drug resistant isolates were mostly from unclustered strains
Germany	Hilleman *et al*, 2005			yes	-
Germany	Kubica *et al*, 2004	6	2	yes	Found a remarkably high rate of Beijing genotype (38.8%) among MDR-TB strains from Germany, with an annual increase during the study period
Estonia	Kruuner *et al*, 2001	39	18	yes	87.5% of all MDR isolates, 67.2% of all isolates with any drug resistance but only 12% of drug sensitive cases were infected with the Beijing genotype
Finland	Puustinen *et al*, 2003	5	1	no	2/4 were drug sensitive; 1 was INH mono-resistant; 1 was resistant to INH, Rif, Strep and EMB
Finland	Martilla *et al*, 2007			no	Only 8.8% of drug resistant isolates were due to Beijing strains

United-Kingdom (Scotland)	Fang *et al*, 1999	15	4	yes	2/2 Beijing strains were drug resistant
Italy	Lari *et al*, 2007	7	7	no	Only one patient (from former Soviet Union) was MDR with no substantial resistance to any other anti-TB drugs was observed among the Beijing strains
Italy (Tuscany)	Rindi *et al*, 2008			no	proportion of isolates resistant to any 1st-line drugs was not different from that of non-Beijing strains isolated in the same study period
Latvia	Tracevska *et al*, 2003	57	4	yes	Beijing genotype not particularly common among drug sensitive isolates in Latvia
Poland	Sajduda *et al*, 2004	25	0	yes	57% of Beijing strains were MDR
Portugal	Perdigão *et al*, 2008	32	14	no	Strains belonging to family Lisboa are responsible for the majority of MDR-TB cases. 53% of these were XDR
Russia	Mokrousov *et al*, 2003	107	4	yes	-
Russia (Archangel Oblast)	Toungoussova *et al*, 2002			yes	41/53 is drug resistant of which 23 is MDR; resistance to Rif and SM significantly associated with Beijing strains but not for EMB and INH
Russia (Archangel prison)	Toungoussova *et al*, 2003			yes	85% of drug resistant strains were part of the Beijing strain family; Beijing strains were more often MDR than were the non-Beijing but difference not significant (did not find a higher rate of MDR phenotype in the Beijing isolates
Russia (Archangel Oblast)	Toungoussova *et al*, 2005			yes	80% of MDR isolates were infected with a Beijing strain; spread of Beijing in Archangel is associated with high rates of drug R to first-line drugs. However, no association was found for second-line drugs; resistance to capreomycin less common in Beijing strains (33% vs 71%)
Russia (Tula Region)	Ignatova *et al*, 2006			yes	76.3% of drug resistant Beijing strains were MDR

Russia (Samara region)	Drobniewski *et al*, 2005			yes	Drug resistant were significantly higher, particularly for MDR-TB among those infected with the Beijing genotype; Resistance to INH, Rif, MDR, SM and EMB 2-fold higher in Beijing versus non-Beijing infected patients
Russia (Samara region)	Balabonova *et al*, 2006			yes	Drug resistance, including MDR, was twice higher among the Beijing strain family
Russia (St. Petersburg)	Narvskaya *et al*, 2002			yes	No evidence of transmission for the non-Beijing MDR isolates was found
Russia	Mokrousov *et al*, 2002			yes	Majority of drug resistant strains from newly diagnosed patients belonged to the Beijing genotype
Russia (Serpukhov prison)	Shemyakin *et al*, 2004			yes	Predominance together with the LAM family among drug resistant isolates derived from prison inmates
Russia (Archangel Oblast)	Toungoussova *et al*, 2004			yes	All drug resistant Beijing strains were MDR whereas only 60% of drug R non-Beijing strains were MDR
Russia (Ukraine)	Nikolayevskyy *et al*, 2006	107	4	yes	Drug sensitivity more common in non-Beijing strains; prevalence of mutations associated with all types of resistance was significantly higher among Beijing strains with strongest association with MDR
Russia	Mokrousov *et al*, 2002			yes	68% of MDR isolates were infected with a Beijing strain; they acquired the mutation TCG521-TTG more readily within the *rpo*B gene
Russia (Kaliningrad)	Mokrousov *et al*, 2008			yes	Beijing strains made 1.6% of all drug sensitive isolates and 28 out of 30 MDR cases; remarkably strong association of MDR as well as of any drug resistance was observed for the Beijing genotype
Russia	Narvskaia *et al*, 2002			yes	Beijing strains were shown to cause more severe unarrested course of drug resistant fibro-cavernous TB
Russia (Ural region)	Kovalev *et al*, 2005			yes	19 out of the 23 MDR isolates belong to strains with the Beijing genotype; Beijing drug resistant isolates were twice as frequent

					as in other groups
Russia (Moscow)	Lipin *et al*, 2007			yes	42.4 and 45.6% of isolates were infected with Beijing and LAM genotypes respectively
Sweden	Ghebremichael *et al*, 2008	6	3	no	Ten of the 46 MDR isolates and the XDR isolate were of the Beijing lineage
Sweden	Brudey *et al*, 2004			no	3/11 Beijing strains were resistant; none were MDR
Spain	Samper *et al*, 2005	30	9	no	Non-Beijing strains made out 94% of all MDR isolates studied. The "B"-strain was the most frequently isolated MDR strain in this study
Spain (Gran Canaria Island)	Caminero *et al*, 2001			no	All Beijing strains were drug senisitve of which 2 developed MDR during treatment
Spain (Elche)	Garcia-Pachon *et al*, 2007			no	Only 1 Beijing strain was found during study period
Spain (Madrid)	De Viedma *et al*, 2006			no	All isolates except for 1 were drug sensitive
Turkey	Durmaz *et al*, 2007	29	-	no	Despite high MDR rates (59%) and 14% of strains were INH mono- resistance, only 1 Beijing strain were found
Turkey (Istanbul)	Koksalan *et al*, 2006			yes	Significant association between Rif resistance and Beijing strains
FAR-EAST ASIA					
China (Beijing)	Mokrousov *et al*, 2006	99	0.3	yes	Rif, PZA, MDR and HRZ patterns were significantly associated with the old Beijing strains; SM weakly associated with new Beijing strains
China (Beijing)	Wei-Wei *et al*, 2007			yes	No significant difference between modern and ancient Beijing strains regarding prevalence of *rpo*B and *kat*G mutation however the proportion of *rpo*B mutants were slightly higher in the modern strains while *kat*G S315T were higher in ancient strains

China (Beijing, Shanghai, Tianjin)	Li *et al*, 2005			yes	Beijing strains significantly more resistant to anti-TB drugs using X^2 test but not with multivariate analysis
China (Hong Kong)	Chan *et al*, 2001			no	No significant association between Beijing and individual first-line anti-TB drugs although there was a tendency for Beijing to be associated with INH sensitive strains
China (Tianjin)	Chai *et al*, 2007			no	-
China (Tibet)	Shi *et al*, 2007			no	Beijing strains were not significantly associated with drug resistance to all of the 4 drugs
Indonesia (Jakarta)	Van Crevel *et al*, 2001	234	1	no	-
Indonesia (West Java and Timor)	Parwati *et al*, 2008			no	Drug resistance patterns of Beijing and non-Beijing isolates were not significantly different; no association found between Beijing strains and drug resistance
Japan	Millet *et al*, 2007	22	0	not available	Neither Beijing nor non-Beijing strains were associated with MDR; total drug resistance = 16%
Korea	Park *et al*, 2005	88	1	yes	Proportion of Beijing strains were significantly higher among MDR isolates than among drug sensitive strains
Singapore	Sun *et al*, 2007	26	3	yes	Rif, SM and MDR isolates more common in Beijing genotype; INH and EMB resistance not associated with any genotype
Thailand (Bangkok)	Prodinger *et al*, 2001	142	11	no	Overall drug resistance to single drugs and MDR were not associated with Beijing
Taiwan	Jou *et al*, 2005	-	-	yes	Prevalence of drug resistance were higher among Beijing strains; significant association between resistance to EMB and INH and Beijing
Vietnam (Hanoi and Ho Chi Minh City)	Tuyen *et al*, 2000	173	5	no	No correlation between the presence of resistance to antimicrobial drug and the banding patterns could be found, however drug resistance were more prevalent in the south

					(where majority of cases are due to Beijing genotype) than north of Vietnam
Vietnam (Hanoi and Ho Chi Minh City)	Anh *et al*, 2000			yes	Beijing were associated with INH and SM resistance
Vietnam (Ho Chi Minh City)	Caws *et al*, 2006			yes	Strongly associated with MDR
China, Japan, Korea and Taiwan	Qian *et al*, 2002	-	-	yes	Mutations Ser531 in the *rpo*B gene were found equally in both Beijing and non-Beijing strains however, Beijing had higher mutation rates for Asp516 and His526
Indonesia, Malaysia, the Philippines, Vietnam	Ang *et al*, 2008	-	-	yes	One third of the MDR isolates belonged to the Beijing strain family
MIDDLE-EAST-CENTRAL ASIA					
Bangladesh	Shamputa *et al*, 2004	225	0	yes	Drug resistance observed more frequently among strains with the Beijing genotype
India	Singh *et al*, 2004	168	1	yes	8 out of 9 strains were drug resistant
India (7 regions)	Singh *et al*, 2007			no	14/21 were drug sensitive
India	Suresh *et al*, 2006			yes/no	Beijing strains is contributing to the MDR epidemic along with EAI3 and CAS1-Delhi strains (same results were obtained with drug S isolates); most predominant clade among MDR isolates is CAS
India (Mumbai)	Almeida *et al*, 2005			yes	Beijing genotype accounted for nearly one third of the MDR isolates. None of the Beijing strains were drug sensitive
India (Bombay)	Mistry *et al*, 2002			yes	-
Iran and Afghanistan	Ramazanzadeh *et al*, 2006	22 and 161	2 and 0	yes	Drug resistance was higher in PGG1 than PGG 2 and 3; All Beijing strains were associated with drug resistance
Iran	Doroudchi *et al*, 2000	22	2	yes	4/5 Beijing strains were drug resistant

Kazakhstan	Kubica *et al*, 2005	130	1	yes	Strong association of Beijing genotype with drug resistance as seen by the significantly lower Beijing rate in the drug sensitive control group
Kazakhstan	Hillemann *et al*, 2005			yes	64% of the MDR strains, 64% of the INH resistant/Rif resistant strains and 40% of drug sensitive strains were characterised as Beijing genotype
Sri Lanka	Rajapaksa *et al*, 2008	60	0	no	12 out of 14 isolates were pan-susceptible
Pakistan	Hasan *et al*, 2006	181	0.3	yes	Beijing strains were highly associated with MDR
Saudi Arabia	Al-Hajoj *et al*, 2007	44	-	yes	Almost a quarter of Beijing strains showed drug resistance and among these more than half (64%) were MDR
Uzbekistan and Turkmenistan	Cox *et al*, 2005	121 and 65	1 and not available	yes	38% of the fully susceptible isolates belonged to the Beijing family while 75% of MDR-TB pts were infected with Beijing strains; stronger association with EMB resistance, followed by Rif and PZA
Uzbekistan and Turkmenistan	Cox *et al*, 2007			yes	Amplification of resistance during treatment associated with Beijing strains: 10 new and 9 retreatment cases' drug profiles changed during treatment of which 9 and 6 were Beijing strains respectively); in cases of pre-existing poly-resistance, Beijing had higher capacity to develop further drug resistance than non-Beijing strains (33% versus 0%)
NORTH AMERICA					
USA (NYC)	Frieden *et al*, 1996	4	11	yes	resistance to kanamycin found in 92% of isolates; resistance to ETH and PZA was common (the latter more common later in the outbreak); Acquisition of resistance to cycloserine less common than to capreomycin and fluoroquinolones
USA (NYC)	Munsiff *et al*, 2003			yes	-
USA (New-York City)	Moss *et al*, 1997			yes	Beijing genotype is the most drug resistant cluster identified in NYC and made up 22% of NYC MDR-TB cases in 1992

USA	Agerton *et al*, 1999			yes	All isolates were MDR
OCEANIA					
Australia (Papua New Guinea)	Gilpin *et al*, 2008	250	4	yes	15/60 isolates were MDR of which all belonged to the Beijing genotype
SOUTH AMERICA					
Brazil (Sao Paulo)	Telles *et al*, 2005	50	12	no	There was no single predominant cluster pattern among resistant strains; 2 out of 4 Beijing strains were MDR
Peru	Escalante *et al*, 1998	162	12	no	-
Venezuela	Artistimuño *et al*, 2007	41	6	no	-
Colombia	Laserson *et al*, 2000	45	2	no	Only 3/11 Beijing isolates were MDR
Peru, Argentina, Brazil, Paraguay, Chile, Colombia, Ecuador	Ritacco *et al*, 2008	-	-	no	No association was found between carrying a strain with the Beijing genotype and having drug resistance or MDR; however 8 out of 10 is MDR
Argentina and Colombia	Olano *et al*, 2007	-	-	no	-
OTHER					
Russia, Europe and East Asia	Mokrousov *et al*, 2006	-	-	yes	Current transmission of MDR in Russia is greatly influenced by the ongoing dissemination of Beijing strains
Texas and Mexico	Quitugua *et al*, 2002	-	-	no	30% of drug resistant Beijing isolates were MDR