Application of spoligotyping in the understanding of the dynamics of *Mycobacterium tuberculosis* strains in high incidence communities

Elizabeth Maria Streicher

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

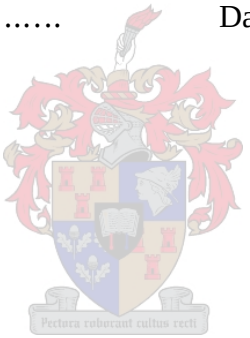
Promoters: Prof T.C. Victor and Prof R.M. Warren

March 2007

Declaration

I, the undersigned, hereby declare that the work contained in this dissertation 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 ……..……..…. Date ……...…….……...



Summary

Tuberculosis (TB) is a global health problem and demands rigorous control management efforts. A dramatic increase in the acquisition and spread of drug resistant TB globally has been observed in recent years. A grim picture has emerged for the control program with the discovery of extreme drug-resistant TB, which is virtually untreatable and is of immense concern for the future of TB control.

In the last decade strain-specific genetic markers have been identified to examine the molecular epidemiology and spread of TB, including IS*6110* DNA-fingerprinting and spoligotyping. Although spoligotyping has less discriminatory power than the gold standard, IS*6110* DNA-fingerprinting, it is simpler, faster and less expensive, as it is PCR-based. Spoligotyping has been applied to enhance our understanding of the dynamics of drug susceptible and drug resistant strains of *Mycobacterium tuberculosis* in high incidence communities, by studying 3 aspects of the TB epidemic: molecular epidemiology of drug resistant TB, recurrent TB and the evolution of *M. tuberculosis*.

By using spoligotyping and other genotypic and phenotypic analysis of drug-resistant *M. tuberculosis* isolates from the Western Cape Province of South Africa showed that drug resistance is widespread and recently transmitted. An emerging drug resistant *M. tuberculosis* outbreak has been identified, termed DRF150, which has specific genotypic characteristics and is resistant to 5 first-line drugs in 45% of the cases. Inappropriate chemotherapy; poor adherence to treatment and prolonged periods of infectiousness due

to the delay in susceptibility testing has led to the development and spread of this drug resistant genotype.

The study demonstrates the ability of the spoligotyping technique to accurately determine the pathogenic mechanism of recurrent disease by spoligotyping, making it useful in large-scale intervention studies. Application of spoligotyping and a newly developed PCR-method showed that the occurrence of multiple infections was higher than what was previously assumed and also more frequent in retreatment cases than new cases. These findings have important implications for the understanding of protective immunity, and the development and testing of new vaccines and drugs.

Various different molecular markers including spoligotyping has been used to reconstruct the evolutionary history of isolates with less than 6 copies of IS*6110* element (termed Low Copy Clade (LCC)), which were previously poor defined. It was also shown that LCC is widely disseminated and play an important role in the global tuberculosis epidemic. Reconstruction of the evolutionary relationship of *M. tuberculosis* Principal Genetic Group 2 strains, identified previously unknown genetic relationships between strain families and laid the foundation to establish correlations between genotype and phenotype.

Spoligotyping signatures, created by evolution of the Direct Repeat region in *M. tuberculosis*, were identified, which will enable the analysis of the strain population structure in different settings and will also enable the rapid identification of strain

families that acquire drug-resistance or escape protective immunity in drug and vaccine trials.

This study contributed to our understanding of the molecular epidemiology of drug resistant TB, recurrent TB and the evolution of *M. tuberculosis* in high incidence communities.

Opsomming

Tuberkulose (TB) is ‘n wereldwye gesondheidsprobleem en benodig streng kontrole en bestuur. ’n Drastiese verhoging in die ontstaan en verspreiding van middelweerstandige TB is wêreldwyd waargeneem oor die afgelope paar jaar. Die identifikasie van uiters weerstandige TB, wat bykans onmoontlik is om te behandel, maak die toekoms van TB kontrole somber.

In die afgelope dekade is TB-stam spesifieke genetiese merkers ontwikkel om molekulêre epidemiologie en verspreiding van TB te ondersoek en sluit IS*6110* DNS-vingerafdrukke and spoligotipering in. Alhoewel spoligotipering nie so goed tussen stamme kan diskrimineer soos die goue standaard, DNS-vingerafdrukke nie, is dit eenvoudiger, vinniger en goedkoper, want dit is PKR gebaseer. Spoligotipering is aangewend om die dinamika van middelsensitiewe en middelweerstandige stamme van *Mycobacterium tuberculosis* in hoë insidensie gemeenskappe beter te verstaan, deur 3 aspekte van die TB epidemie te ondersoek: Molekulêre epidemiologie van middelweerstandige TB, TB-herinfeksie en evolusie van *M. Tuberculosis*.

Spoligotipering en ander genotipiese en fenotipiese analises van middelweerstandige *M. tuberculosis* isolate uit die Wes-Kaap Provinsie van Suid-Afrika wys dat middelweerstandigheid wyd verspreid voorkom. ’n Opkomende middelweerstandige TB uitbraak is geidentifiseer, en word nou DRF150 genoem, wat spesifieke genotipiese karakteristieke het en 45% van die gevalle is weerstandig teen al vyf eerste-linie middels. Onvanpaste behandeling, swak volhouding van behandeling en verlengde aansteeklikheid

as gevolg van stadige weerstandigheids toetsing het gelei tot die onstaan en verspreiding van hierdie weerstandige genotipe.

Daar is gewys dat spoligotipering akuraat die patogeniese meganisme van herhaalde-TB kan klassifiseer, wat is voordelig is vir grootskaalse studies.

Die aanwending van spoligotipering en ’n nuut ontwikkelde PKR-metode het gewys dat die voorkoms van veelvuldige infeksies hoër is as wat voorheen aangeneem is en ook meer dikwels in herhaalde gevalle as nuwe TB gevalle voorkom. Hierde bevindings het belangrike gevolge vir die verstaan van beskermende immuniteit en ontwikkeling en toetse van nuwe vaksines en anti-TB middels.

Verskeie verskillende merkers, insluitend spoligotipering is gebruik om die evolusie van isolate met minder as ses kopieë van IS*6110* element (Lae Kopie Groep (LKG)) te herstruktureer, wat voorheen onbekend was. Die studie wys ook dat LKG wyd verspreid is en ’n belangrike rol in die TB epidemie wêreldwyd speel. Die evolusie geskiedenis van Hoof Genetiese Groep 2 waarvan LKG deel is, is bepaal. Hierdie voorheen onbekende genetiese verwantskap tussen stamme lê die fondamente om die korrelasie tussen genotipiese en fenotipiese eienskappe te ondersoek.

Spoligotiperingseine, wat deur evolusie in die direkte herhaalde lokus in *M. tuberculosis* ontstaan het, is geidentifiseer en sal die analise van stampopulasie struktuur in

verskillende gemeenskappe moontlik maak. Dit stel ons instaat om stamme vinniger te identifiseer wat middelweerstandig is of immuniteit kan ontsnap in vaksine ontwikkeling.

Hierdie studie dra by tot ons kennis en verstaan van molekulêre epidemiologie van middelweerstadige TB, TB-herinfeksie en die evolusie van *M. tuberculosis* in hoë insidensie gemeenskappe.

"Knowing is not enough; we must apply.
Willing is not enough; we must do."

Bruce Lee

Table of contents

Acknowledgements

Thank you to everyone who contributed in any way to this study.

Rob and Tommie, thank you for all the help, support knowledge, and patience during the past 6 years.

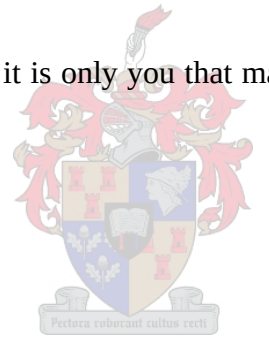
All the students and staff at Department of Medical Biochemistry, especially Lab 453, and all our collaborators and funders, thank you for all the friendliness, enthusiasm and help. This has been another demonstration of Brilliant Teamwork.

Thank you to the communities and TB patients and the Health Care Workers in the Western Cape Province.

To all my friends and family, thank you for all the support, patience and prayers.

Baie dankie Ma en Pa, Jan-Willem, Leensie en Oupa Jan vir al die bystand, gebede, motivering en liefde oor die jare.

Alle eer aan God drie-enig! Lord, it is only you that made everything happen. Ek dank u vir die genade en liefde.



Chapter 1

General Introduction

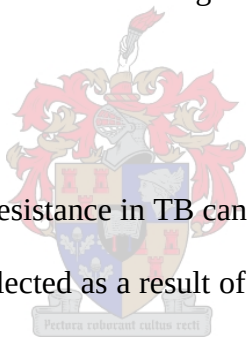
Incidence of Tuberculosis globally and in South Africa

Tuberculosis (TB) remains one of the world's most serious health problems. Approximately one-third of the world's population is infected with *Mycobacterium tuberculosis* and it is the leading cause of death from a single infectious pathogen. According to the World Health Organization Global Tuberculosis Control report 2006, there were 8.9 million new TB cases (140/100 000) and approximately 1.7 million (27/100000) TB deaths in 2004, globally. Treatment success was 82% in 2003, approaching the 85% target set by the WHO. In South Africa the incidence of all cases in 2004 was estimated at an alarming 718/100 000 and the deaths mortality rate was 135/100 000 (51). One of the main reasons for the increasing global burden of disease are the inefficient TB control programs, with low cure rates, because of inadequate and interrupted treatment. The success of TB control programs is also threatened by the emergence of drug-resistant strains of *M. tuberculosis*, especially multidrug-resistant (MDR) strains and the rising HIV epidemic (43). MDR-TB is defined by the WHO as resistance to both Isoniazid (INH) and Rifampin (RIF), two of the primary drugs in the treatment of TB (50). Rapid detection of *M. tuberculosis*, adequate therapy, contact tracing and a proper vaccine are important issues in the control of tuberculosis.

Drug resistant tuberculosis

In order to control the drug resistance epidemic it has become necessary to gain insight into how *M. tuberculosis* develops drug resistance. This knowledge will help us to understand how to prevent the occurrence of drug resistance as well as identifying pathways of genes associated with drug resistance of new drugs. Drug resistance occurs

as a selection process, where inappropriate treatment lead to the selection of natural resistant bacteria to multiply (42). Approximately 12 genes from the genome of *M. tuberculosis* are identified that are linked to antibiotic resistance in TB (19). Mutations in the *katG* gene are associated with resistance to Isoniazid. The *katG* gene encodes for catalase peroxidase, which is needed for the activation of Isoniazid. Resistance to Isoniazid can also be due to mutations in the *inhA*, *kasA*, *oxyR* and *ahpC* genes (19,30). Resistance to Rifampin is caused by mutations in the *rpoB* gene, coding for the ß-subunit of ribonucleic acid (RNA) polymerase (17,31). The *pncA* gene is associated with Pyrazinamide resistance (19,33,34). Resistance for Streptomycin is due to mutations in *rrs* and *rpsL* genes and mutations in the *embB* gene are associated with resistance to Ethambutol. (19,30).



The development of clinical drug resistance in TB can be classified as acquired resistance when drug resistant mutants are selected as a result of ineffective treatment or as primary resistance when a patient is infected with a resistant strain (35).

There is much debate about the relative contribution of acquired and primary resistance to the burden of drug resistant TB in different communities. This controversy focus on whether MDR strains are transmissible or whether the mutations that confer drug resistance also impair the reproductive function of the organism (fitness of the strain). Evidence that MDR strains do have the potential for transmission comes from a series of MDR-TB outbreaks that have been reported over the past decade, in hospitals (3,7,8,11), amongst health care workers (1,18,28) and in prisons (40). Application of molecular

epidemiological methods was central to the identification and description of all these outbreaks.

The most extensive MDR-TB outbreak reported to date occurred in 267 patients from New York, who were infected a Beijing/W strain (12). This cluster of cases included drug resistant isolates that were resistant to all first-line anti-TB drugs. The authors speculate that the delay in diagnosis and administering appropriate therapy resulted in prolonging infectiousness and placed healthcare workers and other hospital residents (or contacts) at risk of infection for nosocomial infection. This difficult to treat strain has subsequently disseminated to other US cities and Paris and the authors showed by using molecular methods, how this initially fully drug susceptible strain clonally expanded to result in a MDR phenotype by sequential acquisition of resistance conferring mutations in several genes (3). Since then, the drug resistant Beijing/W genotype has been the focus of extensive investigations and Beijing drug resistant and susceptible genotypes have been found to be widely spread throughout the world (13), including in South Africa (42) and Russia (25). The Beijing family of strains can be easily identified by a specific spoligotype pattern characterized by the presence of spoligotype spacers 35-43 (2) and characteristic multi-banded IS*6110* restriction fragment-length polymorphism (RFLP) patterns. Although the Beijing strain family is a prominent family in the Western Cape province, a recent outbreak of a specific cluster of the Beijing family in Cape Town was the focus of attention. This cluster of strains all had exactly identical IS*6110*-RFLP patterns and contained a mutation in the promoter of the *inhA* gene, conferring to isoniazid resistance. In addition, 42% of the strains had a mutation in *rpoB* gene,

conferring to Rifampin resistance (20). Although these data led many to propose that Beijing/W strains behaved differently from other strains, recent work suggests that MDR outbreaks are not limited to the Beijing/W genotype. Smaller outbreaks involving other MDR-TB genotypes have been reported in other settings such as the Czech Republic, Portugal and Norway (24,29). However, since much of the MDR burden falls in developing countries in which routine surveillance does not usually include molecular fingerprinting, little is known about the characteristics of circulating drug resistant strains in much of the world. It is therefore possible that there are other MDR strains, as widespread as Beijing/W, which have not been recognized and reported as such.

Molecular epidemiology of tuberculosis

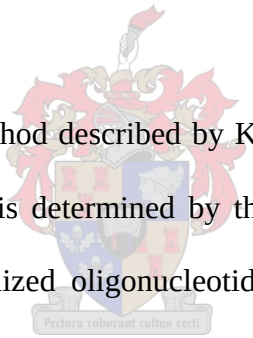
In the last decade strain-specific genetic markers have been identified to examine the molecular epidemiology and spread of drug sensitive and drug resistant TB. These markers have been used as DNA fingerprinting methods such as IS*6110*-RFLP and spoligotyping to genotype strains (43). These studies have shown that there are a wide variety of different strains of *M. tuberculosis* circulating in communities all over the world. Molecular typing methods can therefore be used to identify outbreaks and to facilitate contact tracing. Knowledge gained from this field of molecular epidemiology iss crucial in the understanding of transmission and thus the prevention of transmission of TB strains (26). Typing methods needs to have a high discriminatory power and needs to be rapid, reproducible, easy to perform, inexpensive and be able to use directly on clinical material, but currently, no method meets all these criteria. Most methods used for fingerprinting are based on the IS*6110* insertion sequence and the standardized typing

method for tuberculosis, Restriction Length Polymorphism (RFLP), is based on this IS element and was described in the early 1990's (41). Although this method is considered the gold standard of molecular typing of *M. tuberculosis* because of its high discriminatory power and reproducibility (23), IS*6110*-RFLP requires large amounts of DNA and therefore needs to be cultured for weeks, which makes it time-consuming and requires viable organisms (26). IS*6110*-RFLP is also technically demanding, expensive and needs sophisticated computer software to analyse results (15). Other genotyping methods that have been developed are based on repeat sequences in the genome of *M. tuberculosis* and includes Mycobacterial Interspersed Repetitive Units (MIRU), Polymorphic GC-rich repetitive sequence (PGRS) and the Major Polymorphic tandem Repeat (MPTR) (22,43,44).

The DR locus and Spoligotyping

Spoligotyping is a PCR based technique developed for identification and differentiation of strains of *M. tuberculosis*. Hermans et al. first described the DR region based on *M. bovis* BCG (16) and Groenen et al. suggested this locus for epidemiological studies of the *M. tuberculosis* complex (14). The DR locus consists of multiple conserved 36bp directly repeated sequences (DRs) interspersed by non-repetitive DNA spacers ranging from 35 to 41 bp in length (16,21). To date 94 different spacers have been identified in the *M. tuberculosis* complex, but the 43 spacers originally described are used routinely for the classification of strains. One DR and its neighbouring spacer are termed a direct variable repeat (DVR). It was observed that the order of the spacers is conserved between clinical isolates, although the DR region has been shown to be polymorphic in different clinical

isolates of *M. tuberculosis* (21). Polymorphisms arise from homologous recombination between neighbouring or distant DRs, strand slippage, duplications of DRs, IS*6110* insertions, homologous recombination between adjacent IS*6110* elements and SNPs within the spacers. These polymorphisms make this locus very useful to distinguish *M. tuberculosis* isolates. The function of the DR locus is presently unknown, but the DRs and spacers are apparently present in all isolates and are well conserved among strains, thus may suggest a biological function of the region. Similar motifs have been found in other bacterial genera, but no significant homology with *M. tuberculosis* complex has been identified. The DR region has been identified as a hotspot for integration of the IS*6110*.



According to the standardised method described by Kamerbeek et al. (21), the presence or absence of any of 43 spacers is determined by the hybridisation of PCR-amplified spacer DNA to a set of immobilized oligonucleotides, complementary to the unique spacer DNA sequences. These spacer sequences are derived from the sequences of *M. tuberculosis* H37Rv and *M. bovis* BCG. Amplification and labelling of all the unique spacer DNA sequences within the DR region of a given strain is done by PCR. The primers (of which one is biotinyilated) are complement to the DR sequence and allow the amplification of the spacers between the target DR's. Different length of PCR-product will be obtained as the DR primers can initiate polymeration at any DR, irrespective of the number of DVRs between the primers. The amplified DNA is perpendicular hybridised to spacer oligonucleotides that are covalently linked to a membrane in parallel. Hybridization is detected by chemiluminesence, because one of the primers is

biotinylated and detected through a streptavidin-peroxidase conjugate and a substrate. Strains can be differentiated due to the variation in hybridisation patterns obtained, based on the presence or absence of the DVRs.

An international spoligotyping database has been established by the Pasteur institute of Guadeloupe. The first version described spoligotypes from the Caribbean (37), and in the second version the database was updated to 799 shared types from 3319 isolates, originating mainly from Europe and USA (38). SpolDB3 described spoligotypes from all over the world (11708 isolates from 90 countries), but was still overrepresented with isolates from Europe and North America (65.5% of database). Africa was underrepresented, as most of the isolates were from Zimbabwe (9). The latest version, SpolDB4, consist of 1939 spoligotype patterns from approximately 40 000 clinical isolates originating from 141 countries and was published in June 2006 (5). Our group contributed to this database by submitting spoligotypes from 3207 isolates, from South Africa and also from different countries in Africa (Cameroon, Egypt, Morocco, Tanzania, Uganda and Zimbabwe), thus helping to provide a more representative picture of the world wide description of the *M. tuberculosis* diversity.

IS*6110* RFLP fingerprinting is the reference for standard typing of *M. tuberculosis* strains and is considered to have better discriminatory power than spoligotyping. Despite this limitation spoligotyping remains useful in different applications. The method can be used to monitor specific strains in a geographical area and between countries (5,9,10). It can also be useful in identification of laboratory cross contamination and laboratory error (27)

and the identification and confirmation of dual strain infections (Chapter 4,(49)). It has also been shown to be able to distinguish between relapse or reinfection in the case of a second episode of disease (Chapter 2;(48)). Other advantages of spoligotyping includes the ability to use it on archived material and the identification of different species within the *M. tuberculosis* complex. Spoligotyping is better suited to high-throughput screening of patients in large intervention studies. Because this is a PCR based technique there is no need for culturing of isolates for DNA isolation and thus the overall cost, time and exposure of laboratory workers to the pathogen is greatly reduced. The chance of laboratory cross contamination is also reduced as sub-culturing is not required.

Hypothesis, aim and structure of this study

We hypothesized that the application of spoligotyping will help to improve our understanding of the disease dynamics of *M. tuberculosis* in high incidence communities. The general aim of this dissertation is to apply spoligotyping and other molecular techniques to study 3 main aspects of the TB epidemic: molecular biology of drug resistant TB, recurrent TB and the evolution of *M. tuberculosis*. Each Chapter is structured according to the instructions of the journals in which the articles were or will be published.

To identify characteristics and monitor the spread of drug resistant strains in the Western Cape we established a longitudinal reference database of phenotypic (drug resistance) and genotypic (spoligotyping) data, of all drug resistant isolates collected from 72 clinics in the Western Cape Province. Analysis of this database from January 2001 to February

2002 is described in Chapter 2. We showed that the drug resistance in the region is mainly due to transmission of resistant strains and that four main strain families are primarily responsible for the drug resistant epidemic. This database, together with additional molecular markers has allowed the identification of an outbreak of an emerging drug resistant strain, infecting 64 pulmonary TB cases (Chapter 3). This previously undetected genotype (now designated DRF150) is characterized by 5 IS*6110* insertions, a specific spoligotype and high levels of resistance to the first line TB medications, Isoniazid, Streptomycin and Rifampin. In 45% of the cases it is also resistant to two other first line drugs, Ethambutol and Pyrazinamide.

The efficacy of treatment within different settings and the identification of factors influencing disease dynamics can be determined, if the mechanism of recurrence is established. Recurrence can be due to relapse (the endogenous reactivation of the initial strain) or reinfection (when exogenous infection with a different strain cause a subsequent episode of disease). In chapter 4 we aimed to establish spoligotyping as a technique to identify the mechanism of recurrence, as spoligotyping is better suited in high throughput screening of patients in large intervention studies and does not require prior culture.

It is generally accepted that TB results from a single infection with a single *M. tuberculosis* strain. Using IS*6110*-RFLP, molecular epidemiological studies have shown the presence of a single strain in most cultures collected from TB patients, suggesting that disease is caused by a single strain (infection). It has been shown that multiple strain

infections do occur (4,6,32). However, multiple infections are rarely seen in fingerprinting databases and therefore their significance remains unknown. In chapter 5 we aimed to determine the extent of multiple *M. tuberculosis* infection in sputum specimens collected from new and retreatment TB cases. We developed a PCR method based on comparative genomic data to amplify Beijing and non-Beijing strains and showed that 19% of patients are infected by a Beijing and non-Beijing strain family. The results were compared to spoligotyping results. The methodology set a new precedent for the study of mixed infections and demonstrate the importance of reinfection as a mechanism leading to disease.

Analysis of IS*6110*- RFLP data, has led to the grouping of clinical isolates according to their number of hybridizing bands; high-copy-number strains with > 6 IS*6110* hybridizing bands, and low-copy-number strains with ≤ 6 IS*6110* hybridizing bands (36). This data has been applied to describe clonal expansion in high-copy-number strains (3,45-47) and for their epidemiologic analysis on a global scale (2). Only a limited amount of evolutionary data exists for low-copy-number strains due to their intrinsically limited IS*6110*-RFLP polymorphism. Therefore the mechanisms leading to clonal expansion of the low-copy-number strains remains largely unresolved. Furthermore, it is not known whether these groups of strains are genetically distinct in different geographical regions. In Chapter 6 we have used genotypic data generated using six different methods including spoligotyping to determine the genetic relationship between low-copy-number strains collected in Cape Town, South Africa. This data has been

compared to genotypic data collected from low-copy-number strains cultured in a broad spectrum of geographical settings.

M. tuberculosis isolates can be grouped into 3 principal phylogenetic groups based on polymorphisms in the *katG* and *gyrA* genes. It is thought that group 1 strains are the oldest and gave rise to group 2 strains, which subsequently evolved to generate group 3 strains (39). Within each of these groupings it is thought that strains have evolved different genetic characteristics that can be used to reconstruct their evolutionary history. In Chapter 7, a phylogenetic tree was constructed and associations between evolutionary branches and clinical presentation was assessed.

In chapter 8 we hypothesized that during the evolutionary process strains had evolved distinct signatures detected by spoligotyping, which were specific to a strain family defined by IS*6110*-RFLP. The identification of spoligotype signatures will make it possible to determine the strain population structure in different geographical settings on a global scale. Spoligotype signatures will also be important in monitoring drug and vaccine trials as it will enable the detection of strain families which may have a greater propensity to acquire drug resistance or to escape protective immunity.

Reference List

1. **Beck-Sague, C., S. W. Dooley, M. D. Hutton, J. Otten, A. Breeden, J. T. Crawford, A. E. Pitchenik, C. Woodley, G. Cauthen, and W. R. Jarvis**. 1992. Hospital outbreak of multidrug-resistant Mycobacterium tuberculosis infections. Factors in transmission to staff and HIV-infected patients. JAMA **268**:1280-1286.

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. **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.

4. **Braden, C. R., G. P. Morlock, C. L. Woodley, K. R. Johnson, A. C. Colombel, M. D. Cave, Z. Yang, S. E. Valway, I. M. Onorato, and J. T. Crawford**. 2001. Simultaneous infection with multiple strains of Mycobacterium tuberculosis. Clin.Infect.Dis. **33**:e42-e47.

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

6. **Chaves, F., F. Dronda, M. Alonso-Sanz, and A. R. Noriega**. 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.

7. **Coronado, V. G., C. M. Beck-Sague, M. D. Hutton, B. J. Davis, P. Nicholas, C. Villareal, C. L. Woodley, J. O. Kilburn, J. T. Crawford, T. R. Frieden, and .** 1993. Transmission of multidrug-resistant Mycobacterium tuberculosis among

persons with human immunodeficiency virus infection in an urban hospital: epidemiologic and restriction fragment length polymorphism analysis. J.Infect.Dis. **168**:1052-1055.

8. **Edlin, B. R., J. I. Tokars, M. H. Grieco, J. T. Crawford, J. Williams, E. M. Sordillo, K. R. Ong, J. O. Kilburn, S. W. Dooley, and K. G. Castro**. 1992. An outbreak of multidrug-resistant tuberculosis among hospitalized patients with the acquired immunodeficiency syndrome. N.Engl.J.Med. **326**:1514-1521.

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

10. **Filliol, I., J. R. Driscoll, D. van Soolingen, B. N. Kreiswirth, K. Kremer, G. Valetudie, D. D. Anh, R. Barlow, D. Banerjee, P. J. Bifani, K. Brudey, A. Cataldi, R. C. Cooksey, D. V. Cousins, J. W. Dale, O. A. Dellagostin, F. Drobniewski, G. Engelmann, S. Ferdinand, D. Gascoyne-Binzi, M. Gordon, M. C. Gutierrez, W. H. Haas, H. Heersma, 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 assessed by spoligotyping in an international study. J.Clin.Microbiol. **41**:1963-1970.

11. **Fischl, M. A., R. B. Uttamchandani, G. L. Daikos, R. B. Poblete, J. N. Moreno, R. R. Reyes, A. M. Boota, L. M. Thompson, T. J. Cleary, and S. Lai**. 1992. An outbreak of tuberculosis caused by multiple-drug-resistant tubercle bacilli among patients with HIV infection. Ann.Intern.Med. **117**:177-183.

12. **Frieden, T. R., L. F. Sherman, K. L. Maw, P. I. Fujiwara, J. T. Crawford, B. Nivin, V. Sharp, D. Hewlett, Jr., K. Brudney, D. Alland, and B. N. Kreisworth**. 1996. A multi-institutional outbreak of highly drug-resistant tuberculosis: epidemiology and clinical outcomes. JAMA **276**:1229-1235.

13. **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.

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

15. **Heersma, H. F., K. Kremer, and J. D. van Embden**. 1998. Computer analysis of IS6110 RFLP patterns of Mycobacterium tuberculosis. Methods Mol.Biol. **101**:395-422.

16. **Hermans, P. W., D. van Soolingen, E. M. Bik, P. E. de Haas, J. W. Dale, and J. D. van Embden**. 1991. 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. **59**:2695-2705.

17. **Herrera, L., S. Jimenez, A. Valverde, Garci, M. A. Aranda, and J. A. Saez-Nieto**. 2003. Molecular analysis of rifampicin-resistant Mycobacterium tuberculosis isolated in Spain (1996-2001). Description of new mutations in the rpoB gene and review of the literature. Int.J.Antimicrob.Agents **21**:403-408.

18. **Jereb, J. A., R. M. Klevens, T. D. Privett, P. J. Smith, J. T. Crawford, V. L. Sharp, B. J. Davis, W. R. Jarvis, and S. W. Dooley**. 1995. Tuberculosis in health care workers at a hospital with an outbreak of multidrug-resistant Mycobacterium tuberculosis. Arch.Intern.Med. **155**:854-859.

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

20. **Johnson, R., Warren, R. M., Strauss, O. J., Jordaan, A., Falmer, A. A., Beyers, N., Schaaf, H. S., Murray, M., Cloete, K., van Helden, P. D., and Victor, T. C.** 2006 An outbreak of drug resistant Tuberculosis caused by a Beijing strain in the Western Cape, South Africa. Int.J.Tuberc.Lung Dis. In Press

21. **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.

22. **Kanduma, E., T. D. McHugh, and S. H. Gillespie**. 2003. Molecular methods for Mycobacterium tuberculosis strain typing: a users guide. J.Appl.Microbiol. **94**:781-791.

23. **Kremer, K., D. van Soolingen, R. Frothingham, W. H. Haas, P. W. Hermans, C. Martin, P. Palittapongarnpim, B. B. Plikaytis, L. W. Riley, M. A. Yakrus, J. M. Musser, and J. D. van Embden**. 1999. Comparison of methods based on different molecular epidemiological markers for typing of Mycobacterium

tuberculosis complex strains: interlaboratory study of discriminatory power and reproducibility. J.Clin.Microbiol. **37**:2607-2618.

24. **Kubin, M., M. Havelkova, I. Hyncicova, Z. Svecova, J. Kaustova, K. Kremer, and D. van Soolingen**. 1999. A multidrug-resistant tuberculosis microepidemic caused by genetically closely related Mycobacterium tuberculosis strains. J.Clin.Microbiol. **37**:2715-2716.

25. **Mokrousov, I., I. Filliol, E. Legrand, C. Sola, T. Otten, E. Vyshnevskaya, E. Limeschenko, B. Vyshnevskiy, O. Narvskaya, and N. Rastogi**. 2002. Molecular characterization of multiple-drug-resistant Mycobacterium tuberculosis isolates from northwestern Russia and analysis of rifampin resistance using RNA/RNA mismatch analysis as compared to the line probe assay and sequencing of the rpoB gene. Res.Microbiol. **153**:213-219.

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

27. **Nivin, B., J. Driscoll, T. Glaser, P. Bifani, and S. Munsiff**. 2000. Use of spoligotype analysis to detect laboratory cross-contamination. Infect.Control Hosp.Epidemiol. **21**:525-527.

28. **Pearson, M. L., J. A. Jereb, T. R. Frieden, J. T. Crawford, B. J. Davis, S. W. Dooley, and W. R. Jarvis**. 1992. Nosocomial transmission of multidrug-resistant Mycobacterium tuberculosis. A risk to patients and health care workers. Ann.Intern.Med. **117**:191-196.

29. **Portugal, I., M. J. Covas, L. Brum, M. Viveiros, P. Ferrinho, J. Moniz-Pereira, and H. David**. 1999. Outbreak of multiple drug-resistant tuberculosis in Lisbon: detection by restriction fragment length polymorphism analysis. Int.J.Tuberc.Lung Dis. **3**:207-213.

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

31. **Rattan, A., A. Kalia, and N. Ahmad**. 1998. Multidrug-resistant Mycobacterium tuberculosis: molecular perspectives. Emerg.Infect.Dis. **4**:195-209.

32. **Richardson, M., N. M. Carroll, E. Engelke, G. D. van der Spuy, F. Salker, Z. Munch, R. P. Gie, R. M. Warren, N. Beyers, and P. D. van Helden**. 2002. Multiple Mycobacterium tuberculosis Strains in Early Cultures from Patients in a High-Incidence Community Setting. J.Clin.Microbiol. **40**:2750-2754.

33. **Scorpio, A., P. Lindholm-Levy, L. Heifets, R. Gilman, S. Siddiqi, M. Cynamon, and Y. Zhang**. 1997. Characterization of pncA mutations in pyrazinamide-resistant Mycobacterium tuberculosis. Antimicrob.Agents Chemother. **41**:540-543.

34. **Scorpio, A. and Y. Zhang**. 1996. Mutations in pncA, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculous drug pyrazinamide in tubercle bacillus. Nat.Med. **2**:662-667.

35. **Sepkowitz, K. A., J. Raffalli, L. Riley, T. E. Kiehn, and D. Armstrong**. 1995. Tuberculosis in the AIDS era. Clin.Microbiol.Rev. **8**:180-199.

36. **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 [see comments]. N.Engl.J.Med. **330**:1703-1709.

37. **Sola, C., A. Devallois, L. Horgen, J. Maisetti, I. Filliol, E. Legrand, and N. Rastogi**. 1999. Tuberculosis in the Caribbean: using spacer oligonucleotide typing to understand strain origin and transmission. Emerg.Infect.Dis. **5**:404-414.

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

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

40. **Valway, S. E., S. B. Richards, J. Kovacovich, R. B. Greifinger, J. T. Crawford, and S. W. Dooley**. 1994. Outbreak of multi-drug-resistant tuberculosis in a New York State prison, 1991. Am.J.Epidemiol. **140**:113-122.

41. **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 [see comments]. J.Clin.Microbiol. **31**:406-409.

42. **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.

43. **van Soolingen, D.** 2001. Molecular epidemiology of tuberculosis and other mycobacterial infections: main methodologies and achievements. J.Intern.Med. **249**:1-26.

44. **van Soolingen, D., P. E. de Haas, P. W. Hermans, P. M. Groenen, and J. D. van Embden**. 1993. Comparison of various repetitive DNA elements as genetic markers

for strain differentiation and epidemiology of Mycobacterium tuberculosis. J.Clin.Microbiol. **31**:1987-1995.

45. **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.

46. **Victor, T. C., A. van Rie, A. M. Jordaan, M. Richardson, G. D. Der Spuy, N. Beyers, P. D. van Helden, and R. Warren**. 2001. Sequence polymorphism in the rrs gene of Mycobacterium tuberculosis is deeply rooted within an evolutionary clade and is not associated with streptomycin resistance. J.Clin.Microbiol. **39**:4184-4186.

47. **Warren, R. M., M. Richardson, S. L. Sampson, G. D. van der Spuy, W. Bourn, J. H. Hauman, H. Heersma, W. Hide, N. Beyers, and P. D. van Helden**. 2001. Molecular evolution of Mycobacterium tuberculosis: phylogenetic reconstruction of clonal expansion. Tuberculosis.(Edinb.) **81**:291-302.

48. **Warren, R. M., E. M. Streicher, S. Charalambous, G. Churchyard, G. D. van der Spuy, A. D. Grant, P. D. van Helden, and T. C. Victor**. 2002. Use of spoligotyping for accurate classification of recurrent tuberculosis. J.Clin.Microbiol. **40**:3851-3853.

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

50. **World Health Organization**. 2003. WHO report 2003: Global Tuberculosis Control.

51. **World Health Organization**. 2006. WHO report 2006 Global tuberculosis control: Surveillance, Planning, Financing.

Chapter 2

Genotypic and phenotypic characterization of drug resistant *Mycobacterium tuberculosis* isolates from rural districts of the Western Cape Province of South Africa

Streicher, E.M.[1]; Warren R.M.[1]; Kewley, C.[2]; Simpson, J.[3]; Rastogi, N.[4]; Sola C.[4]; van der Spuy, G.D.[1]; van Helden, P.D.[1]; Victor, T.C.[1,*]

[1]MRC Centre for Molecular and Cellular Biology, Department of Medical Biochemistry, University of Stellenbosch, [2]Brewelskloof Hospital, Worcester, South Africa, [3]National Health Laboratory Services, [4]Institut Pasteur de Guadeloupe, Pointe à Pitre, Guadeloupe.

Published in Journal of clinical Microbiology Vol. 42, No. 2 Feb. 2004, p. 891–894

My contribution to this project:

Planning of project
Spoligotyping of isolates
Construction of database
Analysis of database
Comparison of spoligotypes to the international spoligotyping database in Guadeloupe
Writing of manuscript

ABSTRACT

Genotypic and phenotypic analysis of drug-resistant *Mycobacterium tuberculosis* isolates from the Western Cape Province of South Africa showed that drug resistance is widespread and recently transmitted. Multidrug resistant (MDR) isolates comprise 40% of this collection, and a large pool of isoniazid mono-resistance may be a future source of MDR tuberculosis.

TEXT

By using genotyping methods, outbreaks of multidrug-resistant tuberculosis (MDR-TB) have been identified in hospitals, among health care workers, in prisons, and in communities (1, 3–5, 8, 11, 17, 21), thus focusing attention on MDR-TB as a major public health issue. The most extensive MDR-TB outbreak of Beijing/W-like isolates (9) occurred in New York among 267 patients, the majority of whom were co-infected with human immunodeficiency virus. The Beijing or W-like family of *Mycobacterium tuberculosis* continues to be the focus of extensive investigations as such strains are widely spread throughout the world (2, 10), including South Africa (23) and Russia (15), where it constitutes the major family of MDR-TB isolates. There is a rising concern about the spread of MDR-TB strains from developing to developed countries. However, the bacterial population structure of resistant isolates from both developed and developing countries is not well documented. It is therefore possible that the drug-resistant Beijing/W-like strain or other MDR outbreak-associated strains are widespread but have not been recognized and reported as such. MDR-TB was first recorded in 1985 in the Western Cape Province of South Africa, but no genotype data are available from

these early isolates. Although drug surveillance studies have been done in the Western Cape, they have only provided information on the drug-resistant phenotype (24).

This study describes the genotypic and phenotypic characteristics of drug-resistant isolates (n = 482) from 328 cases collected from Jan 2001 to Feb 2002 (14 months) from 72 clinics in the Boland-Overberg and Southern Cape-Karoo regions of the Western Cape Province, South Africa. The first available drug-resistant isolate of each case as determined by the indirect proportion method on Lowenstein-Jensen medium (13) was used in this study. The age of patients varied from 15 to 73 years (mean, 37 years), and 59% of patients were male. The isolates were categorized into those that were isoniazid (INH) mono-resistant (48%), MDR (40%), and MDR with rifampin mono-resistance also included (12%). Each isolate was genotypically characterized by spoligotyping, using the internationally standardized method (12, 16). Sixty-nine different spoligotype patterns were identified and deposited in the international database at The Pasteur Institute of Guadeloupe (7). Of these, 34 were previously listed, while the remaining 35 types were newly added to this database. The genotypic and phenotypic data were analyzed to identify characteristics of the drug-resistant-TB epidemic in the region. The results indicate that MDR-TB forms a significant part (40%) of the drug resistant-TB epidemic and that drug-resistant TB was spread throughout the region. MDR-TB is present in the whole area but is more prevalent in one town in the Southern Cape region. INH mono-resistance is more prevalent in the Boland, Overberg, and Karoo regions but less represented in the Southern Cape region (Fig. 1).

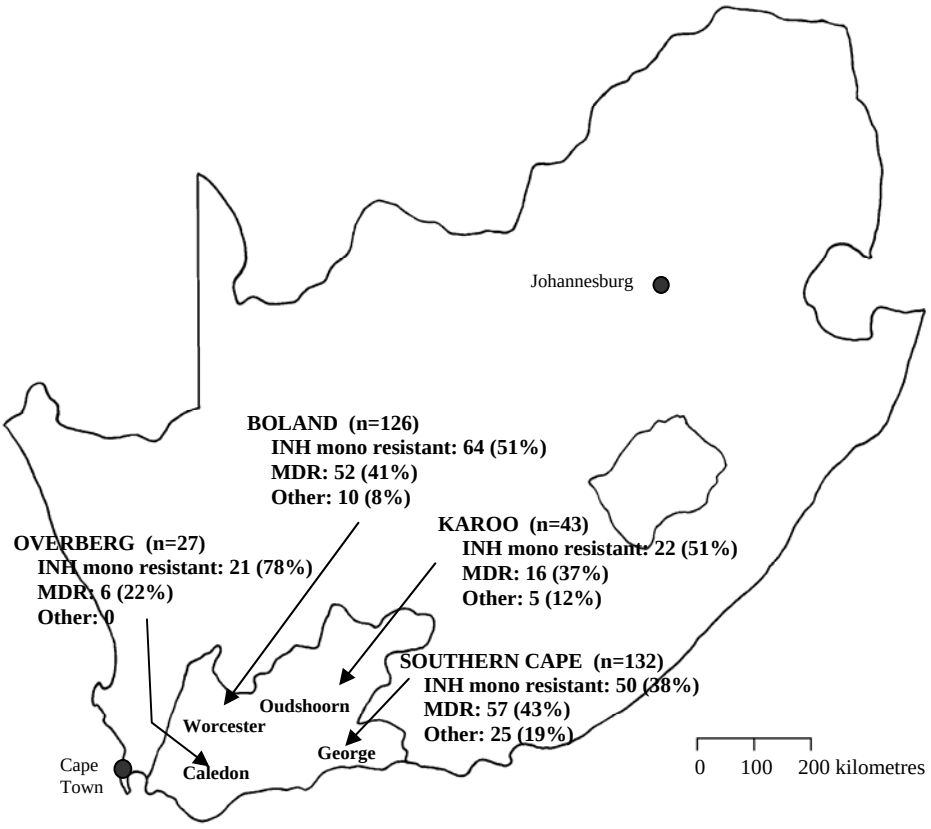
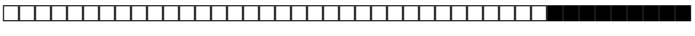
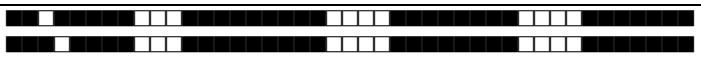
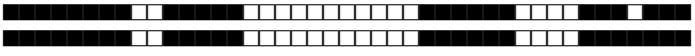
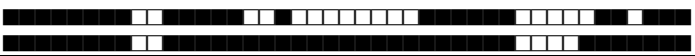
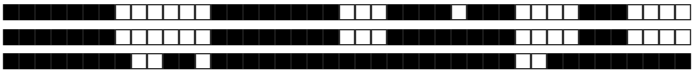


Figure 1 Map of South Africa with the distribution of the three drug resistance groups in 72 clinics of the Western Cape Province

Detailed analysis of the genotypic data showed that more than 80% of the isolates can be grouped based on genotypic spoligotypes and phenotypic drug resistance patterns (Table 1). The numerous clusters suggest that transmission of drug-resistant strains contributes to the spread of drug-resistant TB. This may be an overestimation of the extent of transmission of drug-resistant TB in the region, as the discriminatory power of spoligotyping is less than that of IS*6110* restriction fragment length polymorphism analysis (14). However, the results are similar to those of a IS*6110* restriction fragment length polymorphism analysis study which showed that more than 60% of drug-resistant TB occurred by transmission in urban communities in Cape Town, South Africa (22).

Table 1. Spoligotypes of drug-resistant isolates from two health districts of the Western Cape Province of South Africa.

Int. type [a]	Spoligotype pattern	INH mono^{R}	MDR	Multiple Drug^{R}	Family [b]
1		46	33	12	Beijing n=91 (28%)
npi		1	0	0	
npi		0	1	0	
npi		1	0	0	
npi		1	0	0	
npi		0	1	0	
211		1	0	0	F11 n=40 (12%)
npi		1	0	0	
npi		0	1	0	
npi		0	0	1	
npi		0	2	0	
130		1	0	1	
33		16	10	1	
npi		2	0	0	
71		2	0	0	
npi		2	0	0	F28 n=18 (5%)
npi		1	1	0	
npi		0	1	0	
34		4	3	2	
npi		0	1	0	
npi		1	0	0	
92		9	2	1	
npi		1	0	0	
npi		1	0	0	
348		1	3	0	LCC n=85 (26%)
npi		1	0	0	
npi		0	0	1	
npi (LT115)		0	21	6	
137		3	3	1	
336		0	5	1	
119		8	9	6	
2		0	0	1	Unknown n=94 (29%)
npi		0	1	0	
npi		0	1	0	
20		1	0	0	
npi		1	0	0	
npi		2	0	0	
21		1	0	0	
26		0	1	0	
npi		1	0	0	
npi		1	0	0	
npi		3	2	0	
npi		1	0	0	
npi		1	0	0	
npi		0	1	0	
npi		0	0	1	
npi		1	0	0	
35		0	1	0	
158		1	0	0	
766		0	1	0	
npi		1	0	0	
npi		0	1	0	
39		15	3	0	
811		1	1	0	
npi		2	4	2	
60		0	1	0	
42		3	2	1	
44		0	3	0	
373		2	0	1	
602		0	0	1	

62	1	0	0	
48	1	0	0	
237	0	2	0	
50	0	4	0	
519	1	0	0	
521	1	1	0	
52	0	1	0	
53	8	3	0	
54	3	0	0	
TOTAL (n=328)	157 (48%)	131 (40%)	40 (12%)	

[a] Int. type, international type (according to the international database at the Pasteur Institute of Guadeloupe [7]);
[b] four strain families: Beijing/W-like (direct variable repeats 1 to 34 deleted; correlates to share type 1 in reference 20); F11 (family 11) (direct variable repeats 9 to 11, 21 to 24, and 33 to 36 deleted; correlates to LAM3 family in reference 6); F28 (family 28) (direct variable repeats 9, 10, and 33 to 36 deleted; correlates to the S family in reference 6); LCC (IS*6110* low-copy-number clade) (direct variable repeats 18 and 33 to 36 deleted; correlates to the X family in reference 19).
Abbreviations: npi, not previously identified in spoligotypes DB3 and DB4 (7, 20); npi (LT115), not previously identified in spoligotypes DB3 and DB4 (7, 20), but local type number is 115; INH mono^{R}, INH mono-resistant isolates; Multiple drug^{R}, MDR isolates (rifampin-mono-resistant isolates included).

We speculate that the high level of transmission may in part be exacerbated by the relatively slow culture-based diagnostic procedures. Application of PCR techniques for rapid diagnosis of drug resistance may help to control the ongoing transmission. Based on the spoligotype patterns, the isolates could be grouped into families (Table 1), and the results showed that four strain families were responsible for more than 70% of the drug-resistant-TB epidemic in the region. The Beijing/W-like and the IS*6110* low-copy-number clade spoligotype patterns were the most prevalent drug-resistant isolates. It has been suggested that Beijing isolates are more frequently resistant due to their ability to mutate more rapidly than other strains (10). In neighboring communities in Cape Town, 17% of all patients with TB are infected with a Beijing strain (drug resistant or susceptible) (18). Patients infected with a Beijing/W-like drug-resistant isolate in this study may be overrepresented (28%), suggesting that the Beijing strain family acquires drug resistance mutations more frequently than strain families F11 (12%) and F28 (5%). This is still speculative and needs further investigation. The large pool of INH mono-resistant isolates (48%) is of great concern. INH mono-resistant isolates are mostly from the Beijing/ W-like family (29%). However, many genotypes of the INH mono-resistance

group are also present in the MDR group (Table 1), suggesting that MDR may have developed predominantly from INH mono-resistant isolates by selection. Such selection may easily result in additional MDR-TB in the future. This is supported by the observation that 20% of patients with MDR-TB in this region (since 1990) have had previous infections with mono-resistant or MDR strains (C. Kewley, personal communication). Results from the present investigation showed that cases of MDR-TB are overrepresented in the IS*6110* low-copy-number clade (54%) in comparison to the results seen with the Beijing/W-like (38%), F11 (38%), and F28 (28%) families. Conversely, the IS*6110* low-copy-number clade under-represents INH mono-resistant isolates (27%). Although the IS*6110* low-copy-number clade isolates are found throughout the region, most of the isolates (66%) originate from the Southern Cape. One particular spoligotype (local spoligotype 115) is unique to isolates from the largest town in the region, and this spoligotype was previously not identified in the international spoligotyping database of Guadeloupe, France (7). More than 75% of these isolates are MDR and may represent an outbreak of an emerging MDR-TB strain in this town. This study highlights the drug-resistant-TB epidemic in the Western Cape. It also raises the concern that drug resistance is transmitted and that there is a need for enhanced control strategies, which may include efficient, rapid molecular- based diagnostics.

We thank the IAEA (project SAF 6003 and RAF 6025), the NIH (grant number R21 AI 055800-01), and the Harry Crossley Foundation for financial assistance. We thank T. Dolman and H. Pretorius for preparation of isolates and A. Jordaan for technical assistance.

REFERENCES

1. **Beck-Sague, C., S. W. Dooley, M. D. Hutton, J. Otten, A. Breeden, J. T. Crawford, A. E. Pitchenik, C. Woodley, G. Cauthen, and W. R. Jarvis.** 1992. Hospital outbreak of multidrug-resistant Mycobacterium tuberculosis infections. Factors in transmission to staff and HIV-infected patients. JAMA **268:**1280–1286.

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. **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.

4. **Coronado, V. G., C. M. Beck-Sague, M. D. Hutton, B. J. Davis, P. Nicholas, C. Villareal, C. L. Woodley, J. O. Kilburn, J. T. Crawford, and T. R. Frieden.** 1993. Transmission of multidrug-resistant Mycobacterium tuberculosis among persons with human immunodeficiency virus infection in an urban hospital: epidemiologic and restriction fragment length polymorphism analysis. J. Infect. Dis. **168**:1052–1055.

5. **Edlin, B. R., J. I. Tokars, M. H. Grieco, J. T. Crawford, J. Williams, E. M. Sordillo, K. R. Ong, J. O. Kilburn, S. W. Dooley, and K. G. Castro.** 1992. An outbreak of multidrug-resistant tuberculosis among hospitalized patients with the acquired immunodeficiency syndrome. N. Engl. J. Med. **326:**1514– 1521.

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

7. **Filliol, I., J. R. Driscoll, D. van Soolingen, B. N. Kreiswirth, K. Kremer, G. Valetudie, D. D. Anh, R. Barlow, D. Banerjee, P. J. Bifani, K. Brudey, A. Cataldi, R. C. Cooksey, D. V. Cousins, J. W. Dale, O. A. Dellagostin, F. Drobniewski, G. Engelmann, S. Ferdinand, D. Gascoyne-Binzi, M. Gordon, M. C. Gutierrez, W. H. Haas, H. Heersma, 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 assessed by spoligotyping in an international study. J. Clin. Microbiol. **41:**1963–1970.

8. **Fischl, M. A., R. B. Uttamchandani, G. L. Daikos, R. B. Poblete, J. N. Moreno, R. R. Reyes, A. M. Boota, L. M. Thompson, T. J. Cleary, and S. Lai.** 1992. An outbreak of tuberculosis caused by multiple-drug-resistant tubercle bacilli among patients with HIV infection. Ann. Intern. Med. **117:**177–183.

9. **Frieden, T. R., L. F. Sherman, K. L. Maw, P. I. Fujiwara, J. T. Crawford, B. Nivin, V. Sharp, D. Hewlett, Jr., K. Brudney, D. Alland, and B. N. Kreisworth.** 1996. A multi-institutional outbreak of highly drug-resistant tuberculosis: epidemiology and clinical outcomes. JAMA **276:**1229–1235.

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. **Jereb, J. A., R. M. Klevens, T. D. Privett, P. J. Smith, J. T. Crawford, V. L. Sharp, B. J. Davis, W. R. Jarvis, and S. W. Dooley.** 1995. Tuberculosis in health care workers at a hospital with an outbreak of multidrug-resistant Mycobacterium tuberculosis. Arch. Intern. Med. **155:**854–859.

12. **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.

13. **Kleeberg, H. H., Z. Blackblock, F. Boulahbal, H. L. David, H. Fink, E. M. S. Gatner, J. Grosset, J. Juhlin, J. O. Kilburn, C. G. Kleeberg, F. Mandler, S. R. Pattyn, K. F. Pererson, H. Reutgen, E. H. Runon, H. Saito, K. H. Schroder, M. F. Stander, I. Szabo, S. Takahashi, L. Tripathy, L. Tinka, and B. Vergman.** 1985. A simple method of testing drug susceptibility of Mycobacterium tuberculosis: a report of an international collaborative study. Bull. Int. Union Tuberc. **60:**147–150.

14. **Kremer, K., D. van Soolingen, R. Frothingham, W. H. Haas, P. W. Hermans, C. Martin, P. Palittapongarnpim, B. B. Plikaytis, L. W. Riley, M. A. Yakrus, J. M. Musser, and J. D. van Embden.** 1999. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. J. Clin. Microbiol. **37:**2607–2618.

15. **Mokrousov, I., I. Filliol, E. Legrand, C. Sola, T. Otten, E. Vyshnevskaya, E. Limeschenko, B. Vyshnevskiy, O. Narvskaya, and N. Rastogi.** 2002. Molecular characterization of multiple-drug-resistant Mycobacterium tuberculosis isolates from

northwestern Russia and analysis of rifampin resistance using RNA/RNA mismatch analysis as compared to the line probe assay and sequencing of the rpoB gene. Res. Microbiol. **153:**213–219.

16. **Molhuizen, H. O. F., A. E. Bunschoten, L. M. Schouls, and J. D. A. Van Embden.** 1998. Mycobacteria protocols: rapid detection and simultaneous strain differentiation of *Mycobacterium tuberculosis* complex by spoligotyping. Methods Mol. Biol. **101:**381–393.

17. **Pearson, M. L., J. A. Jereb, T. R. Frieden, J. T. Crawford, B. J. Davis, S. W. Dooley, and W. R. Jarvis.** 1992. Nosocomial transmission of multidrugresistant Mycobacterium tuberculosis. A risk to patients and health care workers. Ann. Intern. Med. **117:**191–196.

18. **Richardson, M., S. W. van Lill, G. D. van der Spuy, Z. Munch, C. N. Booysen, N. Beyers, P. D. van Helden, and R. M. Warren.** 2002. Historic and recent events contribute to the disease dynamics of Beijing-like Mycobacterium tuberculosis isolates in a high incidence region. Int. J. Tuberc. Lung Dis. **6:**1001–1011.

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

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

21. **Valway, S. E., S. B. Richards, J. Kovacovich, R. B. Greifinger, J. T. Crawford, and S. W. Dooley.** 1994. Outbreak of multi-drug-resistant tuberculosis in a New York State prison, 1991. Am. J. Epidemiol. **140:**113–122.

22. **van Rie, A., R. Warren, M. Richardson, R. P. Gie, D. A. Enarson, N. Beyers, and P. D. van Helden.** 2000. Classification of drug-resistant tuberculosis in an epidemic area. Lancet **356:**22–25.

23. **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.

24. **Weyer, K., P. Groenewald, M. Zwarenstein, and C. J. Lombard.** 1995. Tuberculosis drug resistance in the Western Cape. S. Afr. Med. J. **85:**499–504.

Chapter 3

Spread of an emerging *Mycobacterium tuberculosis* drug resistant strain in the Western Cape of South Africa

Elizabeth M. Streicher[1], Thomas C. Victor[1], Carolyn Kewley[2], Annemie M. Jordaan[1], Gian D. van der Spuy[1], Marlein Bosman[3], Hannelie Louw[4], Megan Murray[5], Douglas Young[6], Paul D. van Helden[1], Robin M. Warren[1].

DST/NRF Centre of Excellence in Biomedical Tuberculosis Research / MRC Centre for Molecular and Cellular Biology, Stellenbosch University[1], Department of Health, Worcester[2], Department of Health, George[4], National Health Laboratory Services, Cape Town[3], South Africa, Harvard School of Public Health, Boston, USA[5], and Imperial College, London, UK[6].

Published in International Journal of Tuberculosis and Lung Disease 2007 Feb;11(2):195-201

My contribution to this project:
Identification of isolates part of DRF150
Spoligotyping of isolates
VNTR of isolates
Analysis of data
Writing and editing of manuscript

ABSTRACT

Background

South Africa has a high burden of Drug resistant Tuberculosis (TB).

Methods

Routine drug susceptibility testing was done prospectively over a two year period on *Mycobacterium tuberculosis* isolates in 2 health districts of the Western Province, South Africa. A cluster of drug resistant strains that shared a rare mutation in *katG315* was identified in 64 of the 450 cases who were identified as having been infected with drug resistant TB. Isolates belonging to this cluster were further characterised by phenotyping and genotyping tests. Epidemiologic and clinical characteristics of these cases were used to identify mechanisms leading to the acquisition and spread of this drug resistant strain.

Results

An outbreak of an emerging non-Beijing drug resistant strain, infecting 64 pulmonary TB cases was identified. This previously undetected genotype (now designated DRF150) is characterized by 5 IS*6110* insertions, a specific spoligotype and high levels of resistance to the first line TB medications, Isoniazid, Streptomycin and Rifampin. In 45% of the cases it is also resistant to two other first line drugs, Ethambutol and Pyrazinamide. Key factors leading to the development and spread of this drug resistant genotype were inappropriate chemotherapy; poor adherence to treatment; prolonged periods of infectiousness due to lack of susceptibility testing in new cases and delay in availability of drug susceptibility results in re-treatment cases.

Conclusions

Molecular markers allowed early identification of an emerging non-Beijing drug resistant strain.

Introduction

One of the principal aims of the WHO-recommended Direct Observed Therapy Short course (DOTS) program is to ensure that TB is detected and treated in a consistent manner to prevent the development of drug resistance. Despite widespread implementation of this program in sub-Saharan Africa, there has not yet been a decline in the incidence of TB. Factors that may contribute to the persistence of TB in this setting include poverty, administrative neglect of National Tuberculosis Programmes (NTPs), and the impact of the HIV pandemic on the re-emergence of TB. The emergence of drug-resistant strains of *Mycobacterium tuberculosis* poses an additional burden to the success of NTPs.

Drug resistant TB occurs through two routes. Resistance first arises when inadequate therapy leads to the selection of resistant strains in cases on therapy for TB. When resistance develops through this mechanism, it is termed acquired resistance. Drug resistance can also develop when individuals are infected with a resistant strain. In this case, termed "primary resistance", resistance can be detected both in people who are newly diagnosed and have not previously been treated for TB or in previously treated re-infected cases. It is important to distinguish between primary resistance and acquired resistance as they have different implications to NTPs. There is much debate about the relative contribution of acquired and primary resistance to the burden of drug resistant TB in different communities. This controversy focuses on the relative transmissibility of MDR strains, i.e. whether the mutations that confer drug resistance also impair the reproductive function of the organism. Evidence that MDR strains can be transmitted comes from a series of molecular epidemiological studies of outbreaks reported over the past decade. These outbreaks have been identified in hospitals (2-4,7,11), amongst health care workers (1,9,12), in prisons (15), and in communities (18), and have focused attention on MDR-TB as a major public health issue.

MDR-TB was first identified in the Western Cape area of South Africa in 1985 and within 9 years accounted for 2% of the TB isolates in this region (23). Recent reports have continued to highlight drug resistance in the Western Cape of South Africa and have suggested that much of this disease burden is due to ongoing transmission (14,17,18). In this study we report on clonal expansion and interregional spread of an emerging highly drug resistant strain in communities of South Africa. We investigate the reasons for the rapid spread of this strain and discuss these findings in the context of the NTP.

Materials and methods

This study was approved by the ethics committee of Stellenbosch University, Tygerberg, South Africa and informed consent to participate in this study was obtained from each case.

Study setting

The study was conducted in 2 of the 4 health districts of the Western Province, South Africa from September 2000 to August 2002. Twelve of these clinics are in larger towns (Worcester and George; Figure 1), while the remaining clinics are in rural communities. Drug resistant TB in these two districts is managed by a team of dedicated physicians and primary health care workers devoted to the management of TB in the public sector. The incidence of TB was ~1300 /100 000 in 2002. HIV prevalence in the province as estimated from antenatal screening was <7% in 2001. The geographical study area covers 93 000 km^2 with a predominant mixed race group in an estimated population of 1 348 405 in 2001 (22). TB cases in the Boland/Overberg and Southern Cape/Karoo regions were treated according to the National TB Control Program. Routine drug

susceptibility testing (DST) was performed as described previously(16). DST is done in all patients who do not achieve sputum conversion after 2 months (new patients) or 3 months (retreatment patients), in all re-treatment cases prior to treatment, and in defined risk groups such as contacts of confirmed MDR-TB cases. The minimum inhibition concentration (MIC) of Isoniazid (H), Rifampin (R) and Streptomycin (S) for selected isolates was determined by using different concentrations of the drug in BACTEC 12B medium as described in the manual of the manufacturers (Becton Dickinson). Isolates from the study area that were identified as drug resistant were referred for further genotypic characterization. During the study period 100 fully drug susceptible isolates were collected from the same region and also used for genotyping.

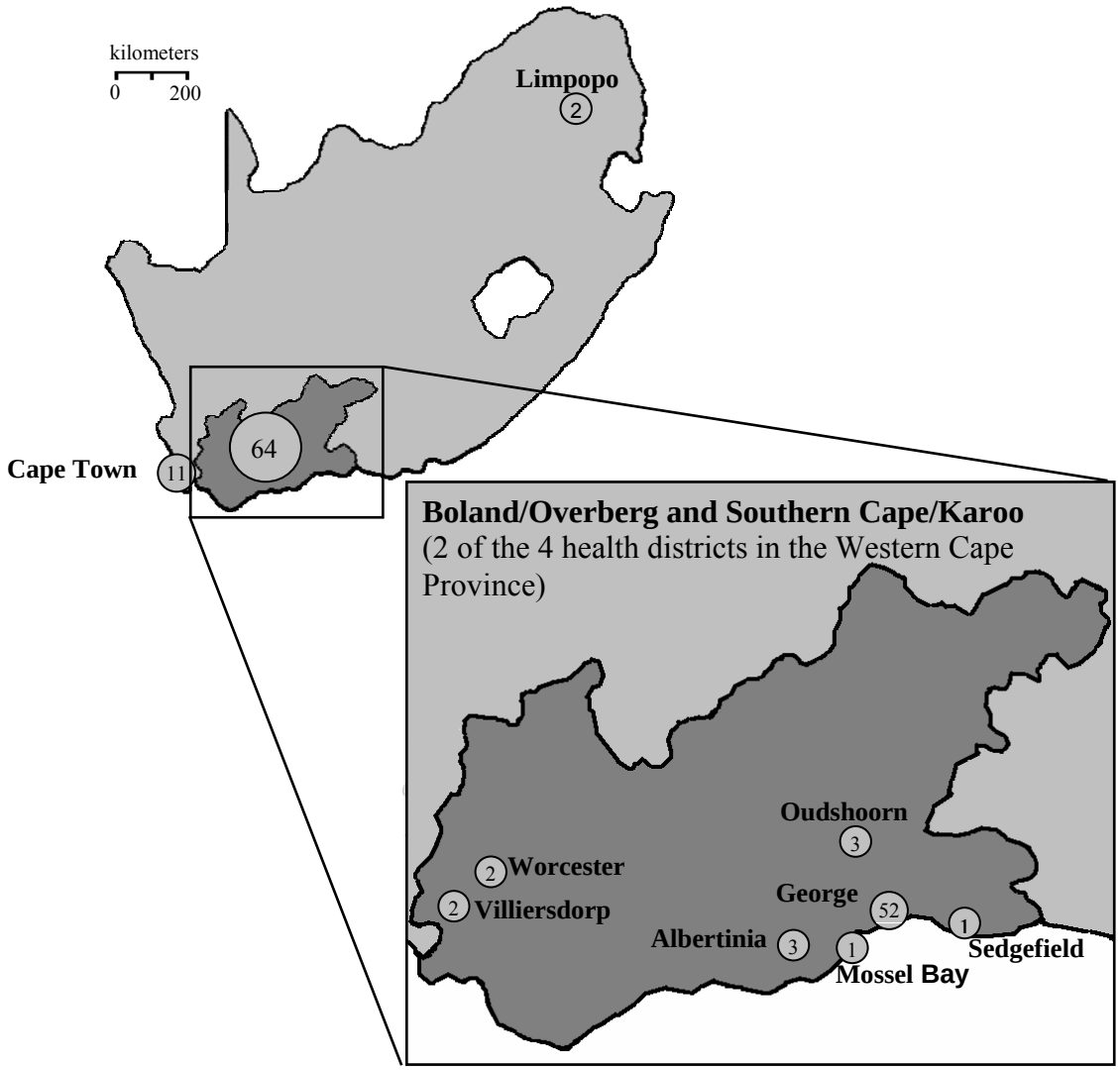


Figure 1. Distribution of currently known cases infected with DRF150 in South Africa
Fifty two of the 64 cases were identified in the town of George and 8 other cases were located in more rural areas within a 100 km radius of this town. The other 4 cases were isolated in rural areas. Cases (n=11) with similar drug resistant infecting isolates were identified in Cape Town (16) previously and in this study 2 cases in the Limpopo Province (~ 2000km north; result from Molecular database, Stellenbosch University, Tygerberg, South Africa).

Genotypic characterization of isolates.

DNA fingerprinting by Spoligotyping and by IS*6110* restriction fragment length polymorphism (RFLP) analysis using IS-3' and IS-5' probes was done by standardized protocols (10,21). Spoligotypes were compared to the SpolDB3 international database containing spoligotypes from more than 90 countries (5). Identification of gene mutations associated with phenotypic drug

resistant patterns was done by a PCR based dot-blot hybridisation method (20). Direct DNA sequencing of selected PCR products was done with an automated ABIPRISM (model 3100, Applied Biosystems) analyser. A cluster is defined as more than one isolate that has exactly the same characteristics.

Epidemiologic and clinical data

Demographic and clinical data were obtained retrospectively through direct patient interviews and medical record review by a trained nurse who utilized standard data-collection instruments. HIV testing was done according to the national guidelines of Voluntary Testing and Counselling (VTC). Cases were considered HIV positive when the initial HIV screening at the clinic was confirmed by Elisa testing. Cases were directly observed and considered adherent to medication if at least 80% of the doses were taken as prescribed by the physician. GraphPad Prism Version 4.03 was used for statistical calculation.

Results

During the study period 450 cases were diagnosed with drug resistant TB by routine DST. Of these, 64/450 cases (14%) were infected with a drug resistant genotype which had a unique mutation at position 315 of the *katG* gene (315*gc*Ser→*ca*Thr; conferring H resistance). In 61/64 (95%) of these, the strain also shared identical mutations at codon 43 in the *rpsL* gene (43*a*Lys→*g*Arg; conferring S resistance). The related clonal structure of these strains was supported by IS*6110*-RFLP fingerprint analysis which showed that 63/64 (98%) of the isolates had identical IS*6110* banding patterns with 5 insertions (Figure 2) while the remaining isolate had 4 of

the 5 insertions, suggesting a deletion of one IS*6110* element. Fully drug susceptible isolates with similar fingerprints lack the corresponding gene mutations.

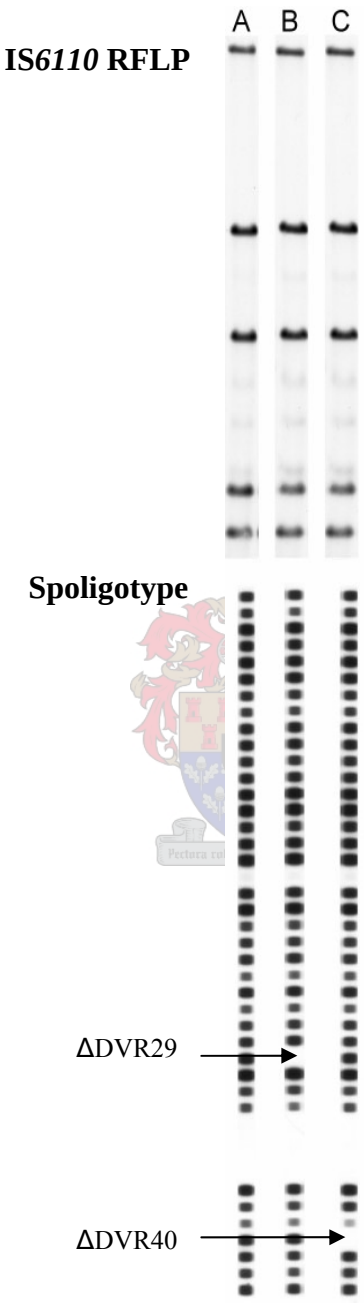


Figure 2. RFLP and Spoligotypes patterns of DRF150
IS*6110*-RFLP analysis showed that 63 of the isolates with the unique mutation at position 315 of the *katG* gene (315gc→ca; Ser→Thr) had identical banding patterns with 5 insertions. The remaining isolate showed 4 of the 5 hybridizing bands (result not shown). Spoligotyping divided the strains into the following three distinct groups: Group A - lack direct variable repeat (DVR) spacers 18, 33-36; Group B - also lack DVR spacer 29; Group C - also lack DVR spacer 40. Strains in Group B do not have shared spoligotypes in the international database at the Pasteur Institut of Guadeloupe (5,6).

Phylogenetic reconstruction of all 64 isolates (Figure 3) predicts that this clone was originally fully drug susceptible and then in a stepwise manner evolved into clones with additional drug resistance mutations. The clones can be divided into 3 groups according to direct variable repeat (DVR) deletions that can be seen in the spoligotyping patterns (Groups A-C, Figure 2 and Figure 3). All the isolates from these 64 cases were resistant to H (MIC = 2.5 ug/ml), 61/64 (95%) were also resistant to S (MIC >1000 ug/ml), 47/64 (73%) were MDR (resistant to HR; MIC for R >100 ug/ml) and 29/64 (45%) of isolates were resistant to as many as 5 anti-TB drugs. Resistance patterns for the defined clusters are shown in Figure 3. From the phylogenetic tree 11 clusters were observed, suggesting recent transmission. Clustering was shown to be significantly associated with the number of resistance markers (Spearman Rank correlation coeff = 0.9276; P value 0.0167). This may be explained by diagnostic delay since initial DST results (H and R) for these cases were available from the routine laboratory between 29-75 days (median 38 days) after the physician's request. The final DST report for second line drugs was available between 31-121 days (median 64 days) after the sample was taken. From the 59 isolates requested to test for second-line drug resistance, 6 were found to be resistant to second-line drugs. In 1999 eleven TB cases infected with isolates with similar phenotypic and genotypic characteristics were identified in Cape Town (16). Two cases with similar drug resistant isolates were now identified from the Limpopo Province (approximately 2000 km north) during the same time period as the current study (Figures 1 and 3). We refer to this Strain as DRF150 (drug resistant Family 150).

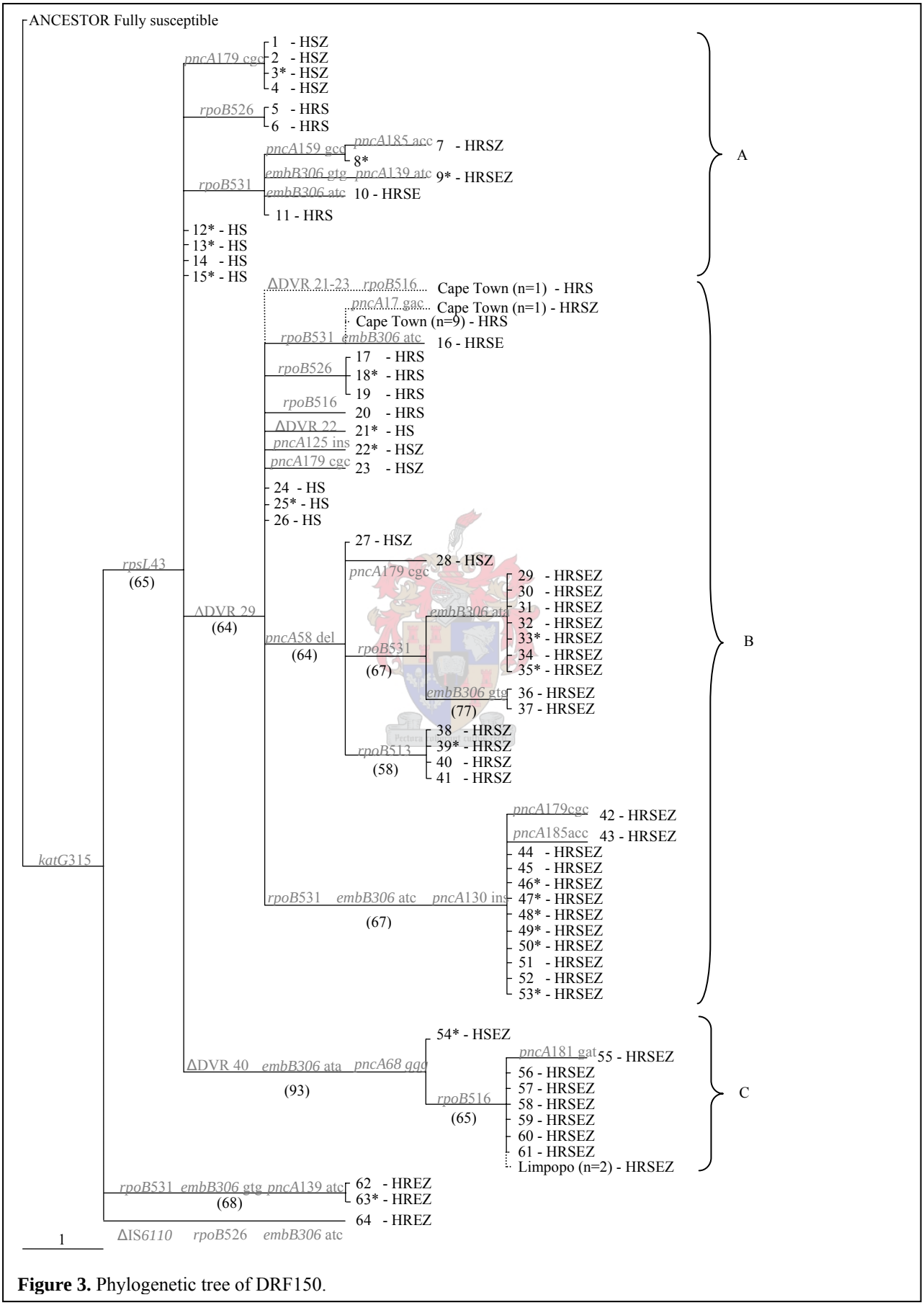


Figure 3. Phylogenetic tree of DRF150.

Figure 3 Phylogenetic tree of DRF150.
For phylogenetic reconstruction, the presence (indicated by "1") or the absence (indicated by "0") of each nucleotide change observed were assigned for all 64 DRF150 isolates. The complete set of evolutionary states were then subjected to phylogenetic analysis using the neighbour joining algorithm (PAUP 4.0*; Phylogenetic Analysis Using Parsimony (*Other Methods) Version 4b10. Sinauer Associates, Sunderland, Massachusetts). Bootstrapping was performed to establish a degree of statistical support for nodes within each phylogenetic reconstruction. A consensus tree was generated using the program contree (PAUP 4.0*) in combination with the majority rule formula. The resulting tree was rooted to the ancestral drug susceptible strain (ANC) which showed the characteristic IS*6110* banding pattern and wild type DNA sequences for the *katG, rpoB, rpsL, embB* and *pncA* genes. Only branches which occurred in > 50 % of the bootstrap trees were included in the final tree and all branches with a zero branch length were collapsed. Bootstrap values are shown at interval nodes in brackets. A,B,C are groupings according to spoligotypes in Figure 2. TB cases infected with similar strains previously identified from Cape Town (n=11)(16) and from the Limpopo Province (n= 2) (approximately 2000 km north; result from Molecular database, Stellenbosch University, Tygerberg, South Africa) during the same time period as the current study are also marked in the tree. New cases are marked with a *. H=Isonizid; R=Rifampin; S=Streptomycin; Z=Pyrazinamide; E=Ethambutol.

Tables 1 and 2 presents the demographics, clinical characteristics and MDR contacts of the 64 cases in the outbreak. Most of the cases were male 38/64 (59%); many had low levels of education, were poorly paid or unemployed. There were more cases ($p<0.0001$) of DRF150 TB amongst those of the South African Coloured ethnicity than among other ethnic groups. Patients' ages ranged from <2 – 62 years with an average of 35 years; 13/64 (20%) of the cases were less then 25 years old and 38/64 (59%) were between 25 and 44 years of age. Previously treated patients constituted 43/64 (67%) of the cases and newly diagnosed cases 21/64 (33%). All the newly diagnosed cases had initially failed therapy on first line drugs. Among those who had been treated for a previous TB episode, 18/43 (42%) had been classified as cured after their first episode of disease, 4/43 (9%) had been considered treatment failures and 10/43 (23%) had not adhered to their first treatment regimen. No documentation on the previous treatment course was available for 9/43 (21%) cases. Treatment outcomes for the 64 cases infected with DRF150 are shown in Table 2. Sixteen of the cases (25%) died, including the 2 cases (2/64; 3.1%) that were HIV infected. Thirty three (52%) of the cases reported contact with a person known to be infected with MDR-TB.

Table 1. Demographic data of cases infected with DRF150

		New cases (%)	Re-treatment cases (%)	All cases (%)
TOTAL		**21 (32.8%)**	**43 (67.2%)**	**64**
Ethnicity	Black	3 (14.3%)	1 (2.3%)	4 (6.3%)
	SAC*	18 (85.7%)	42 (97.7%)	60 (93.8%)
Sex	Female	9 (42.9%)	17 (39.5%)	26 (40.6%)
	Male	12 (57.1%)	26 (60.5%)	38 (59.4%)
Age	0-4	0 (0%)	2 (4.7%)	2 (3.1%)
	5-14	0 (0%)	0 (0%)	0 (0%)
	15-24	4 (19.1%)	7 (16.3%)	11 (17.1%)
	25-34	5 (23.8%)	7 (16.3%)	12 (18.7%)
	35-44	4 (19.1%)	22 (51.2%)	26 (40.6%)
	>45	7 (33.3%)	5 (11.5%)	12 (18.7%)
	Unknown	1 (4.7%)	0 (0%)	1 (1.6%)
Average age		37	34	35

Demographic for 50 of the 64 DRF150 cases were obtained retrospectively through direct patient interviews by a trained study nurse who utilized standard data-collection instruments. Only partial data were available for the remaining 14 cases. SAC* = South African Coloured.

Table 2. Clinical characteristics and MDR contacts of cases infected with DRF150

		New cases (%)	Retreatment cases (%)	All cases (%)
Retreatment	Re-treatment after cure	n.a.	18 (41.9%)	18 (41.9%)
	Re-treatment after interruption	n.a.	10 (23.2%)	10 (23.3%)
	Re-treatment after completion	n.a.	2 (4.7%)	2 (4.7%)
	Re-treatment after failure	n.a.	4 (9.3%)	4 (9.3%)
	unknown	n.a.	9 (20.9%)	9 (20.9%)
Outcome	Treatment interrupted	1 (4.8%)	4 (9.3%)	5 (7.8%)
	Cured	12 (57.1%)	15 (34.9%)	27 (42.2%)
	Patient died	3 (14.3%)	13 (30.2%)	16 (25%)
	Still on treatment	3 (14.3%)	7 (16.3%)	10 (15.6%)
	Refuse further treatment	0 (0%)	1 (2.3%)	1 (1.6%)
	unknown	2 (9.5%)	3 (7%)	5 (7.8%)
HIV+	Positive	1 (4.8%)	1 (2.3%)	2 (3.1%)
	Negative	14 (66.7%)	26 (60.5%)	40 (62.5%)
	Unknown	5 (23.8%)	17 (39.5%)	22 (34.4%)
MDR contact		6 (28.6%)	27 (62.8%)	33 (51.6%)

Clinical data for 50 of the 64 DRF150 cases were obtained retrospectively through medical record review by a trained study nurse who utilized standard data-collection instruments. Only partial data were available for the remaining 14 cases.
n.a. = not applicable

Figure 1 indicates patient location at the time of diagnosis. Fifty two cases were identified in the suburban areas of George (population 150 000) and 8 other cases were located in more rural areas within a 100 km radius of this town. The other 4 cases were isolated in rural areas. The 52 cases in the suburban areas of George represent 52/123 (42%) of the total number of drug resistant cases identified in this urban area during the study period. DRF150 strain continues to be identified at a similar level in this region.

Discussion

This study is the first comprehensive application of molecular epidemiological methods to describe strains of *M. tuberculosis* that gained resistance in a stepwise manner to several front-line anti-TB drugs in South Africa. The community based study identified an outbreak of drug resistant TB involving strains that belong to a newly described genotype – DRF150. This highly drug resistant clone represented almost half of the cases of drug resistant TB identified in the George sub-district during the study period. The young age of disease onset, the spread within families and the strong association with MDR contacts lends further support for primary spread of the disease.

This study highlights the risk of widespread transmission of drug resistant strains that are not of the Beijing/W strain lineage and suggests that circulation of further currently unrecognized drug resistant strains may contribute to the global drug resistant TB epidemic. Our study also suggests that the drug resistance of the DRF150 genotype did not impair the reproductive function (fitness) of this outbreak organism since many transmission events took place during a short time frame. This supports previous observations that strains with mutations in the *katG*315 gene retain virulence and transmissibility (8,13,19). Similarly, acquisition of subsequent mutations do not

appear to have a major fitness cost in DRF150 as demonstrated by the significant positive correlation between the number of mutations and clustering.

We have previously found *M. tuberculosis* isolates in local communities with similar phenotypic and genotypic characteristics as described for DRF150 (14,16,20). However, cultures from cases detected prior to this longitudinal study are not available so it is not possible to establish precisely when and where this outbreak started. The ongoing transmission of this highly drug resistant strain and the amplification of resistance may have been due to sub-optimal chemotherapy, poor adherence to second line chemotherapy, delay in availability of drug susceptibility patterns and/or failure to actively trace contacts. These results suggest that, in areas with high levels of drug resistance, it may be necessary to assess drug susceptibility in all new cases of TB and to promote active cases finding and a dedicated patient centred approach for the treatment of drug resistant TB cases. This may be particularly important in communities, such as described here, where HIV still has to peak. Such a policy would be further strengthened by the use of rapid diagnostic testing by molecular methods. The identification of the DRF150 outbreak would not have been possible without molecular application. It is also important to note that since most of the cases infected with DRF150 are re-treatment cases, they would have been classified as treatment failures with acquired resistance if molecular typing had not been performed. A vast majority of these cases were retreatment cases, and of these patients only 42% were considered initially as cured. That leaves 32% percent of patients who either failed or did not adhere to therapy. This is central to the explanation as to why this particular genotype could have emerged. Since the initial isolate from these patients is not available for typing, a relapse with a resistant organisms cannot be ruled out.

The outbreak strain described in this study differs from the Beijing genotype in that it has a low copy number of IS*6110* insertions and belongs to principal genetic group 2 whereas the Beijing genotype belongs to group 1 (21). Although the rare dinucleotide mutation in *katG*315 can be used as a marker to perform initial screening for cases infected with DRF150, it is a combination of the mutations and DNA fingerprints (spoligotype or IS*6110* RFLP) that defines the DRF150 genotype. Importantly, the same dinucleotide mutation has previously been noted in Beijing/W isolates from New York, but is rarely found in other strain types or in Beijing/W isolates from other regions (2). This, together with the observation that most of the DRF150 genotypes do not have shared spoligotypes in the international spoligotype database (5,6) further support the notion that DRF150 is an emerging drug resistant strain.

Several factors suggest that it will be difficult to eradicate outbreaks such as this one or to contain their spread to new hosts (2). Firstly, the presence of this transmissible drug resistant strain poses a threat to healthcare workers and suggests that rigorous infection control measures will be needed to ensure the safety of those exposed to people with active disease. In addition, many people may have been latently infected with DRF150 and will pose a further risk of reactivation and transmission. By the current detection and treatment strategies in South Africa, resistance to anti-TB drugs will go unnoticed until the DST is requested and obtained. The policy for culture and DST in SA actually calls for routine DST in all patients who do not achieve sputum conversion after 2 months (new patients) or 3 months (retreatment patients), in all retreatment cases prior to treatment, and in defined risk groups such as contacts of confirmed MDR-TB cases. Since DRF150 is resistant to very high concentrations of first line anti-TB drugs, the current DOTS approach in primary drug resistant TB cases will fail to successfully treat this strain. This means that patients on first line drug therapy may remain infectious and continue to transmit MDR-TB

for prolonged periods of time. In addition, repeated exposure to first line drugs may promote acquisition of additional resistance mutations.

The occurrence of this outbreak in a setting in which a good DOTS program was in place testifies to the need to dedicate more resources to intensify basic NTP measures as well as extend them to include specific activities directed at the diagnosis and treatment of MDR-TB. Such measures may include active case finding and evaluation of social networks for sites of transmission, improved communication between laboratory staff and public health workers, and the use of genotyping to identify clustered cases within days.

Acknowledgements

The authors would like to thank South African National Research Foundation (grants GUN 2054278 and DST/NRF Centre of Excellence in Biomedical Tuberculosis Research), the Wellcome Trust (grant Ref. 072402/Z/03/Z), the NIH (grant R21 A155800-01) and the IAEA (grant SAF6008) for financial support. The authors would also like to thank Erica Engelke for MIC analysis, and Nalin Rastogi and Christophe Sola for use of the international spoligotype database.

Reference List

1. **Beck-Sague, C., S. W. Dooley, M. D. Hutton, J. Otten, A. Breeden, J. T. Crawford, A. E. Pitchenik, C. Woodley, G. Cauthen, and W. R. Jarvis**. 1992. Hospital outbreak of multidrug-resistant Mycobacterium tuberculosis infections. Factors in transmission to staff and HIV-infected patients. JAMA **268**:1280-1286.

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. **Coronado, V. G., C. M. Beck-Sague, M. D. Hutton, B. J. Davis, P. Nicholas, C. Villareal, C. L. Woodley, J. O. Kilburn, J. T. Crawford, T. R. Frieden, and .** 1993. Transmission of multidrug-resistant Mycobacterium tuberculosis among persons with human immunodeficiency virus infection in an urban hospital: epidemiologic and restriction fragment length polymorphism analysis. J.Infect.Dis. **168**:1052-1055.

4. **Edlin, B. R., J. I. Tokars, M. H. Grieco, J. T. Crawford, J. Williams, E. M. Sordillo, K. R. Ong, J. O. Kilburn, S. W. Dooley, K. G. Castro, and .** 1992. An outbreak of multidrug-resistant tuberculosis among hospitalized patients with the acquired immunodeficiency syndrome. N.Engl.J.Med. **326**:1514-1521.

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

6. **Filliol, I., J. R. Driscoll, D. van Soolingen, B. N. Kreiswirth, K. Kremer, G. Valetudie, D. D. Anh, R. Barlow, D. Banerjee, P. J. Bifani, K. Brudey, A. Cataldi, R. C. Cooksey, D. V. Cousins, J. W. Dale, O. A. Dellagostin, F. Drobniewski, G. Engelmann, S. Ferdinand, D. Gascoyne-Binzi, M. Gordon, M. C. Gutierrez, W. H. Haas, H. Heersma, 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 assessed by spoligotyping in an international study. J.Clin.Microbiol. **41**:1963-1970.

7. **Fischl, M. A., R. B. Uttamchandani, G. L. Daikos, R. B. Poblete, J. N. Moreno, R. R. Reyes, A. M. Boota, L. M. Thompson, T. J. Cleary, and S. Lai**. 1992. An outbreak of tuberculosis caused by multiple-drug-resistant tubercle bacilli among patients with HIV infection. Ann.Intern.Med. **117**:177-183.

8. **Gillespie, S. H., O. J. Billington, A. Breathnach, and T. D. McHugh**. 2002. Multiple drug-resistant Mycobacterium tuberculosis: evidence for changing fitness following passage through human hosts. Microb.Drug Resist. **8**:273-279.

9. **Jereb, J. A., R. M. Klevens, T. D. Privett, P. J. Smith, J. T. Crawford, V. L. Sharp, B. J. Davis, W. R. Jarvis, and S. W. Dooley**. 1995. Tuberculosis in health care workers at a hospital with an outbreak of multidrug-resistant Mycobacterium tuberculosis. Arch.Intern.Med. **155**:854-859.

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. **Moro, M. L., A. Gori, I. Errante, A. Infuso, F. Franzetti, L. Sodano, and E. Iemoli**. 1998. An outbreak of multidrug-resistant tuberculosis involving HIV-infected patients of two hospitals in Milan, Italy. Italian Multidrug-Resistant Tuberculosis Outbreak Study Group. AIDS **12**:1095-1102.

12. **Pearson, M. L., J. A. Jereb, T. R. Frieden, J. T. Crawford, B. J. Davis, S. W. Dooley, and W. R. Jarvis**. 1992. Nosocomial transmission of multidrug-resistant Mycobacterium tuberculosis. A risk to patients and health care workers [see comments]. Ann.Intern.Med. **117**:191-196.

13. **Pym, A. S., B. Saint-Joanis, and S. T. Cole**. 2002. Effect of katG mutations on the virulence of Mycobacterium tuberculosis and the implication for transmission in humans. Infect.Immun. **70**:4955-4960.

14. **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.

15. **Valway, S. E., S. B. Richards, J. Kovacovich, R. B. Greifinger, J. T. Crawford, and S. W. Dooley**. 1994. Outbreak of multi-drug-resistant tuberculosis in a New York State prison, 1991. Am.J.Epidemiol. **140**:113-122.

16. **van Rie, A., R. Warren, I. Mshanga, A. M. Jordaan, G. D. van der Spuy, M. Richardson, J. Simpson, R. P. Gie, D. A. Enarson, N. Beyers, P. D. van Helden, and T. C. Victor**. 2001. Analysis for a limited number of gene codons can predict drug resistance of Mycobacterium tuberculosis in a high-incidence community. J.Clin.Microbiol. **39**:636-641.

17. **van Rie, A., R. Warren, M. Richardson, R. P. Gie, D. A. Enarson, N. Beyers, and P. D. van Helden**. 2000. Classification of drug-resistant tuberculosis in an epidemic area. Lancet **356**:22-25.

18. **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.

19. **van Soolingen, D., P. E. de Haas, H. R. van Doorn, E. Kuijper, H. Rinder, and M. W. Borgdorff**. 2000. Mutations at amino acid position 315 of the katG gene are associated with high-level resistance to isoniazid, other drug resistance, and successful transmission of Mycobacterium tuberculosis in the Netherlands. J.Infect.Dis. **182**:1788-1790.

20. **Victor, T. C., A. M. Jordaan, A. van Rie, G. D. van der Spuy, M. Richardson, P. D. van Helden, and R. Warren**. 1999. Detection of mutations in drug resistance genes of Mycobacterium tuberculosis by a dot-blot hybridization strategy. Tuber.Lung Dis. **79**:343-348.

21. **Warren, R. M., S. L. Sampson, M. Richardson, G. D. van der Spuy, C. J. Lombard, T. C. Victor, and P. D. van Helden**. 2000. Mapping of IS*6110* flanking regions in clinical isolates of *M. tuberculosis* demonstrates genome plasticity. Mol.Microbiol. **37**:1405-1416.

22. **Western Cape Government.** Cape Gateway: easy access to government information and services. http://www.capegateway.gov.za . 2005.

23. **Weyer, K., P. Groenewald, M. Zwarenstein, and C. J. Lombard**. 1995. Tuberculosis drug resistance in the Western Cape. S.Afr.Med.J. **85**:499-504.

Chapter 4

Use of Spoligotyping for Accurate Classification of Recurrent Tuberculosis

R. M. Warren,[1] E. M. Streicher,[1] S. Charalambous,[2] G. Churchyard,[2] G. D. van der Spuy,[1]A. D. Grant,[3] P. D. van Helden,[1] and T. C. Victor[1]

*MRC Centre for Molecular and Cellular Biology, Department of Medical Biochemistry, Faculty of Health Sciences, University of Stellenbosch, Tygerberg 7505,*1 *and Aurum Health Research, Welkom,*2 *South Africa, and Department of Infectious andTropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom*3

Published in Journal of Clinical Microbiology October 2002;

volume 40, nr. 10 page 3851-3

My contribution to this project:
Planning of project
Spoligotyping of isolates
Comparison of spoligotypes and interpretation
Writing and editing of manuscript

ABSTRACT

The spoligotyping method has become an important tool for the tracking of *Mycobacterium tuberculosis* strains in different epidemiological settings. In this study, we demonstrate the ability of the spoligotyping technique to accurately determine the pathogenetic mechanism of recurrent disease. This methodology has advantages over conventional restriction fragment length polymorphism methods which may be useful in large-scale intervention studies.

TEXT

The direct repeat (DR) region in *Mycobacterium tuberculosis* is characterized by DR sequences interspersed with variable repeat sequences (8), which in combination have been termed DVRs (7). The DR region has been shown to be polymorphic in different clinical isolates of *M. tuberculosis* (8, 21). Polymorphism is generated by IS*6110* insertion, deletion of DVR units by homologous recombination, and duplication of DVR repeat units by strand slippage during DNA replication (5, 19). The nature of polymorphism has been used to genotypically classify clinical isolates by DR restriction fragment length polymorphism (DR-RFLP) to define epidemiological relationships (14, 17). More recently, a PCR-based method was developed which has been termed spoligotyping (7, 9). This method can simultaneously identify *M. tuberculosis* and *M. bovis* species as well as provide a genotypic classification (9). Spoligotyping detects the

presence (or absence) of 43 unique DVR repeat sequences by line-blot hybridization. Strains are differentiated on the basis of the presence of specific variable repeat sequences. Although it is widely accepted that the discriminatory power of the spoligotyping method is lower than that of the internationally standardized RFLP method (10), spoligotyping remains an important tool to genotype clinical isolates in different epidemiological settings (1, 3). These data have been used to monitor the spread of specific strains within defined geographical regions (15) and between different countries (16). The versatility of the spoligotyping method has also allowed certain non-epidemiological questions to be answered. The method was shown to be a useful tool for rapid identification of laboratory error and contamination (4, 12). When used in this context, the objective is to establish whether serial isolates collected from a single patient are genotypically identical or different. A finding of different genotypes is suggestive of contamination or laboratory error. Similarly, spoligotyping has been used to rapidly screen for genotypically identical isolates cultured on the same day, which is suggestive of laboratory contamination. The interpretation of all molecular epidemiological data is based on establishing whether a strain is identical to or different from other strains found within a study community. Strains with shared genotypes are thought to represent ongoing transmission, while strains with unique genotypes are thought to represent reactivation (13). When this principle is applied to serial isolates collected from a single patient, it is possible to relate the genotype of the infecting strain to the genotype of a strain from a prior episode of disease (20). In this context, relapse is defined as the endogenous reactivation of the initial strain and, therefore, the strain genotype for each episode is identical. In contrast, reinfection is defined as a subsequent episode of disease

resulting from exogenous infection with a different strain. Therefore, strains from the different episodes of disease will be genotypically different, provided that there is sufficient genotypic diversity within the *M. tuberculosis* population in the study setting. By investigating this important epidemiological question, it should be possible to determine the efficacy of treatment within different settings and also to identify factors influencing disease dynamics.

In this study, we have investigated whether the spoligotyping method is suitable for accurate identification of pathogenic mechanisms causing recurrent tuberculosis disease. Specimens from 38 patients with recurrent disease after curative treatment were investigated; these patients were drawn from a larger study of recurrent tuberculosis among gold miners in Free State, South Africa. All the patients completed treatment for the first disease episode under conditions of direct observed therapy. Eighteen patients were shown to be clinically cured, with tests showing both smear- and culture-negative results after 6 to 8 months of therapy, while isolates from the remaining 20 patients were smear negative at completion of therapy. On subsequent presentation, tuberculosis disease was diagnosed using the standard case definitions (2). Isolates from each episode were cultured on Lowenstein-Jensen medium, and the cells were resuspended in water and stored at _70°C for periods of up to 3 years. DNA was isolated from subcultures grown on Lowenstein-Jensen medium and genotypically characterized by IS*6110* RFLP analysis using the internationally standardized protocol (18). The IS*6110* banding patterns from isolates representing each episode of disease were compared, and isolates having different banding patterns were defined as representing cases of reinfection, while

isolates having identical banding patterns were defined as representing cases of relapse (Fig. 1). In 16 patients, recurrence was classified as relapse, while in 21 patients, recurrence was classified as reinfection (Table 1). For patient 31, an RFLP result could not be obtained due to the presence of nontuberculous mycobacteria.

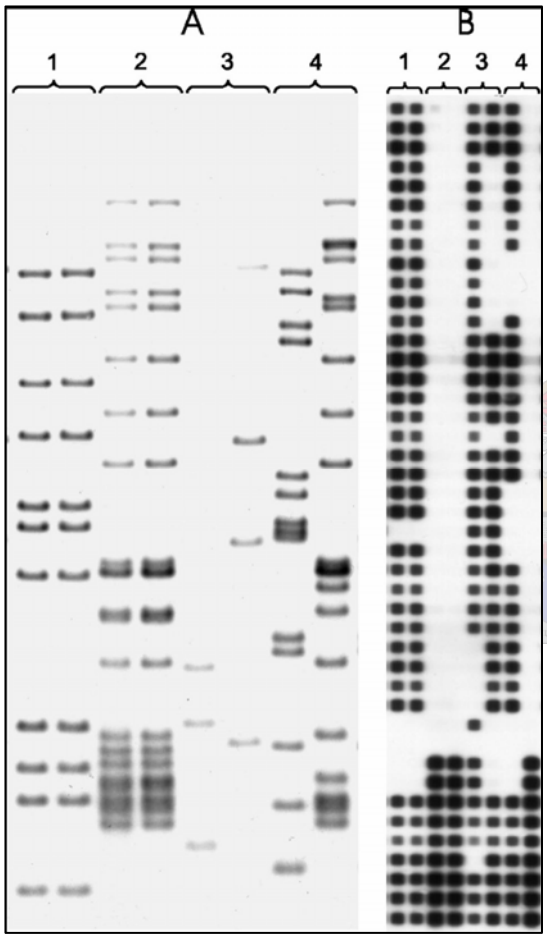


FIGURE. 1. IS*6110* RFLP and spoligotype analysis of clinical isolates collected from different episodes of disease from patients with recurrent disease. (A) IS*6110* RFLP analysis. Pairs of lanes under a single lane number represent the isolates taken from a single patient but different episodes of disease. Lanes 1 and 2 show that the results for the two episodes represented in each pair of lanes are genotypically identical for each of those patients, demonstrating relapse, while lanes 3 and 4 show that the results for the two episodes represented in each pair of lanes are genotypically different for each of those patients, demonstrating reinfection. (B) Spoligotype analysis. Each pair of isolates represented corresponds to the isolates represented under the same lane number in panel A. Lanes 1 and 2 show that the results for the two episodes represented in each pair of lanes are genotypically identical for each of those patients, demonstrating relapse, while lanes 3 and 4 show that the results for the two episodes represented in each pair of lanes are genotypically different for each of those patients, demonstrating reinfection.

The DNA samples were then subjected to spoligotyping in a different laboratory. The recipient laboratory was blinded to the RFLP results. Spoligotyping was done using the internationally standardized method (9). All PCRs were prepared in specially designated areas (in laminar flow hoods) to prevent contamination by amplicons. Each series of

PCRs included water blanks to identify possible reagent contamination. In addition, H37Rv DNA was included as a reference. No amplicon contamination was detected in any of the PCRs, and the H37Rv spoligotype pattern was consistent for each blot and matched the previously reported pattern (19), demonstrating repeatability of the technique. Figure 1 shows an example of the spoligotype pattern for isolates collected from patients with recurrent disease and the classification of relapse and reinfection.

From Table 1 it can be seen that in 36 of 37 (97%) cases, the spoligotype-defined mechanism of recurrence corresponded exactly to that obtained from IS*6110* RFLP data. The results differed in only one case (patient 3), where IS*6110* RFLP analysis had demonstrated an underlying dual infection in the initial infection. This resulted in the detection of a different spoligotype pattern in the second episode of disease. In the remaining patient (patient 31), IS*6110* RFLP data were unable to provide a definition of the mechanism of recurrence, although the spoligotyping method classified the mechanism of recurrence as reinfection. A similar result was obtained when the DNA from isolates from two of the patients was spoligotyped (data not shown) and an RFLP classification could be obtained only after reculturing and DNA extraction. This demonstrates that the PCR-based spoligotype method is sufficiently robust to allow the amplification of the DR region even when there is insufficient DNA for RFLP analysis or the DNA is of a poor quality.

TABLE 1. Genotypic classification of the pathogenic mechanism of recurrent *M. tuberculosis* disease by using IS*6110* RFLP analysis and spoligotyping

Patient	IS*6110* RFLP		Spoligotyping	
	Pattern	**Mechanism**	**Pattern**	**Mechanism**
1	Different	Reinfection	Different	Reinfection
2	Different	Reinfection	Different	Reinfection
3	Same	Relapse/dual	Different	Reinfection
4	Different	Reinfection	Different	Reinfection
5	Different	Reinfection	Different	Reinfection
6	Different	Reinfection	Different	Reinfection
7	Same	Relapse	Same	Relapse
8	Same	Relapse	Same	Relapse
9	Different	Reinfection	Different	Reinfection
10	Same	Relapse	Same	Relapse
11	Different	Reinfection	Different	Reinfection
12	Different	Reinfection	Different	Reinfection
13	Different	Reinfection	Different	Reinfection
14	Different	Reinfection	Different	Reinfection
15	Same	Relapse	Same	Relapse
16	Evolved*a*	Relapse	Same	Relapse
17	Same	Relapse	Same	Relapse
18	Same	Relapse	Same	Relapse
19	Different	Reinfection	Different	Reinfection
20	Different	Reinfection	Different	Reinfection
21	Different	Reinfection	Different	Reinfection
22	Same	Relapse	Same	Relapse
23	Same	Relapse	Same	Relapse
24	Different	Reinfection	Different	Reinfection
25	Same	Relapse	Same	Relapse
26	Different	Reinfection	Different	Reinfection
27	Same	Relapse	Same	Relapse
28	Different	Reinfection	Different	Reinfection
29	Different	Reinfection	Different	Reinfection
30	Different	Reinfection	Different	Reinfection
31	No RFLP	Unknown	Different	Reinfection
32	Same	Relapse	Same	Relapse
33	Different	Reinfection	Different	Reinfection
34	Same	Relapse	Same	Relapse
35	Same	Relapse	Same	Relapse
36	Different	Reinfection	Different	Reinfection
37	Different	Reinfection	Different	Reinfection
38	Evolved*a*	Relapse	Same	Relapse

[a] Paired isolates show only minor changes in the IS*6110* banding, suggesting evolution by transposition.

From the above results, the sensitivity and specificity of the spoligotyping method were calculated to be 100% and 94% (95% confidence interval, 82 to 100%), respectively. This suggests that the spoligotyping results correlate well to the standardized RFLP results when used to determine the pathogenetic mechanism of recurrent disease. This correlation demonstrates that there is sufficient genetic diversity within the DR locus of unrelated isolates (independent of IS*6110* copy number) to allow the differentiation of strains. Furthermore, it is unlikely that such diversity is generated during persistent infection, as previous studies have shown that the stability of the spoligotype is higher than that of the IS*6110* banding pattern (11). However, it is also acknowledged that this method is unable to distinguish dual infection in the first episode from relapse as a result of reactivation of only one of the initial infecting strains. As such dual infections are rare, it is unlikely that this will lead to a significant overestimate of the extent of reinfection. The spoligotyping method has numerous advantages over the standardized RFLP genotyping method in that it is better suited to high-throughput screening of patients in large intervention studies where relapse of tuberculosis must be distinguished from reinfection—for example, in evaluating different antituberculosis drug regimens. It has also been shown that this method can be adapted to genotypically classify strains directly from sputum (6). This would eliminate the need for the subculturing of each isolate for DNA isolation, thereby reducing the overall cost and exposure of laboratory workers to the pathogen. Furthermore, this method should greatly reduce the chance of laboratory contamination, which can result in the incorrect classification of the mechanism of recurrence.

We thank Tygerberg Hospital, the Harry Crossley Foundation, IAEA (projects SAF6/003 and CRP 9925), and Anglogold for financial assistance We thank V. Moloi and D. Sehloho for the preparation and subculturing of tuberculosis specimens and M. Lekitlane for the maintenance of the tuberculosis database.

REFERENCES

1. **Bauer, J., A. B. Andersen, K. Kremer, and H. Miorner.** 1999. Usefulness of spoligotyping to discriminate IS*6110* low-copy-number *Mycobacterium tuberculosis* complex strains cultured in Denmark. J. Clin. Microbiol. **37:**2602–2606.

2. **Corbett, E. L., G. J. Churchyard, T. Clayton, P. Herselman, B. Williams, R. Hayes, D. Mulder, and K. M. De Cock.** 1999. Risk factors for pulmonary mycobacterial disease in South African gold miners. A case-control study. Am. J. Respir. Crit. Care Med. **159:**94–99.

3. **Cronin, W. A., J. E. Golub, L. S. Magder, N. G. Baruch, M. J. Lathan, L. N. Mukasa, N. Hooper, J. H. Razeq, D. Mulcahy, W. H. Benjamin, and W. R. Bishai.** 2001. Epidemiologic usefulness of spoligotyping for secondary typing of *Mycobacterium tuberculosis* isolates with low copy numbers of IS*6110*. J. Clin. Microbiol. **39:** 3709–3711.

4. **de C. Ramos, M., H. Soini, G. C. Roscanni, M. Jaques, M. C. Villares, and J. M. Musser.** 1999. Extensive cross-contamination of specimens with *Mycobacteriumtuberculosis* in a reference laboratory. J. Clin. Microbiol. **37:**916–919.

5. **Fang, Z., N. Morrison, B. Watt, C. Doig, and K. J. Forbes.** 1998. IS*6110*transposition and evolutionary scenario of the direct repeat locus in a group of closely related *Mycobacterium tuberculosis* strains. J. Bacteriol. **180:**2102–2109.

6. **Goyal, M., S. Lawn, B. Afful, J. W. Acheampong, G. Griffin, and R. Shaw.** 1999. Spoligotyping in molecular epidemiology of tuberculosis in Ghana. J. Infect. **38:**171–175.

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

8. **Hermans, P. W., D. van Soolingen, E. M. Bik, P. E. de Haas, J. W. Dale, and J. D. van Embden.** 1991. Insertion element IS*987* from *Mycobacterium bovis* BCG is located in a hot-spot integration region for insertion elements in *Mycobacterium tuberculosis* complex strains. Infect. Immun. **59:**2695–2705.

9. **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.

10. **Kremer, K., D. van Soolingen, R. Frothingham, W. H. Haas, P. W. Hermans, C. Martin, P. Palittapongarnpim, B. B. Plikaytis, L. W. Riley, M. A. Yakrus, J. M.**

Musser, and J. D. van Embden. 1999. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. J. Clin. Microbiol. **37:**2607–2618.

11. **Niemann, S., E. Richter, and S. Rusch-Gerdes.** 1999. Stability of *Mycobacterium tuberculosis* IS*6110* restriction fragment length polymorphism patterns and spoligotypes determined by analyzing serial isolates from patients with drug-resistant tuberculosis. J. Clin. Microbiol. **37:**409–412.

12. **Nivin, B., J. Driscoll, T. Glaser, P. Bifani, and S. Munsiff.** 2000. Use of spoligotype analysis to detect laboratory cross-contamination. Infect. Control Hosp. Epidemiol. **21:**525–527.

13. **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.

14. **Sola, C., L. Horgen, K. S. Goh, and N. Rastogi.** 1997. Molecular fingerprinting of *Mycobacterium tuberculosis* on a Caribbean island with IS*6110* and DRr probes. J. Clin. Microbiol. **35:**843–846.

15. **Sola, C., A. Devallois, L. Horgen, J. Maisetti, I. Filliol, E. Legrand, and N. Rastogi.** 1999. Tuberculosis in the Caribbean: using spacer oligonucleotide typing to understand strain origin and transmission. Emerg. Infect. Dis. **5:**404–414.

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

17. **Torrea, G., G. Levee, P. Grimont, C. Martin, S. Chanteau, and B. Gicquel.** 1995. Chromosomal DNA fingerprinting analysis using the insertion sequence IS*6110* and the repetitive element DR as strain-specific markers for epidemiological study of tuberculosis in French Polynesia. J. Clin. Microbiol. **33:**1899–1904.

18. **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.

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

20. **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.

21. **van Soolingen, D., P. E. de Haas, J. Haagsma, T. Eger, P. W. Hermans, V. Ritacco, A. Alito, and J. D. van Embden.** 1994. Use of various genetic markers in differentiation of *Mycobacterium bovis* strains from animals and humans and for studying epidemiology of bovine tuberculosis. J. Clin. Microbiol. **32:**2425–2433.

Chapter 5

Patients with Active Tuberculosis often Have Different Strains in the Same Sputum Specimen

Robin M. Warren, Thomas C. Victor, Elizabeth M. Streicher, Madalene Richardson, Nulda Beyers, Nicolaas C. Gey van Pittius, and Paul D. van Helden

Medical Research Council Centre for Molecular and Cellular Biology, Department of Medical Biochemistry, and Centre for Tuberculosis Research and Education, Department of Child Health and Paediatrics, Faculty of Health Sciences, Stellenbosch University, Tygerberg, South Africa

Published in American Journal of Respiratory and Critical Care Medicine Vol 169. pp 610–614, 2004

My contribution to this project:

Provision of well characterized standard strains
Spoligotyping of isolates
Comparison and of spoligotypes and interpretation
Development of the method
Writing and editing of manuscript

ABSTRACT

It is generally accepted that tuberculosis results from a single infection with a single *Mycobacterium tuberculosis* strain. Such infections are thought to confer protective immunity against exogenous reinfection. In this study, a novel polymerase chain reaction method was developed to specifically identify *M. tuberculosis* strains belonging to the Beijing and non-Beijing evolutionary lineages in sputum specimens collected from tuberculosis patients resident in an epidemiologic field site in Cape Town, South Africa. The sensitivity and specificity of the polymerase chain reaction–based strain classification method were 100% (95% confidence interval, 85–100%) when compared with DNA fingerprinting and spacer oligotyping (spoligotyping). Application of this method showed that 19% of all patients were simultaneously infected with Beijing and non-Beijing strains, and 57% of patients infected with a Beijing strain were also infected with a non-Beijing strain. Multiple infections were more frequent in retreatment cases (23%) as compared with new cases (17%), but were not associated with sex, age, or smear grading. These results suggest that multiple infections are frequent, implying high reinfection rates and the absence of efficient protective immunity conferred by the initial infection. This finding could influence our understanding of the epidemiology of disease in high-incidence regions and our understanding for vaccine development.

INTRODUCTION

Active tuberculosis is thought to develop as a continuation of the primary infection (primary tuberculosis), or after endogenous reactivation of the primary infection or exogenous reinfection with a second *Mycobacterium tuberculosis* strain (1). Understanding the relative contribution of each of these mechanisms will have important implications for prevention of the development of new cases (2), evaluation of new drugs (3), interpretation of molecular epidemiologic data (4), as well as for the design and evaluation of protective and therapeutic vaccines (3, 5). The relative quantification of exogenous reinfection and endogenous reactivation depends on our ability to accurately differentiate between strains of *M. tuberculosis*. The development of an internationally standardized DNA fingerprinting method (6) has enabled the genotypic classification of *M. tuberculosis* strains with a high level of sensitivity and specificity. Using this methodology, molecular epidemiologic studies have shown the presence of a single strain in most cultures collected from patients with tuberculosis (7, 8), thereby suggesting that disease is caused by a single strain (infection). In contrast, we and others have shown that both human immunodeficiency virus (HIV)– negative and HIV-positive individuals can be infected with more than one strain (*1*) during the same episode (multiple infection) (4, 8–10), (*2*) in different lesions (multiple infection) (11, 12), or (*3*) during successive episodes (reinfection) (3, 9, 13–15). However, multiple infections are rarely found when using the DNA fingerprinting method (7, 8) and therefore their significance remains unknown.

This study aimed to determine the extent of multiple *M. tuberculosis* infection in sputum specimens collected from new and retreatment tuberculosis cases, resident in a community with a high incidence of disease (16), using a polymerase chain reaction (PCR) method based on comparative genomic data. The results are discussed in the context of the bacterial population structure within the human host and how multiple infection could influence the interpretation of molecular epidemiologic data. These findings could have important implications for the future design of anti-tuberculosis vaccines, as well as vaccine and drug trials. This work has been presented at the Keystone Symposium, Taos (2003) (30) and in Prague (2003) (31).

METHODS

Patients and Specimens

During the period March 2000–June 2002, pretreatment sputum specimens were collected only from adult patients (≥ 15 years of age) diagnosed with smear-positive tuberculosis (17). All patients were resident in an epidemiologic field site in Cape Town, South Africa (16). Patient demographics including sex, age, and previous history of tuberculosis were recorded at diagnosis and stored in a database. HIV testing was not routinely done, although a recent survey of 366 new adult smear-positive tuberculosis patients at the field site showed that 10% were HIV co-infected. This study was approved by the Ethics Committee of Stellenbosch University (Tygerberg, South Africa).

Culturing of *M. tuberculosis*

Decontaminated sputum specimens were collected from the National Health Laboratory from both tuberculosis smear positive patients and smear and culture negative individuals (used as controls). Each specimen was centrifuged at 10000xg for 15 min to pellet any bacteria and the supernatant was discarded by gentle aspiration. The bacterial pellet was resuspended in 1 ml of BACTEC medium containing PANTA-Plus (2·4% (w/v)) and incubated with gentle shaking for 7 days at 37°C (without reading) to allow bacterial growth. Thereafter, the bacteria were pelleted by centrifugation at 10000xg for 15 min and the supernatant was aspirated leaving the bacterial cells in a final volume of 200 µl. The bacteria were killed and lysed by heating at 100°C for 20 min. This crude DNA preparation was stored at -20°C. To monitor for any possible cross-contamination within the laminar flow hood, aliquots of BACTEC media were co-processed with every batch of sputum specimens and then incubated at 37°C for 7 days. Thereafter these control samples were processed as described above.

Classification of *M. tuberculosis* Isolates

Sputum specimens were cultured in BACTEC medium (BD, Franklin Lakes, NJ) and crude DNA preparations were obtained after boiling (*see* online supplement for culture and DNA preparation method). To identify a strain belonging to the Beijing evolutionary lineage, the DNA was subjected to PCR amplification using HotStar *Taq* polymerase (Qiagen, Hilden, Germany) and overlapping primer sets complementary to the 3' end of the IS*6110* element and Rv2820 (primer set 1, TTCAACCATCGCCGCCTCTAC and CACCCTCTACTCTGC GCTTTG; primer set 2, ACCGAGCTGATCAAACCCG and

ATGG CACGGCCGACCTGAATGAACC) (*see* online supplement for PCR amplification conditions and fractionation of products). A positive amplification product of 393 base pairs (bp) and 239 bp, respectively, indicated the presence of an IS*6110* insertion in Rv2820 that is unique to the Beijing evolutionary lineage (18) (Figure 1 and Figures 2A and 2B).

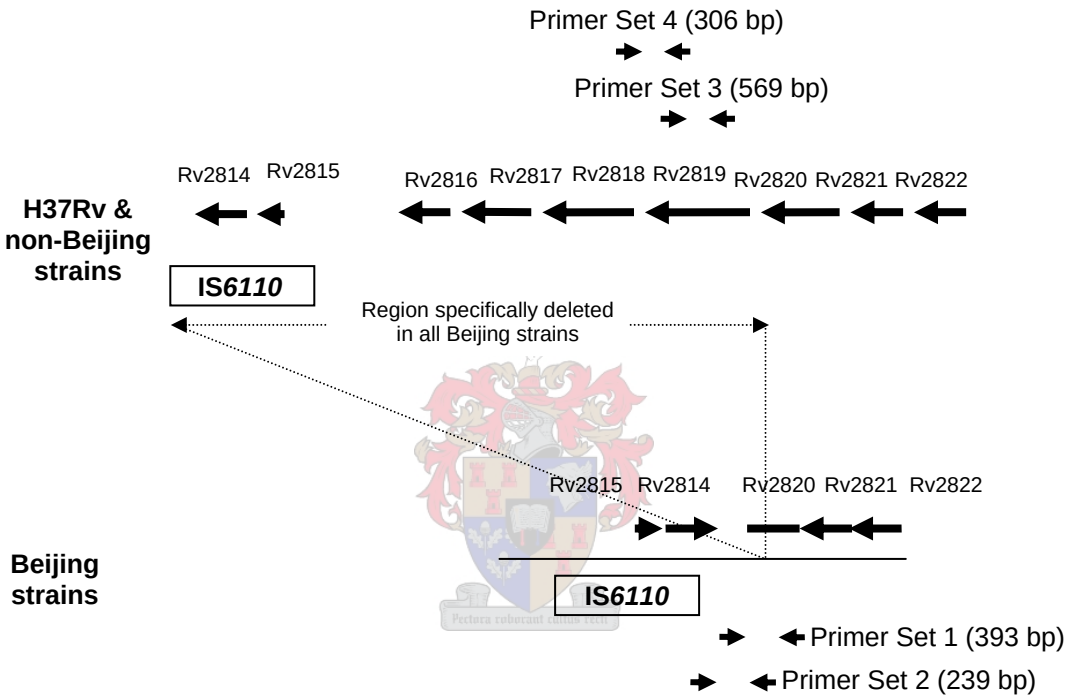


Figure 1. Schematic diagram of the region flanking the direct repeat (DR) region in H37Rv, non-Beijing and Beijing strains.
Large arrows indicate open reading frames according to the annotation of Cole and coworkers (29). The IS*6110* insertion in the DR region is indicated by a *square box* (note: IS*6110* is inverted in the Beijing strain family). The region specifically deleted in all Beijing strains is indicated by *dotted lines* (18, 19). Primer positions (sets 1, 2, 3, and 4) are indicated by *small arrows*.

To determine the presence of *M. tuberculosis* strains other than those belonging to the Beijing evolutionary lineage, the DNA was PCR amplified with overlapping primer sets complementary to Rv2819 (primer set 3, GATCGCTTGTTCTCAGTGCAG and CGAAGGAG TACCACGTGGAG; primer set 4, GGTGCGAGATTGAGGTT CCC and TCTACCTGCAGTCGCTTGTGC). A positive amplification product of 569 and 308 bp,

respectively, indicated the presence of *M. tuberculosis* strain(s) belonging to non-Beijing evolutionary lineages, as the region spanning the genes Rv2816 to Rv2819 (including part of Rv2820) is deleted in all Beijing strains (18, 19) (Figure 1 and Figures 2C and 2D).

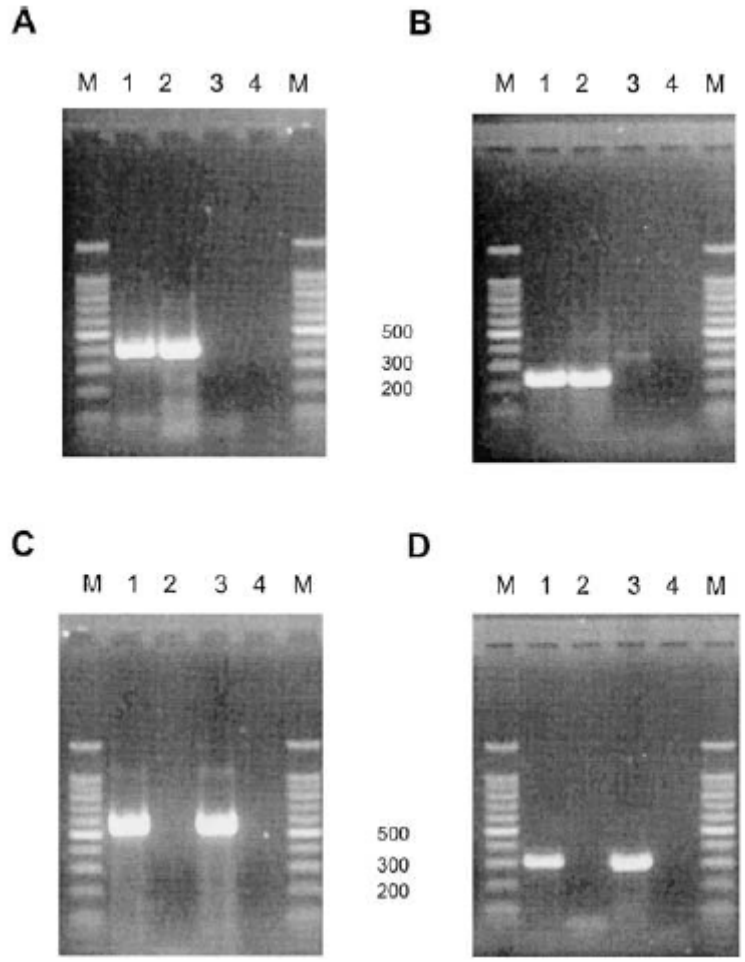


Figure 2. Fractionation of polymerase chain reaction (PCR) amplification products. (*A*) PCR amplification with primer set 1. (*B*) PCR amplification with primer set 2. (*C*) PCR amplification with primer set 3. (*D*) PCR amplification with primer set 4. DNA was visualized after staining with ethidium bromide. *Lanes 1* to *3*, clinical specimens; *lane 4*, H2O; *lane M*, 100-bp marker.

The sensitivity and specificity of amplification for the different primer sets was determined by PCR amplification of DNA templates from genetically distinct *M. tuberculosis* strains characterized by IS*6110* DNA fingerprinting (6) and spacer oligotyping (spoligotyping) (20). In addition, DNA templates from different

mycobacterial species were subjected to PCR amplification to further determine the specificity of the primer sets.

Smear-positive sputum specimens that showed positive amplification products only with primer sets 1 and 2 were classified as single Beijing infections (Figure 2, lane 2). Similarly, specimens that showed amplification products only with primer sets 3 and 4 were classified as single non-Beijing infections (Figure 2, lane 3). Specimens that showed positive amplification products with primer sets 1, 2, 3, and 4, were classified as multiple infections (Figure 2, lane 1). Specimens that showed discordant amplification were assigned as cross-contamination and excluded from the study.

To assess the extent of possible laboratory cross-contamination, control sputum specimens were concurrently collected from individuals who were smear and culture negative and who were resident in the same community. These specimens were incubated in BACTEC medium for 7 days and crude DNA preparations were obtained after boiling. In addition, aliquots of BACTEC medium were co-processed as a second group of negative controls.

PCR amplification conditions.

The PCR reactions were carried out in a total volume of 25 µl, containing 1 µl DNA template, 1 x enzyme buffer, 3.5 mM MgCl2, 4.0 mM dNTP's, 25 pmol of each primer and 0.5 U HotStarTaq DNA polymerase (Qiagen Germany). Amplification was initiated by incubation at 95 °C for 15 min. followed by 45 cycles at 94 °C for 1 min., 62 °C 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 1.5% agarose in 1xTBE pH 8·3 at 3·5V/cm for 4 hours, and visualized by staining with ethidium bromide. 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 conducted in physically separated rooms. Negative controls (water) were interleaved between specimens from smear positive patients to identify reagent or carry over contamination. In addition, negative control specimens from smear and culture negative individuals and BACTEC media controls were also interleaved between specimens from smear positive patients to identify cross-contamination. PCR amplification with overlapping primers sets was used to identify amplicon cross-contamination. The generation of a positive amplification product in any of these control specimens demonstrated the presence of *M. tuberculosis* DNA or PCR amplicons and thereby suggested cross-contamination. A positive control (Beijing strain, primer set 1 and 2; and *M. tuberculosis* H37Rv, primer set 3 and 4) was included to confirm amplification.

Spoligotyping

Spoligotyping was done according to the internationally standardized method (20).

Statistical Methods

The Fisher exact test was used to identify differences between patient demographics, as well as the frequency of multiple infections in new and retreatment cases of tuberculosis. Reproducibility of the PCR classification was calculated according to the Cohen kappa

method. Sensitivity and specificity was calculated with Prism software (GraphPad, San Diego, CA).

RESULTS

Description of Patients

During the period March 2000 to June 2002, 407 adults (≥ 15 years of age) resident within the epidemiologic field site (16) were diagnosed with smear-positive tuberculosis. The average age of these patients was 36.3 years and included 249 (61.2%) male patients and 158 (38.8%) female patients. According to the National Tuberculosis Control Program in line with the Directly Observed Therapy Short-course strategy, diagnosis of tuberculosis is made by sputum smear microscopy in new cases, and by smear microscopy and culture in retreatment cases. We endeavored to recover these sputum specimens from the National Health Laboratory after routine processing; however, some sputum specimens were not identified and were not available for research purposes. Pretreatment sputum specimens were available for analysis from 200 of the 407 patients. Their average age was 35.7 years and included 132 (66%) males and 68 (34%) females. This group of patients did not differ according to the demographic measures of age, sex, and previous history of tuberculosis when compared with patients for whom a sputum specimen was not available.

Validation of the PCR Method

To determine the sensitivity and specificity of the PCR amplification method as outlined in Figure 1, DNA from 59 *M. tuberculosis* cultures was subjected to amplification with primer sets 1, 2, 3, and 4. Primer sets 1 and 2 produced products on amplification of DNA templates (n = 30) only from strains with a characteristic Beijing IS*6110* banding pattern (21) and spoligotype (22). Conversely, primer sets 3 and 4 produced products on amplification of DNA templates (n = 29) only from strains with a non- Beijing genotype. The sensitivity and specificity of each primer set were 100% (95% confidence interval, 85–100%) when compared with the gold standard of the IS*6110* DNA fingerprint (21) and spoligotype (22). The specificity of amplification was further demonstrated by the absence of amplification products when DNA templates from *Mycobacterium avium*, *Mycobacterium abscessus*, *Mycobacterium kansasii*, *Mycobacterium malmoense*, *Mycobacterium perigrinum*, *Mycobacterium smegmatis*, and *Mycobacterium xenopi*, were amplified.

Detection of Multiple Infections

The PCR method was then applied to determine the presence of either Beijing or non-Beijing strains in the sputum specimens of the 200 included patients. PCR amplification of DNA prepared from short-term BACTEC cultures generated products in 192 of the specimens. The absence of amplification in eight specimens could be due either to the DNA concentration being below the limit of detection or to the presence of PCR inhibitors. PCR amplification was highly consistent and the Cohen kappa value for primer sets 1 and 2 and for primer sets 3 and 4 was calculated to be 0.977 and 0.944,

respectively. To avoid an overestimation of multiple infection, patient specimens that showed discordant results (n = 6) were excluded as possible cross-contamination.

Using primer sets 1 and 2, specimens from 61 of the 186 (32.8%) patients showed positive amplification products, demonstrating infections with strain(s) belonging to the Beijing evolutionary lineage (Figures 2A and 2B). PCR amplification using primer sets 3 and 4 demonstrated positive amplification products in 160 (86%) of the patient specimens (Figures 2C and 2D). Comparison between the results obtained with primer sets 1 and 2 and primer sets 3 and 4, showed that 35 (19%) of the specimens were positive for all the primer sets, demonstrating multiple *M. tuberculosis* infection (Table 1). This represents multiple infection in 57% of patients infected with strains belonging to the Beijing lineage. The remaining 26 patient specimens (14%) were positive for amplification with primer sets 1 and 2 only, demonstrating single infection with strains belonging to the Beijing lineage.

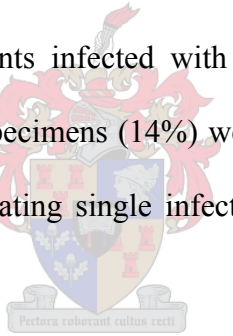


TABLE 1 Stratification of Patients according to Polymerase Chain Reaction-based classification of *Mycobacterium tuberculosis* strains present in the sputum culture

	Number (%)	Male (%)	Female (%)	Smear Grading	Case Definition	Number	Average Age (yr)	Number (<25yr)	Number (≥25yr)
All	186 (100)	126 (68)	60 (32)	S (12) 1 + (34) 2 + (45) 3 + (95)	New	124	35	37	151
					Retreatment	62	36.7		
Single infection	151 (81)	105 (69)	46 (31)	S (10) 1 + (30) 2 + (34) 3 + (77)	New	103	35.8	27	124
					Retreatment	38	37.4		
Multiple infection	35 (19)	21 (60)	14 (40)	S (2) 1 + (4) 2 + (11) 3 + (18)	New	21	32.1	10	25
					Retreatment	14	37.7		

Definition of abbreviation: S = scanty (fewer than 10 bacilli per 100 fields).

Of the 35 specimens assigned as multiple infection by PCR amplification, 9 (26%) showed the presence of both Beijing and non-Beijing spoligotype patterns, confirming multiple infection. Eleven (31%) of these specimens showed only a Beijing spoligotype and 15 (43%) showed only a non-Beijing spoligotype. Of the 26 specimens that showed a single Beijing infection by PCR amplification, all were confirmed by spoligotyping and no underlying non-Beijing spoligotypes were observed. All of the 125 specimens assigned as non-Beijing infections by PCR amplification showed non-Beijing spoligotypes and no underlying Beijing spoligotype could be detected.

The demographics of the 35 patients co-infected with strains belonging to the Beijing and non-Beijing lineages were similar to that of the 151 patients assigned as having single infections (Table 1). Multiple infections were more frequently observed in retreatment cases (23%) as compared with new cases (17%); however, this trend was not significant (Fisher exact odds ratio, 1.4; 95% confidence interval, 0.7–3.1). No significant association could be demonstrated between multiple infection and sex, age, or smear grading (Table 1).

Detection of Laboratory Cross-contamination

To establish whether laboratory cross-contamination contributed to the classification of multiple infections, control samples were analyzed by PCR amplification to identify the presence of *M. tuberculosis* DNA. These control groups included (*1*) decontaminated sputum specimens, collected from individuals who were shown to be smear and culture

negative, and (*2*) BACTEC medium. Both control groups were co-processed during all manipulations in the laminar flow hood as well as during PCR amplification. PCR amplification of the smear- and culture negative group with primer sets 1, 2, 3, and 4 identified 6 of 160 positive specimens. This suggests that the extent of possible cross-contamination in sputum specimens was 3.8%, which is similar to that previously reported (3). No detectable laboratory cross-contamination was observed by PCR amplification of the co-processed negative-control BACTEC medium (n = 37). Of the 125 water control samples included in the study, all were PCR negative, demonstrating the absence of reagent contamination.

DISCUSSION

Over the past decade, molecular genotyping methods have highlighted extensive genotypic heterogeneity among clinical isolates of *M. tuberculosis*, making accurate classification of the different disease-causing strains a complex science (6). Many of these chromosomal polymorphisms have been analyzed, including strain-specific IS*6110* insertion polymorphisms (23, 24) and chromosomal deletions (25–27), to identify possible correlates between genotype and phenotype. In this study we have used these genetic data to describe a method to classify clinical isolates according to the presence or absence of chromosomal markers unique to defined evolutionary lineages. Using this method, it was possible to investigate the epidemiologic phenomenon of multiple

infections in sputum cultures originating from new and retreatment tuberculosis cases resident in an epidemiologic field site in Cape Town, South Africa (16).

This study demonstrates that 19% of the patients included in the study were simultaneously infected with strain(s) belonging to the Beijing and non-Beijing evolutionary lineages. However, among patients infected with a Beijing strain, 57% were also infected with a strain(s) belonging to the non-Beijing evolutionary lineage. This PCR-based estimate of multiple infection was substantially higher than the 4.8% detected by spoligotyping (this study) and the 2.3% reported using DNA fingerprinting (8).

Extensive evaluation of a series of control samples showed that laboratory cross-contamination is an unlikely explanation for the high frequency of multiple infections, although we acknowledge that the observed level of cross-contamination may lead to an overestimate of multiple infection. However, our estimate is also limited by the inability of the method to identify multiple infections with different Beijing strains or different non- Beijing strains. Therefore we conclude that this study represents a conservative estimate of the frequency of multiple infections in the study setting. We acknowledge that the frequency of multiple infection may be vastly different in different settings, as well as in patients with either extra-pulmonary or smear-negative tuberculosis. The high number of multiple infection cases seen in this setting is unlikely to be due to HIV-induced immune deficiency given the low prevalence of HIV and tuberculosis co-infection in the study setting. However, in the absence of HIV testing of all patients, we cannot exclude

HIV as a mechanism enabling multiple infection. In addition, we cannot exclude other factors that may alter the immunity of the patients, such as alcoholism and malnutrition.

The occurrence of multiple infections in 17% of the new tuberculosis cases implies that reinfection in the study setting is extremely high. These multiple infections may occur when both infecting strains present to a "naive" immune response and thereby escape killing (11). An alternative possibility is that of "superinfection." In such cases, we speculate that an ongoing tuberculosis infection may significantly divert the immune response, thereby increasing the overall susceptibility to reinfection. Alternatively, reinfection occurring some time after the initial infection may initiate disease progression and the endogenous reactivation of the primary infection (12). The latter will imply that the primary infection is unable to confer protection against a secondary infection. A similar conclusion was drawn from mouse model experiments after repeated infection with different *M. tuberculosis* strains (28).

The higher proportion of multiple infections in retreatment cases supports previous observations of the importance of reinfection in recurrent tuberculosis (3, 14). However, interpretation of restriction fragment-length polymorphism data in the previous studies depended on the assumption that a patient is infected only with a single strain during each episode of disease. The identification of multiple infections in both new and retreatment cases implies that reinfection studies need to be reevaluated with methodologies that can accurately determine the strain population structure present during each episode. Such

studies will allow for a more accurate quantification of the sterilization efficacy of current and new anti-tuberculosis therapies.

From the data presented in this study it is not possible to predict the order in which the different infections occurred. Therefore, it would be unwise to assume that the re-infecting strain will always be overrepresented during disease progression. As described above, reinfection may reactivate a latent infection, which in turn may then be responsible for disease progression. If this scenario is true, it will have important consequences for the interpretation of molecular epidemiologic data, as it is possible that such cases will present with strains that are different from their source cases even though contact existed (4). Alternatively, multiply infected source cases may infect contacts with an underlying strain (not detected by IS*6110* DNA fingerprinting), thereby making the inference of contact difficult when using only restriction fragment-length polymorphism data.

This study demonstrates that multiple infections are present in patients with active tuberculosis in a high-incidence setting. Most importantly, the initial infection is unable to provide protection against a subsequent infection in this population. This result will have important implications for the understanding of protective immunity and the development and testing of new vaccines and drugs for use in communities where the burden of disease is high. Furthermore, this study highlights the importance of preventing transmission to reduce the risk of exposure and re-exposure of all people to active sources of tuberculosis.

ACKNOWLEDGEMENTS

The authors thank the National Health Laboratory Service regional laboratory for providing access to specimens, and Mrs. L. Pretorius and Mrs. A. Huysamen for excellent technical support and preparation of specimens. The authors also thank Dr. I. Toms (Department of Health, City of Cape Town) and are indebted to the residents of the epidemiologic field site.

REFERENCES

1. Balasubramanian V, Wiegeshaus EH, Taylor BT, Smith DW. Pathogenesis of tuberculosis: pathway to apical localization. *Tuber Lung Dis* 1994; 75:168–178.

2. Mallory KF, Churchyard GJ, Kleinschmidt I, De Cock KM, Corbett EL. The impact of HIV infection on recurrence of tuberculosis in South African gold miners. *Int J Tuberc Lung Dis* 2000;4:455–462.

3. van Rie A, Warren R, Richardson M, Victor TC, Gie RP, Enarson DA, Beyers N, van Helden PD. Exogenous reinfection as a cause of recurrent tuberculosis after curative treatment. *N Engl J Med* 1999;341: 1174–1179.

4. Braden CR, Morlock GP,Woodley CL, Johnson KR, Colombel AC,Cave MD, Yang ZH, Valway SE, Onorato IM, Crawford JT. Simultaneous infection with multiple strains of *Mycobacterium tuberculosis. Clin Infect Dis* 2001;33:E42–E47.

5. Fine PE, Small PM. Exogenous reinfection in tuberculosis. *N Engl J Med* 1999;341:1226–1227.

6. van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, Hermans P, Martin C, McAdamR, Shinnick TM. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* 1993;31:406–409.

7. de Boer AS, Kremer K, Borgdorff MW, de Haas PE, Heersma HF, van Soolingen D. Genetic heterogeneity in *Mycobacterium tuberculosis* isolates reflected in IS*6110* restriction fragment length polymorphism patterns as low-intensity bands. *J Clin Microbiol* 2000;38:4478–4484.

8. Richardson M, Carroll NM, Engelke E, van der Spuy GD, Salker F, Munch Z, Gie RP, Warren RM, Beyers N, van Helden PD. Multiple *Mycobacterium tuberculosis* strains in early cultures from patients in a high-incidence community setting. *J Clin Microbiol* 2002;40:2750–2754.

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

10. Niemann S, Richter E, Rusch-Gerdes S, Schlaak M, Greinert U. Double infection with a resistant and amultidrug-resistant strain of *Mycobacterium tuberculosis. Emerg Infect Dis* 2000;6:548–551.

11. Bates JH, Stead WW, Rado TA. Phage type of tubercle bacilli isolated from patients with two or more sites of organ involvement. *Am Rev Respir Dis* 1976;114:353–358.

12. du Plessis DG, Warren R, Richardson M, Joubert JJ, van Helden PD. Demonstration of reinfection and reactivation inHIV-negative autopsied cases of secondary tuberculosis: multilesional genotyping of *Mycobacterium tuberculosis* utilizing IS*6110* and other repetitive element-based DNA fingerprinting. *Tuberculosis (Edinb)* 2001;81:211–220.

13. Small PM, Shafer RW, Hopewell PC, Singh SP, Murphy MJ, Desmond E, Sierra MF, Schoolnik GK. Exogenous reinfection with multidrugresistant *Mycobacterium tuberculosis* in patients with advanced HIV infection. *N Engl J Med* 1993;328:1137–1144.

14. Caminero JA, Pena MJ, Campos-Herrero MI, Rodriguez JC, Afonso O, Martin C, Pavon JM, Torres MJ, Burgos M, Cabrera P, *et al.* Exogenous reinfection with tuberculosis on a European island with a moderate incidence of disease. *Am J Respir Crit Care Med* 2001;163:717–720.

15. Kruuner A, Pehme L, Ghebremichael S, Koivula T, Hoffner SE, Mikelsaar M. Use of molecular techniques to distinguish between treatment failure and exogenous reinfection with *Mycobacterium tuberculosis. Clin Infect Dis* 2002;35:146–155.

16. Beyers N, Gie RP, Zietsman HL, Kunneke M, Hauman J, Tatley M, Donald PR. The use of a geographical information system (GIS) to evaluate the distribution of tuberculosis in a high-incidence community. *S Afr Med J* 1996;86:40–144.

17. Maher D, Chaulet P, Spinaci S, Harries A. Treatment of tuberculosis: guidelines for national programs, 2nd ed. Geneva, Switzerland:World Health Organization; 1997.

18. Warren RM, Streicher EM, Sampson SL, van der Spuy GD, Richardson M, Nguyen D, Behr MA, Victor TC, van Helden PD. Microevolution of the direct repeat region of *Mycobacterium tuberculosis*: implications for interpretation of spoligotyping data. *JClin Microbiol* 2002;40:4457– 4465.

19. van Rie A, Warren RM, Beyers N, Gie RP, Classen CN, Richardson M, Sampson SL, Victor TC, van Helden PD. Transmission of a multidrugresistant *Mycobacterium tuberculosis* strain resembling "strain W" among noninstitutionalized, human immunodeficiency virus-seronegative patients. *J Infect Dis* 1999;180:1608–1615.

20. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, *et al.* Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol* 1997;35: 907–914.

21. Richardson M, van Lill SW, van der Spuy GD, Munch Z, Booysen CN, Beyers N, van Helden PD, Warren RM. Historic and recent events contribute to the disease dynamics of Beijing-like *Mycobacterium tuberculosis* isolates in a high incidence region. *Int J Tuberc Lung Dis* 2002;6:1001–1011.

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

23. Beggs ML, Eisenach KD, Cave MD. Mapping of IS*6110* insertion sites in two epidemic strains of *Mycobacterium tuberculosis. J Clin Microbiol* 2000;38:2923–2928.

24. Sampson S, Warren R, Richardson M, van der Spuy G, van Helden P. IS*6110* insertions in *Mycobacterium tuberculosis*: predominantly into coding regions. *J Clin Microbiol* 2001;39:3423–3424.

25. Fang Z, Doig C, Kenna DT, Smittipat N, Palittapongarnpim P, Watt B, Forbes KJ. IS*6110*-mediated deletions of wild-type chromosomes of *Mycobacterium tuberculosis. J Bacteriol* 1999;181:1014–1020.

26. Ho TB, Robertson BD, Taylor GM, Shaw RJ, Young DB. Comparison of *Mycobacterium tuberculosis* genomes reveals frequent deletions in a 20 kb variable region in clinical isolates. *Yeast* 2000;17:272–282.

27. Warren RM, Sampson SL, Richardson M, van der Spuy GD, Lombard CJ, Victor TC, van Helden PD. Mapping of IS*6110* flanking regions in clinical isolates of *M. tuberculosis* demonstrates genome plasticity. *Mol Microbiol* 2000;37:1405–1416.

28. Repique CJ, Li A, Collins FM, Morris SL. DNA immunization in a mouse model of latent tuberculosis: effect of DNA vaccination on reactivation of disease and on reinfection with a secondary challenge. *Infect Immun* 2002;70:3318–3323.

29. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE III, *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 1998;393:537–544.

30. Warren RM, Uys P, Victor T, Gey van Pittius NC, Richardson M, Streicher E, Jordaan A, Beyers N, Van Helden PD. Multiple infection is common in a high incidence area: what does it imply? Presented at Keystone Symposium on tuberculosis: integrating host and pathogen biology, Taos, NM, 2003.

31. Warren RM, Victor TC, Streicher EM, Richardson M, Beyers N, Gey van Pittius NC, van Helden PD. Patients with active tuberculosis often have different strains in the same sputum sample. Presented at Third Meeting of Concerted Action Project: new generation genetic markers and techniques for the epidemiology of tuberculosis, Prague, Czech Republic, 2003.

Chapter 6

Clonal Expansion of a Globally Disseminated Lineage of *Mycobacterium tuberculosis* with Low IS*6110* Copy Numbers

R. M. Warren,[1] T. C. Victor,[1] E. M. Streicher,[1] M. Richardson,[1] G. D. van der Spuy,[1] R. Johnson,[1] V. N. Chihota,[1] C. Locht,[2] P. Supply,[2] and P. D. van Helden[1]

MRC Centre for Molecular and Cellular Biology, Department of Medical Biochemistry, Faculty of Health Sciences, University of Stellenbosch, Tygerberg, South Africa,[1] *and Laboratoire des Me´canismes Mole´culaires de la Pathogene`se Bacte´rienne, INSERM U629, Institut Pasteur de Lille, Lille, France*[2]

Published in Journal of Clinical Microbiology, Vol. 42, No. 12 Dec 2004, p5774-5782

My contribution to this project:
Spoligotyping of isolates
Description of different spoligotype characteristics
Comparison of spoligotypes
Writing and editing of manuscript

ABSTRACT

Knowledge of the clonal expansion of *Mycobacterium tuberculosis* and accurate identification of predominant evolutionary lineages in this species remain limited, especially with regard to low-IS*6110*-copy-number strains. In this study, 170 *M. tuberculosis* isolates with <6 IS*6110* insertions identified in Cape Town, South Africa, were characterized by principal genetic grouping, restriction fragment length polymorphism analysis, spoligotyping, IS*6110* insertion site mapping, and variable-number tandem repeat (VNTR) typing. These analyses indicated that all but one of the isolates analyzed were members of principal genetic group 2 and of the same low- IS*6110*-copy-number lineage. The remaining isolate was a member of principal genetic group 1 and a different low-IS*6110*-copy-number lineage. Phylogenetic reconstruction suggests clonal expansion through sequential acquisition of additional IS*6110* copies, expansion and contraction of VNTR sequences, and the deletion of specific direct-variable-repeat sequences. Furthermore, comparison of the genotypic data of 91 representative low-IS*6110*-copy-number isolates from Cape Town, other southern African regions, Europe, and the United States suggests that certain low-IS*6110*-copy-number strain spoligotypes and IS*6110* fingerprints were acquired in the distant past. These clones have subsequently become widely disseminated and now play an important role in the global tuberculosis epidemic.

INTRODUCTION

Sequence analysis of *Mycobacterium tuberculosis* strains collected from different geographical settings has shown that the frequency of mutation in this species is extremely low (31). Single-nucleotide polymorphism (SNP) and variable-number tandem repeat (VNTR)-based analyses have consistently indicated that the global population of *M. tuberculosis* is highly clonal (1, 15, 34). Within this clonal structure, strains can be assigned to one of three principal genetic groups according to SNPs in the *katG* and *gyrA* genes (31). Using an expanded array of synonymous SNPs (sSNPs), these principal genetic groups have been further divided into eight main clusters to depict the evolution of *M. tuberculosis* (15). Simulations have suggested that the branch points in the SNP-derived trees are accurate (1). However, subsequent clonal expansion occurring at each of these branch points remains largely unknown due to the lack of resolution of SNP-based analyses (1).

Various combinations of more variable markers, primarily used for molecular epidemiological studies, have been utilized to identify genotype families which fall into the different clusters defined by SNPs (15). The most commonly used marker is IS*6110*, a transposable element used as a probe in restriction fragment length polymorphism (RFLP) analysis of clinical isolates (35). Genetic relationships between strains have been inferred according to an IS*6110* RFLP Dice similarity index of >65% and the inheritance of other specific polymorphisms. These data have been applied to depict clonal expansion in high-copy-number strains with >6 IS*6110* copies (5, 37, 40, 42) and for their

epidemiological analysis on a global scale (4). However, only a limited amount of evolutionary data exists for low-IS*6110*-copy-number strains with ≤6 IS*6110* hybridizing bands due to their intrinsically limited IS*6110* RFLP polymorphism. One report has presented evidence to show evolution of the IS*6110* banding pattern in the progeny of a low-IS*6110*-copy-number strain (22). More recently, a high degree of congruence was shown to exist between IS*6110* banding patterns and other markers in strains with few IS*6110* copies collected in London, United Kingdom (7). Accordingly, these strains have been classified into three different groups (7). One of these groups (group II) includes the principal genetic group 2 clusters IV and V defined by sSNP analysis, representing strains with one to three and four to six IS*6110* copies, respectively (15). A second distinct group includes strains from principal genetic group 1 cluster I (15), which have been associated with patients from East Africa and Asia (6, 30). However, the process of clonal expansion within these groups of low-IS*6110*-copy-number strains remains largely unresolved. Moreover, the genetic relationships among strains within these groups from different geographical regions are poorly understood.

In this study, we used RFLP (42), principal genetic grouping (31), IS*6110* insertion site mapping (6), spoligotyping (19), and PCR analysis of VNTRs interspersed in multiple loci (14, 23, 24, 33) to determine the genetic relationships among low- IS*6110*-copy-number strains collected in Cape Town, South Africa. These data have been compared to available genotypic data from low-IS*6110*-copy-number strains isolated from other geographical areas in order to better define the evolution of low-IS*6110*-copy-number

strains and the impact of such evolution on the interpretation of molecular epidemiological data.

MATERIALS AND METHODS

Study setting. Between January 1992 and December 1998, *M. tuberculosis* isolates were obtained from patients residing in two suburbs of Cape Town, South Africa (3), as well as a subset of patients residing in the adjoining suburbs. The two suburbs have a population of 38,500 residents within an area of 3.4 km2 and have two health care clinics. In this setting, the average annual incidence of new bacteriologically confirmed tuberculosis cases (culture and/or smear positive) was 313/100,000 (38).

DNA fingerprinting. Genomic DNA from each isolate of *M. tuberculosis* was digested with either PvuII or HinfI, electrophoretically fractionated, and Southern transferred to Hybond N+ (Amersham, Little Chalfont, United Kingdom). The blots containing the PvuII-digested DNA were sequentially hybridized with probes labeled by the enhanced chemiluminescence method that were complementary to the 3' domain of the IS*6110* element (IS-3') (35), the 5' domain of the IS*6110* element (IS-5') (42), the direct repeat (DR) (42), and Marker X (Roche, Basel, Switzerland). Each probe was stripped by denaturation before the next probe was applied. The HinfI Southern blots were hybridized with the 32Plabeled MTB484(1) probe complementary to the polymorphic G+C-rich repeat sequences (PGRS) (41). The autoradiographs were normalized, and the IS-3', IS-5', and DR bands were assigned using GelCompar II software. Cluster analysis

was done using the unweighted pair group method with arithmetic mean and the Dice coefficient (17). Mutations in the IS*6110*-flanking domains were determined as previously described (43). The band corresponding to the IS*6110* insertion in the DR region was identified by aligning the DR and IS-3' autoradiographs (42). The blots probed by MTB484 (1) were visually analyzed by two independent persons (42).

PCR amplification was used to determine the presence or absence of an IS*6110* insertion in the genes Rv0403c, Rv1758, and Rv3018c according to the previously described method (6). IS*6110* insertion in the gene Rv2787c was determined using the primer set 5'-TTCAACCATCGCCGCCTCTAC-3' and 5'-GGCCAAATCCAGCACGGTGAAC-3'.

Mutation analysis. The *M. tuberculosis* isolates were classified into three principal genetic groups according to polymorphisms in the *katG* and *gyrA* genes (31), using the dot blot hybridization method (39).

Spoligotyping. DNA polymorphism in the DR locus was detected in isolates with ≤6 IS*6110* insertion elements by spoligotyping according to a standardized protocol (19).

MIRU-VNTR typing. *M. tuberculosis* isolates were genotyped by PCR amplification of the 12 loci containing VNTRs of elements called mycobacterial interspersed repetitive units (MIRUs) (33) and 9 loci containing VNTRs of other interspersed sequences (14, 21, 24; P. Supply, S. Lesjean, E. Savine, K. Kremer, D. van Soolingen, and C. Locht,

unpublished data) using both manual (33) and automated (32) techniques. The primers against the MIRUVNTR flanking regions were the same as those previously described (33), except that Hex labeling was replaced by Vic labeling. The primers against the other loci are described in Table 1. The samples were subjected to electrophoresis using a 96-well ABI 377 automated sequencer as previously described (32). Sizing of the PCR fragments and assignment of the various VNTR alleles were done using the GeneScan and Genotyper software packages (PE Applied Biosystems) as previously described (32) and based on the data in Table 1. The tables used for VNTR allele scoring are available at http://www.ibl.fr/mirus/mirus.html. Allele assignments in the manual and automated methods were identical.

TABLE 1. Conditions for multiplex PCRs of nine VNTR loci

Multiplex	Conventional designation [a]	VNTR length (bp)	MgCl2 (mM)	PCR primer pairs (5' to 3', with labeling indicated)
Mix E	VNTR2347	57	1.5	GCCAGCCGCCGTGCATAAACCT (FAM) AGCCACCCGGTGTGCCTTGTATGAC
	VNTR2461	57		ATGGCCACCCGATACCGCTTCAGT (VIC) CGACGGGCCATCTTGGATCAGCTAC
	VNTR3171	54		GGTGCGCACCTGCTCCAGATAA (NED) GGCTCTCATTGCTGGAGGGTTGTAC
Mix F	VNTR0424	51	1.5	CTTGGCCGGCATCAAGCGCATTATT GGCAGCAGAGCCCGGGATTCTTC (FAM)
	VNTR0577	58		CGAGAGTGGCAGTGGCGGTTATCT (VIC) AATGACTTGAACGCGCAAATTGTGA
	VNTR1895	57		GTGAGCAGGCCCAGCAGACT (NED) CCACGAAATGTTCAAACACCTCAAT
Mix G	VNTR2401	58	3.0	CTTGAAGCCCCGGTCTCATCTGT (FAM) ACTTGAACCCCACGCCCATTAGTA
	VNTR3690	58		CGGTGGAGGCGATGAACGTCTTC (VIC) TAGAGCGGCACGGGGGAAAGCTTAG
	VNTR4156	59		TGACCACGGATTGCTCTAGT GCCGGCGTCCATGTT (NED)

[a] VNTR 0577, 2461, 4156, and 1895 are described in reference 28 and correspond to ETRC, ETRB, and QUB 4156 and 1895, respectively (14, 25).

Global dissemination. To determine the geographical spread of the low- IS*6110*-copy-number strains, *M. tuberculosis* isolates collected from southern (Western Cape) (*n*=47),

central (Free State) (*n*=4), and northern (Gauteng and Mpumalanga) (*n*=4) regions in South Africa and from Harare and Gweru (*n*=11) in Zimbabwe were subjected to spoligotyping, IS*6110* RFLP, and IS*6110* insertion mapping (6). These genotypic data were compared to previously published genotypic data (IS*6110* banding patterns [visual comparison], IS6*110* insertion points, and spoligotype patterns) for low-IS*6110*-copy-number strains from Europe (United Kingdom [*n*=14] [7] and Denmark [*n*=2] [2]) and the United States (Michigan [*n*=70] [6] and CDC1551 [12]).

Genetic-relationship analysis. The evolutionary state(s) for the RFLP data was assigned according to the presence (indicated by 1) or the absence (indicated by 0) of a hybridizing band. Spoligotypes were assigned according to the presence or absence of spacer sequences, while the VNTR alleles were assigned according to the number of repeats present in the different loci. The complete set of evolutionary states for the different markers was subjected to phylogenetic analysis using the neighbor-joining algorithm Phylogenetic Analysis Using Parsimony (*Other Methods) version 4b10 (Sinauer Associates, Sunderland, Mass.). Bootstrapping was performed to establish a degree of statistical support for nodes within each phylogenetic reconstruction (10). A consensus tree was generated using the program contree [Phylogenetic Analysis Using Parsimony (*Other Methods) version 4b10] in combination with the majority rule formula. The resulting trees were rooted to the principal genetic group 1 isolate [SA CT(67)], a designation derived from the isolate's origin city of Cape Town, South Africa. Only branches which occurred in >50% of the bootstrap trees were included in the final tree, and all branches with a zero branch length were collapsed.

Selection and molecular characterization of strains. Between January 1992 and December 1998, *M. tuberculosis* isolates were obtained from 1,030 patients resident in the suburbs adjoining Cape Town, South Africa. IS*6110* RFLP analysis established that 186 (18.1%) of these patients were infected with a strain containing ≤6 IS*6110* hybridizing bands. No isolate lacking the IS*6110* element was identified in this study setting. Isolates were available from 170 (91%) of these patients for further genotypic analysis.

TABLE 2. Spoligotype patterns of Cape Town *M. tuberculosis* isolates with ≤6 IS*6110* elements

Spoligotype[a]	Octal code[b]	Spoligotype pattern[c]
1	777777777760771	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□□□■■■■■■■■
2	777776777760601	■■■■■■■■■■■■■■■■■■■■■■■■■□■■■■■■■■■■■■■□□□□■■□□□□■
3	777736777760601	■■■■■■■■■■■■■□■■■■□■■■■■■■■■■■■■■■□□□□■■□□□□■
4	677776777760601	■■□■■■■■■■■■■■■■■□■■■■■■■■■■■■■■■□□□□■■□□□□■
5	777776777760771	■■■■■■■■■■■■■■■■■■■■■■■■■□■■■■■■■■■■■■■□□□□■■■■■■■
6	777776777720601	■■■■■■■■■■■■■■■■■■■■■■■■■□■■■■■■■■■■■■□■□□□□■■□□□□■
7	737776777760601	■■■□■■■■■■■■■■■■■■■□■■■■■■■■■■■■■■■□□□□■■□□□□■
8	767776777760601	■■■■■□■■■■■■■■■■■■■□■■■■■■■■■■■■■■■□□□□■■□□□□■
9	777776777760711	■■■■■■■■■■■■■■■■■■■■■■■■■□■■■■■■■■■■■■■□□□□■■■□□■■
10	743776777760601	■■■■□□□■■■■■■■■■■■■■■■■■■□■■■■■■■■■■■■■□□□□■■□□□□■
11	777776777760731	■■■■■■■■■■■■■■■■■■■■■■■■■□■■■■■■■■■■■■■□□□□■■■□■■■
12	700076777760771	■■■□□□□□□□□□■■■■■□■■■■■■■■■■■■■■■□□□□■■■■■■■
13	777776777760740	■■■■■■■■■■■■■■■■■■■■■■■■■□■■■■■■■■■■■■■□□□□■■■■□□□
14	700076774360771	■■■□□□□□□□□□■■■■■□■■■■■■■□□□■■■■□□□□■■■■■■■
15	700076777740371	■■■□□□□□□□□□■■■■■□■■■■■■■■■■■■■■□□□□□■■■■■■■
16	700076777760671	■■■□□□□□□□□□■■■■■□■■■■■■■■■■■■■■■□□□□■■□■■■■
17	777776777560771	■■■■■■■■■■■■■■■■■■■■■■■■■□■■■■■■■■■□■■■□□□□■■■■■■■
18	777776617560771	■■■■■■■■■■■■■■■■■■■■■■■■■□■■□□□■■■■■□■■■□□□□■■■■■■■
19	757777777413731	■■■■□■■■■■■■■■■■■■■■■■■■■■□□□□■□■■■■■□■■■

[a] Spoligotypes are arbitrary designations (see Table 3).
[b] Spoligotype nomenclature according to reference 8.
[c] Closed boxes indicate the presence of DVR sequences, and open boxes indicate the absence of DVR sequences.

Analysis of the *katG* and *gyrA* gene sequences classified 169 of these isolates in principal genetic group 2, while only one isolate was classified as principal genetic group 1 (31). All isolates were then subjected to further analysis using Southern hybridization in combination with probes complementary to IS-3' (35) (Fig. 1) and IS-5' (42), spoligotyping (19) (Table 2), and VNTR allele typing based on 21 independent loci (14, 21, 25, 32, 33) and PGRS RFLP typing (41) (Table 3). This set included MIRU-VNTR loci 2, 4 (ETR-D), 10, 16, 20, 23, 24, 26, 27, 31 (ETR-E), 39, and 40 and VNTR loci 424, 577 (ETR-C), 1895 (QUB-1895), 2347, 2401, 2461(ETR-B), 3171, 3690, and 4156 (QUB-4156) (alias designations are in parentheses). In addition, IS*6110* insertion into the genes Rv0403c, Rv1758, Rv3018c, and Rv2787c was determined by PCR amplification (Fig. 1A). Table 3 summarizes the genotypic data of the different isolates as defined by the combined markers. For this data set, the numbers of principal genetic group 2 genotypes obtained with the different methods were ranked as follows: IS-3' (14 genotypes) < spoligotypes (19 genotypes) = IS-5' (19 genotypes) < IS-3' and IS-5' (22 genotypes) < VNTR loci (38 genotypes) < PGRS (45 genotypes). In combination, these different genotyping methods identified a total of 66 distinct principal genetic group 2 genotypes (Table 3).

TABLE 3. Genotypic classification of Cape Town *M. tuberculosis* isolates with _6 IS*6110* elements

Name	Principal genetic group	IS*6110* Copy no.	IS-3' type	IS-5' type	PGRS type	12 MIRU-VNTR type (allele combination)[a]	9 VNTR allele combination[b]	Spoligotype (octal format)[c]
SA CT(1)	2	1	1	1	1	1 (225125113322)	144442353	1 (777777777760771)
SA CT(2)	2	2	2	2	2	2 (223325153323)	142442383	2 (777776777760601)
SA CT(3)	2	2	2	2	3	3 (224325123422)	242442343	3 (777736777760601)
SA CT(4)	2	2	2	2	3	3 (224325123422)	242442343	2 (777776777760601)
SA CT(5)	2	2	2	2	4	4 (224325143223)	ND	3 (777736777760601)
SA CT(6)	2	2	2	2	4	4 (224325143223)	442442333	2 (777776777760601)
SA CT(7)	2	2	2	2	5	5 (224325143324)	242442343	4 (677776777760601)
SA CT(8)	2	2	2	2	5	5 (224325143324)	242442343	3 (777736777760601)
SA CT(9)	2	2	2	2	5	5 (224325143324)	242442343	2 (777776777760601)
SA CT(10–11)	2	2	2	2	6, 7	6 (224325153223)	442442431	2 (777776777760601)
SA CT(12)	2	2	2	2	2	7 (224325153323)	142442383	2 (777776777760601)
SA CT(13)	2	2	2	2	2	7 (224325153323)	142442383	5 (777776777760771)
SA CT(14)	2	3	3	3[d]	8	8 (223325143323)	242442?3	6 (777776777720601)
SA CT(15)	2	3	3	4	9	9 (224325133324)	442442333	7 (737776777760601)
SA CT(16)	2	3	3	4	9	10 (224325153323)	442442333	7 (737776777760601)
SA CT(17)	2	3	4	5	10	11 (223325163322)	242442333	8 (767776777760601)
SA CT(18)	2	3	4	5	10	12 (224325163322)	242442333	8 (767776777760601)
SA CT(19)	2	3	5	6	11	13 (224325153222)	522442332	9 (777776777760711)
SA CT(20)	2	3	5	7[d]	12	14 (224325153324)	252441343	10 (743776777760601)
SA CT(21)	2	3	6[d]	5	13	15 (223325153223)	442442333	2 (777776777760601)
SA CT(22)	2	3	6[d]	5	14	10 (224325153323)	442442333	2 (777776777760601)
SA CT(23)	2	3	6[d]	5	14	16 (224325153323)	452442333	2 (777776777760601)
SA CT(24)	2	3	7	8	15	17 (223325163433)	ND	2 (777776777760601)
SA CT(25)	2	4	8	9	16	18 (223325143322)	432442333	5 (777776777760771)
SA CT(26–30)	2	4	8	9	16, 17, 18, 19, 20	19 (223325143324)	432442333	5 (777776777760771)
SA CT(31)	2	4	8	9	19	20 (223325143325)	432442333	5 (777776777760771)
SA CT(32–34)	2	4	8	9	17, 21, 22	21 (223325153324)	432442333	11 (777776777760731)
SA CT(35–36)	2	4	8	10	23, 24	22 (223325153224)	332442343	12 (700076777760771)
SA CT(37)	2	4	8	10	23	23 (224325153322)	432442343	12 (700076777760771)
SA CT(38–44)	2	4	8	10	23, 25, 26, 27, 28, 29, 30	24 (224325153324)	432442343	12 (700076777760771)
SA CT(45)	2	4	8	10	31	24 (224325153324)	ND	5 (777776777760771)
SA CT(46–47)	2	4	8	11	32, 33	25 (224325153324)	332442331	12 (700076777760771)
SA CT(48)	2	4	9	12[d]	34	26 (224325153324)	432442333	13 (777776777760740)
SA CT(49–50)	2	4	9	9	34, 35	21 (223325153324)	432442333	13 (777776777760740)
SA CT(51–52)	2	4	9	9	34, 35	27 (224325133324)	432442333	13 (777776777760740)
SA CT(53)	2	4	9	9	34	28 (226325153324)	432442333	13 (777776777760740)
SA CT(54)	2	4	9	10	36	24 (224325153324)	432442343	14 (700076774360771)
SA CT(55)	2	4	9	10	37	29 (224325153325)	432442343	15 (700076777740371)
SA CT(56)	2	4	10	13	38	30 (224325153434)	ND	2 (777776777760601)
SA CT(57)	2	4	11	14	39	31 (224225164434)	ND	16 (700076777760671)
SA CT(58)	2	5	12	15[d]	40	32 (224325143324)	432442334	5 (777776777760771)
SA CT(59)	2	5	12	16	41	33 (224325153322)	432442333	17 (777776777560771)
SA CT(60)	2	5	12	16	42	33 (224325153322)	432442333	15 (777776777760771)
SA CT(61)	2	5	12	16	41	34 (234325133323)	432442333	17 (777776777560771)
SA CT(62)	2	5	12	16	41	35 (234325153323)	ND	5 (777776777760771)
SA CT(63)	2	5	12	17	41	33 (224325153322)	432442333	18 (777776617560771)
SA CT(64)	2	5	13	18	43	36 (224325153222)	532442433	12 (700076777760771)
SA CT(65)	2	5	13	18	44	37 (224325153222)	532442333	12 (700076777760771)
SA CT(66)	2	6	14	19	45	38 (224325164335)	ND	12 (700076777760771)
SA CT(67)	1	5	15	20	46	39 (254316?34613)	ND	19 (757777777413731)

[a] MIRU-VNTR loci according to reference 33.

[b] VNTR loci 424, 577, 1895, 2347, 2401, 2461, 3171, 3690, and 4156 (see Table 1 and references 14 and 24 and P. Supply et al., unpublished data). ND, not determined;
?, missing VNTR allele.
[c] Spoligotypes are arbitrary designations to demonstrate genetic diversity, and octal formats are according to reference 8.
[d] Mutations in the IS*6110*-flanking domains other than in the DR region (43).

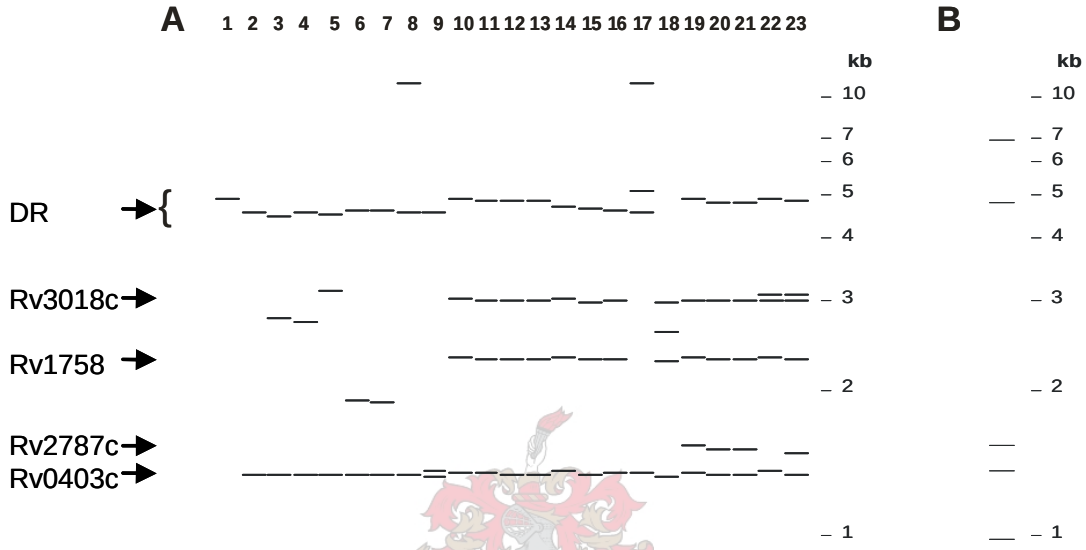


FIG. 1. Southern blot analysis of Cape Town *M. tuberculosis* isolates with ≤6 IS*6110* insertions. Genomic DNA was restricted with *PvuII* and electrophoretically fractionated on agarose gels, and Southern hybridization was done with an enhanced-chemiluminescence labeled IS-3' probe. (A) IS*6110* banding pattern of distinct principal genetic group 2 isolates from Cape Town, South Africa. The lanes are labeled according to the isolate names in Table 3. The band representing the IS*6110* insertion in the DR region and bands representing IS*6110* insertions in genes Rv0403c, Rv1758, Rv2787c, and Rv3018c are indicated. (B) IS*6110* banding pattern of the principal genetic group 1 isolate from Cape Town, South Africa. The lane is labeled according to the isolate name given in Table 3. Co-hybridization between the IS*6110* and DR probes could not be demonstrated due to the presence of a *PvuII* site situated between the 3' domain of the IS*6110* element and the flanking DR region.

Clonal expansion of strains. Combined analysis of the above-mentioned different molecular markers strongly supported the close genetic relatedness of the principal genetic group 2 isolates and their clear distinctness from the single isolate from principal genetic group 1. Six out of 21 VNTR loci (MIRU-VNTR loci 2, 20, 23, and 24 and VNTR loci 2347 and 2401) were fully identical within principal genetic group 2 isolates

analyzed (Table 3), while 7 other loci (MIRU-VNTR loci 4, 16, 27, and 39 and VNTR loci 1895, 2461, and 3171) displayed at most three variations compared to the predominant alleles among these isolates. PCR amplification of the Rv0403c region (Fig. 1A) showed that only the principal genetic group 2 variants with ≥2 and ≤6 IS*6110* elements had an IS*6110* in the same position, suggesting that these isolates were derived from a common ancestor. All principal genetic group 2 spoligotypes showed a deletion of direct-variable-repeat (DVR) sequences 33 to 36 (29), while isolates with ≥2 and ≤6 IS*6110* insertions also all showed an additional DVR 18 deletion (Table 2). DVR 34 was deleted in the principal genetic group 1 isolate (Table 2).

Phylogenetic analysis of the principal genetic group 2 isolates was done based on the whole set of markers, using the principal genetic group 1 member [SA CT(67)] as an outgroup (Fig. 2). The overall branching order of the tree suggests that the principal genetic group 2 isolates evolved from a common progenitor by sequential replicative transposition of IS*6110*, followed, in certain cases, by mutation in the regions flanking the IS*6110* elements (Table 3). According to this tree, these clones first evolved by replicative transposition of the IS*6110* element into Rv0403c (Fig. 1A) and by the deletion of DVR 18 to generate a strain with two IS*6110* insertions (Fig. 2). These two genotypic characteristics, along with the six conserved VNTR loci, were subsequently inherited in all the progeny. In different branches of these progeny, subsequent IS*6110* insertions were identified in Rv1758, Rv3018c, and Rv2787c to generate clonal variants with between three and six IS*6110* insertions (Fig. 2). Along with these events occurred the deletion of DVRs 39 to 42 and 4 to 12 (Table 2 and Fig. 2). Broadly similar pictures

of stepwise acquisition of IS*6110* were obtained by phylogenetic analysis using either IS*6110* RFLP, spoligotyping, and PGRS or VNTR genotypes alone (data not shown).

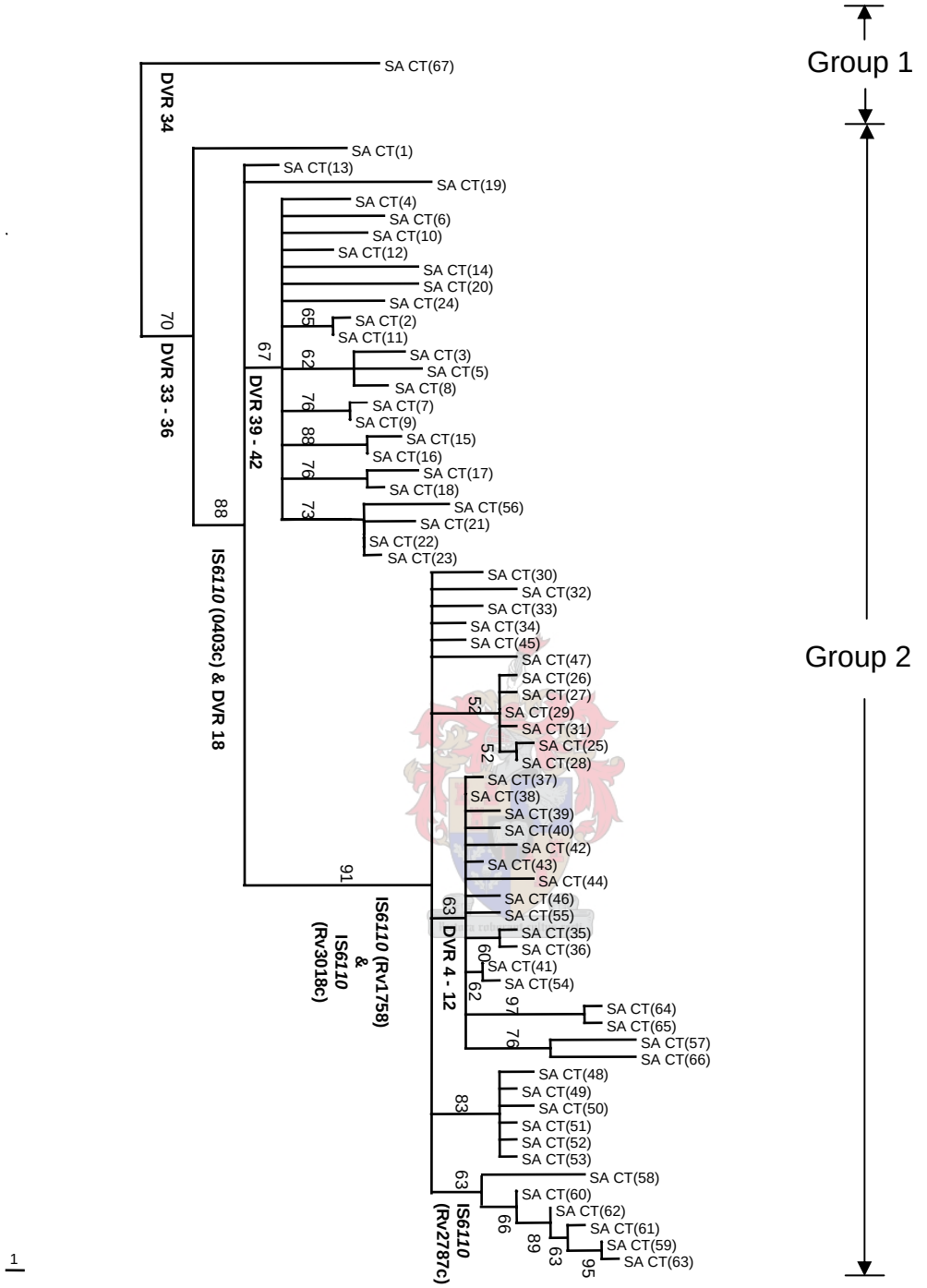


FIG. 2. Phylogenetic tree of Cape Town *M. tuberculosis* isolates with _6 IS*6110* insertions. Genetic data from five different genotyping methods were subjected to phylogenetic analysis using the bootstrapping and neighbor-joining algorithms (see Materials and Methods). The tree was rooted to the principal genetic group 1 clone [isolate SA CT(67)]. Booststrap values are given at internal nodes. IS*6110* insertions in defined genes, as well as DVR deletions, are indicated at the nodes where they occurred. All branches with zero length were collapsed. Isolate names are the same as in Table 3. Principal genetic groups 1 and 2 were assigned according to polymorphisms in the *katG* and *gyrA* genes (31). The scale indicates the number of steps per unit length.

Geographical distribution of strains. To determine the evolutionary relationships among the low-IS*6110*-copy-number strains in different geographical regions, the genotypic data from a representative set of the Cape Town isolates (*n* = 16) were compared with those of a representative set of low- IS*6110*-copy-number isolates from other regions of southern Africa (*n* = 17) (this study), Europe (London, United Kingdom [*n* = 8] [7] and Denmark [*n* = 2] [2]), and the United States (Michigan [*n* = 47] [6] and the CDC1551 reference strain [12]). Phylogenetic analysis using these data indicated a close evolutionary relationship among the principal genetic group 2 isolates from these different geographical regions (Fig. 3). Many of the isolates in the different settings shared identical IS-3' banding patterns, IS*6110* insertion points, and spoligotypes, and only in rare instances were IS*6110* transposition and DVR deletion events found to be unique to a specific geographic region (Fig. 3; IS*6110* insertion in Rv2787c of SA isolates). These observations suggest that these genotype properties were acquired in the distant past prior to global dissemination of the lineage. In contrast, comparison of MIRUVNTR genotypes based on 12 loci, in common with those previously reported (6), failed to identify clones with IS-3', spoligotype, and MIRU-VNTR genotypes identical between Cape Town and Michigan, indicating that the MIRU-VNTR loci are evolving more rapidly. A similar study could not be done for the PGRS genotypes, as the methodology has not been internationally standardized.

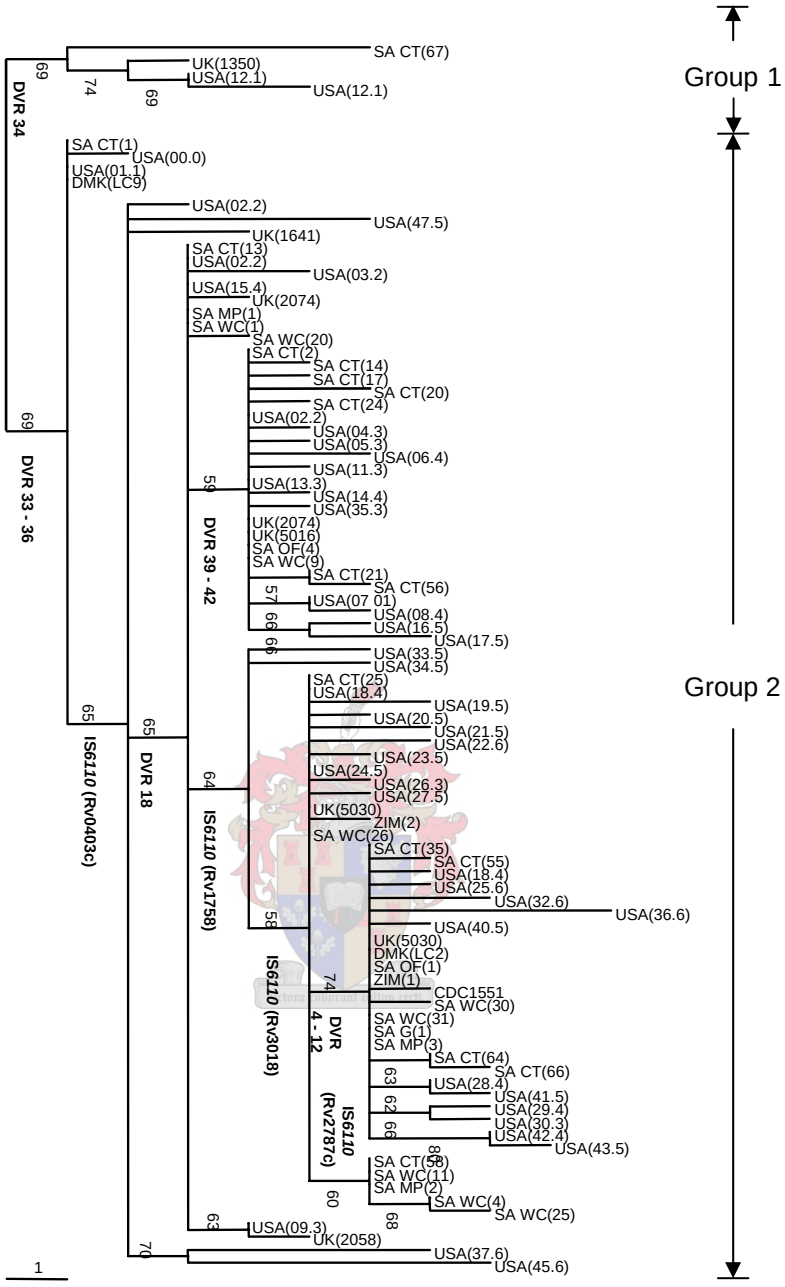


FIG. 3. Phylogenetic tree of *M. tuberculosis* isolates with ≤6 IS*6110* insertions obtained from different geographical regions. Representative isolates (n = 91) cultured in Cape Town, South Africa; Europe (2, 7); and the United States (6, 12) were included. Genetic data from IS*6110* banding patterns, IS*6110* insertion points, and spoligotyping were subjected to phylogenetic analysis using the bootstrapping and neighbor-joining algorithms. The tree was rooted to the principal genetic group 1 isolate [SA CT(67)]. Booststrap values are given at internal nodes. IS*6110* insertions in defined genes, as well as DVR deletions, are indicated at the nodes where they occurred. All branches with zero length were collapsed. The Cape Town isolate names are the same as in Table 3, while the isolates from other regions of southern Africa are labeled as follows: SA WC (South Africa Western Cape), SA OF (South Africa Free Sate), SA G (South Africa Gauteng), SA MP (South Africa Mpumalanga), and ZIM (Zimbabwe). The isolates from Europe are labeled according to the system in references 7 and 2, while the isolates from the United States are labeled according to the system in references 6 and 12. Principal genetic groups 1 and 2 were assigned according to polymorphisms in the *katG* and *gyrA* genes (31). The scale indicates the number of steps per unit length.

DISCUSSION

This study provides evidence that nearly all *M. tuberculosis* isolates with ≤6 IS*6110* elements collected in Cape Town, South Africa, are members of a lineage of the principal genetic group 2. Evidence for this is based on the inheritance of defined polymorphisms, which include (i) principal genetic group 2 classification according to mutations in the *katG* and *gyrA* genes (31) and concordant deletion of DVRs 33 to 36, known to be specific to principal genetic group 2 and 3 strains (29); (ii) the identification of a conserved IS*6110* insertion in Rv0403c (13) and the deletion of DVR 18 from the DR region in principal genetic group 2 strains with ≥2 and ≤6 IS*6110* insertions; and (iii) the presence of six fully conserved VNTR loci. Only one isolate from this setting was identified as being a member of a distinct low-IS*6110*-copy-number lineage of the principal genetic group 1, which corresponds to cluster I (15) or group I (7) and has been primarily associated with patients from East Africa and Asia (6, 30).

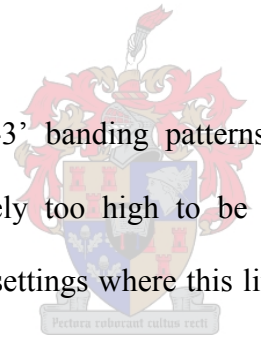
The principal genetic group 2 lineage studied here encompasses the groups referred to as groups II and III (7) or clade X (27) and clusters IV and V defined by sSNP analysis (15). Our phylogenetic analysis, based on fully independent markers in isolates from different geographical areas, supports the notion that strains in this principal genetic group 2 lineage evolved from a common progenitor containing a single IS*6110* element by sequential acquisition of up to five additional IS*6110* copies, as well as by expansion and

contraction of VNTR sequences and the deletion of specific DVRs. Sequential acquisition of additional IS*6110* copies is consistent with the direct evolutionary relationship between the sSNP clusters IV and V, which include strains with one to three and four to six IS*6110* copies, respectively (15). Such congruence between phylogenies inferred from independent sets of markers (within our study or between our study and that of Gutacker et al. [15]) provides strong evidence for the robustness of the inferred phylogeny. Moreover, the deletion of DVRs suggested by our phylogenetic analysis (Fig. 2 and 3) is consistent with previous findings supporting the notion that evolution of the DR region is driven by loss of DVR sequences rather than by their duplication (9, 36, 44).

Interestingly, each IS*6110* transposition event appeared to occur only once within the phylogenetic tree, suggesting divergent evolution. This is in sharp contrast to a previous suggestion that the IS*6110* banding patterns of low-IS*6110*-copy-number strains could have evolved convergently due to the presence of preferential IS*6110* integration sites (13). The limited number of IS*6110* variants identified may suggest that IS*6110* transposition is regulated in this lineage, raising the hypothesis of lineage-specific effects. Regulation of the number of transposable elements, referred to as taming, has been described in eukaryotic genomes and might be a specific mechanism against mutagenic effects induced by these elements (18).

The preservation of certain IS-3' banding patterns and spoligotypes in isolates from Cape Town, other southern African regions, Europe, and the United States suggests that these

markers have remained stable over a long period of time. Therefore, we hypothesize that these genotypes represent clones that evolved in the distant past and have become globally disseminated. Examination of the SpolD database (11) indicates that principal genetic group 2 isolates with the characteristic DVR 33-to-36 and DVR 18 deletions have been isolated in 27 different countries. By comparison, principal genetic group 1 clones with the characteristic DVR 34 deletion have been isolated in 26 countries, with a high prevalence in South Asia. Taken together, these findings suggest that in addition to other well-identified lineages, like W-Beijing (4), the principal genetic group 1 and 2 low-IS*6110*-copy-number lineages now play an important role in the global tuberculosis epidemic.



The inferred stability of the IS-3' banding patterns and of certain spoligotypes for extended periods of time is likely too high to be informative for tracking ongoing transmission between patients in settings where this lineage is predominant. Conversely, comparison of our genotype data with those of Cowan et al. (6) failed to identify strains from Cape Town and Michigan in which the IS-3' banding pattern, spoligotype, and MIRU-VNTR types were identical. Given the stability of MIRU-VNTR genotypes in epidemiologically linked isolates (16, 20, 23, 26), the absence of MIRU-VNTR matching between the two studies is in accordance with the above argument of distant relationships between shared IS-3' types and spoligotypes. Moreover, it supports VNTR typing as a useful tool for epidemiological tracking across various epidemiological settings and bacterial populations. This is consistent with the contention that, as a multilocus-based method, VNTR typing is much less exposed to biases inherent in single loci or in copy

numbers of a single genetic element, such as spoligotyping and IS*6110*-based typing, respectively.

Our study represents a step toward a better understanding of the evolutionary mechanisms modeling the genome in different *M. tuberculosis* lineages and of the different rates at which these events occur. This will provide new insights for the interpretation of molecular epidemiological data and enhance our understanding of how different strains contribute to the tuberculosis epidemic in specific regions and on a global scale.

ACKNOWLEDGMENTS

This study was made possible by grants from the GlaxoSmithKline Action TB Initiative, IAEA (projects SAF6/003 and CRP 9925), the Harry Crossely Foundation, and the National Research Foundation (project 2054201) and from the Ministe`re de la Recherche and Ministe` re des Affaires Etrange`res and the South African Medical Research Council (MRC). The work was also supported by Institut National de la Sante´ et de la Recherche Me´dicale (INSERM), Institut Pasteur de Lille, Re´gion Nord-Pas-de-Calais. P.S. is a Chercheur du Centre National de Recherche Scientifique.

We thank E. Engelke, S. Carlini, M. De Kock, and Frederique De Matos for their technical assistance. We thank S. Charalambous for the provision of clinical isolates collected in the Free State, South Africa.

REFERENCES

1. **Alland, D., T. S. Whittam, M. B. Murray, M. D. Cave, M. H. Hazbon, K. Dix, M. Kokoris, A. Duesterhoeft, J. A. Eisen, C. M. Fraser, and R. D. Fleischmann.** 2003. Modeling bacterial evolution with comparative-genome-based marker systems: application to *Mycobacterium tuberculosis* evolution and pathogenesis. J. Bacteriol. **185:**3392–3399.

2. **Bauer, J., A. B. Andersen, K. Kremer, and H. Miorner.** 1999. Usefulness of spoligotyping to discriminate IS*6110* low-copy-number *Mycobacterium tuberculosis* complex strains cultured in Denmark. J. Clin. Microbiol. **37:**2602– 2606.

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:**40–44.

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. **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.

6. **Cowan, L. S., L. Mosher, L. Diem, J. P. Massey, and J. T. Crawford.** 2002. Variable-number tandem repeat typing of *Mycobacterium tuberculosis* isolates with low copy numbers of IS*6110* by using mycobacterial interspersed repetitive units. J. Clin. Microbiol. **40:**1592–1602.

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

8. **Dale, J. W., D. Brittain, A. A. Cataldi, D. Cousins, J. T. Crawford, J. Driscoll, H. Heersma, T. Lillebaek, T. Quitugua, N. Rastogi, R. A. Skuce, C. Sola, D. van Soolingen, and V. Vincent.** 2001. Spacer oligonucleotide typing of bacteria of the *Mycobacterium tuberculosis* complex: recommendations for standardised nomenclature. Int. J. Tuberc. Lung Dis. **5:**216–219.

9. **Fang, Z., N. Morrison, B. Watt, C. Doig, and K. J. Forbes.** 1998. IS*6110* transposition and evolutionary scenario of the direct repeat locus in a group of closely related *Mycobacterium tuberculosis* strains. J. Bacteriol. **180:**2102– 2109.

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

11. **Filliol, I., J. R. Driscoll, D. van Soolingen, B. N. Kreiswirth, K. Kremer, G. Valetudie, D. A. Dang, 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, M. L. Ho, 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 assessed by spoligotyping in an international study. J. Clin. Microbiol. **41:**1963–1970.

12. **Fleischmann, R. D., D. Alland, J. A. Eisen, L. Carpenter, O. White, J. Peterson, R. DeBoy, R. Dodson, M. Gwinn, D. Haft, E. Hickey, J. F. Kolonay, W. C. Nelson, L. A. Umayam, M. Ermolaeva, S. L. Salzberg, A. Delcher, T. Utterback, J. Weidman, H. Khouri, J. Gill, A. Mikula, W. Bishai, J. W. Jacobs, Jr., J. C. Venter, and C. M. Fraser.** 2002. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. J. Bacteriol. **184:**5479–5490.

13. **Fomukong, N., M. Beggs, H. el Hajj, G. Templeton, K. Eisenach, and M. D. Cave.** 1997. Differences in the prevalence of IS*6110* insertion sites in *Mycobacterium tuberculosis* strains: low and high copy number of IS*6110.* Tuber. Lung Dis. **78:**109–116.

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

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

16. **Hawkey, P. M., E. G. Smith, J. T. Evans, P. Monk, G. Bryan, H. H. Mohamed, M. Bardhan, and R. N. Pugh.** 2003. Mycobacterial interspersed repetitive unit typing of *Mycobacterium tuberculosis* compared to IS*6110*- based restriction fragment length polymorphism analysis for investigation of apparently clustered cases of tuberculosis. J. Clin. Microbiol. **41:**3514–3520.

17. **Hermans, P. W., F. Messadi, H. Guebrexabher, D. van Soolingen, P. E. de Haas, H. Heersma, H. de Neeling, A. Ayoub, F. Portaels, and D. Frommel.** 1995. Analysis of the population structure of *Mycobacterium tuberculosis* in Ethiopia, Tunisia, and The

Netherlands: usefulness of DNA typing for global tuberculosis epidemiology. J. Infect. Dis. **171:**1504–1513.

18. **Jensen, S., M. P. Gassama, and T. Heidmann.** 1999. Taming of transposable elements by homology-dependent gene silencing. Nat. Genet. **21:**209–212.

19. **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.

20. **Kwara, A., R. Schiro, L. S. Cowan, N. E. Hyslop, M. F. Wiser, H. S. Roahen, P. Kissinger, L. Diem, and J. T. Crawford.** 2003. Evaluation of the epidemiologic utility of secondary typing methods for differentiation of *Mycobacterium tuberculosis* isolates. J. Clin. Microbiol. **41:**2683–2685.

21. **Le Fleche, P., M. Fabre, F. Denoeud, J. L. Koeck, and G. Vergnaud.** 2002. High resolution, on-line identification of strains from the *Mycobacterium tuberculosis* complex based on tandem repeat typing. BMC Microbiol. **2:**37. 22. **Mathema, B., P. J. Bifani, J. Driscoll, L. Steinlein, N. Kurepina, S. L. Moghazeh, E. Shashkina, S. A. Marras, S. Campbell, B. Mangura, K. Shilkret, J. T. Crawford, R. Frothingham, and B. N. Kreiswirth.** 2002. Identification and evolution of an IS*6110* low-copy-number *Mycobacterium tuberculosis* cluster. J. Infect. Dis. **185:**641–649.

23. **Mazars, E., S. Lesjean, A. L. Banuls, M. Gilbert, V. Vincent, B. Gicquel, M. Tibayrenc, C. Locht, and P. Supply.** 2001. High-resolution minisatellitebased typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. Proc. Natl. Acad. Sci. USA **98:**1901– 1906.

24. **Roring, S., D. Brittain, A. E. Bunschoten, M. S. Hughes, R. A. Skuce, J. D. van Embden, and S. D. Neill.** 1998. Spacer oligotyping of *Mycobacterium bovis* isolates compared to typing by restriction fragment length polymorphism using PGRS, DR and IS*6110* probes. Vet. Microbiol. **61:**111–120.

25. **Roring, S., A. Scott, D. Brittain, I. Walker, G. Hewinson, S. Neill, and R. Skuce.** 2002. Development of variable-number tandem repeat typing of *Mycobacterium bovis*: comparison of results with those obtained by using existing exact tandem repeats and spoligotyping. J. Clin. Microbiol. **40:**2126– 2133.

26. **Savine, E., R. M. Warren, G. D. van der Spuy, N. Beyers, P. D. van Helden, C. Locht, and P. Supply.** 2002. Stability of variable-number tandem repeats of mycobacterial interspersed repetitive units from 12 loci in serial isolates of *Mycobacterium tuberculosis.* J. Clin. Microbiol. **40:**4561–4566.

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

28. **Smittipat, N., and P. Palittapongarnpim.** 2000. Identification of possible loci of variable number of tandem repeats in *Mycobacterium tuberculosis.* Tuber. Lung Dis. **80:**69–74.

29. **Soini, H., X. Pan, A. Amin, E. A. Graviss, A. Siddiqui, and J. M. Musser.** 2000. Characterization of *Mycobacterium tuberculosis* isolates from patients in Houston, Texas, by spoligotyping. J. Clin. Microbiol. **38:**669–676.

30. **Soini, H., X. Pan, L. Teeter, J. M. Musser, and E. A. Graviss.** 2001. Transmission dynamics and molecular characterization of. *Mycobacterium tuberculosis* isolates with low copy numbers of IS*6110.* J. Clin. Microbiol. **39:**217– 221.

31. **Sreevatsan, S., X. Pan, K. E. Stockbauer, N. D. Connell, B. N. Kreiswirth, T. S. Whittam, and J. M. Musser.** 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc. Natl. Acad. Sci. USA **94:**9869–9874.

32. **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.

33. **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.

34. **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.

35. **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.

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

37. **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.

38. **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.

39. **Victor, T. C., A. M. Jordaan, A. van Rie, G. D. van der Spuy, M. Richardson, P. D. van Helden, and R. Warren.** 1999. Detection of mutations in drug resistance genes of *Mycobacterium tuberculosis* by a dot-blot hybridization strategy. Tuber. Lung Dis. **79:**343–348.

40. **Victor, T. C., A. van Rie, A. M. Jordaan, M. Richardson, G. D. Der Spuy, N. Beyers, P. D. van Helden, and R. Warren.** 2001. Sequence polymorphism in the *rrs* gene of *Mycobacterium tuberculosis* is deeply rooted within an evolutionary clade and is not associated with streptomycin resistance. J. Clin. Microbiol. **39:**4184–4186.

41. **Warren, R., M. Richardson, S. Sampson, J. H. Hauman, N. Beyers, P. R. Donald, and P. D. van Helden.** 1996. Genotyping of *Mycobacterium tuberculosis* with additional markers enhances accuracy in epidemiological studies. J. Clin. Microbiol. **34:**2219–2224.

42. **Warren, R. M., M. Richardson, S. L. Sampson, G. D. van der Spuy, W. Bourn, J. H. Hauman, H. Heersma, W. Hide, N. Beyers, and P. D. van Helden.** 2001. Molecular evolution of *Mycobacterium tuberculosis*: phylogenetic reconstruction of clonal expansion. Tuberculosis **81:**291–302.

43. **Warren, R. M., S. L. Sampson, M. Richardson, G. D. van der Spuy, C. J. Lombard, T. C. Victor, and P. D. van Helden.** 2000. Mapping of IS*6110* flanking regions in clinical isolates of *M. tuberculosis* demonstrates genome plasticity. Mol. Microbiol. **37:**1405–1416.

44. **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.

Chapter 7

Understanding of evolution of Principal Genetic Group 2 strains of *Mycobacterium tuberculosis*

My contribution to this project:	Spoligotyping of isolates *katG*463 and *gyrA*95 screening and classification SNP analysis LSP analysis Identification and analysis of IS*6110* insertion points Construction of a phylogenetic tree to demonstrate evolutionary events Writing and editing of manuscript

ABSTRACT

In order to determine if strain families have developed different unique genetic characteristics, we analysed representative *Mycobacterium tuberculosis* isolates from all the strain families belonging to principal genetic group 2 by spoligotyping, IS*6110* points of insertion, long sequence polymorphisms and single nucleotide polymorphisms to reconstruct phylogenetic relationships. A phylogenetic tree was created using information from the various genetic markers. This tree grouped the different strain families into 4 branches and identified previously unknown relationships between strain families. One of the branches consists of the internationally described LAM family of strains and can be subdivided into 6 strain families, each with its own characteristics. Clinical and epidemiological data of the different lineages identified were compared, but found limited association of clinical properties and epidemiology between different strain families. We showed that a Low Copy Clade progenitor gave rise to high copy number strains, including the Haarlem strain families. This is the first study to demonstrate the relationship between strain families of Principal genetic group 2 and laid the foundation to establish correlations between genotype and phenotype.

INTRODUCTION

The availability of whole genome sequences of different strains and species of *Mycobacterium tuberculosis* has made comparative genomics possible, and has shown this species to be highly clonal. The frequency of mutations in the genome is shown to be low. Three principal genetic groups (PGGs) for *M. tuberculosis* strains based on *katG*463 and *gyrA*95 polymorphisms has been identified (19). This study also proposed that group 2 derived from group 1 and in turn gave rise to group 3. The PGGs have been further divided into eight main clusters to depict the evolution of M. tuberculosis based on synonymous single nucleotide polymorphisms (sSNPs) (12). Topology of the tree was supported by analysis of nonsynonymous SNPs (nsSNP) and SNPs in intergenic regions (iSNPs) in four different settings (11). Phylogeny of *M. tuberculosis* was also described using sSNPs in genes associated with drug resistance and supported the initial PGGs (2). Recently Filliol *et al*. performed phylogenetic analysis to described the evolutionary history of *M. tuberculosis* (8). In this study 6 SNP cluster groups (SCGs) were identified and the authors proposed SNPs as a gold standard for phylogenetic studies. They also described a minimal set of SNPs to identify the SCGs (8).

Subsequent expansion of at the branchpoints of phylogenetic trees based on SNP data is unknown due to the lack of resolution. Various other genetic markers has been used to classify strains of *M. tuberculosis*, to unravel the evolution of this successful pathogen. The most widely used technique to classify strains into is IS*6110*-RFLP, based on the IS*6110* transposable element. Strains are considered part of a family when its RFLP pattern is > 65% similar. A limitation of IS*6110*-RFLP is that two bands of the same size may not represent the same insertion event. However, Mapping of insertion points has been used successfully to

describe clonal expansion of closely related strains. It is highly unlikely that a IS*6110* element will be inserted at the exact same place in the genome during independent events and will rather be inherited. Thus a point of insertion of *IS6110* can be used as a marker to classify strains into strain families. The evolutionary scenario for the family of strains with Low copy numbers of the IS*6110* was described by using various different molecular markers (25). One marker, an insertion of IS*6110* in Rv0403 that was thought to be specific for this group turns out to be also present in some high copy number strain families part of phylogenetic group 2.

The spoligotyping method has also been used to classify strains into strain families. The method is based on the deletion of unique spacer sequences between 36bp direct repeats in the DR region and strains are classified into strain families by specific signatures created by deletions of spacers by homologous recombination (14).

All the described molecular methods have advantages and disadvantages, but in combination should be helpful to unravel the evolutionary scenario of *M. tuberculosis*. In this study we used various genetic markers to analyse and classify phylogenetic group 2 isolates and piece together the evolutionary puzzle.

MATERIALS AND METHODS

M. tuberculosis isolates

M. tuberculosis isolates were cultured from patients with bacteriologically confirmed TB during December 1993 and January 2003 and were resident in 2 adjacent suburbs near Cape Town, South Africa.

IS*6110*-RFLP

Isolates were genotyped according to the internationally standardized IS*6110*-DNA fingerprinting method (22). DNA fingerprints were analyzed with GelCompar software, using UPGMA and Dice coefficients (13). Isolates with an IS*6110* similarity index of ≥65% were grouped into strain families (18).

Spoligotyping

Isolates were Spoligotyped according to the international standardized protocol described by Kamerbeek *et al*. (14)

***GyrA*95 and *katG*463 classification**

Strains within each strain family were classified into PGGs by DNA sequencing of the *katG* and *gyrA* genes (19).

Single Nucleotide Polymorphisms

SNP analysis was done by PCR amplification followed by DNA sequencing according to the method by Baker et al. (2)

Large Sequence Polymorphisms

Regions of Difference as described by Gagneux et al. (9) were analysed by PCR.

Statistical Analysis

The fisher exact test was used to determine the association the different family groups to age, gender, new vs. retreatment cases, disease outcome, smear positivity within 2 months of diagnosis and drug resistance. A p-value of 0.05 was used as the cut-off level for significance.

RESULTS

During the period January 1993 to December 2004 *M. tuberculosis* cultures from 2037 cases from 2 adjacent suburbs near Cape Town, South Africa were genotyped by IS*6110*-DNA fingerprinting. Cluster analysis grouped these isolates into 30 strain families according to a similarity index of ≥65%. PGG analysis, according to polymorphisms in *katG*463 and *GyrA*95 showed 5 families to be representative of PGG 1, 17 families to be representative of PGG 2 and 8 families to be representative of group3 (table 1). Strains classified as PGG 2 formed the cohort for subsequent genotypic analysis. A total of 105 isolates from 17 families were analyzed by spoligotyping, IS*6110* insertion site mapping, Single Nucleotide Polymorphisms and Large Sequence Polymorphisms.

Table 1. Classification of strain families in the 3 Principal Genetic Groups

	Group 1	Group 2	Group 3
Polymorphisms	*katG*463 CTG (Leu) *gyrA*95 ACC (Thr)	*katG*463 CGG (Arg) *gyrA*95 ACC (Thr)	*katG*463 CGG (Arg) *gyrA*95 AGC (Ser)
Families	F20, F25, F27, F29, F31	F1, F2, F3, F4, F6, F7, F9, F10, F11, F13, F14, F15, F19, F24, F26, F28, LCC	F5, F8, F16, F17, F18, F21, F22, F23

LCC = Low Copy Clade

The resolution of differentiation of strain families was found to be depended on the marker used: SNPs < LSP < IS6110 insertion site mapping (see figure 1). Analysis of a combination of all of the markers enabled the prediction of the most parsimonious evolutionary relationship between the different strain families. Four independent evolutionary lineages were observed, which all derived from a single common PGG 2 progenitor. These lineages have previously been named LAM, F28, F3 and a combination of families including, Haarlem and Low Copy Clade (LCC)

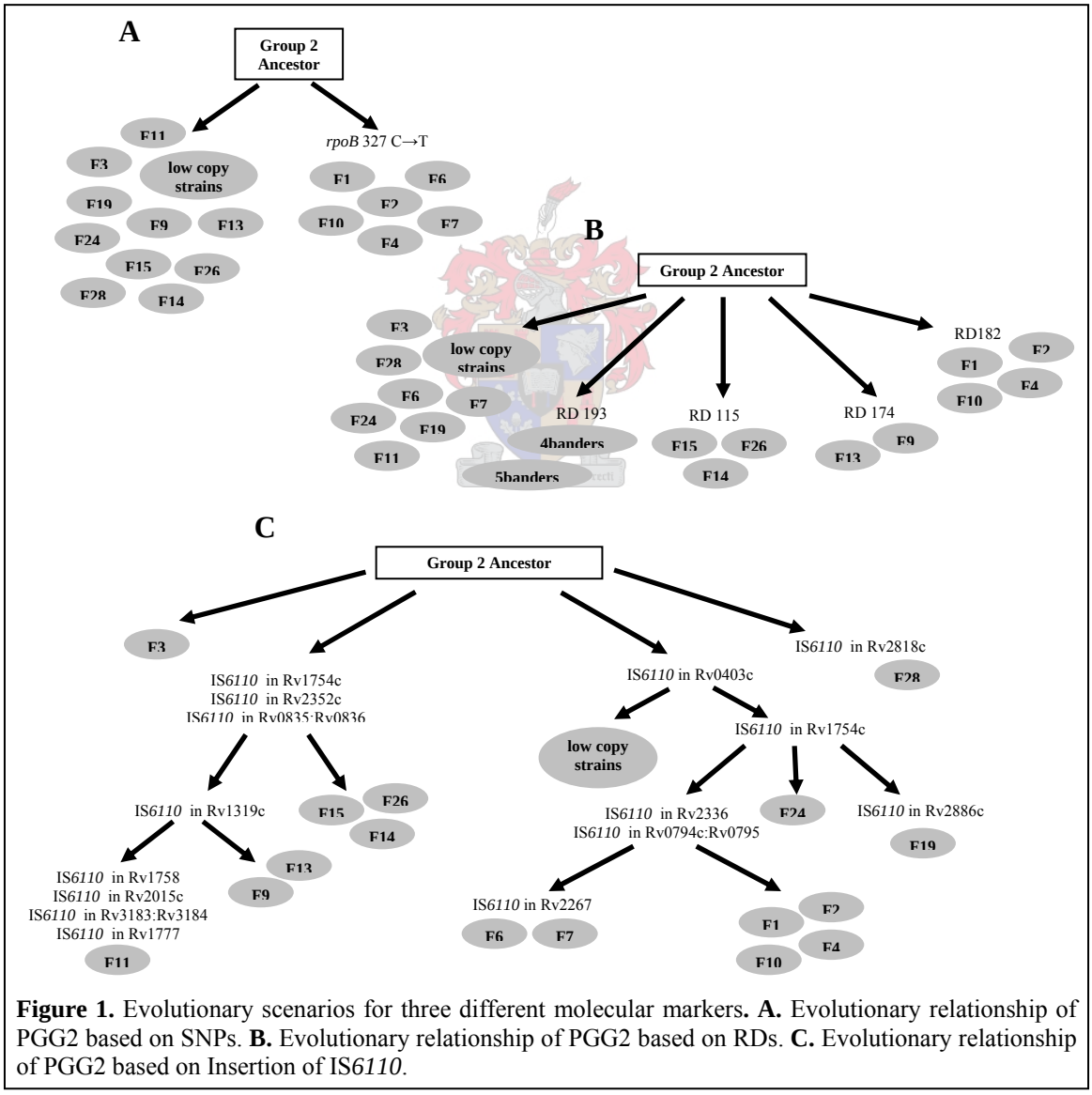


Figure 1. Evolutionary scenarios for three different molecular markers**. A.** Evolutionary relationship of PGG2 based on SNPs. **B.** Evolutionary relationship of PGG2 based on RDs. **C.** Evolutionary relationship of PGG2 based on Insertion of IS*6110*.

The LAM lineage

The LAM lineage was found to consists of 6 strain families (F9, F11, F13, F14, F15, F26), which all share common IS*6110* insertions in Rv1754 (position 1986625), Rv2352 (position 2633841), Rv0835:Rv0836 (position 932204). According to international designation, F11, F13, F15 and F9 corresponds to LAM3, LAM1, LAM4 and LAM11-ZWE, respectively, based on their characteristic spoligotype patterns (3,7) and all of them share the characteristic deletion of DVR 21-24 on their spoligotype profile. The spoligotype signature of F14 and F26 has not been previously classified. The strain families can be differentiated by RDs and additional IS*6110* points of insertions and DVR deletions.

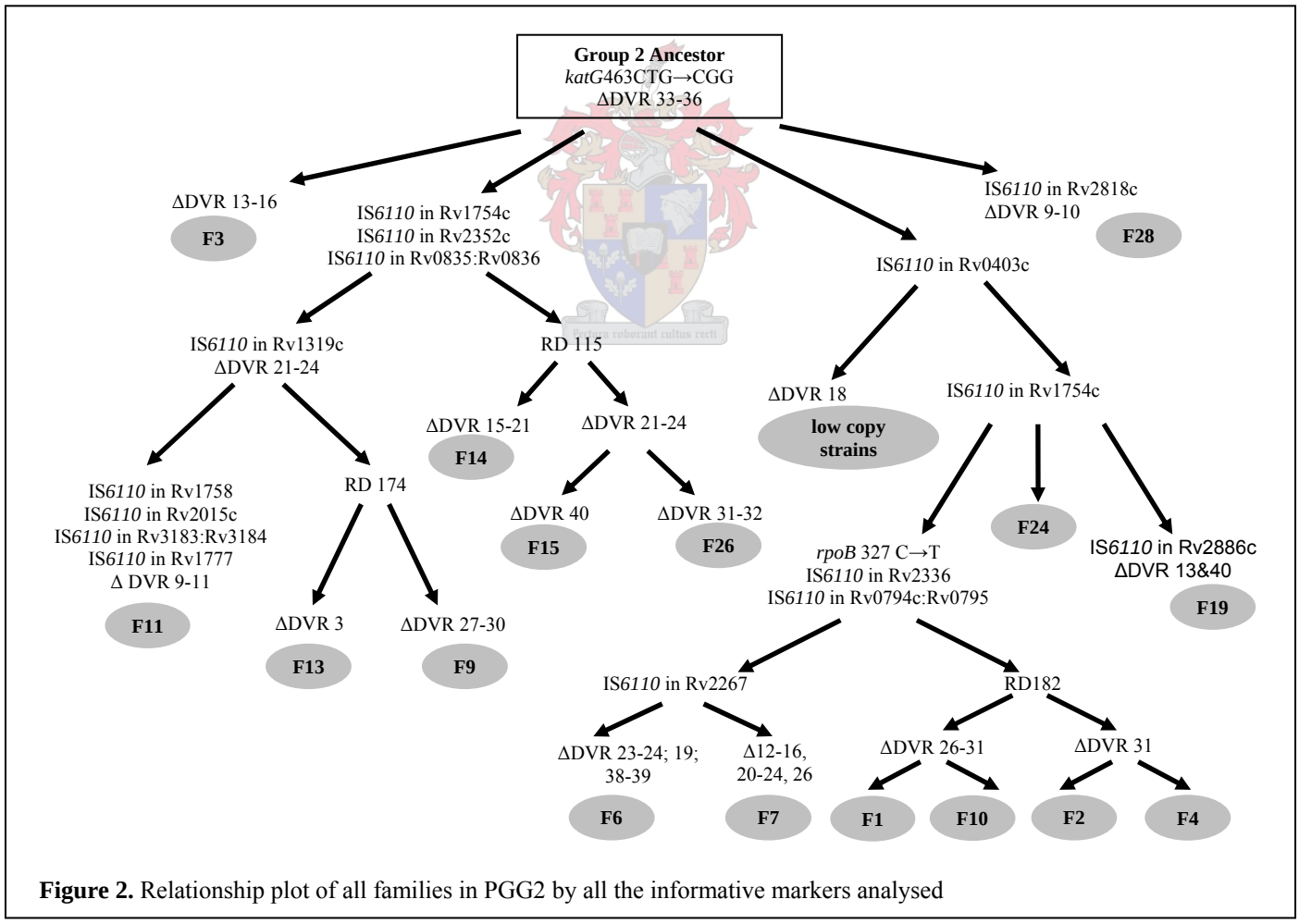


Figure 2. Relationship plot of all families in PGG2 by all the informative markers analysed

LCC and Haarlem lineage

The second lineage consists of strain families that all share a insertion of IS*6110* in Rv0403c (position 483299) as previously described (4,25). Two predominant sub-lineages form part of this lineage: Haarlem and the LCC. The LCC strains are characterized by ≥2 and ≤6 copies of the IS*6110*. and the deletion of DVR 18. In contrast to previous assumption, this study showed that the Low Copy Clade strain with two IS*6110* insertions, was ancestor to 8 high copy number strain families (F1, F2, F4, F6, F7, F10, F19, F24). These high copy number strains all share an IS6110 insertion in Rv1754c (position 1986622) and subsequently evolved into distinct families by additional IS*6110* insertions and DVR deletions (figure 2). The Haarlem strain families (F1, F2, F4, F10) all share deletion of RD182 and have 2 characteristic spoligotype signatures (figure 2). F1 and F10 have the Haarlem 1 spoligotype signature (deletion of DVR 26-31) and F2 and F4 have the Haarlem 3 spoligotype signature (deletion of DVR 31) as described previously (3,7). The Haarlem strain families also share *rpoB*327C→T polymorphism and insertion of IS6110 in Rv2336 (position 2610861) and Rv0794:Rv0795 (position 888787) with families F6 and F7 (figure 2). F6 and F7 therefore forms part of a closely related set of Haarlem-like strain families, and have not been previously described.

F3 lineage and F28 lineage

No characteristic IS6110 insertions have been identified, although F3 has specific spoligotype signature (deletion of DVR 13-16) was found. F28 has specific insertion of IS*6110* in Rv2818c (position 3125706) and deletion of DVR 9-10 in the spoligotype pattern.

Clinical Differences between major groups within PGG 2.

In order to determine the influence of evolution on disease presentation, clinical and demographical data was used to test the hypothesis that evolution has influenced the properties of the different lineages within PGG 2. Table 2 described the clinical characteristics of patients who are infected with strains belonging to 4 major subgroups. These 4 groups represent more than 90% of the isolates from PGG 2 in the study setting. Statistically significant differences were found for gender, outcome, drug resistance and clustering when the 4 groups were compared. The Haarlem families had more females than the LAM super family (p-value = 0.0336; odds ratio = 1.55; 95% CI 1.03-2.34). LCC had higher number of drug resistant cases than the Haarlem families (p-value = 0.0039; odds ratio = 0.1998; 95% CI 0.05929-0.6733) and LAM family (p-value = 0.003; odds ratio = 0.41; 95% CI 0.23-0.73). Treatment failure rates were higher for F28 when compared to Haarlem (p-value = 0.0454; odds ratio = 0.52; 95% CI 0.28-0.97) and LAM families (p-value = 0.01; odds ratio = 0.54; 95% CI 0.34-0.85). Treatment failure rates were also higher for LCC comparing to Haarlem (p-value 0.0316; odds ratio = 1.91; 95% CI 1.08 to 3.36) and LAM families (p-value 0.0016; odds ratio = 1.84; 95% CI 1.26-2.69).

Table 2 clinical characteristics of the major groups within PGG 2

		Haarlem	LAM	F28	LCC
Families		F1; F2; F4; F10	F9; F11; F13; F14; F15; F26	F28	≥2 and <6 copies of IS*6110*
Total		**117**	**443**	**122**	**232**
Age	**Average**	35	33	37	36
Sex	**Female**	61 (52.2%)	184 (41.6%)	61 (50%)	111 (47.9%)
	Male	56 (47.9%)	259 (58.5%)	61 (50%)	121 (52.2%)
Category	**New**	73 (62.4%)	286 (64.6%)	72 (59.1%)	134 (57.8%)
	Retreatment	41 (35.1%)	142 (32.1%)	47 (38.6%)	91 (39.3%)
Outcome	**Successfully treated** [a]	83 (71%)	317 (71.6%)	80 (65.6%)	148 (63.8%)
	Treatment failure [b]	20 (17.1%)	79 (17.9%)	37 (30.4%)	68 (29.4%)
Drug resistance	**Sensitive**	114 (97.5%)	420 (94.9%)	114 (93.5%)	205 (88.4%)
	Resistant	3 (2.6%)	23 (5.2%)	8 (6.6%)	27 (11.7%)
Smear	**Negative**	34 (29.1%)	138 (31.2%)	39 (32%)	75 (32.4%)
	Positive	83 (71%)	305 (68.9%)	83 (68.1%)	157 (67.7%)
Clustering	**Clustered**	95 (81.2%)	341 (77%)	85 (69.7%)	229 (98.8%)
	Unique	22 (18.9%)	102 (23.1%)	37 (30.4%)	3 (1.3%)

[a] Successfully treated – patients that were cured or when treatment were completed without confirmation of cure.
[b] Treatment failure – patients that died, interrupted treatment, and were not treated.

Clustering

Clustering (when two or more fingerprints are exactly the same) can be used as a measure of recent transmission and were based on IS*6110* fingerprint. The LCC showed significantly more clustering than F28 (p-value < 0.0001; odds ratio = 0.03; 95% CI 0.01-0.10), LCC (p-value < 0.0001; odds ratio = 0.06; 95% CI 0.02-0.19) and LAM (p-value < 0.0001; odds ratio = 0.04; 95% CI 0.01-0.14). This maybe explained by the low resolution of IS*6110* banding in LCC.

CONCLUSIONS

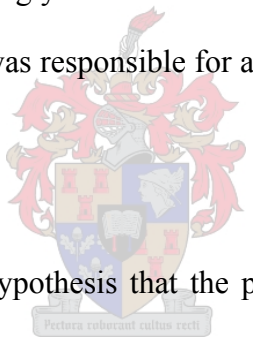
The PGG 2 strain families form an important component of the TB epidemic in Cape Town, South Africa and represent greater than 50% of all the TB cases. Strains with a PGG 2 classification have also been shown to contribute significantly to the global epidemic and have recently been referred to as members of the Euro-American Clade (9). This is the first study to demonstrate evolutionary relationships among members of PGG 2. In this study we used

defined genetic markers (i.e. SNP, LSP, IS*6110* insertion site mapping and spoligotyping) to predict the evolutionary relationships among IS*6110*-RFLP strain families within PGG 2 and demonstrated their evolution from a common ancestor characterised by an IS6110 insertion in the DR region, deletion of DVR's 33-36 and a *katG*463CTG→CGG polymorphism. Our evolutionary scenario predicts 4 major lineages; LAM, F3, F28 and a combination of LCC, Haarlem and other strain families.

The LAM lineage is closely related group, consisting of at least 6 strain families. The term LAM super-families was originally derived from the characteristic deletion of DVRs 21-24 in the DR region (3,7). The LAM super-family have been described globally and different sub-groups of LAM have been found to be more prevalent in different settings. F11 (LAM3) have been shown to be dominant in the Western Cape Province communities in South Africa although a recent study concluded that they were widely distributed on all continents (24). Surprisingly, F9 (LAM11-ZWE) strain family have been found dominant in Central Southern Africa, including Zimbabwe and Zambia ((6), personal communication V. Chihota). In Europe , LAM12/Madrid1 was found to be dominant in Spain (10) and LAM7 which is predominant in Turkey (26). The reason for the dominance of specific strain families in different settings remains speculative, either host genetics has allowed for successful propagation or that the pathogen itself has evolved a higher level of pathogenicity in certain host populations. This supports previous studies which have suggested a relationship between host and pathogen (2).

In contrast to what has been previously assumed, we showed the LCC is the progenitor to strains with more than 6 copies of IS*6110*, including the Haarlem strain families. The Haarlem

strain was initially isolated from a patient from the Netherlands (15) and has subsequently been described in Tunisia (16), in patients from Iran and Afghanistan (23) and also part of an extensively drug resistant outbreak in Iran (17). It has also been suggested that the Haarlem strains have a greater propensity to develop drug resistance though polymorphisms in mismatch repair gene, *mgtC* and their ability to spread has been interpreted to imply a higher level of virulence (1). In the study setting only 117 cases were detected and clustering and or drug resistance was not significantly associated with this group. In a previous study, we demonstrated the epidemiological significance of the LCC in the TB epidemic (25). In this study we showed that drug resistance was associated with this lineage in the urban setting. Similarly, drug resistance was strongly associated with LCC in a rural setting in the Western Cape (21). We also that the LCC was responsible for a MDR outbreak in George, South Africa (20).



In this study we also tested the hypothesis that the properties of closely related strains may influence their epidemiology. However, only limited associations were found. This differs from previous studies, which have shown associations between strain family and site of disease and smear conversion rates (5). In our study setting the large majority of cases have pulmonary disease and therefore and association between pulmonary and extra-pulmonary could not be tested. The absence of an association between clinical characteristics and strains family may be the result that there was an extremely long delay in seeking healthcare.

However, it is also possible that our hypothesis may have lead to the masking of associations. It is possible that more recent evolutionary events unique within either strains families or sub-

lineages thereof may have altered their pathogenic properties. These were not tested in this study. A study on Beijing strains demonstrated that the most recently evolved strain reflected the most pathogenic variant (personal communication: Dr. M. Hanekom).

In summary, we show the complexity of the evolutionary relationship among PGG 2 strains found in Cape Town, South Africa. The predominant lineages have evolved independently from a common progenitor with the Haarlem strain family evolving from the early member of LCC. Only limited associations could be demonstrated between clinical characteristics and lineage.

Reference List

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

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

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

4. **Dale, J. W., H. Al Ghusein, S. Al Hashmi, P. Butcher, A. L. Dickens, F. Drobniewski, K. J. Forbes, S. H. Gillespie, D. Lamprecht, T. D. McHugh, R. Pitman, N. Rastogi, A. T. Smith, C. Sola, and H. Yesilkaya**. 2003. Evolutionary relationships among strains of Mycobacterium tuberculosis with few copies of IS6110. J.Bacteriol. **185**:2555-2562.

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

6. **Easterbrook, P. J., A. Gibson, S. Murad, D. Lamprecht, N. Ives, A. Ferguson, O. Lowe, P. Mason, A. Ndudzo, A. Taziwa, R. Makombe, L. Mbengeranwa, C. Sola, N. Rastogi, and F. Drobniewski**. 2004. High rates of clustering of strains causing tuberculosis in Harare, Zimbabwe: a molecular epidemiological study. J.Clin.Microbiol. **42**:4536-4544.

7. **Filliol, I., J. R. Driscoll, D. van Soolingen, B. N. Kreiswirth, K. Kremer, G. Valetudie, D. D. Anh, R. Barlow, D. Banerjee, P. J. Bifani, K. Brudey, A. Cataldi, R. C. Cooksey, D. V. Cousins, J. W. Dale, O. A. Dellagostin, F. Drobniewski, G. Engelmann, S. Ferdinand, D. Gascoyne-Binzi, M. Gordon, M. C. Gutierrez, W. H. Haas, H. Heersma, G. Kallenius, E. Kassa-Kelembho, T. Koivula, H. M. Ly, A.**

Makristathis, C. Mammina, G. Martin, P. Mostrom, I. Mokrousov, V. Narbonne, O. Narvskaya, A. Nastasi, S. N. Niobe-Eyangoh, J. W. Pape, V. Rasolofo-Razanamparany, M. Ridell, M. L. Rossetti, F. Stauffer, P. N. Suffys, H. Takiff, J. Texier-Maugein, V. Vincent, J. H. de Waard, C. Sola, and N. Rastogi. 2002. Global distribution of Mycobacterium tuberculosis spoligotypes. Emerg.Infect.Dis. **8**:1347-1349.

8. **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.

9. **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 .

10. **Garcia, d., V, E. Bouza, N. Rastogi, and C. Sola**. 2005. Analysis of Mycobacterium tuberculosis genotypes in Madrid and identification of two new families specific to Spain-related settings. J.Clin.Microbiol. **43**:1797-1806.

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

12. **Gutacker, M. M., J. C. Smoot, C. A. Migliaccio, S. M. Ricklefs, S. Hua, D. V. Cousins, E. A. Graviss, E. Shashkina, B. N. Kreiswirth, and J. M. Musser**. 2002. Genome-Wide Analysis of Synonymous Single Nucleotide Polymorphisms in Mycobacterium tuberculosis Complex Organisms. Resolution of genetic relationships among closely related microbial strains. Genetics **162**:1533-1543.

13. **Hermans, P. W., F. Messadi, H. Guebrexabher, D. van Soolingen, P. E. de Haas, H. Heersma, H. de Neeling, A. Ayoub, F. Portaels, and D. Frommel**. 1995. Analysis of the population structure of Mycobacterium tuberculosis in Ethiopia, Tunisia, and The Netherlands: usefulness of DNA typing for global tuberculosis epidemiology [see comments]. J.Infect.Dis. **171**:1504-1513.

14. **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.

15. **Kremer, K., D. van Soolingen, R. Frothingham, W. H. Haas, P. W. Hermans, C. Martin, P. Palittapongarnpim, B. B. Plikaytis, L. W. Riley, M. A. Yakrus, J. M.**

Musser, and J. D. van Embden. 1999. Comparison of methods based on different molecular epidemiological markers for typing of Mycobacterium tuberculosis complex strains: interlaboratory study of discriminatory power and reproducibility. J.Clin.Microbiol. **37**:2607-2618.

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

17. **Masjedi, M. R., P. Farnia, S. Sorooch, M. V. Pooramiri, S. D. Mansoori, A. Z. Zarifi, A. Akbarvelayati, and S. Hoffner**. 2006. Extensively drug-resistant tuberculosis: 2 years of surveillance in Iran. Clin.Infect.Dis. **43**:841-847.

18. **Richardson, M., S. W. van Lill, G. D. van der Spuy, Z. Munch, C. N. Booysen, N. Beyers, P. D. van Helden, and R. M. Warren**. 2002. Historic and recent events contribute to the disease dynamics of Beijing-like Mycobacterium tuberculosis isolates in a high incidence region. Int.J.Tuberc.Lung Dis. **6**:1001-1011.

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

20. **Streicher, E. M., T. C. Victor, C. Kewley, A. Jordaan, G. D. van der Spuy, M. Bosman, H. Louw, M. B. Murray, D. Young, P. van Helden, and R. M. Warren**. 2006. Spread of an emerging Mycobacterium tuberculosis drug resistant strain in the Western Cape of South Africa. Int.J.Tuberc.Lung Dis.

21. **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.

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 [see comments]. J.Clin.Microbiol. **31**:406-409.

23. **Velayati, A. A., P. Farnia, M. Mirsaeidi, and M. M. Reza**. 2006. The most prevalent Mycobacterium tuberculosis superfamilies among Iranian and Afghan TB cases. Scand.J.Infect.Dis. **38**:463-468.

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

25. **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 IS6110 copy numbers. J.Clin.Microbiol. **42**:5774-5782.

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

Chapter 8

Spoligotype signatures in *Mycobacterium tuberculosis* Complex

EM Streicher[1], TC Victor[1], G van der Spuy[1], C Sola[2], N Rastogi[2], PD van Helden[1], RM Warren[1]

[1]DST/NRF Centre of Excellence in Biomedical TB Research, Division of Molecular Biology and Human Genetics, Faculty of Health Sciences, Stellenbosch University;

[2]Institut Pasteur de Guadeloupe, Pointe à Pitre, Guadeloupe.

Published in Journal of Clinical Microbiology 2007 Jan;45(1):237-40

My contribution to this project:	Spoligotyping of isolates Description of spoligotype signatures for different strain families Analysis of data Writing and editing of manuscript

ABSTRACT

Evolution of the Direct Repeat region in *Mycobacterium tuberculosis* has created unique spoligotype signatures specifically associated with IS*6110*-defined strain families. Spoligotyping signatures may enable the analysis of the strain population structure in different settings and will enable the rapid identification of strain families that acquire drug-resistance or escape protective immunity in drug and vaccine trials.

TEXT

Numerous repeat sequences have been identified in the genome of *Mycobacterium tuberculosis* complex, including transposable elements (3), trinucleotide repeats (36), Variable Number Tandem Repeat (VNTR) (10), Mycobacterial Interspersed Repetitive Units (MIRU) (27) and the direct repeat (DR) region (29). The DR region is one of the most extensively studied loci and consists of direct repeat sequences (36bp) interspersed with unique spacer sequences (34-41bp), which together are termed Direct Variable Repeat (DVR) sequences (15,29). The DR region has evolved through deletion of DVR sequences by homologous recombination, single nucleotide mutations and the integration of IS*6110* elements (1,29,34). These events are believed to be unidirectional and occur over time, making the DR region an informative locus to study the evolution and epidemiology of *M. tuberculosis* complex (9,13,35).

Spoligotyping was developed as a genotyping tool to provide information on the structure of the DR region in individual *M. tuberculosis* strains and in different members of the *M. tuberculosis* complex (16). The simplicity of this method has allowed for the

establishment of an international spoligotype database which describes 39,295 entries from 122 countries (4). Alignment of the spoligotype patterns has allowed authors to group isolates according to similarity to create clades or strain families (8). In addition, distinctive spoligotype patterns have been linked to defined species of the *M. tuberculosis* complex (8,18). However, these evolutionary relationships have not been extensively tested with other genotyping methods (9). A positive association between spoligotype and SNP cluster groups (SCG) could not be demonstrated in all instances (9). We hypothesize that the evolution of the DR region has created unique patterns of DVR deletion within the spoligotype and that these patterns are specific to the different IS*6110*-DNA fingerprinting defined strains families within the described SCG's. These unique patterns of DVR deletion have been termed spoligotype signatures.

In order to determine whether evolutionary relationships exist between the IS*6110*-defined families and spoligotype patterns, *M. tuberculosis* isolates from patients resident in the epidemiological field site near Cape Town, were genotyped according to the internationally standardized IS*6110*-DNA fingerprinting method (28). DNA fingerprints were analyzed with GelCompar software, using UPGMA and Dice coefficients (14). Isolates with an IS*6110* similarity index of ≥65% were grouped into strain families (22). In this study, DNA fingerprints from isolates of *M. tuberculosis* were available from 834 TB patients collected between 1993 and 1998, and these were grouped into 33 strain families. Representative strains from each strain family were classified into principal genetic groups (PGG) by DNA sequencing of the *katG* and *gyrA* genes (Table 1) (25).

Table 1 Spoligotyping signatures of *Mycobacterium tuberculosis* complex

PGG [a]	Species name (Strain Family [b]; number cases)	ST nr [c]	Core spoligotype pattern	International family name (reference) [d]
1	*M. africanum*	181	■■■■■□□□■■■■■■■■■■■■■■■■■■■■■■■■□■■■■	AFRI1 (4,8) *
1	*M. africanum*	331	■■■■■■□□□□□■■■■■■■□□□□■■■■■■■■■■□□□■■■■	AFRI2 (4,8) *
1	*M. africanum*	438	■■■■■■□□□□□■■■■■■■■■■■■■■■■■■■■□□□■■■■	AFRI3 (4,8) *
1	*M. bovis*	482	■■□■■■■■□■■■■■■□■■■■■■■■■■■■■■■■■■■■□□□□□	BOV1/*M. bovis*-BCG (4,8) *
1	*M. caprae*	647	□■□□□□□□□□□□□□□□■■■■■■■■■■□■■■■■■■■■□□□□□	BOV4 (4)
1	*M. pinnipedii*	593	□□□■■■■□□□□□□□□□□□□□□■■■■■■■■■■■■■■■□□□□□	PINI1 (4)
1	*M. pinnipedii*	637	□□□□□□□□□□□□□□□□□□□□□□□■■■■■■■■■■■■■■□□□□□	PINI2 (4)
1	*M. canettii*	592	□□□□□□□□□□□□□□□□□□□□□□□□□□□■□□□□□■□□□□□□□	canettii (4)
1	*M. microti*	539	□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□■■□□□□□	microti (4)
1	*M. tuberculosis* (F27; n = 6)	1	□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□■■■■■■■■	Beijing (12) *
1	*M. tuberculosis* (F29; n = 318)	1	□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□■■■■■■■■	Beijing (12) *
1	*M. tuberculosis* (F31; n = 31)	1	□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□■■■■■■■■	Beijing (12) *
1	*M. tuberculosis* (F25; n = 12)	21	■■■□□□□■□■■■■■■■■■■■■■□□□□□□□□□□□□■■■■■■■■	CAS1/Kili (2,4,8,24)
1	*M. tuberculosis* (F20; n = 9)	26	■■■□□□□■■■■■■■■■■■■■■■□□□□□□□□□□□□□□■■■■■■	CAS1/Delhi (2,4,8,24)
1	*M. tuberculosis*	288	■■■□□□□■■■■■■■■■■■■■■■□□□□□□□□□□□□□□■■■■■■■■	CAS2 (2,4,24) *
1	*M. tuberculosis*	19	■■□■■■■■■■■■■■■■■■□□■■■■■■■□□□□■□■■■■■■■■■	Manila/EAI2 (4,6) *
1	*M. tuberculosis*	236	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□□□□■□■■■■■■■■■	Ancesterial Mtb/EAI (26)
2	*M. tuberculosis* (F26; n = 11)	1241	■■■■■■■■■■■■■■■■■■■■■■■□□□□■■■■■■□□□□□□■■■■■■■	
2	*M. tuberculosis* (F13; n = 26)	20	■■□■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■□□□□■■■■■■■	LAM1 (4,8) *
2	*M. tuberculosis* (F11; n = 304)	33	■■■■■■■■□□□■■■■■■■■■■■□□□□■■■■■■■□□□□■■■■■■■	LAM3 (8,31) *
2	*M. tuberculosis* (F9; n = 28)	59	■■■■■■■■■■■■■■■■■■■■■■■□□□□■■□□□□■■□□□□■■■■■■■	LAM11-ZWE(4,7)
2	*M. tuberculosis* (F15; n = 17)	60	■■■■■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■□□□□■■■□■■■	LAM4 (4,8)
2	*M. tuberculosis*	209	■■■■■■■■□□□□□□■■■■■■■□□□□■■■■■■■□□□□■■■■■■■	LAM12/Madrid 1 (4,11)
2	*M. tuberculosis*	41	■■■■■■■■■■■■■■■■■■■■■■■□□□□■□□■■■■■□□□□■■■■■■■	LAM7-TUR (4,37)
2	*M. tuberculosis*	61	■■■■■■■■■■■■■■■■■■■■■■■■□□□■■■■■■■□□□□■■■■■■■	LAM10-Cameroon (4,20,21)
2	*M. tuberculosis* (F14; n = 52)	765	■■■■■■■■■■■■■■■■■■■□□□□□□■■■■■■■■■■□□□□■■■■■■■	
2	*M. tuberculosis* (F30; n = 1)	2309	■■■■■■■□□□■■■■■■■■■□□□□□□■■■■■■□□□□□□■■■■■■	
2	*M. tuberculosis* (F28; n = 122)	34	■■■■■■■■□□■■■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■	S/PZA^{R}-Quebec (4,8,19) *
2	*M. tuberculosis* (F7; n = 25)	136	■■■■■■■■■□□□□□■■■□□□□■□■■■■■■□□□□■□□■■■■	
2	*M. tuberculosis* (F3; n = 4)	102	■■■■■■■■■■□□□□■■■■■■■■■■■■■■■■□□□□■■■■■■■	
2	*M. tuberculosis* (F19; n = 6)	37	■■■■■■■■■■■□■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■	T3 (8)
2	*M. tuberculosis* (F6; n = 16)	39	■■■■■■■■■■■■■■■■■□■■□□■■■■■■■■□□□□■□□■■■■	T4-CE1(4)
2	*M. tuberculosis* (LCC; n = 230) [e]	119	■■■■■■■■■■■■■■■■■□■■■■■■■■■■■■■■□□□□■■■■■■■	X1 family (4,8,35)
2	*M. tuberculosis* (1banders; n = 21) [f g]	53	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■	T1 (8) *
2	*M. tuberculosis* (F1; n = 4)	47	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□□□□□□■□□□□■■■■■■■	Haarlem 1 (4,8,17) *
2	*M. tuberculosis* (F10; n = 6)	47	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□□□□□□■□□□□■■■■■■■	Haarlem 1 (4,8,17) *
2	*M. tuberculosis* (F2; n = 57)	50	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□■□□□□■■■■■■■	Haarlem 3 (4,8,17)
2	*M. tuberculosis* (F4; n = 47)	50	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□■□□□□■■■■■■■	Haarlem 3 (4,8,17)
2	*M. tuberculosis* (F24; n = 24)		No consistent signature deletion identified	
3	*M. tuberculosis* (F22; n = 4) [g]	53	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■	T1 (8) *
3	*M. tuberculosis* (F5; n = 2)	52	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□□□□■■■□■■■	T2 (8)
3	*M. tuberculosis* (F8; n = 4)	1067	■■■■■■□□□□□□■■■■■■■□□□■■■■■■■□□□□■■■□□□□	
3	*M. tuberculosis* (F12; n = 4)		■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□■■■■■■■□□□□■■■■■■■	
3	*M. tuberculosis* (F16; n = 16)		No consistent signature deletion identified	
3	*M. tuberculosis* (F17; n = 12)		No consistent signature deletion identified	
3	*M. tuberculosis* (F18; n = 31)		No consistent signature deletion identified	
3	*M. tuberculosis* (F21; n = 29)		No consistent signature deletion identified	
3	*M. tuberculosis* (F23; n = 30)		No consistent signature deletion identified	
3	*M. tuberculosis*	58	■■■■■■■■■■■■■■■■■■■■■□■■□■■■■■■■■■□□□□■■■■■■■	Madrid 2 (11)

[a] PGG = Principal genetic groups according to *katG*463 and *GyrA*95 classification (25)
[b] Strain classification of local families for *Mycobacterium tuberculosis* according to Richardson et al. (22)
[c] Shared-Type number as described in SpolDB4 (4)
[d] * = match the SPOTCLUST algorithm using RIM (32).
[e] LCC = Low copy clade (35). These strains have 2-6 bands on the IS*6110* fingerprint pattern
[f] The strains in this strain family has one copy of IS*6110* (35).
[g] Although there is a uniform spoligotype pattern, no signature other than the deletion of spacers other than 33-36 is present.

At least one strain representing each IS*6110* banding pattern within a strain family was subjected to spoligotyping using the internationally standardized method (16). In this study a spoligotype signature was defined as the deletion of either a single DVR or multiple DVRs unique to all members of a specific strain family. Random deletions of DVRs were ignored, as they probably represent recent evolutionary events which occurred after the evolutionary event which generated the signature and were not inherited by all progeny. Spoligotype signatures were compared to previously published spoligotype data (2,4,6-8,11,12,17,19-21,24,26,31,35,37).

Table 1 shows that the deletion of DVR's 33-36 was common to all members of PGG 2 and 3. Twenty-seven of 33 (82%) strain families had spoligotype signatures. Eighteen of these spoligotype signatures were unique to a specific strain family suggesting that these strain families had evolved independently. In contrast, five strain families (F9, F11, F13, F15, F26) shared a distinct spoligotype signature (deletion of DVR's 21-24), as well as, having evolved their own defining signature, suggesting that these families are closely related and had evolved from a common progenitor. This evolutionary relationship was supported by IS*6110* insertion site mapping (33). Comparison of these spoligotypes with the SpolDB4 database shows that these families form part of the previously described Latino-American and Mediterrannean (LAM) family (8). Family F11 has the characteristic LAM3 signature, F13 corresponds to LAM1 and F15 corresponds to LAM4 and F9 corresponds to LAM11-ZIM (Table 1) (4). Review of the literature showed that spoligotype signatures were also identified in the other members of the *M. tuberculosis* complex (Table 1) (4,16,23,29,30).

Spoligotype signatures were absent in 6 strain families (Table 1). The absence of a spoligotype signature was associated with PGG 2 and 3 (Table 1). These strains were previously described as the T strain family, thereby confirming a previous report which demonstrated that the T strain family spoligotypes were distributed in a number of different SCG's (9). Although the T strain family remains ill-defined, certain clones with in the T strain family have been characterized, by specific deletions of DVRs (Table 1 (4)). The presence of other defining signatures within the different T strain family members cannot be excluded as yet unrecognised signatures may have evolved within the DR regions of these strains and fall outside of the 43 DVR sequences routinely analysed.

The visual method of defining a spoligotype signature compared well with the previously described SPOTCLUST algorithm (32).

We acknowledge that the definition of a spoligotype signature used in this study has certain limitations as it is possible that extensive deletion of DVR sequences may have lead to convergence, although such events are rare (34). Similarly, deletion of a small number of DVR's may lead to convergence resulting in misclassification. However, in this study no cases of misclassification were detected. Together this highlights a potential limitation in the use of spoligotypes in phylogenetic analysis.

Despite these limitations, spoligotyping remains a highly informative genotyping method and the identification of strain family specific signatures provides an important marker for clonality. The classification of most strains into distinct evolutionary groups will enable the rapid stratification of patients in studies aimed to identify pathogenic

characteristics associated with the disease causing strain (5,9). Furthermore, using this method it will be possible to rapidly identify emerging strain families. We propose that the identification of spoligotype signatures will provide a means to determine the strain population structure in different geographical settings and on global scale. Spoligotype signatures will also be an important tool in the monitoring of drug and vaccine trials as it will enable the detection of strain families which may have a greater propensity to acquire drug resistance or to escape protective immunity.

Acknowledgements:

The authors would like to thank South African National Research Foundation (grants GUN 2054278 and DST/NRF Centre of Excellence in Biomedical Tuberculosis Research), the Wellcome Trust (grant Ref. 072402/Z/03/Z) and the IAEA (grant SAF6008) for financial support.

Reference List

1. **Aga, R. S., E. Fair, N. F. Abernethy, K. DeRiemer, E. A. Paz, L. M. Kawamura, P. M. Small, and M. Kato-Maeda**. 2006. Microevolution of the Direct Repeat Locus of Mycobacterium tuberculosis in a Strain Prevalent in San Francisco. J.Clin.Microbiol. **44**:1558-1560.

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

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

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

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

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

7. **Easterbrook, P. J., A. Gibson, S. Murad, D. Lamprecht, N. Ives, A. Ferguson, O. Lowe, P. Mason, A. Ndudzo, A. Taziwa, R. Makombe, L. Mbengeranwa, C. Sola, N. Rastogi, and F. Drobniewski**. 2004. High rates of clustering of strains causing tuberculosis in Harare, Zimbabwe: a molecular epidemiological study. J.Clin.Microbiol. **42**:4536-4544.

8. **Filliol, I., J. R. Driscoll, D. van Soolingen, B. N. Kreiswirth, K. Kremer, G. Valetudie, D. D. Anh, R. Barlow, D. Banerjee, P. J. Bifani, K. Brudey, A. Cataldi, R. C. Cooksey, D. V. Cousins, J. W. Dale, O. A. Dellagostin, F.**

Drobniewski, G. Engelmann, S. Ferdinand, D. Gascoyne-Binzi, M. Gordon, M. C. Gutierrez, W. H. Haas, H. Heersma, G. Kallenius, E. Kassa-Kelembho, T. Koivula, H. M. Ly, A. Makristathis, C. Mammina, G. Martin, P. Mostrom, I. Mokrousov, V. Narbonne, O. Narvskaya, A. Nastasi, S. N. Niobe-Eyangoh, J. W. Pape, V. Rasolofo-Razanamparany, M. Ridell, M. L. Rossetti, F. Stauffer, P. N. Suffys, H. Takiff, J. Texier-Maugein, V. Vincent, J. H. de Waard, C. Sola, and N. Rastogi. 2002. Global distribution of Mycobacterium tuberculosis spoligotypes. Emerg.Infect.Dis. **8**:1347-1349.

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. **Frothingham, R. and W. A. Meeker-O'Connell**. 1998. Genetic diversity in the Mycobacterium tuberculosis complex based on variable numbers of tandem DNA repeats. Microbiology **144 (Pt 5)**:1189-1196.

11. **Garcia, d., V, E. Bouza, N. Rastogi, and C. Sola**. 2005. Analysis of Mycobacterium tuberculosis genotypes in Madrid and identification of two new families specific to Spain-related settings. J.Clin.Microbiol. **43**:1797-1806.

12. **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.

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

14. **Hermans, P. W., F. Messadi, H. Guebrexabher, D. van Soolingen, P. E. de Haas, H. Heersma, H. de Neeling, A. Ayoub, F. Portaels, and D. Frommel**. 1995. Analysis of the population structure of Mycobacterium tuberculosis in Ethiopia, Tunisia, and The Netherlands: usefulness of DNA typing for global tuberculosis epidemiology [see comments]. J.Infect.Dis. **171**:1504-1513.

15. **Hermans, P. W., D. van Soolingen, E. M. Bik, P. E. de Haas, J. W. Dale, and J. D. van Embden**. 1991. 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. **59**:2695-2705.

16. **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.

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

18. **Molhuizen, H. O., A. E. Bunschoten, L. M. Schouls, and J. D. van Embden**. 1998. Rapid detection and simultaneous strain differentiation of Mycobacterium tuberculosis complex bacteria by spoligotyping. Methods Mol.Biol. **101**:381-394.

19. **Nguyen, D., P. Brassard, J. Westley, L. Thibert, M. Proulx, K. Henry, K. Schwartzman, D. Menzies, and M. A. Behr**. 2003. Widespread pyrazinamide-resistant Mycobacterium tuberculosis family in a low-incidence setting. J.Clin.Microbiol. **41**:2878-2883.

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

21. **Niobe-Eyangoh, S. N., C. Kuaban, P. Sorlin, J. Thonnon, V. Vincent, and M. C. Gutierrez**. 2004. Molecular characteristics of strains of the cameroon family, the major group of Mycobacterium tuberculosis in a country with a high prevalence of tuberculosis. J.Clin.Microbiol. **42**:5029-5035.

22. **Richardson, M., S. W. van Lill, G. D. van der Spuy, Z. Munch, C. N. Booysen, N. Beyers, P. D. van Helden, and R. M. Warren**. 2002. Historic and recent events contribute to the disease dynamics of Beijing-like Mycobacterium tuberculosis isolates in a high incidence region. Int.J.Tuberc.Lung Dis. **6**:1001-1011.

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

24. **Singh, U. B., N. Suresh, N. V. Bhanu, J. Arora, H. Pant, S. Sinha, R. C. Aggarwal, S. Singh, J. N. Pande, C. Sola, and P. Seth**. 2004. Predominant tuberculosis spoligotypes, Delhi, India. Emerg.Infect.Dis. **10**:1138-1142.

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

26. **Sun, Y. J., A. S. Lee, S. T. Ng, S. Ravindran, K. Kremer, R. Bellamy, S. Y. Wong, D. van Soolingen, P. Supply, and N. I. Paton**. 2004. Characterization of ancestral Mycobacterium tuberculosis by multiple genetic markers and proposal of genotyping strategy. J.Clin.Microbiol. **42**:5058-5064.

27. **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.

28. **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 [see comments]. J.Clin.Microbiol. **31**:406-409.

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

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

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

32. **Vitol, I., J. Driscoll, B. Kreiswirth, N. Kurepina, and K. P. Bennett**. 2006. Identifying Mycobacterium tuberculosis complex strain families using spoligotypes. Infect.Genet.Evol.

33. **Warren, R. M., S. L. Sampson, M. Richardson, G. D. van der Spuy, C. J. Lombard, T. C. Victor, and P. D. van Helden**. 2000. Mapping of IS*6110* flanking regions in clinical isolates of *M. tuberculosis* demonstrates genome plasticity. Mol.Microbiol. **37**:1405-1416.

34. **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.

35. **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 IS6110 copy numbers. J.Clin.Microbiol. **42**:5774-5782.

36. **Wiid, I. J., C. Werely, N. Beyers, P. Donald, and P. D. van Helden**. 1994. Oligonucleotide (GTG)5 as a marker for Mycobacterium tuberculosis strain identification. J.Clin.Microbiol. **32**:1318-1321.

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

Chapter 9

General Conclusion

With the introduction of anti-TB drugs in the 1940s and especially after rifampin was introduced in the late 1960s, a steady decline in TB rates was observed (3) and predictions were made that it was only a matter of time before TB would be eliminated. Two major factors contributed to the failure of eliminating TB – coinfection with HIV and the development of drug resistance (11). Before the 1990s the only markers to study the epidemiology of TB were drug susceptibility patterns and phage types (4) and many questions in TB epidemiology remained unresolved. After the identification of the IS*6110* transposable element (7,8) and the complete sequence of the *M. tuberculosis* genome (2), a series of genotyping methods were developed that contributed to the understanding of TB epidemiology. Genotyping methods were central in studying the effects of recent transmission, exogenous reinfection and reactivation of latent disease and also in the identification of drug resistant outbreaks of TB (5,6,9-12).

Molecular epidemiology of drug resistant TB

Through the establishment of a Longitudinal database of drug resistant TB isolates with phenotypic, genotypic and clinical data the drug resistant epidemic in the Western Cape Province, South Africa was described. The drug resistance epidemic is mostly due to transmission of drug resistant strains and about 40% of the drug resistant strains are MDR. An outbreak of an emerging drug resistant strain was identified and the genotypic characteristics of strains and clinical and demographical characteristics of cases infected were described. Inappropriate chemotherapy; poor adherence to treatment; prolonged periods of infectiousness due to delay of susceptibility testing led to the development and transmission of this drug resistant genotype.

Recurrent TB

We showed that spoligotyping can be applied successfully to differentiate between the mechanisms of recurrence, relapse and reinfection, which are important factors in determining the efficacy of treatment within different settings and also to identify factors influencing disease dynamics.

Application of spoligotyping and a newly developed PCR-method showed that occurrence of multiple infections are higher than what was previously assumed and also more frequent in retreatment cases than in new cases. These results suggest that the initial infection is unable to provide protection against a subsequent infection in this population and will have important implications for the understanding of protective immunity and the development and testing of new vaccines and drugs for use in communities where the burden of disease is high.

Evolution of *M. tuberculosis*

Although IS*6110*-fingerprinting was central in evolutionary studies, the Low copy strains remained an ill defined group on the basis of IS*6110*-fingerprinting. This study demonstrated the evolutionary scenario of strains with low copy numbers of IS*6110*, which were previously poor defined, based on different molecular markers, including spoligotyping. It was also shown that the LCC is widely disseminated and plays an important role in the global tuberculosis epidemic. We also showed the evolutionary

scenario of the principle genetic group 2 strains of which the low copy strains are part, based on different molecular markers. Previously unknown genetic relationships between strain families were identified and it was shown, in contrast to what was previously assumed, that a LCC ancestor was the progenitor of high IS*6110* copy number strains. Spoligotyping signatures, created by evolution of the Direct Repeat region in *M. tuberculosis,* were identified in different strain families of *M. tuberculosis* in our local setting, Spoligotype signatures will enable the analysis of the strain population structure in different settings and the rapid identification of strain families that acquire drug-resistance or escape protective immunity in drug and vaccine trials.

.

Future studies

The studies described in this dissertation laid the foundation for further research in the following ways: a longitudinal database of drug resistant TB in the Western Cape Province was established. This database can be expanded to isolates from the rest of South Africa and also from other countries in Africa. The database can be applied to monitoring the strain dynamics which impact the control of TB over different time periods and will be important in the development of molecular markers for early detection of drug resistant TB outbreaks, especially Extensively/Extreme drug resistant TB Strains (XDR-TB), which are resistant to first-line as well as second-line drugs and is almost impossible to treat (1).

In this study, methods to successfully identify the mechanisms of recurrence and to measure the extent of dual strain infections were established. This can now be applied in

different study settings to identify the influences that different strain families have on reinfection. The studies on the evolution of *M. tuberculosis* strain families laid the foundations to study phenotypic differences and is an important foundation for new vaccine development. The spoligotype signatures identified and described can in future be applied in other settings to describe an overview of the strain dynamics in these settings.

This study contributed to our understanding of the molecular biology of drug resistant TB, recurrent TB and the evolution of *M. tuberculosis* in high incidence communities.

Reference List

1. **Centers for Disease control and Prevention (CDC)**. 2006. Emergence of Mycobacterium tuberculosis with extensive resistance to second-line drugs--worldwide, 2000-2004. MMWR Morb.Mortal.Wkly.Rep. **55**:301-305.

2. **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.

3. **Espinal, M. A.** 2003. The global situation of MDR-TB. Tuberculosis.(Edinb.) **83**:44-51.

4. **Gruft, H., R. Johnson, R. Claflin, and A. Loder**. 1984. Phage-typing and drug-resistance patterns as tools in mycobacterial epidemiology. Am.Rev.Respir.Dis. **130**:96-97.

5. **Johnson, R., Warren, R. M., Strauss, O. J., Jordaan, A., Falmer, A. A., Beyers, N., Schaaf, H. S., Murray, M., Cloete, K., van Helden, P. D., and Victor, T. C.** An outbreak of drug resistant Tuberculosis caused by a Beijing strain in the Western Cape, South Africa. Int.J.Tuberc.Lung Dis. 2006.

6. **Murray, M. and E. Nardell**. 2002. Molecular epidemiology of tuberculosis: achievements and challenges to current knowledge. Bull.World Health Organ **80**:477-482.

7. **Thierry, D., A. Brisson-Noel, V. Vincent-Levy-Frebault, S. Nguyen, J. L. Guesdon, and B. Gicquel**. 1990. Characterization of a Mycobacterium tuberculosis insertion sequence, IS6110, and its application in diagnosis. J.Clin.Microbiol. **28**:2668-2673.

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

9. **van Rie, A., T. C. Victor, M. Richardson, R. Johnson, G. D. van der Spuy, E. J. Murray, N. Beyers, N. C. Gey Van Pittius, P. D. van Helden, and R. M. Warren**. 2005. Reinfection and Mixed Infection Cause Changing Mycobacterium tuberculosis Drug-resistance Patterns. Am.J.Respir.Crit Care Med.

10. **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.

11. **van Soolingen, D.** 2001. Molecular epidemiology of tuberculosis and other mycobacterial infections: main methodologies and achievements. J.Intern.Med. **249**:1-26.

12. **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.

Other publications

Journal articles

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

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

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

W. A. Githui, A. M. Jordaan, E. S. Juma, P. Kinyanjui, F. G. Karimi, J. Kimwomi, H. Meme, P. Mumbi, **E. M. Streicher**, R. Warren, P. D. van Helden, and T. C. Victor. 2004. **Identification of MDR-TB Beijing/W and other Mycobacterium tuberculosis genotypes in Nairobi, Kenya.** Int.J.Tuberc.Lung Dis. 8:352-360.

R. M. Warren, **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.

Book Chapter

R. Johnson*, **E.M. Streicher***, G.E. Louw, R.M. Warren, P.D. van Helden, T.C. Victor. 2005. **Drug Resistance in Mycobacterium tuberculosis.** In Mycobacterial Molecular Microbiology. Edited by Tanya Parish, Eds Horizon Scientific Press. (Chapter 5, Pages 169-197).
* **equal contribution as principle author.**