IDENTIFICATION OF MECHANISMS REGULATING THE INTRA CELLULAR CONCENTRATION OF RIFAMPICIN IN *MYCOBACTERIUM TUBERCULOSIS*

by

Margaretha de Vos

Dissertation presented for the degree of Doctor Philosophy in Medical Sciences (Molecular Biology) in the Faculty of Medicine and Health Sciences at Stellenbosch University

Promoter: Prof Thomas Calldo Victor

Co-Promoter: Prof Robin Mark Warren

March 2013

DECLARATION

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated) that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: March 2013

SUMMARY

Rifampicin resistance in clinical isolates of *Mycobacterium tuberculosis* develops through selection of bacterial variants harbouring mutations in the *rpoB* gene. These mutations infer a fitness-cost in the absence of antibiotic pressure, however, fitness-levels of rifampicin-resistant strains can be restored by compensatory mutations in *rpoA* and *rpoC.* This study was the first to investigate the epidemiological relevance of these compensatory mutations in clinical *M. tuberculosis* isolates collected in South Africa. Through targeted DNA sequencing, we demonstrated a strong association between *rpoC* mutations and transmission, and the *rpoB* S531L mutation. Our study emphasises the epidemiological relevance of compensatory evolution in response to the emergence of rifampicin resistance, and illustrates how compensatory mutations may be selected as a function of epistatic interactions.

Recently a hypothesis has been developed which suggests that the activation of efflux systems through exposure to rifampicin may explain the observed spectrum of rifampicin resistance phenotypes. To elucidate whether rifampicin dependent activation of efflux systems also increases energy production, the RNA expression profiles of candidate energy metabolism genes were investigated. This study demonstrated that rifampicin exposure induced an overall increase in the expression of energy metabolism genes. Our findings suggest that the response to rifampicin is not universal and may depend on other genomic mutations. From these results we conclude that the stress response induced by exposure to rifampicin increases the energy production which fuels efflux activity thereby enabling the cell to extrude rifampicin in an energy dependent manner. This also provides a platform to explain the mechanism by which the newly developed drug, TMC207, increases the rate of culture conversion when used in combination with second-line anti-TB drugs. We propose that inhibition of ATP synthesis by TMC207 will deprive the efflux pumps and transporter genes of energy, which will result in the accumulation of second-line anti-TB drugs within the bacilli, leading to more efficient binding of the second-line drugs to their targets and ultimately to cell death.

To identify the genetic basis governing the level of rifampicin resistance, we sequenced the genomes of MDR clinical isolates and *in vitro* generated rifampicin resistant mutants. Only minor genetic changes in addition to the *rpoB* mutation were identified in the genomes of *in vitro* rifampicin resistant mutants which displayed varying levels of resistance. This suggests that these mutants may either use alternative regulatory mechanisms or have acquired SNPs outside the genetic regions investigated in this study to modulate rifampicin resistance levels. In contrast, the genomes of clinical MDR isolates from the Low Copy Clade showed considerable variability in genes involved in cell wall, cellular processes and lipid metabolism, while the genomes from the Beijing Clade displayed

variability in genes known to confer drug resistance and compensatory mechanisms. These results suggest that the structure and processes of the cell wall, as well as lipid metabolism plays a critical role in determining the intra-cellular concentration of rifampicin. Finally, this study illustrated the complexity in the physiology of *M. tuberculosis* resistant to rifampicin, whereby multiple mechanisms are employed by the bacteria to modulate its resistance levels.

OPSOMMING

Rifampisien weerstandigheid in kliniese isolate van *Mycobacterium tuberculosis* ontwikkel deur die seleksie van bakteriële variante wat mutasies in die *rpoB* geen het. Alhoewel hierdie mutasies lei tot ‘n afname in fiksheid van die bakterieë in die teenwoordigheid van antibiotika, kan die fiksheids vlakke van rifampisien weerstandige stamme herstel word deur vergoedende mutasies in *rpoA* en *rpoC*. Hierdie is die eerste studie wat die epidemiologiese relevansie van hierdie vergoedende mutasies in kliniese *M. tuberculosis* isolate wat in Suid-Afrika versamel is, ondersoek. Deur middel van doelgerigte DNA volgordebepaling het ons ‘n sterk assosiasie tussen *rpoC* mutasies en transmissie, en die *rpoB* S31L mutasie getoon. Hierdie studie beklemtoon die epidemiologiese relevansie van regstellende evolusie na aanleiding van die ontwikkeling van rifampisien weerstandigheid en illustreer hoe regstellende mutasies geselekteer mag word as ‘n funksie van epistatiese interaksies.

‘n Hipotese is onlangs ontwikkel wat voorstel dat blootstelling aan rifampisien uitvloei sisteme in die bakterium aktiveer, wat moontlik die waargenome spektrum van rifampisien weerstandige fenotipes kan verklaar. Ons het die RNA uitdrukkingsprofiele van kandidaat-energiemetabolisme gene ondersoek om te bepaal of rifampisien afhanklike aktivering van uitvloei sisteme ook energieproduksie verhoog. Hierdie studie demonstreer dat rifampisien-blootstelling ‘n algehele verhoging in die uitdrukking van energiemetabolisme gene induseer. Ons bevindinge stel voor dat die reaksie van die sel op rifampisien blootstelling nie universeel is nie, en moontlik ook afhanklik is van ander genomiese mutasies. Uit hierdie resultate kan ons aflei dat die stres respons wat geïnduseer word deur rifampisien-blootstelling energieproduksie verhoog, wat weer die uitvloei aktiwiteit aanvuur, en gevolglik die sel in staat stel om rifampisien op ‘n energie-afhanklike wyse uit te dryf. Dit bied ook ‘n basis om die meganisme te verklaar waardeur die nuwe middel, TMC207, die tempo van kultuuromskakeling verhoog wanneer dit saam met tweede-linie anti-TB middels gebruik word. Ons stel voor dat die inhibisie van ATP sintese deur TMC207 die uitvloeipompe en transporteerder gene van energie ontneem. Gevolglik veroorsaak dit ‘n ophoping van tweedelinie anti-TB middels binne-in die bakterium, wat geleentheid bied vir meer effektiewe binding tussen die middels en hulle teikens en uiteindelik seldood veroorsaak.

Ons het DNA volgordes bepaal van die genome van MDR kliniese isolate en *in vitro* selekteerde rifampisienweerstandige mutante om sodoende die genetiese grondslag waarop die vlak van rifampisienweerstandigheid beheer word, te identifiseer. Slegs klein verskille, bo en behalwe die *rpoB* mutasie, is geïdentifiseer in die genome van *in vitro* rifampisien weerstandige mutante wat verskillende vlakke van weerstandigheid getoon het. Dit dui aan dat hierdie mutante of ander

regulatoriese meganismes gebruik, of hulle het enkelnukleotied polimorfismes buite die genetiese area wat in hierdie studie ondersoek is, waarmee rifampisien weerstandigheid gemoduleer word. In teenstelling hiermee het die genome van kliniese MDR isolate van die "Low Copy Clade" aansienlike variasie getoon in gene wat betrokke is by die selwand, sellulêre prosesse en lipiedmetabolisme. Verder het die genome van die Beijing genotipe variasie in gene getoon wat betrokke is by middelweerstandigheid en regstellende meganismes. Hierdie resultate dui aan dat die struktuur en prosesse van die selwand, asook lipiedmetabolisme, 'n kritiese rol speel in die bepaling van die intrasellulêre konsentrasie van rifampisien. Opsommend, hierdie studie toon verskeie meganismes aan wat deur die bakterieë gebruik word om weerstandigheidsvlakke te moduleer en die kompleksiteit van die fisiologie van *M. tuberculosis* wat weerstandig is teen rifampisien.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to the following people and institutions, without whom this thesis would not have been possible.

- My promoter Prof Tommie Victor, co-promoter Prof Rob Warren and mentor Dr Gail Louw for their excellent discussions and suggestions, patience and guidance in writing this thesis.

- All my colleques and friends at the Division of Molecular Biology and Human Genetics, in particular my "lab family" in Lab 453, Dr Borna Müller and Dr Monique Williams.

- Everyone who was part of my bioinformatics journey:
  Prof Gavin Sherlock, who gave me the opportunity to visit his group at Stanford University for training in the analysis of whole genome sequencing data and perl programming.
  Mrs Anzaan Dippenaar, Mr Keith Siame and Ms Mmakamohelo Direko (SANBI, UWC) for their assistance and guidance in the analysis of whole genome sequencing data.
  Mrs Michelle Daya for providing phython scripts for the analysis of the data.

- The National Research Foundation (NRF), the South African Medical Research Coucil (MRC), the Harry Crossley Foundation, the Aurum Institute and Stellenbosch Univeristy for financial support.

- My parents (Hennie Bester, Hannetjie Bester, Pieter de Vos and Diana de Vos), my sisters (Christelle Döman, Hannelie Bester and Este de Vos) and all my friends for their love, support and prayers.

- My best friend and husband, Pierrie de Vos, for his love and support through the best and worst of times.

God has blessed me immensely and I am grateful for everyone who was part of this adventure.

Jesus Christus, my Verlosser, aan U kom al die eer.

I can do all things through Christ who strengthens me.
Philippians 4:13

TABLE OF CONTENTS

DETAILED CONTENTS

LIST OF ABBREVIATIONS

°C	:	Degree Celsius
µl	:	microlitres
2-DE	:	2-Dimensional gel electrophoresis
ABC	:	ATP binding cassette
ADC	:	Albumin dextrose catalase
ATP	:	Adenosine triphospate
bp	:	base pairs
BSA	:	Bovine serum albumin
BWA	:	Burrows-Wheeler Aligner
cDNA	:	Complementary DNA
CFUs	:	Colony forming units
dH_2O	:	Distilled water
ddH_2O	:	Double distilled water
DNA	:	Deoxyribonucleic acid
dNTP	:	Deoxyribonecleotide triphosphate
DST	:	Drug susceptibility testing
EDTA	:	Ethylenediaminetetraacetic acid
g	:	Grams
GATK	:	Genome Analysis Toolkit
HIV	:	Human immunodeficiency virus
Indels	:	Short insertions and deletions
IS	:	Insertion sequence
LAM	:	Latin-American and Mediterranean
LCC	:	Low Copy Clade
LJ	:	Loewenstein Jensen
M. tuberculosis	:	*Mycobacterium tuberculosis*

MALDI-TOF	:	Matrix Assisted Lazer Desorption/Ionization Time of Flight
MDR	:	Multi Drug Resistant
MFS	:	The Major Facilitator Super family
MGIT	:	Mycobacterial Growth Indicator Tube
MIC	:	Minimum Inhibitory Concentration
ml	:	millilitres
mM	:	millimolar
mRNA	:	Messenger RNA
OADC	:	Oleic Albumin Dextrose Catalas
OD	:	Optical density
PCR	:	Polymerase chain reaction
QRT-PCR	:	Quantitative REAL-TIME PCR
RFLP	:	Restriction Fragment Length Polymorphism
rpm	:	Revolutions per minute
RNA	:	Ribonucleic acid
RND	:	Resistance-Nodulation-cell Division
RRDR	:	RIF Resistance Determining Region
rRNA	:	Ribosomal RNA
SA	:	South Africa
SAM	:	Sequence Alignment/Map
SDS	:	Sodium dodecyle sulphate
SMR	:	Small Multidrug Resistance
SNP	:	Single nucleotide polymorphism
TB	:	Tuberculosis
TDR	:	Total drug resistant
TBE	:	Tris/Borate/EDTA
TE	:	Tris/EDTA
Tm	:	Melting temperature

U	:	Units
V	:	Volt
XDR	:	Extreme Drug Resistant
ZN	:	Ziehl-Neelsen

CHAPTER 1
GENERAL INTRODUCTION

1.1. Background

South Africa is among the countries most severely affected by multidrug-resistant tuberculosis (MDR-TB), which is defined by the resistance to at least the two most potent anti-TB drugs namely isoniazid and rifampicin. It was estimated that 9.6% of all Tuberculosis cases in South Africa in 2008 were MDR-TB (227). Molecular epidemiological analyses of MDR and extensively drug resistant (XDR) strains (i.e. MDR strains with additional resistance to any fluoroquinolone and one of the injectable drugs kanamycin, amikacin or capreomycin) suggest that the high incidence of drug-resistant TB in South Africa is largely driven by transmission (92, 149).

Rifampicin is considered as one of the most important first line drugs used for the treatment of TB due to its strong sterilizing activity against both growing and stationary phase bacilli (89, 233). Rifampicin targets and subsequently inhibits RNA synthesis by binding to the β-subunit of the RNA polymerase (208). *Mycobacterium tuberculosis* (*M. tuberculosis*) can acquire resistance to rifampicin through mutations in the gene encoding its target, *rpoB* (152). Mutations in an 81 base pair core region of the *rpoB* gene known as the rifampicin Resistance Determining Region (RRDR) account for approximately 95% of rifampicin resistance detected in clinical isolates (152). The most frequently mutated codons being at mutations at amino acid positions 526 and 531, while mutations at 511, 516, 518, 522 and 533 are less frequently observed in the clinical setting (152, 158). A number of studies have shown that the acquisition of rifampicin resistance is associated with a fitness-cost (26, 63, 120). Drug resistance associated fitness cost can describe as reduced growth compared to the wild type bacteria in the absence of drug pressure. However, in some cases clinical isolates harbouring the *rpoB* S531L mutation had no fitness-cost compared to parent wild-type isolates (26, 63). For these clinical isolates it was suggested that either the *rpoB* S531L had no fitness-cost in these particular genetic backgrounds, or that secondary mutations were acquired to ameliorate the initial low fitness-cost of the *rpoB* S531L mutation (63). Compensatory mutations overcome the fitness-cost associated with drug resistance mutations, resulting in an overall fitness level comparable to a wild-type strain in the absence of antibiotics (11, 62, 63). Recently, a novel set of putative compensatory polymorphisms in RNA polymerase genes (*rpoA* and *rpoC*) were described in rifampicin-resistant *M. tuberculosis* isolates (40). In another study, whole genome sequencing revealed that 14 out of 24 of *M. tuberculosis* clinical isolates that harboured an *rpoB* mutation also carried non-synonymous polymorphisms in *rpoA* or *rpoC* (34). Since *rpoA/C* is highly conserved in *M. tuberculosis*, it is therefore assumed that these polymorphisms are recently acquired. Finally, a recent study using genetic reconstruction and competitive growth experiments in *Salmonella enteric* (*S. eterica*) has shown that secondary mutations in *rpoA*, *rpoB* and *rpoC* were necessary and sufficient for compensating fitness-costs incurred by rifampicin resistance mutations (28). The epidemiological relevance of compensatory mutations in *rpoA/C* is however unknown.

The frequency and level of resistance (as expressed by the minimum inhibitory concentration (MIC) values) of mutations in *rpoB* selected by the Luria Delbrück fluctuation assay has been studied in both clinical and laboratory strains of *M. tuberculosis* (82, 120, 132). These studies illustrated that mutations at amino acid positions 531 and 526 in *rpoB* are generally associated with higher resistant levels, whereas others (516, 522) generally display lower resistant levels. Important to note is that in some cases resistant levels were shown to be variable with mutants selected from the same progenitor and harbouring identical *rpoB* mutations (82, 120). However, for most mutations only a minimum MIC cut-off value was given in the respective studies and the true variability of resistance levels in *in vitro* mutants remains unclear (82, 120).

Similarly, rifampicin resistance levels were also shown to be highly variable in clinical MDR and mono-rifampicin resistant isolates, independent of the *rpoB* mutation and genetic background (114). It was also observed that the inhibition of efflux pumps and/or transporter proteins significantly restored rifampicin susceptibility in these clinical isolates, while differential expression of efflux and transporter genes were observed after rifampicin exposure (114). These results suggested the importance of these efflux and transporter proteins in modulating rifampicin resistance levels. In addition, when investigating the proteomic response of MDR isolates after rifampicin exposure, an increase in the abundance of proteins essential for the production of adenosine triphospate (ATP) were observed (M Bester, MSc thesis, 2009). This stimulated the formation of a new hypothesis which proposed that increased energy production is required to drive metabolic processes (including efflux pumps) needed to modulate the level of rifampicin resistance (M Bester, MSc thesis, 2009).

1.2. Problem Statement

The findings described above indicate that drug resistance in *M. tuberculosis* is more complicated than previously thought. Recently the concept of one mutation conferring one resistance phenotype was challenged by the observation that closely related clinical and *in vitro* rifampicin resistant *M. tuberculosis* isolates harbouring the same *rpoB* mutation display varying levels of rifampicin resistance (114). For these isolates it was suggested that biological mechanisms in addition to the *rpoB* mutation are responsible for the varying rifampicin resistant levels.

1.3. Overall Hypothesis

The MIC of rifampicin in clinical and *in vitro* rifampicin resistant isolates of *M. tuberculosis* is defined by mutations other than those in the *rpoB* gene and that the expression of these mutations define gene function and regulation, which in turn control the intracellular concentration of rifampicin.

1.4. Overall Aim

To identify novel mechanisms which influence the level of rifampicin resistance in clinical isolates and *in vitro* selected rifampicin resistant mutants of *M. tuberculosis.*

1.5. Objectives

1.5.1. To investigate the association of compensatory mutations in *rpoA/C* with the degree of antibiotic resistance in clinical isolates collected from a high incidence setting of MDR-TB in South Africa.

1.5.2. To investigate the expression levels of energy metabolism genes in response to rifampicin exposure in rifampicin resistant clinical *M. tuberculosis* isolates.

1.5.3. To characterize the genotypes and phenotypes of rifampicin resistance in *in vitro* rifampicin resistant mutants selected from the Beijing and H37Rv genetic backgrounds.

1.5.4. To identify genetic variants resulting in increased rifampicin resistance using whole genome sequencing of *in vitro* and clinical rifampicin resistant *M. tuberculosis* isolates.

CHAPTER 2
LITERATURE REVIEW

DRUG RESISTANCE BEYOND CLASSICAL MUTATIONS IN MYCOBACTERIA

2.1. Introduction

International guidelines recommend that drug susceptible TB should be treated with four first-line drugs (isoniazid, rifampicin, ethambutol, pyrazinamide) for 2 months followed by isoniazid and rifampicin for 4 months (224). However, non-compliance, the use of low quality drugs, malabsorption, and inappropriate treatment associated with diagnostic delay may lead to the spontaneous emergence of drug resistant TB (69, 93, 178, 228). Drug resistance has the potential to compromise the efficacy of treatment given the limited number of anti-TB drugs that are currently available. To date, four categories of drug resistance have been described for the pathogen *M. tuberculosis* (105); (i) Mono-resistance is defined as resistance to a single anti-TB drug, (ii) multiple resistance is defined as resistance to more than one of the first line drugs, including isoniazid or rifampicin, (iii) multi-drug resistant (MDR) TB is defined as *M. tuberculosis* strains that are resistant to at least isoniazid and rifampicin, whereas (iv) extensively drug resistant (XDR)-TB is defined as MDR-TB with additional resistance to second-line anti-TB drugs; a fluoroquinolone and any one of the injectable drugs i.e. kanamycin, amikacin and capreomycin (178, 235). More recently, a further resistance category has been proposed, namely totally drug resistant TB (TDR) which is defined as resistance to all first line and second-line drugs (219). However, this definition is controversial since not all of the second-line drug susceptibility testing methods have been standardized (101, 161).

It is well established that mycobacteria can acquire resistance to both first and second line anti-TB drugs as a function of genetic alterations in target genes, summarized in Table 2.1. Identifying the mode of action (and conversely the mechanism of resistance) of an antibiotic largely depends on whether the antibiotic displays a broad or narrow-spectrum activity against bacterial species. In the case of antibiotics displaying a broad-spectrum activity, one can assume that the mechanism of action would be similar in most bacterial species. This is true for anti-TB drugs such as rifampicin, streptomycin and fluoroquinolones where the mode of action in mycobacteria were derived from studies done in other bacterial species such as *Escherichia coli* (72, 73). Therefore, similar resistance mechanisms in *M. tuberculosis* would also be assumed. In contrast, antibiotics displaying a specific anti-mycobacterial activity might interact with targets that are unique to mycobacteria; needs to be converted to an active form by enzymes unique to mycobacteria; or requires uptake systems unique to mycobacteria. Anti-TB drugs that belong to this category are capreomycin, ethambutol, isoniazid, ethionamide and pyrazinamide. The mechanism of action and resistance was identified by studying the effect of these drugs on the bacterial cell. Scanning electron microscopy showed that isoniazid has as specific effect on mycolic acid synthesis and the observation that isoniazid resistant isolates had attenuated catalase-peroxidase activity prompted the identification of mutations conferring resistance in *katG* (184, 225). Similarly, the loss of pyrazinamidase activity was associated with pyrazinamide resistance (98). The mode of action of capreomycin was derived from structurally similar antibiotic,

viomycin, which is known to inhibit protein synthesis in *E. coli* (90, 191). Since both antibiotics target the 70S ribosome, the unique activity of capreomycin against *M. tuberculosis* is most likely explained by an unique mycobacterial uptake system (38).

The central dogma for drug resistance in *M. tuberculosis* has always involved the concept that one mutation results in one resistant phenotype. However, this concept has been challenged by the following: Firstly, it is well known that some mycobacterial species displays intrinsic resistance to anti-TB drugs through various cellular processes including a more adapted cell envelope, drug modifying enzymes and efflux mechanisms (14, 209). Secondly, it is also well documented that a certain percentage of drug resistant isolates do not harbour mutations in the classical resistance conferring genes (Table 2.2), suggesting that other resistance mechanisms exist for these isolates. Lastly, it has also been shown that mutations in classical drug resistance conferring genes results in various levels of resistance. In such instances the level of resistance, defined as the minimum inhibitory concentration (MIC), is the minimum concentration of an antibiotic that is required to kill 99% of the bacterial cells in that culture (27). This differs from the critical concentration which is the *in vitro* drug concentration that is able to kill 95% of bacterial isolates which have not been previously exposed to the drug (61). These varying phenotypes for a given drug can be attributed by the mutation itself, genetic background or additional genetic variants. These observations suggests that resistant phenotypes are more complex than previously thought and emphasise the need for high throughput technology to identify novel resistance mechanisms in *M. tuberculosis*.

This review will discuss evidence challenging the current dogma that one mutation results in one resistance phenotype, as well as the introduction and use of whole genome sequencing to identify novel mechanisms which modulates varying resistance phenotypes.

2.2. Evidence that challenge the dogma

2.2.1. Intrinsic ancient mechanisms displayed by other Mycobacteria

Intrinsic resistance is defined as the natural resistance displayed by bacteria to an antibiotic. The primary mechanisms underlying intrinsic resistance in mycobacteria include i) low permeability of the cell wall, ii) efflux systems, iii) the expression of antibiotic modifying and degrading enzymes, and iv) the expression of target-modifying enzymes (42, 43). These mechanisms may work independently or synergistically to protect the bacterium against the toxic effect of antibiotics (139). Such phenomena have been described in *Mycobacterium abscessus* for example, where a combination of low cell wall permeability and aminoglycoside-modifying enzymes confer resistance to aminoglycosides (138). The intrinsic mechanisms are discussed below.

2.2.1.1. The Mycobacterial cell envelope

The mycobacterial cell envelope is one of the characteristic features of mycobacteria and provides some level of resistance to most drugs and chemotherapeutic agents (87), since it forms the primary layer of protection against such extracellular compounds, including host defence factors such as lysozome. Additionally, it directly contributes to the virulence of mycobacteria such as those causing TB and leprosy (103). The impaired permeability of mycobacterial cell wall is likely due to high hydrophobicity and a unique arrangement of long-chain lipophylic mycolic acids, glycopeptidolipids, trehalose dimycolates, arabinogalactans and peptidoglycans (30, 87). This limits the permeation of both hydrophobic and hydrophilic molecules. It has been shown that the deletion of a reductase (encoded by MSMEG4722) in *M. smegmatis* involved in the final step of mycolic biosynthesis resulted in an increased sensitivity to hydrophobic drugs such as rifampicin (decrease in MIC from 16 µg/ml to 0.125 µg/ml), which are able to diffuse directly through the lipid membranes (23). Similarly, treatment of mycobacterium species with bacitracin resulted in increased cell wall permeability and a faster reduction of cell growth was observed when treated in combination with a hydrophobic rifamycin (118). This suggested that cell wall distortions caused by bacitracin resulted in enhanced diffusion of rifamycin into the cell (118). In *M. smegmatis* the rate of rifampicin influx is greatly reduced when compared to *M. tuberculosis* and *M. aurum*, indicating that reduced influx is a component of intrinsic rifampicin resistance in *M. smegmatis*(148).

The cell wall also contains porin-like proteins which have channel-forming activities for nutrient uptake and the export of waste products (102). Lethal amounts of hydrophilic antibiotics are sometimes able to pass through these porin-channels into the mycobacterial cell; however it has been shown that permeability rates of these channels are 100 to 1000 fold lower than those found in *Escherichia.coli* (88). The deletion of a porin, *mspA*, in *M. smegmatis* resulted in a 16 fold increase in β-lactam resistance, a 10 fold increase in vancomycin resistance and a 3 fold increase in rifampicin resistance (192). Conversely, over-expression of *mspA* in *M. bovis* and *M. tuberculosis* increased

susceptibility to ampicillin (2 fold decrease in MICs) (117), thereby demonstrating the inability of this hydrophilic compound to cross the mycobacterial cell wall and the importance of the *mspA* porin in intrinsic resistance to this β-lactam antibiotic. Therefore, the development of drugs that could alter the mycobacterial cell wall permeability and the introduction thereof in the current TB treatment regimen could be of great importance.

2.2.1.2. Antibiotic modifying/inactivating enzymes

Acetylation

Mycobacteria produce enzymes that inactivate antibiotics by chemical modification (209). Arylamine N-acetyltransferases (NATs) are enzymes that catalyse the transfer of the acetyl group from acetyl coenzyme A (AcCoA) to the free amino group of arylamines and hydrazines. The human N-Acetyl transferase (NAT2) is able to inactivate isoniazid by transferring an acetyl group to its free amino group forming an acetylamide (145). The *M. tuberculosis* Nat (*nat*, *Rv3566c*) enzyme has also been implicated in inactivating isoniazid since heterologous expression of *M. tuberculosis* Nat in *M. smegmatis* resulted in a three-fold increase in isoniazid resistance (145), while increased sensitivity to isoniazid was observed in a *M. smegmatis* Δ*nat* knockout mutant (146). It was also demonstrated that isoniazid is an excellent substrate for *M. smegmatis* Nat (168), and based on the high degree of sequence homology between the *M. smegmatis* Nat and *M. tuberculosis* Nat it was suggested that isoniazid might also be a good substrate for *M. tuberculosis* Nat. This is however controversial, since a recent study has demonstrated through kinetic characterization that isoniazid is a poor substrate for the *M. tuberculosis* Nat enzyme (180). Since these experiments were done *in vitro*, it can also be speculated that these findings might differ *in vivo*.

Some *M. tuberculosis* clinical isolates with wildtype *rpsL* and *rrs* genes exhibit a low-level streptomycin resistance phenotype (MIC 25- 50 µg/ml), which is still above clinical resistance of 2 µg/ml (58, 125, 133). This suggests that alternative mechanisms conferring streptomycin resistance may exist. In certain mycobacterial species (i.e. *M. fortuitum*, *M. chelonae*, *M. tuberculosis* and *M. smegmatis*) aminoglycoside resistance (in the absence of the known drug-resistance causing mutations) can be explained by the expression of acetyltransferases (4, 5, 126). Acetyltransferases transfer a functional group (phosphate, nucleotide or acetate) to the aminoglycoside structure, thereby preventing binding of the aminoglycoside to the ribosome (42, 90). Three types of transferases have been demonstrated: O-phosphotransferases, O-nucleotidyltransferases and N-acetyltransferases (111). The *aac(2')-Id* gene, encoding for an N-acetyltransferase in *M. fortuitum*, conferred low level resistance to aminoglycosides when introduced into *M. smegmatis* mc^2155 (5). Over-expression of the *aac(2')-Id* gene resulted in higher aminoglycoside MIC values, suggesting that the intrinsic level of aminoglycoside resistance is determined in part by the intracellular concentration of Aac(2')-ld (126). In *M. fortuitum,* the *aph(3")-lc* gene, encoding for a 3'-O-phosphotransferase, confers resistance to

streptomycin. No homologue of this gene is evident in other mycobacterium species (155). In *Pseudomonas aeruginosa* the *aph(3')-llb* gene (encoding for an aminoglycoside-phosphotransferase) is under the positive control of a surrogate regulator HpaA (231). Therefore the transcriptional regulation of *aph(3")-lc* needs to be investigated further in *M. fortuitum* to fully understand this resistance mechanism.

Ribosylation

Many fast growing mycobacteria, including *M. smegmatis, M chelonae, M. flavescens, M. vaccae* and *M. parafortuitum*, have been reported to be intrinsically resistant to rifampicin, in the absence of resistance-causing mutations in the *rpoB* gene (77). These mycobacteria inactivate rifampicin by ribosylation, whereby ADP-ribosyl transferase transfers an ADP-ribose unit to an amino acid residue on rifampicin (20). The enzyme responsible for this reaction is encoded by the gene *arr-ms.* Constitutive expression of *arr-ms* in *M. smegmatis* increases the rifampicin MIC to 32 µg/ml (54), which is significantly higher than the critical concentration (2 µg/ml) used to define rifampicin resistance in *M. tuberculosis* (208). Over-expression of recombinant *arr-ms* in *E.coli* resulted in increased MICs for different rifamycins to ≥ 512 µg/ml (rifabutin, rifampin, rifamycin, rifaximin) (20). This gene is absent from the *M. tuberculosis* genome thereby explaining the lower rifampicin MIC in this pathogen.

2.2.1.3. Efflux systems

The role of efflux pumps in intrinsic and acquired resistance is not fully understood and limited data exists for their role in intrinsic resistance to antibiotics currently included in the TB-regimen. It has been observed that inhibitors of efflux pumps mechanisms have limited effect on changing the rifampicin MIC in wildtype *M. smegmatis, M tuberculosis* and *M. aurum* isolates (148). This suggests that the genes encoding for these efflux pumps are expressed at a basal level to maintain the normal homeostasis within the cell (148). However, evidence suggests that the expression of efflux pump genes can also be induced to result in elevated levels of acquired drug resistance (113). This will be discussed later (section 2.2.3.7).

The Major Facilitator Superfamily (MFS) efflux pump, P55, has been shown to contribute to intrinsic resistance to antibiotics in *M. bovis* (25, 181). The gene, *Rv1410c*, encoding its homolog in *M. tuberculosis,* forms an operon with *R1411c* which encodes for a lipoprotein LprG (157). It has been shown that this operon plays a role in the extrusion of antibiotics. In addition, it is part of the oxidative stress response and is required to maintain normal growth characteristics (157). Over-expression of P55 in *M. smegmatis* resulted in an 8-fold increase in the MIC for streptomycin. The use of an energy uncoupler (Carbonyl cyanide m-chlorophenyl hydrazone) or efflux pump inhibitors (verapamil and reserpine) on these P55 expressing cells resulted in a decrease in the streptomycin

MIC. The deletion of P55 in *M. bovis* BCG resulted in decreased MICs for rifampicin (0.0032 µg/ml to 0.0004 µg/ml), ethambutol (4 µg/ml to 2 µg/ml), and vancomycin (4 µg/ml to 0.25 µg/ml) (157), indicating a role for P55 in intrinsic resistance to these drugs. Interestingly, the deletion of P55 in *M. smegmatis* did not affect the rifampicin MIC, however the MICs for isoniazid and ethambutol were decreased 2 fold (24). This suggests that intrinsic resistance to a drug is determined by a combination of mechanisms and is therefore closely linked to genetic background.

The Stp protein, encoded by *Rv2333c* in *M. tuberculosis*, is another pump responsible for intrinsic resistance in *M. bovis* BCG (encoded by *Mb2361c)*. The deletion of this pump in *M. bovis* BCG resulted in a 2 and 4 fold decrease in spectinomycin and tetracycline resistance respectively (156) and an increase in [^{3}H]tetracycline accumulation. Over-expression of *Rv2333c* in *M. bovis* BCG using a multicopy plasmid resulted in a 2 fold increase in spectinomycin and tetracycline resistance (156).

2.2.1.4. Genetic polymorphisms of target genes

Certain mycobacterial species such as *M. abscessus*, *M. leprae* and *M. chelonea* are naturally resistant to ethambutol due to a variant amino acid motif in the Ethambutol Region of Drug Resistance (ERDR) of the *embB* gene, resulting in an ethambutol MIC ≥ 64 µg/ml (8). Introduction of this variant in susceptible *M. smegmatis* (with an ethambutol MIC of 0.5 µg/ml) resulted in a 500-fold increase in the ethambutol MIC (256 µg/ml) (8). No other naturally occurring variants in target genes have been reported for mycobacteria.

2.2.2. Classical drug resistance mutations conferring varying levels of drug resistance

For most anti-TB antibiotics evidence suggests that classical drug resistance causing mutations confer varying levels of resistance for the respective antibiotic. These will be discussed below and summarized in Table 2.3.

2.2.2.1. Isoniaizid – *katG, inhA, inhA* promoter mutations

Isoniazid is one of the most important first line anti-TB drugs. On entry into the cell by passive diffusion, isoniazid is activated by the catalase-peroxidase enzyme encoded by *katG* to form isonicotinic acyl anions and other reactive oxygen species such as superoxide, peroxide and hydroxyl radicals (232). KatG is considered to be important to protect the bacteria against oxidative stress, specifically peroxidases produced by phagocytic cells. The metabolic products that are generated by the KatG-mediated isoniazid activation reacts with NAD (H) to form an isoniazid-NAD (P) adduct. This adduct binds to and inactivates the protein encoded by *inhA* which is an enoyl-acyl carrier protein (ACP) reductase, involved in mycolic acid synthesis (159). Inactivation of inhA therefore leads to the inhibition of cell wall synthesis and cell death. Certain mutations in the *katG* gene confer high level resistance to isoniazid (MIC > 5 µg/ml) and occur in 50-95% of all isoniazid resistant isolates (1, 75, 152). Alternatively, mutations at different positions in *inhA*, encoding for the drug

target of isoniazid, will affect the structure of InhA, resulting in isoniazid resistance (152). These structural mutations are rarely seen in clinical isolates and generally confer low level isoniazid resistance (MIC < 0.5 µg/ml) (154, 178, 222). The promoter regulating the expression of *inhA* has been identified to be positioned upstream from the *mabA*/*inhA* operon (17). Various mutations in this promoter region (-15 C to T, -17 G to T and -8 T to C) (15) lead to the over-expression of *inhA*, resulting in an increase in the concentration of InhA which partially overcomes the toxic metabolic by products generated by KatG (17, 104, 116). These *inhA* mutations are associated with low-level INH resistance (MIC 0.2 -1 µg/ml) (75).

2.2.2.2. Rifampicin – *rpoB* mutations

Since nucleotide substitutions in *rpoB* result in a decreased binding affinity of RNA polymerase for rifampicin, the level of rifampicin resistance is dependent on the position of the mutation in the *rpoB* gene (82). Although most *rpoB* mutations cause resistance above the critical concentration of 2 µg/ml, the MIC for rifampicin has been reported to range from 32 to 256 µg/ml on 7H10 solid media (Table 2.3) (82). In clinical isolates, mutations at codon Ser531 and His526 account for more than 75% of rifampicin resistance and these mutations are associated with high-level rifampicin resistance (MIC > 32 µg/ml). However, mutations found at positions 511, 516, 518 and 522 are associated with lower-level resistance (233). Some reports have described rifampicin MICs in liquid media of 1-2 µg/ml as "lower-level resistance", where the critical concentration is defined at 1 µg/ml (74, 216, 217) . It is suggested that a higher dose of treatment might still be feasible for isolates with MICs up to 1 µg/ml (216).

2.2.2.3. Pyrazinamide – *pncA* mutations

Pyrazinamide is an important first-line drug for the treatment of TB as it appears to be important for the killing of latent *M. tuberculosis* (81). Pyrazinamide needs to be hydrolised by pyrazinamidases to its active form, pyrazinoic acid (98, 177). Pyrazinamidases are expressed constitutively in pyrazinamide susceptible *M. tuberculosis* isolates. Mutations in the gene, *pncA,* encoding for pyrazinamidases resulted in decreased pyrazinamidases activity, thereby causing resistance to pyrazinamide. These mutations are scattered throughout the *pncA* gene, with some degree of clustering in regions that contain the active (D8, K96, A134 and C138) and metal-binding (D49, H51 and H71) sites of pyrazinamidases (51). Mutations in the metal-binding site (D49N, H51R) results in the lowest enzymatic activity and are associated with the highest pyrazinamide resistance (MIC of 400 µg/ml and > 800 µg/ml respectively), where mutations near the active site (D12A, D12G, F49L) showed pyrazinamide resistance levels of 400 µg/ml(177). Even though this study showed that resistance levels are negatively correlated with enzymatic activity, it was also demonstrated that only 27.3% of resistance variability is explained by enzymatic activity, suggesting that mechanisms other

than low pyrazinamidase activity should occur frequently and are associated with *pncA* mutations (177).

2.2.2.4. Ethambutol – *embB* mutations

Amino acid substitutions at *embB*306 are associated with ethambutol resistance, however the change in level of resistance is determined by the nature of the amino acid substitution at that position (188). For example, it has been shown that Met306Leu and Met306Val substitutions are associated with high level ethambutol resistance (40 µg/ml), while Met306Ile substitutions are associated with low level ethambutol resistance (20 µg/ml) (86, 188), where the critical concentration for ethambutol is defined at 5 µg/ml (67). Mutations at codon 306 in *in vitro* selected EMB resistant mutants conferred EMB resistance with a MIC of 8 µg/ml, which is significantly lower than what is found in clinical isolates. This suggests that additional mutations at currently unknown loci contribute towards the high level EMB resistance (162).

2.2.2.5. Streptomycin – *rpsL, rrs* mutations

Streptomycin inhibits protein synthesis by binding to the 30S subunit of bacterial ribosome, with the site of action at the ribosomal protein S12 and 16S rRNA (233). Thus mutations in *rpsL*, encoding for the S12 protein, and *rrs*, encoding for 16S rRNA, confers resistance to streptomycin (58). Mutations in the *rpsL* gene are associated with high level streptomycin resistance, with Lys43Arg mutations resulting in an MIC >1000 µg/ml, while Lys88Arg mutations result in MICs ranging from 250 to > 1000 µg/ml (125, 140). Mutations in the *rrs* gene are associated with intermediate levels of streptomycin resistance (MIC 50 – 500 µg/ml) (125).

2.2.2.6. Ofloxacin – *gyrA* mutations

Ofloxacin is a fluoroquinolone used as a second-line drug for the treatment of MDR-TB. In *M. tuberculosis* it targets and inactivates a type II DNA polymerase (DNA gyrase) thereby introducing negative supercoils in the circular DNA molecule. Mutations in the genes encoding for the DNA gyrase, *gyrA* and *gyrB*, are responsible for conferring resistance to fluoroquinolone (152). Asp94Gly substitutions in the *gyrA* gene are associated with high level ofloxacin resistance (MIC >20 µg/ml) (94, 202), whereas Ala90Val and Ser91Pro substitutions are associated with a lower level of ofloxacin resistance (MIC ≤ 6 µg/ml) (36, 183). It is speculated that the Asp94Gly mutation has the greatest influence on the binding constant between the DNA gyrase and the fluoroquinolone, resulting in a higher MIC. In contrast, isolates harbouring Asp94Gly substitutions displayed lower level moxifloxacin resistance (MIC 1-2 µg/ml), but still considered as resistant since the critical concentration of moxifloxacin is 0.25 µg/ml (119, 183). Since these susceptibility levels are below the peak serum concentration (4 µg/ml) it is suggested that moxifloxacin can still be used for the treatment of ofloxacin resistant isolates (183).

2.2.3. Non-classical mutations modulating the levels of drug resistance

There is cumulative evidence that other genes (other than the classical drug resistance conferring genes) have the ability of modulating the levels of resistance. These mechanisms will be discussed below and summarized in Table 2.3.

2.2.3.1. Isoniazid – *furA* mutations

As discussed above, isoniazid is a prodrug which is activated by KatG. Therefore, one can argue that the MIC for isoniazid is defined by the rate at which isoniazid is activated by KatG, rather than the interaction of its activated metabolic products with downstream cellular processes (150). KatG, catalase-peroxidase enzyme, is considered to be an important protective mechanism against oxidative stress, specifically peroxidases produced by phagocytic cells (106). Accordingly, the isoniazid MIC is determined largely by the rate of expression of *katG* and the functionality of the encoded KatG. In contrast to *E. coli*, expression of *katG* in members of the *M. tuberculosis* complex *(M. tuberculosis*, *M. bovis, M. africanum* and *M. microti)* is not regulated by the LysR transcriptional activator,*oxyR* (194, 195), since *oxyR* is deleted in these members. Thus, the regulation of *katG* is thought to occur through alternative regulatory proteins present in the proteome of these mycobacteria (47, 48, 179). Previous studies done in non-mycobacterial species have suggested that expression of oxidative stress genes is coupled to iron metabolism via the ferric uptake regulator, Fur (229). This prompted the investigation of the relationship between the *fur* orthologues in *M. tuberculosis* and *katG* expression. Whole genome sequencing of *M. tuberculosis* revealed that a *fur*-like gene, *furA,* is positioned 40 bp upstream of *katG* and that both genes are co-transcribed (48, 143, 150). Subsequently it has been shown that the *katG* locus is genetically linked to the *furA* gene in all mycobacteria (229). This suggested that FurA might act as a transcriptional regulator of a subset of oxidative stress response genes (48, 143). Further investigation of this region revealed that *katG* is transcribed from two promoters (127). One promoter controls transcripts produced during early infection and contains both the *furA* and *katG*, whereas the second transcript contains only *katG* and is produced during extended growth in macrophages. It is suggested that due to the presence of only one catalase in *M. tuberculosis*, the bacteria has adapted to express *katG* from two promoters depending on the environmental inputs and physiological demands (127). Deletion of *furA* results in an increase in *katG* expression and hypersensitivity to isoniazid (decreased MIC) (229), suggesting that FurA acts as a repressor of *katG* expression (150). This is consistent with the observation that FurA negatively regulates its own expression by binding to a region upstream of *furA* (163). To date, only two *furA* mutant clinical isolates have been reported (150). Since these strains also have mutations in the *katG* gene, the direct effect of these *furA* mutations on the level of isoniazid resistance could not be determined, although it is tempting to speculate that the *furA* mutations may modulate the level of isoniazid resistance by altering *katG* expression. Recently, mutations were identified in clinical *M. tuberculosis* isolates in the *furA-katG* intergenic region (-7 G to A, -10 A to C, -12 G to A) were

associated isoniazid resistance (13). It was shown that the introduction of these mutations (-7 G to A, -10 A to C) in isogenic strains resulted in decreased expression of *katG* and increased isoniazid resistance (MIC 0.2 µg/ml) (12).

A second regulatory mechanism for *katG* has been proposed in *M. tuberculosis* involving a LysR-type regulator, OxyS (49). It has been shown that OxyS is an oxidative stress response regulator in *M. tuberculosis* that binds to the immediate region upstream of *katG* in the terminal region of *furA* (109). Over expression of *oxyS* in *M. smegmatis* resulted in a two fold increase in *katG* expression, while no significant change was observed in the expression of *furA* (109). These results suggested that *katG* is negatively regulated by OxyS. Over expression or reduction in the expression of OxyS did however not change susceptibility to isoniazid (49).

2.2.3.2. Streptomycin – *gidB* mutations

Various studies have reported the association of mutations in *gidB* (a putative 16S rRNA) in *M. tuberculosis* clinical isolates and low level streptomycin resistance (226). However, in various reports mutations in *gidB* were present in both streptomycin resistant and susceptible isolates, questioning the association with streptomycin resistance (140). Recently, the deletion of this gene resulted in a 16 fold increase in the streptomycin MIC (226). Complementation of the Δ*gidB* mutant with a wild type copy of *gidB* resulted in the restoration of streptomycin susceptibility. Additionally, it was shown that clinical isolates harbouring a *gidB* mutation had on average a 4 fold higher streptomycin MIC than those without a *gidB* mutation. However, part of the population of *gidB* mutation-containing isolates had a MIC of less than or equal to the critical concentration of streptomycin, and were thus classified as streptomycin sensitive. Pharmacodynamic analysis and Monte Carlo simulation of the streptomycin resistant *gidB* mutation-containing isolates suggested that these isolates will respond poorly with streptomycin treatment, even though these isolates demonstrated only low level streptomycin resistance. Drug susceptibility testing in 7H9 liquid media and LJ resulted in discordant results. This raised the question whether the critical concentration is set too low for LJ testing, resulting in isolates being scored as sensitive, while they are in fact clinically resistant (226).

2.2.3.3. Ethambutol – *embR* mutations

The level of ethambutol resistance may be influenced by the level of expression of the *embCAB*-operon, encoding for the drug target. This operon is present in all mycobacterial species; however, in *M. avium* the *emb* gene cluster contains only the *embAB* genes with an additional gene, *embR*, immediately upstream of the *embAB* genes (21, 153). *EmbR* encodes for a putative transcriptional activator belonging to the OmpR-like family (9) and is a multi-domain protein that possesses a DNA binding winged helix-turn-helix domain, a bacterial transcription activation domain and a forkhead-associated (FHA) domain (176). In the *M. tuberculosis* and *M. smegmatis* genome, the *embR* homolog

is located 2 MB from the *embCAB* locus, and is hypothesized to modulate the level of *embCAB* expression and thereby arbinosyltransferase activity (39, 153).

The Ser/Thr protein kinase, PnkH is another protein involved in the transcriptional regulation of the *embCAB* operon (129). PnkH is a sensory protein which becomes autophosphorylated in response to external stimuli thereby increasing the protein kinase activity (129). In turn, PnkH phosphorylates the C-terminal FHA domain of EmbR. FHA domains are capable of mediating protein-protein interactions through P-Thr (phosphothreonine) recognition and often participate in STPK-dependant signal-transduction pathways (53). The phosphorylation of EmbR enhances the binding activity towards the promoter regions of *embCAB in M. tuberculosis* (129, 176). Since ethambutol resistance depends on the abundance of EmbB and EmbA in the cell, increased expression of *embCAB* through increased activation by the PnkH-EmbR pair is predicted to increase EMB resistance (21, 176). This is supported by the observation that expression of PknH but not *pknH* harbouring the K45M mutation (this mutation in the ATP binding site abrogates the kinase activity of PknH) in *M. smegmatis* increased expression of *embCAB* (176). The ethambutol MICs of the *M. smegmatis* PknH K45M mutant increased threefold (0.75 µg/ml) when compared with the *M. smegmatis* strains harbouring the wild type kinase (0.25µg/ml). Conversely, mutations in the FHA domain of EmbR have been associated with ethambutol resistance (153). Recently the *embR* gene of 44 resistant isolates were screened and mutations were observed in 16/44 (36%) isolates at 23 different codons (190). The effect of these mutations on the level of ethambutol resistance, however, was not determined.

Recently a protein, EmbR2 (130), identified in *M. tuberculosis* CDC1551 was shown to inhibit the PknH-dependant phophorylation of EmbR. *MT3428*, the gene encoding for EmbR2, was annotated as a putative transcriptional regulator and shown to be absent from *M. tuberculosis* strain H37Rv. The gene is also present in other mycobacterial genomes such as *M. bovis* and *M. bovis* BCG. EmbR2 and EmbR share 55% sequence identity over 381 residues. In contrast to *embR*, which is genetically linked to *pknH* (39, 130), *embR2* is found to be genetically distant to *pknH* and is not phophorylated by PknH. Increasing concentrations of EmbR2, inhibited the phosphorylation of EmbR in a dose-dependent manner (130).

2.2.3.4. Ethionamide – *ethA, ethR* mutations

Ethionamide is a pro-drug that is activated by EthA (a FAD-containing monooxygenase enzyme), to generate metabolic by-products which exert a toxic effect upon the synthesis of the mycolic acid constituents of the mycobacterial cell wall (19, 45). The expression of *ethA* is negatively regulated by its neighbouring gene, *ethR* (19, 45). EthR is a member of the TetR/CamR family of transcriptional regulators which binds cooperatively as a homo-octamer to the *ethA* operator, 5 to 16 nucleotides upstream from the *ethA* start codon (55). Thus, expression of *ethR* defines the level of expression of

ethA which in turn determines the level the intrinsic resistance to ethionamide (60). This is supported by the observation that over expression of *ethR* leads to repression of expression of *ethA* and to an increase in the MIC of ethionamide from 15 µg/ml to 250 µg/ml in *M. smegmatis* (19, 45). Conversely, it is suggested that conditions leading to the down-regulation of *ethR* expression or a mutation in the *ethR* gene increases ethionamide susceptibility (lowering the MIC) (19). The deletion of *ethR* in *M. bovis* BCG increased the sensitivity to ethionamide from an MIC of 2µg/ml to an MIC < 0.6 µg/ml (19). Similarly, the addition of benzylacetone, which binds to EthR thereby prevented its binding to the *ethA* operator, appeared to cause a potent bacteriostatic effect when co-administrated with ethionamide (60).

2.2.3.5. Amikacin – e*is* promoter mutations

It is known that certain mutations in the 16S rRNA gene, *rrs*, can cause high-level resistance to kanamycin with MICs exceeding 80µg/ml (121). The *rrs*1401A/G mutation can also confer cross resistance to other second-line drugs such as amikacin and capreomycin (121). However, up to 80% of kanamycin-resistant clinical isolates which display low-level resistance to kanamycin, do not have mutations in the *rrs* gene and do not exhibit cross-resistance (100, 204). Recently, a previously unidentified mechanism of kanamycin resistance was described which resulted as a consequence of point mutations in the promoter region of, *eis,* the enhanced intracellular gene (230). Eis shares sequence homology and secondary structural characteristics with the GCN5-related N-acetyltranferase (GNAT) superfamily of acetyltransferases, which includes the aminoglycoside N-acetyltransferase (166, 220). Expression analysis revealed that isolates harbouring these promoter mutations have 180 fold higher transcript levels than the wild type H37Rv laboratory strain, and introduction of this mutation in H37Rv increased its resistance to kanamycin to between 20-25 µg/ml. Eis acetylates kanamycin at a rate 3.3 fold higher than that of amikacin, which explains the observed absence of cross-resistance to kanamycin and amikacin in isolates harbouring *eis* mutations. Therefore, testing of suspect drug resistant cases can be important as isolates harbouring *eis* promoter mutations and resistant to kanamycin can still be treated with amikacin.

2.2.3.6. Ciprofloxacin – *lfrA, lfrR*

The *lfrA* gene in *M. smegmatis* encodes a transporter which confers low-level resistance to fluoroquinolone ciprofloxacin when over expressed (by cloning *lfrA* into a multicopy vector) (110, 206). Conversely, the disruption of *lfrA* resulted in a 2 fold decrease of the MICs for ciprofloxacin (110, 167). *lfrR*, an open reading frame (570 bp), upstream of the *lfrA* gene shares homology with several TetR transcriptional proteins (108). TetR transcriptional regulators are characterized by a conserved DNA binding domain (helix-turn-helix at the N-terminal) and a ligand binding domain (C-terminal region). Binding of an inducing ligand to the C-terminal region results in conformational changes in the N-terminal region, reducing its affinity to its target promoter DNA (78). Transcript analysis has revealed that the *lfrR* and *lfrA* genes are organized as an operon, with a promoter 220

base pairs upstream from *lfrR* (31). LfrR represses *lfrA* expression by binding directly to the promoter region of *lfrR-lfrA* (31). Various LfrA substrates (Acriflavine and Ethidiumbromide) were found to up-regulate the expression of the transporter by specifically promoting the dissociation of the repressor-operator complex (31). Addition of excess amount of acriflavine (2X MIC) abolished the binding of LfrR to the 43bp DNA region, suggesting that acriflavine acts as a ligand inducer for LfrR. Although ciprofloxacin acts as a substrate of *lfrA* it showed no interaction with LfrR, demonstrating that not all substrates of *lfrA* act as inducers. Since there are no known homologs of *lfrA* in *M. tuberculosis* it is speculated that the regulation of other efflux pumps may be involved in resistance to fluoroquinolones in *M. tuberculosis* (43, 44). A two to eight fold reduction in ofloxacin MICs were observed in ofloxacin resistant isolates (harbouring an *gyrA* mutation) when treated with efflux pump inhibitors cyanide m-chlorophenyl hydrazone (CCCP), 2,4-dinitrophenol (DNP), and verapamil (56, 182).

2.2.3.7. Unknown regulatory mutations leading to induced resistance

Prolonged exposure of fully susceptible and rifampicin-mono resistant clinical *M. tuberculosis* isolates to sub-inhibitory concentrations of isoniazid (0.1 µg/ml) resulted in increased isoniazid MICs (exceeding 64 µg/ml), which is considered high-level resistance (115). The increased resistance phenotype could also be reduced by means of efflux pump inhibitors (thioridazine, chlorpromazine and verapamil). Increased gene expression of efflux systems (*mmpl7, Rv1258c*, *P55, mmr* and *Rv2459*) correlated well with enhanced real-time efflux of a common efflux substrate, ethidium bromide. Further exposure of these isolates to isoniazid resulted in the selection of mutations and deletions in *katG* along with sustained increased efflux activity. It was therefore suggested that efflux systems play a role in the development of isoniazid resistance by allowing a subpopulation of the bacterial cells to survive during exposure to sub-inhibitory concentrations of isoniazid. During this period, mutations conferring high-level and stable isoniazid resistance can be selected for. Even though a significant increase in the expression of efflux genes were observed, their activity seemed promiscuous as an association between the extrusion of isoniazid and a specific gene could not be made (115). Inconsistent up-regulation of efflux pump and transporter genes following rifampicin exposure was also observed among closely related MDR *M. tuberculosis* isolates (114). In that study it was also shown that that the exposure of rifampicin mono-resistant and MDR isolates to 2 µg/ml rifampicin for 7 days induced a 10 fold increase in the ofloxacin MIC (114). Even though the regulatory mechanisms responsible for the antibiotic-induced efflux-mediated response described above are unknown, several other studies have shown the up regulation of efflux and transporters genes following antibiotic exposure, with accompanied increased resistance in some cases (18, 68, 144).

2.3. The use of whole genome sequencing in identifying novel drug resistance causing mechanisms

Several reviews have described expanding applications of next generation sequencing techniques in microbial research (10, 70, 80). Whole genome sequencing of the *M. tuberculosis* laboratory strain, H37Rv, was a revolutionary turning point for modern mycobacterial research and provided important insights into the biology of the species (39). With the development of next generation sequencing technologies, whole genome sequencing of additional *M. tuberculosis* laboratory and clinical isolates have allowed for large scale comparative and evolutionary studies (40, 124, 174).

In the pre-seqeuncing era, *M. tuberculosis* clinical strains have always been described as "monomorphic" with low sequence variation between strains (3). *M. tuberculosis* isolates from the same strain cluster (as characterized by IS*6110* RFLP) were considered as identical with little genomic diversity (218). Whole genome sequencing has however shown that there is high genomic diversity between seemingly closely related *M. tuberculosis* isolates. Comparative whole genome sequencing of two Beijing isolates with identical DNA fingerprints, identified 160 SNPs and one deletion differentiating the two isolates (142). Additionally, re-sequencing and comparative genomic analysis of the laboratory strains H37Rv and H37Ra identified 30 SNPs and 74 insertion-deletion events accumulated during low number of passages which possibly resulted in attenuation and gain of resistance to lysis (234). Re-sequencing of six H37Rv laboratory strains collected from multiple laboratories also revealed several evolutionary events occurring in the *in vitro* environment (85). These two studies demonstrated how the genomes of *M. tuberculosis* isolates continuously evolve, even in a controlled environment.

The mutation rate accompanied with *in vivo* acquired drug resistance within single patient sputum samples or within patient-to-patient transmission events, however, appears to be very low. Re-sequencing of drug sensitive isolates collected from a five-patient transmission chain revealed that only one SNP occurred over six years and three transmissions, while only four mutational events (2 SNPs a tandem repeat change and IS*6110* transposition change) occurred sequentially in a single patient (173). In a separate study, only two point mutations (*katG* S315T and *rpoB* D516Y) were identified in non-repetitive regions between serial sputum isolates collected from a single patient over a period of 12 months (171). These sputum isolates evolved from fully sensitive to MDR over a period of 12 months. The two points mutations identified are known to confer isoniazid and rifampicin resistance, and therefore the lack of any additional mutations suggested a low mutation rate per generation in the acquisition of MDR (171). In contrast, whole genome sequencing of MDR and XDR outbreak strains from HIV positive patients from KwaZulu Natal, South Africa, identified 22 SNPs which were unique to the XDR genome or shared by only the MDR and XDR genomes (134). These SNPs were not located in any of genes known to confer resistance. Of the 22 SNPs identified,

12 were found to be associated with a drug resistance phenotype or compensatory mechanism and found to be unique to the KwaZulu Natal clade. Investigation of the flanking regions of these SNPs in other *M. tuberculosis* strains revealed that only one SNP were monophyletic, suggesting that these SNPs were markers for strain-phylogeny. It is thought that a large degree of genetic diversity observed between lineages (or clades) is likely to have a functional meaning (76). This is also supported by the observation that closely related *Streptococcus* M3 strains with seemingly modest genetic changes have significant different transcriptomes (22).

While whole genome sequencing has proven to be a valuable tool for the identification of mutations conferring resistance for new anti-TB drugs when sequencing is done from *in vitro* selected mutants (14, 66, 205), most studies using whole genome sequencing as an investigative tool have highlighted certain limitations in the current technology (59). Firstly, polymorphisms in transposable elements, repetitive regions (such as the PE-PPE genes), gene duplications and genetic regions absent in the reference genome are difficult to detect by current next-generation sequencing technologies. Secondly, comparative genomic analysis has limited discriminating power to distinguish adaptive mutations from neutral or compensatory. Lastly, the whole genome sequencing studies mentioned above were done from a mixture of bacterial cells which might include sub-populations with varying genotypes and phenotypes. Given the low sensitivity and coverage depth of some of these studies' sequencing strategies, SNPs acquired in adapting sub-populations will be mistaken for sequencing errors. A recent study has investigated heterozygous mapping results by using statistical approaches to distinguish sequencing errors from real polymorphisms (201). This study was able to show the fixation of adaptive mutations in evolving *M. tuberculosis* sputum populations. Another option to identify heterogeneity would be to sequence from single colony forming units, or to increase the sequencing coverage to accurately detect low-frequency variants. Thus the success of future next generation experiments will depend on the careful planning of the study design to increase the identification of essential variants in adapting bacterial populations.

2.4. Concluding remarks

It was generally thought that specific mutations in target genes result in specific drug resistance levels. However, recent reports show that drug resistance is more complex than the concept that a single mutation in a target gene is the sole mechanism modulating resistance against a given compound. The multi-faceted aspect of the mechanisms that influence the intracellular drug concentration is illustrated by the concept of varying levels of drug resistance for a single mutation in one target gene. Even though whole genome sequencing has also proven to be a valuable tool for the identification of mutations conferring resistance for new anti-TB drugs, the identification of variants essential for the evolution of drug resistance is hindered by several limiting factors presented by the current genome sequencing technologies and analysis strategies. Additionally, the prediction of the

functional relevance of these identified variants is still a daunting task and very few studies have made an association between genetic variants and phenotype. Finally, this review highlights the importance of an anti-TB regimen which targets a broad range of cellular mechanisms. New drugs complementing the efficacy of existing drugs should be included in regimens to shorten and improve treatment outcomes. Additionally, the inclusion of existing drugs currently not used in the TB treatment regimen might improve treatment outcome.

Table 2.1: Summary of genes associated with drug resistance in *M. tuberculosis*.

	Drug		Drug target	Mutations	Enzyme	Reference
First line drugs	Isoniazid (INH)		InhA	*katG*	Catalase peroxidase	(153, 154)
				ahpC	Alkyl hydroperoxide	
				inhA	fatty acid enoyl acyl carrier protein reductase A	
				kasA	β-ketoacyl-ACP	
				Ndh	NADH dehydrogenease	
	Rifampicin (RIF)		β subunit RNA polymerase	*rpoB*	β subunit RNA polymerase	(35, 208)
	Pyrazinamide (PZA)		No specific target	*pncA*	Pyrazinomidase	(175)
	Ethambutol (EMB)		EmbB	*embCAB*	arabinosyl transferase	(211)
Second line drugs	**Amino glycosides**	Streptomycin (STR)	16S rRNA ribosomal subunits	*rpsL*	16S rRNA ribosomal subunits	(7, 58, 121, 122)
		Capreomycin (CAP)		*rrs*		
		Kanamycin (KAN)		*tlyA*		
		Amikacin (AMI)				
	Ethionamide (ETH)		InhA	*inhA*	fatty acid enoyl acyl carrier protein reductase A	(55, 131)
				ethA	flavin monooxygenase	
				ethR	Transcriptional regulator	
	Fluoroquinolones (FQ)		DNA gyrase	*gyrA*	DNA gyrase	(50)
				gyrB		

Table 2.2: Percentage of drug resistance clinical *M. tuberculosis* isolates with no classical mutations

Drug	Percentage of drug resistance clinical *M. tuberculosis* isolates with no classical mutation	Reference
Isoniazid	**9-30**	(32, 152, 154)
Rifampicin	**2-5**	(32, 151, 210)
Ethambutol	**21**	(32)
Pyrazinamide	**15**	(32, 193)
Ciprofloxacin/ofloxacin	**8**	(32)
Kanamycin	**17-20**	(32, 65)
Amikycin	**10-20**	(32, 65)
Capreomycin	**20-28**	(32, 65)

Table 2.3: Classical and non-classical mutations modulating the level of resistance

Drug		Locus	Mutation	MIC	Reference
Isoniazid	Classical	*katG*		High level > 5 µg/ml	(1, 75, 152)
		InhA		Low level < 0.5µg/ml	(154)
		InhA promoter	-15 C to T, -17 G to T, -8 T to C	Low level 0.2-1 µg/ml	(125)
	Non-Classical	*furA-katG* intergenic	-7 G to A, -10 A to C	Increased resistance (MIC 0.2 µg/ml)	(12, 143, 150, 229)
Rifampicin	Classical	*rpoB*	Ser531Leu	>64 µg/ml	(219)
				>512 µg/ml	(41)
				> 32* µg/ml	(120)
				≥ 32 - ≥ 256* µg/ml	(82)
				10-170 µg/ml	(114)
			His526Tyr	10-64 µg/ml	(189)
				>512µg/ml	(41)
				> 32µg/ml *	(120)
				≥32 - ≥256µg/ml *	(82)

			Leu511Pro	Low level	(233)
			Asp516Val	Low level	(233)
			Ser522Leu	8-16 µg/ml *	(82)
				>16 µg/ml *	(120)
Pyrazinamide	Classical	p*ncA*	Asp49Asn, His51Arg	> 400 µg/ml	(177)
			Asp12Ala, Asp12Gly, Phe49Leu	> 400 µg/ml	(177)
Ethambutol	Classical	*embB*	Met306Leu, Met306Val	High level 40 µg/ml	(67, 86)
			Met306Ile	Low level 20 µg/ml	(67, 86)
	Non-Classical	*embR* *pnkH* *embR2*	Unknown	Unknown	(9, 129, 130)
Streptomycin	Classical	*rpsL*	Lys43Arg	> 1000 µg/ml	(140, 153)
			Lys88Arg	250 - >1000 µg/ml	(153)
		rrs		50 -500 µg/ml	(125)
	Non-Classical	*gidB*		Low level ≤ 16 µg/ml	(226)
Amikacin	Classical	*rrs*		> 80 µg/ml	(121)
	Non-Classical	*eis*	Promoter mutations	Unknown	(166, 230)
Ofloxacin	Classical	g*yrA*	Asp94Gly	> 20 µg/ml	(94, 183, 202)
			Ala90Val and Ser91Pro	Low level < 6 µg/ml	(183)
Ciprofloxacin	Non-classical	*lfrA-lfrR*	Unknown	Unknown	(108, 110, 167, 206)
Ethionamide	Non-Classical	*EthR*	Unknown	Unknown	(19, 60)

Legend to Table 2.3

* In vitro selected mutants

CHAPTER 3

MATERIALS AND METHODS

3.1. Epidemiological relevance of compensatory mutations in *rpoA*/C

3.1.1. *M. tuberculosis* strain selection

Sputum specimens from Tuberculosis (TB) suspect cases were submitted to the National Health Laboratory Services (NHLS) in Green Point, South Africa for routine culture phenotypic drug susceptibility testing (DST) as described previously (197). Tuberculosis cases confirmed by the NHLS were subsequently subjected to genotypic testing by targeted gene sequencing in *rpoB*, *katG*, *inhA*, *inhA* promoter region, *embB, pncA*, gyrA and *rrs* (112, 197, 221), Spoligotyping (95, 198, 199) and IS*6110* restriction fragment length polymorphism (RFLP) DNA fingerprinting (215). All specimens were stored at the Department of Biomedical Sciences, Stellenbosch University to form a longitudinal culture bank (112, 197, 221). Well characterized *M. tuberculosis* isolates were selected from the longitudinal culture bank and included both sensitive (n=67) and drug-resistant (n=298) *M. tuberculosis* isolates collected between 2000-2010 from patients resident in the Western Cape, South Africa, (37, 197). This isolate panel represented a mixture of different drug resistant phenotypes including: isoniazid mono-resistant, rifampicin mono-resistant, MDR *sensu stricto* (excluding pre-XDR and XDR cases), pre-XDR and XDR. The genetic population structure of this sample set is summarized in Table 3.1. Isolates belonging to the Beijing family were further sub-characterised into "typical" and "atypical" Beijing isolates by PCR (196). Isolates belonging to the atypical Beijing strain family were previously termed the R86 genotype (84). Clustered isolates were defined as having identical RFLP patterns while non-clustered patterns were unique and not detected in any other isolate of the full collection.

Table 3.1: Population structure of clinical isolates selected in the Western Cape, South Africa

Family	Subfamily	DS	INH	MULTI	RIF	MDR	PRE-XDR	XDR	All
BEIJING	Typical	21	8	1	9	26	10	1	76
BEIJING	Atypical (R86)	5	1	0	1	12	14	31	64
BEIJING	Unknown	0	1	0	8	3	1	0	13
LCC	All	29	5	7	2	21	4	9	77
LAM	All	11	8	1	16	20	9	4	69
T	All	0	9	0	15	5	6	2	37
Unknown	All	1	3	0	3	11	4	0	22
F28	All	0	0	0	0	2	0	0	2
HAARLEM	All	0	0	0	0	0	0	2	2
CAS	All	0	0	0	0	0	1	0	1
EAI	All	0	0	0	1	0	0	0	1
F26	All	0	0	0	0	1	0	0	1
Total		67	35	9	55	101	49	49	365

Legend to Table 3.1:
DS: Drug susceptible; INH: Isoniazid mono-resistant; MULTI: Rifampicin or isoniazid mono-resistant + other resistance to at least one other drug (not MDR); RIF: Rifampicin mono-resistant; MDR: Multi drug resistant
XDR: Extensive drug resistant; LCC: Low Copy Clade lineage; LAM: Latin American Mediterranean lineage
CAS: Central Asian Strains; EAI: East Africa India lineage; F26: S family

3.1.2. Target gene sequencing

3.1.2.1. Preparation of cultures for the preparation of DNA templates for PCR amplification

A 500 µl aliquot of the BACTEC MGIT 960 culture of each isolate was boiled in a 1.5 ml Eppendorf tube (Merck, New Jersey, USA) for 20 minutes at 100°C to release crude DNA templates.

3.1.2.2. Primer design

Oligonucleotide primers (Table 3.2) were designed using Primer3 software (160) using the whole genome sequence of the *M. tuberculosis* H37Rv reference strain (http://genolist.pasteur.fr/TubercuList). Primers spanning the entire *rpoA* locus (Rv3457c) (-15 bp upstream, + 20bp downstream of Rv3457c), as well as the region encoding the RpoA-RpoC interaction site in *rpoC* (Rv0668, bp 735-1680) (40) were designed.

Table 3.2: Oligonucleotide primers used in this study for the detection of nucleotide substitutions in *rpoA* and *rpoC*

		Primer sequence	Product size (base pairs)	T_m (°C)*
***rpoA*1**	**Forward**	5' gcattccagtcgattccatc 3'	676	60.43
	Reverse	5' ccaagatcgccttctgatgt 3'		60.22
***rpoA*2**	**Forward**	5' ggacgtcgaaaggaagaaga 3'	639	59.41
	Reverse	5' gtctccacgtccaggatcag 3'		60.68
rpoC	**Forward**	5' cgaaaacctctaccgcgaac 3'	992	62.02
	Reverse	5' gcgacaggatgttgttggag 3'		61.67

Legend to Table 3.2:
* DNA melting temperature

3.1.2.3. Polymerase Chain Reaction (PCR)

PCR reactions were performed under the following thermocycling conditions: The PCR reactions were executed in 25 µl reaction volumes. The PCR master-mix consisted of 1x Q-solution, 1x reaction buffer, 2mM $MgCl_2$, 200µM of each deoxyribonucleotide triphospate (dNTP), 50 µM Forward primer, 50µM Reverse primer, 1.25 U Hotstar Taq polymerase, 1.25 µl of the crude DNA template, and H_2O to make up a final volume of 25 µl. The reaction mixtures were heated in the thermal cycler (GeneAmp PCR System 2400, Applied Biosystems, Foster City, CA, USA) at the following conditions: an initial denaturing step at 95ºC for 15 minutes, followed by 35 cycles of a denaturing step at 94ºC for 1 minute, an annealing step at T_m (62ºC) for 1 minute, an extension step at 72ºC for 2 minutes, and a final extension step at 72ºC for 15 minutes. All PCR experiments included a negative control to which no DNA template was added.

3.1.2.4. Gel electrophoresis

A 1.5% Agarose gel was prepared by dissolving 1.5 g Agarose in 100 ml 1x-TBE (pH 8.3) buffer by boiling. Once the mixture had cooled, 5 µl ethidiumbromide (10 mg/ml) was added. The agarose was cast and allowed to cool to room temperature. The PCR product (25 µl) was mixed with 5 µl of blue loading dye (0.25% Xylene Cyanol, 30% glycerol) and loaded onto the gel. A 100 bp DNA ladder (GeneRuler, Fermentas) was loaded into a well adjacent to the samples to determine the size of the amplified DNA samples. The gel was run at 150 V for approximately 1-2 hours in 1x TBE buffer and visualized under ultra violet light using the Kodak Digital Science Electrophoresis Documentation and Analysis System 120 (VilberLourmat, France).

3.1.2.5. Sanger Sequencing

PCR products were purified and sequenced with the ABI PRISM DNA Sequencer model 377, Perkin Elmer at the Central Analytic Facility of Stellenbosch University. DNA sequences were then aligned and compared to the relevant genomic sequence of *M. tuberculosis* H37Rv reference strain (http://genolist.pasteur.fr/TubercuList) using DNAman version 4.1 (LynnonBiosoft, Quebec, Canada) to determine if polymorphisms are present in the genes domains of interest.

3.1.3. Statistical Analysis

Statistical analysis was done in collaboration with Dr B Müller (Biomedical Sciences, Stellenbosch University). To test for an association between compensatory mutations and clustering, IS*6110* RFLP patterns of the 365 isolates included were compared with RFLP patterns obtained from our full sample bank consisting of 2003 RFLP typed isolates with 564 different RFLP types identified. Statistical analyses were performed in IC Stata 10.0 (StataCorp LP, College Station, Texas, USA). Chi-squared analyses, univariate and multivariate logistic regression analyses were done to identify associations of compensatory mutations in *rpoC* with clustering, level of drug resistance, strain genotype, and different rifampicin resistance mutations in *rpoB*. A p-value <0.05 was considered statistically significant.

3.2. Investigation of the expression of energy metabolism genes in response to rifampicin exposure

3.2.1. Strain selection

MDR (n=4) and rifampicin mono-resistant clinical isolates (n=2) from genetically distinct evolutionary lineages (Beijing and Low Copy Clade (LCC)) were selected. The characteristics of these isolates were previously described (114). These isolates were characterised by an identical *rpoB* mutation at codon 531 (Ser531Leu) and were selected based on their rifampicin MICs (Table 3.3).

Table 3.3: Characteristics of the MDR and rifampicin mono-resistant isolates selected.

Resistance profile	Isolate	Family	Cluster	*rpoB* mutation	Other resistance causing mutations	Rifampicin MIC (µg/ml)
Rifampicin mono-resistant	R721	Beijing	209	Ser531Leu	None	70
	R160	LCC	338			80
MDR-TB isolates	R257	Beijing	220	Ser531Leu	*inhA* -15 *embB*306 Met306Val	30
	R179				*pncA*103 Tyr103Ter	140
	R439*	LCC	DRF150	Ser531Leu	*katG*315 Ser315Thr *embB*306 Met306Ile *pncA* Del58	60
	R451*					170

Legend to Table 3.3:
* Isolates selected for protein analysis

3.2.2. Culture of *M. tuberculosis* isolates

Preparation of freezer stock

All *Mycobacterium tuberculosis* strains were initially grown on Löwenstein-Jensen (LJ) solid medium with continuous aeration for approximately three weeks at 37°C. After three weeks the buff coloured colonies that had a breadcrumb appearance were scraped from the LJ's. Primary cultures were set up by inoculation of the scraped colonies into 5 ml of 7H9 Middlebrook medium (Becton, Dickinson and Company, Sparks, USA) (supplemented with 10 % albumin-dextrose-catalase (ADC); 0.2 % (v/v) glycerol (Merck Laboratories, Saarchem, Gauteng, SA) and 0.1 % Tween80 (Becton, Dickinson and Company, Sparks, USA. These primary cultures were grown in filtered screw cap tissue culture flasks (Greiner Bio-one, Maybachstreet, Germany) without shaking at 37°C to mid-logarithmic phase (OD_{600}= 0.6-0.8). Subsequently, the primary cultures were inspected for contamination by Ziehl-Neelsen (ZN) gram staining and cultured on blood agar plates. *M. tuberculosis,* an acid fast bacterium, will retain dyes when heated and treated with acidified organic compounds, and will therefore appear pink in a contrasting background when the ZN test is done. *M. tuberculosis* does not grown on blood agar within two days. The primary cultures were then sub-cultured in 10 ml 7H9 Middlebrook medium supplemented with ADC and incubated at 37°C. After approximately 2 weeks of growth (OD_{600} = 0.7-0.8), the secondary sub-cultures were re-inspected for contamination and a 25 % glycerol stock (1:1 v/v, 500 µl culture and 500 µl 50 % glycerol) was prepared and stored at -80°C.

3.2.3. Protein Analysis

As described previously (M. Bester, MSc thesis, 2009), a proteomic approach was followed to identify differential protein expression in clinical MDR *M. tuberculosis* isolates after rifampicin exposure. These clinical isolates (R451 and R439, see Table 3.3) displayed varying levels of rifampicin resistance. Briefly, two-dimensional electrophoresis and Matrix Assisted Lazer Desorption/Ionization Time of Flight (MALDI-TOF) analysis were used to identify differential protein expression after 24 hours of rifampicin (2 µg/ml) exposure. Proteins showing increased level of abundance after rifampicin exposure and which are involved in energy metabolism are shown in Table 3.4.

Table 3.4: Proteins with increased expression involved in energy metabolism

Accession nr	Gene	Protein	Protein Function *
Rv3029c	*fixA*	Electron transfer flavoprotein beta- subunit	Transfers electrons to main respiratory chain via ETF-ubiquinone oxidoreductase (ETF dehydrogenase)
Rv1308	*atpA*	ATP synthase alpha chain	Alpha chain (Regulatory unit) of F_OF_1 ATP synthase, which produces ATP from ADP in presence of proton gradient across membrane.
R1446c	*opcA*	Putative oxpp cycle protein	May be involved in functional assembly of glucose 6-phosphate dehydrogenase
Rv0462	*lpdC*	Dihydrolipoamide dehydrogenase lpd	Component of the alpha-ketoacid dehydrogenase complex
Rv0363c	*fba*	Fructose bisphosphate aldolase	Involved in glycolysis (catalytic activity)
Rv0896	*gltA2*	Citrate synthase GltA2	Involved in tricarboxylic acid cycle (Krebs cycle, catalytic activity)
RV1307	*atpH*	ATP synthase delta chain ATPH	F_OF_1 ATP synthase subunit delta. Either transmits conformational changes from CF(0) into CF(1) or is implicated in proton conduction

Legend to Table 3.4:
* Protein function according to http://genolist.pasteur.fr/TubercuList

3.2.4. RNA extraction and gene expression analysis

RNA extraction and gene expression analysis was done in collaboration with Dr GE Louw (Biomedical Sciences, Stellenbosch University).

3.2.4.1. RNA extraction from *M. tuberculosis* cultures

Each experimental culture was set up by inoculating 800 µl of the secondary sub-culture into 80ml enriched 7H9 medium (1:100 dilution) and incubating at 37°C until mid-log phase (OD_{600}= 0.7-0.8). The 80 ml culture was then divided into 2 x 40 ml cultures. Rifampicin was added to one 40 ml culture to a final of concentration of 2 µg/ml (internationally defined critical concentration). No anti-

TB drugs were added to the other 40 ml culture (control). Both cultures were then incubated at 37 °C for a further 24 hours.

After antibiotic exposure, RNA was extracted as described previously (114). Briefly, five volumes of 5M GITC (Sigma-Aldrich, St Louis, Germany) were added to one volume of each *M. tuberculosis* culture. The cultures were centrifuged (Eppendorf, Centrifuge 5810R) in 50 ml tubes (LASEC, South Africa) at 3 000 rpm for 20 minutes at 20°C. The supernatant was discarded and the bacterial pellet was transferred to a 2 ml Eppindorf tube (Merck, New Jersey, USA) and centrifuged at 13 000 rpm for 1 minute at room temperature. After centrifugation, the pellet was re-suspended in 1 ml TRIzol LS reagent (Invitrogen, CA, USA). The suspension was transferred to a 2 ml screw capped tube containing silica beads (IEPSA, Medical diagnostics, South Africa) and ribolysed using a FastPrep FP120 ribolyser (Bio101 SAVANT, Vista, USA) at 6 m/s for 4 x 20 seconds with 1 minute of cooling on ice between intervals. The tube was incubated for 5 minutes at room temperature before the lysed suspension was centrifuged at 12 000 rpm for 10 minutes. The supernatant and cellular debris was transferred to a 2 ml Phase Lock gel tube (Merck New Jersey, USA) containing 300 µl chloroform (Sigma-Aldrich, St Louis, Germany). The tube was inverted rapidly for 15 seconds, followed with periodically inversion for an additional 2 minutes. The tube was incubated for 5 minutes at room temperature and then centrifuged at 12 000 rpm for 10 minutes. The top aqueous layer was then transferred to a clean 1.5 ml tube containing an equal volume of isopropanol and incubated overnight at 4ºC. Precipitated nucleic acids were collected by centrifugation at 12 000 g for 30 minutes at 4ºC. After centrifugation, the RNA pellet was washed with 1 ml 70 % ethanol, inverted several times and centrifuged at 12 000 rpm for 10 minutes at 4°C. After centrifugation, the ethanol was aspirated and the RNA pellet was allowed to air-dry at room temperature. The RNA pellet was dissolved in 70 µl RNase-free water (Ambion, Applied Biosystems, CA, USA) and stored at -80°C for subsequent experiments.

3.2.4.2. DNAse treatment of extracted RNA

Contaminating chromosomal DNA was digested with RQ1 RNase-free DNase (Promega, WI, USA) according the instructions of the manufacturers. 15 µl of RNA was added to 4 µl DNase and 4 µl DNase Buffer, followed by incubation for 30 minutes at 37°C. After incubation the DNase treated RNA was stored at -80°C for further use.

3.2.4.3. Phenol purification

The DNase treated sample was made up to a final volume of 200 µl with RNAse free water. An equal volume of phenol:chloroform (4:1) (Sigma-Aldrich, St Louis, Germany) was added to the diluted RNA and mixed by gentle inversion. The RNA/phenol:chloroform mix was incubated for 10 minutes on ice. After incubation, the mixture was centrifuged for 10 minutes at 12 000 rpm at room

temperature. The top layer was transferred to a clean tube and 1/10 volume of sodium acetate (pH 5.2) (Merck, New Jersey, USA) and 2.5 volumes of 100 % ethanol (Merck, New Jersey, USA) were added to the RNA. The RNA was incubated for 1 hour (or overnight) at 4°C. The precipitated nucleic acids were collected by centrifugation at 12000 rpm for 30 minutes at 4°C. After centrifugation, the pellet was washed with 1 ml 70% ethanol, inverted several times and centrifuged at 12 000 rpm for 10 minutes at 4°C. Thereafter, the ethanol was aspirated and the purified RNA pellet was allowed to air-dry at room temperature. The RNA pellet was dissolved in 70 µl of RNase-free water and stored at -80°C till further use. The RNA quality and quantity were assessed by using the Experion analyzer with the Experion Software version 2.01 (Bio-Rad Laboratories, Hercules, CA 94547) and PCR amplification using the extracted purified RNA as the template in combination with the primers listed in Table 3.5.

Table 3.5: Primers used to assess the quality of the extracted RNA

Gene name		Primer sequence	Product size (base pairs)	T_m (°C)*
Rv2629c	Forward	5′ TAGGATCCATGCGATCAGAACGTCTCCG 3′	695	62
	Reverse	5′ GATAAGCTTCTAGGATCTATGGCTGCCGAGT 3′		62

Legend to Table 3.5:
* DNA melting temperature

3.2.4.4. cDNA synthesis of purified RNA

cDNA was synthesized from 1 µg highly purified RNA using the QuantiTect Reverse Transcriptase kit according to the manufacturer's instructions (Southern Cross Biotechnologies). Briefly, 1 µg RNA was mixed with 2 µl genomic DNA Whipeout Buffer and RNase-free water (to a volume of 14 µl) and incubated for 2 minutes at 42°C. After incubation Quantiscript Reverse Transcriptase, Quantiscript RT buffer, and RT Primer Mix was added to the RNA mix and incubated for 15 minutes at 42°C. After incubation the mix was incubated for 3 minutes at 95ºC to inactivate the Quantiscript Reverse Transcriptase.

3.2.4.5. Primer design of Quantitative Real Time PCR (QRT-PCR) of candidate genes

PCR primers (Table 3.6) for quantitative reverse transcriptase analysis of the candidate genes in Table 3.4 in addition to genes (*Rv1312, Rv1303*) flanking the ATP synthase operon, its regulator *Rv1846c* and the housekeeping gene 16s rRNA were designed against the whole genome sequence of the *M. tuberculosis* H37Rv reference strain (http://genolist.pasteur.fr/TubercuList) using Primer software 3 version 0.2 (Whitehead Scientific, South Africa).

3.2.4.6. Quantitative Real Time PCR

A standard Master mix was set up by adding 2µl of Light Cycler Fast start DNA Master PLUS SYBR Green I reaction mix (Roche Applied Science, Germany), 2 µl forward and reverse primer (10µM) of each candidate gene (Table 3.6), 11µl of RNase-free water. Subsequently, 9 µl aliquotes of the master mix was added to 1 µl cDNA. A no template control was added to assess contamination. RT-qPCR was done by using the Lightcycler 2.0 instrument (Roche Applied Science). A four step PCR parameter protocol was used: (i) Activation program (95°C for 15 minutes); (ii) RT PCR program repeated for 40 cycles (95°C for 15s for denaturation, 56/58^{o}C for 30s for annealing and 72°C for 30s for polymerase elongation); (iii) Melting curve program (95°C for 0s, 60°C for 15s, 90°C for 0s with a heating rate of 0.1°C/s) and (iv) Cooling down program of 40°C for 10s. Each RT-qPCR experiment was done on duplicate biological samples that were each assayed in triplicate.

Table 3.6: Primers used for the gene expression assessment of the energy metabolism genes

Gene		Primer sequence (5' to 3')	Gene length	Product size (bp)	Tm (^{o}C)*
fixA	**Forward**	5' ACGACCATCGTGGTCCTGAT 3'	801	146	58
	Reverse	5'ATCTGTAGCGCTTCCTCCAC 3'			
atpA	**Forward**	5' GAGGAGCAAGTGGTTTCGAT 3'	1650	141	58
	Reverse	5'GTCCCGGATCTCAGTCAAAA 3'			
opcA	**Forward**	5' CGGTCAACAAGAAGCTCGAC 3'	912	115	56
	Reverse	5' TCGATGGACTCTTCCAGCAT 3'			
lpdC	**Forward**	5' GCGAGGTGACCTTCGACTAC 3'	1395	112	58
	Reverse	5' TGGATCTCGGTGATCTTGTTC 3'			
Fba	**Forward**	5' GGGGTCAAAGACATGGTGAC 3'	1035	192	60
	Reverse	5' CGACTGGAACAAAGGATTGC 3'			
gltA2	**Forward**	5' GGTCAAATTGATGGGTTTCG 3'	1296	176	58
	Reverse	5' TGCGTTCGATGAAGTAGTCG 3'			
atpH	**Forward**	5' GAGGCGAGTCAAGCTCACAC 3'	1341	113	60
	Reverse	5' TGGGCCTCTAGTTGTTCTGC 3'			
Rv1312	**Forward**	5' GCGTTCTACCGGCTTTCTAGT 3'	444	141	56
	Reverse	5' GCACAGTTCCACAACGACAA 3'			
Rv1303	**Forward**	5' CTGACGGTCGGGATGTTCTT 3'	486	108	56
	Reverse	5' TTTTAACGGGTGCTCTTTGG 3'			
Rv1846c	**Forward**	5' GGTCATGGACCATTTGTGG 3'	417	140	58
	Reverse	5' GCAGCACGAGGTTCTTCTTT 3'			

Legend to Table 3.6:
* DNA melting temperature

3.2.4.7. Statistical analysis

Both 16S rRNA and *sigA* were included as reference genes, but 16S rRNA was used for normalization of mRNA levels, because in the experiments 16S rRNA expression levels were the most stable. The level of gene expression of each individual gene was quantified by the delta-delta Ct calculation in which the relative abundance of the target gene was normalized relative to the levels of the reference RNA transcripts (16S rRNA). Data analyses were done according to the delta-delta C_T equation R=2 -(LCT sample -LCT control). Only experiments with a standard deviation of <0.5 were included for analysis. Significant fold changes were identified based on The Relative Expression Software Tool - 384 (REST-384©) that assigns significance with a significance level of 5% (147).

3.3. Selection and characterization of rifampicin resistant *in vitro* mutants

3.3.1. Strain selection

To investigate the genotype and phenotype of *in vitro* generated rifampicin resistant mutants in a clinical strain background, a pan susceptible clinical isolate from the Beijing family (cluster 208 (K636)) and the laboratory strain H37Rv (ATCC 27294) were selected as progenitor strains for the Luria Delbrück Fluctuation assay. Freezer stocks of these isolates were prepared as described in section 3.2.2.

3.3.2. Selection of rifampicin resistant mutants by the Luria Delbrück Fluctuation assay

3.3.2.1. Growth curves

A 1:100 dilution of the freezer stock was inoculated in 7H9 Middlebrook media supplemented with ADC. The culture was incubated at 37°C for 28 days and the growth of the culture was assessed with daily OD_{600} readings. Additionally, 1 ml of the culture was serially diluted by factors of 10 at different time points (day 1, 2, 3, 7, 14, 21, 28 after initial inoculation) and plated on 7H10 Middlebrook (Becton, Dickinson and Company, Sparks, USA) solid media supplemented with OADC (Becton, Dickinson and Company, Sparks, USA). Plates were incubated at 37 °C for 21 days after which the Colony forming Units (CFUs) on each plate were determined.

3.3.2.2. Selection of *in vitro* rifampicin resistant mutants by the Luria Delbrück Fluctuation assay

In vitro mutants were selected according to a modified Luria Delbrück fluctuation essay (132). Briefly, a 1:100 dilution of the freezer stock was cultured in 7H9 Middlebrook media supplemented with ADC and incubated at 37°C to an OD of 0.8 (corresponding to approximately 10^8 CFU/ml). The culture was diluted in 500 ml of 7H9 Middlebrook media supplemented with ADC (to yield approximately 10^3 CFU/ml) and 5 ml aliquots were dispensed in 100 tissue culture flasks and incubated for 32 days at 37°C. The entire culture volume were transferred to 50 ml centrifuge tubes and centrifuged for 5 minutes at 3000 rpm. The supernatant were aspirated and the bacterial pellet was

suspended in 1 ml of 0.05% Tween. The tube was centrifuged for 5 minutes at 3000 rpm and the supernatant were aspirated, leaving a small amount of liquid. The bacterial pellet was re-suspended and glycerol stocks (25%, 1:1 v/v, 500µl culture and 500µl 50% glycerol) was prepared from half of the volume of the progenitor culture before rifampicin selection. The other half volume was plated on 7H10 solid Middlebrook media supplemented with OADC and contained 2µg/ml rifampicin. The plates were incubated at 37°C for 28 days. Single colonies were picked from these plates and sub-cultured in 5 ml 7H9 Middlebrook media supplemented with ADC until a growth OD_{600} of 0.8 was reached. Cultures were inspected for contamination and a 25% glycerol stock (1:1 v/v, 500µl culture and 500 µl 50% glycerol) was prepared and stored at -80°C.

3.3.3. Targeted gene sequencing

To determine the frequency of mutations in *rpoB*, the Rifampicin Resistance Determining Region (RRDR) was PCR amplified using DNA templates prepared from the glycerol stocks as described in section 3.1.2.1. The resulting amplicons were purified and sequenced using the Sanger method at Central Analytic Facility, Stellenbosch University.

3.3.3.1. Primer design

Primers (Table 3.7) for the amplification of the RRDR in the *rpoB* genes were designed using Primer software 3 version 0.2 (Whitehead Scientific, South Africa), and the whole genome sequence of the *M. tuberculosis* H37Rv reference strain (http://genolist.pasteur.fr/TubercuList).

Table 3.7: Primers used for the amplification of the RRDR

		Primer Sequence	Product size (bp)	Tm (°C)*
rpoB	**Forward**	5' TGGTCCGCTTGCACGAGGGTCAGA 3'	437	78
	Reverse	5' CTCAGGGGTTTCGATCGGGCACAT 3'		76

Legend to Table 3.7:

* DNA melting temperature

3.3.3.2. PCR (Touch down)

The PCR reactions were executed in 25 µl reaction volumes. The PCR master-mix consisted of 1X reaction buffer, 2mM $MgCl_2$, 200 µM of each deoxyribonucleotide triphospate (dNTP), 50 µM Forward primer, 50 µM Reverse primer, 1.25 U Hotstar Taq polymerase, 1.25 µl of the crude DNA template, and H_2O to make up a final volume of 25 µl. The reaction mixtures were heated in the thermal cycler (GeneAmp PCR System 2400, Applied Biosystems, Foster City, CA, USA) at the following conditions: an initial denaturing step at 95ºC for 15 minutes, followed by 2 cycles of a denaturing step of 94ºC for 1 minute and an annealing step at T_m (72ºC) for 1 minute. This is followed by another 2 cycles of a denaturing step at 94ºC for 1 minute, an annealing step at T_m (71ºC) for 1

minute. This is followed by another 2 cycles of a denaturing step at 94ºC for 1 minute, an annealing step at T_m (70ºC) for 1 minute. This is followed by 40 cycles of a denaturing step at 94ºC for 1 minute, an annealing step at T_m (62ºC) for 1 minute, an extension step at 72ºC for 1 minutes, and a final extension step at 72ºC for 10 minutes.

Subsequent gel electrophoresis and Sanger sequencing were done as described in section 3.1.2.4 and 3.1.2.5.

3.3.4. Rifampicin MIC determination

The Rifampicin MIC for the different *in vitro* mutants (generated by the Luria Delbruck Fluctuation assay) was determined in the BACTEC MGIT 960 instrument (BD Bioscience, MD, USA) according to the manufacturer instructions. Briefly, to initiate a starter culture, 800 µl of OADC were inoculated with 200 µl of the *M. tuberculosis* isolate freezer stock in the MGIT 960 media. The culture was incubated in the MGIT instrument at 37°C until a positive growth reading was reached. Two days after positive growth was reached, a drug free growth control tube was prepared. This was done by diluting the positive culture suspension (1:100) using sterile saline and inoculating 500 µl of this dilution into a MGIT tube supplemented with 800 µl OADC. Subsequently, 500 µl of the undiluted positive culture suspension was inoculated in MGIT tubes supplemented with 800µl OADC and 100 µl of the specific rifampicin concentration (with final concentration ranging between 10 and 200 µg/ml). The MGIT culture tubes were placed in the BACTEC MGIT 960 instrument and growth was continuously monitored using EpiCenter (version 5.75A) TBeXist software (BD Bioscience, Erembodegem, Belgium) for approximately 14 days.

The EpiCenter, version 5.75A, equipped with the TB exist module, were used for the analysis of the data generated by the MGIT 960 instrument. Results were interpreted as follows. At the time when the growth control tube reached a growth unit (GU) > 400, a strain would be considered as resistant if the GU of drug containing tube was ≥ 100. If the GU of the drug containing tube was < 100, the tube was incubated for a further 7 days. If the GU was still <100 after 7 days, the strain would be considered as sensitive. If the GU reached > 100 within the 7 days (after the growth control reached a GU of >400), the strain would be considered as intermediate resistant to the respective drug concentration (186).

The MIC is defined the minimum concentration of an antibiotic that is required to kill 99% of the bacterial cells in that culture (27). However, for this study we defined the MIC as the level of rifampicin resistance, which is reflected as the highest concentration of rifampicin where growth was still detected.

3.3.5. Statistical analysis

A two sample proportional test (confidence level of 95%) using the R statistical package (http://www.r-project.org) were used to test whether there was a statistical difference in the frequencies of the different *rpoB* mutations between the different genetic backgrounds and findings from previous studies (82).

3.4. Identification of genetic variants responsible for varying rifampicin resistance levels

3.4.1. Strain selection

To identify genetic variants resulting in increased rifampicin resistance, two *in vitro* rifampicin resistant mutants and five clinical MDR isolates with varying rifampicin resistance were selected for whole genome sequencing on the Illumina platform (Table 3.8).

Table 3.8: *M. tuberculosis* isolates selected for whole genome sequencing

	Isolate	Evolutionary lineage	Spoligotype	IS*6110* cluster type	*rpoB* mutation	Other classical resistance causing mutations	Rifampicin MIC (µg/ml)
***In vitro* mutants**	N37S1	Beijing	2	208	S531L	None	>100
	N36S1	Beijing			S531L		60
Clinical Isolates	K636	Beijing	2	208	Wild type	None	< 2
	R179	Beijing	2	220	S531L	*inhA* -15 prom *emb*B306 Met306Val *pncA* 103 Tyr103Ter	140
	R257	Beijing					30
	R451	Low Copy Clade	115	DRF150	S531L	katG315 Ser315Thr embB306 Met306Ile pncA Del58	170
	R439	Low Copy Clade					60

3.4.2. DNA extraction

DNA was extracted as previously described (223). In summary, 100 ul of freezer stock from each isolate were inoculated onto two petri dishes containing 7H10 media supplemented with OADC and incubated at 37°C for two weeks or until sufficient growth was observed. Thereafter, the outer surface of the petri dishes were decontaminated with Incidin Plus (Ecolab, Minnesota, USA) and placed in a biosafety autoclave bag and transferred to a prewarmed fan oven at 80°C for 1h to ensure heat killing. Colonies were gently scraped with a disposable 10 µl loop from the 7H10 media surface and transferred to a 50 ml tube containing approximately 20 glass beads (4mm in diameter) and 6 millilitres of extraction buffer (5% sodium glutamate, 50 mM Tris-HCl (pH 7.4) and 25mM EDTA)). The 50 ml tubes containing the extraction buffer cell suspension were vortexed to disrupt the colonies. Lysozyme (25 mg, Roche, Germany) and RNAse A (50 µg, Roche, Germany) were then added and incubated after gentle mixing for 2h at 37°C to degrade the cell wall and to digest RNA. Thereafter,

600 µl of 10x proteinase K buffer (5% sodium dodecyl sulphate, 100nM Tris-HCl (pH 7.8), 50mM EDTA) were added and the suspension was incubated for a further 16 hours at 45°C to digest all bacterial proteins. An equal volume of phenol/chloroform/isoamyl alcohol (24:23:1) was added and the mixed intermittently over a period of 2 hours at room temperature. After centrifugation at 3000 x g for 20 minutes, the aqueous phase was aspirated and an equal volume of chloroform/isoamyl alcohol (24:1) was added. After centrifugation at 3000 rpm for 20 minutes, the resultant DNA was precipitated with the addition of 600 µl 3 M sodium acetate (pH 5.2) and an equal volume of isopropanol and immediately collected on a glass rod. The DNA was washed with 70% ethanol, the air-dried pellet, re-dissolved in 300 µl TE (10 mM Tris-HCl (pH 8.0), 1mM EDTA) and stored at -20°C. The concentration of the extracted DNA was measured using the NanoDrop spectrophotometers (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

3.4.3. Illumina sequencing

Sequencing of the *M. tuberculosis* clinical isolates (n=5) and *in vitro* selected mutants (n=2) were done by Partners Healthcare Center for Personalized Genetic Medicine, Harvard Medical School, Boston, USA using the Illumina sequencing platform version 1.5 to generate pair-end reads with a size of 100 base pairs. DNA libraries were prepared by Dr. Gail Louw (Department of Biomedical Sciences).

3.4.4. Computational analysis of whole genome sequences

Computational analysis was done in collaboration with Prof G Sherlock (Stanford University, USA), Dr D Kvitek (Stanford University, USA) and Mmakamohelo Direko (SANBI, UWC). The workflow for the analysis of the whole genome sequences produced on the Illumina platform is summarized in Figure 3.1.

3.4.4.1. Data sources

The fastq files of the genomes sequenced were acquired from the Gateway for Integration of Genomic-Proteomic Applications and Data (GIGPAD) platform (http://pcpgm.partners.org/it-solutions/gigpad). The reference genome (*M. tuberculosis* H37Rv) in fasta format, genome summary information and genome sequences per gene was downloaded from http://genome.tbdb.org/annotation/genome/tbdb/MultiDownloads.html.

3.4.4.2. Fastq file format

The fastq format (Figure 3.2) is a text-based format for the storing of biological sequences and the corresponding quality information.

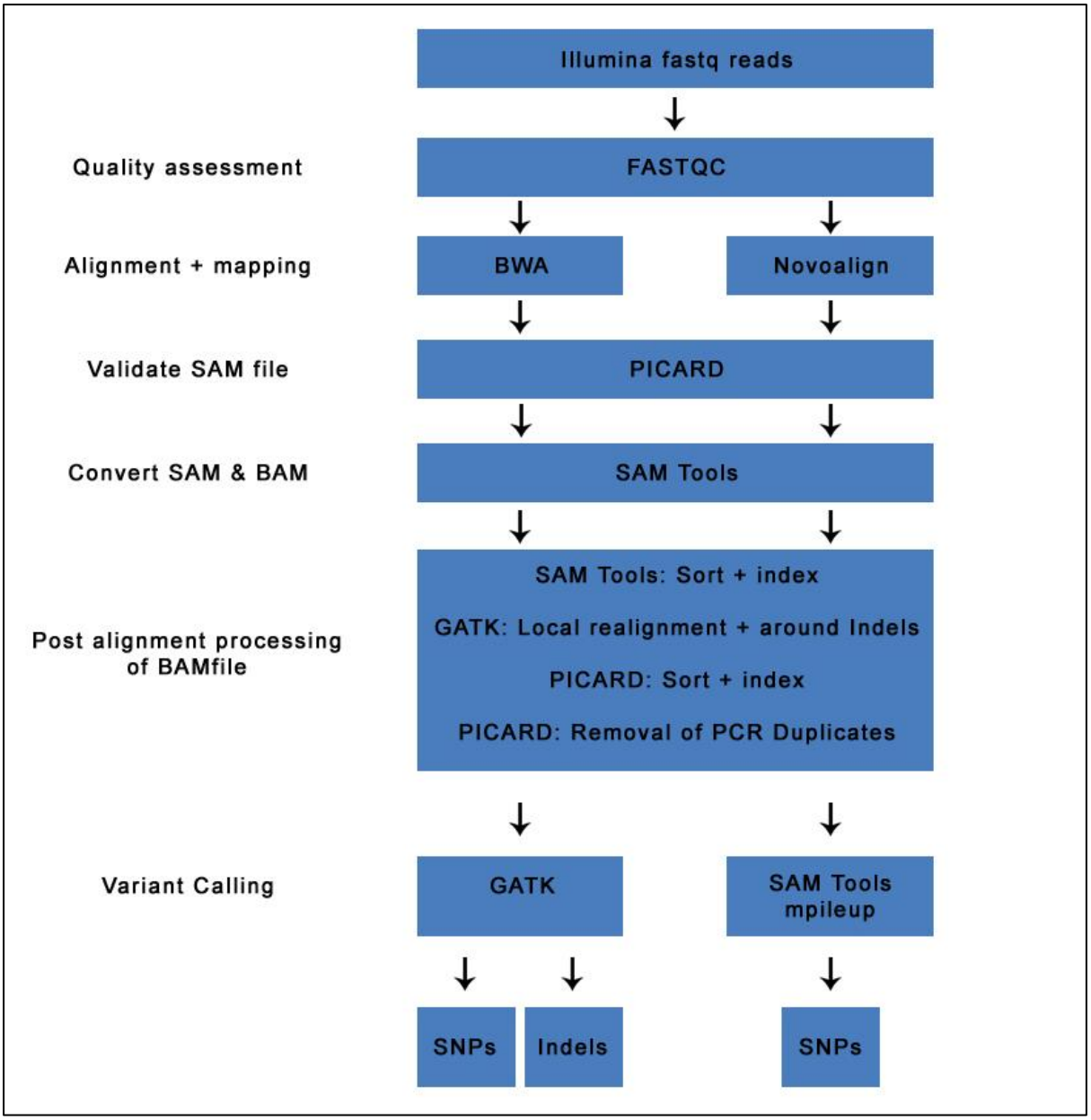

Figure 3.1: Summarized workflow for the computational analysis of whole genome sequences

Legend to Figure 3.1:

Computational tools are highlighted in blue

```
1 @ILLUMINA-C90280_0037_FC:1:1:12188:986#CGATGT/1
2 NGATCCTGGGCAGCTACCGATTCAGCGCCTTCCGCAGCGACAAGACCGCGCCCAAAGACGCCGGACTCCGCAAAATCACCGTGCTCTGCTGTGCAAAGGA
3 +
4 BGKKGJJJJH_____PVWVV____________________TVVTTV_________BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB
```

Figure 3.2: Representative of a typical read from a fastq file produced by the Illumina sequencing platform version 1.5.

Legend for Figure 3.2:

1. Begins with a '@' character, which is followed by the sequence identifier
2. The raw sequence letters
3. Begins with a '+' character and is optionally followed by the same sequence identifier
4. Encodes for the quality values of the sequence in line 2.

3.4.4.3. Quality assessment and quality control of the fastq reads

The raw sequence reads were assessed with the FastQC software to validate the quality of the reads and to determine whether the data included any factors that would influence the subsequent analysis. This java-based program takes fastq files as input and the results obtained are produced in a HTML format. FastQC analysed the raw data through a 7 steps module which include the following:

1. The general statistics included a description of the platform used, input file type, the amount of reads processed, read length and percentage GC content.
2. Calculated the per read quality score using the means scores of all the reads.
3. Calculated the per base sequence content to determine the distribution of the four nucleotides throughout the reads.
4. Calculated the GC content throughout the reads and compared it to a theoretical normal distribution and mean value.
5. Calculated the possibility of contamination in the reads.
6. Calculated the number of uncalled bases throughout the reads.
7. Calculated the amount of duplicate sequences.

3.4.4.4. Read alignment and mapping

The raw sequence reads were aligned to the reference genome of *M. tuberculosis* H37Rv with two respective aligners, Burrows-Wheeler aligner (BWA) (107) and Novoalign (http://www.novocraft.com/main/index.php). Using two different aligners will minimize the identification of false positive variants since the aligners use different algorithms. Both aligners produce a Sequence Alignment/Map (SAM) format. This format is compact in size and allows for most operations on the alignment to work on a stream without loading the entire alignment into the memory. This format can also be indexed, allowing the efficient and fast retrieval of all reads at a specific chromosomal locus.

Burrows-Wheeler aligner (BWA)

BWA aligns relative short reads to a reference genome by implementing two algorithms: BWA-short (for reads shorter than 200bp) and BWA-SW (for longer reads around 100kbp). BWA-short was used for this study's analysis since it does gapped global alignment, supports paired-end reads and produces results with a low error rate (< 3%). It also requires an indexed reference genome for the alignment step. This was done by using the "faidx" command from SAMtools. BWA takes fastq reads as input and uses the "aln" command to find suffix array (SA) coordinates of good hits for every read. Subsequently the "sampe" command was used to convert the SA coordinates into chromosomal coordinates and thereby producing a SAM file. Default command lines were used for this alignment procedure.

Novoalign

Novoalign also aims to produce fast and accurate alignments. This software also requires the reference genome to be indexed. This was done by the "Novoindex" command, using default values of a k-mer indexing size of 13 and a indexing step size of 1. Novoalign takes fastq reads as input and uses the Needleman Wunsh algorithm to compute the alignment. The software also does global gapped alignment and for this analysis the default value of 6 was used, thus allowing six mismatches per alignment.

3.4.4.5. SAM file validation

The "ValidateSamFile" command from Picard (http://picard.sourceforge.net) was used to report on the validity of the SAM file.

3.4.4.6. Converting SAM file to BAM file

SAM Tools (http://samtools.sourceforge.net) provide a set of various utilities for manipulating and post-processing of SAM files. This includes tools to sort, merge, and index SAM files. The SAMtools command "view" were used to convert the SAM alignment files to the binary (BAM) format

3.4.4.7. Alignment statistics

SAMtools was used to compute statistics about the alignment process. The "flagstat" command calculates the total number of reads used as input in the mapping software. Additionally it also calculates: the number of duplicate reads; the number of reads that mapped to the reference genome and the number of reads which properly paired when aligned to the reference genome.

3.4.4.8. Post alignment processing of BAM files

BAM files were subsequently processed to correct for errors incorporated during the alignment step.

Coordinate sorting and indexing of BAM files

The SAMTools functions "sort" and "index" were used to convert the BAM file in a format that is easy to manage and read. The BAM file is sorted by coordinate to avoid loading extra alignments into computational memory.

Local realignment around Indels (short insertions and deletions)

Insertions and deletions might influence the alignment of reads to its reference genome, which may result in many bases mismatching the reference near the misalignment (which could be misinterpreted for single nucleotide polymorphisms (SNPs). To minimize the number of mismatching bases across all reads, the Genome Analysis Toolkit (GATK) (Broad Institute, Cambridge, UK) were used to locally realign misaligned reads. The first step in this process includes the identification of small

intervals which were misaligned, by using the Realigner Target Creator tool. Subsequently, the IndelRealigner tool was used to realign the suspicious intervals to the reference genome, and thereby fixing the misaligned reads.

Co-ordinate sorting and indexing of Realigned BAM files

The realigned BAM files were sorted with Picard's (http://picard.sourceforge.net/) "sortsam" function and indexed with Samtools' "index" function.

Removal of PCR duplicates

Duplicate reads may be produced with PCR amplification during library construction. To mitigate the biased introduced by PCR amplification, the Picard command "Mark Duplicates" were used to locate duplicate molecules in the BAM file, which were flagged in the output BAM file.

3.4.4.9. Variant calling

The H37Rv reference genome was used to identify SNPs and short insertions and deletions (Indels) in the respective genomes. For this purpose two different variant callers were used to minimize the identification of false positive variants. Variants identified with both callers were thus used for subsequent analysis.

GATK

The UnifiedGenotyper tool from GATK was used for SNP and Indel calling and produces an output in the Variant call format (vcf) format. Stand_call_conf were set to 50. This will allow a minimum phred-scaled confidence threshold of 50, variants with a confidence value ≥ 50 are emitted as called sites. Stand_emit_conf were set to 10. Variants with a phred-scale confidence value ≥ 10 but less than the calling threshold of 50 are emitted but marked as filtered. The output vcf file contains information about the position, the alternative sequence and thePHred scaled probability that the alternative polymorphism exist at this position. The Phred scale (Q) is defined as a property which is logarithmically related to the probability that there might be an error in the actual base-calling. $Q = -10\log_{10}$, a value of 10 indicates that there is a 1 in 10 chance of error. The vcf file also contains information specific to the alternative base, including the number of reads bridging that position, as well as the number of reads that contained the reference and alternative base at that position.

Sam Tools

SAMtools' function "mpileup" were used create a pileup of the reads relative to the reference genome and subsequently identifies SNPs relative to the reference sequence. Default parameters were used. Indels called by SAM tools were found to be inconsistent and after considered evaluation, it was decided that indels called by SAM tools would be discarded from the subsequent analysis. The vcf file

contains information about the position, the alternative sequence and quality score in PHRED scale of each variant. The file also contains specific information about the variant, including the number of reads bridging that position.

3.4.5. Extraction of overlapping SNPs generated by the various pipelines

As discussed in previous sections two different aligners (BWA and Novoalign) and two different variant callers (GATK and SAM Tools) where used to minimize the identification of false positives variants (Figure 3.1). Therefore, four strategies (pipelines) where used to identify SNPs, these where namely BWA-GATK, BWA-SAMTools, Novoalign-GATK and Novoalign-SAMTools. For indel calling, only two pipelines where used: BWA-GATK and Novoalign-GATK. Thus variants were only considered in further analysis if identified with all pipelines. In-house python scripts written by Mrs Michelle Daya (Department of Biomedical Sciences, Stellenbosch University) were used to extract overlapping SNPs generated by all four pipelines (Figure 3.3). For this purpose SNPs present in both vcf files created by BWA-GATK and BWA-SAM Tools were first extracted. Likewise, SNPs present in the both vcf files created by Novoalign-GATK and Novoalign-SAMTools were also extracted. SNPs present in both of these newly created vcf files were extracted to create a final vcf file that contain high confidence SNPs identified by all four pipelines. A similar approach was followed to extract Indels present in both vcf files created by BWA-GATK and Novoalign-GATK. The statistical package R (http://www.r-project.org) was used to create Venn diagrams to illustrate the distribution of variants created with the respective pipelines.

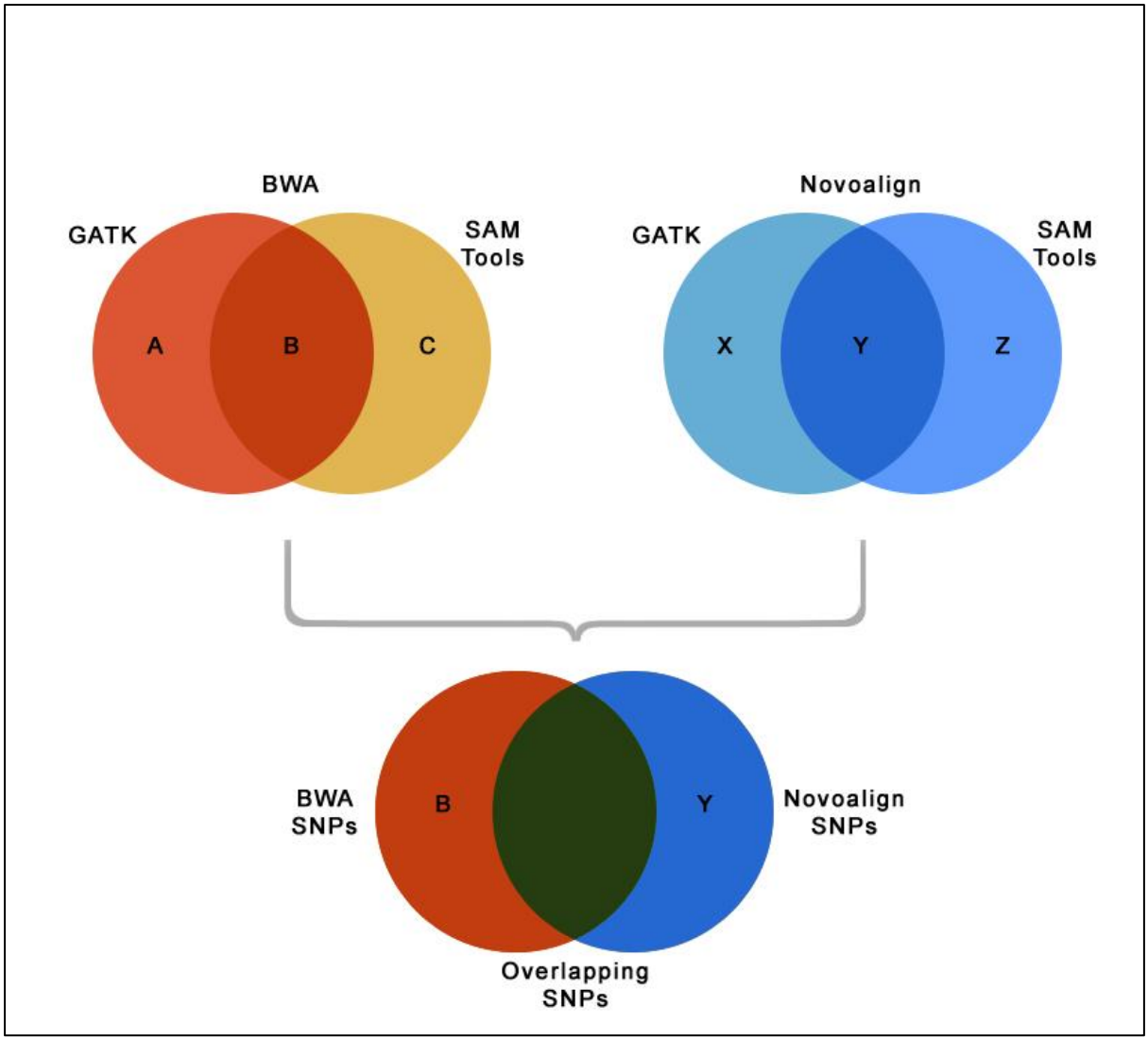

Figure 3.3: Strategy followed to identify high confidence SNPs from the four described pipelines

Legend to Figure 3.3
High confidence SNPs identified by the four described pipelines are shown in green.

3.4.6. Reconstruction of a phylogenetic tree

A python script written by Mrs Michelle Daya (Biomedical Sciences, Stellenbosch University) were used create a concatenated sequence of all polymorphic sites identified for each genome. The concatenated sequences were then aligned using Bio-Edit (v 7.1.3) (71) and used to construct a maximum likelihood phylogenetic tree (standard parameters) using MEGA 5..05 (207).

3.4.7. Annotation and functional classification of SNPs/Indels

Perl scripts were written to: 1) annotate the identified SNPs, 2) calculate the predicted aminoacid change for SNPs located in genes, 3) annotate the identified Indels 4) calculate the effect of indels on the reading frame of the respective genes, and 5) classify each variant according to its cellular function.

3.4.8. Comparison of variants between genomes

In-house python scripts written by Mrs Michelle Daya were used to compare the high confidence variants from the sequenced genomes.

3.4.9. Validation of high confidence variants

High confidence variants identified in the comparison of sequenced genomes were first validated by visualization using GenomeView (2). Subsequently primers (Table 3.9) were designed to amplify and sequence the surrounding regions of the variants as described in sections 3.1.2.2 – 3.1.2.5.

Table 3.9: Primers used to validate high confidence variants

		Primer sequence	Product size (base pairs)	T_m (°C)*
glpK	**Forward**	5'gtggatcctggaaaatgtcg 3'	509	60.32
	Reverse	5' gtacacgggtttagcgttcc 3'		59.50
malQ	**Forward**	5' atcccgatgcctcaggtgt 3'	395	62.87
	Reverse	5' gatgatgtggtcgatgcgta 3'		60.51
TB16.3	**Forward**	5' gatccaggcgaggtgatg 3'	544	60.16
	Reverse	5' cggtggataccaccaacact 3'		60.59
fadD12	**Forward**	5' gaaaccctcaagcagcatgtc 3'	309	62.10
	Reverse	5' ggcgcctaggtacactcctg 3'		62.11
lpqN	**Forward**	5' ttcaacatcaagaccgacagc 3'	644	61.23
	Reverse	5' ttagggcgtgatggtcgtc 3'		62.03
lpqQ	**Forward**	5' cggaaccgactggtcagagt 3'	329	62.64
	Reverse	5'gccggttcaatacccaacac 3'		62.45
mprA	**Forward**	5'tcgctttccttcaatggcta 3'	362	60.85
	Reverse	5' caggtcggagaacctcatgg 3'		62.98
ndhA	**Forward**	5' cacagcgaccagactgaagg 3'	363	62.03
	Reverse	5' accgtggtcatgtccatcaa 3'		62.25
rpoB	**Forward**	5' agggtcagaccacgatgacc 3'	695	62.36
	Reverse	5' gacgagggcacgtactcca 3'		62.27
Rv0283	**Forward**	5' cggtgagttcagcgacga 3'	400	62.25
	Reverse	5' gggccgaatcaacgagaata 3'		62.21
Rv1225c	**Forward**	5' gtgcttcctggtgaacaacg 3'	390	61.69
	Reverse	5' ccgatcatgaccatctgctg 3'		62.64
Rv1226c	**Forward**	5' ttggtttttcaccacctaccg 3'	332	62.01
	Reverse	5'ataccgcaaccacgacgact 3'		62.82
Rv3108	**Forward**	5' tgtcagcgtttgaggacgat 3'	647	61.81
	Reverse	5'gcgcctcaagagagtttcgt 3'		61.99

Legend to Table 3.9: * Melting temperature

CHAPTER 4

RESULTS

4.1. Epidemiological relevance of compensatory mutations in *rpoA*/C

A well-characterized convenience sample of 67 drug-susceptible and 298 drug-resistant, clinical *M. tuberculosis* isolates was tested for the presence of putative compensatory mutations in *rpoA* and *rpoC*. Initial screening of 243 rifampicin-resistant isolates from this collection identified eight isolates (3.2%) harbouring distinct mutations in *rpoA* (Figure 4.1A*)*. These mutations were D57N, T63T, S165I, D190G, and E319K. Given the infrequent occurrence of *rpoA* mutations in this sample set, the remainder of the collection was not further assessed and mutations identified in *rpoA* were excluded from further analyses.

In contrast, non-synonymous mutations in *rpoC* were detected in 23.8% of the rifampicin-resistant clinical isolates (60/252 isolates), which all also harboured mutations in RRDR of *rpoB* (Figure 4.1B). However, mutations in *rpoC* were absent from all 113 rifampicin-sensitive isolates tested. Altogether, eight different non-synonymous mutations in *rpoC* (G332R, F452C, D485Y, V483G, V483A, I491T, Q523E, and H525Q) accounted for all non-synonymous mutations detected. In addition, one synonymous substitution (A542A) was identified. This mutation was previously reported to be a phylogenetic marker for the LAM family of *M. tuberculosis* strains and therefore excluded from further analyses (40).

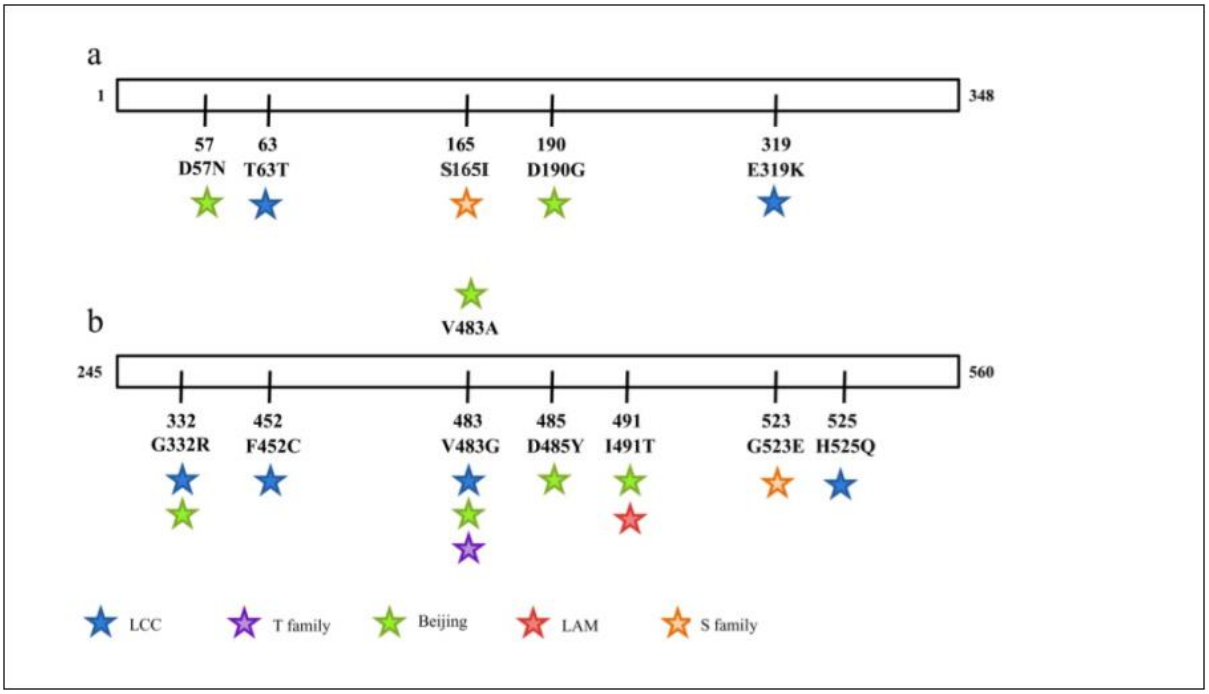

Figure 4.1 Synonymous and non-synonymous mutations identified in *rpoA* (a) and *rpoC* (b) of isolates collected from the Western Cape, South Africa. Each star indicates the presence of a mutation in a specific family of strains.

The non-synonymous mutations detected in the *rpoC* gene showed clear evidence of convergent evolution, suggesting they were positively selected among rifampicin-resistant strains. Specifically, genotyping of our collection of clinical *M. tuberculosis* isolates highlighted the independent emergence of putative compensatory mutations in *rpoC*. Three (G332R, V483G, I491T) of the eight non-synonymous mutations detected, independently evolved in phylogenetically distantly related strain families as defined by spoligotyping (Figure 4.1). Five of the eight compensatory mutations in *rpoC* (G332R, V483G, V483A, I491T, Q523E) were reported previously in MDR *M. tuberculosis* strains from various countries including Russia, Ghana, Abkhazia/Georgia, Kazakhstan and Uzbeksitan (34, 40). Mutations at amino-acid positions 452, 485 and 525 of *rpoC* were also detected previously, albeit with different amino-acid changes (previously, F452S, D485N, D485H, H525N while here, F452C, D485Y and H525Q) (34, 40).

To test whether mutations in *rpoC* were associated with increased transmissibility of drug-resistant *M. tuberculosis*, we compared the presence of *rpoC* mutations between isolates belonging to a recognized IS*6110* RFLP cluster versus isolates showing unique RFLP patterns. Among rifampicin-resistant isolates with clustered RFLP patterns, 30.7% (50/163 isolates; 95% CI: 23.7%-38.4%) harboured mutations in *rpoC*; significantly more than among rifampicin-resistant isolates with non-clustered RFLP patterns (5/40 isolates; 12.5%; 95% CI: 4.2%-26.8%; Fisher's exact test $p<0.05$) (Figure 4.2). Mutations in *rpoC* were detected in one third of the RFLP clusters present among rifampicin-resistant isolates (in 12/35 RFLP clusters) (Figure 4.2). Two of these RFLP clusters showed two distinct *rpoC* mutations and a third cluster harboured four different *rpoC* mutations. Thus, mutations in *rpoC* were acquired at least 17 times among the 35 RFLP clusters while only 5 of 40 non-clustered RFLP types harboured *rpoC* mutations.

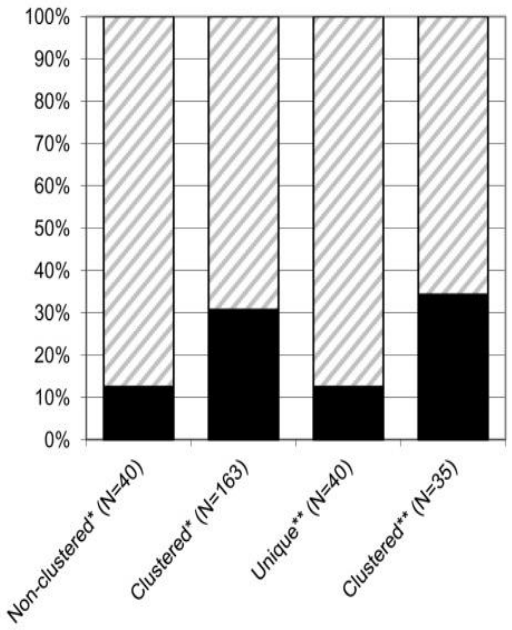

Figure 4.2 Proportion of rifampicin-resistant isolates or IS*6110* RFLP types/clusters harbouring *rpoC* mutations

* Number of isolates classified as either clustered or non-clustered based on IS*6110* RFLP type harbouring an *rpoC* mutation

** Number of IS*6110* RFLP types classified as either unique or clustered harbouring an *rpoC* mutation

Comparing the proportion of *rpoC* mutations across isolates with various degrees of drug resistance showed that this proportion increased significantly towards higher levels of drug resistance with 42.6% of XDR-TB isolates harbouring non-synonymous *rpoC* mutations (p=0.001; Table 4.1). Similarly, *rpoC* mutations were non-randomly distributed among different families of strains. In particular, a genotype of "atypical" Beijing strains known as the R86 cluster showed a four times higher proportion of isolates harbouring *rpoC* mutations than isolates belonging to other genotypes (14.4% versus 56.9% among rifampicin resistant isolates; p<0.001; Table 4.1).

Interestingly, the presence of non-synonymous mutations in *rpoC* was significantly associated with the *rpoB* S531L mutation (p<0.001; Table 4.1). Among isolates harbouring this mutation, 38.3% also harboured *rpoC* mutations while only 4.9% of the isolates with other *rpoB* mutations exhibited mutations in *rpoC*. A broad range of *rpoC* mutations was detected among isolates with an *rpoB* S531L mutation with the *rpoC* V483G mutation being most frequently detected. Only two isolates with *rpoC* mutations did not show an *rpoB* S531L mutation. Instead, these isolates harboured the L511P and D516V mutations in *rpoB* and both showed a V483G mutation in *rpoC*. A multivariate logistic regression model adjusted for different types of *rpoB* mutations, strain genotype and degree of drug resistance, confirmed strong independent associations between compensatory mutations in *rpoC*

and the R86 genotype on the one hand, and the *rpoB* S531L mutation on the other hand (Table 4.1). By contrast, the association between XDR-TB and *rpoC* mutations was not supported in this multivariate model, probably because it was confounded by an underlying association between the R86 genotype and XDR-TB.

Table 4.1 Logistic regression analysis for factors influencing the presence of compensatory mutations in *rpoC*

Characteristics	Variants	Unadjusted			Adjusted[a]		
		Freq	OR	p-value	Freq	OR	p-value
Drug resistance group[b]	R mono	55	N/A	N/A	12	N/A	N/A
	MDR *s.s.*	101	5.7	0.022	64	0.4	0.365
	Pre-XDR	49	18.3	<0.001	28	1.0	0.963
	XDR	47	19.6	<0.001	43	0.5	0.536
Strain family[c]	Non-R86	180	N/A	N/A	91	N/A	N/A
	R86 (Atypical)	58	7.8	<0.001	56	44.3	<0.001
Resistance mutation[c]	*Other rpoB mutations*	41	N/A	N/A	41	N/A	N/A
	rpoB S531L mutation	107	12.1	0.001	107	92.0	<0.001

Legend to Table 4.1:
Freq: Frequency of isolates showing a given characteristic included in the logistic regression analysis
OR: Odds-ratio
N/A: Not applicable
DS: Drug-sensitive
H mono: isoniazid mono-resistant
Poly-R: poly-resistant but non-MDR
R mono: rifampicin mono-resistant
MDR s.s.: Multidrug-resistant sensu stricto (excluding identified pre-XDR and XDR isolates)
[a]Analysis adjusted for drug resistance group, strain family and rifampicin resistance mutation
[b]Mutations in *rpoC* were only present among rifampicin resistant isolates. Other drug resistance groups were therefore excluded from the analysis
[c]Only rifampicin-resistant isolates were included

To further investigate the link between the *rpoB* S531L mutation and compensatory mutations in *rpoC,* we analyzed two previously published datasets; one from high-MDR-TB burden countries in Central Asia and one from a global collection of MDR *M. tuberculosis* isolates (40). Among isolates with an *rpoB* S531L mutation from these collections, 33.3% and 21.7% harboured *rpoC* mutations, respectively. Conversely, in only 4.3% and 2.3% of the isolates with other *rpoB* resistance mutations, respectively, *rpoC* mutations were detected. This association remained significant when adjusted for the different *M. tuberculosis* lineages in a multiple logistic regression analysis (p=0.001 and p=0.015 for the high-burden and global dataset, respectively). The fact that we could observe the same association of *rpoB* S531L with *rpoC* mutations in three independent datasets supports a biological basis for this association.

4.2. Investigation of the expression of energy metabolism genes in response to rifampicin exposure

In the initial proteomic study, an increase in the abundance of proteins involved in energy metabolism (*fixA, gltA2, opcA, lpdC, fba, atpA* and *atpH*) were detected in two MDR clinical *M. tuberculosis* isolates after exposure to 2 µg/ml rifampicin (M Bester, MSc thesis, 2009). Subsequent transcriptomic analysis by QRT-PCR confirmed a significant up regulation of two genes (*atpA* and *atpH*) in response to rifampicin exposure. These two genes encode for catalytic subunits of the F_1F_0 ATP synthase enzyme (16). This enzyme is responsible for the production of ATP by the conversion of electrochemical potential energy generated by the proton motive force, into chemical energy in the form of ATP (57). The F_1F_0 ATP synthase enzyme is encoded by a ten gene *atpBEFHAGDC* operon together with *Rv1303* and *Rv1312*, which is reported to be negatively regulated by *Rv1846c* (164). To test the hypothesis that *M. tuberculosis* increased its ATP production after rifampicin exposure, QRT-PCR was done to determine whether the expression levels of the energy metabolism genes (previously identified,Chapter 3.2.3, Table 3.4), *Rv1303*, *Rv1312* and *Rv1846c* were dependent on the genetic background of the strain, the level of rifampicin resistance and/or drug resistance profile. For this purpose isolates were selected from the Beijing and LCC lineage, displaying resistant profiles that varied from rifampicin-mono resistant to MDR (Chapter 3.2.1, Table 3.3).

Significant differential gene expression of the majority of the genes irrespective of genetic background and drug resistance profiles was observed between isolates cultured in the presence of 2 µg/ml rifampicin (Table 4.2). Four genes (*atpA, atpH, fba* and *Rv1846c*) displayed similar expression levels in all six isolates tested after rifampicin exposure. Of these *atpA*, *atpH* and *fba* showed significant up regulation ($p<0.05$, > 2 fold) while expression of *Rv1846c* remained constant. The remaining six genes *(gltA2, fixA, opcA, Rv1303, Rv1312, lpdC)* showed isolate specific gene expression with significant up regulation in at least one of genes in the isolates tested. Only one gene, *gltA*, showed significant down regulation in one of the rifampicin mono-resistant isolate after rifampicin exposure.

Analysis of the rifampicin mono-resistant isolates showed that six genes (*opcA, lpdC, fba, atpA, atpH* and *Rv1303*) were significantly up regulated after rifampicin exposure. Additionally, three genes showed isolate specific gene expression, with the *fixA* significantly up regulated in the LCC isolate and *gltA2* and *Rv1312* significantly differentially expressed in the Beijing isolate.

When separating the analysis in terms of rifampicin resistance levels in the MDR isolates, a distinct gene expression profile was observed between the high and low rifampicin MIC isolates. In the low RIF MIC MDR isolates *lpdC*, *fba*, *atpA* and *atpH* were significantly up regulated after rifampicin

exposure while in the high MIC MDR isolates and additional two genes were significantly up regulated (*gltA and opcA*). The Beijing and LCC high MIC strains also displayed strain specific significant up regulation for *fixA* and *lpdC* respectively.

Table 4.2: Differential expression profile of candidate genes

	MDR				Rifampicin mono-resistant	
	Beijing		LCC		Beijing	LCC
	R179 (High RIF MIC[a])	R257 (Low RIF MIC[b])	R451 (High RIF MIC[c])	R439 (Low RIF MIC[d])	R721[e]	R160[f]
fixA	4.13*	-1.19	1.52	1.91	1.25	2.23*
gltA2	3.64*	-1.17	2.23*	1.16	-4.61[#]	1.36
opcA	2.03*	1.37	2.12*	1.07	6.66*	2.23*
lpdC	-1.08	8.95*	2.55*	2.18*	5.26*	2.27*
fba	2.12*	2.12*	2.04*	3.12*	2.38*	5.24*
atpA	3.51*	2.00*	2.22*	4.07*	4.50*	2.04*
atpH	2.62*	3.14*	4.12*	3.72*	2.76*	2.17*
Rv1303	-1.18	1.16	1.22	1.16	2.86*	2.24*
Rv1312	1.08	1.11	1.25	1.85	3.97*	1.02
Rv1846c	-1.16	-1.54	-1.99	-1.21	-1.92	-1.23

Legend to Table 4.2:
RIF: Rifampicin
[a] Rifampicin MIC: 170 µg/ml, [b] Rifampicin MIC: 60 µg/ml, [c] Rifampicin MIC: 140 µg/ml, [d] Rifampicin MIC: 30 µg/ml, [e] Rifampicin MIC: 70 µg/ml, [f] Rifampicin MIC: 80 µg/ml
* Significant up regulation (and highlighted in grey)
[#] Significant down regulation (and highlighted in blue)

4.3. Selection and characterization of rifampicin resistant *in vitro* mutants

Spontaneous *in vitro* mutants resistant to rifampicin were selected from a pan susceptible Beijing cluster 208 clinical isolate (K636) and a H37Rv laboratory strain of *M. tuberculosis* (Chapter 3.3) for subsequent genotypic and phenotypic characterisation (Chapter 3.3.3 and 3.3.4). In total 54 Beijing and 94 H37Rv rifampicin resistant mutants were selected from twenty and 100 7H10 Middlebrook plates containing 2µg/ml rifampicin, respectively. Targeted sequencing of the RRDR of the *rpoB* gene revealed that 52 of the 54 Beijing resistant mutants harboured mutations, while all 94 H37Rv resistant mutants harboured mutations in the RRDR of the *rpoB* gene (Appendix, Table A1 and A2). Altogether, nine mutations in the RRDR of the *rpoB* gene (Asp516Val, Ser522Leu, Ser522Trp, His526Asp, His526Arg, His526Tyr, Ser531Trp, Ser531Leu) (Figure 4.3) accounted for all mutations detected in the strains from the two different genetic backgrounds. The frequency of these mutations is shown in Figure 4.3, together with what was previously found by Huitric *et al* in Table 4.3 (82). Mutations at position 526 accounted for 56.5% of all mutations detected in the Beijing resistant

mutants, with the amino acid change to aspartic acid being the most dominant (29.6%). Similarly, mutations at position 526 accounted for 69.1% of all mutations detected in the H37Rv resistant mutants, with the amino acid change to tyrosine being the most frequent (29.6%). In contrast, only 18.5% and 13.8% of the Beijing and H37Rv resistant mutants harboured mutations at position 531 respectively, while 11.1% and 15.97% harboured mutations at position 522, respectively. Only one H37Rv resistant mutant harboured a mutation (Asp to Val) at position 516. Four Beijing resistant mutants harboured multiple mutations in the RRDR of *rpoB*, namely His526Arg/Ser531Trp, His526Asp/Ser531Leu, His526His/Pro/Asp/Tyr, His526His/Arg (Appendix, Table A1). These mutants might reflect a mixed population of single colonies. No significant statistical difference was observed for the frequencies of the different *rpoB* mutants selected from the two genetic backgrounds. However, the frequency of mutants selected from the H37Rv background harbouring the His526Asp, H526Arg, Ser531Leu and Ser522Trp mutations were significantly different from a previous study (82) (two sample proportion test, $P < 0.05$). Similarly, mutants selected from the Beijing background harbouring His526Asp and Ser522Trp mutations showed significant differences in their frequency when compared to a previous study (82) (two proportion test, $P < 0.05$).

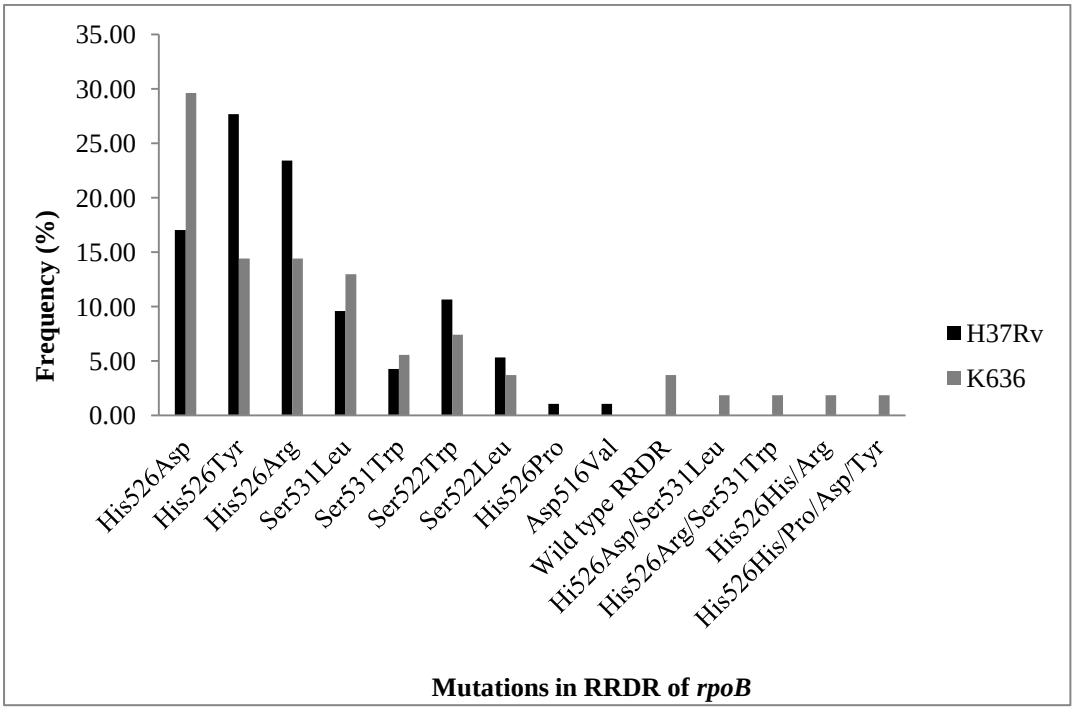

Figure 4.3: Frequency of *rpoB* mutations observed in *in vitro* mutants selected from a pan susceptible Beijing cluster 208 (K636) clinical isolate and the laboratory strain, H37Rv

Table 4.3: Frequency of *rpoB* mutations detected in *in vitro* mutants selected from a pan susceptible Beijing cluster 208 (K636) clinical isolate and the laboratory strain, H37Rv

	Frequency of *rpoB* mutations			
	H37Rv		Beijing	
***rpoB* mutation**	**Current study (%)**	**Huitric *et al* (%)***	**Current study (%)**	**Huitric *et al* (%)***
His526Asp	16 (17.0)[a]	3 (3)[a]	16 (29.6)[b]	3 (3.4)[b]
His526Tyr	26 (27.7)	22 (22)	8 (14.8)	20 (22.4)
His526Arg	22 (23.4)[a]	10 (10)[a]	8 (14.8)	16 (18.0)
Ser531Leu	9 (9.6)[a]	35 (35)[a]	7 (13.0)	20 (22.4)
Ser531Trp	4 (4.3)	6 (6)	3 (5.6)	4 (4.5)
Ser522Trp	10 (10.6)[a]	0 (0)[a]	4 (7.4)[b]	0[b]
Ser522Leu	5 (5.3)	13 (13)	2 (3.7)	12 (13.5)
His526Pro	1 (1.1)	0	0	1 (1.12)
Asp516Val	1 (1.1)	0	0	2 (2.25)
Hi526Asp/Ser531Leu	0	0	1 (1.9)	0
His526Arg/Ser531Trp	0	0	1 (1.9)	0
His526His/Arg	0	0	1 (1.9)	0
His526His/Pro/Asp/Tyr	0	0	1 (1.9)	0
Wild type RRDR	0	0	2 (3.7)	0

Legend for Table 4.3:
* Reference: (82)
[a] Significant difference in the frequency of respective *rpoB* mutation in the H37Rv background between current study and Huitric *et al* (highlighted in grey)
[b] Significant difference in the frequency of respective *rpoB* mutation in the Beijing background between current study and Huitric *et al* (highlighted in blue)

Since mutations at positions 531 and 526 in the *rpoB* gene are most frequently seen in clinical *M. tuberculosis* isolates (152), mutants harbouring mutations at these positions were selected for rifampicin MIC determination using the BACTEC MGIT 960 system. Additionally mutants selected from the H37Rv and harbouring mutations at position 522 were also subjected for MIC determination. The rifampicin MICs observed for the respective *rpoB* mutations are shown in Table 4.4. Rifampicin MICs were found to be highly variable among the Beijing and H37Rv mutants harbouring identical *rpoB* mutations. Mutants harbouring mutations at position 526 displayed higher rifampicin MIC values than the 531 mutants, with the majority of 526 mutants displayed rifampicin MICs that exceeds 200 µg/ml. In contrast, mutants harbouring the Ser531Leu mutation showed MICs that varied between 60 and greater than 125 µg/ml, while mutants harbouring the Ser531Trp mutations displayed varying rifampicin between 100 µg/ml and greater than 200 µg/ml. Mutants harbouring the Ser522Leu mutation displayed rifampicin MICs lower or equal to 2 µg/ml, while mutants harbouring the Ser522Trp mutation displayed rifampicin MICs between 50 and 100 µg/ml. Varying levels of resistance also existed between the Beijing resistant mutants sharing identical *rpoB* and selected from a common progenitor. Together these results suggest that mechanisms other than the *rpoB* mutation contribute to higher level of rifampicin resistance.

Table 4.4: Rifampicin MIC values for mutants selected from a Beijing clinical and H37Rv laboratory strain

rpoB mutation	Strain Background	Clone name	Plate number	Rifampicin MIC (µg/ml)
His526Asp	Beijing	K636.53	26	>100
		K636.20	27	> 150
		K636.50	39	>150
		K636.36	44	>200
	H37Rv	20	20	>200
		25	25	100 < x <150
		32	32	>200
His526Tyr	Beijing	K636.21	2	>200
		K636.7	25	>200
		K636.24	25	>200
		K636.17	29	>200
		K636.40	39	>200
		K636.35	44	150 < x <200
	H37Rv	3	3	>200
		6	6	>150
		27	27	>200
		76	76	>200
		91	91	>200
		92	92	>200
His526Arg	H37Rv	16	16	>200
		28	28	>200
		34	34	>200
		36	36	>200
		70	70	150 < x< 200
Ser531Leu	Beijing	K636.1	36	60
		K636.2	37	100 < x <150
		K636.44	37	>100
		K636.45	38	>125
		K636.3	42	80
		K636.48	42	50
	H37Rv	17	17	>100
		30	30	100 < x <150
		67	67	100 < x < 200
		75	75	50 < x < 100
Ser531Trp	Beijing	K636.43	25	>150
		K636.33	43	>100
	H37Rv	43	43	>200
Ser522Leu	H37Rv	18	18	< 2
		26	26	2
Ser522Trp	H37Rv	2	2	>50
		9	9	50 <x < 100
		72	72	> 50

4.4. Identification of genetic variants responsible for varying rifampicin resistance levels

To identify genetic variants responsible for modulating the rifampicin resistance levels in *M. tuberculosis* the genomes of seven rifampicin resistant isolates were sequenced on the Illumina platform (Chapter 3.4, Table 3.8). These seven genomes included strains from evolutionary distinct lineages, i.e. two *in vitro* selected rifampicin resistant clones from the progenitor Beijing lineage strain K636 (N37S1 and N36S1) and four clinical MDR isolates from the Beijing (R451, R439) and LCC lineages (R257, R179), respectively.

4.4.1. Quality of raw reads and mapping coverage

FastQC was used to generate a comprehensive quality control report for each of the genome sequences (Appendix, Figure A1-A6) that confirmed the quality and quantity of sequencing data. The quantity of these reads corresponded to an average coverage of 422 fold (341 - 477) for the seven genomes and GC content of the sequencing data ranged between 63 and 65% as expected for *M. tuberculosis* genomes (Table 4.5). A deterioration of the sequence quality of the reads was observed in the last 30 bp. Though a possibility would be those low-quality reads, we believed that the depth and quality of the reads would be sufficient for mapping and SNP calling. The paired-end reads from each genome were aligned to H37Rv using BWA and Novoalign. Subsequently, the SAM Tools "flagstat" tool was used to calculate the mapping coverage of each aligner (Table 4.6). The percentage of reads mapping to the reference genome using Novoalign ranged between 62.53 and 83.07%, while BWA produced a mapping percentage of between 56.81 and 73.21%.

Table 4.5: Statistical results produced by FastQC for each genome

Isolate name	Genetic background	Drug resistance	Total reads (base pair)	Read length	Raw coverage (x)	GC content (%)
N36S1	Beijing	Rifampicin-mono (*in vitro*)	18541342	100	420.29	63
N37S1		Rifampicin-mono (*in vitro*)	15080134	100	341.83	64
K636		Pan susceptible	17310702	100	392.40	65
R451	LCC	MDR	21166742	100	479.80	64
R439		MDR	18533476	100	420.11	65
R257	Beijing	MDR	18685916	100	423.57	64
R179		MDR	21057802	100	477.33	65

Table 4.6: Alignment statistics generated by SAM Tools for Novoalign and BWA

	Novoalign SAM Tools statistics			BWA SAM Tools statistics		
Isolate name	**Total nr reads ***	**Reads mapped (%)**	**Reads properly paired to reference (%)**	**Total nr reads ***	**Reads mapped (%)**	**Reads properly paired to reference (%)**
N36S1	18541342	13160743 (70.98)	12973790 (69.97)	18541342	12261205 (66.13)	11329964 (61.11)
N37S1	15080134	9504621 (63.03)	9428962 (62.53)	15080134	9084970 (60.24)	8566518 (56.81)
K636	17310702	14061826 (81.23)	13883056 (80.2)	17310702	13423591 (77.55)	12673136 (73.21)
R451	21166742	16514674 (78.02)	16382912 (77.4)	21166742	15818582 (74.73)	14996350 (70.85)
R439	18533476	14706215 (79.35)	14561646 (78.57)	18533476	14173632 (76.48)	13419810 (72.41)
R257	18685916	14534376 (77.78)	14425222 (77.2)	18685916	13747583 (73.57)	12919820 (69.14)
R179	21057802	17617283 (83.66)	17493194 (83.07)	21057802	17024433 (80.85)	16227538 (77.06)

Legend to Table 4.6: * Total number of reads that passed quality control

4.4.2. Variants identified using H37Rv as reference genome

As described in chapter 3.4.4 and 3.4.5, different strategies were used to identify high confidence SNPs and indels for each genome against H37Rv as the reference genome. The number of SNPs and indels identified by each strategy for each genome is summarized in Table 4.7. SNPs and indels identified by the respective pipelines were then extracted for further analysis using Python scripts. A representation of the strategies followed to identify high confidence SNPs for each genome is shown in Figure 4.4. In comparison with H37Rv, the number of high confidence SNPs identified in the seven genomes ranged between 1065 and 1560, whereas the number of high confidence indels identified ranged between 93 and 166. The size of the short insertions and deletions ranged between 1 and 29 base pairs.

Table 4.7: Number of SNPs/Indels identified by the different strategies used in this study

		Beijing 208			Beijing 220		LCC DRF150	
		K636	**N36S1**	**N37S1**	**R179**	**R257**	**R451**	**R439**
SNPs	BWA-GATK	6793	11462	11299	6236	6083	7285	8637
	BWA-SAM Tools	2189	3287	3396	2081	2315	1915	1845
	Overlapping BWA	**1964**	**3029**	**3124**	**1908**	**2063**	**1778**	**1709**
	Novoalign-GATK	1750	1716	1763	1878	1876	1216	1219
	Novoalign-SAM Tools	1724	1702	1684	1752	1875	1195	1192
	Overlapping Novoalign	**1610**	**1585**	**1586**	**1649**	**1686**	**1129**	**1119**
	Overlapping BWA-Novoalign*	**1533**	**1531**	**1530**	**1560**	**1607**	**1068**	**1065**
Indels	BWA-GATK	327	334	330	360	301	289	287
	Novoalign-GATK	193	171	166	218	195	118	121
	Overlapping BWA-Novoalign*	**147**	**139**	**140**	**166**	**148**	**93**	**96**

Legend to Table 4.7: * High confidence variants used for further analysis

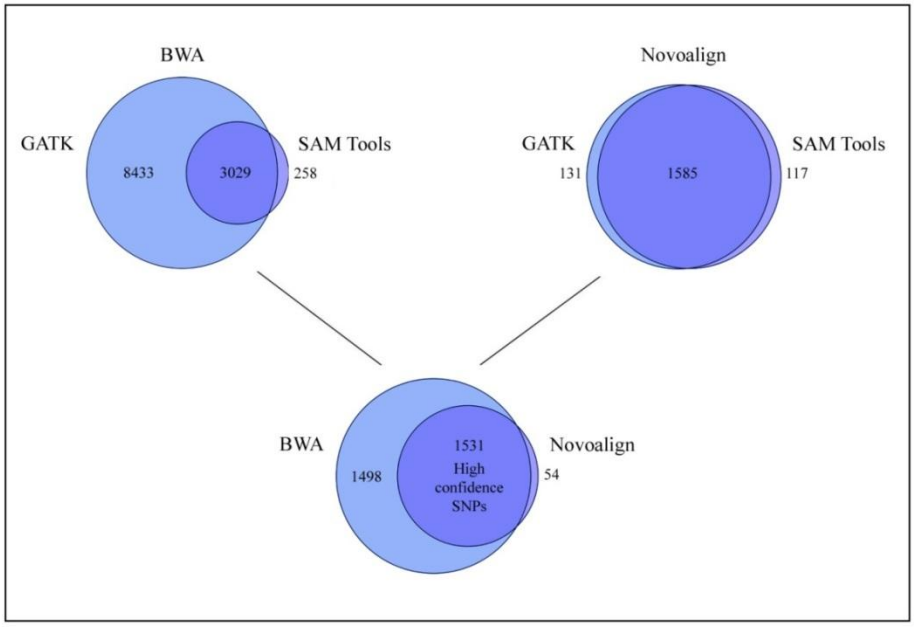

Figure 4.4: Representation of the strategies used to identify high confidence SNPs in one of the 7 genomes (the *in vitro* mutant N36S1)

Subsequently, Perl scripts were used to annotate the variants, categorise them according to cellular function and to predict the effect of the variants on protein level. Approximately 50 % of the SNPs were identified in the coding regions classified as conserved hypothetical proteins, cell wall and cellular processes, or intermediary metabolism and respiration (Table 4.8). The highest number of indels was identified in intergenic regions, as well as in genes classified to the PE-PPE family of proteins (Table 4.9). High confidence SNPs from each genomes were subsequently used for the reconstruction of a phylogenetic tree (Figure 4.5).

Table 4.8: SNPs identified in functional categories in all the *M. tuberculosis* isolates using H37Rv as reference genome

	Beijing 208			Beijing 220		LCC DRF150	
	K636 (%)	N36S1(%)	N37S1(%)	R179(%)	R257(%)	R451(%)	R439(%)
Cell wall and cell processes	274 (17.9)	274 (17.9)	271 (17.7)	276 (17.7)	277 (17.2)	173 (16.2)	171 (16.1)
Conserved hypothetical	233 (15.2)	353 (15.3)	234 (15.3)	243 (15.6)	244 (15.2)	161 (15.1)	164 (15.4)
Hypothetical	4 (0.3)	3 (0.2)	4 (0.3)	4 (0.3)	5 (0.3)	1 (0.1)	1 (0.1)
Information pathways	80 (5.2)	80 (5.2)	80 (5.2)	84 (5.38)	83 (5.2)	50 (4.7)	50 (4.7)
Insertion seqs and phages	46 (3.0)	48 (3.1)	46 (3.0)	40 (2.56)	43(2.67)	29 (2.72)	31 (2.9)
Intermediary metabolism and respiration	273 (17.8)	271 (17.7)	272 (17.8)	282(18.07)	284 (17.7)	188 (17.6)	186 (17.5)
Lipid metabolism	134 (8.7)	134 (8.8)	133 (8.7)	132 (8.46)	132 (8.2)	101 (9.5)	186 (17.5)
PE/PPE families	208 (13.6)	208 (13.6)	204 (13.3)	210 (13.5)	241 (15.0)	177 (16.6)	98 (9.2)
Regulatory	54 (3.5)	54 (3.5)	54 (3.5)	58 (3.7)	59 (3.7)	29 (2.72)	173 (16.2)
Virulence, detoxification and adaptation	45 (2.9)	45 (2.9)	45 (2.9)	45(2.9)	49 (3.05)	32 (3.00)	29 (2.7)
Intergenic	182 (11.9)	179 (11.7)	187 (12.2)	187 (12.0)	191 (11.9)	127 (11.9)	131 (12.3)
Total	**1533**	**1531**	**1530**	**1561**	**1608**	**1068**	**1065**

Table 4.9: Indels identified in functional categories in all *M. tuberculosis* isolates using H37Rv as reference genome

	Beijing 208			Beijing 220		LCC DRF150	
	K636(%)	N36S1(%)	N37S1(%)	R179(%)	R257(%)	R451(%)	R439(%)
Cell wall and cell processes	19 (12.9)	18 (13.0)	20 (14.3)	21 (12.7)	20 (13.5)	9 (9.7)	10 (10.4)
Conserved hypothetical	21 (14.3)	20 (14.3)	21 (15.0)	28 (16.9)	22 (14.9)	14 (15.1)	14 (14.6)
Hypothetical	2 (1.4)	2 (1.4)	2 (1.4)	2 (1.2)	2 (1.4)	1 (1.1)	1 (1.0)
Information pathways	1 (0.7)	1 (0.7)	1 (0.7)	1 (0.6)	1 (0.7)	1 (1.1)	1 (1.0)
Insertion seqs and phages	4 (2.7)	4 (2.9)	4 (2.9)	4 (2.4)	4 (2.7)	1 (1.1)	1 (1.0)
Intermediary metabolism and respiration	11 (7.5)	12 (8.6)	12 (8.6)	11 (6.6)	12 (8.1)	13 (14.0)	13 (13.5)
Lipid metabolism	8 (5.4)	8 (5.8)	8 (5.7)	11 (6.6)	9 (6.1)	4 (4.3)	6 (6.3)
PE/PPE families	**34 (23.1)**	**24 (17.3)**	**23 (16.4)**	**36 (21.7)**	**27 (18.2)**	**20 (21.5)**	**18 (18.8)**
Regulatory	5 (3.4)	6 (4.3)	6 (4.3)	9 (5.4)	8 (5.4)	1 (1.1)	1 (1.0)
Virulence, detoxification and adaptation	4 (2.7)	5 (3.6)	4 (2.7)	4 (2.4)	4 (2.7)	0 (0.0)	0 (0.0)
Intergenic	**38 (25.8)**	**39 (28.1)**	**39 (27.9)**	**39 (23.5)**	**39 (26.3)**	**29 (31.2)**	**31 (32.3)**
Total	**147**	**139**	**140**	**166**	**148**	**93**	**96**

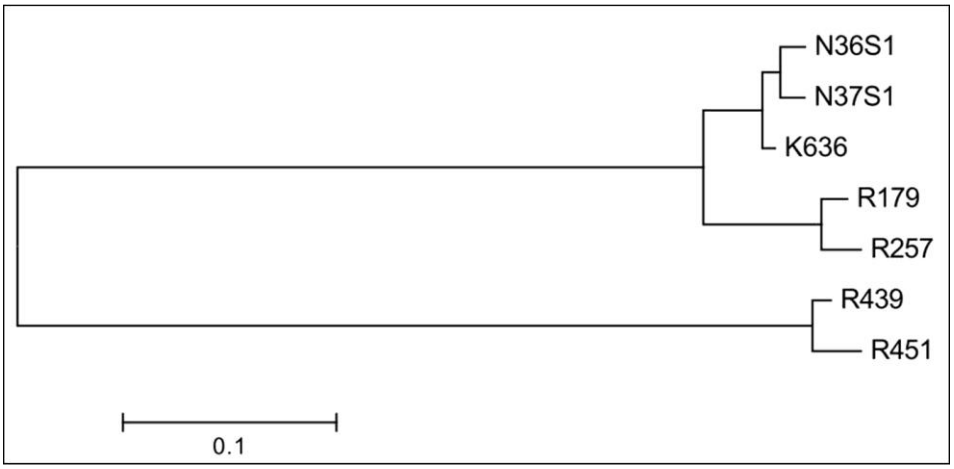

Figure 4.5: Phylogenetic tree constructed by high confidence SNPs identified in all the isolates

The above analysis implies that the observed variants were acquired in the LCC and Beijing isolates. This is however not the case as H37Rv is not the progenitor of these two lineages. In order to identify SNPs unique to the Beijing or LCC lineage, high confidence SNPs present in all seven genomes were extracted from the high confidence SNPs files. The 539 SNPs present in all Beijing and LCC isolates represents changes that occurred in H37Rv (Figure 4.6). A total of 850 SNPs were found to be unique to members of the Beijing lineage, while 485 SNPs were found to be unique to members of the LCC lineage. Functional categorization of these SNPs revealed that the highest proportion of SNPs to the total number of genes in each functional class was identified in PE-PPE genes (Table 4.10). Thereafter, SNPs in lipid metabolism play a significant role in the adaptation of the pathogen, which may reflect different levels of pathogenicity for the members of the three lineages. Similarly, genes functionally classified in information pathways and cell wall and cell processes was also shown to be under evolutionary pressure.

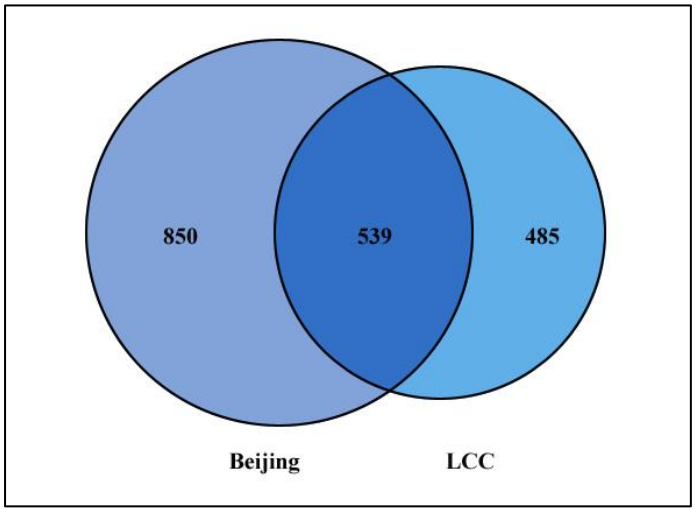

Figure 4.6: Distribution of high confidence SNPs between the Beijing and LCC lineage relative to H37Rv

Table 4.10: Functional categorization of high confidence SNPs unique to H37Rv, Beijing and LCC lineages

	Number of genes[a]	H37Rv unique (%*)	Beijing unique (%*)	LCC unique (%*)
Cell wall and cell processes	773	97 (12.5)	148 (19.1)	69 (8.9)
Conserved hypothetical	1031	86 (8.3)	132 (12.8)	74 (7.2)
Hypothetical	16	1 (6.3)	2 (12.5)	0 (0.0)
Information pathways	242	30 (12.4)	47 (19.5)	20 (8.3)
Insertion seqs and phages	147	9 (6.1)	28 (19.04)	20 (13.6)
Intermediatry metabolism and respiration	936	92 (9.8)	154 (16.5)	92 (9.8)
Lipid metabolism	271	55 (20.3)	67 (24.7)	41 (15.1)
PE-PPE	168	69 (41.1)	113 (67.3)	83 (49.4)
Regulatory	198	20 (10.1)	32 (16.2)	9 (4.5)
Virulence, detoxification and adaptation	238	16 (6.7)	26 (10.9)	14 (5.9)
Total	**4020**	**539**	**850**	**485**

Legend to Table 4.10:

[a] Number of genes belonging to each functional category in *M. tuberculosis*

* Proportion of SNPs to the total number of genes classified to each functional category

4.4.3. Comparative analysis of the high confidence variants between genomes

Python scripts were used to compare the high confidence variants between the different genomes and GenomeView was used to eliminate false positives. Due to the high GC content found in transposable and repetitive elements (i.e. PE-PPE/PGRS and IS*6110*), reads containing these elements are difficult to map to the reference genome. Therefore variants identified in these regions were excluded from further analysis since these regions are high GC rich.

4.4.3.1. Increased rifampicin resistance acquired in *in vitro* selected mutants

To identify genetic variants which might result in increased level of rifampicin resistance acquired *in vitro*, the genomes of two *in vitro* selected rifampicin resistant mutants (N36S1, N37S1) and its Beijing pan susceptible progenitor (K636) were compared. Only one SNP in addition to the *rpoB* S531L mutation was identified to be unique to the genome of N36S1, after comparing the high confidence variants (relative to H37Rv) identified in the genomes of N36S1 (rifampicin MIC: 60 µg/ml) and K636 (Table 4.11). This synonymous SNP was identified in a conserved hypothetical protein encoded by *TB16.3.* One deletion in *glpK,* encoding for a glycerol kinase, was identified to be unique to N36S1. This deletion causes a frame shift which will result in a premature stop codon (Table 4.12).

Two SNPs in addition to the *rpoB* S531L were identified to be unique to the genome of N37S1, after comparing the high confidence variants (relative to H37Rv) identified in the genomes of N37S1 (rifampicin MIC: > 100 µg/ml) and K636 (Table 4.11). Two of these SNPs were identified in genes *TB16.3* and *malQ.* The gene *malQ* is annotated as a 4-alpha-glucanotransferase and is involved in

glycogen degradation. Similar to the genome of N36S1, the deletion at position 573 in *glpK* was also identified in the genome of N37S1 (Table 4.12).

Table 4.11: SNPs found to be unique to the genomes of the *in vitro* mutants N36S1 and N37S1

Locus	Gene	Base change	Amino acid change	Gene description	Functional Group	SNP present in: N36S1	SNP present in: N37S1
Rv0667	*rpoB*	C1349T	Ser450Leu*	DNA-directed RNA polymerase beta chain	Information pathways	Yes	Yes
Rv1781c	*malQ*	G1221T	Val407Val	4-alpha-glucanotransferase	Intermediary metabolism and respiration	No	Yes
Rv2185c	*TB16.3*	T303C	Gly101Gly	Conserved hypothetical protein	Conserved hypothetical	Yes	Yes

Legend for Table 4.11:
* Amino acid change according to *M. tuberculosis* gene sequence. Would translate to S531L according to the *E. coli* gene sequence

Table 4.12: Indels found to be unique to the genomes of the *in vitro* mutants N36S1 and N37S1

Locus	Gene	Position in gene	Indel	WT ORF	MUT ORF	Functional ORF	Description	Functional Group
Rv3696c	*glpK*	573	C (deletion)	1-518	1-252	1-191	Glycerol kinase	Intermediary metabolism and respiration

Legend for Table 4.12:
WT = wild type
MUT = mutant
ORF = open reading frame

Variants identified unique to the *in vitro* rifampicin resistant mutants were validated by targeted DNA sequencing of the two mutants and the pan susceptible progenitor isolate. PCR amplification and Sanger sequencing confirmed the presence of the SNP in *TB16*.3 in both mutants (N37S1 and N36S1) the SNP in *malQ* in N37S1. The deletion identified in *glpK* in both mutant genomes was also present in the progenitor isolate, K636, and therefore was excluded as a candidate responsible for modulating rifampicin resistance levels. The SNPs in *malQ* and *TB16.3* were screened in the remaining panel of 52 *in vitro* rifampicin resistant mutants and their progenitor culture (Section 3.3.2.2) before rifampicin selection (Appendix Table A1). The V407V change in *malQ* was identified in only one additional mutant, K636.44 (rifampicin MIC > 100 µg/ml). This mutant shares the same progenitor before rifampicin selection as N37S1. The G101G SNP in *TB16*.3 was identified in 20 of the 52 mutants screened. These 20 mutants represented 10 of the 20 progenitors before rifampicin selection. An additional four progenitors also harboured this SNP, while the mutants selected from these progenitors harboured the wild type nucleotide. This suggested that the progenitors before rifampicin selection comprised a mix population with both wild type and mutant base codons in *TB16.3*.

4.4.3.2. Increased rifampicin resistance acquired *in vivo* in Beijing strain background

To identify genetic variants which might result in increased rifampicin resistance acquired *in vivo,* the high confidence variants (relative to H37Rv) identified in the two closely related Beijing clinical isolates, R179 (Rifampicin MIC: 140 µg/ml) and R257 (Rifampicin MIC: 30 µg/ml), were compared. In total, five non-synonymous and four synonymous SNPs were unique to the isolate with the high rifampicin MIC isolate R179 (Table 4.13). Of the five non-synonymous SNPs, three were identified in genes known to confer drug resistance in *M. tuberculosis*, namely *rpoB* (additional to the *rpoB*S531L)*, rpsL, ethA.* The two remaining non-synonymous SNPs were identified in *lpqN* and *Rv3108*, which encodes for a lipoprotein and hypothetical protein respectively. Additionally, four Indels were identified unique to the genome of R179 (Table 4.14). All four indels were identified in non-essential genes (170) involved in various cellular functions including lipid metabolism (*echA*), regulation (*mprA*) and those with unknown function (*Rv2449c, Rv3195*). Only one of the four indels (one base pair deletion in *mprA*) would result in a frame shift and premature stop codon.

Table 4.13: SNPs found to be unique to the genome of R179

Locus	Gene	Base change	Amino acid change	Gene description	Functional Group
Rv0583c	*lpqN*	G235A	Asp7Asn	Lipoprotein	Cell wall and cell processes
Rv0667	*rpoB*	C1147T	Pro483Ser	DNA-directed RNA polymerase beta chain	Information pathways
Rv0682	*rpsL*	A128G	Lys43Arg	30S ribosomal protein S12	Information pathways
Rv1057		G21A	Arg7Arg	Conserved hypothetical protein	Conserved hypothetical
Rv2850c		C1044T	Asp348Asp	Magnesium chelatase	Intermediary metabolism and respiration
Rv2946c	*pks1*	C1134G	Pro378Pro	Polyketide synthase	Lipid metabolism
Rv3108		A245G	Asp82Gly	Hypothetical protein	Conserved hypothetical
Rv3546	*fadA5*	G681C	Gly227Gly	Acetyl-CoA acetyltransferase	Lipid metabolism
Rv3854c	*ethA*	T409C	C137Arg	Monooxygenase	Intermediary metabolism and respiration

Legend to Table 4.13:
Non-synonymous SNPs are highlighted in grey

Table 4.14: Indels found to be unique to the genome of R179

Locus	Gene	Position in gene	Indel	Change	WT ORF	MUT ORF	Description	Functional group
Rv0222	*echA1*	332	TGGCGGG (deletion)	In frame deletion	1-263	1-261	Enoyl-CoA hydratase	Lipid metabolism
Rv0981	*mprA*	165	A (deletion)	Truncated protein	1-231	1-80	Mycobacterial persistence regulator	Regulatory
Rv2449c		488	GCCGTC (deletion)	In frame deletion	1-420	1-418	Conserved hypothetical protein	Conserved hypothetical
Rv3195		588	AACTGGG C (deletion)	In frame deletion	1-473	1-470	Conserved hypothetical protein	Conserved hypothetical

Legend for Table 4.14:
WT = wild type
MUT = mutant
ORF = open reading frame

To identify genetic variants which might result in decreased level of rifampicin resistance acquired *in vivo,* variants identified unique to the genome of the low rifampicin MIC isolate of R257 was also investigated (Table 4.15). From the nine SNPs identified as unique to low MIC isolate R257, six were non-synonomous and involved in various cellular functions including information pathways (*rpoC*), virulence, detoxification and adaptation (*katG, mce3F*), intermediary metabolism and respiration (*moeB2*) and those with unknown function (*Rv0311).* In addition, two Indels (Table 4.16) unique to the genome of R257 were identified in non-essential genes *ethA* and *pncA* involved in intermediary metabolism and respiration (170). Mutations in these two genes are known for conferring resistance to anti-TB drugs ethionamide and pyrazinamide.

Table 4.15: SNPs found to be unique in the genome of R257

Locus	Gene	Base change	Amino acid change	Gene function	Functional Group
Rv0169	*mce1A*	G183A	Gly61Gly	MCE-family protein	Virulence, detoxification and adaptation
Rv0311		T83C	Ile28Thr	Hypothetical protein	Conserved hypothetical
Rv0668	*rpoC*	T1450G	Trp484Gly	DNA-directed RNA polymerase beta chain	Information pathways
Rv0823c		C960T	Leu320Leu	Transcriptional regulator	Regulatory
Rv1046c		G59A	Arg20His	Hypothetical protein	Hypothetical
1472359*	Intergenic	A/C-	-	-	-
Rv1908c	*katG*	T610C	Trp204Arg	Catalase-peroxidase-peroxynitritase T	Virulence, detoxification and adaptation
Rv1971	*mce3F*	C686A	Ala229Glu	MCE-family protein	Virulence, detoxification and adaptation
Rv3116	*moeB2*	G137A	Cys46Glx	Molybdenum cofactor biosynthesis protein	Intermediary metabolism and respiration

Legend to Table 4.15:
Non-synonymous SNPs are highlighted in grey
* Genome position

Table 4.16: Indels found to be unique to the genome of R257

Locus	Gene	Position in gene	Indel	Change	WT ORF	MUT ORF	Description	Functional group
Rv2043c	*pncA*	467	GCACCCT G (deletion)	Frame shift	1-187	Extends to upstream gene	Pyrazinamidase	Intermediary metabolism and respiration
Rv3854c	*ethA*	66	T (deletion)	Truncated protein	1-490	1-62	Monooxygenase	Intermediary metabolism and respiration

Legend for Table 4.16:
WT = wild type
MUT = mutant
ORF = open reading frame

4.4.3.3. Increased rifampicin resistance acquired *in vivo* in LCC strain background

To identify genetic variants which might result in increased rifampicin resistance acquired *in vivo*, the high confidence variants (relative to H37Rv) identified in the two closely related LCC clinical isolates, R451 (Rifampicin MIC: 170 µg/ml) and R439 (Rifampicin MIC: 60 µg/ml) were compared. In total, seven non-synonymous and three synonymous SNPs were identified to be unique to the genome of high MIC isolate R451 (Table 4.17). The seven non-synonymous SNPs were in genes involved in various processes including cell wall and cell processes (*Rv0283*, *lpqQ*, *Rv1081c*, *Rv1226c*); lipid metabolism (*omt);* virulence, detoxification and adaptation (*cstA)* and intermediary metabolism and respiration (*icd1)*. Additionally, two intragenic Indels and one intergenic indel were unique to the genome of R451 (Table 4.18). The Indels identified in *ndhA* and *FadD12* results in a frameshift and premature stop codon. *NdhA* is annotated as a Membrane NADH dehydrogenase involved in intermediary metabolism and respiration, while *FadD12* is annotated as a Long-chain fatty-acid-CoA ligase and involved in fatty acid metabolism.

Table 4.17: SNPs found to be unique to the genome of R451

Locus	Gene	Base change	Amino acid change	Gene function	Functional Group
Rv0283		G67A	Val23Ile	Conserved membrane protein	Cell wall and cell processes
Rv0440	*groEL*	G1350A	Gln450Gln	60 kda chaperonin 2 groEL2	Virulence, detoxification and adaptation
Rv0787		C270A	Ala90Ala	Hypothetical protein	Conserved hypothetical
Rv0835	*lpqQ*	A485G	Asp162Gly	Lipoprotein	Cell wall and cell processes
Rv1081c		C237A	Asp79Glu	Conserved membrane protein	Cell wall and cell processes
Rv1153c	*omt*	T467C	Val156Ala	O-methyltransferase	Lipid metabolism
Rv1226c		C247T	Arg83Trp	Transmembrane protein	Cell wall and cell processes
Rv2236c	*cobD*	G882A	Arg294Arg	Cobalamin biosynthesis transmembrane protein	Intermediary metabolism and respiration
Rv2289	*cdh*	C426T	Asn142Asn	cdp-diacylglycerol pyrophosphatase	Lipid metabolism
Rv3063	*cstA*	G472T	Gly158Cys	Carbon starvation protein A	Virulence, detoxification and adaptation
Rv3339c	*icd1*	G607A	Gly203Ser	Isocitrate dehydrogenase	Intermediary metabolism and respiration

Legend to Table 4.17:
Non-synonymous SNPs are highlighted in grey

Table 4.18: Indels found to be unique to the genome of R451

Locus	Gene	Position in gene	Indel	Change	WT ORF	MUT ORF	Description	Functional group
Rv0392c	*ndhA*	53	C (insertion)	Truncated protein	1-471	1-22	Membrane NADH dehydrogenase	Intermediary metabolism and respiration
Intergenic	-		GT (insertion)	-	-	-	-	-
Rv1427c	*fadD12*	1559	TG (deletion)	Frame shift	536	Extends to upstream gene	Long-chain fatty-acid-CoA ligase	Lipid metabolism

Legend for Table 4.18:
WT = wild type
MUT = mutant
ORF = open reading frame

To identify genetic variants which might result in decreased rifampicin resistance acquired *in vivo,* variants identified to be unique to the genome of R439 was also investigated (Table 4.19). Only four SNPs were identified to be unique to the genome of R439, from which only two were identified as non-synonymous. These two SNPs were identified in *Rv0284* and *far*, encoding for proteins involved in cell wall and cell processes and lipid metabolism respectively. Additionally two Indels were also unique to the genome of R439 (Table 4.20). These two Indels were in the *plsB2* and *plsC* genes which are involved in lipid metabolism.

Table 4.19: SNPs found to be unique to the genome of R439

Locus	Gene	Base change	Amino acid change	Gene function	Functional Group
Rv0284		T581G	Leu194Arg	Conserved membrane protein	Cell wall and cell processes
Rv0855	*far*	C802A	Arg286Ser	Fatty-acid-CoA racemase	Lipid metabolism
Rv2623	*TB31.7*	C504G	Gly168Gly	Conserved hypothetical protein	Virulence, detoxification and adaptation
Rv2940c	*mas*	C5979T	Val1993Val	Multifunctional mycocerosic acid synthase membrane-associated	Lipid metabolism

Legend to Table 4.19:
Non-synonymous SNPs are highlighted in grey

Table 4.20: Indels found to be unique to the genome of R439

Locus	Gene	Position in gene	Indel	Change	WT ORF	MUT ORF	Description	Functional group
Rv2482c	*plsB2*	1119	CT (deletion)	Truncated	1-790	1-384	Glycerol-3-phosphate acyltransferase	Lipid metabolism
Rv2483c	*plsC*	1494	A (deletion)	Frame shift	1-581	Extends to upstream gene	Bifunctional phospholipid biosynthesis enzyme	Lipid metabolism

Legend for Table 4.20:
WT = wild type
MUT = mutant
ORF = open reading frame

CHAPTER 5

DISCUSSION

5.1. Epidemiological relevance of compensatory mutations in *rpoA/C*

Several recent studies established the role of compensatory mutations in the genes *rpoA* and *rpoC* to alleviate the fitness-cost incurred by rifampicin resistance-conferring mutations in *rpoB*. However, as yet, little is known about the epidemiological relevance of compensatory evolution in drug-resistant *M. tuberculosis*. This study is the first to show that putative compensatory mutations in *rpoC* are prevalent among rifampicin-resistant isolates in a high-burden setting in South Africa, and strongly associated with transmission of rifampicin-resistant strains. Moreover, the presented data confirms the convergent evolution of specific putative compensatory *rpoC* mutations, which is consistent with positive selection of these mutations as shown previously (40). Taken together, our findings suggest that compensatory evolution in *M. tuberculosis* is of epidemiological relevance and contributes to the spread of drug resistance.

This study has several limitations. The sample analyzed consists of a convenience sample which does not accurately represent the overall population structure of *M. tuberculosis* in this setting. Thus, the true proportions of *M. tuberculosis* isolates with *rpoC* mutations may deviate from the proportions herein reported. However, these limitations do not affect the main conclusions of this study. Furthermore, defining transmission chains on the basis of identical IS*6110* RFLP patterns is not ideal. For example, two of the five isolates showing non-clustered RFLP types belonged to the R86 genotype, as suggested by other phylogenetic markers (Appendix, Table A3). Thus, rather than being rarely transmitted, these strains are representative of new RFLP variants, which emerged from a fast-spreading clone of *M. tuberculosis*, herein identified to be associated with *rpoC* mutations. Similarly, the remaining isolates with non-clustered RFLP patterns harbouring *rpoC* mutations also belonged to well-recognized strain families. Thus, this study probably overestimates the proportion of putative compensatory mutations among relatively infrequently transmitted strains.

Multivariate logistic regression analysis revealed a strong association between the presence of *rpoC* mutations and strain genotype (OR=44.3; $p<0.001$) as well as the specific rifampicin resistance mutation acquired (OR=92.0; $p<0.001$) (Table 4.1). The observation that the variability in terms of the presence of *rpoC* mutations is critically influenced by purely genetic properties, perhaps illustrates how the direction of compensatory evolution is shaped by epistatic interactions with the strain genetic background. Indeed, studies in several bacterial species have shown how genetic background and primary drug resistance mutations predetermine subsequently acquired mutations, favoured within a particular strain background (172, 185, 212, 213).

The association between the *rpoB* S531L and *rpoC* mutations could be explained by at least three models. Firstly, compensatory mutations in *rpoC* could act by restoring structural interactions between the β'- and β-subunit of the RNA polymerase, distorted after the acquisition of a resistance

mutation in *rpoB*. In this scenario, resistance mutations other than the *rpoB* S531L mutation could affect the structural properties of the RNA polymerase complex in different ways, such that mutations in *rpoC* are not able to restore the interaction between β'- and β-subunits. This could explain the relative underrepresentation of *rpoC* mutations in strains harbouring such mutations. In *Salmonella enterica* distinct mutations in *rpoC* have been shown to compensate for the *rpoB* R529C mutation (28). Secondly, the *rpoB* S531L mutation may allow for a wider range of compensatory mutations, including mutations in *rpoC*, while other rifampicin resistance mutations could require more specific adaptations outside the genetic regions herein investigated. This is supported by the variety of compensatory mutations detected among strains with an *rpoB* S531L mutation. Finally, *rpoB* S531L has been shown to cause a small fitness defect compared to other *rpoB* mutations (63). Hence, such a low fitness cost might be easier to compensate than other rifampicin resistance mutations. More work is required to differentiate between these various hypotheses.

5.2. Investigation of the expression of energy metabolism genes in response to rifampicin exposure

This is the first study to investigate the expression levels of energy metabolism genes in rifampicin resistant clinical isolates of *M. tuberculosis* in response to rifampicin exposure. While the majority of genes showed significant differential expression after rifampicin exposure, three genes (*atpH, atpA* and *fba*) were found to be up regulated irrespective of genetic background of the strain, degree of drug resistance or level rifampicin resistance (MIC). Interestingly, the expression levels of *Rv1303* and *Rv1312* were not responsive to rifampicin despite up regulation of other genes in the operon (*atpA* and *atpH*) This suggests either the existence of an internal promoter within the operon which independently regulates expression of the *atpBEFHAGDC* genes or that *Rv1303* and *Rv1312* RNA is rapidly degraded within the cell (short half-life).

Our observation of the absence of a change in the expression level of *Rv1846c* is counterintuitive given that previous studies have concluded that this gene encodes a protein which acts as a negative regulator of the *atpBEFHAGDC* operon. The absence of repression may be explained by an increase in the rate at which Rv1846c was cleaved thereby preventing binding to the promoter region of the operon. However, the expression level of the gene *Rv1255c* responsible for cleavage of Rv1846c was not tested in this study (Chip-seq data, www.tbdb.org). In this study we also showed that expression of *fba* was up regulated in response to rifampicin. *fba* encodes for the fructose-biphosphate aldolase, a key enzyme in glycolysis (169), suggesting an increased production of ATP through this pathway as well. It is predicted that an investigation of other genes involved in glycolysis might reveal whether a general response of glycolysis genes is necessary to modulate rifampicin resistance levels.

In this study we observed a clear distinction between the expression levels of the selected genes between the high and low rifampicin MIC strains. The overall increase in expression levels in the high MIC strains suggests that these strains have a higher energy requirement and thus have adapted their transcriptional response to increase energy production. Two genes, *gltA* and *opcA*, displayed significant up regulation in both high MIC MDR isolates, while no significant change observed for these genes in the low MIC MDR isolates. *GltA2* encodes for the citrate synthase enzymes, which is the pace-making enzyme of the Citric acid cycle, which involves a series of chemical reactions essential for energy production through aerobic respiration (83). *OpcA* encodes for a protein essential for the assembly of the glucose-6-phosphate dehydrogenase (203). This enzyme is the rate limiting enzyme in the pentose phosphate pathway, which is an alternative pathway to glycolysis for the production of energy from glucose-6-phospate (79). We hypothesise that the up regulation of both these rate limiting enzymes of the Citric acid cycle and pentose phosphate pathway in the high rifampicin MIC isolates would result in increased rate of the production of ATP.

The expression profiles between the MDR and mono-rifampicin resistant isolates were vastly different, with more genes being significantly up regulated in both mono-rifampicin resistant isolates. This may suggest that the energy requirements of *M. tuberculosis* adjust with the acquisition of additional drug resistance causing mutations. However, the effect of these mutations regarding energy metabolism remains unknown.

The overall observation of increased expression of energy metabolism genes supports our hypothesis that rifampicin resistant isolates, when exposed to rifampicin, have increased energy requirements. The recently reported up regulation of ATP dependent efflux and transporter genes in response to rifampicin in the same MDR isolates suggests that these metabolic processes are linked (114). These findings suggest that rifampicin resistant bacilli are responding to the presence of rifampicin by increasing efflux in an energy dependent manner and thereby maintaining the intracellular concentration of rifampicin below a toxic threshold. This may have important ramification for the TB treatment guidelines as we have shown that inappropriate treatment of MDR-TB with rifampicin induced cross-resistance to second-line anti-TB drugs (114).

The observation that certain environmental stimuli leads to the elevation of ATP dependent efflux activity provides a platform to propose a hypothesis to explain the increased rate in culture conversion of MDR-TB patients who were treated with a standardized second-line anti-TB drug regimen in combination with the promising new drug TMC207 relative to those only receiving the standardized regimen. TMC207 inhibits ATP synthesis by binding to the AtpE subunit of the mycobacterial F_1F_0 ATP synthase (14, 99). We propose that inhibition of ATP synthesis will deprive the efflux pumps and transporter genes of energy thereby effectively inhibiting the cells ability to regulate the

intracellular concentration of toxic compounds. Consequently, there will be an accumulation of second-line anti-TB drugs with the bacilli leading to more efficient binding of the drugs to their targets and cell death.

5.3. Selection and characterization of rifampicin resistant *in vitro* mutants

Several studies have investigated the frequency of mutations in the RRDR of *rpoB* in *in vitro* rifampicin resistant mutants and although the level of rifampicin resistance was investigated in these mutants, it was not clear what the contribution of the *rpoB* mutation to the level of rifampicin resistance (82, 120). Here we aimed to characterize the frequency of *rpoB* mutations in *in vitro* mutants selected from two distinct *M. tuberculosis* genetic backgrounds. Secondly, we also aimed to characterize the rifampicin resistance phenotype of these mutants by investigating the level of rifampicin resistance.

Previous studies have shown that the Ser531Leu and His526Tyr mutations in *rpoB* occur at the highest frequency among *in vitro* generated rifampicin resistant mutants. In addition, these studies have inferred that these two mutations are associated with the lowest fitness cost in *in vitro* resistant mutants (63) and therefore a possible correlation between the mutation frequency and fitness cost can be made. In this study mutations at amino acid position 526 were found to be overrepresented in both genetic backgrounds, with the frequency of the His526Asp mutant significantly different from what was observed in a previous study (82). Mutants selected from the H37Rv background showed also significant differences in frequency for three additional mutations (His526Arg, Ser531Leu and Ser522Trp), while mutants selected from the Beijing background only showed significant differences in frequency for one additional mutation (Ser522Trp). We therefore suggest that the laboratory conditions and strain genetic backgrounds used in this setting affected the type and frequency of mutants selected.

Mutations at amino acid position 526 in the *rpoB* also displayed higher levels of resistance above 200 µg/ml with less variability in both genetic backgrounds. In contrast, mutations at amino acid position 531 showed higher variability between 50 and 150 µg/ml in the Beijing background, whereas mutants selected from the H37Rv harbouring 531 mutations showed less variability, with rifampicin MICs remaining ranging between 100 and 200 µg/ml. These results suggest that the *rpoB* mutation alone does not account for the varying rifampicin levels detected in *in vitro* rifampicin resistant mutants.

5.4. Identification of genetic variants responsible for varying rifampicin resistance levels

This is the first study to investigate genomic variation and its association with varying rifampicin resistance levels in closely related *M. tuberculosis* isolates. *M. tuberculosis* strains can be divided into three principles genetic groups (PGG) based on polymorphisms found at position 463 in *katG* and position 95 in *gyrA*, with the Beijing lineage belonging to PGG 1, LCC to PGG 2 and H37Rv to PGG 3 (187). Our analysis of whole genome sequencing data from seven *M. tuberculosis* isolates, identified on average 1552 high confidence SNPs and 148 high confidence indels in Beijing isolates when compared to the reference genome, H37Rv. Similarly, the comparison between H37Rv and the LCC isolates identified an average of 1066 SNPs and 94 indels. A total number of 850 and 485 of all SNPs identified relative to H37Rv were found to be unique to the Beijing and LCC lineages, respectively. Additionally, 539 SNPs were unique genetic changes acquired in the genome of H37Rv. This highlights the importance knowing the evolutionary direction when using H37Rv as the assembly template. Furthermore, it must be noted that when using this method of assembly then certain genetic information will be lost which includes IS*6110* insertion positions and regions absent in H37Rv.

Analysis of the SNPs unique to the respective lineages showed that the highest number of SNPs (as a proportional of the number of genes classified to each functional category (http://tuberculist.epfl.ch/)), were localized in PE-PPE genes. These genes account for approximately 10% of the *M. tuberculosis* genome and are known to be highly polymorphic and hypothesized to be involved in antigenic variation (29, 97). More recently, numerous other functions have been assigned to this family of proteins including virulence and modulation of the host immune response as well as hydrolase activity (6, 136, 137, 200). In addition this analysis in this study showed that changes in genes involved in lipid metabolism, information pathways and cell wall and cellular processes play a predominant role in the evolutionary adaptation of this pathogen. This suggests that these cellular processes are under evolutionary pressure which may explain phenotypic differences observed between the clinical strains.

In order to identify genetic variants responsible for increasing rifampicin resistance in the *in vitro* resistant *M. tuberculosis* mutants, the genomes of two *in vitro* selected rifampicin resistant mutants and their progenitor pan susceptible Beijing isolate were compared. After excluding variants identified in repetitive and transposable regions, only two synonymous SNPs and one deletion were acquired in addition to the *rpoB* Ser531Leu mutation in the mutant genomes. The functional relevance of the two SNPs in *TB16*.3 and *malQ* remains unclear, as they do not alter the protein structures. The low frequency of the Val407Val SNP in *malQ* in the remaining set of the *in vitro* rifampicin resistant mutants (2/54 mutants) suggest that this would be a non-universal mechanism for increased rifampicin

resistance levels. In contrast the Gly101Gly SNP in *TB16*.3 was identified in 38% (20/54) which represented 70% (14/20) of the progenitors before rifampicin selection. Three scenarios could explain the occurrence of this SNP in multiple *in vitro* mutants harbouring different *rpoB* mutations and rifampicin resistance levels. Firstly, it is possible that this genetic event arose convergently in 14 of the 20 pre-rifampicin selected progenitors, which would suggest the importance of this SNP under the specific selection pressure. Secondly, it is also possible that this SNP was acquired randomly with a minimal fitness cost during the first culturing step of the Luria-Delbrück fluctuation assay. In the latter case, this SNP would not be associated with the evolution of rifampicin resistance. Thirdly, the possibility also exist that this variant was present in a proportion of the progenitor population.

Currently, we do not understand the effect of synonymous SNPs on the regulation of cellular processes and homeostasis of the cell. While synonymous SNPs are known to effect mRNA stability, mRNA processing, mRNA maturation and substrate specificity, thereby altering gene expression and protein function in mammalian cells, little is known about the effect of these SNPs in Mycobacteria (33, 52, 91, 96, 128). It is also suggested that SNPs could affect the binding sites of small regulatory RNAs, thereby affecting the gene expression of the target (64, 141). The limited number of variants identified in this study might suggest that increased rifampicin resistance *in vitro* might be due to non-genetic (epigenetic) regulatory mechanisms. However, it is acknowledged that the genome sequencing methodologies used in this study have certain limitations; a) mutations in regions absent in H37Rv would not be identified, b) repetitive regions (such as the PE-PPE protein families) are not fully represented, and c) the impact of transposon elements was not assessed. In particular, the gene sequences of the PE-PPE protein families have been shown with Sanger sequencing to be highly polymorphic (123, 214). There is increasing evidence that these proteins may modulate the structure of the cell wall and thereby may influence permeability towards certain drugs (46, 165).

This study showed that high genomic diversity exists between isolates previously termed "closely related" based on IS*6110* RFLP type and targeted DNA sequencing of genes conferring drug resistance. In total 25 variants differentiated the two MDR Beijing isolates, whereas 20 variants differentiated the MDR LCC isolates. A previous study showed through with whole genome sequencing that two clinical isolates sharing identical IS*6110* RFLP fingerprints are different at 131 separate loci (142). That study raised the question to whether standard typing measures are of sufficient discriminatory power to differentiate between strains, which might result in misinterpretation of the transmission dynamics of *M. tuberculosis* outbreaks. Thus the difference in the levels of rifampicin resistance in our study may reflect the accumulation of a number of different mutations, some of which will not have been detected by our analysis method.

In summary, the LCC genomes showed that 61.43% (10/14) of the variants (non-synonymous SNP and indels) were identified in genes involved in cell wall and cell processes and lipid metabolism. We therefore propose that structural and integrity differences in the cell wall caused by these variants might influence rifampicin resistance levels. A recent study have shown the presence of potential adapting mutations in genes related to cell wall biosynthesis in clinical drug resistant isolates (201). Structural changes in proteins involved in lipid metabolism might result in increased energy production through the catabolism of host lipids or increased production (anabolism) of lipids reincorporated into the lipid rich cell wall. Structural modulation of the cell wall might influence the permeability towards certain drugs, whereas increased production of ATP can be utilized by other cellular processes such as efflux systems to modulate levels of resistance. In contrast, the Beijing genomes showed less variability in these cellular processes (2/17, 11.77%) and instead showed higher variability in the genes conferring drug resistance or compensatory mechanisms, with 41.18% (7/17) of the variants (non-synonymous SNPs and Indels) identified in *rpoB, rpsL, ethA, pncA, rpoC* and *katG*. From these seven variants, three have been previously reported, with the SNP in *rpsL* known to confer streptomycin resistance (135). Non-synonymous SNPs in *rpoC* have also been described as a putative compensatory mechanism for rifampicin resistance (40). Mutations at position 564 in *rpoB* (as identified in R257) outside of the RRDR have also previously been reported, albeit with the amino acid change from proline to leucine (74). The clinical isolate harbouring this mutation also harboured a *rpoB* Glu562Gly mutation and displayed a rifampicin MIC below 1 µg/ml (74). It has also been shown in *S. enterica* that the acquisition of the *rpoB* Pro564Ser mutation in addition to the Arg526Cys mutation resulted in decreased rifampicin resistance (two fold) and increased growth fitness (measured as the doubling time) (28).

The findings from the MDR isolates are in conflict with previous results which showed that rifampicin stimulated the expression of efflux pump and transporter genes (114). However, no association between specific transporters/pumps and high rifampicin resistance levels could be made. Despite this, it would be envisaged that a genetic component for increased efflux expression would be present. Our failure to identify such a mechanism could be due to the limitations associated with the study (see above and below). An alternative explanation could be that the *rpoB* mutation affects the functionality of the RNA polymerase by changing the promoter specificity and in so doing alters the transcriptome. This notion is supported by the observation that rifampicin mono-resistant *M. tuberculosis* isolates harbouring mutations at codon 526 displayed delayed induction of efflux mechanisms after rifampicin exposure compared to isolates harbouring mutations at codon 531 (114).

From these results it is evident that the genetic component for increased rifampicin resistance is more complex than previously thought with no clear pathway employed by the bacteria to survive in higher concentrations of rifampicin. Furthermore, the effect of additional resistance conferring mutations (as

seen in the MDR isolates) on the mutations identified remains unknown. A further confounding factor is that whole genome sequencing of the MDR isolates were done from a representative scrape of the total population of bacterial cells which may include sub-populations with varying rifampicin resistance levels. Thus the MIC determination may not reflect the true population structure as only clones with the highest level of resistance within the culture will be measured. This implies that the genetic variants observed may not reflect the sub-population with the highest MIC as all genetic variants with confidence levels of <95% will have been excluded from the analysis. Therefore, we suggest that future studies using whole genome sequencing should be done on single colony forming units to exclude heterogeneity represented by the mixed bacterial population. In addition, the combined use of the four different strategies for the analysis of whole genome sequence data would minimize the identification of false positives. Furthermore, future studies should include functional analysis of the variants; however, experimental planning will be limited by the prediction of the number and order in which these variants were acquired by *M. tuberculosis*.

Finally, this study only investigated varying levels of rifampicin resistance in isolates harbouring the *rpoB* Ser531Leu mutation. The question remains whether similar modulatory mechanisms are used by isolates harbouring mutations at different loci in the RRDR, and why certain mutations (i.e. at position 516 and 533) remains associated with lower rifampicin resistance levels.

CHAPTER 6

CONCLUSION

Despite the availability of effective anti-TB drugs, TB still remains an urgent global health problem (236). This is in part due to the emergence of spontaneous drug resistance by *M. tuberculosis* which is driven by several factors such as delayed diagnosis, the use of low quality drugs, malabsorption and inappropriate treatment (69, 93). While the acquisition of drug resistance is usually associated with a fitness cost, *M tuberculosis* has displayed the ability to adapt its genome to compensate for this loss (62). Therefore, the complexity of rifampicin resistance extends beyond the acquisition of mutations in *rpoB* and previous studies have suggested that mutations in *rpoC* ameliorate the fitness cost inferred by the *rpoB* mutations. This is the first study to investigate the prevalence of putative compensatory *rpoC* mutations in drug resistant clinical isolates from high TB burden settings in South Africa. This study also aimed to investigate whether mutations in *rpoC* has an epidemiological impact on the transmission of drug resistant *M. tuberculosis* clinical isolates. Results from this study demonstrated that mutations in *rpoC* are strongly associated with transmission of rifampicin-resistant strains in this setting. We therefore postulate that *rpoC* mutations facilitate the restoration of fitness in some clinical strains of drug-resistant *M. tuberculosis,* thereby enhancing their ability to spread. The fact that not all successful MDR/XDR strains carry such mutations suggests that alternative mechanisms of fitness compensation might exist. The observation of an association between strain genotype as well as the *rpoB* S531L mutation with *rpoC* mutations supports a role for mutation-specific epistatic effects in driving the compensatory events described in this study.

The complexity of rifampicin resistance is further inferred by *M. tuberculosis* displaying varying levels of rifampicin resistance in clinical isolates with identical *rpoB* mutations (82, 114). This suggests that other mechanisms in addition to *rpoB* mutations are responsible for the modulation of rifampicin resistance. Recently it was also demonstrated that the level of resistance can be lowered by the inclusion of efflux pump inhibitors, suggesting efflux and transporter mechanisms as strong candidates for the modulation of the intra-cellular concentration of rifampicin (114). A previous study done in our group also showed increased expression of proteins involved in the production of energy after rifampicin exposure (M Bester, MSc Thesis, 2009). To further investigate these findings, the expression profiles of genes involved in different pathways of ATP production were studied in *M. tuberculosis* isolates from different genetic backgrounds which displayed different level of rifampicin resistance. These results suggest that the energy requirements of *M. tuberculosis* adapts with the acquisition of additional drug resistance, with more energy metabolism genes being up-regulated in the mono-rifampicin resistant isolates when compared to MDR isolates. Secondly, MDR isolates with high rifampicin MICs also displayed an overall increase in the expression of energy metabolism genes when compared to the MDR isolates displaying low rifampicin MICs. This suggests that isolates with high rifampicin MICs have a higher energy requirement and therefore have adapted their transcriptional response to increase energy production. These findings together with the observation of rifampicin induced efflux activity let to the development of a model to explain the mechanism(s)

whereby *M. tuberculosis* regulates the intra-cellular concentration of rifampicin. We propose that increased energy produced upon rifampicin exposure will fuel efflux systems, resulting in decreased intra-cellular concentration of rifampicin. In contrast, inhibition of ATP synthesis (by TMC207) will deplete efflux and transporter systems from their energy requirements resulting in the accumulation of anti-TB drugs within the bacilli. The high intracellular concentration will lead to more efficient binding of the drugs to their targets leading to cell death.

To further investigate mechanisms responsible for defining the level of rifampicin resistance, *in vitro* rifampicin resistant mutants which displayed different levels of rifampicin resistance (despite having identical *rpoB* mutations) were generated and sequenced. Analysis of these clones showed that the level of rifampicin resistance varied significantly amongst clones that shared identical *rpoB* mutations. Whole genome sequencing and bioinformatics analysis of two *in vitro* Ser531Leu mutants with varying rifampicin resistance levels showed that minor genetic changes are acquired with increased rifampicin resistance.

In a separate experiment we showed using whole genome sequencing that MDR isolates presumed to be identical (shared identical RFLP fingerprints and drug resistance conferring mutations) were genetically distinct. This questions the use of RFLP methods to define epidemiological relationships between *M. tuberculosis* isolates (142). Furthermore, whole genome sequencing results suggest that strains from different clades employs distinct strategies to modulate rifampicin resistance levels.

Although whole genome sequencing has proved to be a useful technique in unravelling the evolution of drug resistance (40, 142), this study also highlighted the importance of meticulous study design as we were not able to clearly define mechanisms beyond cell wall and lipid metabolism. Despite this limitation, this is a general theme seen in other studies investigating the physiology of drug resistance and therefore we believe that the structure of the cell wall plays a critical role in the transport of molecules in and out of the cell. To identify more detailed mechanisms, it would be essential to redesign our study to analyse single clones. Preliminary data from our group suggests that clinical isolates reflect a spectrum of different phenotypes with respect to the level of rifampicin resistance. In that ongoing study it has been shown that analysis of single colony forming units allowed the identification of clones with definitive levels of resistance. Analysis of these clones with whole genome sequencing is expected to provide novel insight into the mechanisms of defining the level of resistance without the complicating factors of mixed populations, which would not be detected by current whole genome sequencing techniques. This is in line as proposed by Ford and colleagues that whole genome sequencing studies need to be cleverly designed to answer complex questions (59).

In summary this was the first study to investigate the role of compensatory *rpoC* mutations in an epidemiological context. This study highlights the importance of compensatory mechanisms in the spread of *M. tuberculosis* isolates. This was the first study to sequence the whole genomes of clinical and *in vitro* rifampicin resistant *M. tuberculosis* isolates displaying varying levels of rifampicin resistance. Together with our findings of increased expression of energy metabolism genes after rifampicin exposure, these results indicates that rifampicin resistance is more complex than previously envisaged as *M. tuberculosis* uses both genetic and regulatory mechanisms to define the level of rifampicin resistance. The influence of genetic variants on the transcriptome (and vice versa) is however largely unknown. Understanding the physiology of *M. tuberculosis* in response to rifampicin exposure may aid in the development of drugs to improve the efficacy the current anti-TB drugs, such as efflux and ATP (energy metabolism) inhibitors, which may shorten and improve treatment outcomes.

CHAPTER 7

FUTURE STUDIES

Experimental Work:

1. Determine if the ATP synthase operon are encoded by a single polycistronic DNA. This could be done by PCR amplification of intragenic regions.
2. Determine if there are any internal promoters inside the coding region. This could be done by cloning the intragenic region upstream of a reporter gene and then measuring expression of the reporter in *M. smegmatis*.
3. Whole genome sequencing analysis from single colonies displaying variable rifampicin resistance phenotypes. This will be done single colony forming units reflecting the extreme MIC's from 5 isolates which will be selected based on: a) isolates having the same *rpoB* mutation but different genetic backgrounds, and b) isolates having different *rpoB* mutations.
4. Whole genome sequencing analysis of the MDR isolates investigated in this study will be repeated by including a closely related susceptible "progenitor" isolate, as well as using different reference genomes for alignment and mapping.
5. Identify genetic variants resulting in increased resistance in isolates harbouring *rpoB* mutations which are regarded to display lower rifampicin resistance levels. For this single colonies will be selected from clinical isolates displaying lower rifampicin resistance levels (MIC $\leq$ 10 μg/ml) using the Luria Delbrück Fluctuation assay and a rifampicin selection concentration in increments of 10x above the isolates' rifampicin MIC. Whole genome sequencing and RNA sequencing will be done to identify mechanisms involved in increased rifampicin resistance.

Publications:

1. **Title:** Putative compensatory mutations in the *rpoC* gene of rifampicin-resistant *Mycobacterium tuberculosis* are associated with ongoing transmission.
   Authors: de Vos M, Müller B, Black PA, Borrell S, van Helden PD, Warren RM, Gagneux S, Victor TC
   Journal: Antimicrobial Agents and Chemotherapy
   Status: Accepted 3 December 2012

2. **Title:** Increased rifampicin resistance requires up-regulation of energy metabolism.
   Authors: de Vos M, Louw GE, Black PA, Grobbelaar M, Gey van Pittius NC, van Helden PD, Victor TC, Warren RM
   Status: Manuscript in advance stage of preparation

3. **Title**: Drug resistance beyond classical mutations in Mycobacteria.
   Authors: de Vos M, Louw GE, Williams MJ, Gey van Pittius NC, van Helden PD, Warren RM, Victor TC
   Status: Manuscript in advance stage of preparation

APPENDIX

BUFFERS AND SOLUTIONS

Culture media:

ADC mycobacterial liquid culturing media supplement:

25 g Bovine serium albumin (BSA) Fraction V; 10 g glucose; 0.75 ml catalase. The supplement was prepared in ddH_2O (500 ml), sterilised through a 0.22µm filtration unit and stored at 4°C.

Middlebrook 7H9 liquid medium:

4.7 g 7H9 medium; 2 ml glycerol; 0.5 ml polysorbate (Tween80). The broth was prepared in ddH_2O (900 ml) and autoclaved for 15 min at 121°C.

Middlebrook 7H10 solid medium:

19 g 7H10 medium; 5 ml glycerol. The solid media was prepared in ddH_2O (900 ml) and autoclaved for 15 min at 121°C.

Protein extraction:

Extraction buffer:

5% sodium glutamate, 50 mM Tris-HCl (pH 7.4), 25 mM EDTA

Proteinase K:

5% sodium dodecyl sulphate, 100nM Tris-HCl (pH 7.8), 50 mM EDTA

TE buffer:

10 mM Tris-HCl (pH 8.0), 1mM EDTA

Drug concentrations:

Rifampicin:

80 mg Rifampicin (Sigma); 10 % Dimethyl sulfoxide (DMSO)

Gel electrophoresis:

10 X TBE Buffer (pH 8.3):

108 g 0.45M Tris; 55 g 0.44M boric acid; 7.4 g 10mM EDTA in 1 L dH_2O

Loading dye:

0.25 % xylene cyonol; 50 % glycerol

TABLES AND FIGURES

Table A1: Beijing *in vitro* mutants' genotypes for *rpoB*, *malQ* and *TB16*.3

Progenitor	Mutant name	*rpoB*	*malQ*	*TB16.3*
1	**Pre-Rif culture**	**WT**	**WT**	**Gly101Gly**
	K636.18	His526Asp	WT	WT
	K636.22	His526Asp	WT	Gly101Gly
	K636.23	Ser531Trp	WT	WT
2	**Pre-Rif culture**	**WT**	**WT**	**WT**
	K636.21	His526Tyr	WT	WT
	K636.26	Ser522Leu	WT	WT
12	**Pre-Rif culture**	**WT**	**WT**	**Gly101Gly**
	K636.13	Ser522Trp	WT	Gly101Gly
	K636.54	H526R	WT	Gly101Gly
21	**Pre-Rif culture**	**WT**	**WT**	**WT**
	K636.30	Ser522Trp	WT	WT
	K636.38	His526Arg	WT	WT
25	**Pre-Rif culture**	**WT**	**WT**	**WT**
	K636.4	His526Arg	WT	WT
	K636.5	His526Arg	WT	WT
	K636.6	His526Arg	WT	WT
	K636.7	His526Tyr	WT	WT
	K636.8	His526His/Pro/Asp/Tyr	WT	WT
	K636.9	Ser522Trp	WT	WT
	K636.24	His526Tyr	WT	WT
	K636.43	Ser531Trp	WT	WT
26	**Pre-Rif culture**	**WT**	**WT**	**Gly101Gly**
	K636.11	WT	WT	WT
	K636.12	His526Asp	WT	WT
	K636.46	His526Asp	WT	WT
	K636.52	His526Asp	WT	WT
	K636.53	His526Asp	WT	WT
27	**Pre-Rif culture**	**WT**	**WT**	**WT**
	K636.15	His526Asp	WT	WT
	K636.20	His526Asp	WT	WT
28	**Pre-Rif culture**	**WT**	**WT**	**Gly101Gly**
	K636.42	His526Asp	WT	WT
29	**Pre-Rif culture**	**WT**	**WT**	**Gly101Gly**
	K636.17	His526Tyr	WT	WT
	K636.28	His526Arg	WT	WT
	K636.29	His526Arg	WT	WT
	K636.51	His526His/Arg	WT	WT
30	**Pre-Rif culture**	**WT**	**WT**	**Gly101Gly +WT**
	K636.31	His526Asp	WT	WT

33	**Pre-Rif culture**	**WT**	**WT**	**Gly101Gly**
	K636.25	His526Asp	WT	Gly101Gly
36	**Pre-Rif culture**	**WT**	**WT**	**WT**
	K636.1 (N36S1)	Ser531Leu	WT	Gly101Gly
37	**Pre-Rif culture**	**WT**	**WT**	**Gly101Gly +WT**
	K636.2 (N37S1)	Ser531Leu	Val407Val	Gly101Gly
	K636.10	WT	WT	Gly101Gly
	K636.44	Ser531Leu	Val407Val	Gly101Gly
	K636.47	His526Asp		Gly101Gly
38	**Pre-Rif culture**	**WT**	**WT**	**Gly101Gly**
	K636.45	Ser531Leu	WT	Gly101Gly
39	**Pre-Rif culture**	**WT**	**WT**	**Gly101Gly +WT**
	K636.16	His526Tyr	WT	Gly101Gly
	K636.40	His526Tyr	WT	Gly101Gly
	K636.50	His526Asp	WT	WT
40	**Pre-Rif culture**	**WT**	**WT**	**WT**
	K636.49	Ser531Leu	WT	Gly101Gly
41	**Pre-Rif culture**	**WT**	**WT**	**Gly101Gly**
	K636.32	His526Asp	WT	WT
	K636.39	Ser522Leu	WT	Gly101Gly
42	**Pre-Rif culture**	**WT**	**WT**	**Gly101Gly +WT**
	K636.3	Ser531Leu	WT	Gly101Gly
	K636.19	His526Arg	WT	Gly101Gly
	K636.37	His526Asp + Ser531Trp	WT	Gly101Gly +WT
	K636.48	Ser531Leu	WT	Gly101Gly
43	**Pre-Rif culture**	**WT**	**WT**	**WT**
	K636.33	Ser531Trp	WT	Gly101Gly
	K636.41	His526Asp	WT	Gly101Gly
44	**Pre-Rif culture**	**WT**	**WT**	**Gly101Gly +WT**
	K636.14	Ser522Trp	WT	WT
	K636.27	His526Tyr	WT	WT
	K636.34	His526Arg+ Ser531Trp	WT	WT
	K636.35	His526Tyr	WT	WT
	K636.36	His526Asp	WT	WT

Table A2: H37Rv *in vitro* mutants' genotype for *rpoB*

Clone name	*rpoB* mutation	Clone name	*rpoB* mutation	Clone name	*rpoB* mutation	Clone name	*rpoB* mutation
1	His526Tyr	26	Ser522Leu	51	His526Arg	76	His526Tyr
2	Ser522Trp	27	His526Tyr	52	Ser531Leu	77	His526Tyr
3	His526Tyr	28	His526Arg	53	Ser522Trp	78	His526Asp
4	His526Asp	29	Ser522Trp	54	Ser531Leu	79	His526Arg
5	His526Tyr	30	Ser531Leu	55	His526Tyr	80	His526Tyr
6	His526Tyr	31	His526Arg	56	His526Asp	81	Ser531Trp
7	His526Tyr	32	His526Asp	57	Ser522Leu	82	His526Arg
8	His526Tyr	34	His526Arg	58	His526Arg	83	His526Arg
9	Ser522Trp	35	His526Asp	59	Ser522Leu	84	His526Arg
10	His526Tyr	36	His526Arg	60	Ser531Leu	85	His526Arg
11	Ser531Trp	37	Ser531Trp	62	His526Asp	86	His526Arg
12	His526Tyr	38	His526Asp	63	His526Arg	87	His526Tyr
13	Ser522Leu	39	Ser522Trp	64	His526Asp	88	Ser522Trp
14	Asp516Val	40	Ser531Leu	65	His526Asp	89	Ser522Trp
16	His526Arg	41	His526Arg	66	Ser531Leu	90	His526Arg
17	Ser531Leu	42	Ser522Trp	67	Ser531Leu	91	His526Tyr
18	Ser522Leu	43	Ser531Trp	68	His526Tyr	92	His526Tyr
19	His526Arg	44	His526Tyr	69	His526Asp	93	His526Arg
20	His526Asp	45	His526Arg	70	His526Arg	94	His526Tyr
21	His526Tyr	46	His526Tyr	71	His526Arg	95	His526Tyr
22	His526Asp	47	His526Tyr	72	Ser522Trp	96	His526Pro
23	His526Arg	48	His526Tyr	73	His526Asp	97	His526Asp
24	Ser522Trp	49	His526Tyr	74	His526Tyr		
25	His526Asp	50	His526Asp	75	Ser531Leu		

Table A3: Distribution of SNPs in *rpoC* in clustered and unique RFLP types

RFLP clustering	*rpoB* mutation	*rpoC* mutation	Lineage	Atypical Beijing	Drug resistance group	Frequency
Clustered	531TTG	D485Y	BEIJING	Non-atypical Beijing	MDR	2
Clustered	531TTG	G332R	LCC	Non-atypical Beijing	MDR	1
Clustered	531TTG	I491T	BEIJING	Non-atypical Beijing	Rifampicin mono-resistant	1
Clustered	531TTG	I491T + A542A	LAM	Non-atypical Beijing	Pre-XDR	1
Clustered	531TTG	V483A	BEIJING	Atypical Beijing	Pre-XDR	1
Clustered	531TTG	V483G	BEIJING	Atypical Beijing	XDR	18
Clustered	531TTG	V483G	BEIJING	Atypical Beijing	Pre-XDR	5
Clustered	531TTG	V483G	BEIJING	Atypical Beijing	MDR	4
Clustered	531TTG	V483G	BEIJING	Non-atypical Beijing	MDR	1
Clustered	531TTG	V483G	LCC	Non-atypical Beijing	Rifampicin mono-resistant	1
Clustered	531TTG	V483G	T	Non-atypical Beijing	XDR	1
Clustered	531TTG	V483G	T	Non-atypical Beijing	MDR	1
Clustered	531TTG	A542A	LAM	Non-atypical Beijing	MDR	10
Clustered	531TTG	A542A	LAM	Non-atypical Beijing	Pre-XDR	2
Clustered	531TTG	A542A	LAM	Non-atypical Beijing	Rifampicin mono-resistant	1
Clustered	531TTG	A542A	LAM	Non-atypical Beijing	XDR	1
Clustered	531TTG	A542A	LCC	Non-atypical Beijing	MDR	1
Clustered	531TTG	Wild-type	BEIJING	Non-atypical Beijing	MDR	12
Clustered	531TTG	Wild-type	LCC	Non-atypical Beijing	XDR	6
Clustered	531TTG	Wild-type	BEIJING	Non-atypical Beijing	Pre-XDR	5
Clustered	531TTG	Wild-type	LCC	Non-atypical Beijing	MDR	5
Clustered	531TTG	Wild-type	BEIJING	Atypical Beijing	MDR	3
Clustered	531TTG	Wild-type	BEIJING	Atypical Beijing	XDR	1
Clustered	531TTG	Wild-type	BEIJING	Non-atypical Beijing	XDR	1
Clustered	531TTG	Wild-type	EAI	Unknown	Rifampicin mono-resistant	1
Clustered	531TTG	Unknown	LCC	Non-atypical Beijing	XDR	2
Clustered	511CCG	V483G	BEIJING	Atypical Beijing	MDR	1
Clustered	511CCG	Wild-type	LCC	Non-atypical Beijing	MDR	1
Clustered	516GTC	V483G	BEIJING	Atypical Beijing	Pre-XDR	1
Clustered	516GTC	Wild-type	BEIJING	Atypical Beijing	XDR	10
Clustered	516GTC	Wild-type	BEIJING	Atypical Beijing	Pre-XDR	4
Clustered	516GTC	Wild-type	BEIJING	Atypical Beijing	MDR	2
Clustered	516GTC	Wild-type	LCC	Non-atypical Beijing	XDR	1
Clustered	516TAC	Wild-type	BEIJING	Atypical Beijing	Pre-XDR	1
Clustered	516TAC	Wild-type	LCC	Non-atypical Beijing	Pre-XDR	1
Clustered	516TAC	Wild-type	LCC	Non-atypical Beijing	MDR	1

RFLP clustering	*rpoB* mutation	*rpoC* mutation	Lineage	Atypical Beijing	Drug resistance group	Frequency
Clustered	526CGC	Wild-type	BEIJING	Non-atypical Beijing	MDR	1
Clustered	526GAC	A542A	LAM	Non-atypical Beijing	Rifampicin mono-resistant	1
Clustered	526GAC	Wild-type	BEIJING	Non-atypical Beijing	Pre-XDR	1
Clustered	526TAC	Wild-type	BEIJING	Atypical Beijing	Rifampicin mono-resistant	1
Clustered	526TAC	Wild-type	BEIJING	Atypical Beijing	MDR	1
Clustered	526TAC	Wild-type	BEIJING	Non-atypical Beijing	MDR	1
Clustered	526TAC	Wild-type	BEIJING	Non-atypical Beijing	Pre-XDR	1
Clustered	526TAC	Wild-type	LCC	Non-atypical Beijing	MDR	1
Clustered	531CAG	Wild-type	BEIJING	Non-atypical Beijing	MDR	1
Clustered	531TTC	A542A	LAM	Non-atypical Beijing	MDR	1
Clustered	533CCG	Wild-type	BEIJING	Non-atypical Beijing	Pre-XDR	1
Clustered	533CCG	Wild-type	BEIJING	Non-atypical Beijing	MDR	1
Clustered	Wild-type	A542A	LAM	Non-atypical Beijing	Isoniazid mono-resistant	3
Clustered	Wild-type	A542A	LAM	Non-atypical Beijing	XDR	2
Clustered	Wild-type	A542A	LAM	Non-atypical Beijing	Rifampicin mono-resistant	2
Clustered	Wild-type	Wild-type	BEIJING	Non-atypical Beijing	Isoniazid mono-resistant	7
Clustered	Wild-type	Wild-type	LCC	Non-atypical Beijing	Poly-resistant	6
Clustered	Wild-type	Wild-type	BEIJING	Atypical Beijing	Drug-sensitive	5
Clustered	Wild-type	Wild-type	LCC	Non-atypical Beijing	Drug-sensitive	5
Clustered	Wild-type	Wild-type	BEIJING	Non-atypical Beijing	Rifampicin mono-resistant	4
Clustered	Wild-type	Wild-type	BEIJING	Non-atypical Beijing	Poly-resistant	1
Clustered	Wild-type	Wild-type	BEIJING	Unknown	Pre-XDR	1
Clustered	Wild-type	Wild-type	LAM	Non-atypical Beijing	Poly-resistant	1
Clustered	531TTG	D485Y	BEIJING	Non-atypical Beijing	Pre-XDR	1
Clustered	531TTG	D485Y	BEIJING	Non-atypical Beijing	MDR	1
Clustered	531TTG	F452C	LCC	Non-atypical Beijing	Pre-XDR	1
Clustered	531TTG	H525Q	LCC	Non-atypical Beijing	MDR	1
Clustered	531TTG	V483G	T	Non-atypical Beijing	Pre-XDR	6
Clustered	531TTG	V483G	T	Non-atypical Beijing	MDR	1
Clustered	Unknown/mix	A542A	LAM	Non-atypical Beijing	Drug-sensitive	7
Clustered	Unknown/mix	A542A	LAM	Non-atypical Beijing	MDR	3
Clustered	Unknown/mix	A542A	LAM	Non-atypical Beijing	Pre-XDR	2
Clustered	Unknown/mix	Wild-type	LCC	Non-atypical Beijing	Drug-sensitive	24

RFLP clustering	*rpoB* mutation	*rpoC* mutation	Lineage	Atypical Beijing	Drug resistance group	Frequency
Clustered	Unknown/mix	Wild-type	BEIJING	Non-atypical Beijing	Drug-sensitive	19
Clustered	Unknown/mix	Wild-type	LCC	Non-atypical Beijing	MDR	8
Clustered	Unknown/mix	Wild-type	BEIJING	Non-atypical Beijing	MDR	3
Clustered	Unknown/mix	Wild-type	LCC	Non-atypical Beijing	Pre-XDR	2
Clustered	Unknown/mix	Wild-type	BEIJING	Non-atypical Beijing	Pre-XDR	1
Clustered	Unknown/mix	Wild-type	BEIJING	Unknown	Isoniazid mono-resistant	1
Clustered	Unknown/mix	Wild-type	BEIJING	Unknown	MDR	1
Clustered	Unknown/mix	Wild-type	LCC	Non-atypical Beijing	Poly-resistant	1
Clustered	Unknown/mix	Wild-type	LCC	Non-atypical Beijing	Rifampicin mono-resistant	1
Clustered	Unknown/mix	Wild-type	unknown	Non-atypical Beijing	Drug-sensitive	1
Clustered	Unknown/mix	Wild-type	unknown	Non-atypical Beijing	Pre-XDR	1
Clustered	Unknown/mix	Wild-type	unknown	Non-atypical Beijing	Isoniazid mono-resistant	1
Non-clustered	531TTG	I491T + A542A	LAM	Non-atypical Beijing	Pre-XDR	1
Non-clustered	531TTG	V483G	BEIJING	Atypical Beijing	MDR	1
Non-clustered	531TTG	A542A	LAM	Non-atypical Beijing	MDR	3
Non-clustered	531TTG	A542A	LAM	Non-atypical Beijing	Rifampicin mono-resistant	3
Non-clustered	531TTG	A542A	LAM	Non-atypical Beijing	Pre-XDR	2
Non-clustered	531TTG	Wild-type	BEIJING	Non-atypical Beijing	MDR	3
Non-clustered	531TTG	Wild-type	HAARLEM	Non-atypical Beijing	XDR	1
Non-clustered	531TTG	Wild-type	T	Non-atypical Beijing	Rifampicin mono-resistant	1
Non-clustered	531TTG	Wild-type	T	Non-atypical Beijing	MDR	1
Non-clustered	531TTG	Wild-type	unknown	Non-atypical Beijing	Rifampicin mono-resistant	1
Non-clustered	511CCG	Wild-type	unknown	Non-atypical Beijing	MDR	1
Non-clustered	516TAC	Wild-type	BEIJING	Atypical Beijing	XDR	1
Non-clustered	526CGC	Wild-type	unknown	Non-atypical Beijing	MDR	1
Non-clustered	526TAC	Wild-type	BEIJING	Non-atypical Beijing	Rifampicin mono-resistant	1
Non-clustered	526TAC	Wild-type	HAARLEM	Non-atypical Beijing	XDR	1
Non-clustered	533CCG	Wild-type	BEIJING	Non-atypical Beijing	Rifampicin mono-resistant	1
Non-clustered	Wild-type	A542A	LAM	Non-atypical Beijing	Isoniazid mono-resistant	5
Non-clustered	Wild-type	A542A	LAM	Non-atypical Beijing	Rifampicin mono-resistant	1
Non-clustered	Wild-type	A542A	unknown	Non-atypical Beijing	Isoniazid mono-resistant	1
Non-clustered	Wild-type	Wild-type	LCC	Non-atypical Beijing	Isoniazid mono-resistant	4
Non-clustered	Wild-type	Wild-type	BEIJING	Non-atypical Beijing	Rifampicin mono-resistant	2

RFLP clustering	*rpoB* mutation	*rpoC* mutation	Lineage	Atypical Beijing	Drug resistance group	Frequency
Non-clustered	Wild-type	Wild-type	BEIJING	Non-atypical Beijing	Isoniazid mono-resistant	1
Non-clustered	Wild-type	Wild-type	LCC	Unknown	Isoniazid mono-resistant	1
Non-clustered	531TTG	Q523E	F28	Non-atypical Beijing	MDR	1
Non-clustered	531TTG	V483G	BEIJING	Atypical Beijing	Pre-XDR	1
Non-clustered	531TTG	V483G	T	Non-atypical Beijing	XDR	1
Non-clustered	Unknown/mix	A542A	LAM	Non-atypical Beijing	Drug-sensitive	4
Non-clustered	Unknown/mix	A542A	LAM	Non-atypical Beijing	MDR	3
Non-clustered	Unknown/mix	A542A	F26	Non-atypical Beijing	MDR	1
Non-clustered	Unknown/mix	A542A	unknown	Non-atypical Beijing	MDR	1
Non-clustered	Unknown/mix	Wild-type	BEIJING	Non-atypical Beijing	Drug-sensitive	2
Non-clustered	Unknown/mix	Wild-type	unknown	Non-atypical Beijing	MDR	2
Non-clustered	Unknown/mix	Wild-type	BEIJING	Atypical Beijing	XDR	1
Non-clustered	Unknown/mix	Wild-type	F28	Non-atypical Beijing	MDR	1
Non-clustered	Unknown/mix	Wild-type	T	Non-atypical Beijing	MDR	1
Non-clustered	Unknown/mix	Wild-type	unknown	Non-atypical Beijing	Pre-XDR	1
Unknown	531TTG	V483G	BEIJING	Atypical Beijing	Pre-XDR	1
Unknown	531TTG	V483G	T	Non-atypical Beijing	MDR	1
Unknown	531TTG	A542A	LAM	Non-atypical Beijing	XDR	1
Unknown	531TTG	A542A	unknown	Non-atypical Beijing	MDR	1
Unknown	Wild-type	A542A	LAM	Non-atypical Beijing	Rifampicin mono-resistant	8
Unknown	Wild-type	Wild-type	BEIJING	Unknown	Rifampicin mono-resistant	6
Unknown	Wild-type	Wild-type	T	Non-atypical Beijing	Isoniazid mono-resistant	3
Unknown	Wild-type	Wild-type	T	Non-atypical Beijing	Rifampicin mono-resistant	1
Unknown	Wild-type	Wild-type	unknown	Non-atypical Beijing	Isoniazid mono-resistant	1
Unknown	Wild-type	Wild-type	unknown	Non-atypical Beijing	Pre-XDR	1
Unknown	Wild-type	Wild-type	unknown	Unknown	Rifampicin mono-resistant	1
Unknown	531TTG	G332R + A542A	BEIJING	Unknown	MDR	1
Unknown	531TTG	V483G	LAM	Non-atypical Beijing	Pre-XDR	1
Unknown	531TTG	V483G	LCC	Non-atypical Beijing	MDR	1
Unknown	Unknown/mix	A542A	unknown	Non-atypical Beijing	MDR	2
Unknown	Unknown/mix	Wild-type	T	Non-atypical Beijing	Rifampicin mono-resistant	13
Unknown	Unknown/mix	Wild-type	T	Non-atypical Beijing	Isoniazid mono-resistant	6
Unknown	Unknown/mix	Wild-type	unknown	Non-atypical Beijing	MDR	3

RFLP clustering	*rpoB* mutation	*rpoC* mutation	Lineage	Atypical Beijing	Drug resistance group	Frequency
Unknown	Unknown/mix	Wild-type	BEIJING	Unknown	Rifampicin mono-resistant	2
Unknown	Unknown/mix	Wild-type	BEIJING	Unknown	MDR	1
Unknown	Unknown/mix	Wild-type	BEIJING	Unknown	Isoniazid mono-resistant	1
Unknown	Unknown/mix	Wild-type	CAS	Non-atypical Beijing	Pre-XDR	1
Unknown	Unknown/mix	Wild-type	LCC	Non-atypical Beijing	MDR	1
Unknown	Unknown/mix	Wild-type	unknown	Non-atypical Beijing	Pre-XDR	1
Unknown	Unknown/mix	Wild-type	unknown	Non-atypical Beijing	Rifampicin mono-resistant	1

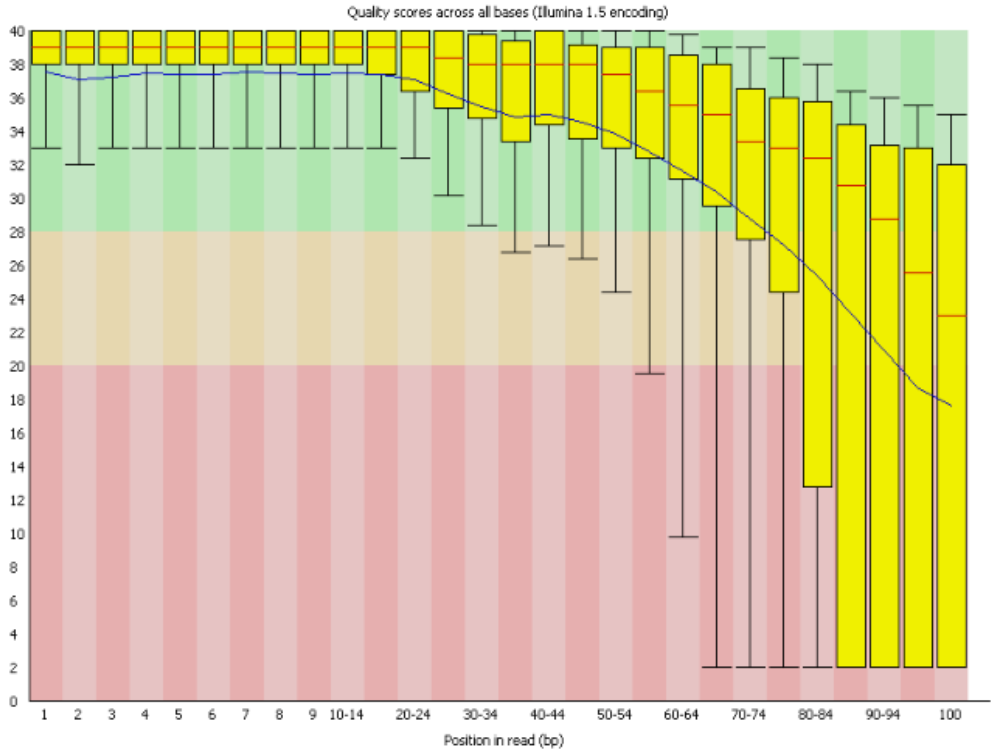

Figure A1: BoxWhisper plot for quality scores across all bases

Legend to Figure A1:

The blue line within the yellow boxes represents the median quality value for the respective bases. This graph indicates good quality calls for the first 75 base pairs, where it starts to drop toward the end of the read. This is normal as the quality of calls will degrade as the run progress.

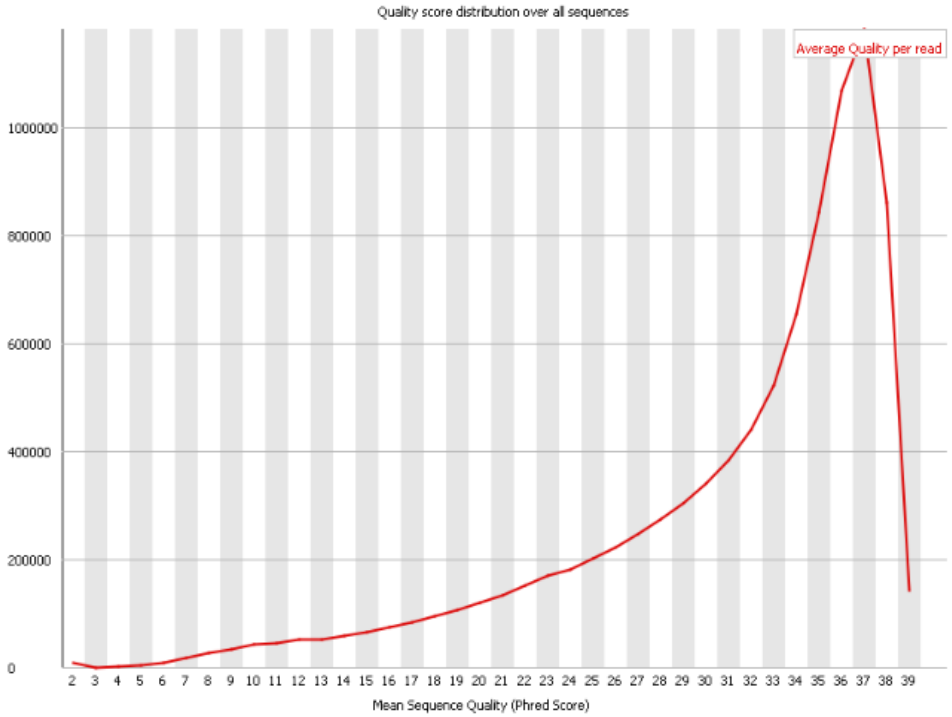

Figure A2: Distrubtion of quality scores across all sequences in the library
Legend to Figure A2:
The mean quality (Phred) score is calculated across all bases in the library. Only a small subset of sequences displayed a mean quality score below 24, with the majority of sequences diplaying good quality values

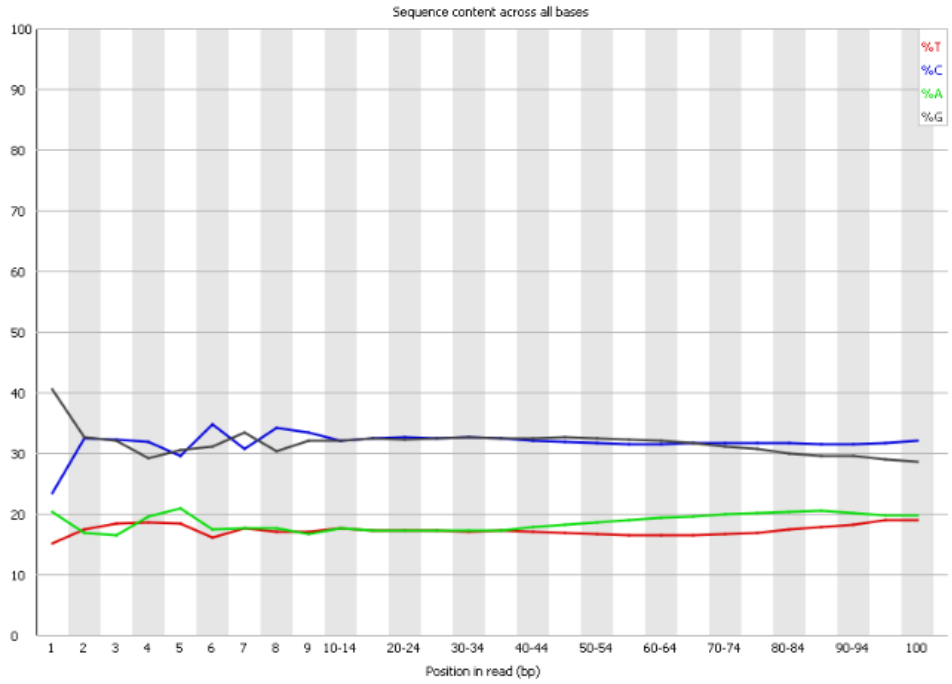

Figure A3: Sequence content across all bases in the library
Legend to Figure A3:
This figure indicates higher proportion of G and C bases in this library, which is expected of a *M. tuberculosis* genome with high GC content

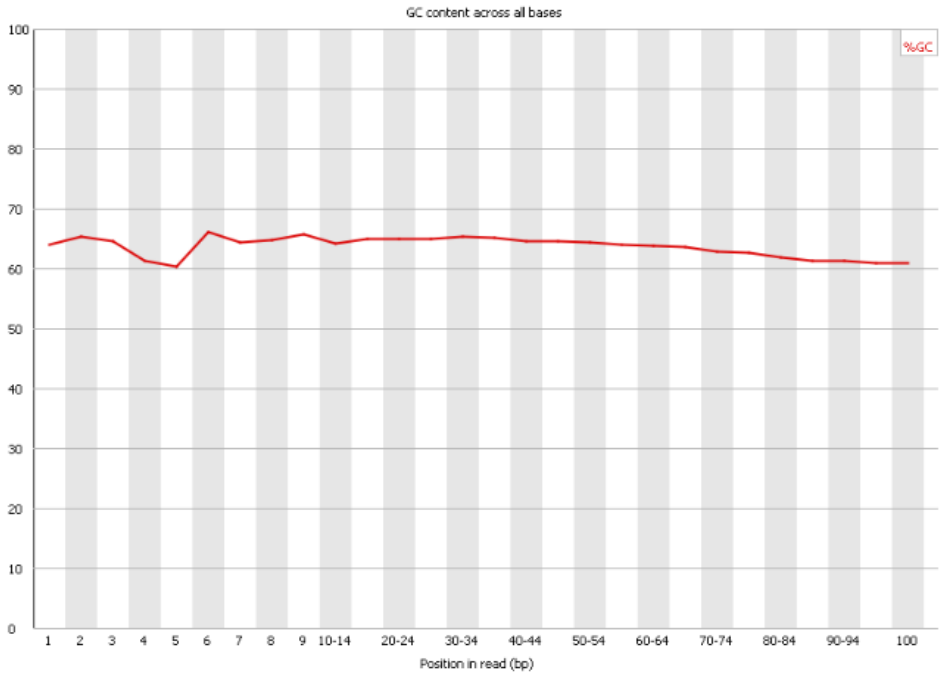

Figure A4: The mean percentage of the GC content across all bases
Legend to Figure A4:
The horizontal plot across the graph indicates minimal change in the average GC content across all bases. It also confirms the expected high GC content of M. tuberculosis which is above 60%

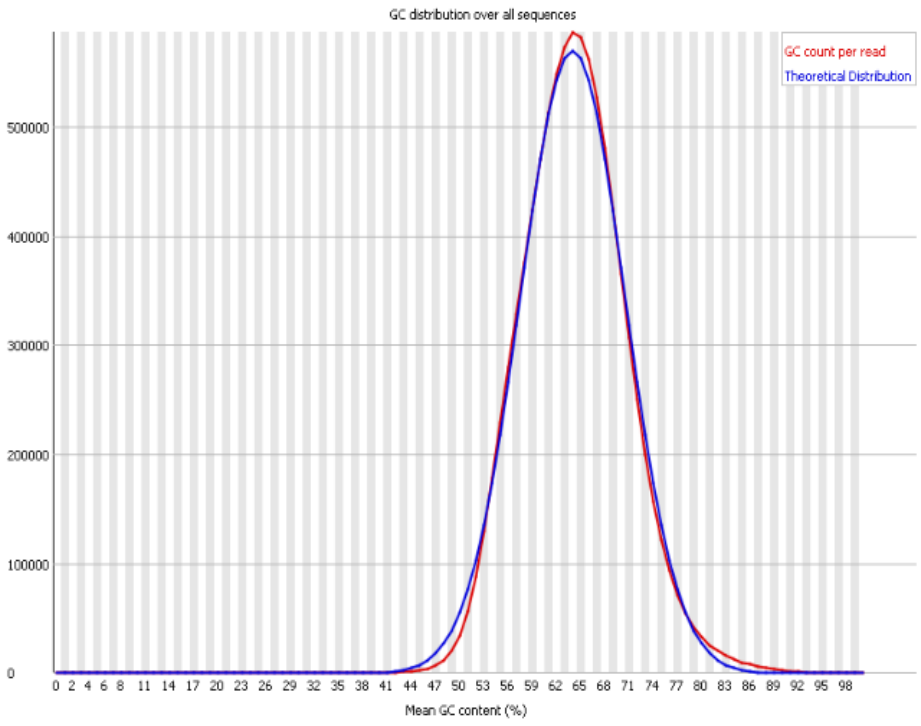

Figure A5: The mean distribution of the percentage GC content across all sequences
Legend to Figure A5:
A normal distribution of the GC content is observed with the central peak corresponding to the overall GC content of this genome. This plot also indicates that there were no contaminants in the sequenced library

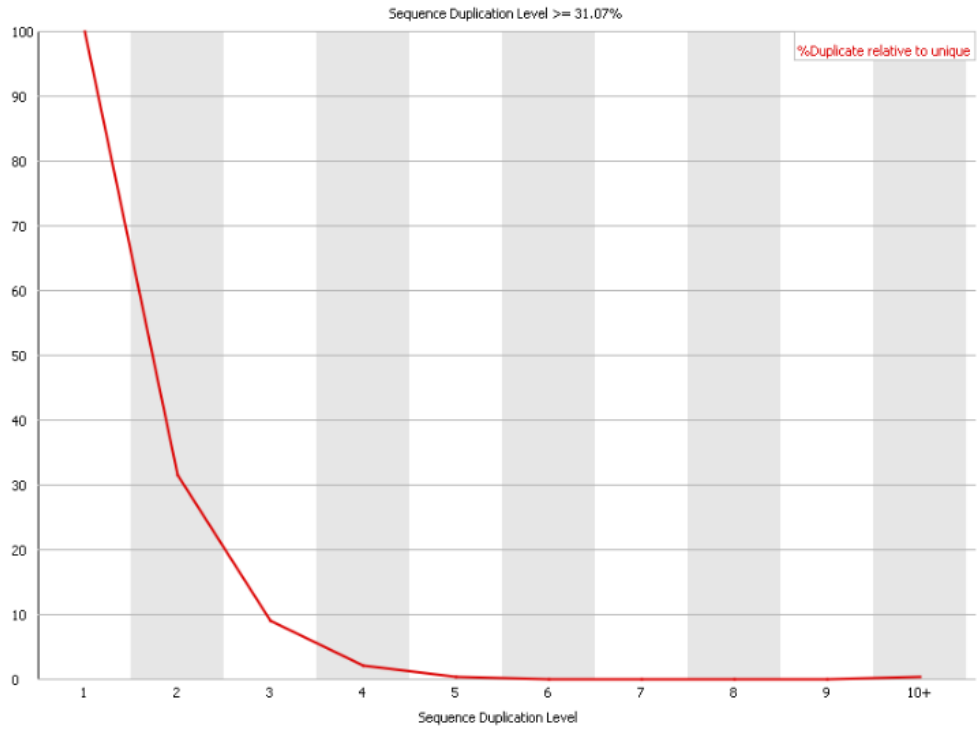

Figure A6: Sequence duplication levels in the library sequenced

Legend to Figure A6:

This plot indicates that a high percentage of duplicate sequences are present. This however might be due to technical duplication (PCR over amplification) or over-sequencing (high fold coverage).

Reference List

1. **Abe, C., I. Kobayashi, S. Mitarai, M. Wada, Y. Kawabe, T. Takashima, K. Suzuki, L.-H. Sng, S. Wang, H. H. Htay, and H. Ogata**. 2008. Biological and molecular characteristics of *Mycobacterium tuberculosis* clinical isolates with low-level resistance to isoniazid in Japan. J. Clin. Microbiol. **46**:2263–2268.
2. **Abeel, T., T. Van Parys, Y. Saeys, J. Galagan, and Y. Van de Peer**. 2012. GenomeView: a next-generation genome browser. Nucleic Acids Res. **40**:e12.
3. **Achtman, M.** 2008. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. Annu. Rev. Microbiol. **62**:53–70.
4. **Aínsa, J. A., C. Martin, B. Gicquel, and R. Gomez-Lus**. 1996. Characterization of the chromosomal aminoglycoside 2'-N-acetyltransferase gene from Mycobacterium fortuitum. Antimicrob. Agents Chemother. **40**:2350–2355.
5. **Aínsa, J. A., E. Pérez, V. Pelicic, F. X. Berthet, B. Gicquel, and C. Martín**. 1997. Aminoglycoside 2'-N-acetyltransferase genes are universally present in mycobacteria: characterization of the aac(2')-Ic gene from *Mycobacterium tuberculosis* and the aac(2')-Id gene from *Mycobacterium smegmatis*. Mol. Microbiol. **24**:431–441.
6. **Akhter, Y., M. T. Ehebauer, S. Mukhopadhyay, and S. E. Hasnain**. 2012. The PE/PPE multigene family codes for virulence factors and is a possible source of mycobacterial antigenic variation: perhaps more? Biochimie **94**:110–116.
7. **Alangaden, G. J., B. N. Kreiswirth, A. Aouad, M. Khetarpal, F. R. Igno, S. L. Moghazeh, E. K. Manavathu, and S. A. Lerner**. 1998. Mechanism of resistance to amikacin and kanamycin in *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. **42**:1295–1297.
8. **Alcaide, F., G. E. Pfyffer, and A. Telenti**. 1997. Role of *embB* in natural and acquired resistance to ethambutol in mycobacteria. Antimicrob. Agents Chemother. **41**:2270–2273.
9. **Alderwick, L. J., V. Molle, L. Kremer, A. J. Cozzone, T. R. Dafforn, G. S. Besra, and K. Fütterer**. 2006. Molecular structure of EmbR, a response element of Ser/Thr kinase signaling in *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. U.S.A. **103**:2558–2563.
10. **Amini, S., and S. Tavazoie**. 2011. Antibiotics and the post-genome revolution. Curr. Opin. Microbiol. **14**:513–518.
11. **Andersson, D. I., and D. Hughes**. 2010. Antibiotic resistance and its cost: is it possible to reverse resistance? Nat. Rev. Microbiol. **8**:260–271.
12. **Ando, H., T. Kitao, T. Miyoshi-Akiyama, S. Kato, T. Mori, and T. Kirikae**. 2011. Downregulation of *katG* expression is associated with isoniazid resistance in *Mycobacterium tuberculosis*. Mol. Microbiol. **79**:1615–1628.
13. **Ando, H., Y. Kondo, T. Suetake, E. Toyota, S. Kato, T. Mori, and T. Kirikae**. 2010. Identification of *katG* mutations associated with high-level isoniazid resistance in *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. **54**:1793–1799.
14. **Andries, K., P. Verhasselt, J. Guillemont, H. W. H. Göhlmann, J.-M. Neefs, H. Winkler, J. V. Gestel, P. Timmerman, M. Zhu, E. Lee, P. Williams, D. de Chaffoy, E. Huitric, S. Hoffner, E. Cambau, C. Truffot-Pernot, N. Lounis, and V. Jarlier**. 2005. A Diarylquinoline Drug Active on the ATP Synthase of *Mycobacterium tuberculosis*. Science **307**:223–227.
15. **Baker, L. V., T. J. Brown, O. Maxwell, A. L. Gibson, Z. Fang, M. D. Yates, and F. A. Drobniewski**. 2005. Molecular analysis of isoniazid-resistant *Mycobacterium tuberculosis* isolates from England and Wales reveals the phylogenetic significance of the ahpC -46A polymorphism. Antimicrob. Agents Chemother. **49**:1455–1464.
16. **Bald, D., and A. Koul**. 2010. Respiratory ATP synthesis: the new generation of mycobacterial drug targets? FEMS Microbiol. Lett. **308**:1–7.
17. **Banerjee, A., E. Dubnau, A. Quemard, V. Balasubramanian, K. S. Um, T. Wilson, D. Collins, G. de Lisle, and W. R. Jacobs Jr**. 1994. *inhA*, a gene encoding a target for isoniazid and ethionamide in *Mycobacterium tuberculosis*. Science **263**:227–230.

18. **Banerjee, S. K., K. Bhatt, S. Rana, P. Misra, and P. K. Chakraborti**. 1996. Involvement of an efflux system in mediating high level of fluoroquinolone resistance in *Mycobacterium smegmatis*. Biochem. Biophys. Res. Commun. **226**:362–368.
19. **Baulard, A. R., J. C. Betts, J. Engohang-Ndong, S. Quan, R. A. McAdam, P. J. Brennan, C. Locht, and G. S. Besra**. 2000. Activation of the pro-drug ethionamide is regulated in mycobacteria. J. Biol. Chem. **275**:28326–28331.
20. **Baysarowich, J., K. Koteva, D. W. Hughes, L. Ejim, E. Griffiths, K. Zhang, M. Junop, and G. D. Wright**. 2008. Rifamycin antibiotic resistance by ADP-ribosylation: Structure and diversity of Arr. Proc. Natl. Acad. Sci. U.S.A. **105**:4886–4891.
21. **Belanger, A. E., G. S. Besra, M. E. Ford, K. Mikusová, J. T. Belisle, P. J. Brennan, and J. M. Inamine**. 1996. The *embAB* genes of Mycobacterium avium encode an arabinosyl transferase involved in cell wall arabinan biosynthesis that is the target for the antimycobacterial drug ethambutol. Proc. Natl. Acad. Sci. U.S.A. **93**:11919–11924.
22. **Beres, S. B., R. K. Carroll, P. R. Shea, I. Sitkiewicz, J. C. Martinez-Gutierrez, D. E. Low, A. McGeer, B. M. Willey, K. Green, G. J. Tyrrell, T. D. Goldman, M. Feldgarden, B. W. Birren, Y. Fofanov, J. Boos, W. D. Wheaton, C. Honisch, and J. M. Musser**. 2010. Molecular complexity of successive bacterial epidemics deconvoluted by comparative pathogenomics. Proc. Natl. Acad. Sci. U.S.A. **107**:4371–4376.
23. **Bhatt, K., S. K. Banerjee, and P. K. Chakraborti**. 2000. Evidence that phosphate specific transporter is amplified in a fluoroquinolone resistant *Mycobacterium smegmatis*. Eur. J. Biochem. **267**:4028–4032.
24. **Bianco, M. V., F. C. Blanco, B. Imperiale, M. A. Forrellad, R. V. Rocha, L. I. Klepp, A. A. Cataldi, N. Morcillo, and F. Bigi**. 2011. Role of P27 -P55 operon from *Mycobacterium tuberculosis* in the resistance to toxic compounds. BMC Infect Dis **11**:195.
25. **Bigi, F., A. Alito, M. I. Romano, M. Zumarraga, K. Caimi, and A. Cataldi**. 2000. The gene encoding P27 lipoprotein and a putative antibiotic-resistance gene form an operon in *Mycobacterium tuberculosis* and Mycobacterium bovis. Microbiology (Reading, Engl.) **146 (Pt 4)**:1011–1018.
26. **Billington, O. J., T. D. McHugh, and S. H. Gillespie**. 1999. Physiological cost of rifampin resistance induced in vitro in *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. **43**:1866–1869.
27. **Böttger, E. C.** 2011. The ins and outs of *Mycobacterium tuberculosis* drug susceptibility testing. Clinical Microbiology and Infection **17**:1128–1134.
28. **Brandis, G., M. Wrande, L. Liljas, and D. Hughes**. 2012. Fitness-compensatory mutations in rifampicin-resistant RNA polymerase. Molecular microbiology.
29. **Brennan, M. J., and G. Delogu**. 2002. The PE multigene family: a “molecular mantra” for mycobacteria. Trends Microbiol. **10**:246–249.
30. **Brennan, P. J., and H. Nikaido**. 1995. The envelope of mycobacteria. Annu. Rev. Biochem. **64**:29–63.
31. **Buroni, S., G. Manina, P. Guglierame, M. R. Pasca, G. Riccardi, and E. De Rossi**. 2006. LfrR is a repressor that regulates expression of the efflux pump LfrA in *Mycobacterium smegmatis*. Antimicrob. Agents Chemother. **50**:4044–4052.
32. **Campbell, P. J., G. P. Morlock, R. D. Sikes, T. L. Dalton, B. Metchock, A. M. Starks, D. P. Hooks, L. S. Cowan, B. B. Plikaytis, and J. E. Posey**. 2011. Molecular detection of mutations associated with first- and second-line drug resistance compared with conventional drug susceptibility testing of *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. **55**:2032–2041.
33. **Capon, F., M. H. Allen, M. Ameen, A. D. Burden, D. Tillman, J. N. Barker, and R. C. Trembath**. 2004. A synonymous SNP of the corneodesmosin gene leads to increased mRNA stability and demonstrates association with psoriasis across diverse ethnic groups. Hum. Mol. Genet. **13**:2361–2368.

34. **Casali, N., V. Nikolayevskyy, Y. Balabanova, O. Ignatyeva, I. Kontsevaya, S. R. Harris, S. D. Bentley, J. Parkhill, S. Nejentsev, S. E. Hoffner, R. D. Horstmann, T. Brown, and F. Drobniewski**. 2012. Microevolution of extensively drug-resistant tuberculosis in Russia. Genome Res. **22**:735–745.
35. **Chaves, F., M. Alonso-Sanz, M. J. Rebollo, J. C. Tercero, M. S. Jimenez, and A. R. Noriega**. 2000. *rpoB* mutations as an epidemiologic marker in rifampin-resistant *Mycobacterium tuberculosis*. Int. J. Tuberc. Lung Dis. **4**:765–770.
36. **Chen, J., Z. Chen, Y. Li, W. Xia, X. Chen, T. Chen, L. Zhou, B. Xu, and S. Xu**. 2012. Characterization of gyrA and gyrB mutations and fluoroquinolone resistance in *Mycobacterium tuberculosis* clinical isolates from Hubei Province, China. Braz J Infect Dis **16**:136–141.
37. **Chihota, V. N., B. Müller, C. K. Mlambo, M. Pillay, M. Tait, E. M. Streicher, E. Marais, G. D. van der Spuy, M. Hanekom, G. Coetzee, A. Trollip, C. Hayes, M. E. Bosman, N. C. Gey van Pittius, T. C. Victor, P. D. van Helden, and R. M. Warren**. 2012. Population structure of multi- and extensively drug-resistant *Mycobacterium tuberculosis* strains in South Africa. J. Clin. Microbiol. **50**:995–1002.
38. **Chopra, I., and P. Brennan**. 1997. Molecular action of anti-mycobacterial agents. Tuber. Lung Dis. **78**:89–98.
39. **Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry 3rd, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead, and B. G. Barrell**. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature **393**:537–544.
40. **Comas, I., S. Borrell, A. Roetzer, G. Rose, B. Malla, M. Kato-Maeda, J. Galagan, S. Niemann, and S. Gagneux**. 2012. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. Nat. Genet. **44**:106–110.
41. **Cummings, M. P., and M. R. Segal**. 2004. Few amino acid positions in *rpoB* are associated with most of the rifampin resistance in *Mycobacterium tuberculosis*. BMC Bioinformatics **5**:137.
42. **Danilchanka, O., C. Mailaender, and M. Niederweis**. 2008. Identification of a novel multidrug efflux pump of *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. **52**:2503–2511.
43. **De Rossi, E., J. A. Aínsa, and G. Riccardi**. 2006. Role of mycobacterial efflux transporters in drug resistance: an unresolved question. FEMS Microbiol. Rev. **30**:36–52.
44. **De Rossi, E., P. Arrigo, M. Bellinzoni, P. A. E. Silva, C. Martín, J. A. Aínsa, P. Guglierame, and G. Riccardi**. 2002. The multidrug transporters belonging to major facilitator superfamily in *Mycobacterium tuberculosis*. Mol. Med. **8**:714–724.
45. **DeBarber, A. E., K. Mdluli, M. Bosman, L. G. Bekker, and C. E. Barry 3rd**. 2000. Ethionamide activation and sensitivity in multidrug-resistant *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. U.S.A. **97**:9677–9682.
46. **Delogu, G., C. Pusceddu, A. Bua, G. Fadda, M. J. Brennan, and S. Zanetti**. 2004. Rv1818c-encoded PE_PGRS protein of *Mycobacterium tuberculosis* is surface exposed and influences bacterial cell structure. Molecular Microbiology **52**:725–733.
47. **Deretic, V., W. Philipp, S. Dhandayuthapani, M. H. Mudd, R. Curcic, T. Garbe, B. Heym, L. E. Via, and S. T. Cole**. 1995. *Mycobacterium tuberculosis* is a natural mutant with an inactivated oxidative-stress regulatory gene: implications for sensitivity to isoniazid. Mol. Microbiol. **17**:889–900.
48. **Deretic, V., J. Song, and E. Pagán-Ramos**. 1997. Loss of oxyR in *Mycobacterium tuberculosis*. Trends Microbiol. **5**:367–372.

49. **Domenech, P., N. Honoré, B. Heym, and S. T. Cole**. 2001. Role of OxyS of *Mycobacterium tuberculosis* in oxidative stress: overexpression confers increased sensitivity to organic hydroperoxides. Microbes Infect. **3**:713–721.
50. **Drlica, K., and M. Malik**. 2003. Fluoroquinolones: action and resistance. Curr Top Med Chem **3**:249–282.
51. **Du, X., W. Wang, R. Kim, H. Yakota, H. Nguyen, and S. H. Kim**. 2001. Crystal structure and mechanism of catalysis of a pyrazinamidase from Pyrococcus horikoshii. Biochemistry **40**:14166–14172.
52. **Duan, J., M. S. Wainwright, J. M. Comeron, N. Saitou, A. R. Sanders, J. Gelernter, and P. V. Gejman**. 2003. Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. Hum. Mol. Genet. **12**:205–216.
53. **Durocher, D., I. A. Taylor, D. Sarbassova, L. F. Haire, S. L. Westcott, S. P. Jackson, S. J. Smerdon, and M. B. Yaffe**. 2000. The molecular basis of FHA domain:phosphopeptide binding specificity and implications for phospho-dependent signaling mechanisms. Mol. Cell **6**:1169–1182.
54. **Ellis, M. W., and J. S. Lewis 2nd**. 2005. Treatment approaches for community-acquired methicillin-resistant Staphylococcus aureus infections. Curr. Opin. Infect. Dis. **18**:496–501.
55. **Engohang-Ndong, J., D. Baillat, M. Aumercier, F. Bellefontaine, G. S. Besra, C. Locht, and A. R. Baulard**. 2004. EthR, a repressor of the TetR/CamR family implicated in ethionamide resistance in mycobacteria, octamerizes cooperatively on its operator. Mol. Microbiol. **51**:175–188.
56. **Escribano, I., J. C. Rodríguez, B. Llorca, E. García-Pachon, M. Ruiz, and G. Royo**. 2007. Importance of the efflux pump systems in the resistance of *Mycobacterium tuberculosis* to fluoroquinolones and linezolid. Chemotherapy **53**:397–401.
57. **Feniouk, B. A., T. Suzuki, and M. Yoshida**. 2006. The role of subunit epsilon in the catalysis and regulation of FOF1-ATP synthase. Biochim. Biophys. Acta **1757**:326–338.
58. **Finken, M., P. Kirschner, A. Meier, A. Wrede, and E. C. Böttger**. 1993. Molecular basis of streptomycin resistance in *Mycobacterium tuberculosis*: alterations of the ribosomal protein S12 gene and point mutations within a functional 16S ribosomal RNA pseudoknot. Mol. Microbiol. **9**:1239–1246.
59. **Ford, C., K. Yusim, T. Ioerger, S. Feng, M. Chase, M. Greene, B. Korber, and S. Fortune**. 2012. *Mycobacterium tuberculosis*--heterogeneity revealed through whole genome sequencing. Tuberculosis (Edinb) **92**:194–201.
60. **Frénois, F., A. R. Baulard, and V. Villeret**. 2006. Insights into mechanisms of induction and ligands recognition in the transcriptional repressor EthR from *Mycobacterium tuberculosis*. Tuberculosis (Edinb) **86**:110–114.
61. **Fu, L. M., and T. M. Shinnick**. 2007. Understanding the action of INH on a highly INH-resistant *Mycobacterium tuberculosis* strain using Genechips. Tuberculosis (Edinb) **87**:63–70.
62. **Gagneux, S.** 2009. Fitness cost of drug resistance in *Mycobacterium tuberculosis*. Clin. Microbiol. Infect. **15 Suppl 1**:66–68.
63. **Gagneux, S., C. D. Long, P. M. Small, T. Van, G. K. Schoolnik, and B. J. M. Bohannan**. 2006. The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. Science **312**:1944–1946.
64. **Georg, J., and W. R. Hess**. 2011. cis-Antisense RNA, Another Level of Gene Regulation in Bacteria. Microbiol Mol Biol Rev **75**:286–300.
65. **Georghiou, S. B., M. Magana, R. S. Garfein, D. G. Catanzaro, A. Catanzaro, and T. C. Rodwell**. 2012. Evaluation of Genetic Mutations Associated with *Mycobacterium tuberculosis* Resistance to Amikacin, Kanamycin and Capreomycin: A Systematic Review. PLoS One **7**.
66. **Grzegorzewicz, A. E., H. Pham, V. A. K. B. Gundi, M. S. Scherman, E. J. North, T. Hess, V. Jones, V. Gruppo, S. E. M. Born, J. Korduláková, S. S. Chavadi, C. Morisseau, A. J. Lenaerts,**

R. E. Lee, M. R. McNeil, and M. Jackson. 2012. Inhibition of mycolic acid transport across the *Mycobacterium tuberculosis* plasma membrane. Nature Chemical Biology **8**:334–341.

67. **Gumbo, T.** 2010. New susceptibility breakpoints for first-line antituberculosis drugs based on antimicrobial pharmacokinetic/pharmacodynamic science and population pharmacokinetic variability. Antimicrob. Agents Chemother. **54**:1484–1491.
68. **Gupta, A. K., V. P. Reddy, M. Lavania, D. S. Chauhan, K. Venkatesan, V. D. Sharma, A. K. Tyagi, and V. M. Katoch**. 2010. *jefA* (Rv2459), a drug efflux gene in *Mycobacterium tuberculosis* confers resistance to isoniazid & ethambutol. Indian J. Med. Res. **132**:176–188.
69. **Gupta, R., and M. Espinal**. 2003. A prioritised research agenda for DOTS-Plus for multidrug-resistant tuberculosis (MDR-TB). Int. J. Tuberc. Lung Dis. **7**:410–414.
70. **Hall, N.** 2007. Advanced sequencing technologies and their wider impact in microbiology. J. Exp. Biol. **210**:1518–1525.
71. **Hall, T.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series **41**:95–98.
72. **Hancock, R. E.** 1981. Aminoglycoside uptake and mode of action--with special reference to streptomycin and gentamicin. I. Antagonists and mutants. J. Antimicrob. Chemother. **8**:249–276.
73. **Hartmann, G., K. O. Honikel, F. Knüsel, and J. Nüesch**. 1967. The specific inhibition of the DNA-directed RNA synthesis by rifamycin. Biochim. Biophys. Acta **145**:843–844.
74. **Hauck, Y., M. Fabre, G. Vergnaud, C. Soler, and C. Pourcel**. 2009. Comparison of Two Commercial Assays for the Characterization of *rpoB* Mutations in *Mycobacterium tuberculosis* and Description of New Mutations Conferring Weak Resistance to Rifampicin. J. Antimicrob. Chemother. **64**:259–262.
75. **Hazbón, M. H., M. Brimacombe, M. Bobadilla del Valle, M. Cavatore, M. I. Guerrero, M. Varma-Basil, H. Billman-Jacobe, C. Lavender, J. Fyfe, L. García-García, C. I. León, M. Bose, F. Chaves, M. Murray, K. D. Eisenach, J. Sifuentes-Osornio, M. D. Cave, A. Ponce de León, and D. Alland**. 2006. Population genetics study of isoniazid resistance mutations and evolution of multidrug-resistant *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. **50**:2640–2649.
76. **Hershberg, R., M. Lipatov, P. M. Small, H. Sheffer, S. Niemann, S. Homolka, J. C. Roach, K. Kremer, D. A. Petrov, M. W. Feldman, and S. Gagneux**. 2008. High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. PLoS Biol. **6**:e311.
77. **Hetherington, S. V., A. S. Watson, and C. C. Patrick**. 1995. Sequence and analysis of the *rpoB* gene of *Mycobacterium smegmatis*. Antimicrob. Agents Chemother. **39**:2164–2166.
78. **Hinrichs, W., C. Kisker, M. Düvel, A. Müller, K. Tovar, W. Hillen, and W. Saenger**. 1994. Structure of the Tet repressor-tetracycline complex and regulation of antibiotic resistance. Science **264**:418–420.
79. **Ho, H., M. Cheng, and D. T. Chiu**. 2007. Glucose-6-phosphate dehydrogenase--from oxidative stress to cellular functions and degenerative diseases. Redox Rep. **12**:109–118.
80. **Hu, B., G. Xie, C.-C. Lo, S. R. Starkenburg, and P. S. G. Chain**. 2011. Pathogen comparative genomics in the next-generation sequencing era: genome alignments, pangenomics and metagenomics. Brief Funct Genomics **10**:322–333.
81. **Hu, Y., A. R. Coates, and D. A. Mitchison**. 2006. Sterilising action of pyrazinamide in models of dormant and rifampicin-tolerant *Mycobacterium tuberculosis*. Int. J. Tuberc. Lung Dis. **10**:317–322.
82. **Huitric, E., J. Werngren, P. Juréen, and S. Hoffner**. 2006. Resistance levels and *rpoB* gene mutations among in vitro-selected rifampin-resistant *Mycobacterium tuberculosis* mutants. Antimicrob. Agents Chemother. **50**:2860–2862.

83. **Hull, E. P., M. E. Spencer, D. Wood, and J. R. Guest**. 1983. Nucleotide sequence of the promoter region of the citrate synthase gene (gltA) of Escherichia coli. FEBS Letters **156**:366–370.
84. **Ioerger, T. R., Y. Feng, X. Chen, K. M. Dobos, T. C. Victor, E. M. Streicher, R. M. Warren, N. C. Gey van Pittius, P. D. Van Helden, and J. C. Sacchettini**. 2010. The non-clonality of drug resistance in Beijing-genotype isolates of *Mycobacterium tuberculosis* from the Western Cape of South Africa. BMC Genomics **11**:670.
85. **Ioerger, T. R., Y. Feng, K. Ganesula, X. Chen, K. M. Dobos, S. Fortune, W. R. Jacobs Jr, V. Mizrahi, T. Parish, E. Rubin, C. Sassetti, and J. C. Sacchettini**. 2010. Variation among genome sequences of H37Rv strains of *Mycobacterium tuberculosis* from multiple laboratories. J. Bacteriol. **192**:3645–3653.
86. **Jadaun, G. P. S., R. Das, P. Upadhyay, D. S. Chauhan, V. D. Sharma, and V. M. Katoch**. 2009. Role of embCAB gene mutations in ethambutol resistance in *Mycobacterium tuberculosis* isolates from India. Int. J. Antimicrob. Agents **33**:483–486.
87. **Jarlier, V., and H. Nikaido**. 1994. Mycobacterial cell wall: structure and role in natural resistance to antibiotics. FEMS Microbiol. Lett. **123**:11–18.
88. **Jarlier, V., and H. Nikaido**. 1990. Permeability barrier to hydrophilic solutes in Mycobacterium chelonei. J Bacteriol **172**:1418–1423.
89. **Jindani, A., C. J. Doré, and D. A. Mitchison**. 2003. Bactericidal and Sterilizing Activities of Antituberculosis Drugs during the First 14 Days. Am. J. Respir. Crit. Care Med. **167**:1348–1354.
90. **Johansen, S. K., C. E. Maus, B. B. Plikaytis, and S. Douthwaite**. 2006. Capreomycin binds across the ribosomal subunit interface using tlyA-encoded 2'-O-methylations in 16S and 23S rRNAs. Mol. Cell **23**:173–182.
91. **Johnson, A. D., D. Wang, and W. Sadee**. 2005. Polymorphisms affecting gene regulation and mRNA processing: broad implications for pharmacogenetics. Pharmacol. Ther. **106**:19–38.
92. **Johnson, R., R. M. Warren, G. D. van der Spuy, N. C. Gey van Pittius, D. Theron, E. M. Streicher, M. Bosman, G. J. Coetzee, P. D. van Helden, and T. C. Victor**. 2010. Drug-resistant tuberculosis epidemic in the Western Cape driven by a virulent Beijing genotype strain. Int. J. Tuberc. Lung Dis. **14**:119–121.
93. **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.
94. **Kam, K. M., C. W. Yip, T. L. Cheung, H. S. Tang, O. C. Leung, and M. Y. Chan**. 2006. Stepwise decrease in moxifloxacin susceptibility amongst clinical isolates of multidrug-resistant *Mycobacterium tuberculosis*: correlation with ofloxacin susceptibility. Microb. Drug Resist. **12**:7–11.
95. **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.
96. **Kimchi-Sarfaty, C., J. M. Oh, I.-W. Kim, Z. E. Sauna, A. M. Calcagno, S. V. Ambudkar, and M. M. Gottesman**. 2007. A "silent" polymorphism in the MDR1 gene changes substrate specificity. Science **315**:525–528.
97. **Kohli, S., Y. Singh, K. Sharma, A. Mittal, N. Z. Ehtesham, and S. E. Hasnain**. 2012. Comparative genomic and proteomic analyses of PE/PPE multigene family of *Mycobacterium tuberculosis* H37Rv and H37Ra reveal novel and interesting differences with implications in virulence. Nucleic Acids Res **40**:7113–7122.
98. **Konno, K., F. M. Feldmann, and W. McDermott**. 1967. Pyrazinamide susceptibility and amidase activity of tubercle bacilli. Am. Rev. Respir. Dis. **95**:461–469.
99. **Koul, A., L. Vranckx, N. Dendouga, W. Balemans, I. Van den Wyngaert, K. Vergauwen, H. W. H. Göhlmann, R. Willebrords, A. Poncelet, J. Guillemont, D. Bald, and K. Andries**. 2008.

Diarylquinolines are bactericidal for dormant mycobacteria as a result of disturbed ATP homeostasis. J. Biol. Chem. **283**:25273–25280.

100. **Krüüner, A., P. Jureen, K. Levina, S. Ghebremichael, and S. Hoffner**. 2003. Discordant resistance to kanamycin and amikacin in drug-resistant *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. **47**:2971–2973.
101. **Krüüner, A., M. D. Yates, and F. A. Drobniewski**. 2006. Evaluation of MGIT 960-based antimicrobial testing and determination of critical concentrations of first- and second-line antimicrobial drugs with drug-resistant clinical strains of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **44**:811–818.
102. **Lanéelle, M.-A., and M. Daffé**. 2009. Mycobacteria Protocols, p. 1–11. *In* Humana Press.
103. **Langford, K. W., B. Penkov, I. M. Derrington, and J. H. Gundlach**. 2011. Unsupported planar lipid membranes formed from mycolic acids of *Mycobacterium tuberculosis*. J. Lipid Res. **52**:272–277.
104. **Larsen, M. H., C. Vilchèze, L. Kremer, G. S. Besra, L. Parsons, M. Salfinger, L. Heifets, M. H. Hazbon, D. Alland, J. C. Sacchettini, and W. R. Jacobs Jr**. 2002. Overexpression of *inhA*, but not kasA, confers resistance to isoniazid and ethionamide in *Mycobacterium smegmatis*, M. bovis BCG and M. tuberculosis. Mol. Microbiol. **46**:453–466.
105. **Laurenzo, D., and S. A. Mousa**. 2011. Mechanisms of drug resistance in *Mycobacterium tuberculosis* and current status of rapid molecular diagnostic testing. Acta Tropica **119**:5–10.
106. **Lee, H.-I., J.-H. Yoon, J.-S. Nam, Y.-M. Kim, and Y.-T. Ro**. 2010. Cloning, expression and characterization of the catalase-peroxidase (*KatG*) gene from a fast-growing Mycobacterium sp. strain JC1 DSM 3803. J. Biochem. **147**:511–522.
107. **Li, H., and R. Durbin**. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25**:1754–1760.
108. **Li, X.-Z., L. Zhang, and H. Nikaido**. 2004. Efflux pump-mediated intrinsic drug resistance in *Mycobacterium smegmatis*. Antimicrob. Agents Chemother. **48**:2415–2423.
109. **Li, Y., and Z.-G. He**. 2012. The Mycobacterial LysR-Type Regulator OxyS Responds to Oxidative Stress and Negatively Regulates Expression of the Catalase-Peroxidase Gene. PLoS One **7**.
110. **Liu, J., H. E. Takiff, and H. Nikaido**. 1996. Active efflux of fluoroquinolones in *Mycobacterium smegmatis* mediated by LfrA, a multidrug efflux pump. J. Bacteriol. **178**:3791–3795.
111. **Llano-Sotelo, B., E. F. Azucena Jr, L. P. Kotra, S. Mobashery, and C. S. Chow**. 2002. Aminoglycosides modified by resistance enzymes display diminished binding to the bacterial ribosomal aminoacyl-tRNA site. Chem. Biol. **9**:455–463.
112. **Louw, G. E., R. M. Warren, P. R. Donald, M. B. Murray, M. Bosman, P. D. Van Helden, D. B. Young, and T. C. Victor**. 2006. Frequency and implications of pyrazinamide resistance in managing previously treated tuberculosis patients. Int. J. Tuberc. Lung Dis. **10**:802–807.
113. **Louw, G. E., R. M. Warren, N. C. Gey van Pittius, C. R. E. McEvoy, P. D. Van Helden, and T. C. Victor**. 2009. A balancing act: efflux/influx in mycobacterial drug resistance. Antimicrob. Agents Chemother. **53**:3181–3189.
114. **Louw, G. E., R. M. Warren, N. C. Gey van Pittius, R. Leon, A. Jimenez, R. Hernandez-Pando, C. R. E. McEvoy, M. Grobbelaar, M. Murray, P. D. van Helden, and T. C. Victor**. 2011. Rifampicin reduces susceptibility to ofloxacin in rifampicin-resistant *Mycobacterium tuberculosis* through efflux. Am. J. Respir. Crit. Care Med. **184**:269–276.
115. **Machado, D., I. Couto, J. Perdigão, L. Rodrigues, I. Portugal, P. Baptista, B. Veigas, L. Amaral, and M. Viveiros**. 2012. Contribution of Efflux to the Emergence of Isoniazid and Multidrug Resistance in *Mycobacterium tuberculosis*. PLoS One **7**.
116. **Madison, B. M., S. H. Siddiqi, L. Heifets, W. Gross, M. Higgins, N. Warren, A. Thompson, G. Morlock, and J. C. Ridderhof**. 2004. Identification of a *Mycobacterium tuberculosis* strain with stable, low-level resistance to isoniazid. J. Clin. Microbiol. **42**:1294–1295.

117. **Mailaender, C., N. Reiling, H. Engelhardt, S. Bossmann, S. Ehlers, and M. Niederweis**. 2004. The MspA porin promotes growth and increases antibiotic susceptibility of both Mycobacterium bovis BCG and *Mycobacterium tuberculosis*. Microbiology **150**:853–864.
118. **Malaviya, A., and J. Gomes**. 2008. Enhanced biotransformation of sitosterol to androstenedione by Mycobacterium sp. using cell wall permeabilizing antibiotics. J. Ind. Microbiol. Biotechnol. **35**:1235–1239.
119. **Malik, S., M. Willby, D. Sikes, O. V. Tsodikov, and J. E. Posey**. 2012. New Insights into Fluoroquinolone Resistance in *Mycobacterium tuberculosis*: Functional Genetic Analysis of gyrA and gyrB Mutations. PLoS One **7**.
120. **Mariam, D. H., Y. Mengistu, S. E. Hoffner, and D. I. Andersson**. 2004. Effect of *rpoB* mutations conferring rifampin resistance on fitness of *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. **48**:1289–1294.
121. **Maus, C. E., B. B. Plikaytis, and T. M. Shinnick**. 2005. Molecular analysis of cross-resistance to capreomycin, kanamycin, amikacin, and viomycin in *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. **49**:3192–3197.
122. **Maus, C. E., B. B. Plikaytis, and T. M. Shinnick**. 2005. Mutation of *tlyA* confers capreomycin resistance in *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. **49**:571–577.
123. **McEvoy, C. R. E., R. Cloete, B. Müller, A. C. Schürch, P. D. van Helden, S. Gagneux, R. M. Warren, and N. C. Gey van Pittius**. 2012. Comparative analysis of *Mycobacterium tuberculosis* pe and ppe genes reveals high sequence variation and an apparent absence of selective constraints. PLoS ONE **7**:e30593.
124. **McGuire, A. M., B. Weiner, S. T. Park, I. Wapinski, S. Raman, G. Dolganov, M. Peterson, R. Riley, J. Zucker, T. Abeel, J. White, P. Sisk, C. Stolte, M. Koehrsen, R. T. Yamamoto, M. Iacobelli-Martinez, M. J. Kidd, A. M. Maer, G. K. Schoolnik, A. Regev, and J. Galagan**. 2012. Comparative analysis of mycobacterium and related actinomycetes yields insight into the evolution of *Mycobacterium tuberculosis* pathogenesis. BMC Genomics **13**:120.
125. **Meier, A., P. Sander, K. J. Schaper, M. Scholz, and E. C. Böttger**. 1996. Correlation of molecular resistance mechanisms and phenotypic resistance levels in streptomycin-resistant *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. **40**:2452–2454.
126. **Mick, V., M. J. Rebollo, A. Lucía, M. J. García, C. Martín, and J. A. Aínsa**. 2008. Transcriptional analysis of and resistance level conferred by the aminoglycoside acetyltransferase gene *aac*(2')-Id from *Mycobacterium smegmatis*. J. Antimicrob. Chemother. **61**:39–45.
127. **Milano, A., F. Forti, C. Sala, G. Riccardi, and D. Ghisotti**. 2001. Transcriptional regulation of furA and *katG* upon oxidative stress in *Mycobacterium smegmatis*. J. Bacteriol. **183**:6801–6806.
128. **Mishra, K. C., C. de Chastellier, Y. Narayana, P. Bifani, A. K. Brown, G. S. Besra, V. M. Katoch, B. Joshi, K. N. Balaji, and L. Kremer**. 2008. Functional role of the PE domain and immunogenicity of the *Mycobacterium tuberculosis* triacylglycerol hydrolase LipY. Infect. Immun. **76**:127–140.
129. **Molle, V., L. Kremer, C. Girard-Blanc, G. S. Besra, A. J. Cozzone, and J.-F. Prost**. 2003. An FHA phosphoprotein recognition domain mediates protein EmbR phosphorylation by PknH, a Ser/Thr protein kinase from *Mycobacterium tuberculosis*. Biochemistry **42**:15300–15309.
130. **Molle, V., R. C. Reynolds, L. J. Alderwick, G. S. Besra, A. J. Cozzone, K. Fütterer, and L. Kremer**. 2008. EmbR2, a structural homologue of EmbR, inhibits the *Mycobacterium tuberculosis* kinase/substrate pair PknH/EmbR. Biochem. J. **410**:309–317.
131. **Morlock, G. P., B. Metchock, D. Sikes, J. T. Crawford, and R. C. Cooksey**. 2003. ethA, *inhA*, and *katG* loci of ethionamide-resistant clinical *Mycobacterium tuberculosis* isolates. Antimicrob. Agents Chemother. **47**:3799–3805.
132. **Morlock, G. P., B. B. Plikaytis, and J. T. Crawford**. 2000. Characterization of Spontaneous, In Vitro-Selected, Rifampin-Resistant Mutants of *Mycobacterium tuberculosis* Strain H37Rv. Antimicrob Agents Chemother **44**:3298–3301.

133. **Morris, S., G. H. Bai, P. Suffys, L. Portillo-Gomez, M. Fairchok, and D. Rouse**. 1995. Molecular mechanisms of multiple drug resistance in clinical isolates of *Mycobacterium tuberculosis*. J. Infect. Dis. **171**:954–960.
134. **Motiwala, A. S., Y. Dai, E. C. Jones-López, S.-H. Hwang, J. S. Lee, S. N. Cho, L. E. Via, C. E. Barry, and D. Alland**. 2010. Mutations in extensively drug-resistant *Mycobacterium tuberculosis* that do not code for known drug-resistance mechanisms. J. Infect. Dis. **201**:881–888.
135. **Nair, J., D. A. Rouse, G. H. Bai, and S. L. Morris**. 1993. The rpsL gene and streptomycin resistance in single and multiple drug-resistant strains of *Mycobacterium tuberculosis*. Mol. Microbiol. **10**:521–527.
136. **Nair, S., A. D. Pandey, and S. Mukhopadhyay**. 2011. The PPE18 protein of *Mycobacterium tuberculosis* inhibits NF-κB/rel-mediated proinflammatory cytokine production by upregulating and phosphorylating suppressor of cytokine signaling 3 protein. J. Immunol. **186**:5413–5424.
137. **Nair, S., P. A. Ramaswamy, S. Ghosh, D. C. Joshi, N. Pathak, I. Siddiqui, P. Sharma, S. E. Hasnain, S. C. Mande, and S. Mukhopadhyay**. 2009. The PPE18 of *Mycobacterium tuberculosis* interacts with TLR2 and activates IL-10 induction in macrophage. J. Immunol. **183**:6269–6281.
138. **Nessar, R., E. Cambau, J. M. Reyrat, A. Murray, and B. Gicquel**. 2012. Mycobacterium abscessus: a new antibiotic nightmare. Journal of Antimicrobial Chemotherapy.
139. **Nguyen, L., and C. J. Thompson**. 2006. Foundations of antibiotic resistance in bacterial physiology: the mycobacterial paradigm. Trends Microbiol. **14**:304–312.
140. **Nhu, N. T. Q., N. T. N. Lan, N. T. N. Phuong, N. van V. Chau, J. Farrar, and M. Caws**. 2012. Association of streptomycin resistance mutations with level of drug resistance and *Mycobacterium tuberculosis* genotypes. Int. J. Tuberc. Lung Dis. **16**:527–531.
141. **Nicoloso, M. S., H. Sun, R. Spizzo, H. Kim, P. Wickramasinghe, M. Shimizu, S. E. Wojcik, J. Ferdin, T. Kunej, L. Xiao, S. Manoukian, G. Secreto, F. Ravagnani, X. Wang, P. Radice, C. M. Croce, R. V. Davuluri, and G. A. Calin**. 2010. SINGLE NUCLEOTIDE POLYMORPHISMS INSIDE microRNA TARGET SITES INFLUENCE TUMOR SUSCEPTIBILITY. Cancer Res **70**:2789–2798.
142. **Niemann, S., C. U. Köser, S. Gagneux, C. Plinke, S. Homolka, H. Bignell, R. J. Carter, R. K. Cheetham, A. Cox, N. A. Gormley, P. Kokko-Gonzales, L. J. Murray, R. Rigatti, V. P. Smith, F. P. M. Arends, H. S. Cox, G. Smith, and J. A. C. Archer**. 2009. Genomic Diversity among Drug Sensitive and Multidrug Resistant Isolates of *Mycobacterium tuberculosis* with Identical DNA Fingerprints. PLoS One **4**.
143. **Pagán-Ramos, E., J. Song, M. McFalone, M. H. Mudd, and V. Deretic**. 1998. Oxidative stress response and characterization of the oxyR-ahpC and furA-*katG* loci in Mycobacterium marinum. J. Bacteriol. **180**:4856–4864.
144. **Pasca, M. R., P. Guglierame, F. Arcesi, M. Bellinzoni, E. De Rossi, and G. Riccardi**. 2004. Rv2686c-Rv2687c-Rv2688c, an ABC fluoroquinolone efflux pump in *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. **48**:3175–3178.
145. **Payton, M., R. Auty, R. Delgoda, M. Everett, and E. Sim**. 1999. Cloning and characterization of arylamine N-acetyltransferase genes from *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*: increased expression results in isoniazid resistance. J. Bacteriol. **181**:1343–1347.
146. **Payton, M., C. Gifford, P. Schartau, C. Hagemeier, A. Mushtaq, S. Lucas, K. Pinter, and E. Sim**. 2001. Evidence towards the role of arylamine N-acetyltransferase in *Mycobacterium smegmatis* and development of a specific antiserum against the homologous enzyme of *Mycobacterium tuberculosis*. Microbiology (Reading, Engl.) **147**:3295–3302.
147. **Pfaffl, M. W.** 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. **29**:e45.

148. **Piddock, L. J., K. J. Williams, and V. Ricci**. 2000. Accumulation of rifampicin by Mycobacterium aurum, *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. J. Antimicrob. Chemother. **45**:159–165.
149. **Pillay, M., and A. W. Sturm**. 2007. Evolution of the Extensively Drug-Resistant F15/LAM4/KZN Strain of *Mycobacterium tuberculosis* in KwaZulu-Natal, South Africa. Clin Infect Dis. **45**:1409–1414.
150. **Pym, A. S., P. Domenech, N. Honoré, J. Song, V. Deretic, and S. T. Cole**. 2001. Regulation of catalase-peroxidase (KatG) expression, isoniazid sensitivity and virulence by furA of *Mycobacterium tuberculosis*. Mol. Microbiol. **40**:879–889.
151. **Rahim, Z., C. Nakajima, R. Raqib, K. Zaman, H. P. Endtz, A. G. M. van der Zanden, and Y. Suzuki**. Molecular mechanism of rifampicin and isoniazid resistance in *Mycobacterium tuberculosis* from Bangladesh. Tuberculosis.
152. **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.
153. **Ramaswamy, S. V., A. G. Amin, S. Göksel, C. E. Stager, S. J. Dou, H. El Sahly, S. L. Moghazeh, B. N. Kreiswirth, and J. M. Musser**. 2000. Molecular genetic analysis of nucleotide polymorphisms associated with ethambutol resistance in human isolates of *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. **44**:326–336.
154. **Ramaswamy, S. V., R. Reich, S.-J. Dou, L. Jasperse, X. Pan, A. Wanger, T. Quitugua, and E. A. Graviss**. 2003. Single nucleotide polymorphisms in genes associated with isoniazid resistance in *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. **47**:1241–1250.
155. **Ramón-García, S., C. Martín, J. A. Aínsa, and E. De Rossi**. 2006. Characterization of tetracycline resistance mediated by the efflux pump Tap from Mycobacterium fortuitum. J. Antimicrob. Chemother. **57**:252–259.
156. **Ramón-García, S., C. Martín, E. De Rossi, and J. A. Aínsa**. 2007. Contribution of the Rv2333c efflux pump (the Stp protein) from *Mycobacterium tuberculosis* to intrinsic antibiotic resistance in Mycobacterium bovis BCG. J. Antimicrob. Chemother. **59**:544–547.
157. **Ramón-García, S., C. Martín, C. J. Thompson, and J. A. Aínsa**. 2009. Role of the *Mycobacterium tuberculosis* P55 efflux pump in intrinsic drug resistance, oxidative stress responses, and growth. Antimicrob. Agents Chemother. **53**:3675–3682.
158. **Rattan, A., A. Kalia, and N. Ahmad**. 1998. Multidrug-resistant *Mycobacterium tuberculosis*: molecular perspectives. Emerging Infect. Dis. **4**:195–209.
159. **Rawat, R., A. Whitty, and P. J. Tonge**. 2003. The isoniazid-NAD adduct is a slow, tight-binding inhibitor of *InhA*, the *Mycobacterium tuberculosis* enoyl reductase: adduct affinity and drug resistance. Proc. Natl. Acad. Sci. U.S.A. **100**:13881–13886.
160. **Rozen, S., and H. Skaletsky**. 2000. Primer3 (version 0.4.0).
161. **Rüsch-Gerdes, S., G. E. Pfyffer, M. Casal, M. Chadwick, and S. Siddiqi**. 2006. Multicenter laboratory validation of the BACTEC MGIT 960 technique for testing susceptibilities of *Mycobacterium tuberculosis* to classical second-line drugs and newer antimicrobials. J. Clin. Microbiol. **44**:688–692.
162. **Safi, H., R. D. Fleischmann, S. N. Peterson, M. B. Jones, B. Jarrahi, and D. Alland**. 2010. Allelic exchange and mutant selection demonstrate that common clinical embCAB gene mutations only modestly increase resistance to ethambutol in *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. **54**:103–108.
163. **Sala, C., F. Forti, E. Di Florio, F. Canneva, A. Milano, G. Riccardi, and D. Ghisotti**. 2003. *Mycobacterium tuberculosis* FurA autoregulates its own expression. J. Bacteriol. **185**:5357–5362.
164. **Sala, C., A. Haouz, F. A. Saul, I. Miras, I. Rosenkrands, P. M. Alzari, and S. T. Cole**. 2009. Genome-wide regulon and crystal structure of BlaI (Rv1846c) from *Mycobacterium tuberculosis*. Mol. Microbiol. **71**:1102–1116.

165. **Sampson, S. L.** 2011. Mycobacterial PE/PPE Proteins at the Host-Pathogen Interface. Clinical and Developmental Immunology **2011**:1–11.
166. **Samuel, L. P., C.-H. Song, J. Wei, E. A. Roberts, J. L. Dahl, C. E. Barry 3rd, E.-K. Jo, and R. L. Friedman**. 2007. Expression, production and release of the Eis protein by *Mycobacterium tuberculosis* during infection of macrophages and its effect on cytokine secretion. Microbiology (Reading, Engl.) **153**:529–540.
167. **Sander, P., E. De Rossi, B. Böddinghaus, R. Cantoni, M. Branzoni, E. C. Böttger, H. Takiff, R. Rodriquez, G. Lopez, and G. Riccardi**. 2000. Contribution of the multidrug efflux pump LfrA to innate mycobacterial drug resistance. FEMS Microbiol. Lett. **193**:19–23.
168. **Sandy, J., S. Holton, E. Fullam, E. Sim, and M. Noble**. 2005. Binding of the anti-tubercular drug isoniazid to the arylamine N-acetyltransferase protein from *Mycobacterium smegmatis*. Protein Sci **14**:775–782.
169. **Santangelo, M. de la P., P. M. Gest, M. E. Guerin, M. Coinçon, H. Pham, G. Ryan, S. E. Puckett, J. S. Spencer, M. Gonzalez-Juarrero, R. Daher, A. J. Lenaerts, D. Schnappinger, M. Therisod, S. Ehrt, J. Sygusch, and M. Jackson**. 2011. Glycolytic and Non-glycolytic Functions of *Mycobacterium tuberculosis* Fructose-1,6-bisphosphate Aldolase, an Essential Enzyme Produced by Replicating and Non-replicating Bacilli. J. Biol. Chem. **286**:40219–40231.
170. **Sassetti, C. M., D. H. Boyd, and E. J. Rubin**. 2001. Comprehensive identification of conditionally essential genes in mycobacteria. Proc. Natl. Acad. Sci. U.S.A. **98**:12712–12717.
171. **Saunders, N. J., U. H. Trivedi, M. L. Thomson, C. Doig, I. F. Laurenson, and M. L. Blaxter**. 2011. Deep resequencing of serial sputum isolates of *Mycobacterium tuberculosis* during therapeutic failure due to poor compliance reveals stepwise mutation of key resistance genes on an otherwise stable genetic background. J. Infect. **62**:212–217.
172. **Schrag, S. J., V. Perrot, and B. R. Levin**. 1997. Adaptation to the fitness costs of antibiotic resistance in Escherichia coli. Proc. Biol. Sci. **264**:1287–1291.
173. **Schürch, A. C., K. Kremer, A. Kiers, O. Daviena, M. J. Boeree, R. J. Siezen, N. H. Smith, and D. van Soolingen**. 2010. The tempo and mode of molecular evolution of *Mycobacterium tuberculosis* at patient-to-patient scale. Infect. Genet. Evol. **10**:108–114.
174. **Schürch, A. C., K. Kremer, R. M. Warren, N. V. Hung, Y. Zhao, K. Wan, M. J. Boeree, R. J. Siezen, N. H. Smith, and D. van Soolingen**. 2011. Mutations in the regulatory network underlie the recent clonal expansion of a dominant subclone of the *Mycobacterium tuberculosis* Beijing genotype. Infection, Genetics and Evolution **11**:587–597.
175. **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.
176. **Sharma, K., M. Gupta, M. Pathak, N. Gupta, A. Koul, S. Sarangi, R. Baweja, and Y. Singh**. 2006. Transcriptional control of the mycobacterial embCAB operon by PknH through a regulatory protein, EmbR, in vivo. J. Bacteriol. **188**:2936–2944.
177. **Sheen, P., P. Ferrer, R. H. Gilman, J. López-Llano, P. Fuentes, E. Valencia, and M. J. Zimic**. 2009. Effect of pyrazinamidase activity on pyrazinamide resistance in *Mycobacterium tuberculosis*. Tuberculosis (Edinb) **89**:109–113.
178. **Shenoi, S., and G. Friedland**. 2009. Extensively drug-resistant tuberculosis: a new face to an old pathogen. Annu. Rev. Med. **60**:307–320.
179. **Sherman, D. R., P. J. Sabo, M. J. Hickey, T. M. Arain, G. G. Mahairas, Y. Yuan, C. E. Barry 3rd, and C. K. Stover**. 1995. Disparate responses to oxidative stress in saprophytic and pathogenic mycobacteria. Proc. Natl. Acad. Sci. U.S.A. **92**:6625–6629.
180. **Sikora, A. L., B. A. Frankel, and J. S. Blanchard**. 2008. Kinetic and chemical mechanism of arylamine N-acetyltransferase from *Mycobacterium tuberculosis*. Biochemistry **47**:10781–10789.

181. **Silva, P. E., F. Bigi, M. P. Santangelo, M. I. Romano, C. Martín, A. Cataldi, and J. A. Aínsa**. 2001. Characterization of P55, a multidrug efflux pump in Mycobacterium bovis and *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. **45**:800–804.
182. **Singh, M., G. P. S. Jadaun, Ramdas, K. Srivastava, V. Chauhan, R. Mishra, K. Gupta, S. Nair, D. S. Chauhan, V. D. Sharma, K. Venkatesan, and V. M. Katoch**. 2011. Effect of efflux pump inhibitors on drug susceptibility of ofloxacin resistant *Mycobacterium tuberculosis* isolates. Indian J. Med. Res. **133**:535–540.
183. **Sirgel, F. A., R. M. Warren, E. M. Streicher, T. C. Victor, P. D. Van Helden, and E. C. Böttger**. 2012. gyrA Mutations and Phenotypic Susceptibility Levels to Ofloxacin and Moxifloxacin in Clinical Isolates of *Mycobacterium tuberculosis*. J. Antimicrob. Chemother.
184. **Slayden, R. A., and C. E. Barry 3rd**. 2000. The genetics and biochemistry of isoniazid resistance in *Mycobacterium tuberculosis*. Microbes Infect. **2**:659–669.
185. **Sousa, A., S. Magalhães, and I. Gordo**. 2012. Cost of antibiotic resistance and the geometry of adaptation. Mol. Biol. Evol. **29**:1417–1428.
186. **Springer, B., K. Lucke, R. Calligaris-Maibach, C. Ritter, and E. C. Böttger**. 2009. Quantitative Drug Susceptibility Testing of *Mycobacterium tuberculosis* by Use of MGIT 960 and EpiCenter Instrumentation. J. Clin. Microbiol. **47**:1773–1780.
187. **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.
188. **Sreevatsan, S., K. E. Stockbauer, X. Pan, B. N. Kreiswirth, S. L. Moghazeh, W. R. Jacobs Jr, A. Telenti, and J. M. Musser**. 1997. Ethambutol resistance in *Mycobacterium tuberculosis*: critical role of *embB* mutations. Antimicrob. Agents Chemother. **41**:1677–1681.
189. **Srivastava, K., R. Das, P. Jakhmola, P. Gupta, D. S. Chauhan, V. D. Sharma, H. B. Singh, A. S. Sachan, and V. M. Katoch**. 2004. Correlation of mutations detected by INNO-LiPA with levels of rifampicin resistance in *Mycobacterium tuberculosis*. Indian J. Med. Res. **120**:100–105.
190. **Srivastava, S., A. Ayyagari, T. N. Dhole, K. K. Nyati, and S. K. Dwivedi**. 2009. emb nucleotide polymorphisms and the role of *embB*306 mutations in *Mycobacterium tuberculosis* resistance to ethambutol. Int. J. Med. Microbiol. **299**:269–280.
191. **Stanley, R. E., G. Blaha, R. L. Grodzicki, M. D. Strickler, and T. A. Steitz**. 2010. The structures of the anti-tuberculosis antibiotics viomycin and capreomycin bound to the 70S ribosome. Nat. Struct. Mol. Biol. **17**:289–293.
192. **Stephan, J., C. Mailaender, G. Etienne, M. Daffé, and M. Niederweis**. 2004. Multidrug Resistance of a Porin Deletion Mutant of *Mycobacterium smegmatis*. Antimicrob Agents Chemother **48**:4163–4170.
193. **Stoffels, K., V. Mathys, M. Fauville-Dufaux, R. Wintjens, and P. Bifani**. 2012. Systematic Analysis of Pyrazinamide-Resistant Spontaneous Mutants and Clinical Isolates of *Mycobacterium tuberculosis*. Antimicrobial agents and chemotherapy.
194. **Storz, G., and S. Altuvia**. 1994. OxyR regulon. Meth. Enzymol. **234**:217–223.
195. **Storz, G., L. A. Tartaglia, and B. N. Ames**. 1990. The OxyR regulon. Antonie Van Leeuwenhoek **58**:157–161.
196. **Strauss, O. J., R. M. Warren, A. Jordaan, E. M. Streicher, M. Hanekom, A. A. Falmer, H. Albert, A. Trollip, E. Hoosain, P. D. van Helden, and T. C. Victor**. 2008. Spread of a low-fitness drug-resistant *Mycobacterium tuberculosis* strain in a setting of high human immunodeficiency virus prevalence. J. Clin. Microbiol. **46**:1514–1516.
197. **Streicher, E. M., I. Bergval, K. Dheda, E. C. Böttger, N. C. Gey van Pittius, M. Bosman, G. Coetzee, R. M. Anthony, P. D. van Helden, T. C. Victor, and R. M. Warren**. 2012. *Mycobacterium tuberculosis* population structure determines the outcome of genetics-based second-line drug resistance testing. Antimicrob. Agents Chemother. **56**:2420–2427.

198. **Streicher, E. M., T. C. Victor, G. van der Spuy, C. Sola, N. Rastogi, P. D. van Helden, and R. M. Warren**. 2007. Spoligotype signatures in the *Mycobacterium tuberculosis* complex. J. Clin. Microbiol. **45**:237–240.
199. **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.
200. **Sultana, R., K. Tanneeru, and L. Guruprasad**. 2011. The PE-PPE domain in mycobacterium reveals a serine α/β hydrolase fold and function: an in-silico analysis. PLoS ONE **6**:e16745.
201. **Sun, G., T. Luo, C. Yang, X. Dong, J. Li, Y. Zhu, H. Zheng, W. Tian, S. Wang, C. E. Barry, J. Mei, and Q. Gao**. 2012. Dynamic population changes in *Mycobacterium tuberculosis* during acquisition and fixation of drug resistance in patients. J. Infect. Dis.
202. **Sun, Z., J. Zhang, X. Zhang, S. Wang, Y. Zhang, and C. Li**. 2008. Comparison of gyrA gene mutations between laboratory-selected ofloxacin-resistant *Mycobacterium tuberculosis* strains and clinical isolates. Int. J. Antimicrob. Agents **31**:115–121.
203. **Sundaram, S., H. Karakaya, D. J. Scanlan, and N. H. Mann**. 1998. Multiple oligomeric forms of glucose-6-phosphate dehydrogenase in cyanobacteria and the role of OpcA in the assembly process. Microbiology (Reading, Engl.) **144 (Pt 6)**:1549–1556.
204. **Suzuki, Y., C. Katsukawa, A. Tamaru, C. Abe, M. Makino, Y. Mizuguchi, and H. Taniguchi**. 1998. Detection of kanamycin-resistant *Mycobacterium tuberculosis* by identifying mutations in the 16S rRNA gene. J. Clin. Microbiol. **36**:1220–1225.
205. **Tahlan, K., R. Wilson, D. B. Kastrinsky, K. Arora, V. Nair, E. Fischer, S. W. Barnes, J. R. Walker, D. Alland, C. E. Barry 3rd, and H. I. Boshoff**. 2012. SQ109 targets MmpL3, a membrane transporter of trehalose monomycolate involved in mycolic acid donation to the cell wall core of *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. **56**:1797–1809.
206. **Takiff, H. E., M. Cimino, M. C. Musso, T. Weisbrod, R. Martinez, M. B. Delgado, L. Salazar, B. R. Bloom, and W. R. Jacobs Jr**. 1996. Efflux pump of the proton antiporter family confers low-level fluoroquinolone resistance in *Mycobacterium smegmatis*. Proc. Natl. Acad. Sci. U.S.A. **93**:362–366.
207. **Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar**. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28**:2731–2739.
208. **Taniguchi, H., H. Aramaki, Y. Nikaido, Y. Mizuguchi, M. Nakamura, T. Koga, and S. Yoshida**. 1996. Rifampicin resistance and mutation of the *rpoB* gene in *Mycobacterium tuberculosis*. FEMS Microbiol. Lett. **144**:103–108.
209. **Telenti, A.** 1998. Genetics and pulmonary medicine. 5. Genetics of drug resistant tuberculosis. Thorax **53**:793–797.
210. **Telenti, A., P. Imboden, F. Marchesi, D. Lowrie, S. Cole, M. J. Colston, L. Matter, K. Schopfer, and T. Bodmer**. 1993. Detection of rifampicin-resistance mutations in *Mycobacterium tuberculosis*. Lancet **341**:647–650.
211. **Telenti, A., W. J. Philipp, S. Sreevatsan, C. Bernasconi, K. E. Stockbauer, B. Wieles, J. M. Musser, and W. R. Jacobs Jr**. 1997. The emb operon, a gene cluster of *Mycobacterium tuberculosis* involved in resistance to ethambutol. Nat. Med. **3**:567–570.
212. **Toprak, E., A. Veres, J.-B. Michel, R. Chait, D. L. Hartl, and R. Kishony**. 2012. Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. Nat. Genet. **44**:101–105.
213. **Trindade, S., A. Sousa, K. B. Xavier, F. Dionisio, M. G. Ferreira, and I. Gordo**. 2009. Positive epistasis drives the acquisition of multidrug resistance. PLoS Genet. **5**:e1000578.
214. **Uplekar, S., B. Heym, V. Friocourt, J. Rougemont, and S. T. Cole**. 2011. Comparative Genomics of esx Genes from Clinical Isolates of *Mycobacterium tuberculosis* Provides Evidence for Gene Conversion and Epitope Variation ▽. Infect Immun **79**:4042–4049.

215. **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.
216. **Van Ingen, J., R. Aarnoutse, G. de Vries, M. J. Boeree, and D. van Soolingen**. 2011. Low-level rifampicin-resistant *Mycobacterium tuberculosis* strains raise a new therapeutic challenge. Int. J. Tuberc. Lung Dis. **15**:990–992.
217. **Van Klingeren, B., M. Dessens-Kroon, T. van der Laan, K. Kremer, and D. van Soolingen**. 2007. Drug Susceptibility Testing of *Mycobacterium tuberculosis* Complex by Use of a High-Throughput, Reproducible, Absolute Concentration Method. J Clin Microbiol **45**:2662–2668.
218. **Van Soolingen, D., K. Kremer, and E. Vynycky**. 2002. New Perspectives in the Molecular Epidemiology of Tuberculosis, p. 17–45. *In* S.H.E. Kaufmann, and H. Hahn (eds.), Issues in Infectious Diseases. KARGER, Basel.
219. **Velayati, A. A., M. R. Masjedi, P. Farnia, P. Tabarsi, J. Ghanavi, A. H. Ziazarifi, and S. E. Hoffner**. 2009. Emergence of new forms of totally drug-resistant tuberculosis bacilli: super extensively drug-resistant tuberculosis or totally drug-resistant strains in iran. Chest **136**:420–425.
220. **Vetting, M. W., L. P. S de Carvalho, M. Yu, S. S. Hegde, S. Magnet, S. L. Roderick, and J. S. Blanchard**. 2005. Structure and functions of the GNAT superfamily of acetyltransferases. Arch. Biochem. Biophys. **433**:212–226.
221. **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.
222. **Warren, R. M., E. M. Streicher, N. C. Gey van Pittius, B. J. Marais, G. D. van der Spuy, T. C. Victor, F. Sirgel, P. R. Donald, and P. D. van Helden**. 2009. The clinical relevance of Mycobacterial pharmacogenetics. Tuberculosis (Edinb) **89**:199–202.
223. **Warren, R., M. de Kock, E. Engelke, R. Myburgh, N. Gey van Pittius, T. Victor, and P. van Helden**. 2006. Safe *Mycobacterium tuberculosis* DNA extraction method that does not compromise integrity. J. Clin. Microbiol. **44**:254–256.
224. **WHO**. 2010. WHO | Guidelines for treatment of tuberculosis, fourth edition.
225. **Winder, F. G., and P. B. Collins**. 1970. Inhibition by isoniazid of synthesis of mycolic acids in *Mycobacterium tuberculosis*. J. Gen. Microbiol. **63**:41–48.
226. **Wong, S. Y., J. S. Lee, H. K. Kwak, L. E. Via, H. I. M. Boshoff, and C. E. Barry**. 2011. Mutations in gidB Confer Low-Level Streptomycin Resistance in *Mycobacterium tuberculosis*. Antimicrobial Agents and Chemotherapy **55**:2515–2522.
227. **World Health Organization**. 2010. Multidrug and extensively drug resistant TB (M/XDR-TB) 2010 GLOBAL REPORT ON SURVEILLANCE AND RESPONSE.
228. **Yew, W. W., and C. C. Leung**. 2008. Management of multidrug-resistant tuberculosis: Update 2007. Respirology **13**:21–46.
229. **Zahrt, T. C., J. Song, J. Siple, and V. Deretic**. 2001. Mycobacterial FurA is a negative regulator of catalase-peroxidase gene *katG*. Mol. Microbiol. **39**:1174–1185.
230. **Zaunbrecher, M. A., R. D. Sikes Jr, B. Metchock, T. M. Shinnick, and J. E. Posey**. 2009. Overexpression of the chromosomally encoded aminoglycoside acetyltransferase eis confers kanamycin resistance in *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. U.S.A. **106**:20004–20009.
231. **Zeng, L., and S. Jin**. 2003. *aph(3')-IIb*, a gene encoding an aminoglycoside-modifying enzyme, is under the positive control of surrogate regulator HpaA. Antimicrob. Agents Chemother. **47**:3867–3876.
232. **Zhang, Y., B. Heym, B. Allen, D. Young, and S. Cole**. 1992. The catalase-peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*. Nature **358**:591–593.

233. **Zhang, Y., and W. W. Yew**. 2009. Mechanisms of drug resistance in *Mycobacterium tuberculosis*. Int. J. Tuberc. Lung Dis. **13**:1320–1330.
234. **Zheng, H., L. Lu, B. Wang, S. Pu, X. Zhang, G. Zhu, W. Shi, L. Zhang, H. Wang, S. Wang, G. Zhao, and Y. Zhang**. 2008. Genetic basis of virulence attenuation revealed by comparative genomic analysis of *Mycobacterium tuberculosis* strain H37Ra versus H37Rv. PLoS ONE **3**:e2375.
235. 2007. Extensively drug-resistant tuberculosis--United States, 1993-2006. MMWR Morb. Mortal. Wkly. Rep. **56**:250–253.
236. WHO | Multidrug and extensively drug-resistant TB (M/XDR-TB): 2010 global report on surveillance and response. WHO.