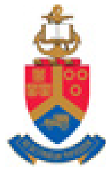


The evolution of drug resistance in the different lineages of the *Mycobacterium tuberculosis* complex.

Hleliwe Hlanze (u17148597)

Submitted in partial fulfilment of the degree:

BSc(Masters) Bioinformatics

Department of Biochemistry, Genetics and Microbiology, University of Pretoria

January 2024

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Declaration

I declare that the mini-dissertation/dissertation/thesis, which I hereby submit for the degree of Hleliwe Hlanze at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at another university. Where secondary material is used, this has been carefully acknowledged and referenced in accordance with university requirements. I am aware of university policy and implications regarding plagiarism.

SIGNATURE: DATE: 01/15/2024

© University of Pretoria

Acknowledgments

I would like to thank God for the strength and wisdom He gave me in the process of completing my MSc.

To my family thank you for your support, love, encouragement, and understanding from the start of my journey with this postgraduate programme.

I would also like to thank my supervisor Professor Oleg Reva for his guidance, mentorship patience, and sacrifice in helping me with my degree.

To my colleagues at the Center for Bioinformatics and Computational Biology, thank you for your encouragement and help when times were challenging.

Lastly, I would like to thank the NRF for funding this project as well as the University of Pretoria.

© University of Pretoria

Summary

The eradication of drug resistance (DR) in *Mycobacterium tuberculosis* (Mtb) has been proven to be challenging. Several DR mechanisms that are not fully understood have been suggested to play significant roles in the acquisition, maintenance, and spread of DR-Mtb strains. A DR mechanism of interest in this study was epistasis; an important evolutionary process suggested to play a role in the evolution of DR in Mtb. This mechanism of interest results in co-adapted alleles in the Mtb population meaning that the acquisition of new mutations are not independent of each other. Our study aimed at modelling evolutionary trajectories of DR in the human adapted lineages of the *M. tuberculosis* complex. To accomplish this an extensive search of sequence data in public databases as well as metadata was done. The data then went through quality analyses that yielded a dataset of 9388 sequences that represented the diversity of Mtb isolates from different lineages. Additionally, statistical analysis identified significant non-random associations between linked common and lineage specific DR mutations. This analysis enabled the identification of epistatic interactions among pre-requisite/stepping stone mutations, intermediate mutations, DR mutations, and compensatory mutations, all of which were suggested to play a role in the evolution of DR in Mtb. Lastly, evolutionary trajectories towards DR were modelled for the overall Mtb population and the different lineages

Identifying linked common associations resulted in the identification of 64 universal markers, in functional and non-functional genes, of DR to more than three antibiotics. Functional mutations included a DR mutation and other mutations involved in key biological processes such as lipid metabolism, and redox metabolism suggested to have a strong association with multiple drug resistance. Non-functional genes were largely comprised of highly mutable *pe* and *ppe* genes. Associations among these markers were then determined to investigate the likelihood of the occurrence of a "mutator phenotype" necessary for the acquisition of a multiple drug resistance phenotype. Indeed when these associations were modelled they demonstrated the occurrence of a "mutator" phenotype as a result of PE_PGRS genes before the acquisition of multiple drug resistance.

Further analysis was done to determine lineage specific non-random associations between polymorphisms in the various lineages. Epistatic interactions among polymorphisms formed multiple evolutionary pathways that resulted in various evolutionary trajectories. Several evolutionary paths within the various lineages were established and showed the importance of the various categories of mutations and their role in DR acquisition and compensation. Several

© University of Pretoria

pre-requisite mutations in genes of known function and hypotheticals were suggested to be important in the acquisition of high level resistance. Intermediate mutations were also suggested to play a role in the acquisition of several evolutionary trajectories. The study also identified several pairs of DR mutations that formed positive epistatic interactions promoting further resistance to additional drugs. Compensatory mutations in several genes were suggested to be important in the amelioration of fitness costs as a consequence of DR acquisition. Ultimately these results showed the importance of epistasis in the evolution of DR in Mtb.

This study then concluded by emphasising the importance of understanding evolutionary trajectories towards DR in Mtb because this would lead to the improvement of diagnostics, treatment outcomes, and the effective use of antibiotics. The study also highlighted the importance of further investigation into inter-lineage diversity and its role in the evolution of DR. Furthermore, much progress has been made in the fight against the tuberculosis disease, however, significant gaps remain in our understanding of the pathogenesis of this bacterium.

© University of Pretoria

Table of contents

LIST OF ABBREVIATIONS AND ACRONYMS

LIST OF TABLES

LIST OF FIGURES

PUBLICATION

DEFINITIONS

KEYWORDS

Chapter1: Introduction....1

1.1 Background....1

1.2 Problem statement....2

1.3 Work hypothesis and statement of objectives....3

1.4 Chapter overviews....3

Chapter2: Literature review....5

2.1 Introduction....5

2.2 Brief history of TB historical basis of TB treatment, and resistance....7

2.2.1 The discovery of anti-tuberculous drugs....7

2.2.2 The emergence of MDR and XDR....8

2.2.3 Totally drug resistant TB....9

2.3 TB treatment 10

2.3.1 First line drugs....12

2.3.2 Second line drugs....13

2.4 Challenges with the treatment of TB....16

2.4.1 Drug resistance development....17

2.4.2 Patient adherence....17

© University of Pretoria

2.4.3 Mtb's state of dormancy……………………………………………18

2.4.4 Development of new drugs…………………………………………18

2.4.5 TB vaccine……………………………………………………...19

2.5 TB diagnosis…………………………………………………………19

2.5.1 Is whole genome sequencing an alternative to TB diagnosis?.........19

2.6 Genetic diversity of *Mycobacterium tuberculosis*……………………….21

2.6.1 Lineage 1……………………………………………………..22

2.6.2 Lineage 2……………………………………………………..22

2.6.3 Lineage 3……………………………………………………..23

2.6.4 Lineage 4……………………………………………………..23

2.6.5 Lineage 5 and Lineage 6……………………………………..23

2.6.6 Lineage 7……………………………………………………..23

2.6.7 Lineages 8 and 9……………………………………………..24

2.7 The evolution of drug resistance in Mtb………………………………24

2.7.1 Intrinsic resistance…………………………………………..26

2.7.2 Acquired drug resistance……………………………………...29

2.7.3 Epistasis……………………………………………………...30

2.7.4 Fitness cost and Compensatory evolution in essence are part of sign epistasis……………………………………………………34

2.7.5 Clonal interference………………………………………….36

2.8 Conclusion………………………………………………………….37

Chapter 3: Methods and Materials………………………………………38

3.1 Introduction………………………………………………………..38

3.2 Data collection……………………………………………………...39

3.3 Mapping and variant calling………………………………………41

© University of Pretoria

3.4 Split package…………………………………………………………………42

3.4.1 Identify_polymorphism.py………………………………………42

3.4.2 Fastq2vcf.py……………………………………………………43

3.4.3 Annotation.py…………………………………………………44

3.4.4 Split.py…………………………………………………………45

3.4.5 Risk.py…………………………………………………………46

3.5 Data visualization…………………………………………………………..48

Chapter 4: Identification of Universal lineage-independent markers of multidrug resistance in *M. tuberculosis*……………………………………………………………49

4.1 Introduction…………………………………………………………………..49

4.2 Universal markers of antibiotic resistance in functional genes………………49

4.3 Mutations in PE/PPE and PE_PGRS genes and their associations with multidrug resistance…………………………………………………………………..53

4.4 Epistatic interactions among universal markers of drug resistance in *Mycobacterium tuberculosis*…………………………………………………………..57

Chapter 5: Identification of lineage-specific markers of antibiotic resistance and reconstruction of epistatic interactions between them……………………………………59

5.1 Introduction…………………………………………………………………..59

5.2 Epistatic interactions specific for Lineage 1.2…………………………………63

5.3 Epistatic interactions specific for Lineage 2……………………………………65

5.4 Epistatic interactions specific for Lineage 3……………………………………67

5.5 Epistatic interactions specific for Lineage 4……………………………………68

5.6 Epistatic interactions specific for Lineage 4.1…………………………………69

5.7 Epistatic interactions specific for Lineage 4.1(Haarlem)……………………...70

5.8 Epistatic interactions specific for Lineage 4.2(Ural)…………………………..72

5.9 Epistatic interactions specific for Lineage 4.6(Uganda)………………………73

© University of Pretoria

Chapter 6: Discussion and Conclusion……………………………………………………..75

6.1 Discussion…………………………………………………………………..75

6.2 Conclusive remark………………………………………………………....78

References……………………………………………………………………………...80

Appendix…………………………………………………………………………….....95

© University of Pretoria

Abbreviations:

AMK: amikacin

CFZ: clofazimine

CP: capreomycin

CS: cycloserine

DOTS: directly observed therapy

DS: drug susceptible

DR: drug resistant/resistance

EMB: ethambutol

ETH: ethionamide

FQ: fluoroquinolones

INH: isoniazid

KAN: kanamycin

LFX: levofloxacin

LZD: linezolid

MXF: moxifloxacin

MtbC: mycobacterium tuberculosis complex

OFX: ofloxacin

PAS: para-aminoglycosides

PTH: prothionamide

PZA: pyrazinamide

STM: streptomycin

TB: tuberculosis

RIF: rifampicin

© University of Pretoria

WGS: whole genome sequencing

WHO: World Health Organization

XDR-TB: extensively drug-resistant tuberculosis

XXDR-TB: extremely drug resistant tuberculosis

Acronyms:

PATRICK: Pathosystems Resource Integration Center

List of tables:

Table 2: Antibiotics used to treat tuberculosis together with their modes of action and drug resistance mutations.

Table 3: Numbers of isolates from each lineage of *M. tuberculosis* collected for the analysis.

Table 4.1: Universal markers of drug resistance in *M. tuberculosis* found in functional genes.

Table 4.2: Universal markers of drug resistance in *M. tuberculosis* found in PE/PPE and non-functional genes.

Table 5: The most discriminative polymorphisms that distinguish between resistant and susceptible strains in various *Mycobacterium tuberculosis* lineages.

List of figures:

Figure 2.1: The geographical distribution of the different lineages of *Mycobacterium tuberculosis* complex.

Figure 2.2: Acquired and intrinsic mechanisms of resistance found in *Mycobacterium tuberculosis*.

Figure 2.3: Figure adapted from Trauner et al. (2014) showing putative epistatic interactions among genes found within *M. tuberculosis* genomes.

Figure 2.4: The different forms of epistasis and their effect on fitness (↑ high fitness) (↓ low fitness).

Figure 2.5: Shows the evolutionary trajectory of DR from the acquisition of mutations to clonal interference.

Figure 3.1: Analytical workflow of the project.

© University of Pretoria

Figure 3.2: The command prompt interface of the script *identify_polymorphisms*

Figure 3.3: Command prompt interface of *fastq2vcf*.

Figure 3.4: The command prompt interface of the script annotation.

Figure 3.5: The command prompt interface of the script *split*.

Figure 3.6: The command prompt interface of the script *risk*.

Figure 3.7: Two pairs of polymorphisms with the first pair showing omnidirectional dependency with the likelihood of a drug resistance (DR) mutation (KatG R463L) occurring when a mutation has already occurred in a secondary polymorphic site (stp D69Y). The likelihood of a mutation occurring in the opposite direction is less likely ($B \rightarrow b_{|a} < 0.5$ and $A \rightarrow a_{|b} > 0.5$). The second pair shows bidirectional dependency with the likelihood of a DR mutation occurring when a mutation has already occurred in a secondary polymorphic site (PPE35L896S) and a mutation occurring in a secondary polymorphic site when a mutation has already occurred in the DR mutation (both $B \rightarrow b_{|a}$ and $A \rightarrow a_{|b} > 0.5$). This indicates genetic drift and such pairs were removed from the analysis.

Figure 4.1: Epistatic interactions among mutations in global markers of drug resistance. The green node represents pre-requisite mutations, the blue nodes are intermediate mutations, and the pink nodes are final compensatory mutations. Epistatic links to known DR mutations are highlighted by orange edges. Nodes are labelled by the respective gene name and codon position number following the colon.

Figure 5.1: Epistatic interactions showing various evolutionary paths towards drug resistance (DR) in lineage 1.2. The green nodes represent a pre-requisite mutation, the blue nodes an intermediate mutation, the orange nodes represent DR mutations, and the pink nodes and final/compensatory mutation. Epistatic links to and from known DR mutations are highlighted by orange edges. Nodes are labeled by the respective gene name and codon position number following the horizontal line.

Figure 5.2: Epistatic interactions showing various evolutionary paths towards drug resistance (DR) in lineage 2. The green nodes represent a pre-requisite mutation, the blue nodes an intermediate mutation, the orange nodes represent DR mutations, and the pink nodes and final/compensatory mutation. Epistatic links to and from known DR mutations are highlighted

© University of Pretoria

by orange edges. Nodes are labelled by the respective gene name and codon position number following the horizontal line.

Figure 5.3: Epistatic interactions showing various evolutionary paths towards drug resistance (DR) in lineage 3. The green nodes represent a pre-requisite mutation, the blue nodes an intermediate mutation, the orange nodes represent DR mutations, and the pink nodes and final/compensatory mutation. Epistatic links to and from known DR mutations are highlighted by orange edges. Nodes are labelled by the respective gene name and codon position number following the horizontal line.

Figure 5.4: Epistatic interactions showing the mutational path in lineage 4. The orange node represents a drug resistance (DR) mutation and the pink node a final/compensatory mutation. The epistatic interaction from the known DR mutation is highlighted by the orange edge. Nodes are labelled by the respective gene name and codon position number following the horizontal line.

Figure 5.5: Epistatic interactions showing various evolutionary paths towards drug resistance (DR) in lineage 4.1. The green nodes represent a pre-requisite mutation, the blue nodes an intermediate mutation, the orange nodes represent DR mutations, and the pink nodes and final/compensatory mutation. Epistatic links to and from known DR mutations are highlighted by orange edges. Nodes are labelled by the respective gene name and codon position number following the horizontal line.

Figure 5.6: Epistatic interactions showing various evolutionary paths towards drug resistance (DR) in lineage 4.1(Haarlem). The blue nodes are intermediate mutations, the orange nodes represent DR mutations and/or pre-requisite mutations, and the pink nodes are final/compensatory mutations. Epistatic links to and from known DR mutations are highlighted by orange edges. Nodes are labelled by the respective gene name and codon position numbers following the horizontal line.

Figure 5.7: Epistatic interactions showing various evolutionary paths towards drug resistance (DR) in lineage 4.2 (Ural). The green nodes represent a pre-requisite mutation, the blue nodes an intermediate mutation, the orange nodes represent DR mutations, and the pink nodes and final/compensatory mutation. Epistatic links to and from known DR mutations are highlighted by orange edges. Nodes are labelled by the respective gene name and codon position number following the horizontal line.

© University of Pretoria

Figure 5.8: Epistatic interactions showing various evolutionary paths towards drug resistance (DR) in lineage 4.6 (Uganda). The green nodes represent a pre-requisite mutation, the blue nodes an intermediate mutation, the orange nodes represent DR mutations, and the pink nodes and final/compensatory mutation. Epistatic links to and from known DR mutations are highlighted by orange edges. Nodes are labelled by the respective gene name and codon position number following the horizontal line.

Publication: The Epistatic Landscape of Antibiotic Resistance of Different Clades of *Mycobacterium tuberculosis*.

Definitions:

Acquired drug resistance: drug resistance that occurs either through mutations or horizontal gene transfer mediated by phages, plasmids, or transposon elements

Attributable risk statistic:

Dormancy: refers to a non-replicating state in *Mycobacterium tuberculosis* with low-to-absent metabolic activity.

Epistasis: is a phenomenon where the effect of one mutation depends on the presence of another mutation in the same genome thus affecting the organism's phenotype.

Evolutionary path: a sequences of genetic changes that form chains in which each step facilitates the next, favouring, step by step, a significant increase in resistance.

Evolutionary trajectory: refers to the order of changes that result in the evolution of one organismal (phenotypic or genotypic) state from another.

Negative epistasis: the fitness of a double mutant is lower than expected consequently constraining the evolution towards drug resistance by aggravating its cost.

Hetero-resistance: occurs when some bacterial populations are susceptible and others are resistant.

Latency: refers to a state of asymptomatic infection without apparent disease.

Linkage disequilibrium: A measure of whether alleles at two loci coexist in a population in a non-random fashion. Alleles that are in LD are found together on the same haplotype more often than would be expected under a random combination of alleles.

© University of Pretoria

Pathogenesis: describes the mechanisms by which a disease develops, progresses, and either persists or is resolved.

Persistence: bacterial cells that are not susceptible to the drug, but do not possess resistance genes.

Positive epistasis: the fitness of a double mutant is higher than expected consequently promoting the evolution of drug resistance while minimising its cost.

Primary resistance: resistance that can be transmitted from person to person.

Secondary resistance: resistance bacteria acquire through DR mutations such as Single Nucleotide Polymorphisms and insertions and/or deletions

Sign epistasis: when a mutation may be beneficial or deleterious depending on the genetic background.

Virulence: damage that TB infection causes to the host.

Keywords: *Mycobacterium tuberculosis*, epistasis, evolution, drug resistance, compensatory evolution, pre-requisite mutations, lineage.

© University of Pretoria

Chapter1: Introduction

1.1 Background

DR Tom Frieden, the former director of the Centers for Disease Control and Prevention, warned, "If we are not careful, we will soon be in a post-antibiotic era. And for some patients and for some microbes, we are already there."(1). Currently, we have a global problem of microbes that are resistant to almost every antibiotic available for treatment (2,3). Yes, the discovery of antibiotics revolutionized the treatment of bacterial infectious diseases enabling the eradication of deadly pathogens, unfortunately, it is well-established that antibiotic use results in the evolution of drug resistance (DR) (4,5). The main drivers of antibiotic resistance have been due to the misuse and overuse of antibiotics; social, political, and economic factors affecting access to proper treatment, as well as the lack of new antibiotics (2,6-8).

Additionally, there are a wide variety of biochemical and physiological mechanisms that may be responsible for the emergence of resistance (4). The lack of knowledge of the complex processes that contribute to the emergence of resistance is one of the many reasons why we struggle to prevent and treat resistant infectious diseases (4). Thus, it is of utmost importance that we understand the pathogenesis of bacteria and the mechanisms with which they can, evolve towards DR, sustain transmission, and evade fitness costs associated with DR; to slow the evolution of pathogens and effectively control their spread (7,9).

Mycobacterium tuberculosis (Mtb) is an example of such a microbe whose eradication has been made difficult due to the emergence of resistant strains (7). Drug resistant tuberculosis (DR-TB) has undermined the progress made over the years in controlling the TB epidemic (10). The treatment of this disease has burdened healthcare systems and resulted in major economic repercussions and societal stigmas (4). The non-adherence to treatment and poor functioning healthcare systems have often been attributed to the emergence of TB resistance, however even in well-functioning public healthcare systems where patients adhere to treatment the evolution of DR still occurs (8). Additionally, the TB microbe has a low mutation rate, a highly clonal genome, and a limited genetic diversity yet, it is a successful pathogen that demonstrates a remarkable adaptive potential (8,11,12). This suggests that the evolution of DR is more complex than previously assumed and that the genetic basis of DR, which is poorly understood, could explain why successful DR-TB strains emerge (8,11-13).

© University of Pretoria

Antituberculous drugs target essential functions, meaning that the occurrence of a resistance mutation would result in a fitness cost (3). In light of this, how does Mtb harbour multiple DR mutations while retaining the ability to transmit, infect, persist, and cause disease (3)? Mtb clonal genome favours a mechanism that could explain the evolutionary trajectory towards DR. Epistasis is a mechanism that involves the spread of co-adapted alleles in an Mtb population that may contribute to the evolutionary success of these resistant strains (7). Compensatory evolution is another mechanism that involves the compensation of fitness costs as a result of DR mutations; believed to be important in the successful spread, persistence, and maintenance of DR in Mtb (14,15).

The genetic diversity of Mtb, causing TB in humans, consists of human-adapted lineages suggested to differ in their geographic distribution, biological fitness, virulence, and propensity to acquire DR (7,16). Compensatory evolution, the varied frequencies of mutations among lineages, the co-occurrence of highly resistant mutations in successful clinical isolates, and lineage specific DR mechanisms suggest the occurrence of epistasis (8,17,18). Hence, epistatic interactions could also explain how in the context of its genetic diversity Mtb can modulate the evolution of DR (19,20). Understanding how antibiotic resistance evolves and spreads is key to improving antibiotic treatment (15).

1.2 Problem statement

Mycobacterium tuberculosis (Mtb) has been in existence for centuries and its eradication has proven to be a challenge. Mtb has resulted in more deaths than any other infectious disease and is currently the most deadly pathogen from a single infectious agent. Drug resistance (DR) has resulted in the loss of many lives, burdened healthcare systems, and challenged economies. The discovery of antibiotics and the use of combination therapy to treat TB seemed to limit the evolution of DR. However, with this preventative measure put in place, resistance still occurred. Mtb's biology is unique and at face value suggests that it would not be a problematic pathogen to eradicate. However, this bacterium's persistence continues to baffle scientists and healthcare practitioners. Mtb employs several mechanisms to evade its hosts' defence mechanisms, most of which are not clearly understood. To combat the problem of the emergence of resistant Mtb strains; DR mechanisms need to be clearly understood.

There are several mechanisms suggested to result in resistance but, it is not clearly understood what determines the evolutionary trajectory towards resistance in Mtb. The genetic diversity of Mtb, though limited, is suggested to play an important role in the emergence of resistance.

© University of Pretoria

Understanding the path towards resistance, while considering the genetic diversity of DR, is important because the evolution of DR can potentially be predicted and treatment regimens adjusted to limit the evolution of DR. Extrinsic factors that result in the emergence of DR are more clearly defined, however, the genetic basis of resistance in Mtb is not so clear. Epistasis is a phenomenon suggested to play a significant role in the evolutionary trajectory towards resistance and the experimental evidence for this is lacking therefore the statistical validation of association between mutations may be key in determining the steps Mtb takes to become resistant, transmit, and thrive in the human host.

1.3 Work hypothesis and statement of objectives

The work hypothesis of this project was that the epistatic interactions between mutations determine the evolutionary trajectory of the development of drug resistance in the human adapted lineages of the *Mycobacterium tuberculosis* complex.

This study aimed to model the evolutionary trajectories of drug resistance in the human-adapted lineages of the *Mycobacterium tuberculosis* complex.

To achieve the aim, the following objectives were defined:

- To create a local database of genome sequences representing different Mtb lineages and provide metadata on their antibiotic resistance. The database will be created using public databases and publications to unbiasedly represent the distribution of drug resistance mutations in different Mtb lineages.
- To determine the statistical significance of non-random associations between linked common and lineage-specific drug resistance mutations.
- To model the evolutionary pathways towards the development of drug resistance in the overall Mtb population and different lineages.

1.4 Chapter overviews

Chapter 1: This chapter provides an overview of the problems this study investigates. It states the purpose of the study and presents how the aim of the study will be accomplished.

Chapter 2: This chapter reviews the existing literature on the various factors that have contributed to the evolution of drug resistance in Mtb. This chapter concludes with research gaps, which are addressed in the study.

© University of Pretoria

Chapter 3: This chapter presents the methodology that is used in the study. It describes how data was collected and the detailed statistical analysis steps used to determine evolutionary trajectories leading to DR in the different Mtb lineages.

Chapter 4: This chapter reports on the discovered universal marker genetic polymorphisms that may predict the likelihood of resistance to several antibiotics irrespective of the lineage designation. It also determines that evolution towards multi-drug resistance begins with a switch to the "mutator" phenotype, which leads to the accumulation of mutations in PE_PGRS genes.

Chapter 5: In this chapter, we report on the evolutionary trajectories toward drug resistance of the various lineages in the *M. tuberculosis* complex.

Chapter 6: This chapter discusses the accomplished objectives of this study and concludes with the implications of our findings.

© University of Pretoria

Chapter 2: Literature Review

2.1 Introduction

A major concern when treating diseases that require antibiotics is the emergence of incurable strains (10). Unfortunately, the use of antibiotics presented a new selection pressure that led to the emergence of drug resistance (DR) (10). Currently, there is a rise of deadly bacteria that are difficult to treat due to the evolution towards DR (3). Tuberculosis (TB) is a curable disease, however, its total eradication is challenged by the emergence of DR and lately by the COVID-19 pandemic (7,21). This has resulted in a global epidemic of DR-TB that has challenged most of the progress made in controlling the TB epidemic (8,22). Since the discovery of the first drugs for TB treatment, a significant amount of progress has been made, but the lack of new drugs, better treatment regimes, limited healthcare resources in low-income countries, poor diagnostics, lack of knowledge on DR mechanisms, and many more factors have resulted in the rise of DR-TB (23-27).

TB is caused by *Mycobacterium tuberculosis* (Mtb) which forms part of the *Mycobacterium tuberculosis* complex (MtbC) (28). This group of Mycobacteria is made up of closely related human and animal-adapted bacterial species and sub-species (28). For this study, the introduction will focus on the human-adapted members of the MtbC, which are *Mycobacterium tuberculosis sensu stricto* and *Mycobacterium africanum* (29). The genetic diversity of Mtb is relatively low, it currently only comprises 9 lineages including two recently proposed lineages 8 and 9 (28,30). Lineage 8 is believed to be an as-yet-unknown ancestral stage between the MtbC and its progenitor *M. cannetii* - like mycobacteria, and lineage 9 seems to be restricted to East Africa (28,30). Mtb is an obligate pathogen that has had a long history of co-evolution with humans (7). The genetic background of Mtb is suggested to influence several aspects of the evolution of DR (8). Particularly, the propensity of some lineages to acquire resistance and the fitness costs associated with the acquisition of DR may vary depending on the genetic background of a strain (8).

The treatment of TB comprises drugs that are used in combination. Drug susceptible (DS) TB is normally treated with a combination of drugs isoniazid (INH), rifampicin (RIF), ethambutol (EMB), and pyrazinamide (PZA) which are administered for 6 months (12). If resistance to INH and RIF occurs then this is termed Multi-Drug-Resistant TB (MDR) (12,31). MDR is treated using fluoroquinolones (FQ) and injectable aminoglycosides (AG) (12). Previously,

© University of Pretoria

XDR was defined as MDR with additional resistance to any FQ and at least one of the second-line injectable drugs (12,31). Currently, XDR is defined as MDR/RR-TB with resistance to any FQ and resistance to one of the drugs levofloxacin (LFX), moxifloxacin (MFX), bedaquiline (BDQ), and linezolid (LZD), which is different from the previous definition of XDR (32,33). In this study, the old definition of XDR is used. The successful treatment of MDR or XDR is low with a success rate of ~54% and ~28% respectively (8). Unfortunately, the ongoing epidemics of MDR and XDR threaten to make TB incurable (24). DR can be acquired (secondary resistance) if bacteria harbour DR mutations, which include Single Nucleotide Polymorphisms (SNPS) and insertions and/or deletions (INDELs) (34,35). DR can also be transmitted from person to person, this is known as primary resistance (26).

The evolution of DR is a result of an interplay of extrinsic (social and economic factors) and intrinsic (acquired and innate DR) factors (27,36,37). The failure to effectively treat TB risks the amplification of resistance to additional antibiotics and in turn, prolonging opportunities for transmission (34). Furthermore, to treat TB effectively there is a need for new diagnostic methods that will provide comprehensive genomic information on the known mechanisms of DR to improve treatment outcomes (38). Whole Genome Sequencing (WGS) is a tool often used in research to study the genomes of various organisms (9). This tool can be an alternative diagnostic device however, the lack of standardization and bioinformatics expertise of this technology limits its use in the clinical setting (39,40).

The successful emergence of DR even in well-functioning healthcare systems suggests that the genetic basis of resistance in Mtb is not fully understood (8). Epistasis is an important evolutionary process and has been suggested to play a role in the evolution of antibiotic resistance (41). Mtb's highly clonal genome favours epistasis in that polymorphisms are in some form of linkage disequilibrium (7,12,14,42). The use of combination therapy in Mtb was intended to limit the evolution of DR and, if resistance mutations were to occur, there would be a fitness cost (23,43). Unfortunately, this notion was challenged by the suggested emergence of compensatory mutations believed to compensate for any fitness cost incurred (41). Compensatory evolution may play a role in the persistence and maintenance of DR mutations in Mtb (26). The relationship between compensatory evolution and epistasis affects the evolution of MDR in Mtb and investigating these factors will aid in better predictions of the future trajectory of MDR-TB (26).

© University of Pretoria

2.2 Brief history of TB historical basis of TB treatment, and resistance

Tuberculosis has a long history of co-evolution with humans (7). The first occurrence of TB can be dated back to Egyptian mummies in 5000 BC (44). Descriptions of this disease can be also found in writings from China, India, and the Bible (44). The TB pathogen is one of the oldest diseases known to affect humans (45). Mtb may have killed more people than any other microbial pathogen (45). This disease reached epidemic proportions in the 18th and 19th centuries in Europe and North America (45). Currently, TB is one of the top ten leading causes of death globally. A brief look into the history of TB and the current challenges in the treatment of this disease brings to light a grim reality that we still have a long way to go before TB is fully eradicated.

2.2.1 The discovery of anti-tuberculous drugs

TB was first isolated by Robert Koch on March 24, 1882, and it would only be about sixty years later that a drug called streptomycin (SM) would be discovered in 1943 that could treat TB (5,46). SM was effective for the first few months, but soon thereafter, there were reports of patients who were resistant to the drug, deeming SM ineffective (8,47). It was then realized through drug trials that monotherapy was not effective resulting in the rapid onset of resistance (5,8). The golden age (1950's – 1960's) of antibiotics led to the discovery of new drugs. During this period new anti-TB drugs namely: isoniazid (1951) (clinical use from 1952), pyrazinamide (1952), cycloserine (1954) (clinical use from 1964), ethionamide (1956) (clinical use from 1965), rifampin/rifampicin (1957) (clinical use from 1966), ethambutol (1962), and fluoroquinolones (1962), were discovered (5,48,49). The discovery of new drugs was welcome, but with every discovery of a new drug; the selection of mutations that conferred resistance to this drug occurred (5). Combinatorial therapy was then introduced by the British Medical Research Council (BMRC) to suppress resistance (5,8). The initial set of drugs used for combinatorial therapy were PAS and SM in 1950, with time as new drugs were discovered, modifications were made to this treatment regimen (8,50). In the early 1970's, a combination of the first-line drugs: INH, RIF, EMB, and PZA, was introduced. Clinical trials conducted by the BMRC proved these drugs to be highly effective against DS-TB, and since the 1980's these drugs have been used to treat DS-TB (3,7). This not only improved treatment outcomes, it also shortened the duration of treatment from 18 months or more to 9 or 6 months (8). There was great optimism in the healthcare community that TB would ultimately be eliminated (5).

© University of Pretoria

2.2.2 The emergence of MDR and XDR.

In the 1950s, Middlebrook and Mitchison conducted studies that hypothesized that DR would always confer a fitness cost due to observations in guinea pigs infected with INH-resistant strains that were less virulent than guinea pigs infected with INH-susceptible strains (40,51-53). This resulted in a central dogma that DR would not challenge the global control of TB and would consequently result in reduced deaths and a decrease in transmission of resistant strains (5,40,54). Additionally, mathematical models suggested that MDR would not be a global problem but a localized one (24,26). It was also believed that MDR strains were not capable to transmit enough to establish a sustainable infection chain (8,55)

There was, however, a notable rise in outbreaks of mono and multidrug resistance, particularly in the 1970s, from various parts of the world, challenging the earlier studies that the fitness cost would be a limiting factor in the evolution of resistant strains (5,40). Even though outbreaks were occurring, especially in low to middle-income countries, disease severity was no longer a major concern as it was in the previous years (5). Additionally, the first outbreaks of MDR-TB mainly occurred with homeless individuals and people co-infected with HIV (56). It was believed that less fit MDR strains would only infect immunocompromised individuals (56). The introduction of combination therapy resulted in a sharp decline in TB cases throughout the world and it was generally thought that TB would no longer be a public health concern (23,43). Consequently, there was a reduction in funding for TB research, which resulted in slowing down drug discovery, development of diagnostics, media coverage of DS or DR-TB, and vaccine research (5,23).

A significant rise in MDR outbreaks, in the mid-1980s to early 1990s, led to a renewed interest in TB (23). In response to this, the WHO, in a press release, declared TB as a global health emergency in 1993 (8,57). A notable outbreak of MDR TB caused by the W strain (Beijing strain) in New York challenged the concepts and the central dogma of decreased transmissibility of MDR strains (54,55,58,59). The WHO introduced the Directly Observed Therapy Strategy (DOTS) to address the rise of MDR (54). This involved the use of sputum smear microscopy to detect the presence or the absence of TB, political commitment to control TB, drugs administered under the observation of a health care professional, uninterrupted supply of drugs, and assessment of treatment results (54,60). However, sputum smear microscopy could not differentiate Mtb from other acid fast-bacteria and lacked sensitivity to detect DR in patients, consequently, in some instances DR patients were treated using first-line

© University of Pretoria

drugs (47,54,61). To improve on this strategy, the DOTS-plus strategy was introduced, which involved the use of second-line drugs to treat MDR (62). Unfortunately, the second-line drug treatment is expensive and patients experience adverse side effects, which makes it difficult for patients to adhere to the treatment (62).

It was just a matter of time before the resistance to second-line drugs would be reported. In 2005, an outbreak of XDR was reported in the Tugela Ferry hospital in KZN, South Africa (54). This XDR strain was sufficiently fit to transmit and cause disease in both immunocompetent and immunocompromised individuals (12,31). Retrospective analysis studies determined that XDR-TB was present in South Africa as early as 1992 in the Western Cape and the XDR-TB strain causing the Tugela outbreak was present from 2001 (37).

2.2.3 Totally drug resistant TB

A study done in Iran in 2009 reported a group of 15 patients with resistance to all anti-TB drugs tested (27,63). The authors of this study used the terms extremely drug-resistant (XXDR)-TB and totally drug-resistant (TDR)-TB (27). Other countries have also reported cases of total drug resistance (TDR) in India, Italy, and South Africa (27,64-66). However, the term 'totally drug resistant' is a debated issue that has not been clearly defined for TB and therefore, not yet recognized by the WHO (16,27). Currently, cases where patients are reported to be resistant to all test TB drugs are defined as XXDR-TB, mainly because in vitro drug susceptibility testing (DST) is technically challenging and has limitations (27). To use the term TDR, a consensus has to be reached regarding the DST protocol for defining MDR and XDR cases. Additionally, a thorough investigation of the appropriate testing methods, resistance minimum inhibitory concentration (MIC) value definitions, and acceptable level of reproducibility of the results must be achieved (27,67). The emergence of XDR and TDR-TB re-emphasizes the need for new drug regimens and/or alternative strategies to combat the spread of DR-TB (16).

Currently, the WHO uses five antibiotic resistance categories of which XDR and pre-XDR are the new definitions, while MDR remains as it was defined before (33). The five resistance categories are INH-resistant TB, RR-TB (Rifampicin resistant), MDR, pre-XDR, and XDR (33). Pre-XDR is RR and MDR with additional resistance to any FQ (32,33).

© University of Pretoria

2.3 TB treatment

The currently recommended treatment for DS-TB is 6 months (4 months INH, RIF, EMB, and PZA; 2 months INH and RIF) (16,27). Resistance to the first line drugs requires 9 – 20 months to treat with a decreased success rate when compared to DS-TB and an even lower success rate when treating XDR-TB (16).

Drugs treating TB can either be bacteriostatic (antibiotics that inhibit growth) or bactericidal (antibiotics that kill bacteria) (68). Among the first-line drugs, RIF, INH, and PZA are bactericidal, whereas EMB is bacteriostatic (27). Most of the second-line drugs exhibit bactericidal activity, although some are only bacteriostatic (27). Many anti-TB drugs are administered as prodrugs including INH, ethionamide (ETH), prothionamide (PTH), pretomanid (PTM), delamanid (DEL), pyrazinamide (PZA), and para-aminosalicylic acid (PAS) (69). "A prodrug is a substance that has no pharmacological effect and needs to be converted into another structural form via enzymatic or chemical reaction, to become active and exhibit therapeutic properties to the body" (69). The list of antibiotics together with their mode of action, target, and main drug resistance mutations can be found in Table 2.1.

Table 2: Antibiotics used to treat tuberculosis together with their modes of action and drug resistance mutations

Drug	**Mode of action**	**Target**	**Main drug resistance Mutation/s**
Isoniazid	Inhibits mycolic acid biosynthesis	Mycolic acid	*katG*
Rifampicin	Inhibits RNA polymerase	RNA polymerase	*rpoB*
Ethambutol	Inhibits cell wall synthesis	Arabinosyl transferases	embCAB operon
Pyrazinamide	Inhibits membrane transport		*pncA*

© University of Pretoria

Ethionamide and prothionamide	Inhibits mycolic acid biosynthesis	mono-oxygenase enzyme	*ethA*, *ethR*, and *inhA*
Fluoroquinolones	Inhibit DNA synthesis	DNA gyrase and topoisomerase IV	*gyrA* and *gyrB*
Para-aminosalicylic acid	Inhibits folate synthesis	Thymidylate synthase, Dihydrofolate synthase, Dihydrofolate reductases(70)	*thyA*
Clofazimine	Releases reactive oxygen species upon re-oxidation by oxygen	NADH dehydrogenase	*rv0678*
Bedaquiline,	Inhibits mycobacterial ATP synthase	ATP synthase	*atpE,*
Delamanid & pretomanid	Inhibit mycolic acid synthesis		
Linezolid	Inhibits an early step in protein synthesis	50S ribosomal subunit	*rrl* & *rplC*
Streptomycin	Inhibits protein synthesis	ribosomal protein S12 and 16S rRNA	*rpsL, rrs* & *gidB*
Kanamycin & amikacin	Inhibits protein synthesis	30S ribosomal subunit	*rrs*
Capreomycin	Inhibits protein synthesis	30S and 50S ribosome subunits	*rrs*

© University of Pretoria

2.3.1 First line drugs

- Isoniazid (INH)

INH, chemically known as isonicotinic acid hydrazide, is a synthetic drug that was first described in 1912. It shows bactericidal activity against TB (69). It is one of the most effective, widely used, and affordable drugs for treating TB (69). The molecular mechanisms of INH resistance involve several genes in multiple biosynthetic networks and pathways (69,71). INH is activated by a catalase peroxidase enzyme encoded by the gene *katG* (18). The active INH products are then targeted by enzymes namely, enoyl acyl carrier protein (ACP), reductase (InhA), and beta-ketoacyl ACP synthase (KasA) that in turn inhibit mycolic acid biosynthesis (70-72). Resistance to INH is mainly attributed to mutations in *katG*. Other mutations have been reported in *inhA*, *ahpC*, *kasA*, *ndh*, *iniABC, fadE, furA,* Rv1592c, and Rv1772 (18,71). The mechanism of action of INH is complex and it is not exactly clear how this drug's activity results in cell death (71,73)

- Rifampicin (RIF)

RIF forms part of a group of drugs called rifamycins produced by *Amycolatopsis rifamycinica* (74). RIF is a semi-synthetic drug that binds to RNA polymerase inhibiting the process of transcription (46,72). RIF demonstrates a remarkable sterilising ability and its use in combination with INH and EMB resulted in shortening the duration of treatment from 18-24 months to 9 months (75,76) Most antibiotics require active growth and metabolism to exert their anti-bacterial effects. RIF retains its activity against slow-growing, and even non-replicating, Mtb bacilli (3,77). This is important to treat TB where low metabolic activity and/or non-replication are considered key factors in persistent Mtb infection (3). Resistance to RIF is mostly due to mutations in the *rpoB* gene that encodes for the β-subunit of the DNA dependent RNA polymerase (78,79). The 81 bp region of the gene is termed as the resistance determining region (RRDR) that contributes to 96% of resistance phenotypes (72). Currently, RIF is being studied to determine if higher doses of this drug could shorten the current duration of treatment (75).

- Pyrazinamide (PZA)

PZA works remarkably well against persistent bacterial populations (80). PZA is one of the few drugs that can inhibit semi-dormant bacteria inside macrophages where the environment is acidic with a pH value varying from 4.5 and 6.2 (69). Additionally, the inclusion of PZA

© University of Pretoria

reduced the duration of treatment from 9 months to 6 months (76). The mechanisms of PZA are not yet well understood (43,46,80). What is known about PZA's mode of action is that it is hydrolyzed to pyrazinoic acid (POA) in the mycobacterial cytoplasm by pyrazinamidase/nicotinamidase (PZase) encoded by the *pncA* gene (81,82). Once activated, POA disrupts the bacterial membrane energetics thereby inhibiting membrane transport (18,80). The POA is then pumped out of the bacterial cell by a weak efflux mechanism (18). This creates an acidic environment and the POA is protonated allowing for reabsorption into the cell, resulting in cellular damage (18,46). Mutations throughout the *pncA* gene have been described as a cause of resistance to PZA (69).

Mtb is exposed to acidic conditions *in vivo* as evidenced by the fact that the first-line TB drug PZA kills the pathogen only at low pH *in vitro* (83). Reproducibility of DST for PZA is challenging because PZA requires a standardized low pH assay (pH < 6.6), which is not ideal for bacterial growth and significantly slows the growth rate (69,84). It can lead to false resistance results (69,84). Additionally, the pH of the culture medium is not constant and can increase with the ammonia produced by the PZA bacterial metabolism, therefore inactivating the drug (69).

- Ethambutol (EMB)

EMB is active against actively replicating bacteria and interferes with the cell wall's arabinogalactan synthesis by targeting the arabinosyl transferases (46). Mutations in the *embCAB* operon encoding arabinosyl transferases are responsible for about 70% of clinically relevant EMB resistance (85).

2.3.2 Second line drugs

- Ethionamide (ETH) and prothionamide (PTH)

These two drugs (thiomides) are closely related compounds and are interchangeable in the TB chemotherapy regime (69). Both drugs are activated by a mono-oxygenase enzyme (EthA), encoded by the *ethA* gene that inhibits *inhA* (reductase), a shared target with the related drug INH, ultimately inhibiting mycolic acid biosynthesis (69,86). Resistance to ETH and PTH is conferred by mutations in *ethA*, *ethR*, and *inhA* (18,87). Cross resistance of ETH and INH results from mutations in the *inhA* gene (18,87). The *mshA* gene encoding a glycosyltransferase involved in mycothiol biosynthesis is also suggested to confer resistance to ETH (88).

- Fluoroquinolones (FQs)

© University of Pretoria

FQs belong to the quinolone class of antibiotics that target two type II bacterial topoisomerases, DNA gyrase and topoisomerase IV; enzymes that are important for replication, transcription, recombination, and chromosomal supercoiling (80,89). The antibiotics of this class used for DR-TB treatment comprise of 2nd generation FQs, ciprofloxacin, and ofloxacin (CIP & OFX), 3rd generation FQs levofloxacin (LEV), and 4th generation FQs gatifloxacin and moxifloxacin (GAT & MXF) (90). Fourth-generation FQs are being proposed as first line drugs to shorten DS-TB treatment (91). Individuals who have been treated with FQs for non-Tb infections could drive the acquisition of FQ resistance in chronically infected patients (22,89). FQ resistance is associated with mutations in either two DNA gyrase subunits encoded by *gyrA* and *gyrB* (79). Mutations in the quinolone resistance-determining region (QRDR) of *gyrA* determine the majority of clinically relevant FQ resistance (80,89). Heteroresistance, the coexistence of sensitive and resistant organisms in the same patient, has recently been shown to be surprisingly common in FQ-resistant Mtb isolates (89).

- Para-aminosalicylic acid (PAS)

PAS is highly specific for acid-fast bacteria (43). PAS was one of the first antibiotics used in the treatment of TB together with INH and SM (18). The mechanism of action and resistance are not entirely clear in PAS (92). It is suggested that it inhibits folate synthesis and iron utilization (92). It is also suggested that there could be additional mechanisms of action for PAS (91,93). Mutations associated with PAS are found in the *thyA*, *folC*, *dfrA*, and *ribD* genes (92).

- Clofazimine (CLO)

CLO is commonly used for the treatment of leprosy (94). The precise mechanism of action of CLO is not clearly understood (94). A study done by Lechartier and Cole (2015) in *M. smegmatis* suggests that it is probably a prodrug, reduced by NADH dehydrogenase to release reactive oxygen species upon re-oxidation by oxygen. Resistance to CLO has been attributed to non-target mutations in Rv0678, leading to the efflux of the drug (18).

- Bedaquiline (BDQ), Delamanid (DLM) and pretomanid (PTM)

BDQ, DLM, and PTM are drugs that have been introduced to the second-line treatment of MDR-TB (96). BDQ inhibits mycobacterial ATP synthase by suppressing bacterial respiration (18,96). The drug is therefore active against the dormant TB bacterium (18). Mutations in the

© University of Pretoria

atpE, *mmpR5,* and *pepQ* genes confer resistance to BDQ (96,97). DLM and PTM belong to the nitroimidazole class of antibiotics (18). Both drugs inhibit mycolic acid synthesis in actively replicating bacteria (96). Mutations in *ddn*, *fgd*, *fbiA*, *fbiB*, *fbiC*, and *fbiD* confer resistance to DLM or PTM (97).

- Linezolid (LZD)

LZD was the first member of the oxalizolidinones to be recommended to treat MDR/XDR (98). LZD targets ribosomes in the peptidyltransferase center on the 50S ribosomal subunit, in turn inhibiting an early step in protein synthesis and tRNA binding (18,94). Resistance to LZD has been associated with mutations in the *rrl* (encoding 23S rRNA) and *rplC* (encoding ribosomal protein L3) genes (98).

- Streptomycin (SM)

SM is an aminoglycoside that is active against slow-growing bacilli and targets the ribosomal protein S12 and 16S rRNA; components of the 30S subunit of the ribosome (18,91). This interaction prevents translation, in turn inhibiting protein synthesis (18). Mutations in the *rpsL* and *rrs* genes, encoding the ribosomal protein S12 and the 16S rRNA respectively, are common mutations believed to cause resistance to SM (18). Additional mutations in the *gidB* gene, encoding 7-methylguanosine methyltransferase specific for the 16S rRNA confer low-level resistance to SM (91).

- Kanamycin (KM), amikacin (AMK) and capreomycin (CAP)

KM and AMK are aminoglycosides while CAP is a cyclic peptide (99,100). Despite their belonging to different classes of antibiotics, all of them inhibit protein synthesis (99,100). KM and AMK target the A-site on the 16S ribosomal RNA of the 30S ribosome (99). Although they have different specificities toward this region, they both alter its conformation (99). CAP on the other hand is suggested to target a small inter-subunit bridge; the interface of the small and large sub-units of the ribosome (100,101). Resistance to these drugs is associated with mutations in the 1400 bp region of the *rrs* gene encoding 16S rRNA (18,99). Additional mutations in the *tlyA* gene confer resistance to CAP, and mutations in the *eis* promoter region and *whiB7* gene confer resistance to KM (18,46).

© University of Pretoria

2.3.3 Challenges with the treatment of TB

TB has a complex pathology that makes it a difficult disease to treat. Its complex pathology is a result of an interplay of biological and social determinants (17,26). These include variably effective control programs, the circulation of counterfeit drugs, the presence of co-morbidities in patients (e.g. HIV infection), numerous patient-related factors, including societal, immunological, and many less well-understood bacterial properties such as bacterial physiology, genetic factors, and population biology (12,26). The sophisticated infrastructure to test for DS, MDR, and XDR is not readily available in resource-limited settings (8). Low-income countries struggle to get the appropriate resources, the expertise for molecular testing, and, in some instances, they have to rely on other countries to perform DST for the second-line drugs often resulting in the delay of treatment or the empirical administration of treatment (26,102,103).

Incorrect treatment can have dire consequences for the patient and population thus, early detection and correct treatment are vital in controlling the occurrence of resistance (104). Incorrect treatment can result in the positive selection of resistant strains thus leading to the creation of novel resistance mutations and increasing the number of ineffective drugs used (104). To reduce the circulation of resistant strains, patients need to be treated with sufficient active drugs to prevent the selection of additional resistance in turn improving treatment outcomes (105). DST is the most widely used method for the diagnosis of Mtb (9). The accuracy of this test is crucial in making sure that patients are not exposed to ineffective drugs (104). DST involves the use of a solid or liquid culture for the identification and phenotypic DST of an Mtb strain (106,107). The challenge with this technique is the long turnaround time. It can take weeks or even months to obtain results due to the slow growth rate of Mtb (9). It requires a high biosafety level and lacks reproducibility for certain drugs such as PZA and EMB (9,104,106,107). Additional challenges include variable MIC concentrations that in some instances lead to phenotypic drug tolerance (108). Using a liquid culture is an alternative to reduce the turnaround time to ~10 days (9).

To quicken diagnosis, the WHO endorsed the use of the Xpert Mtb/ RIF test for RIF resistance, and line probe assays (LPA) were recommended to detect resistance to 1st and 2nd line drugs (32,102,109). These tests are limited to only a small number of loci and the phenotypic binary reporting (sensitive/resistant) does not inform on the degree of intermediate resistance (72,102,110). Several other diagnostic techniques have been endorsed by the WHO to quicken

© University of Pretoria

the diagnosis of TB as discussed by Archarya et al. (2020). What is clear, is that new rapid and accurate testing methods are needed to effectively treat DS and DR-TB.

2.3.4 Drug resistance development

Resistance can be acquired de novo (secondary resistance) in patients undergoing treatment either because of the poor administration of treatment, lack of patient adherence, interrupted drug supply, or patient-dependent pharmacodynamic and pharmacokinetic properties of the drugs administered (24,26). An additional way in which resistance could be acquired is through the direct transmission of a resistant strain also known as primary resistance (24,26). Cross resistance between drugs can occur because drugs target related metabolic pathways and these drugs could also interact resulting in resistance (26). Selection of DR strains can occur with intermittent exposure to suboptimal drug concentration levels initially producing hetero-resistance and then finally fully resistant bacterial populations (26). The treatment of other diseases has also contributed to the resistance in drugs such as FQ that treat other respiratory infections (24). DR can also vary depending on the locations of bacterial populations in the body and the characteristics of the drug (7). Depending on the antibiotic in question, there may be multiple mechanisms of resistance (8).

The interactions between the pathogen and the host contribute to the development of resistance (8). This interaction is increasingly recognized to influence the DR evolution (11). Pharmacogenomics studies have revealed that the current dosage of drugs is not enough to have the desired sterilizing concentrations in all patients and drug penetrance into TB lesions is variable in turn facilitating the evolution of DR (11). Within and between patient heterogeneity means that some cases of TB are unintentionally exposed to drug treatment that is at sub-inhibitory concentrations (8).

2.3.5 Patient adherence

Treatment of TB is very lengthy and exhaustive (24). Patients struggle to adhere to the treatment because TB drugs are toxic and patients often experience adverse side effects that include gastrointestinal problems, hepatotoxicity, peripheral neuropathy, psychiatric disorders, ototoxicity, nephrotoxicity, etc, (111). Mtb is unlike other bacteria; it is slow-growing with a generation time of 18-24h (83). Colonies are formed on agar within 3-4 weeks (83). The TB drug regimen had not been changed for decades. Additionally, the duration of treatment is long:

© University of Pretoria

4-6 months for DS-TB and even longer for MDR and XDR making it difficult for patients to adhere to the treatment (112). Recently the WHO has newly recommended using a 4-month regimen of INH, rifapentine, MXF, and PZA (2 months of INH, rifapentine, MXF, PZA, and two months on INH, rifapentine, and MXF) (113). Tolerable drugs and shorter drug regimens will encourage patient adherence (32).

2.3.6 Mtb's state of dormancy

Most antibiotics rely on the ability of bacteria to replicate for their action, but Mtb's non-replicating state makes it phenotypically resistant to bactericidal antibiotics (25). The immune system of the human host tempers with the ability of antibiotics to deliver swift and a durable cure for the disease (114). Once the TB bacterium is in the granulomas of the hosts' lungs, it can shut down its central metabolism, terminate replication, and thereby transit to a state of dormancy making it a difficult therapeutic target (25,58,83). TB therapy is long due to TB's state of dormancy and has created a major hurdle in the successful treatment of DS and DR-TB (25). This non-replicating state is referred to as phenotypic DR and in this context, the phenotypic DR is related to the physiologic state of dormancy of the bacterium and is independent of genetic mutations (83,114).

2.3.7 Development of new drugs

A message that is iterated over and over again in literature is the need for new drugs and improved treatment regimens (7). Currently, challenges that impede the progress of drug discovery include: (a) biological mechanisms that involve mycobacterial persistence and latency are not clearly understood and there are currently no effective drugs against persisters; (b) limited validated animal models that can reliably predict precise treatment outcomes in newly identified drugs; (c) lengthy clinical trials make it very difficult to predict the efficacy of new drug regimens; (d) there is a lack of trial sites in high TB burden countries that have a and (e) lack of investment in TB drug discovery programs due to insufficient profit return opportunity (23). Furthermore, there is currently a limited understanding of Mtb biology during its infection with the human host (115).

The ability of Mtb to survive in different microenvironments within the host is not entirely clear (115). New drugs should be effective against various physiological states of Mtb, as the different local conditions (acidic pH, hypoxia, higher carbon monoxide, or nutrient scarcity) potentially change their metabolic activity and induce drug tolerance (115). In September 2022, the WHO reported in the 2022 TB report that there were 26 drugs for the treatment of TB

© University of Pretoria

disease in Phase I, Phase II, or Phase III trials (33). Sixteen of these drugs include 16 new chemical entities, of which two drugs have received accelerated regulatory approval, one drug that was recently approved by the US Food and Drug Administration under the limited population pathway for antibacterial and antifungal drugs, and six repurposed drugs (33). In Phase II and Phase III trials, there are various combination regimens with new or repurposed drugs (33).

2.3.8 TB vaccine

BCG is the only approved vaccine that has been proven to be effective in protecting against TB infection in children (116). However, BCG is not as effective against pulmonary TB in adolescents and young adults, who are the main source of transmission (116,117). The discovery of a new vaccine could potentially be used as an adjunct to chemotherapy against DR-TB (116). Exposure to the Mtb pathogen does not induce sufficient protective immunity and not enough is known about the innate and adaptive human cellular response that is crucial for the clearance of Mtb infections (116,118). To develop a new vaccine, the cellular processes involved in Mtb susceptibility and pathogenesis need to be further investigated (116). In August 2021, there were 14 vaccine candidates in clinical trials: two in Phase I, eight in Phase II, and four in Phase III (33). These vaccine candidates aim to prevent TB infection, and TB disease and to help improve the outcomes of TB treatment (33). There is currently no effective vaccine, fuelling the need for new treatment regimes, drugs, and diagnostics to slow the evolution of DR and limit the transmission of DR variants (8).

2.4 TB diagnosis

2.4.1 Is whole genome sequencing an alternative to TB diagnosis?

Whole genome sequencing (WGS) is a tool often used in research to further understand the complex pathology of TB (9). WGS involves culturing fresh or frozen sputum specimens on solid or liquid media, extracting DNA from the cells, DNA library preparation and sequencing using short read technologies such as Illumina platforms or long read technologies such as SMRT PacBio and Oxford Nanopore, mapping the reads to a reference genome (usually H37rv accessed at the NCBI under NC_000962.3) and detecting variants be it SNPs and/or indels (107,119,120). The Mtb genome is relatively small (~4.4 MB), it constitutes a single chromosome and is well-suited for sequencing (72,107). WGS is a tool that provides

© University of Pretoria

comprehensive information on the Mtb genome (39,106). The advantages of incorporating this tool in TB diagnostics include: providing susceptibility results for both first and second line drugs, assisting clinicians in predicting DR profiles, monitoring the initiation of treatment, and predicting DR acquisition by identifying mutations, that are associated with DR (9,103,107,119). Moreover, the use of WGS has been constructive in understanding the DR-TB evolution, the estimation of the genetic diversity of TB, and the identification of mechanisms involved in DR (119,121).

Several studies have shown the potential use of WGS in clinical settings (103). A study done by Walker et al. (2015) used a catalogue of genetic mutations obtained from WGS data to predict phenotypic DST (110). Chan et al. (2013) showed the potential of using rapid WGS to reduce the time of detection for XDR tuberculosis from weeks to days (122). Nikolayevsky et al. (2016) conducted a systematic analysis of 12 studies and established that WGS had a higher discriminatory power in comparison to conventional genotyping methods (IS6110 RFLP, 24 VNTR, spoligotyping) to predict recent person-to-person transmissions (123). Computational algorithms such as the Resistance Sniffer program created by Muzondiwa et al. (2020) use WGS data to predict DR in Mtb isolates (104). In Uganda, Ssengooba et al. (2016) showed the potential use of WGS in complementing drug resistant surveys or surveillance with good specificity in comparison to phenotypic DST (124). Genetic identification techniques are key to achieving greater speed and improving the accuracy of diagnosis (46). In well-resourced countries, such as the UK and the Netherlands, the use of WGS for individualized therapy and WGS-based surveillance has already made its way into the clinical setting (107).

The studies have shown the importance of using WGS for the diagnosis, surveillance of disease transmission, and prediction of DR in Mtb (9). There are, however, several issues that have to be resolved if WGS is to be fully incorporated into the clinical setting (40,125). The need to culture before sequencing is time-consuming, therefore, extracting DNA directly from sputum would be ideal (119,126). Currently, several WGS data analysis workflows exist and they differ in terms of their scope, reporting lines, and the lack of standardisation; this, in turn, makes it difficult to perform cross-comparisons and the rigorous validation of these pipelines (107). Moreover, there is a need to understand the relationship between the phenotype and the genotype of the resistant strains in order to make concrete conclusions (39,119). Additionally, a catalogue of high-quality resistance-determining variants is needed to ensure the clinical utility and use of these tools (127). The use of WGS is also hindered by the lack of bioinformatics skills among clinical microbiologists, therefore the data has to be presented in

© University of Pretoria

a clear easily interpretable manner to clinicians and programme managers (39,40). WGS, however, cannot replace phenotypic susceptibility testing for all antibiotics because of an incomplete understanding of the genetic basis of DR (122). Phenotypic tests will still be important in the detection of DR patterns that are difficult to interpret, however, the goal should be to detect all genetic variants that are associated with DR (107).

2.5 Genetic diversity of *Mycobacterium tuberculosis*

The genetic diversity of MtbC includes indels, duplications, and non-synonymous SNPs (24). These differences resulted in the typing of MtbC into the various lineages. The typing method IS*6110* RFLP was used as the first gold standard for fine typing of MtbC (128). Newer methods are now also being used that include PCR-based methods known as spoligotyping, MIRU-VNTR, and methods based on the detection of large sequence polymorphisms (LSPs) or single nucleotide polymorphisms (SNPs) (128,129). These techniques have been instrumental in the identification of the 7 major lineages (1-7) and the recent lineages 8 and 9 (129). Mtb's long history of co-evolution with humans suggests that human populations and MtbC strains differ in their susceptibility/resistance and virulence characteristics (42). Studies have suggested that these lineages differ in terms of their geographic distribution, virulence phenotypes, immune response regulation, transmissibility, and disease severity, hence differing in epidemiologic and clinical profiles (129,130).

Lineages 2, 3, and 4 are often referred to as evolutionarily "modern" (Figure 2.1) because they harbour a deletion in the genomic region known as TbD1 (131). Lineages 1, 5, and 6 are referred to as "ancient" and do not harbour this deletion (29). The presence or absence of this deletion should not be necessarily interpreted as indicating an evolutionary time dimension (42). Lineage 7, on the other hand, is an intermediate lineage (29). The lack of HGT in Mtb means that a deletion cannot be reacquired, hence they can be used as phylogenetic markers to classify Mtb (42). Some lineages are more widely distributed whereas others are more restricted to a particular geographic region (Figure 2.1); it is not fully understood why this is the case (42).

An important question is how genetic diversity in Mtb translates to phenotypic diversity (24,131). The role of strain diversity in the clinical setting is not entirely clear (24). What makes this a challenge is that many additional variables would need to be taken into account and strain

© University of Pretoria

classification is still a challenge due to the lack of standardization of the methods used to type Mtb (24).

2.5.1 Lineage 1

Lineage 1 (L1) also known as the Indo-oceanic/East African-Indian lineage is defined by the deletion RD239 (12,132)3-1. L1 shows an intermediate geographic range (133). L1 is likely to be among the most genetically diverse lineages and demonstrates an inter-lineage diversity (134,135). It is phylogeographically restricted to East Africa, South Asia, Southeast Asia, and the Philippines (134).

L1 is often not associated with resistance and has a reduced potential for transmissibility and virulence, presenting a restricted geographical distribution (132,136). Studies highlighted by Netikul et al. (2021) found that patients infected with L1 are generally older and have a higher case fatality rate, even though L1 is generally less DR than modern lineages (135). L1 isolates were shown to grow slower than L2 but induce a stronger cytokine response in macrophage culture (135). Furthermore, L1 Mtb isolates may also be distinct by antigens, probably due to their long history of co-evolution with humans (135).

2.5.2 Lineage 2

Lineage 2 (L2) also known as the Beijing/East Asian lineage is characterised by the RD105 deletion (137). L2 is often reported to be associated with DR and the most successful lineage in causing TB disease (17,138). The emergence of L2 genotype strains and their association with MDR further supports the fact that they possess a strong capability to acquire DR mutations (139). Data from Hakamata et al. (2020) demonstrated that higher mutation rates of Mtb L2 strains during human infection likely explain the higher adaptability and propensity for DR acquisition (140). The biological basis of how the mutation rate in L2 could affect the acquisition of DR is still to be determined (17,24).

L2 consists of a variety of sub-lineages that might differ phenotypically (128,141). These differences suggest the existence of biogeographical diversity and differences in pathogenicity (138,142).

© University of Pretoria

2.5.3 Lineage 3

Lineage 3 (L3), also known as the Central Asian strains (CAS), occurs predominantly in India, Central Asia, and East Africa (135,143). This lineage shows an intermediate geographical range and is characterized by the deletion RD750 (133,137). L3 also demonstrates inter-lineage diversity (143). L3 has been reported to be the dominant strain associated with MDR in South Asian, North, and East African countries and is also a potential driver of MDR in other parts of the world (144,145).

2.5.4 Lineage 4

Lineage 4 (L4) also known as the Euro-American lineage is one of the most widely distributed lineages among the MtbC together with L2 (131). It is prevalent in Europe, the Americas, Africa, and the Middle East and is characterised by the *pks* 15/1 7-bp deletion (137). L4 appears to be more virulent and transmissible together with L2 in comparison to the other lineages (42). This, however, is not universally true for all Mtb isolates of this lineage as much variation exists in L4 (42). Stucki et al. (2016) determined that L4 consists of 10 sublineages, where some are geographically restricted (e.g. lineage 4.6.1/Uganda and Lineage 4.6.2/Cameroon), and others are more widely distributed (42).

2.5.5 Lineage 5 and Lineage 6

Lineage 5 (L5) and Lineage 6 (L6) are restricted to West Africa (135). L5 and L6 are also known as *M. africanum* West Africa 1 and West Africa 2, respectively, and almost only occur in West Africa or in recent immigrants from those regions (131). Lineage 6 occurs mostly in the Western part of West Africa, whereas Lineage 5 is commonly found further to the East in regions bordering the Gulf of Guinea (131). L5 and L6 geographical restriction to West Africa is not understood and is suspected to be due to adaptation to West African populations (146). L5 and L6 are characterized by the RD711 and RD702 deletions respectively (137). Strains from L5 and L6 grow slower and are less virulent in comparison to the other lineages (133).

2.5.6 Lineage 7

Lineage 7 (L7) is almost only observed in Ethiopia or recent Ethiopian emigrants (147). L7 is associated with prolonged patient delay and slow growth in vitro (148). A study done by Yimer et al. (2015) in the Amhara region of Ethiopia found that L7 is associated with pulmonary

© University of Pretoria

tuberculosis (PTB) in this region and infection with L7 did not cause severe symptoms, thus patients were more likely to delay seeking medical attention (149).

2.5.7 Lineages 8 and 9

Recently two new proposed lineages 8 and 9 have been discovered. Lineage 8 is believed to be an as-yet-unknown ancestral stage between the MtbC and its progenitor *M. cannetii* - like mycobacteria (30). A population genomic and phylogeographical analysis of 675 genomes by Coscolla et al. (2021) identified unclassified genomes that were proposed to be named MtbC lineage 9 (28). Lineage 8 and Lineage 9 seem to be restricted to East Africa (135).

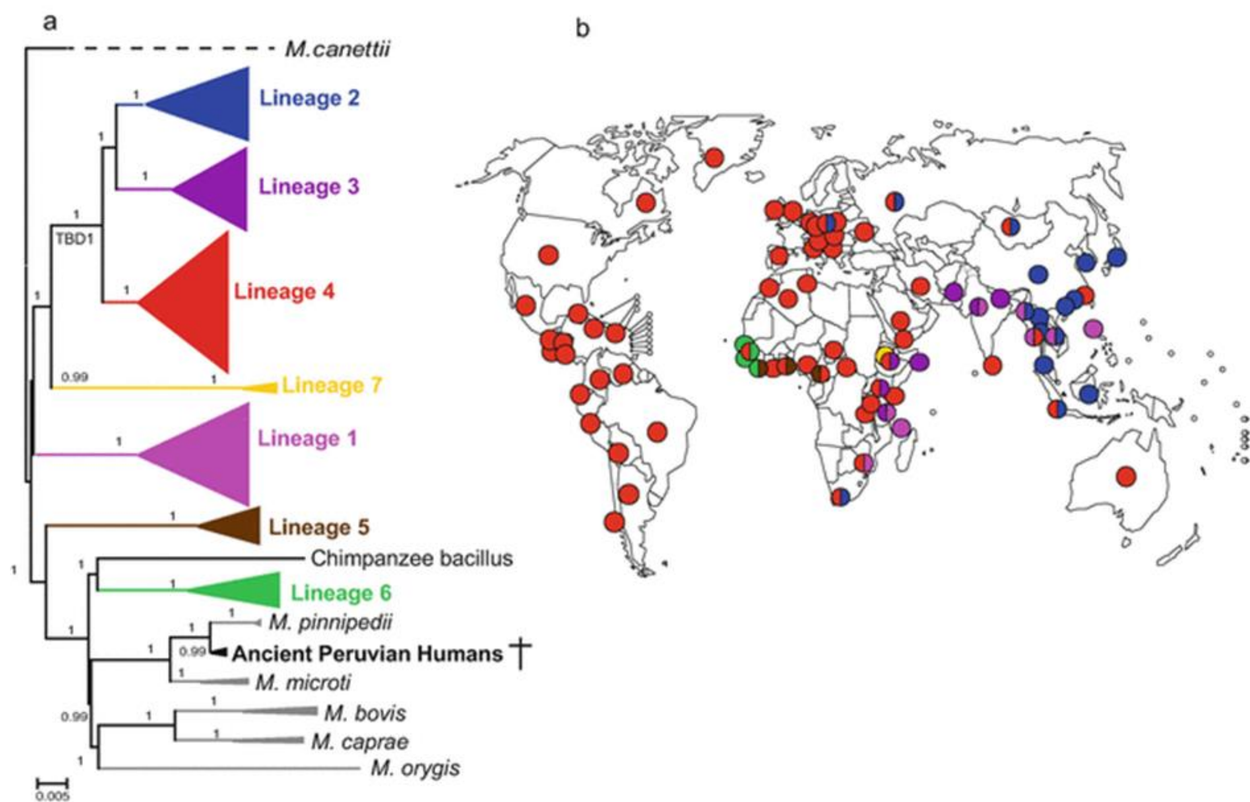


Figure 2.1: The geographical distribution of the different lineages of the *Mycobacterium tuberculosis* complex (131).

2.7 Development of drug resistance in Mtb

The evolution of pathogens has resulted in major economic repercussions, loss of life, prolonged treatment, and increased hospitalizations, and is the biggest concern in public health care around the world (150,151). Understanding how these deadly strains emerge and achieve

© University of Pretoria

sustained transmission is key in determining preventative measures, the containment of Mtb as well as avoiding future epidemics (151).

The evolution of DR in TB has been largely attributed to extrinsic factors. Extrinsic factors include the social and economic factors that have resulted in the development of DR-TB (7,8). These include an interrupted drug supply, low-quality drugs, patient non-adherence, bad quality TB control programs, and major political events (7,8,16). These, however, are not sufficient in explaining the evolution of DR-TB, because even in well-functioning healthcare systems, where there is strict adherence to treatment, DR emerging still occurs (8). There are additional mechanisms of resistance that include intrinsic, acquired resistance, and less clear epistasis that drive the evolution of DR (7). How these mechanisms lead to the emergence and the fixation of resistance strains is not well understood (16). Several hypotheses of DR development are illustrated in Figure 2.2.

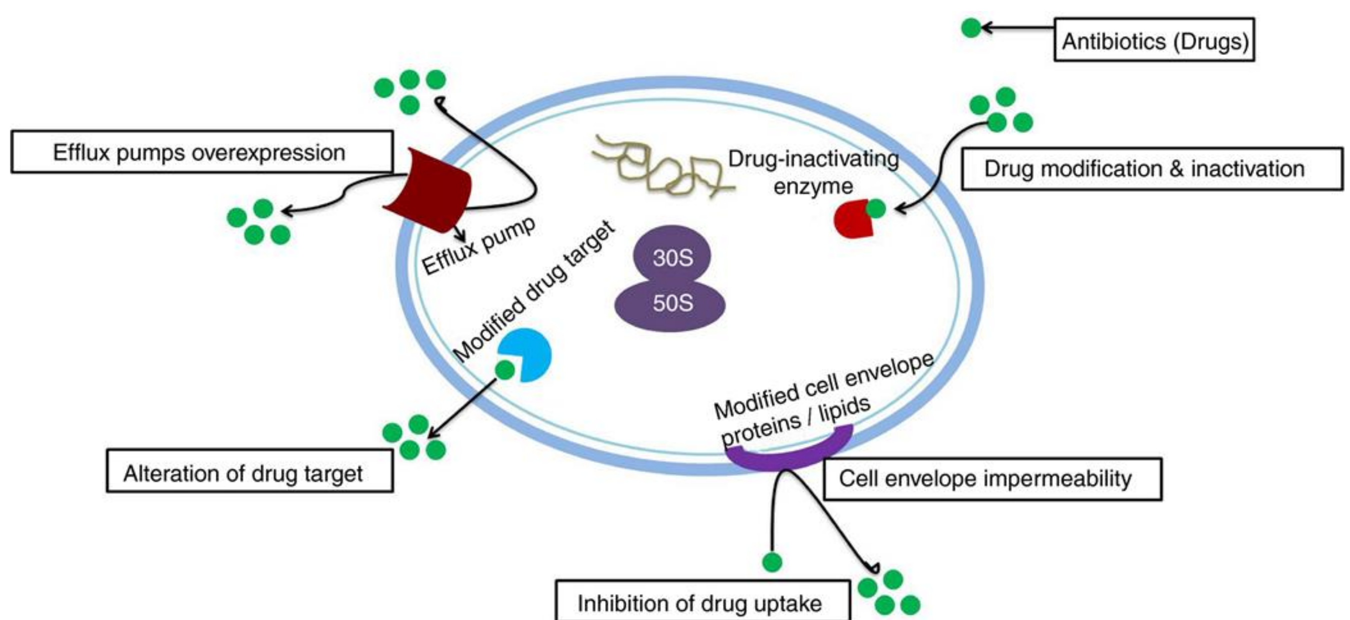


Figure 2.2: Acquired and intrinsic mechanisms of resistance found in *Mycobacterium tuberculosis*.

© University of Pretoria

2.7.1 Intrinsic resistance

There is no denying that a significant amount of progress has been made in the treatment of TB, however, antibiotics provided a new selection pressure that led to the survival, emergence, and amplification of DR mutants among pathogens, which already have a broad intrinsic resistance spectrum (38). This not only limits the use of the current antituberculous drugs but also makes the development of new drugs more difficult (43). Intrinsic resistance (Figure 2.2) prevents access to the target and accounts for resistance that is not conferred by DR mutations (70). Intrinsic factors include efflux pumps, drug-altering enzymes, and the impermeability of the cell wall (152).

- **Impermeability of the cell envelope**

One of the ways in which TB bacilli are classified is according to the composition of their cell wall (153). Gram-negative bacteria are surrounded by a thin peptidoglycan cell wall, and the cell wall is surrounded by an outer membrane containing lipopolysaccharides (153). Gram-positive bacteria do not contain the outer membrane and are surrounded by layers of peptidoglycan many times thicker than those that are found in Gram-negative bacteria (153). Mtb on the other hand is an acid fast Gram-positive bacterium with an unusually thicker and more hydrophobic cell wall, with a periplasmic space similar to that of Gram-negative bacteria, due to the presence of an array of different lipids (8,43,91). As a result, the glycosylation observed in the membrane and cell envelope associated proteins, lipids, and lipogylcans might contribute to the virulence, phenotypic variability across Mtb lineages, and DR development (16). Studies have reported that the cell wall of Mtb makes it difficult for antibiotics to penetrate (8,26,36,54). A recent model showed that the cell envelope is divided into three parts: capsule, cell wall, and cell membrane (27). The outer capsule is mainly formed of proteins, glucan, and a small number of lipids (27). The cell wall consists of the outer mycomembrane (MM), arabinogalactan (AG), and inner peptidoglycan (PG) (27). MM has two leaflets, the outer one is formed of lipids such as trehalose mycolates, phospholipids, glycopeptidolipids, and lipoglycans, and the inner leaflet is composed of long-chain mycolic acids (MA) (27,43). This array of lipids is what makes the cell envelope difficult for diffusion of even hydrophobic molecules, such as rifamycins, macrolides, fluoroquinolones and tetracyclines (27).

There is little known on changes in the cell envelope composition of DR Mtb. Some studies have observed differences in cell envelope thickness between susceptible and resistant strains (16). The treatment of MDR and XDR strains is also suggested to be challenged by cell wall

© University of Pretoria

composition changes. Velayati et al. (2009) observed using electron transmission microscopy (TEM) differences in cell wall thickness between susceptible and resistant strains. MDR showed a difference of ~2 nm and XDR showed a difference of ~4.6 nm in cell wall thickness compared to DS strains (16,154). However, in this publication, little information was provided on whether the thicker cell envelope is a common feature of the resistant strains (16). Other studies have also looked into the difference in lipid profiles between resistant and susceptible strains (16). Using a high-throughput mass spectrometry-based lipidomic approach, Pal et al. (2017) determined that MDR-Mtb strains have increased levels of free fatty acids, TDM, glycerophospholipids, altered glycerolipids, and unique and distinct lipid signatures when compared to DS isolates (16,155). Several studies have looked in depth at the importance of cell wall integrity in Mtb (156-159).

- **Porins**

Pore forming proteins called porins are mounted on the outer layer of the cell wall and facilitate the entry of small molecules, nutrients, and hydrophilic compounds, that are important for Mtb viability and replication (27,43,160). Mtb does encode for two porin-like proteins (OmpA and Rv1698), but the role of porins in DR is lacking (8,27).

- **Efflux pumps**

Efflux pumps are a group of cell envelope glycoproteins that are important for bacterial metabolism, physiology, transport of nutrients, removal of toxins wastes, and transportation of signalling molecules through the cell envelope (16,27,161). They are crucial for expelling drug molecules when they enter the bacterial cell (16,27,161). Putative drug efflux pumps sometimes called transporters in Mtb are grouped into five super-families namely: (i) ATP-binding cassette (ABC) superfamily, (ii) major facilitator superfamily (MFS), (iii) small multidrug resistance (SMR) family, (iv) resistance-nodulation-cell division (RND) superfamily, and (v) multidrug and toxic compound extrusion (MATE) family (27,152,161,162). Intrinsic resistance due to efflux pumps has been observed in a number of studies where there is an overexpression of efflux pumps due to drug exposure (27).

Mtb efflux pumps are rapidly upregulated within hours after experiencing antibiotic stress and usually confer low level resistance and play a significant role in evolving to higher levels of resistance (16,77,161). The upregulated efflux pumps provide a selective advantage that allows Mtb to survive and replicate under sub-optimal drug concentrations until further development of several other resistance-associated mutations confer clinically relevant drug-resistant (16).

© University of Pretoria

They may be stepping stones for high-level resistance (8). The overexpression of efflux pumps in resistant strains as a result of antibiotic stress has been reported by several studies and these findings emphasized the importance of understanding the mechanisms that drive regulation of efflux pumps (27).

The inhibition of drug efflux pumps is a strategy to attack persistent Mtb and shorten therapy (163). The use of efflux pump inhibitors as alternative antimicrobials or adjuvants can potentially reduce the speed of DR development during TB therapy (152). The use of inhibitors is well-established in other bacteria, but their use against Mtb is not entirely clear. There are drugs of this class that have been suggested to be included in treating TB, such as carbonyl cyanide m-chlorophenyl hydrazine (CCCP), verapamil (VP), thioridazine (TZ) and chlorpromazine (CPZ), which can decrease Mtb resistance to RIF (161,164). Among them, verapamil has been shown as the most potent efflux inhibitor (161,164).

- **Drug degradation and modification**

Once a drug has entered the cell envelope, it may be cleaved by enzymes and rendered ineffective (8). Drug degradation and modification are well-studied DR mechanisms in Mtb (27,91). Mtb is intrinsically resistant to β-Lactam antibiotics, which reduce the number of drugs available to treat Mtb (8,26,27). β-Lactam antibiotics target transpeptidases called penicillin-binding proteins (PBPs), which are important in cell wall synthesis (27,165). β-Lactamases produced by Mtb hydrolyse the β-lactam ring rendering the drug ineffective (27). The ineffectiveness of β-lactam drugs can be attributed to many factors and these include the impermeability of the cell wall and the low affinity of PBPs (27,43). The low permeability of the cell wall results in reduced access by the antibiotics and the low affinity of PBPs does not allow drugs to bind to the target (27).

A new mechanism of resistance may explain how ~10% of INH resistance is not conferred by DR mutations (166). A study done by Arun et al. (2020) suggested a new mechanism of INH resistance that involves acetyltransferase Rv2170, which may inactivate INH (166). This enzyme was able to catalyse the transfer of the acetyl group to INH from acetyl coenzyme A (acetyl-CoA), breaking down INH to isonicotinic acid and acetylhydrazine (166). This mechanism still needs to be investigated in INH-resistant clinical strains (166).

Intrinsic resistance to aminoglycosides is due to acetyltransferase Eis (enhances intracellular survival) (36,43). Eis was initially found to play a role in the survival of Mtb in macrophages and was found to also play a role in manipulating host innate immunity against bacterial

© University of Pretoria

infection (43). Eis also acylates CAP and KAN as well, so Eis has a dual function of protecting Mtb from host immunity and antibiotics, which implies coevolution of virulence and antibiotic resistance in Mtb (27,43). Mutations in the promoter of *eis* sometimes lead to over-expression and confer low-level resistance to KAN, but not to AMK (27). The expression of the ribosomal methyltransferase (Erm) has been attributed natural resistance mechanism against the macrolide class of antibiotics (36,160). Mtb also reduces the binding of macrolides and lincosamides to ribosomes and renders a natural resistance to this drug (43). Two main classes of aminoglycoside-modifying enzymes in mycobacteria include acetyltransferase and phosphotransferase (27).

- **Target alteration and target mimicry**

Target mimicry is an interesting DR mechanism in Mtb whose aim is to inhibit FQ's mode of action (27,29,43). FQs kill bacterial cells by inhibiting DNA replication, transcription, and repair by binding to DNA gyrase and topoisomerase (43). Mtb is suggested to develop resistance through MfpA (Mycobacterium fluoroquinolone resistance protein A) (29). MfpA closely resembles the double helix structure of DNA (43). It is proposed that MfpA binds to DNA gyrase and prevents FQ from binding in turn inhibiting the drug's action (29,43). The physiological effect and the significance of this protein to FQ resistance remain unknown (43).

2.7.2 Acquired drug resistance

Acquired DR can occur either through mutations or horizontal gene transfer mediated by phages, plasmids, or transposon elements (Figure 2.2) (26). HGT has not been reported in Mtb, therefore Mtb is highly clonal and its thick cell wall makes gene exchange difficult (54). As a result, DR is mainly due to chromosomal mutations (7,54). Chromosomal mutations confer resistance through modification or overexpression of the drug targets and by prevention of the prodrug activation (8). The most commonly occurring mutations conferring DR are non-synonymous SNPs and to a lesser degree synonymous SNPs and INDELS (26). DR mutations are often found in genes that encode enzymes that are directly targeted by antibiotics, in regulatory regions of these genes, or in gene products that are involved in the activation of pro-drugs (26).

© University of Pretoria

- **Drug target alteration**

The interaction between a drug and its target is highly specific. Point mutations in genes that encode drug targets preclude drug binding to the target resulting in a decrease in drug potency (160). Antibiotics target central metabolic pathways, including DNA, RNA, and protein synthesis, which are essential for the survival of the bacterium within the host (160). Because the drug targets are vital for the survival of the bacterium, they are highly conserved and the regions suitable for DR mutations within the target proteins are limited (7,8). Resistance due to drug target alterations confers resistance to RIF, INH, FQ, aminoglycosides, PAS, and SM (18).

- **Loss of pro-drug activation**

Some TB drugs require activation by intracellular enzymes to exert bactericidal/bacteriostatic activity (17,167). Mutations in the genes that encode drug-activating enzymes result in suppression or loss of enzyme activity; in turn conferring resistance to the respective drug (17,167). The prevention of prodrug activation results in resistance to INH, RIF, EMB, PAS, DEL, and PRE (8,17). If the activation of the TB- drug is not essential to the survival/growth of the bacterium, then the target region of DR conferring mutations in the polymorphic gene is larger meaning that many point mutations and INDELS may occur in this gene (8).

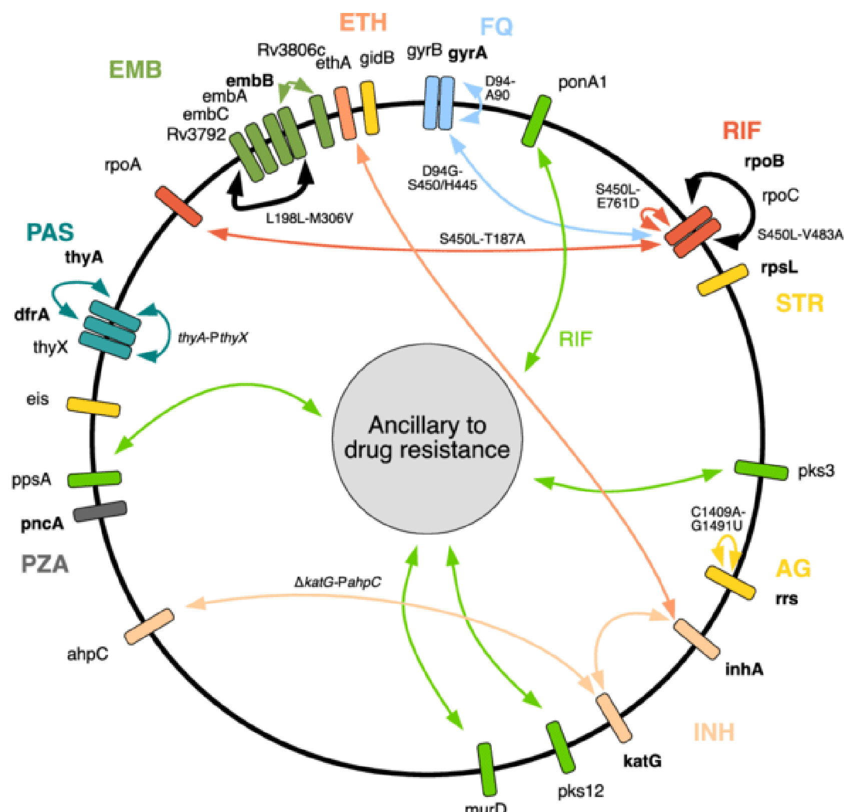
2.7.3 Epistasis

Epistasis is a mechanism that plays an important role in the evolution of organisms and the evolution of DR (7). In the context of DR, epistasis is a phenomenon where the effect of one mutation depends on the presence of another mutation in the same genome thus affecting the organism's phenotype (24,41). The epistatic interactions may include DR mutations, associated fitness costs, and the necessity of intermediate and compensatory mutations to reduce the fitness cost. The consequences of epistatic interactions could affect the levels of resistance, the frequency of resistance, or the relative fitness of resistant mutants, the dynamics of evolution, constraint types, and the order of mutations that can be fixed by natural selection (168). Consequently, epistasis affects the dynamics of evolution, for example, by causing a mutation to have a smaller fitness cost in the presence of another compensatory mutation (169). The clonal nature of Mtb favours the epistasis in that the absence of HGT means that all SNPs in a single Mtb genome are in linkage disequilibrium (LD) and hence the co-adapted alleles are

© University of Pretoria

spread together in the population (7,14,42). Consequently, new mutations arise not independently from each other (42,170). They represent linked components of epistasis relevant to the evolution of DR in TB. These include positive, negative, and sign epistasis (7,29).

An analysis of epistatic interactions could shed light as to why and how Mtb demonstrates a remarkable adaptive potential (20). The evolutionary trajectory of DR in Mtb can be described by determining the order and how quickly DR mutations occur during the adaptation (20). Describing the evolutionary trajectory of DR in Mtb can be achieved by identifying epistatic interactions and their directions (17). These interactions would determine the order in which the DR mutations arise by determining the optimal mutational pathway for resistance to meaningfully arise (17). A review by Trauner et al. (2014) discussed how an epistatic interaction can reduce fitness costs, increase the level of resistance, and accommodate for changes that occur in the physiology of Mtb when resistance occurs (Figure 2.3) (171). Understanding how epistatic interactions drive the evolution of DR in Mtb will help anticipate future trajectories and prevent the progression of an even deadlier disease (26). Studies that look into epistasis can provide vital information to slow or limit the evolution of DR (20).



© University of Pretoria

Figure 2.3: Figure adopted from Trauner et al. (2014) showing putative epistatic interactions among genes found within *M. tuberculosis* genomes.

- **Positive epistasis.**

In positive epistasis, the fitness of a double mutant is higher than expected consequently promoting the evolution of MDR while minimising its cost (11,20,24). In other words, positive epistasis promotes the evolution of DR (24). The emergence of DR in Mtb is a stepwise acquisition of mutations and studies have shown the preference for a combination of DR mutations that result in higher *in vivo* fitness (17). Positive epistasis between DR mutations has been suggested by several studies, where combinations of at least two specific mutations, such as *rpoB*531, *katG*315, *rpsL*43, *embB*306, and *gyrA*94, were favoured (7,24,41,172). Spies et al. (2013) demonstrated that the double mutants *rpsL* K43R/*katG*-S315T, *rpsL*-K43R/*rpoB*-S531L, and *rpoB*-S531L/*katG*-S315T, are frequently detected in clinical isolates, and grow faster than drug-susceptible strains (173). Interactions between DR mutations may restore or even increase the biological fitness of resistant strains in comparison to drug susceptible ones (7). Although not much is known about interactions between DR mutations in Mtb, they could play an important role in the emergence and evolution of MDR and XDR Mtb (7).

- **Negative epistasis**

In negative epistasis, the fitness of a double mutant is lower than expected consequently constraining the evolution of MDR by aggravating its cost (11,20,24). In light of this combining antibiotics where their respective DR-conferring mutations interact negatively, may result in a reduction in biological fitness (26). In light of this, a possibility would be to combine drugs where the respective DR-conferring mutations interact negatively, thus lowering the fitness of strains carrying both DR determinants (26)

- **Sign epistasis**

Genetic interaction can have positive or negative effects on the phenotype or fitness of an organism depending on the genetic background (169,174,175). The genetic background could be in the context of a resistant genetic background, compensatory, and the genetic diversity of the strains. This is a primary example of how epistasis could modulate the evolution towards DR (17). Sign epistasis can limit the number of evolutionary paths available to a population, it can constrain the type and order of selected mutations that can be fixed by natural selection

© University of Pretoria

and it can also make adaptive trajectories dependent upon the first random mutation (20,169). This is so because in sign epistasis if mutations are individually advantageous, this may have profound effects on the trajectory of DR-evolution, but if the combination of these individually advantageous mutations is deleterious, one would not expect to see these combinations persist in the population (20). Thus, sign epistasis can constrain the number of paths that are available for adaptation (20).

When looking at biological fitness in the context of sign epistasis; a mutation may be beneficial or have no effect in a specific genetic background, but deleterious in a different genetic background (20,29). In a reciprocal sign epistasis, the interacting mutations change their signs, for example when mutations that are individually deleterious become beneficial in combination with other mutations, or when two individually beneficial mutations become deleterious in a different genetic background (Figure 2.4) (20,29).

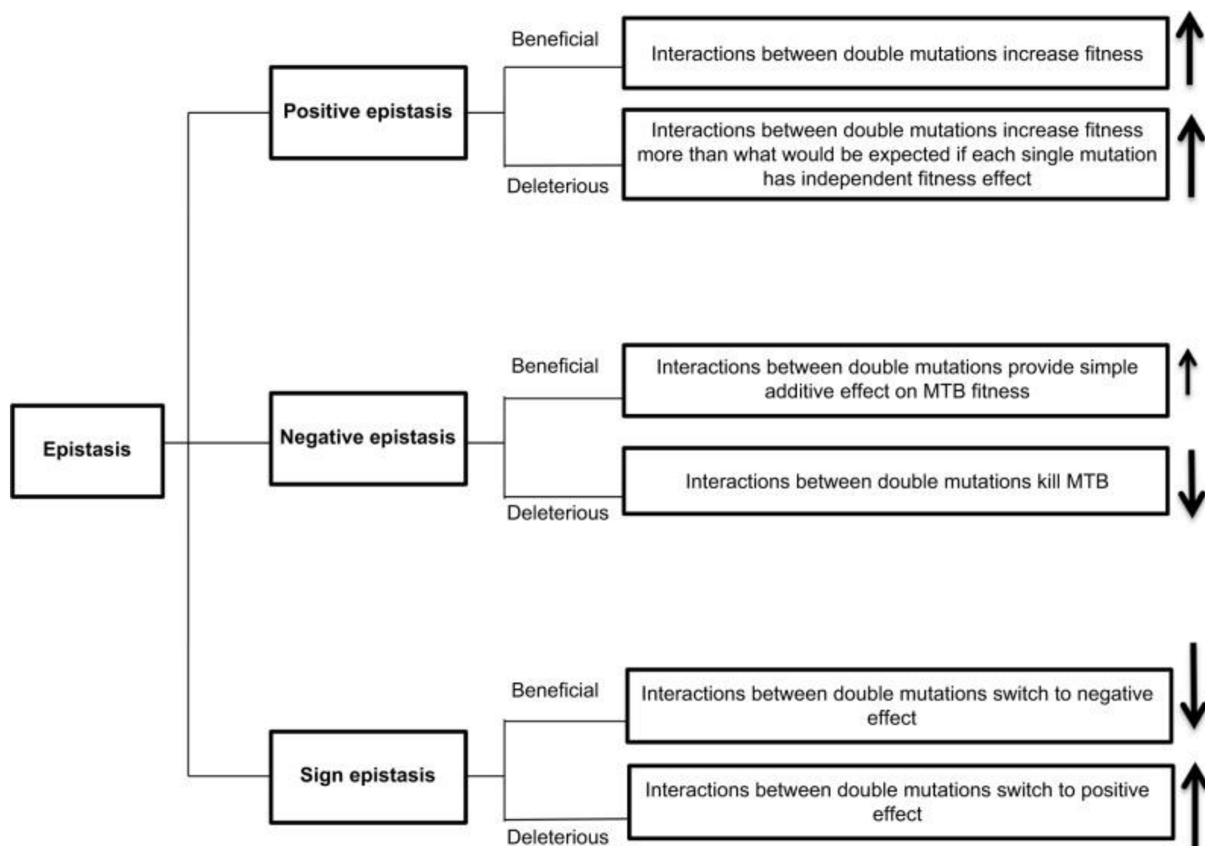


Figure 2.4: The different forms of epistasis and their effect on fitness (↑ high fitness) (↓ low fitness) (29).

© University of Pretoria

2.7.4 Fitness cost and Compensatory evolution in essence are part of the sign epistasis

TB drugs target essential or highly conserved genes (7,173). Mutations that occur in genes that code for these key pathways may have adverse effects on the normal physiological state of the bacterium in the absence of the drug resulting in a fitness cost (3,7,164). In bacteria, the acquisition of resistance is often associated with a biological cost because of the essentiality of the highly conserved antibiotic targets (173). Initially, there may be a fitness cost due to the acquisition of a DR mutation, but as the bacterium continues to evolve, it acquires additional mutations that ameliorate the initial fitness costs over time (173). Studies have also reported mutations that result in no or low fitness costs that are often positively selected for in clinical isolates (26,173). The extent of the fitness cost may also depend on a mutation as well as on multiple mutations termed the genetic background of the strain (7).

The clonal nature of Mtb means that negative fitness effects can only be alleviated through reversion or compensation because deleterious mutations cannot be replaced through HGT (42). It was demonstrated that fitness plays an important role in the emergence of resistant bacteria (26). There is a growing amount of evidence on the role of compensatory evolution in the acquisition and maintenance of resistance, highlighting the importance of determining the interaction between DR mutations and compensatory mutations (14). Fitness costs conferred by DR mutations can be reduced or eliminated by compensatory/second-site mutations (20). Compensation is often said to occur after the initial mutation, however, compensation can occur before or after the initial mutation (20). Under the concept of sign epistasis, compensatory mutations that occur in a DR-sensitive background that are deleterious are unlikely to segregate independently of the resistance mutation (19). This raises an important question on Mtb's ability to harbour multiple drug resistance mutations while retaining its ability to infect, persist, and cause disease (3).

The best-known examples of compensatory mutations are found in the RNA polymerase of RIF-resistant strains, which have been shown to compensate for the deleterious effects of RpoB mutations in the absence of RIF (29,42). Mutations within *rpoA* and *rpoC*, encoding the α and β-subunits of the RNA polymerase compensate for the fitness costs associated with the resistance mutations in *rpoB* (16,26,27,51). The intragenic mutation in *rpoB*, V615M, increased the rate of transcription elongation and compensated for the defective RNA polymerase *rpoB* having S531L substitution (27). Compensatory mutations in *rpoA* and *rpoC* genes are associated with the improved transmissibility of the DR strains as evidenced by

© University of Pretoria

clonal exaptation of the strains carrying these particular mutations that appeared to restore the fitness of the resistance strains to the wild-type function (171). This was investigated in a study by Cohen et al. (2015) on the XDR outbreak in Tugela Ferry, KwaZulu-Natal, South Africa, in 2005. They studied the long evolutionary history of the MDR strain and revealed compensatory mutations associated with RIF resistance, which may have contributed to the success of this clone dating back to 1957 during the early onset of TB treatment (176).

Newer work by Zimic et al. (2020) identified 35 additional putatively novel compensatory mutations in *rpoC* genes from Mtb isolates in Peru, which were RIF-resistant (177). Another study by Coll et al. (2018) identified a *pnaB2* mutation that may compensate for PZA resistance conferred by *pncA* mutations and also found mutations in the *thyX–hsdS.1* promoter that compensates for PAS resistance conferred by *thyA* however, experimental validation is required (34). In Mtb; the well-known *katG* mutation at codon position 315 is suggested by multiple studies to be compensated for by the overexpression of *ahpC* (7,8,16,24,26,29). Resistance to aminoglycosides (AMGs) is conferred by mutations in 16S rRNA (26). If mutations they destabilize the rRNA secondary structure, which can be compensated by a secondary mutation in the same gene (26,51). Compensatory mutations in the *ahpC* promoter and 16S rRNA are rarely seen in clinical isolates of Mtb (26,27).

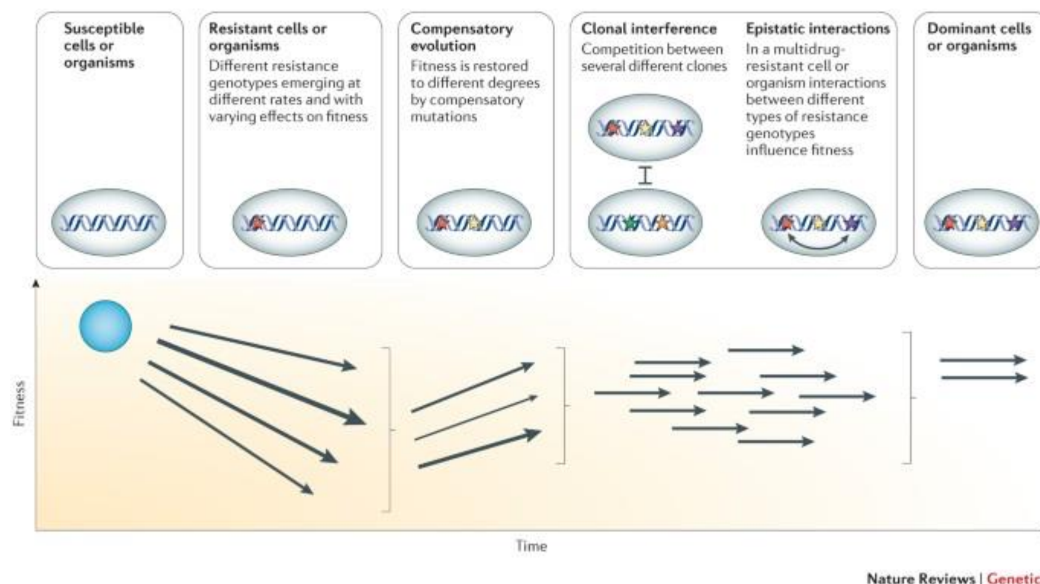


Figure 2.5: Shows the evolutionary trajectory of DR from the acquisition of mutations to clonal interference (178).

© University of Pretoria

2.7.5 Clonal interference.

In populations where HGT does not occur, each mutation remains in the genetic background where it originally occurred (150,179,180). If multiple positive traits are selected for different genetic backgrounds, these traits cannot be recombined (150,179). Instead, they have to compete against each other. This process is known as clonal interference (150,179). Clonal interference results in an occurrence of dominant clones that can adapt in response to antibiotic selective pressures (16,27).

Sun et al. (2012) performed a study to investigate this process in 3 TB patients by tracking DR allele frequencies in these patients across 2 time points. The 1st patient had no DR mutations at the onset of treatment and was susceptible to all drugs, but at the second time point of sampling, they detected 4 segregating DR mutations (*katG* V1A, D94N, S315R, and *inhA* C-15T) that conferred resistance to INH (150,181). At the final time point of sampling, sequencing results showed that 94.2% of the reads within the sample had a single DR mutation (*katG* D94G), which suggested that this mutant had a higher fitness in comparison with the other three mutants (< 5%), including those containing DR mutations (*katG* G125V and L611R), which was not present at the second time point (150,181). This could mean that while Mtb rapidly acquires resistance mutations, different DR strains compete against each other due not HGT/recombination (150,179). The second patient had an Mtb population fixed for the DR mutation (*rpoB* L533P) but remained sensitive to RIF (150,181). Eighteen months later, the patient had a population that was dominated by another DR mutation (*rpoB* H526Y) and was now resistant to RIF, which suggests that successive sweeps of alternative DR mutations can lead to MDR strains (150,181).

Similarly, Eldholm et al.(2014) monitored the evolution of an XDR strain from a susceptible ancestor in a single patient (7). They showed that DR-associated mutations were acquired multiple times by individual clones, but only one of them successfully outcompeted and replaced the other clones (7). This suggests that adaptive mutants are fixed and become dominant while others are lost due to competition (7,27,179,182). An illustration of this process can be seen in Figure 2.5.

© University of Pretoria

2.8 Conclusion

How TB has evolved DR mechanisms to recognize and respond to antibiotics is largely unknown (43). In the quest to create new drugs, there is a great need to understand the underlying mechanisms that drive the DR evolution (27,151). Understanding the pathology of Mtb is difficult, as this bacterium remains elusive (27). To eradicate or at least control TB, it is important to understand the mechanisms that result in the pathogenesis of TB and how it survives selective pressures resulting from antibiotic use (27). To avoid the emergence of such deadly strains, understanding how these strains achieve sustained transmission especially before they become a healthcare problem, could inform for prevention and containment (151). Furthermore, understanding the mechanisms shaping the transmission of Mtb strains can provide a lead about the potential approaches for TB control (143).

© University of Pretoria

Chapter 3: Materials and Methods

3.1 Introduction

Whole genome sequencing (WGS) studies of *Mycobacterium tuberculosis* (Mtb) worldwide, have led to the identification of its genetic diversity and have provided some insight into its evolution and transmission (121). The sequencing of whole Mtb genomes provides comprehensive genetic information with which the genetic basis of drug resistance (DR) can be determined (39,72,107). The genetic diversity of Mtb can affect the baseline susceptibility towards drugs, it can modulate the acquisition and prevalence of DR, and affect the resulting phenotypic effects of DR (183). Studying the interplay between the genetic diversity of Mtb and the emergence of DR could provide important information on how to restrict the evolution of DR in Mtb (183). Unfortunately, studies on lineage-specific differences in Mtb are biased towards lineages 2 and 4, and not much is known about the other remaining lineages.

Studies on compensatory evolution have suggested the importance of compensatory mutations in the development and maintenance of DR in Mtb. Reports on compensatory mutations are mostly focused on first-line drugs, not much is known about compensatory mutations that exist in second line drugs (184). Furthermore, mutations often labelled as pre-requisite/stepping stone/intermediate mutations are suggested to affect the level of resistance ultimately leading to higher levels of resistance (184). The identification of low level resistance conferring mutations such as pre-requisite mutations can lead to the prevention or reversal of the resistance (185). Unfortunately, there are no well-curated databases of mutations causing intermediate levels of resistance (108).

In this study, we collected an extensive amount of sequence data and used statistical approaches to establish epistatic interactions between pre-requisite/stepping stone mutations, intermediate mutations, DR mutations, and compensatory mutations in the context of the genetic background of sequences (184). Firstly, linkage disequilibrium (LD) was used to establish non-random associations between mutations, then attributable risk statistics were used to determine directional dependencies of mutations and then statistically validated evolutionary paths towards resistance were determined. Additionally, universal markers of DR irrespective of the genetic background of sequences were also determined.

The data of this study was collected from multiple sources meaning that this is a meta-analysis study. Unfortunately, most studies investigating the evolution of DR in Mtb are done in vitro

© University of Pretoria

and not in vivo. This could limit understanding of the clinical relevance of these associations. However, the elucidation of epistatic interactions could explain the trends of DR seen in the clinical setting.

3.2 Data collection

An extensive search for sequence data and corresponding metadata was undertaken by searching through the NCBI database. Research articles reporting on WGS studies in Mtb from various countries were collected with the use of provided accession numbers of sequences as well as the metadata of sequence files. The metadata of these files had to include the phenotypic drug susceptibility (DS) profiles of each isolate and, optionally, the lineage they belonged to. These were the criteria with which sequence data was included for the downstream analysis. Sequence files in fastQ format were collected from the cryptic consortium project as well as metadata that included the DS profiles and lineage designations. The cryptic consortium project involves a study that analysed whole-genome sequences of Mtb from 16 countries and their associated phenotypes of resistance (186). The study aimed at predicting the susceptibility of sequences to the first-line drugs (186).

Variant call files (VCF) were also collected from the Pathosystems Resource Integration Centre (PATRICK) database and the Tuberculosis Antibiotic Resistance Catalogue Project (TB ARC). The PATRICK database was designed to assist scientists in the research of infectious diseases (187). This database provides information on bacterial genomes, data associated with the relevant genome analyses, computational tools, and a platform for bioinformatics analysis (187). The TB ARC project involved an international collaboration of researchers whose aim was to create a catalogue of mutations responsible for DR in Mtb to inform diagnostics (188). Lastly, VCF files were collected from the Genome-wide Mycobacterium tuberculosis variation (GMTV) database. This database consists of whole genome sequences from around the world, but mostly from the Russian Federation (189). The metadata included DS profiles and lineage designations. If the lineage designation of a strain was unknown, it was determined using the Resistance sniffer program. This program uses a computational algorithm that was developed to predict DR and lineage designations in Mtb isolates (104).

Although the number of sequenced Mtb genomes in public databases is huge and keeps growing, the majority of them are not provided with DS profiles and other metadata; therefore,

© University of Pretoria

these sequences were not suitable for this study. The initial number of sequence files (Illumina reads) collected was 17000+. The metadata and filtering of reads limited the number of suitable files. The lack of metadata largely excluded a significant number of the originally selected files. Additionally, files belonging to animal adapted lineages were removed. In total, 9388 fastQ or VCF files were selected from public sources, including the type strain H37Rv, which was assumed to be susceptible to all antibiotics. Numbers of obtained whole genome sequences of Mtb strains belonging to different lineages are shown in Table 3.

The development of this collection was aimed at improving our knowledge of how antibiotic resistance is associated with specific DR and supplementary mutations in the global Mtb population and in the different lineages. The DR mutations conferring resistance to the first and second line antibiotics were the focus of this study. An attempt was made to include all lineages of the *M. tuberculosis* complex. No DR mutations were found in lineage 7. Indeed, lineage 7 generally is not associated with a DR phenotype and largely these strains are susceptible to the antibiotics used against Mtb (190). This is why lineage 7 was excluded from this analysis.

Table 3: Numbers of isolates from each lineage of *M. tuberculosis* collected for the analysis.

Lineage	Count
Lineage_1 (unidentified to sub-lineages)	714
Lineage_1.2	26
Lineage_2 (Beijing)	2532
Lineage_3 (CAS)	1593
Lineage_4 (unidentified to sub-lineages)	1157
Lineage_4.1	3
Lineage_4 (Cameroon)	115
Lineage_4.1 (Ghana)	19
Lineage_4.1 (Haarlem)	782
Lineage_4.3 (LAM)	925
Lineage_4 (S-type)	298
Lineage_4 (Tur)	140
Lineage_4.6(Uganda)	31
Lineage_4.2 (Ural)	272
Lineage_4.1 (X-type)	707
Lineage_5	20
Lineage_6	54
Grand Total	9388

© University of Pretoria

3.3 Mapping and variant calling

Bioinformatics pipelines created for this study were guided by Walter et al. (2020), Bush et al (2020), and Olson et al. (2015) publications (191,192). Specifically, Olson et al. (2015) and Bush et al. (2020) described processes and parameter settings for mapping fastQ reads against reference genomes, variant calling, and filtering. Walter et al. (2020) suggestions were useful for the parameterization of variant inclusion and bash script writing for the creation of analytical pipelines. These publications provided this study with a guideline to optimize the use of bioinformatics software tools (Bowtie2 v2.3.4.1 & BCFTools v1.7) to limit false variant calls for both Single Nucleotide polymorphisms (SNPs) and Insertions and Deletions (INDELS) (193,194).

DNA reads in fastQ format were mapped against the H37Rv genome reference sequence using Bowtie2. False variant calls were removed and checked for duplications using Sambamba v0.5.0 (195). Sambamba is a software tool designed by Tarasov et al. (2015) for faster processing of SAM, BAM, and CRAM sequence alignment files (195). Variant calling and filtering were done using BCFtools. Stringent filtering thresholds were used to limit false variant calls: SNP gap of 3 (-g3), indel gap of 10 (-G10), exclude (-e) SNPs that have < 150 QUAL score, a DP less than 20, AF < 0.95, SP > 30, < 5 reads mapped at that position and reads in each direction that have < 1 read in support of the variant. The detailed bash script showing how read mapping, variant calling and variant filtering was done are provided in Appendix 1.

The aforementioned steps were done on a Linux-based server. Figure 3.1 shows a summary of the bioinformatics workflow used to produce analysis-ready VCF files. Most bioinformatics workflows used to determine variants typically remove PE and PPE regions, because they result in a large number of false variant calls (106,107). Thus, PE and PPE genes were excluded from consideration. Several in-house scripts were written on Python 3 to identify and validate genetic polymorphisms associated with DR development in Mtb lineages.

© University of Pretoria

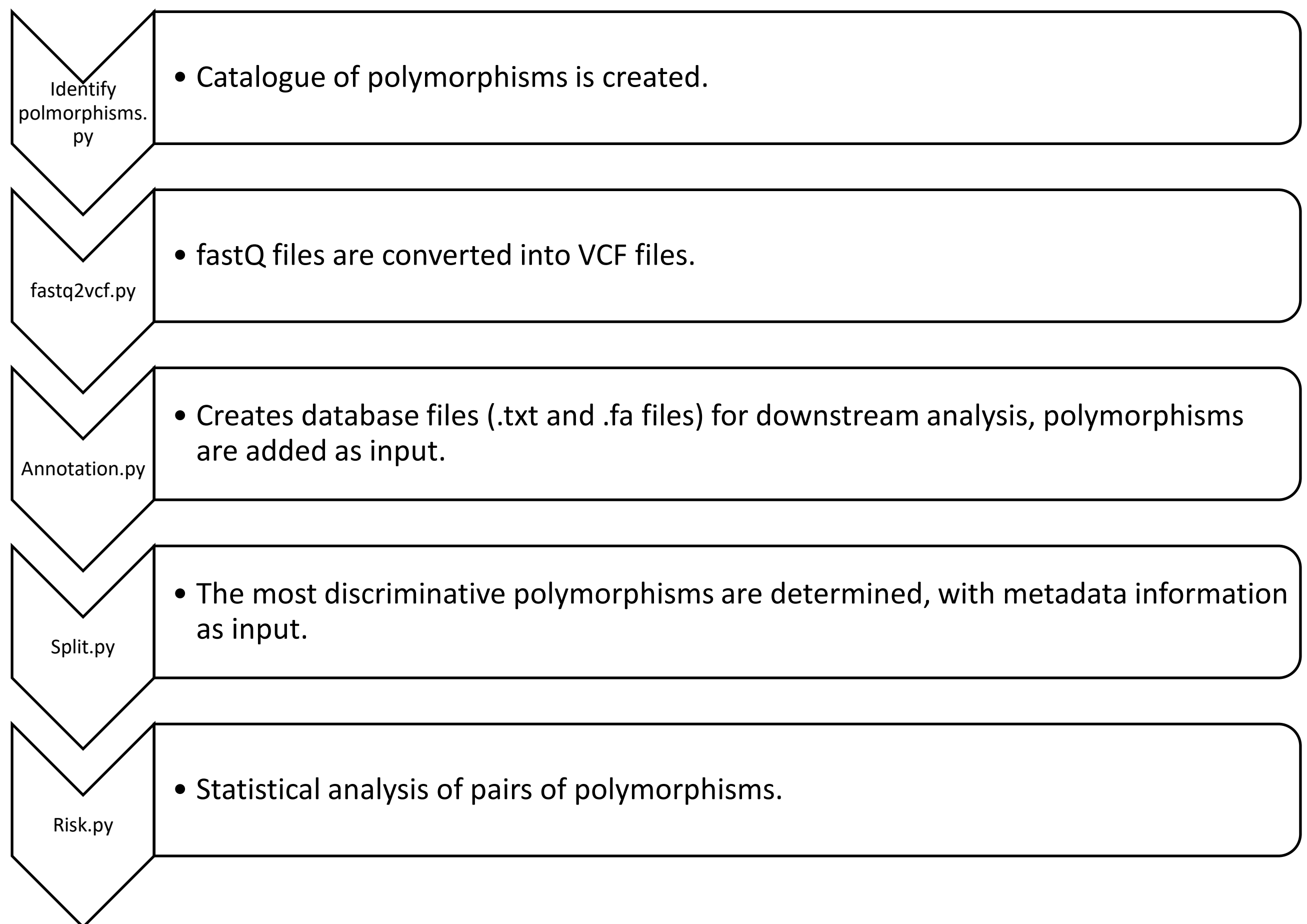


Figure 3.1: Analytical workflow of the project.

3.4 Split package

Statistical analysis was done with the use of an in-house collection of Python scripts called Split. These scripts were adapted from a previous NRF funded project on DR in Mtb. These scripts were published in the book chapter by van Niekerk et al. (2018). The package consists of five scripts/subprograms written in Python 3. The scripts can be run either on a local computer or on a Linux based server which can provide a better solution due to the large size of our datasets.

3.4.1 Identify_polymorphism.py

This script converts between different formats of representation of mutations in literature. Some studies represent polymorphic sites either by their location in the reference genome, i.e 2086672, or by the nucleotide sequence, or by the gene name and the locus tag, i.e glcB|310n or Rv1837c|310n respectively, or by the codon number and coding sequence, i.e GlcB|104 or Rv1258c|40 respectively. All of these polymorphic representations refer to the same mutation.

© University of Pretoria

This program may also be used to check the amino acid substitutions caused by a nucleotide substitution, if any. The command prompt interface of the script is shown in Figure 3.2.

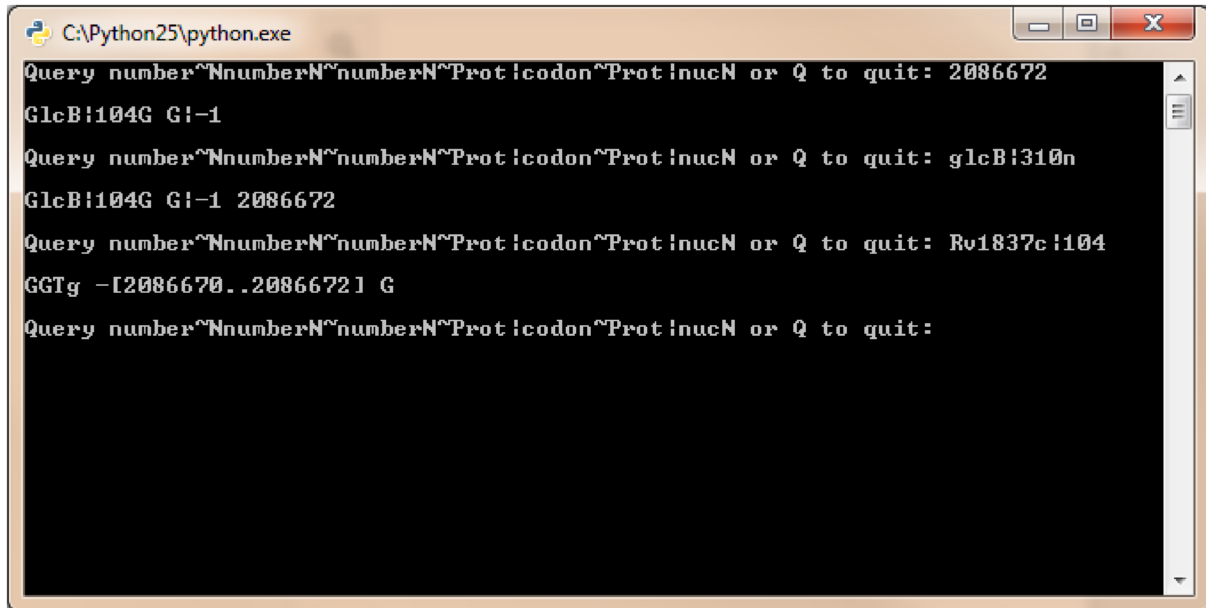


Figure 3.2: The command prompt interface of the script *identify_polymorphisms*.

3.4.2 Fastq2vcf.py

Most sequence data available is stored in a fastQ file. The script fastq2vcf converts fastQ files to VCF files using the R module Rsubread. It maps fastQ files against the provided reference genome, performs variant calling, and generates a VCF file. A command prompt interface of *fastq2vcf* collecting initial program run parameters is shown in Figure 3.3.

© University of Pretoria

```
SPE - fastq2vcf.py - Press Ctrl+Break to stop - C:\Python25\python.exe "fastq2vcf.py"

FASTQ mapping to reference 2020/04/10

Settings for this run:

  I     Project folder          :
  P     Single of paired end?   : paired-end

Y to accept these settings, type the letter for one to change or Q to quit

?i

Enter name of the project folder in 'fastq'? myreads

FASTQ mapping to reference 2020/04/10

Settings for this run:

  I     Project folder          : myreads
  P     Single of paired end?   : paired-end

Y to accept these settings, type the letter for one to change or Q to quit

?y

       Rsubread 1.34.7

//================================ setting =================================\\
||                                                                          ||
||             Index name : refseq                                          ||
||            Index space : base space                                      ||
||            Index split : no-split                                        ||
||       Repeat threshold : 100 repeats                                     ||
```

Figure 3.3: Command prompt interface of *fastq2vcf*.

3.4.3 Annotation.py

This script was used to produce database files used for downstream analysis by analysing the provided VCF files stored in a folder indicated as a program run parameter. The program produced two database files; one was an annotated database of polymorphisms found in the VCF files and the second one consisted of pseudo-sequences of amino acids at the polymorphic sites in fasta format. An example of a portion of the output of annotated polymorphisms and pseudo-sequences of amino acids can be found in Appendix 3 and 4 respectively. The command prompt interface of the script *annotation* is shown in Figure 3.4.

© University of Pretoria

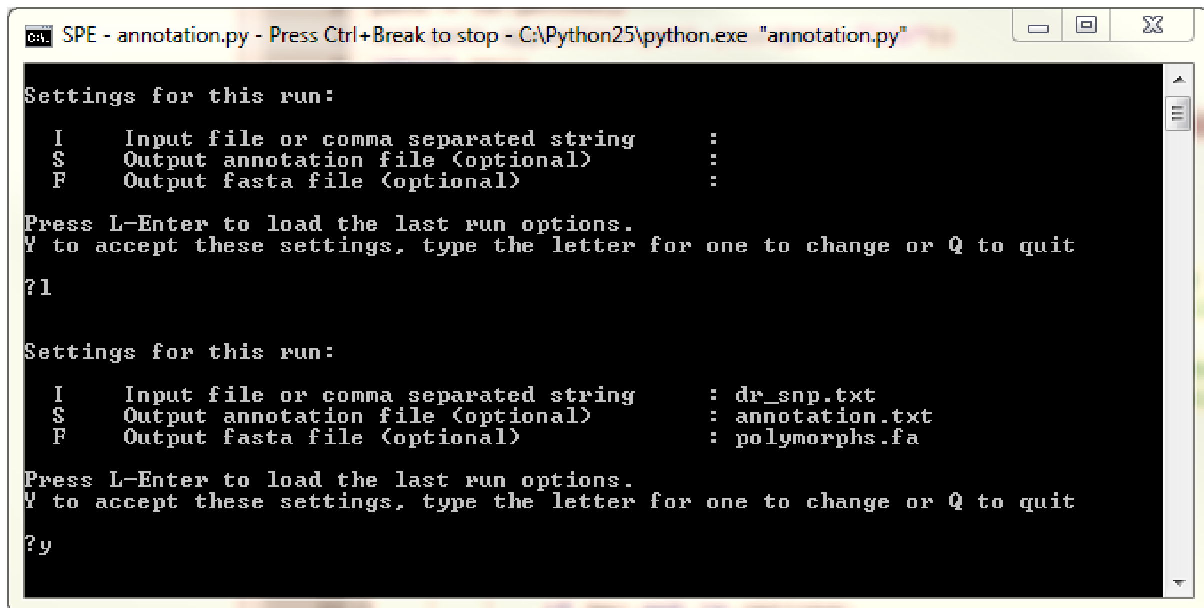


Figure 3.4: The command prompt interface of the script *annotation*.

3.4.4 Split.py

Metadata on the drug susceptibility profiles of each VCF file and the annotated polymorphisms generated by annotation.py were added as input into this program. A sample of the metadata can be found in appendix 2. This program allowed for the identification of the most powerful marker mutations to distinguish between 2 groups (for example, resistant or susceptible, or belonging to different lineages) of strains. The command prompt interface of the script *split* is shown in Figure 3.5.

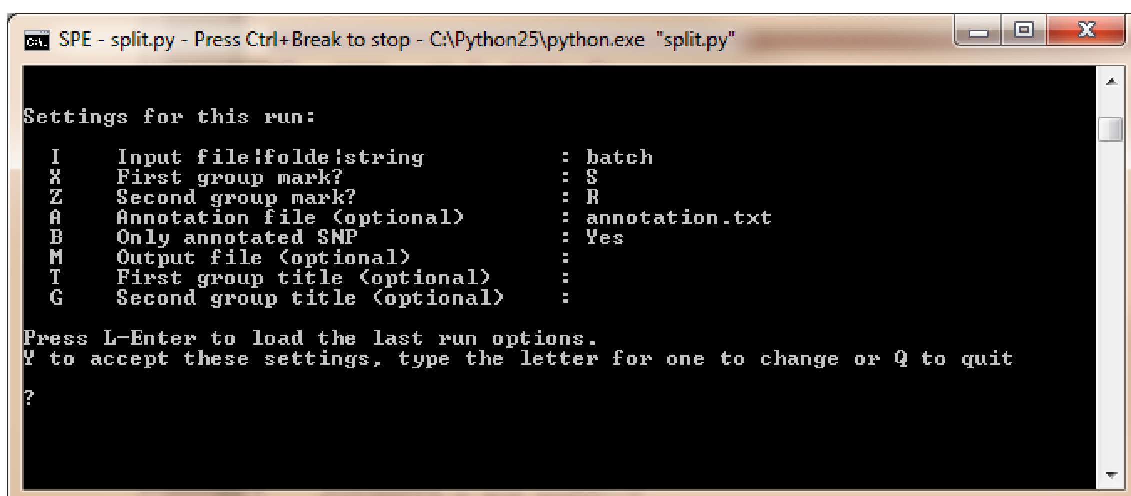


Figure 3.5: The command prompt interface of the script *split*.

© University of Pretoria

The most discriminative polymorphisms that were used as marker polymorphisms and could differentiate between susceptible and resistant strains were identified using the power value as shown in Eq. 1, which was published in a paper by Muzondiwa et al. 2020 (104).

$$Power_k = 1 - \frac{A \cap B}{\min(N_A, N_B)} \qquad (1)$$

where A∩B is the number of strains in the lineages *A* and *B* sharing the same allelic state of the locus *k*; N_A and N_B – sample sizes of the lineages *A* and *B*, respectively. Power values range from 0 to 1. The polymorphisms that had power values greater than 0.6 were selected for the downstream analysis.

3.4.5 Risk.py

This script identifies dependencies between pairs of mutations with the use of a linkage disequilibrium value (LD), chi2 test value, and the attributable risk statistic. Dependencies between pairs of mutations enabled the identification of epistatic links among mutations. The LD established if two non-random mutations were linked and the chi square test statistically validated links. The attributable risk statistic provided directional dependencies that enabled the identification of pre-requisite/initial, intermediate and compensatory mutations. Pre-requisite mutations were identified as initial mutations that had no mutations that pre-ceded them; this would identify the initial mutation at the start of an evolutionary path. Intermediate mutations had mutations that preceded and followed them. Compensatory mutations were mutations that were the last/final mutations to occur in an evolutionary path.

The fasta file and annotated polymorphisms produced by the script *annotation*, and pairs of marker polymorphisms predicted by the script *split*, are added as input to this subprogram. The output produced by this program shows the descriptors of the first and second polymorphic sites, for example, KatG|463R(-2) versus Stp|69D(-1) as shown in Figure 3.6 in the following format: gene|codon location, amino acid substitution, DNA strand of gene location and the number of substituted nucleotide within the codon in parentheses. The program then shows a contingency table in square brackets of how many times these two mutations co-existed in the same genome. The contingency table is followed by calculated linkage disequilibrium, Chi2, and attributable risk values of dependences of the first polymorphism from the second one (A-

© University of Pretoria

>a|b) and the second from the first one (B->b|a). For each dependence, the standard error is calculated as shown in the command prompt interface of the script *risk* in Figure 3.6.

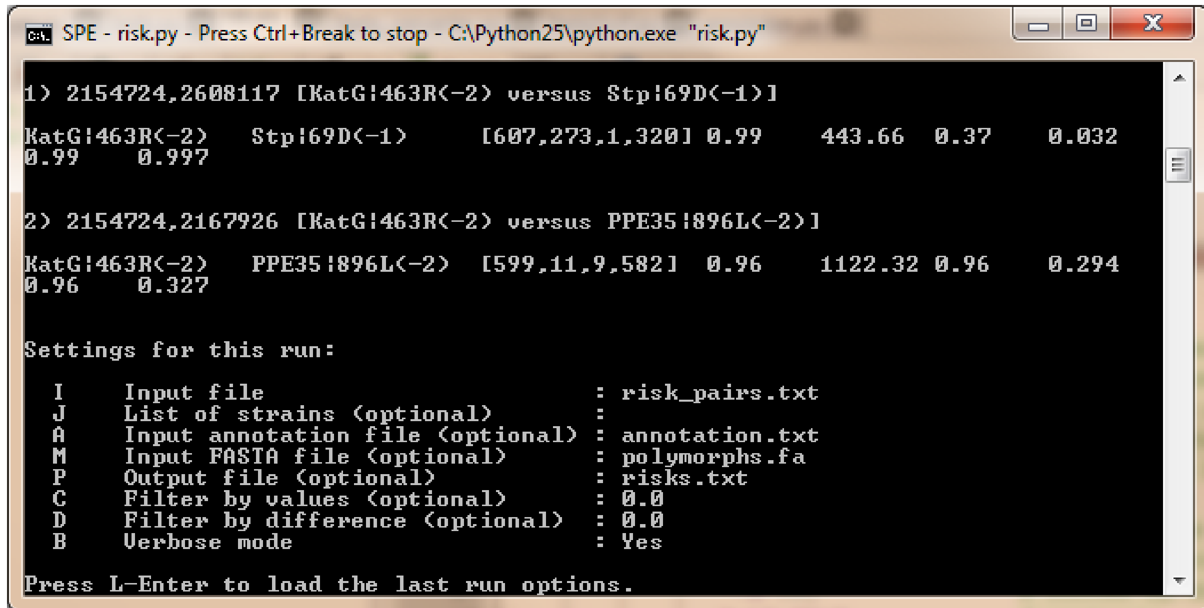


Figure 3.6: The command prompt interface of the script *risk*.

Pairs of the most discriminative polymorphisms were selected to determine any possible epistatic interactions. These pairs were added as input to the risk.py sub-program that implements Levin's attributable risk statistic algorithm (196) as explained in a previous publication by van Niekerk et al. (2018) and shown in Eq. 2.

$$R_a = \frac{F_{AB}F_{ab} - F_{aB}F_{Ab}}{(F_{AB}+F_{aB})(F_{aB}+F_{ab})} \qquad (2)$$

Where *Ra* is an attributable value of mutations A → a if the mutation B → b already took place and F_{xy} are frequencies of allele combinations among sampled Mtb strains.

Pairs of mutations that showed similar dependencies in both directions (A->a|b) and B->b|a), were excluded as these indicated the occurrence of genetic drift. Furthermore, polymorphisms that had negative attributable risk statistics showed negative epistasis meaning these associations among polymorphisms were incompatible with each other showing negative epistasis. These are less likely to occur in combination and thus were also excluded from the analysis. Pairs of mutations useful for this study were those, with attributable risk values greater

© University of Pretoria

than 0.5 recorded only for one dependence (either A->a|b) or B->b|a), which indicate the direction of epistatic relations. This concept is illustrated in Figure 3.7 adapted from the publication by van Niekerk et al. (2018).

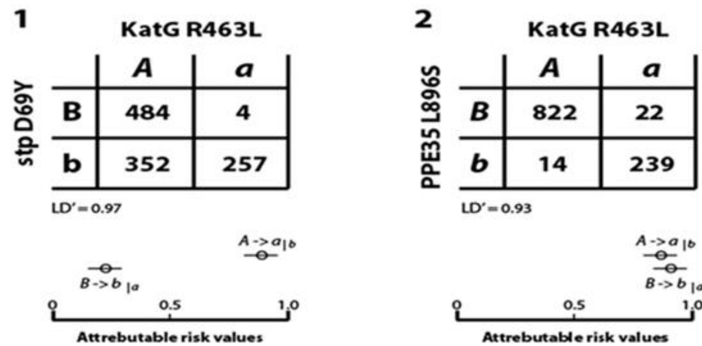


Figure 3.7: Two pairs of polymorphisms with the first pair showing omnidirectional dependency with the likelihood of a drug resistance (DR) mutation (KatG R463L) occurring when a mutation has already occurred in a secondary polymorphic site (stp D69Y). The likelihood of a mutation occurring in the opposite direction is less likely ($B \rightarrow b_{|a} < 0.5$ and $A \rightarrow a_{|b} > 0.5$). The second pair shows bidirectional dependency with the likelihood of a DR mutation occurring when a mutation has already occurred in a secondary polymorphic site (PPE35L896S) and a mutation occurring in a secondary polymorphic site when a mutation has already occurred in the DR mutation (both $B \rightarrow b_{|a}$ and $A \rightarrow a_{|b} > 0.5$). This indicates genetic drift and such pairs were removed from the analysis.

3.5 Data visualization

Output files containing statistical associations among polymorphisms produced from the *risk* script were added as input in the form of a text table to the software Cytoscape version 3.1.10. Cytoscape is an open-source software often used to visualise complex networks. Epistatic interactions included had a linkage disequilibrium value greater than 0.5, a chi-square value greater than 6.63 corresponding to a p-value of <0.01, and attributable risk statistical values showing omnidirectional dependencies greater than 0.5. Some polymorphisms with high power values were not included in the analysis of epistatic interactions due to the absence of statistically significant associations.

© University of Pretoria

Chapter 4: Identification of Universal lineage-independent markers of multidrug resistance in *M. tuberculosis*

4.1 Introduction

The identification of resistance genes is vital for the eradication of the tubercular disease. Mutations in genes at specific loci in the *Mycobacterium tuberculosis* (Mtb) genome are reported to confer resistance and are suggested as drug targets in the development of anti-TB drugs (197). The extensive use of antibiotics posed a selection pressure on Mtb and, in response, Mtb has acquired mutations over time that have enabled it to survive and thrive in unfavourable environments (198). Mutations conferring drug resistance (DR) are often found in drug targets or drug metabolism mechanisms (199). Well-known DR mutations are found in the genes *rpoB*, *inhA*, *katG*, *pncA*, *gyrA,* and *gyrB*, which confer resistance to rifampicin (RIF), isoniazid (INH), pyrazinamide (PZA) and fluoroquinolones (FQ's) respectively (198).

Often studies develop catalogues of DR-associated mutations to aid in diagnostics. A major challenge with the creation of these databases is that they are not well maintained/curated and in most cases lack clinical metadata (200). Therefore, reliable and comprehensive catalogues of genetic markers for DR-TB are lacking (200). Most reported DR mutations are associated with resistance to one drug or at most several drugs. Unfortunately, there is a limited amount of available evidence on mutations found in a single locus conferring resistance to multiple drugs (197). In this chapter, we identified universal markers of DR that suggest resistance to more than 3 antibiotics. We then used the attributable risk statistic to statistically validated epistatic interactions among the functional and non-functional genes to determine any evolutionary paths that would result in a "mutator phenotype" being required prior to the development of multidrug resistance.

4.2 Universal markers of antibiotic resistance in functional genes.

We found universal markers that may predict the likelihood of DR to more than three drugs independent of Mtb lineage designation. These links between allelic states of the identified polymorphic loci and DR yielded high power association values. A portion of the detailed description of each polymorphism can be found in Appendix 5. In total, 64 such polymorphic

© University of Pretoria

sites were identified within protein coding genes, out of which 47 sites were found in highly mutable PE/PPE regions and 17 were located in functional genes. Mutations in functional genes associated with multidrug resistance are shown in Table 4.1.

Table 4.1: Universal markers of drug resistance in *M. tuberculosis* found in functional genes

Locus tag	Gene name	Polymorphic Codon
Rv1269c	Conserved secreted protein	29
Rv0766c	Cytochrome P450 *cyp123*	278, 279
Rv3480c	Possible triacylglycerol synthase (diacylglycerol acyltransferase)	318
Rv1928c	Probable short-chain type dehydrogenase/reductase	190
Rv0045c	Possible hydrolase	83
Rv3854c	Monooxygenase *ethA*	337
Rv0823c	Possible transcriptional regulatory protein	322
Rv3919c	glucose-inhibited division protein *gidB*	65
Rv3346c	Conserved transmembrane protein	19
Rv1522c	Transmembrane transport protein *mmpL12*	549
Rv3093c	Hypothetical oxidoreductase	207
Rv3761c	Acyl-CoA dehydrogenase *fadE36*	303
Rv2330c	Lipoprotein *lppP*	20

© University of Pretoria

Rv1446c	OXPP cycle protein *opcA*	14
Rv2434c	Probable conserved transmembrane protein	317

We found statistically significant associations with the gene *cyp123* at codon position 279, which was associated with resistance to amikacin (AM), capreomycin (CM), ethambutol (EMB), isoniazid (INH), moxifloxacin (MFX), ofloxacin (OFX), prothionamide (PTO), pyrazinamide (PZA), rifampicin (RIF), and streptomycin (SM). Cyp123 forms part of the 20 subfamilies of cytochrome 450s present in Mtb (201,202). A GWAS study done by Lagutkin et al. (2022) found that SNPs in the Cyp123 gene were associated with resistance to STM, INH, and RIF (203). It was however not found at the codon position 279 but at M192A. A significant association between the mutation in FadE36 at codon position 303 with resistance to EMB, INH, PZA, RIF, and SM was found. FadE genes are suggested to function as acyl-CoA dehydrogenases, which play a role in lipid metabolism (204). The association of FadE36 to DR was not reported in the literature.

The gene *gidB* encodes an S-adenosyl methionine (SAM)-dependent 7-methylguanosine (m7G) methyltransferase, which performs methylation of 16S rRNA that is required for proper functionality of ribosomes. SM inhibits the ribosomal translational proofreading by binding the methylated site in 16S rRNA. Consequently, mutations in *gidB* prevent 16S rRNA methylation and reduce the binding affinity of the antibiotic resulting in SM resistance. However, amino acid substitutions at codon position 65 of these genes were not reported as significant for SM resistance. In our study, the Gid 65th-codon alterations were associated with DR to AM, EMB, OFX, PTO, PZA, and SM. The association of this mutation with the resistance to many antibiotics with different modes of action suggests that alteration in methylation of the 16S rRNA subunit is a key in Mtb's survival strategies.

MmpL12 is one of the transmembrane transports of a large MmpL family of transporters found in Mtb. A transmembrane transporter MmpL3 has been suggested to increase resistance to drugs that inhibit these transporter proteins (205). Studies on these proteins will improve knowledge of DR mechanisms and provide new drug targets (205). In MmpL12 there are no associations with DR found to date in the literature. In our study, however, we found a strong association of mutations in MmpL12 at codon position 549 with EMB, INH, PZA, RIF, and SM. This finding defines cell wall integrity as a cornerstone of resistance to various antibiotics.

© University of Pretoria

The hydrolase Rv0045 is suggested to be an esterase that involves lipid metabolism in Mtb (206,207) which leads us to one another confirmation that the adjustment of bacterial cellular membrane integrity is the central mechanism of increased drug resistance. In our results, we found that mutations in Rv0045 at codon position 83 have strong associations with resistance to AM, CM, EMB, MFX, OFX, PTO, and PZA. There are no confirmations of this fact in literature. Another mutable protein involved in cell wall biosynthesis is LppP lipoprotein. This protein was reported to be important for the survival of Mtb in macrophages and a 7 bp deletion in this gene was found to indicate a higher risk of DR (208). Our study showed that mutations in this gene at codon position 20 were associated with resistance to EMB, INH, PZA, RIF, and SM.

Rv0823c is a transcription regulatory protein. In our study, we found polymorphisms in Rv0823 at codon position 322 to be associated with resistance to AM, CM, EMB, MFX, OFX, PTO, and PZA. A study done by Raman and Chandra (2008) investigating possible pathways for DR determined that Rv0823 would be a suitable drug target to counter the emergence of DR, however, mechanisms of DR increasing due to mutations in this gene are not clear (209). Rv1269c encodes a secreted protein of unknown (203). In the same GWAS done by Lagutkin et al. (2022), they found that an 8 bp insertion was associated with resistance to SM, RIF, EMB, AM, and FQs. In our study, mutations in this gene at codon position 29 were associated with resistance to AM, CM, EMB, INH, MFX, PZA, RIF, and SM.

OpcA is a protein involved in the Oxidative Pentose Phosphate (OxPP) cycle and it is suggested to be overexpressed in INH-resistant clinical isolates (210,211). In our analysis, we did not find any direct association between the INH resistance and the OpcA mutations at codon position 14, but we found strong associations between these mutations and resistance to AM, CM, MFX, OFX, and PTO. The role of OpcA in Mtb is not clear, but in cyanobacteria, this protein is directly involved in the oxidative stress response which explains the over-expression of this gene as a factor of INH resistance by lowering the redox potential that prevents oxidation of the INH-prodrug in converting it into an active compound (212). Our study showed that fine-tuning the redox potential of the cell due to mutations in OpcA plays a significant role in increased resistance of Mtb to many antibiotics. Another protein, Rv1928c, encoding a dehydrogenase/reductase, is also involved in the maintenance of the cellular redox potential. This gene is suggested to play an important role in redox metabolism, which is suggested to be important for INH resistance in a study by Furio et al. (2021) (213). In our results, the mutations in Rv1928c at codon position 190 showed no associations with INH resistance but there were

© University of Pretoria

links with Mtb resistance to CM, EMB, MFX, OFX, PTO, PZA, and SM. This finding once more demonstrates that redox potential maintenance is a key point of multidrug resistance in Mtb.

Rv2434c encodes a conserved transmembrane protein and its association with DR is unclear. Mutations at codon position 317 were found in this study to be associated with resistance to AM, CM, MFX, OFX, and PTO. Rv3093c encodes an oxidoreductase that forms part of a gene cluster (Rv3093c-Rv3095c) (214). Increased expression of this gene cluster was said to be associated with ethionamide (ETH) resistance (214). The direct association of mutations in Rv3093c at codon position 207 with antibiotic resistance has not yet been established, but our study showed strong associations of these mutations with EMB, INH, PZA, RIF, and SM resistance. Rv3346c encodes a conserved transmembrane protein. No associations of mutations in this gene with DR have been reported except for the upregulation of this gene after INH exposure (215). We found that mutations in Rv3346c at codon position 19 were associated with EMB, INH, PZA, RIF and, SM resistance. Most likely Rv3093c and Rv3346c are two other genes involved in oxidative stress response in Mtb.

Rv3480c encodes a possible triacylglycerol synthase and is suggested to be a biomarker to differentiate between active and latent TB; therefore, it is potentially associated with INH resistance (216,217). We found an association between mutations in Rv3480c at codon position 318 and resistance to AM, CM, EMB, MFX, OFX, PTO, PZA, and RIF. The molecular or biological mechanisms of this association with drug resistance remained unclear.

EthA monooxygenase is commonly known to confer resistance to ETH (69,86). Mutations in *ethA* can lead to reduced EthA enzyme activity, thereby diminishing the activation of ETH prodrug. We have found that mutations at codon position 337 in this gene were associated with resistance to many other antibiotics, namely: AM, CM, EMB, OFX, PTO, PZA, and SM. The involvement of this enzyme in multidrug resistance has not been reported before.

4.3 Mutations in PE/PPE and PE_PGRS genes and their associations with multidrug resistance

PE and PPE are a family of genes that make up 10% of the Mtb genome and they are named after conserved proline (P) and glutamic acid (E) residues (218,219). The technology available in studying *pe* and *ppe* genes is challenged by the high GC content of these genes in turn making

© University of Pretoria

sequencing, alignment, and cloning these genes challenging (218). These genes are generally excluded from bioinformatics pipelines due to the difficulty of mapping these genes to the appropriate Mtb reference genome (218). Long-read sequencing is an alternative technology that is suggested to improve the understanding of *pe*/*ppe* genes (218). The PE subfamily is further divided into PE and the highly repetitive PE_polimorphic GC-rich sequences (PE_PGRS) proteins (219,220).

PE/PPE proteins are suggested to contribute to the virulence of Mtb and are suggested to have a close association with the ESX secretion systems (219,221). Most likely these proteins are secreted by different ESX systems with the majority being secreted by the ESX-5 secretion system (222). Type VII secretion systems (T7SS) are important virulence determinants that have been explored as novel drug and vaccine targets (223-225). There are five T7SS (ESX-1 – ESX-5) (226). The biological functions of PE and PPE proteins are not clearly understood; it is believed they function as outer membrane nutrient transport proteins, and play a role in host-pathogen interactions or immune evasion (218,227).

Our study showed strong associations between site-specific mutations in these genes and multidrug resistance (Table 4.2). While these associations were statistically strong, they did not necessarily mean the existence of specific functional links. PPE and PE_PGRS genes are mutational hotspots in Mtb genomes. We hypothesized that Mtb evolution towards multidrug resistance starts with a rise of sub-populations within the pathogen characterized by an increased rate of mutations known also as the "mutator phenotype". Thus, an increase in the rate of mutations will first be reflected in genomic loci, where the rate of mutation initially is high, i.e. in "mutation hotspots" such as PPE and PE_PGRS genes. Abnormally high levels of mutations in these genes, even though no specific functional links may exist, serve as a worrisome indicator of the mutator phenotype and speedy evolution of the pathogen towards multidrug resistance acquisition. To check this hypothesis, attributable risk associations between the above-mentioned mutations in functional and PPE/PE_PGRS genes were analysed and shown in Table 4.2.

Table 4.2: Universal markers of drug resistance in *M. tuberculosis* found in PE/PPE and non-functional genes

Locus tag	**Gene name**	**Codon**

© University of Pretoria

Rv1243c	PE_PGRS23	475
Rv1452c	PE_PGRS28	687
Rv0872c	PE_PGRS15	604
Rv1068c	PE_PGRS20	267
Rv0872c	PE_PGRS15	606
Rv3345c	PE_PGRS50	933
Rv1067c	PE_PGRS19	429
Rv3350c	PPE56	64
Rv1818c	PE_PGRS33	410
Rv1452c	PE_PGRS28	689
Rv2293c	hypothetical protein	220
Rv0278c	PE_PGRS3	552
Rv0278c	PE_PGRS3	203
Rv0578c	PE_PGRS7	859
Rv1883c	hypothetical protein	74
Rv0278c	PE_PGRS3	204
Rv0278c	PE_PGRS3	207
Rv1067c	PE_PGRS19	427
Rv1754c	hypothetical protein	12
Rv1243c	PE_PGRS23	475
Rv3344c	PE_PGRS49	443
Rv3344c	PE_PGRS49	138
Rv0305c	PPE6	782
Rv3344c	PE_PGRS49	139

© University of Pretoria

Rv3347c	PPE55	1486
Rv1067c	PE_PGRS19	332
Rv1803c	PE_PGRS32	546
Rv0963c	hypothetical protein	52
Rv0278c	PE_PGRS3	209
Rv1067c	PE_PGRS19	332
Rv0578c	PE_PGRS7	867
Rv3350c	PPE56	60
Rv3345c	PE_PGRS50	740
Rv1840c	PE_PGRS34	144
Rv1883c	hypothetical protein	75
Rv2923c	hypothetical protein	126
Rv1883c	hypothetical protein	74
Rv2975c	hypothetical protein	83
Rv1067c	PE_PGRS19	284
Rv1928c	Probable short-chain type dehydrogenase/reductase	134
Rv0872c	PE_PGRS15	604
Rv1450c	PE_PGRS27	364
Rv2262c	hypothetical protein	330
Rv1818c	PE_PGRS33	415
Rv0872c	PE_PGRS15	606
Rv1450c	PE_PGRS27	696

© University of Pretoria

4.4 Epistatic interactions among universal markers of drug resistance in *M. tuberculosis*

The attributable risk assessment approach was used to investigate possible dependences between universal multidrug resistance mutations identified in our study. The statistical approach allows comparison of likelihoods that one mutation can be a prerequisite of other mutations or that a mutation occurred only when other mutations took place thus enabling the reconstructing of evolutionary pathways showing the acquisition of antibiotic resistance by different strains. The results of this analysis are shown in Fig. 4.1.

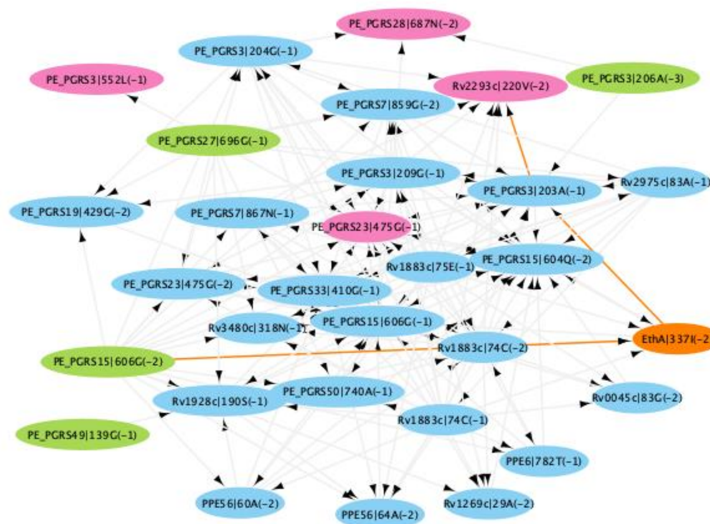


Figure 4.1: Epistatic interactions among mutations in global markers of drug resistance. The green node represents pre-requisite mutations, the blue nodes are intermediate mutations, and the pink nodes are final compensatory mutations. Epistatic links to known DR mutations are highlighted by orange edges. Nodes are labelled by the respective gene name and codon position number following the colon.

Figure 4.1 shows arrows indicating the chronological order of mutation acquisition from prerequisite and intermediate mutations to DR and compensatory mutations. Notably, mutations in the PE_PGRS genes act as initial nodes in the evolution towards multidrug resistance, supporting our hypothesis that pathogen evolution begins with a shift to a 'mutator phenotype' and the initial mutations firstly occur in the highly mutable, mostly non-functional, regions of PE_PGRS genes. Subsequent mutations, depicted as intermediate in the network, are found in the genes linked to cell wall integrity and oxidative stress response. A unique node,

© University of Pretoria

represented by the EthA gene mutations, receives all incoming arrows, suggesting it finalizes the multidrug resistance phenotype by potentially mitigating the fitness costs of the earlier mutations. This statistical analysis of attributable risk provides new insights into the complex pathways of antibiotic resistance development in *M. tuberculosis*. The following chapter delves deeper into the development of antibiotic resistance across different *M. tuberculosis* lineages.

© University of Pretoria

The evolution of drug resistance in the different lineages of the *Mycobacterium tuberculosis* complex.

Chapter 5: Identification of lineage-specific markers of antibiotic resistance and the reconstruction of epistatic interactions between them.

5.1 Introduction

Epistatic interactions provide a means by which we can determine optimal mutational pathways that lead to resistance and the compensation of fitness in the various lineages of Mtb (26). Ideally, the optimal evolutionary path should limit any fitness costs and this is dependent on the initial mutation and the pre-existing genetic context (17). In our results, we first created multiple subsets categorised by the lineage of our initial database to determine the most discriminative polymorphisms that distinguish between resistant and susceptible strains in various Mtb lineages, to study drug resistance (DR) evolution in these lineages. Discriminative polymorphisms had power values greater than 0.6. This power value indicated that there was a 60 percent chance that these polymorphisms were strongly associated with the antibiotic in question. Table 5.1 details the most discriminative polymorphisms that distinguish between resistant and susceptible strains in various Mtb lineages. A portion of the detailed information on the strains and their DR profile per lineage can be found in the appendix (Appendix 2).

Furthermore, we identified the likelihood of epistatic interactions through statistical analysis that enabled us to model evolutionary trajectories towards DR. Evolutionary trajectories were established with epistatic interactions that formed many evolutionary pathways with a direction

© University of Pretoria

between pre-requisite, intermediate, DR, and final/compensatory mutations. Direction in this context means determining the likelihood of DR mutations in question occurring when a mutation has already occurred or that a mutation occurs when a DR mutation has already occurred. The mutations discussed in our results are nonsynonymous mutations that resulted in amino acid changes. Pre-requisite mutations also known as stepping stone mutations may or may not be associated with DR. If associated with DR, they are suggested to confer lower/intermediate levels of resistance ultimately resulting in higher levels of resistance when another DR mutation takes place (228,229). Intermediate mutations form epistatic interactions between the pre-requisite and the final/compensatory mutations. Some intermediate mutations may play a compensatory role since compensation could either occur before or after DR.

Final mutations in our study were suggested to have a compensatory role. Not all final mutations had confirmed compensatory roles in literature and their function was not clear. Compensatory mutations can be alternative or additional mutations that occur in intra or intergenic loci (26,29). They affect bacterial epidemiology and treatment outcomes by preventing DR bacteria from being outcompeted by the original drug susceptible (DS) bacteria even when antibiotics are not present (26). These mutations do not contribute to DR on their own (11).

Our results show that there are several evolutionary trajectories that a strain can pursue. "Evolutionary trajectory refers to the order of changes that resulted in the evolution of one organismal (phenotypic or genotypic) state from another" (230). Multiple evolutionary paths resulted in various evolutionary trajectories. "Pathways are sequences of genetic changes that form chains in which each step facilitates the next, favouring, step by step, a significant increase in resistance" (231). Furthermore, there was a high diversity of mutations in MtbC in our results, however, only a few were linked to DR. Not all evolutionary paths were associated with DR, and not all lineages had known evolutionary paths leading to DR, hence they were not included for discussion. A detailed network of all interactions can be found in Appendix 36-43. Some mutations were associated with DR, whereas others occurred in virulence factors, which are crucial for survival of Mtb within the human host.

In our results, we found that the DR phenotype in different Mtb lineages is associated with genetic polymorphisms in different genes (Table 5.1). We then analysed associations between other lineage specific polymorphisms and DR mutations. The results of this study are presented in the following sections of this chapter.

© University of Pretoria

Table 5: The most discriminative polymorphisms that distinguish between resistant and susceptible strains in various *Mycobacterium tuberculosis* lineages.

Lineage	Antibiotic associations	DR polymorphisms
1	RIF, EMB, PZA, INH, and SM	*Gid*,
1.2	RIF, EMB, ETH, INH and SM	*WhiB6*, *EmbB*, *WmbC*, *Rv1258c*, *MmpL5*, *MshA*, *PncA*, *Rpsl*, *RpoB*, *KatG*, *GyrA*, *EmbC*
2	AM, CM, CS, EMB, ETH, INH, KM, MXF, OFX, PAS, PTO, PZA, RIF and SM	*EthA*, *WhiB6*, *GyrA*, *EmbB*, *PncA*,
3	INH, PTO, PZA, EMB, SM and RIF	*KatG*, *Gid*, *GyrA*, *MshA*, *RpoB*, *MmpL5*, *RpsL*, *PncA*
4	RIF, EMB, PZA, INH, and SM	*Gid*, *PncA*
4.1	INH and RIF	*GyrA*, *MshA*, *RpoB*, *RpoC*, *MmpL5*, *KatG*, *EmbC*, *Rv1258c*
4.6(Cameroon)	EMB and PZA	*RpoB*

© University of Pretoria

4.1(Ghana)	PZA	*RpoB*, *EmbB*, *PncA*, *UbiA*
4.1(Haarlem)	AM, CM, EMB, INH, MFX, OFX, PTO, PZA, RIF and SM	*UbiA*, *PncA*, *EmbB*, *EthA*, *Gid*
4.3(LAM)	INH, EMB, RIF, and PZA	*KatG*
4(S-type)	INH, EMB, RIF, and PZA	None
4(Tur)	INH, EMB, and, RIF	None
4.6(Uganda)	PZA	*EmbB*, *PncA*, *KatG*, *RpoB*, *EmbR*, *Rv2752c*, *EmbA*
4.1(X-type)	AM, CM, EMB, ETH, INH, KM, MXF, OXF, PTO, PZA, RIF, and SM	
4.2(Ural)	AM, CM, EMB, ETH, INH, KM, MFX, OFX, PTO, PZA, RIF, and SM	*Gid*, *RpoC*, *CcsA*, *RpsL*, *RpoB*,
5	RIF, EMB, PZA, and SM	*RpoB*, *EmbB*, *RpsL*
6	EMB, SM and PZA	*KatG*

A portion of the detailed list of all polymorphisms associated with each lineage can be found in Appendix 6-22. Additional information showing a portion of the statistical analysis of each epistatic pair can be found in Appendix 23-34.

© University of Pretoria

5.2 Epistatic interactions specific for Lineage 1.2

Lineage 1.2 has the largest diversity of evolutionary paths leading to DR when compared to other lineages collected in our study. Lineage 1 is largely known to be susceptible to antibiotic treatment. However, its sublineages 1.2.1 and 1.1.2 contribute significantly to the global burden of Mtb (135).

Figure 5.1: Epistatic interactions showing various evolutionary paths towards drug resistance (DR) in lineage 1.2. The green nodes represent pre-requisite mutations, the blue nodes are intermediate mutations, the orange nodes represent DR mutations, and the pink nodes are final/compensatory mutations. Epistatic links to and from known DR mutations are highlighted by orange edges. Nodes are labelled by the respective gene name and codon position numbers following the horizontal line.

Fig. 5.1 shows a rather complex network of pathways leading from many prerequisite and intermediate mutations to DR mutations ending up with a few compensatory mutations. The roles of the prerequisite mutations are difficult to understand as many of them occur in hypotheticals: Rv1873|3, Rv3818|86, Rv3896c|193, Rv3856c|306, Rv3897c|166, and Rv3912|213. It could be that these mutations did not have any specific biological consequences but simply reflected the transition of the strain to a mutator phenotype that initiated further DR evolution. This hypothesis is confirmed by the fact that one of the prerequisite mutations,

© University of Pretoria

MuT4|47, was found in the *mut4* gene associated with the mutator phenotype. In several publications, the MutT4 gene was suggested as a virulence factor of the Beijing lineage (lineage 2)(232,233).

Other prerequisite mutations were located in functional genes: possible glycerolphosphodiesterase Rv2277c|6; EccC2|258 ATPase, EccE2|206 and Rv3900c|21, components of the ESX-2 type VII secretion system; EsxE|80, a putative ESAT-6 like protein; conserved membrane protein Rv3888c|56; a polyketide synthase Pks7|1102; a phospholipase C4 Rv0074|46; and a UDP-glucose 4-epimerase GalE1|268. The association of several prerequisite mutations with the ESX-2 type VII secretion system indicates its importance for the survival of the Mtb pathogen and its role in diminishing the fitness cost; however, the exact function is not entirely clear (234). EsxE together with EsxF play a role in toxin secretion as reported in a paper by Tak et al. (2021); however, the association of these genes with the ESX secretion system is not entirely clear (235). Rv3895c forms part of a genomic island (Rv3884c–Rv3896c) suggested to contain the Esx2 gene cluster (236). Mutations in these genes were prerequisite mutations to a variety of pathways toward the evolution of DR to several antibiotics. Perhaps mutations in these genes improve the survival of Mtb within its host under antibiotic treatment.

Usually, one or several prerequisite mutations lead to DR mutations either directly, or through one or two intermediate mutations. Intermediate mutations occurred in triacylglycerol synthase Rv1760|219, uridylyltransferase GlnD|231, conserved transmembrane proteins Rv0064|549 and Rv1920|253, aldehyde dehydrogenase Rv0223c|454, EsxP|3 protein, Fad32|227 protein, and YidC translocase. One mutation in the efflux pump Rv3728|493 was marked in Figure 5.1 as an intermediate mutation; however, it can be considered a DR mutation preceding several other DR mutations. Rv3728 is believed to be overexpressed in multi-drug resistance isolates resistant to INH, EMB and SM in a study done by Gupta et al. (2010) and overexpressed during exposure to KM in a study done by Lee et al. (2021) (237,238).

Mutations known from the literature as directly associated with DR were deemed as DR mutations in Figure 5.1. They usually were located in the second or third nodes in the DR evolutionary pathways. These mutations included: GyrA|95 and GyrA|668 mutations in DNA gyrase that confers resistance to MXF (Moxifloxacin) and LEV (Levofloxacin); KatG|463 that encodes a catalase-peroxidase-peroxynitritase and confers resistance to INH; MmpL5|767 conferring resistance to bedaquiline (BDQ) and clofazimine (CZF); RpoB|731 encoding DNA-

© University of Pretoria

directed RNA polymerase that is associated with resistance to rifampicin (RIF); RpsL|43 that encodes 30S ribosomal protein S12 and confers resistance to SM; MshA|187 glycosyltransferase conferring resistance to INH and ETH; EmbB|406 conferring resistance to EMB; PncA|31 encoding pyrazinamidase/nicotinamidase that confers resistance to PZA; and mutation in transcriptional regulator WhiB6|53 conferring resistance to CAP, AMI, and SM. Many DR mutations formed paths of multidrug resistance evolution in Lineage 1.2 strains.

Many DR evolutionary pathways ended with the mutation DnaA|1 which looks like a general compensatory mutation in this lineage. However, the role of this mutation at the beginning of the chromosomal replication initiator protein is unclear. Other identified compensatory mutations were located in conserved membrane proteins Rv0012|258 and Rv2169c|121, monooxygenase Rv0897c|92, ABC transporter ModB|179 and hypotheticals Rv0454|56. In the case of DR mutations in GyrA, compensatory mutations GyrA|94 and GyrA20 also occurred.

Compensatory mutations identified in lineage 1.2 played a compensatory role in several resistance genes. A study done by Hicks et al. (2020) using the GWAS approach to identify DR mutations determined that the *dnaA* mutant conferred intermediate levels of resistance to INH enabling the bacterium to survive during INH treatment (239). In contrast, our study found DnaA to be perhaps a compensatory mutation instead to a number of mutations that confer resistance.

In this lineage, we also identified positive epistatic interactions among DR mutations, these included *katG*315/*embB*406, *embB*406/*rpoB*731, *rpoB*731/*mmpL5* 767, *mmpL5* 767/*katG*463, *gyrA*94/*katG*315, *pncA*31/*katG*463, *pncA*31/*embB*406, *pncA*31/*katG*315, *mmpL5* 948/*katG*463, *rpsL*43/*katG*315, *rpsL*43/*rpoB*731 which were some of the many positive interactions we found.

5.3 Epistatic interactions specific for Lineage 2

Figure 5.2 shows epistatic interactions revealed for strains belonging to Lineage 2 (Beijing). DR evolutionary pathways begin with several prerequisite mutations, which are located in EccE2|297 a conserved component of the type VII secretion system, and EspK|262 a secretion-associated protein. These prerequisite mutations lead directly to two DR mutations EmbB|295 conferring resistance to EMB, and WhiB6|51 conferring SM resistance. These mutations lead

© University of Pretoria

to several compensatory mutations including two mutations in hypothetical proteins Rv0095c|128 and 137.

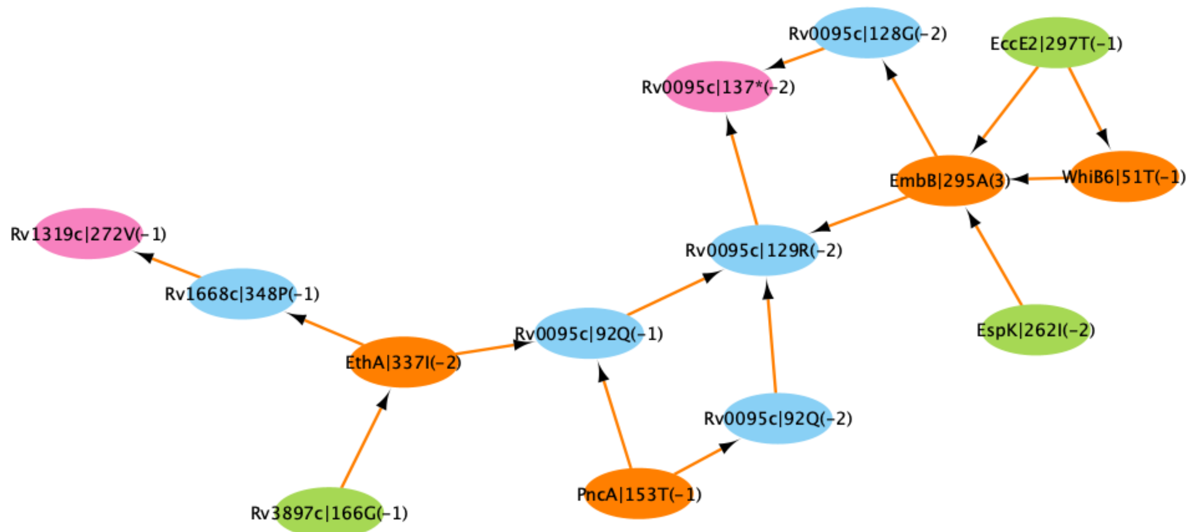


Figure 5.2: Epistatic interactions showing various evolutionary paths towards drug resistance (DR) in lineage 2. The green nodes represent pre-requisite mutations, the blue nodes are intermediate mutations, the orange nodes represent DR mutations, and the pink nodes are final/compensatory mutations. Epistatic links to and from known DR mutations are highlighted by orange edges. Nodes are labelled by the respective gene name and codon position numbers following the horizontal line.

DR mutation EthA|337 that encodes monooxygenase EthA and confers resistance to ETH depends on a preliminary mutation Rv3897c|166 in a hypothetical protein. Another DR mutation PncA|153 encosing pyrazinamidase/nicotinamidase confers resistance to PZA. This mutation is not dependent on preliminary mutations but leads to multiple mutations in the hypothetical protein Rv0095c, which were common for all DR mutations in this lineage. DR mutation EthA|337 has another set of compensatory mutations one in an ABC transporter Rv1668c|348 and the other in an adenylate cyclase Rv1319c|272.

Interestingly, lineage 2 according to our results showed no significant associations among DR mutations as suggested in the literature (135). This could be because our results were limited to the data collected, lineage 2 also demonstrates intra-lineage diversity suggested to have variable DR patterns, a data set of sub-lineage lineage 2 strains may be required, and/or there were no statically significant associations found in these mutations. The DR mutations specific

© University of Pretoria

to this lineage included EmbB, PncA, and WhiB6. The compensatory mutation Rv0095c had no known associations with DR and compensatory evolution in the literature.

5.4 Epistatic interactions specific for Lineage 3

Mtb strains of Lineage 3 are characterized by a tendency to MDR acquisition (Figure 5.3). Multiple DR evolutionary pathways start from four prerequisite mutations leading to a network of multiple intermediate mutations, which possibly play also the role of compensatory mutations since in most cases DR mutations were final nodes in the respective DR pathways.

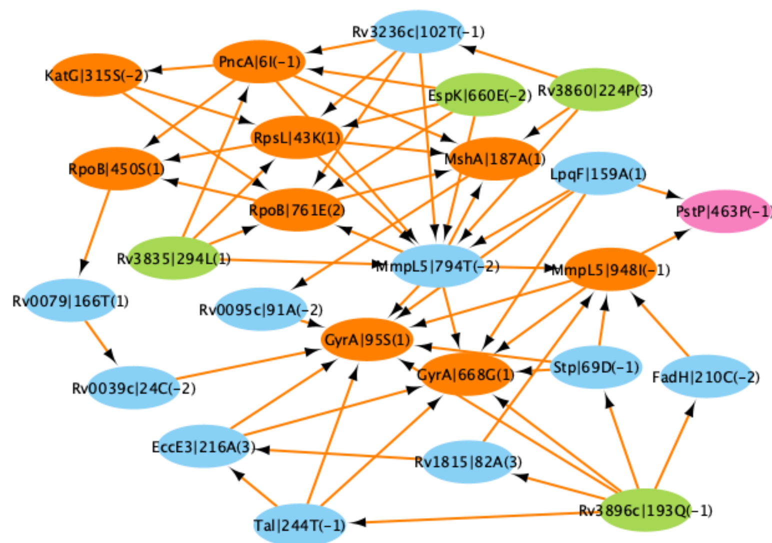


Figure 5.3: Epistatic interactions showing various evolutionary paths towards drug resistance (DR) in lineage 3. The green nodes represent pre-requisite mutations, the blue nodes are intermediate mutations, the orange nodes represent DR mutations, and the pink nodes are final/compensatory mutations. Epistatic links to and from known DR mutations are highlighted by orange edges. Nodes are labelled by the respective gene name and codon position numbers following the horizontal line.

The four preliminary mutations were located in hypotheticals Rv3896c|193, Rv3835|294, and Rv3860|224, and in the ESX-1 secretion-associated protein EspK|660. The central intermediate polymorphic protein with the biggest number of incoming and outcoming arrows is the MmpL5|794 transmembrane protein, which was reported to be associated with bedaquiline (BDQ) and clofazimine (CFZ) resistance (240). Many other intermediate mutations were within functional genes: FadH|210 a probable NADPH-dependent 2,4-dienoyl-CoA reductase; Stp|69 an integral membrane drug efflux protein, which was associated with DR in *M. bovis*

© University of Pretoria

(213,241); a transaldolase Tal|244; a lipoprotein LpqF|159; EccE3|216 type VII secretion system; and several others.

The following DR mutations were discovered: MmpL5|948 (BDQ and CFZ resistance), GryA|668 (MFX and LEV resistance), PncA|6 (PZA resistance), MshA|187 (ETH and INH resistance), RpsL|43 (SM resistance), and KatG|350 (INH resistance).

Mutations LpqF|159 and MmpL5|948 required a compensatory mutation PstP|436 encoding phosphoserine/threonine phosphatase (239). The study done by Li et al. (2020) found that phenotypically the overexpression of PstP led to resistance to ETH in *M. smegmatis* (242). In Mtb, this is yet to be determined (242).

In lineage 3, strong positive epistatic links were found among DR mutations *mmpL5* 948/*gyrA*668, *rpsL*43/*mshA*187, *pncA*6/*mshA*187, *katG*315/*pncA*6, and *rpsL*43/*katG*315.

5.5 Epistatic interactions specific for Lineage 4

The only strong epistatic link common for all isolates of lineage 4 begins at the PncA|136 mutation encodes for a pyrazinamidase/nicotinamidase and that confers drug resistance to pyrazinamide to the final/compensatory mutation OpcA|12 encodes for a putative OXPP cycle protein as shown in Figure 5.4.

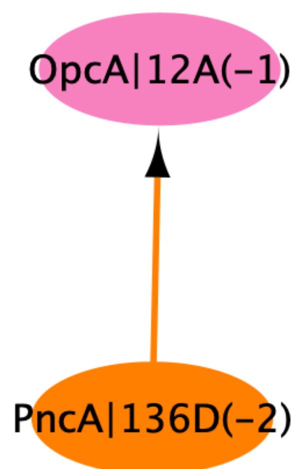


Figure 5.4: Epistatic interactions showing the mutational path in lineage 4. The orange node represents a drug resistance (DR) mutation and the pink node a final/compensatory mutation. The epistatic interaction from the known DR mutation is highlighted by the orange edge. Nodes

© University of Pretoria

are labelled by the respective gene name and codon position number following the horizontal line.

5.6 Epistatic interactions specific for Lineage 4.1

Two networks of DR evolution were detected in strains of this lineage (Figure 5.5).

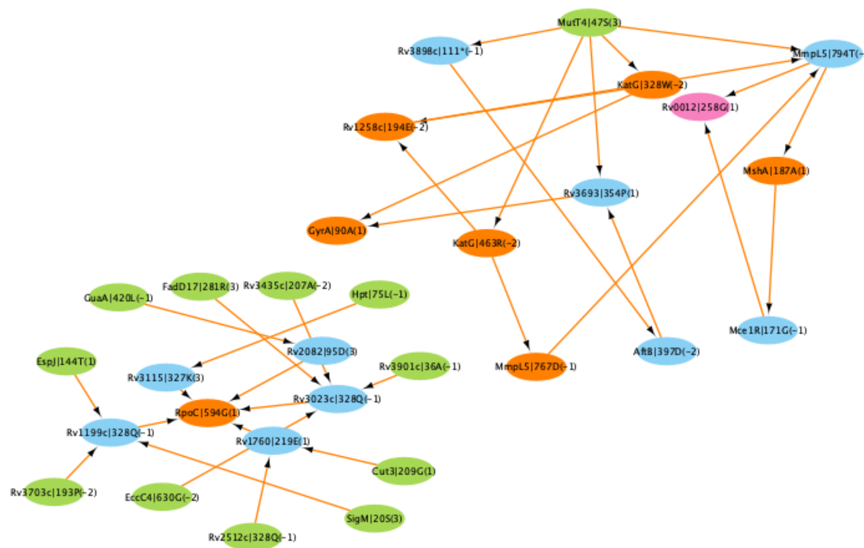


Figure 5.5: Epistatic interactions showing various evolutionary paths towards drug resistance (DR) in lineage 4.1. The green nodes represent pre-requisite mutations, the blue nodes are intermediate mutations, the orange nodes represent DR mutations, and the pink nodes are final/compensatory mutations. Epistatic links to and from known DR mutations are highlighted by orange edges. Nodes are labelled by the respective gene name and codon position numbers following the horizontal line.

Remarkably, the evolutionary paths of the larger network begin with the pre-requisite mutation in mutator protein MutT4|47. It paves the path for the next intermediate mutations in a hypothetical protein Rv3898c|111, a conserved membrane protein Rv3693|354, a transmembrane transport protein MmpL5|794, and two DR mutations, KatG|328 and KatG|463, both conferring resistance to INH. The two latter DR mutations preceded further MDR development by the acquisition of the following DR mutations: GyrA|90 (FQ resistance), efflux TAP protein Rv1258c|194 (RIF resistance), Mmpl5|767 (BDQ and CFZ resistance). Another DR mutation of this network, MshA|187, conferring INH and ETH resistance, requires preliminary mutations MmpL5|794 and in the transcriptional regulator Mce1|171. Then this DR mutation is followed by a compensatory mutation Rv0012c|258 a conserved membrane

© University of Pretoria

protein. Mce1 forms part of a family of proteins called Mce (mammalian cell entry) proteins that are important for the survival of Mtb in macrophages in the early stages of infection (243).

One of the intermediate mutations in this network is located in arabinofuranosyltransferase AftB|397. According to the WHO DR catalogue, *aftB* is associated with resistance to Amikacin (AMI) and capreomycin (CAP), however, the respective mutations were not at codon position 397.

The second network is centered on a DR mutation RpoC|594 rendering resistance to RIF. This mutation needs a complex genetic background of 11 prerequisite and 5 intermediate mutations in many genes that include the top-level transcriptional regulator SigM|205, cutinase precursor Cut3|209, hypoxanthine-guanine phosphoribosyltransferase Hpt|75, GMP synthase GauA|420, Fatty-acid-CoA synthetase FadD17|281, a secretion-associated protein EspJ|144 and others.

Regarding the SigM transcriptional factor, it is known that it activates a signal cascade of transcriptional regulation crucial for the survival of Mtb in macrophages (244,245). The transcription process in Mtb is controlled by 13 sigma factors (244). Primary sigma factors SigA and SigB control the basal expression of genes, the rest of the sigma factors regulate gene expression depending on environmental conditions (244). FadD17 is a fatty-acid-CoA synthetase belonging to the FadD family, which is important for the biosynthesis of mycolic acids (246).

In lineage 4.1 we also identified positive epistatic interactions among DR genes *gyrA*90/*katG*328, *mmpL5* 767/*katG*436 *katG*328/Rv1258c194, Rv1258c|*katG*463.

5.7 Epistatic interactions specific for Lineage 4.1 (Haarlem)

This lineage is the most frequent in Europe and North America. The network of epistatic interactions is shown in Figure 5.6.

© University of Pretoria

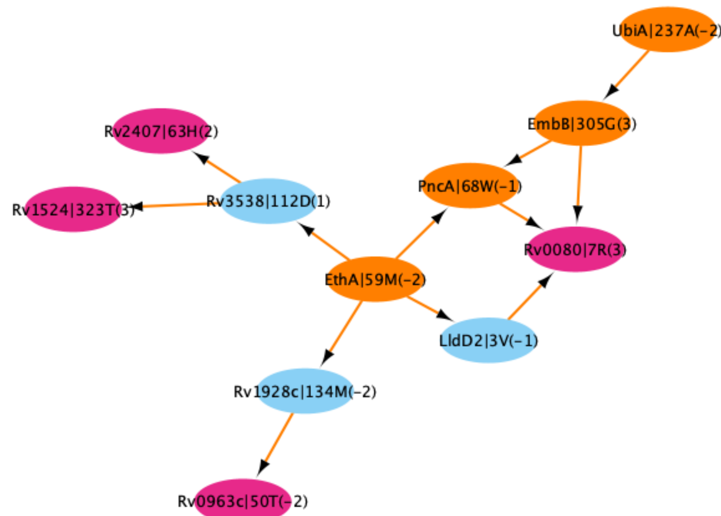


Figure 5.6: Epistatic interactions showing various evolutionary paths towards drug resistance (DR) in lineage 4.1(Haarlem). The blue nodes are intermediate mutations, the orange nodes represent DR mutations and/or pre-requisite mutations, and the pink nodes are final/compensatory mutations. Epistatic links to and from known DR mutations are highlighted by orange edges. Nodes are labelled by the respective gene name and codon position numbers following the horizontal line.

Four of the most common DR mutations found in the Haarlem lineage are EthA|59 (ETH resistance), PncA|68 (PZA resistance), EmbB|305 (EMB resistance), and UbiA|237 (EMB resistance). It looks like the strains of this lineage have the necessary epistatic background to acquire these mutations without any additional pre-requisite mutations. The DR mutations created an ordered path of MDR development (Figure 5.6). However, these mutations are followed by several compensatory mutations in various genes. Mutations in L-lactate dehydrogenase LldD2|3, short-chain type dehydrogenase/reductase Rv1928c|134, and 3-hydroxyacyl-thioester dehydratase HtdY Rv3538|112 were intermediates, which led to the final compensatory mutations in glycosyltransferase Rv1524|323 and in three hypotheticals. Also, the function of the gene Rv2407 is not known, a GWAS study by Lagutkin et al. (2022) determined that a 1bp deletion in this gene was associated with resistance to STM, INH, and FQ. In Mtb, epistatic interactions between DR mutations for a single drug can result in the multi-level acquisition of high-level resistance (171). Safi et al. (2013) conducted an in vitro study on the multi-step accumulation of EMB DR-conferring mutations that resulted in high-level resistance to EMB. They showed that the accumulation of mutations in the *Rv3806c/ubiA*, *Rv3792*, *embB,* and *embC* genes produced a variable number of MIC concentrations, ultimately

© University of Pretoria

resulting in high-level EMB resistance (228). Furthermore, Cohen et al. (2015) found that *ubiA* mutations emerged before EMB resistance conferred by *embB* (176). In our results, ubiA mutations led to an evolutionary path towards *embB* codon 305 known to confer high-level resistance in the lineage 4.1 (Haarlem). Determining pre-requisite mutations could enable the prediction of high-level resistance, improve treatment, and the discovery of new drugs (108,229).

5.8 Epistatic interactions specific for Lineage 4.2 (Ural)

Mtb strains of Ural lineage are abundant in Central and North Asia. The epistatic relationships between DR mutations in this lineage are shown in Figure 5.7.

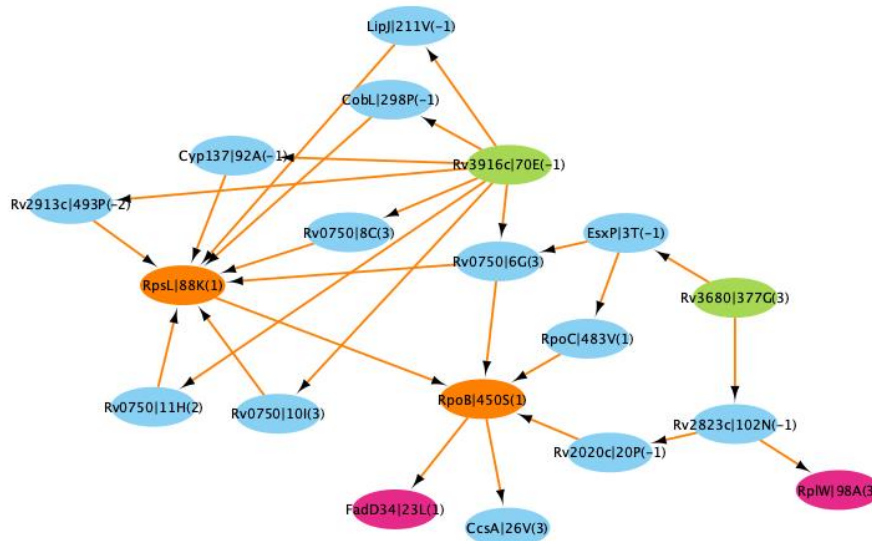


Figure 5.7: Epistatic interactions showing various evolutionary paths towards drug resistance (DR) in lineage 4.2 (Ural). The green nodes represent pre-requisite mutations, the blue nodes are intermediate mutations, the orange nodes represent DR mutations, and the pink nodes are final/compensatory mutations. Epistatic links to and from known DR mutations are highlighted by orange edges. Nodes are labelled by the respective gene name and codon position numbers following the horizontal line.

In contrast to the Haarlem strains, the evolution towards DR in the Ural strains required multiple prerequisite and intermediate mutations. Two common DR mutations were RpsL|88 (SM resistance) and RpoB|4505 (RIF resistance). DR evolution started from two basic prerequisite mutations in the anion transporter ATPase Rv3680|377 and the hypothetical Rv3916c|70. These mutations were followed by a series of mutations in various genes (Figure

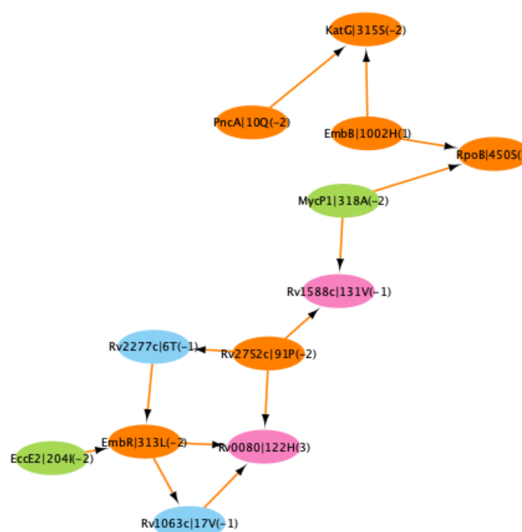
© University of Pretoria

5.7), many of which were hypotheticals. One functional gene mutation was located in fatty-acid-CoA ligase FadD34|23. FadD34 forms part of several oxidative stress genes found in Mtb (247). Rv2823c in a study done by Guo et al. (2022) was suggested to be associated with BDQ resistance, as it was included in the list of hitchhiking mutations found in BDQ-resistant isolates (248). Other functional intermediate and compensatory mutations were found in lignin peroxidase LipJ|221, precorrin-6Y C(5,15)-methyltransferase CobL|298, cytochromes Cyp137|92, D-amino acid aminohydrolase Rv2913c|463, DNA-directed RNA polymerase RpoC|483, ESAT-6 like protein EsxP|3, lignin peroxidase LipJ|211 and the cytochrome C-type biogenesis protein CcsA|26. Mutations in cytochromes may be important to cope with the oxidative stress generated by the host immune system (247). In response to oxidative stress Mtb produces reactive oxygen species ROS and reactive nitrogen intermediates RNIs, the FadD34 gene forms part of genes involved in oxidative stress because FadD34 modulates fatty acid synthase thereby regulating oxidative metabolism (247).

Compensation of DR can happen before or after the acquisition of a DR mutation. In our study, we found a compensatory mutation in the *rpoC* gene at codon position 483 before the DR mutation in the *rpoB* gene at codon position 450 in the Ural lineage. This compensatory mutation has been extensively described in several studies on its importance in compensating for resistance to RIF (249,250).

5.9 Epistatic interactions specific for Lineage 4.6 (Uganda)

Mtb pathogens of the Uganda lineage are frequent in Western and Central Africa. The epistatic interactions between DR mutations of this lineage are shown in Figure 5.8.



© University of Pretoria

Figure 5.8: Epistatic interactions showing various evolutionary paths towards drug resistance (DR) in lineage 4.6 (Uganda). The green nodes represent pre-requisite mutations, the blue nodes are intermediate mutations, the orange nodes represent DR mutations, and the pink nodes are final/compensatory mutations. Epistatic links to and from known DR mutations are highlighted by orange edges. Nodes are labelled by the respective gene name and codon position numbers following the horizontal line.

MDR development in these strains occurred through the independent accumulation of two mutations: PncA|10, which rendered PZA resistance, and EmbB|100, which rendered EMB resistance. These mutations were followed by two other DR mutations: KatG|315 (INH resistance) and RpoB (RIF resistance). The latter DR mutation was dependent on a prerequisite mutation in MycP1|318. Gene *mycP* encodes extracellular membrane-bound proteases called mycosins that are essential for the export and cleavage of the type VII-secreted virulence-associated proteins involved in the pathogenesis of *M. tuberculosis* (251).

Two other DR mutations of these strains required a specific arrangement of prerequisite and compensatory mutations, which included the EccE2|204 type VII secretion system protein, glycerolphosphodiesterase Rv2277c|6, and two hypotheticals Rv1063c|17 and Rv0080|122. This epistatic background allowed the acquisition of two DR mutations: EmbR|313 (EMB resistance) and Rv2752c|91 (INH resistance). The latter mutation in the gene encoding RNase J was described as a stepping-stone mutation associated with the MDR phenotype. It was found that the deletion of the RNase J caused more Mtb cells to survive standard antibiotic treatment (229). This MDR development in these strains requires one another common compensatory mutation Rv1588c|131 in the encoded conserved protein of unknown function.

© University of Pretoria

Chapter 6: Discussion and Conclusion

6.1 Discussion

Our study aimed at modelling the evolutionary trajectories of Mtb towards drug resistance (DR). We first did an extensive search in literature to unbiasedly represent the distribution of Mtb genomes among the human adapted lineages of the *M. tuberculosis* complex. We then established linked common polymorphisms by determining universal markers of DR in Mtb irrespective of lineage designation and, we determined lineage specific associations among lineages to model evolutionary trajectories toward DR. The extensive search for unbiased Mtb lineages found in several databases yielded 9388 genomes used for our analysis. Additionally, 37000 polymorphisms for analysis were found in our sequence data. To ensure that indeed a mutation was a DR mutation the WHO catalogue of DR mutations for both first and second line drugs was used.

Our second objective was aimed at determining linked common polymorphisms found in Mtb isolates irrespective of lineage designation. We identified 64 universal markers of DR to more than 3 antibiotics. Most markers in functional genes had confirmed associations with DR in literature. We also identified that redox potential maintenance is key in maintaining multiple DR in Mtb. Furthermore, we hypothesised that Mtb evolution towards multidrug resistance starts with a rise of sub-populations within the pathogen characterized by an increased rate of mutations known also as the "mutator phenotype". In our analysis, we confirmed that mutations in PE_PGRS genes are indeed the initial mutations that occur in the evolution towards multidrug resistance.

Lastly, we determined evolutionary trajectories towards a DR phenotype in Mtb. Little is known about epistatic interactions between DR mutations in Mtb (7). Whole genome sequencing (WGS) studies have shown that there seems to be a preference for a combination of certain DR resulting in higher *in vivo* fitness (17). These combinations suggest the occurrence of epistasis and could play a role in the successful emergence of MDR and XDR (7,51). Our results from the various lineages analysed, lineages 1.2, lineages 3, lineages 4, 4.1, 4.1 (Haarlem), 4.2 (Ural), and 4.6 (Uganda) showed several combinations of associated DR mutations forming positive epistatic interactions suggested reducing the fitness cost resulting from DR mutation acquisition (7). We found several stable combinations of mutations,

© University of Pretoria

including those confirmed in literature *rpsL*43/*katG*315, *rpsL*43/*rpoB*531, and *rpoB*531/*katG*315 (173). Therefore, the acquisition of one or additional DR mutations may compensate for fitness costs caused by the initial DR mutation (26).

We expected to see statistically significant associations among DR mutations in lineage 2 known as the Beijing lineage, but, interestingly, we did not find any statistically significant associations among DR mutations found in the collected strains of these lineages. Lineage 2 is often associated with a higher propensity to acquire DR mutations. This may be due to the inter-lineage diversity within this lineage that is suggested to have differences in the likelihood of DR resistance and an increased dataset may find higher frequencies of DR polymorphisms that may lead to better statistical associations with DR. Lineage 2 is the most widely studied on the effects of its diversity on DR phenotypes epidemiologically, clinically, and experimentally (252).

Lineage 2 is often linked with treatment failure, disease relapses, and outbreaks of MDR-TB (128,253). While hyper virulence has been associated with Lineage 2, this is not the case with all strains belonging to this lineage (252). This lineage is associated with hypo and hypervirulence (252,254,255). The differences in these strains can be attributed to genetically encoded discrepancies, epigenetic modification, and natural variation in protein expression (252). Therefore the genetic heterogeneity of large samples needs to be checked and if necessary reduced to sub-lineage level.

Depending on the genetic background of strains, a certain combination of mutations may be beneficial in one background but deleterious in another (20). This is a primary example of how epistasis could modulate the evolution towards DR (17). This type of epistasis not only affects the size but also the sign (beneficial or deleterious) of the fitness costs of mutations (169). In the concept of signed epistasis, one mutational pathway may be either beneficial in a particular genetic background or the same mutational pathway may be detrimental or less beneficial in a different background (169). In our results, we determined several mutational pathways unique to the lineages analyzed in our study. These mutational pathways are suggested to be specific to the lineage in question and play a vital role in the evolutionary trajectory toward a resistant phenotype.

Epistasis is necessary for pathogenic bacteria to improve their fitness cost (27). DR has long been believed to be associated with a reduction in fitness because antibiotics target essential and highly conserved genes (8,16,43,51). When mutations occur in these targeted genes, it

© University of Pretoria

results in the inactivation or the prevention of drug activity and in turn, affects the normal function of the gene (51). The clonal nature of Mtb means that negative fitness effects can only be alleviated through reversion or compensation because deleteriously mutated loci cannot be replaced through the transfer of intact loci from other strains (42). The reversion of resistance in Mtb is rarely observed hence a higher likelihood of the occurrence of compensatory evolution (26,27). Compensatory mutations of known and unknown functions in our study were found to play a role in the evolution of DR. The commonly known *rpoA* gene is suggested to compensate for fitness costs associated with rifampicin (RIF) resistance. In our study, we detected mutations in the *rpoA* gene at codon position 483 preceding the DR mutation in the *rpoB* gene at codon position 450 in the Ural lineage. The prerequisite mutation can be necessary to reduce the fitness cost to an acceptable level before the occurrence of the DR mutation.

The best-known examples of compensatory mutations are found in *rpoA* and *rpoC*, encoding the α and β'subunits of the RNA polymerase, which compensate for fitness costs associated with the resistance mutations in *rpoB* (16,26,27,51). Additionally, the combination of *rpoB* mutations with compensatory mutations in *rpoA* and *rpoC* is associated with the improved transmissibility of strains as evidenced by clonal exaptation of strains carrying these particular mutations that appear to restore resistance strains to wild-type function (171,176). Unfortunately, data on compensatory mutations is still limited and mainly focuses on INH and RIF (7).

The initial/pre-requisite mutations are important when determining an optimal path toward resistance (17). Pre-requisite mutations also known as stepping stone mutations may or may not be associated with DR. If associated with DR, they are suggested to confer lower/intermediate levels of resistance ultimately resulting in higher levels of resistance when other mutations take place (228,229). The level of resistance is key and determines the success of a strain and the process of epistasis is suggested to affect the level of resistance (168). Low-level resistance may enable the Mtb clonal lines to survive treatment providing an opportunity to acquire high-level resistance (229). The range of mutations that result in intermediate levels of resistance are not yet known and their clinical implication is not entirely clear (239).

Several studies have found stepping-stone mutations that were suggested to imply the success of DR in Mtb strains (229). In our results, lineage 4.6 (Uganda), has a pre-requisite mutation in Rv2752c towards resistance to the drug ethambutol (EMB). Efflux pumps are often

© University of Pretoria

suggested to be stepping stones to high-level resistance (77). In our study, we found prerequisite mutations in several efflux pumps and genes associated with increased levels of resistance, such as the Rv1258c/Tap gene in lineage 1.2 and the *ubiA* gene (171). Studies have found that *ubiA* genes emerge before the high level of ethambutol (EMB) resistance (176,228). In our results, *ubiA* mutations led to an evolutionary path towards *embB* substitution in codon 305 known to confer the high-level resistance in the lineage 4.1 (Haarlem). Determining pre-requisite mutations could enable predicting the development of high-level resistance, improving treatment protocols, and discovery of new drugs (108,229).

Not all polymorphisms found were directly linked to DR; some of them were found in genes playing important roles in the virulence and the pathogenesis of Mtb in the various lineages of Mtb (243). We identified mutations in genes involved in secretion systems, cell wall synthesis, and sigma factors all suggested to be important for the virulence of Mtb. Nevertheless, our analysis showed that these mutations were also associated with the evolutionary pathways towards antibiotic resistance. The full set of virulence factors found in each lineage is provided in Appendix 35.

Ultimately our study confirms that the evolutionary process of acquiring DR in Mtb involves a sequence of steps instead of occurring at once (256).

6.2 Conclusive remark

We must be able to predict the evolutionary trajectories of DR in Mtb. This in turn would improve the dosing regimens of antibiotics and ensure the effective use of antibiotics (168). Little or no experimental data is there that describes epistatic interactions that exist in Mtb, hence, it was appropriate to empirically determine associations among mutations in our study. It would be ideal to show the full sequence of mutational events and therefore determine the evolutionary consequences of each epistatic interaction in Mtb (169). Sequence divergent data may be used to study the role of epistasis or the use of experimental evolution, however, determining the exact order of mutations and their individual and combined effects on fitness is challenging even in simple organisms (41,169).

Experimental infection models provide clear evidence that members of the *M. tuberculosis* complex differ in virulence and immunogenicity in experimental infection models, but how this affects the progression toward disease in humans is not clear (128). Studies have yielded

© University of Pretoria

contradicting results on the effect of Mtb on strain diversity in clinical settings, hence we can only cautiously suggest whether Mtb diversity indeed does affect its human host (128). The disparity between in vitro isolate studies and clinical isolates may be due to the long evolution within the human body, the clonal interference, the variability in drug pressure, and the parallel evolution (7).

Further investigations on compensatory mutations in other antibiotics and not just INH and RIF are needed to determine the real influence of compensatory mutations in the DR evolution and the amelioration of the fitness cost in Mtb (26). This is crucial because identifying and inhibiting compensatory mechanisms may result in the disruption of DR acquisition and improve treatment outcomes (27). Even though recent improvements made in TB treatment are generous and must be acknowledged, however, significant gaps in our understanding of Mtb pathogenesis remain (8,223). Ultimately, understanding the epistatic interactions among DR mutations, the role of compensatory mutations, and the genetic background of Mtb will provide insight into the evolution of DR in Mtb (7).

© University of Pretoria

References

1. Timbrook, T. T., Hurst, J. M., and Bosso, J. A. (2016) Impact of an antimicrobial stewardship program on antimicrobial utilization, bacterial susceptibilities, and financial expenditures at an academic medical center. *Hospital Pharmacy* **51**, 703-711
2. Ventola, C. L. (2015) The antibiotic resistance crisis: part 1: causes and threats. *Pharmacy and therapeutics* **40**, 277
3. Koch, A., Mizrahi, V., and Warner, D. F. (2014) The impact of drug resistance on Mycobacterium tuberculosis physiology: what can we learn from rifampicin? *Emerging microbes & infections* **3**, 1-11
4. Davies, J., and Davies, D. (2010) Origins and evolution of antibiotic resistance. *Microbiology and molecular biology reviews* **74**, 417-433
5. Keshavjee, S., and Farmer, P. E. (2012) Tuberculosis, drug resistance, and the history of modern medicine. *New England Journal of Medicine* **367**, 931-936
6. Minssen, T., Outterson, K., Van Katwyk, S. R., Batista, P. H. D., Chandler, C. I., Ciabuschi, F., Harbarth, S., Kesselheim, A. S., Laxminarayan, R., and Liddell, K. (2020) Social, cultural and economic aspects of antimicrobial resistance. *Bulletin of the World Health Organization* **98**, 823
7. Nguyen, Q. H., Contamin, L., Nguyen, T. V. A., and Bañuls, A. L. (2018) Insights into the processes that drive the evolution of drug resistance in Mycobacterium tuberculosis. *Evolutionary applications* **11**, 1498-1511
8. Gygli, S. M., Borrell, S., Trauner, A., and Gagneux, S. (2017) Antimicrobial resistance in Mycobacterium tuberculosis: mechanistic and evolutionary perspectives. *FEMS microbiology reviews* **41**, 354-373
9. Chopra, K., Sidiq, Z., Hanif, M., and Dwivedi, K. K. (2020) Advances in the diagnosis of tuberculosis-Journey from smear microscopy to whole genome sequencing. *indian journal of tuberculosis* **67**, S61-S68
10. Perdigão, J., and Portugal, I. (2019) Genetics and roadblocks of drug resistant tuberculosis. *Infection, Genetics and Evolution* **72**, 113-130
11. O'Neill, M. B., Mortimer, T. D., and Pepperell, C. S. (2015) Diversity of Mycobacterium tuberculosis across evolutionary scales. *PLoS Pathog* **11**, e1005257
12. Gagneux, S. (2017) Strain variation in the Mycobacterium tuberculosis complex: its role in biology, epidemiology and control. Springer
13. Muzondiwa, D., Hlanze, H., and Reva, O. N. (2021) The Epistatic Landscape of Antibiotic Resistance of Different Clades of Mycobacterium tuberculosis. *Antibiotics* **10**, 857
14. Warner, D. F., Koch, A., and Mizrahi, V. (2015) Diversity and disease pathogenesis in Mycobacterium tuberculosis. *Trends in microbiology* **23**, 14-21
15. MacLean, R. C., and San Millan, A. (2019) The evolution of antibiotic resistance. *Science* **365**, 1082-1083
16. Allue Guardia, A., Garcia, J. I., and Torrelles, J. B. (2021) Evolution of drug-resistant Mycobacterium tuberculosis strains and their adaptation to the human lung environment. *Frontiers in Microbiology* **12**, 137
17. Borrell, S., and Trauner, A. (2017) Strain diversity and the evolution of antibiotic resistance. *Strain Variation in the Mycobacterium tuberculosis Complex: Its Role in Biology, Epidemiology and Control*, 263-279
18. Dookie, N., Rambaran, S., Padayatchi, N., Mahomed, S., and Naidoo, K. (2018) Evolution of drug resistance in Mycobacterium tuberculosis: a review on the molecular determinants of resistance and implications for personalized care. *Journal of Antimicrobial Chemotherapy* **73**, 1138-1151
19. Fenner, L., Egger, M., Bodmer, T., Altpeter, E., Zwahlen, M., Jaton, K., Pfyffer, G. E., Borrell, S., Dubuis, O., and Bruderer, T. (2012) Effect of mutation and genetic background on drug

© University of Pretoria

resistance in Mycobacterium tuberculosis. *Antimicrobial agents and chemotherapy* **56**, 3047-3053

20. Wong, A. (2017) Epistasis and the evolution of antimicrobial resistance. *Frontiers in Microbiology*, 246
21. Dheda, K., Perumal, T., Moultrie, H., Perumal, R., Esmail, A., Scott, A. J., Udwadia, Z., Chang, K. C., Peter, J., and Pooran, A. (2022) The intersecting pandemics of tuberculosis and COVID-19: population-level and patient-level impact, clinical presentation, and corrective interventions. *The Lancet Respiratory Medicine*
22. Casali, N., Nikolayevskyy, V., Balabanova, Y., Harris, S. R., Ignatyeva, O., Kontsevaya, I., Corander, J., Bryant, J., Parkhill, J., and Nejentsev, S. (2014) Evolution and transmission of drug-resistant tuberculosis in a Russian population. *Nature genetics* **46**, 279-286
23. Bhat, Z. S., Rather, M. A., Maqbool, M., Lah, H. U., Yousuf, S. K., and Ahmad, Z. (2017) Cell wall: a versatile fountain of drug targets in Mycobacterium tuberculosis. *Biomedicine & Pharmacotherapy* **95**, 1520-1534
24. Borrell, S., and Gagneux, S. (2011) Strain diversity, epistasis and the evolution of drug resistance in Mycobacterium tuberculosis. *Clinical Microbiology and Infection* **17**, 815-820
25. Connolly, L. E., Edelstein, P. H., and Ramakrishnan, L. (2007) Why is long-term therapy required to cure tuberculosis? *PLoS medicine* **4**, e120
26. Müller, B., Borrell, S., Rose, G., and Gagneux, S. (2013) The heterogeneous evolution of multidrug-resistant Mycobacterium tuberculosis. *Trends in Genetics* **29**, 160-169
27. Singh, R., Dwivedi, S. P., Gaharwar, U. S., Meena, R., Rajamani, P., and Prasad, T. (2020) Recent updates on drug resistance in Mycobacterium tuberculosis. *Journal of applied microbiology* **128**, 1547-1567
28. Coscolla, M., Gagneux, S., Menardo, F., Loiseau, C., Ruiz-Rodriguez, P., Borrell, S., Otchere, I. D., Asante-Poku, A., Asare, P., and Sánchez-Busó, L. (2021) Phylogenomics of Mycobacterium africanum reveals a new lineage and a complex evolutionary history. *Microbial genomics* **7**, 000477
29. Al-Saeedi, M., and Al-Hajoj, S. (2017) Diversity and evolution of drug resistance mechanisms in Mycobacterium tuberculosis. *Infection and drug resistance* **10**, 333
30. Ngabonziza, J. C. S., Loiseau, C., Marceau, M., Jouet, A., Menardo, F., Tzfadia, O., Antoine, R., Niyigena, E. B., Mulders, W., and Fissette, K. (2020) A sister lineage of the Mycobacterium tuberculosis complex discovered in the African Great Lakes region. *Nature communications* **11**, 1-11
31. Wang, F., Shao, L., Fan, X., Shen, Y., Diao, N., Jin, J., Sun, F., Wu, J., Chen, J., and Weng, X. (2015) Evolution and transmission patterns of extensively drug-resistant tuberculosis in China. *Antimicrobial agents and chemotherapy* **59**, 818-825
32. Gill, C. M., Dolan, L., Piggott, L. M., and McLaughlin, A. M. (2022) New developments in tuberculosis diagnosis and treatment. *Breathe* **18**
33. Organization, W. H. (2022) Global tuberculosis report 2021: supplementary material.
34. Coll, F., Phelan, J., Hill-Cawthorne, G. A., Nair, M. B., Mallard, K., Ali, S., Abdallah, A. M., Alghamdi, S., Alsomali, M., and Ahmed, A. O. (2018) Genome-wide analysis of multi-and extensively drug-resistant Mycobacterium tuberculosis. *Nature genetics* **50**, 307-316
35. Müller, B., Chihota, V. N., Pillay, M., Klopper, M., Streicher, E. M., Coetzee, G., Trollip, A., Hayes, C., Bosman, M. E., and Gey van Pittius, N. C. (2013) Programmatically selected multidrug-resistant strains drive the emergence of extensively drug-resistant tuberculosis in South Africa. *PloS one* **8**, e70919
36. Eldholm, V., and Balloux, F. (2016) Antimicrobial resistance in Mycobacterium tuberculosis: the odd one out. *Trends in microbiology* **24**, 637-648
37. Cox, H., Dickson-Hall, L., Jassat, W., Moshabela, M., Kielmann, K., Grant, A., Nicol, M., Black, J., Mlisana, K., and Vanleeuw, L. (2017) Drug-resistant tuberculosis in South Africa: history,

© University of Pretoria

progress and opportunities for achieving universal access to diagnosis and effective treatment. *South African Health Review* **2017**, 157-167

38. Perdigao, J., and Portugal, I. (2019) Genetics and roadblocks of drug resistant tuberculosis. *Infection, Genetics and Evolution* **72**, 113-130
39. Bradley, P., Gordon, N. C., Walker, T. M., Dunn, L., Heys, S., Huang, B., Earle, S., Pankhurst, L. J., Anson, L., and De Cesare, M. (2015) Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis. *Nature communications* **6**, 1-15
40. Dheda, K., Gumbo, T., Maartens, G., Dooley, K. E., McNerney, R., Murray, M., Furin, J., Nardell, E. A., London, L., and Lessem, E. (2017) The epidemiology, pathogenesis, transmission, diagnosis, and management of multidrug-resistant, extensively drug-resistant, and incurable tuberculosis. *The lancet Respiratory medicine* **5**, 291-360
41. Borrell, S., Teo, Y., Giardina, F., Streicher, E. M., Klopper, M., Feldmann, J., Müller, B., Victor, T. C., and Gagneux, S. (2013) Epistasis between antibiotic resistance mutations drives the evolution of extensively drug-resistant tuberculosis. *Evolution, medicine, and public health* **2013**, 65-74
42. Brites, D., and Gagneux, S. (2017) The nature and evolution of genomic diversity in the Mycobacterium tuberculosis complex. *Strain variation in the Mycobacterium tuberculosis complex: its role in biology, epidemiology and control*, 1-26
43. Nguyen, L. (2016) Antibiotic resistance mechanisms in M. tuberculosis: an update. *Archives of toxicology* **90**, 1585-1604
44. Barberis, I., Bragazzi, N. L., Galluzzo, L., and Martini, M. (2017) The history of tuberculosis: from the first historical records to the isolation of Koch's bacillus. *Journal of preventive medicine and hygiene* **58**, E9
45. Martini, M., Besozzi, G., and Barberis, I. (2018) The never-ending story of the fight against tuberculosis: from Koch's bacillus to global control programs. *Journal of preventive medicine and hygiene* **59**, E241
46. Viñuelas-Bayón, J., Vitoria, M. A., and Samper, S. (2017) Rapid diagnosis of tuberculosis. Detection of drug resistance mechanisms. *Enfermedades infecciosas y microbiologia clinica (English ed.)* **35**, 518-526
47. Keshavjee, S., Gelmanova, I., Pasechnikov, A., Mishustin, S., Andreev, Y., Yedilbayev, A., Furin, J., Mukherjee, J., Rich, M., and Nardell, E. (2008) Treating multidrug-resistant tuberculosis in Tomsk, Russia: developing programs that address the linkage between poverty and disease. *Annals of the New York Academy of Sciences* **1136**, 1-11
48. Hutchings, M. I., Truman, A. W., and Wilkinson, B. (2019) Antibiotics: past, present and future. *Current opinion in microbiology* **51**, 72-80
49. Bankier, R. G. (1965) Psychosis associated with cycloserine. *Canadian Medical Association Journal* **93**, 35
50. Kerantzas, C. A., and Jacobs Jr, W. R. (2017) Origins of combination therapy for tuberculosis: lessons for future antimicrobial development and application. *MBio* **8**, e01586-01516
51. Borrell, S., and Gagneux, S. (2009) Infectiousness, reproductive fitness and evolution of drug-resistant Mycobacterium tuberculosis [State of the art]. *The International Journal of Tuberculosis and Lung Disease* **13**, 1456-1466
52. Mitchison, D. (1954) Tubercle bacilli resistant to isoniazid. *British medical journal* **1**, 128
53. Middlebrook, G., and Cohn, M. L. (1953) Some observations on the pathogenicity of isoniazid-resistant variants of tubercle bacilli. *Science* **118**, 297-299
54. Pillay, M., and Sturm, A. W. (2007) Evolution of the extensively drug-resistant F15/LAM4/KZN strain of Mycobacterium tuberculosis in KwaZulu-Natal, South Africa. *Clinical infectious diseases* **45**, 1409-1414
55. Gillespie, S. H. (2001) Antibiotic resistance in the absence of selective pressure. *International journal of antimicrobial agents* **17**, 171-176

© University of Pretoria

56. Emane, A. K. A., Guo, X., Takiff, H. E., and Liu, S. (2021) Highly transmitted M. tuberculosis strains are more likely to evolve MDR/XDR and cause outbreaks, but what makes them highly transmitted? *Tuberculosis* **129**, 102092
57. Nathavitharana, R. R., and Friedland, J. S. (2015) A tale of two global emergencies: tuberculosis control efforts can learn from the Ebola outbreak. Eur Respiratory Soc
58. Gillespie, S. H. (2002) Evolution of drug resistance in Mycobacterium tuberculosis: clinical and molecular perspective. *Antimicrobial agents and chemotherapy* **46**, 267-274
59. Glynn, J. R., Whiteley, J., Bifani, P. J., Kremer, K., and van Soolingen, D. (2002) Worldwide occurrence of Beijing/W strains of Mycobacterium tuberculosis: a systematic review. *Emerging infectious diseases* **8**, 843
60. Bansal, R., Sharma, D., and Singh, R. (2018) Tuberculosis and its treatment: an overview. *Mini reviews in medicinal chemistry* **18**, 58-71
61. Acharya, B., Acharya, A., Gautam, S., Ghimire, S. P., Mishra, G., Parajuli, N., and Sapkota, B. (2020) Advances in diagnosis of Tuberculosis: an update into molecular diagnosis of Mycobacterium tuberculosis. *Molecular biology reports* **47**, 4065-4075
62. Sterling, T. R., Lehmann, H. P., and Frieden, T. R. (2003) Impact of DOTS compared with DOTS-plus on multidrug resistant tuberculosis and tuberculosis deaths: decision analysis. *Bmj* **326**, 574
63. Velayati, A. A., Masjedi, M. R., Farnia, P., Tabarsi, P., Ghanavi, J., ZiaZarifi, A. H., and Hoffner, S. E. (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
64. Udwadia, Z. F., Amale, R. A., Ajbani, K. K., and Rodrigues, C. (2012) Totally drug-resistant tuberculosis in India. *Clinical Infectious Diseases* **54**, 579-581
65. Klopper, M., Warren, R. M., Hayes, C., van Pittius, N. C. G., Streicher, E. M., Müller, B., Sirgel, F. A., Chabula-Nxiweni, M., Hoosain, E., and Coetzee, G. (2013) Emergence and spread of extensively and totally drug-resistant tuberculosis, South Africa. *Emerging infectious diseases* **19**, 449
66. Migliori, G., De Iaco, G., Besozzi, G., Centis, R., and Cirillo, D. (2007) First tuberculosis cases in Italy resistant to all tested drugs. *Weekly releases (1997–2007)* **12**, 3194
67. Horne, D. J., Pinto, L. M., Arentz, M., Lin, S.-Y. G., Desmond, E., Flores, L. L., Steingart, K. R., and Minion, J. (2013) Diagnostic accuracy and reproducibility of WHO-endorsed phenotypic drug susceptibility testing methods for first-line and second-line antituberculosis drugs. *Journal of clinical microbiology* **51**, 393-401
68. Nemeth, J., Oesch, G., and Kuster, S. P. (2015) Bacteriostatic versus bactericidal antibiotics for patients with serious bacterial infections: systematic review and meta-analysis. *Journal of Antimicrobial Chemotherapy* **70**, 382-395
69. Laborde, J., Deraeve, C., and Bernardes-Génisson, V. (2017) Update of antitubercular prodrugs from a molecular perspective: mechanisms of action, bioactivation pathways, and associated resistance. *ChemMedChem* **12**, 1657-1676
70. Nasiri, M. J., Haeili, M., Ghazi, M., Goudarzi, H., Pormohammad, A., Imani Fooladi, A. A., and Feizabadi, M. M. (2017) New insights in to the intrinsic and acquired drug resistance mechanisms in mycobacteria. *Frontiers in microbiology* **8**, 681
71. Unissa, A. N., Subbian, S., Hanna, L. E., and Selvakumar, N. (2016) Overview on mechanisms of isoniazid action and resistance in Mycobacterium tuberculosis. *Infection, Genetics and Evolution* **45**, 474-492
72. Phelan, J., Coll, F., McNerney, R., Ascher, D. B., Pires, D. E., Furnham, N., Coeck, N., Hill-Cawthorne, G. A., Nair, M. B., and Mallard, K. (2016) Mycobacterium tuberculosis whole genome sequencing and protein structure modelling provides insights into anti-tuberculosis drug resistance. *BMC medicine* **14**, 31

© University of Pretoria

73. Vilchèze, C., and Jacobs Jr, W. R. (2019) The isoniazid paradigm of killing, resistance, and persistence in Mycobacterium tuberculosis. *Journal of molecular biology* **431**, 3450-3461
74. Brucoli, F., and D McHugh, T. (2021) Rifamycins: do not throw the baby out with the bathwater. Is rifampicin still an effective anti-tuberculosis drug? *Future Medicinal Chemistry* **13**, 2129-2131
75. Grobbelaar, M., Louw, G. E., Sampson, S. L., van Helden, P. D., Donald, P. R., and Warren, R. M. (2019) Evolution of rifampicin treatment for tuberculosis. *Infection, Genetics and Evolution* **74**, 103937
76. Murray, J. F., Schraufnagel, D. E., and Hopewell, P. C. (2015) Treatment of tuberculosis. A historical perspective. *Annals of the American Thoracic Society* **12**, 1749-1759
77. Sheikh, B. A., Bhat, B. A., Mehraj, U., Mir, W., Hamadani, S., and Mir, M. A. (2021) Development of new therapeutics to meet the current challenge of drug resistant tuberculosis. *Current pharmaceutical biotechnology* **22**, 480-500
78. Liu, Y., Behr, M., Small, P., and Kurn, N. (2000) Genotypic Determination of Mycobacterium tuberculosisAntibiotic Resistance Using a Novel Mutation Detection Method, the Branch Migration Inhibition M. tuberculosis Antibiotic Resistance Test. *Journal of clinical microbiology* **38**, 3656-3662
79. Nikolayevsky, V., Brown, T., Balabanova, Y., Ruddy, M., Fedorin, I., and Drobniewski, F. (2004) Detection of mutations associated with isoniazid and rifampin resistance in Mycobacterium tuberculosis isolates from Samara Region, Russian Federation. *Journal of clinical microbiology* **42**, 4498-4502
80. Miotto, P., Cirillo, D. M., and Migliori, G. B. (2015) Drug Resistance inMycobacterium tuberculosis: Molecular Mechanisms Challenging Fluoroquinolones and Pyrazinamide Eff ectiveness. *Chest* **147**, 1135-1143
81. Lamont, E. A., Dillon, N. A., and Baughn, A. D. (2020) The bewildering antitubercular action of pyrazinamide. *Microbiology and Molecular Biology Reviews* **84**, e00070-00019
82. Scorpio, A., and Zhang, Y. (1996) Mutations in pncA, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculous drug pyrazinamide in tubercle bacillus. *Nature medicine* **2**, 662-667
83. Gengenbacher, M., and Kaufmann, S. H. (2012) Mycobacterium tuberculosis: success through dormancy. *FEMS microbiology reviews* **36**, 514-532
84. Zhang, Y., Permar, S., and Sun, Z. (2002) Conditions that may affect the results of susceptibility testing of Mycobacterium tuberculosis to pyrazinamide. *Journal of medical microbiology* **51**, 42-49
85. Xiang, X., Gong, Z., Deng, W., Sun, Q., and Xie, J. (2021) Mycobacterial ethambutol responsive genes and implications in antibiotics resistance. *Journal of Drug Targeting* **29**, 284-293
86. Morlock, G. P., Metchock, B., Sikes, D., Crawford, J. T., and Cooksey, R. C. (2003) ethA, inhA, and katG loci of ethionamide-resistant clinical Mycobacterium tuberculosis isolates. *Antimicrobial agents and chemotherapy* **47**, 3799-3805
87. Thee, S., Garcia-Prats, A., Donald, P., Hesseling, A., and Schaaf, H. (2016) A review of the use of ethionamide and prothionamide in childhood tuberculosis. *Tuberculosis* **97**, 126-136
88. Vilchèze, C., and Jacobs Jr, W. (2014) Resistance to isoniazid and ethionamide in Mycobacterium tuberculosis: genes, mutations, and causalities. Microbiol Spectr 2: MGM2-0014-2013. MGM2-0014-2013
89. Zhang, D., Gomez, J. E., Chien, J.-Y., Haseley, N., Desjardins, C. A., Earl, A. M., Hsueh, P.-R., and Hung, D. T. (2016) Genomic analysis of the evolution of fluoroquinolone resistance in Mycobacterium tuberculosis prior to tuberculosis diagnosis. *Antimicrobial agents and chemotherapy* **60**, 6600-6608
90. Scoper, S. V. (2008) Review of third-and fourth-generation fluoroquinolones in ophthalmology: in-vitro and in-vivo efficacy. *Advances in therapy* **25**, 979-994

© University of Pretoria

91. Almeida Da Silva, P. E., and Palomino, J. C. (2011) Molecular basis and mechanisms of drug resistance in Mycobacterium tuberculosis: classical and new drugs. *Journal of antimicrobial chemotherapy* **66**, 1417-1430
92. Li, G., Zhang, J., Jiang, Y., Zhao, L.-l., Liu, H., Li, M., Zhao, X., and Wan, K. (2020) Cross-resistance of isoniazid, para-aminosalicylic acid and pasiniazid against isoniazid-resistant Mycobacterium tuberculosis isolates in China. *Journal of global antimicrobial resistance* **20**, 275-281
93. Satta, G., Witney, A. A., Begum, N., Ortiz Canseco, J., Boa, A. N., and McHugh, T. D. (2020) Role of whole-genome sequencing in characterizing the mechanism of action of para-aminosalicylic acid and its resistance. *Antimicrobial Agents and Chemotherapy* **64**, e00675-00620
94. Kadura, S., King, N., Nakhoul, M., Zhu, H., Theron, G., Köser, C. U., and Farhat, M. (2020) Systematic review of mutations associated with resistance to the new and repurposed Mycobacterium tuberculosis drugs bedaquiline, clofazimine, linezolid, delamanid and pretomanid. *Journal of Antimicrobial Chemotherapy* **75**, 2031-2043
95. Lechartier, B., and Cole, S. T. (2015) Mode of action of clofazimine and combination therapy with benzothiazinones against Mycobacterium tuberculosis. *Antimicrobial agents and chemotherapy* **59**, 4457-4463
96. Peloquin, C. A., and Davies, G. R. (2021) The treatment of tuberculosis. *Clinical Pharmacology & Therapeutics* **110**, 1455-1466
97. Gómez-González, P. J., Perdigao, J., Gomes, P., Puyen, Z. M., Santos-Lazaro, D., Napier, G., Hibberd, M. L., Viveiros, M., Portugal, I., and Campino, S. (2021) Genetic diversity of candidate loci linked to Mycobacterium tuberculosis resistance to bedaquiline, delamanid and pretomanid. *Scientific reports* **11**, 1-13
98. Pi, R., Liu, Q., Jiang, Q., and Gao, Q. (2019) Characterization of linezolid-resistance-associated mutations in Mycobacterium tuberculosis through WGS. *Journal of Antimicrobial Chemotherapy* **74**, 1795-1798
99. Krause, K. M., Serio, A. W., Kane, T. R., and Connolly, L. E. (2016) Aminoglycosides: an overview. *Cold Spring Harbor perspectives in medicine* **6**, a027029
100. Palomino, J. C., and Martin, A. (2014) Drug resistance mechanisms in Mycobacterium tuberculosis. *Antibiotics* **3**, 317-340
101. Stanley, R. E., Blaha, G., Grodzicki, R. L., Strickler, M. D., and Steitz, T. A. (2010) The structures of the anti-tuberculosis antibiotics viomycin and capreomycin bound to the 70S ribosome. *Nature structural & molecular biology* **17**, 289-293
102. Chisompola, N. K., Streicher, E. M., Muchemwa, C. M. K., Warren, R. M., and Sampson, S. L. (2020) Molecular epidemiology of drug resistant Mycobacterium tuberculosis in Africa: a systematic review. *BMC infectious diseases* **20**, 1-16
103. Iketleng, T., Lessells, R., Dlamini, M. T., Mogashoa, T., Mupfumi, L., Moyo, S., Gaseitsiwe, S., and de Oliveira, T. (2018) Mycobacterium tuberculosis next-generation whole genome sequencing: opportunities and challenges. *Tuberculosis research and treatment* **2018**
104. Muzondiwa, D., Mutshembele, A., Pierneef, R. E., and Reva, O. N. (2020) Resistance sniffer: an online tool for prediction of drug resistance patterns of mycobacterium tuberculosis isolates using next generation sequencing data. *International Journal of Medical Microbiology* **310**, 151399
105. Aakre, C. D., and Laub, M. T. (2012) Asymmetric cell division: a persistent issue? *Developmental cell* **22**, 235-236
106. Cabibbe, A. M., Walker, T. M., Niemann, S., and Cirillo, D. M. (2018) Whole genome sequencing of Mycobacterium tuberculosis. *European Respiratory Journal* **52**
107. Meehan, C. J., Goig, G. A., Kohl, T. A., Verboven, L., Dippenaar, A., Ezewudo, M., Farhat, M. R., Guthrie, J. L., Laukens, K., and Miotto, P. (2019) Whole genome sequencing of Mycobacterium tuberculosis: current standards and open issues. *Nature Reviews Microbiology* **17**, 533-545

© University of Pretoria

108. Stanley, S., Liu, Q., and Fortune, S. M. (2022) Mycobacterium tuberculosis functional genetic diversity, altered drug sensitivity, and precision medicine. *Frontiers in Cellular and Infection Microbiology*, 1478
109. Dewhare, S. S. (2021) Drug resistant tuberculosis: Current scenario and impending challenges. *Indian Journal of Tuberculosis*
110. Walker, T. M., Kohl, T. A., Omar, S. V., Hedge, J., Elias, C. D. O., Bradley, P., Iqbal, Z., Feuerriegel, S., Niehaus, K. E., and Wilson, D. J. (2015) Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study. *The Lancet infectious diseases* **15**, 1193-1202
111. Prasad, R., Singh, A., and Gupta, N. (2019) Adverse drug reactions in tuberculosis and management. *indian journal of tuberculosis* **66**, 520-532
112. Silva, D. R., Mello, F. C. d. Q., and Migliori, G. B. (2020) Shortened tuberculosis treatment regimens: what is new? *Jornal Brasileiro de Pneumologia* **46**
113. Organization, W. H. (2022) *WHO consolidated guidelines on tuberculosis. Module 4: treatment-drug-susceptible tuberculosis treatment*, World Health Organization
114. Gold, B., and Nathan, C. (2017) Targeting phenotypically tolerant Mycobacterium tuberculosis. *Microbiology spectrum* **5**, 5.1. 27
115. Fernandes, G. F., Thompson, A. M., Castagnolo, D., Denny, W. A., and Dos Santos, J. L. (2022) Tuberculosis Drug Discovery: Challenges and New Horizons. *Journal of Medicinal Chemistry*
116. Wilkie, M. E., and McShane, H. (2015) TB vaccine development: where are we and why is it so difficult? *Thorax* **70**, 299-301
117. Scriba, T. J., Netea, M. G., and Ginsberg, A. M. (2020) Key recent advances in TB vaccine development and understanding of protective immune responses against Mycobacterium tuberculosis. in *Seminars in Immunology*, Elsevier
118. Martin, C., Aguilo, N., Marinova, D., and Gonzalo-Asensio, J. (2020) Update on TB vaccine pipeline. *Applied Sciences* **10**, 2632
119. Nimmo, C., Shaw, L. P., Doyle, R., Williams, R., Brien, K., Burgess, C., Breuer, J., Balloux, F., and Pym, A. S. (2019) Whole genome sequencing Mycobacterium tuberculosis directly from sputum identifies more genetic diversity than sequencing from culture. *BMC genomics* **20**, 389
120. Via, L. E., Cho, S.-N., Hwang, S., Bang, H., Park, S. K., Kang, H. S., Jeon, D., Min, S. Y., Oh, T., and Kim, Y. (2010) Polymorphisms associated with resistance and cross-resistance to aminoglycosides and capreomycin in Mycobacterium tuberculosis isolates from South Korean patients with drug-resistant tuberculosis. *Journal of clinical microbiology* **48**, 402-411
121. Advani, J., Verma, R., Chatterjee, O., Pachori, P. K., Upadhyaya, P., Singh, R., Yadav, J., Naaz, F., Ravikumar, R., and Buggi, S. (2019) Whole genome sequencing of Mycobacterium tuberculosis clinical isolates from India reveals genetic heterogeneity and region-specific variations that might affect drug susceptibility. *Frontiers in microbiology* **10**, 309
122. Chan, J. Z.-M., Sergeant, M. J., Lee, O. Y.-C., Minnikin, D. E., Besra, G. S., Pap, I., Spigelman, M., Donoghue, H. D., and Pallen, M. J. (2013) Whole-Genome Sequencing for Rapid Susceptibility Testing of M. tuberculosis.
123. Nikolayevskyy, V., Kranzer, K., Niemann, S., and Drobniewski, F. (2016) Whole genome sequencing of Mycobacterium tuberculosis for detection of recent transmission and tracing outbreaks: A systematic review. *Tuberculosis* **98**, 77-85
124. Ssengooba, W., Meehan, C. J., Lukoye, D., Kasule, G. W., Musisi, K., Joloba, M. L., Cobelens, F. G., and de Jong, B. C. (2016) Whole genome sequencing to complement tuberculosis drug resistance surveys in Uganda. *Infection, Genetics and Evolution* **40**, 8-16
125. Satta, G., Lipman, M., Smith, G., Arnold, C., Kon, O., and McHugh, T. (2018) Mycobacterium tuberculosis and whole-genome sequencing: how close are we to unleashing its full potential? *Clinical Microbiology and Infection* **24**, 604-609

© University of Pretoria

126. Dohál, M., Porvazník, I., Pršo, K., Rasmussen, E. M., Solovič, I., and Mokrý, J. (2020) Whole-genome sequencing and Mycobacterium tuberculosis: Challenges in sample preparation and sequencing data analysis. *Tuberculosis* **123**, 101946
127. Schleusener, V., Köser, C. U., Beckert, P., Niemann, S., and Feuerriegel, S. (2017) Mycobacterium tuberculosis resistance prediction and lineage classification from genome sequencing: comparison of automated analysis tools. *Scientific reports* **7**, 1-9
128. Coscolla, M., and Gagneux, S. (2010) Does M. tuberculosis genomic diversity explain disease diversity? *Drug Discovery Today: Disease Mechanisms* **7**, e43-e59
129. Albanna, A. S., Reed, M. B., Kotar, K. V., Fallow, A., McIntosh, F. A., Behr, M. A., and Menzies, D. (2011) Reduced transmissibility of East African Indian strains of Mycobacterium tuberculosis. *PloS one* **6**, e25075
130. Le Hang, N. T., Hijikata, M., Maeda, S., Miyabayashi, A., Wakabayashi, K., Seto, S., Diem, N. T. K., Yen, N. T. T., Van Duc, L., and Thuong, P. H. (2021) Phenotypic and genotypic features of the Mycobacterium tuberculosis lineage 1 subgroup in central Vietnam. *Scientific Reports* **11**, 1-12
131. Coscolla, M., and Gagneux, S. (2014) Consequences of genomic diversity in Mycobacterium tuberculosis. in *Seminars in immunology*, Elsevier
132. Duarte, T. A., Nery, J. S., Boechat, N., Pereira, S. M., Simonsen, V., Oliveira, M., Gomes, M. G. M., Penha-Gonçalves, C., Barreto, M. L., and Barbosa, T. (2017) A systematic review of East African-Indian family of Mycobacterium tuberculosis in Brazil. *Brazilian Journal of Infectious Diseases* **21**, 317-324
133. Coscolla, M. (2017) Biological and epidemiological consequences of MTBC diversity. *Strain variation in the Mycobacterium tuberculosis complex: its role in biology, epidemiology and control*, 95-116
134. Conceição, E. C., Refrégier, G., Gomes, H. M., Olessa-Daragon, X., Coll, F., Ratovonirina, N. H., Rasolofo-Razanamparany, V., Lopes, M. L., Van Soolingen, D., and Rutaihwa, L. (2019) Mycobacterium tuberculosis lineage 1 genetic diversity in Pará, Brazil, suggests common ancestry with east-African isolates potentially linked to historical slave trade. *Infection, Genetics and Evolution* **73**, 337-341
135. Netikul, T., Palittapongarnpim, P., Thawornwattana, Y., and Plitphonganphim, S. (2021) Estimation of the global burden of Mycobacterium tuberculosis lineage 1. *Infection, Genetics and Evolution* **91**, 104802
136. Guimarães, A. E. d. S., Sharma, A., Furlaneto, I. P., Rutaihwa, L., Cardoso, J. F., da Conceição, M. L., Spinassé, L. B., Machado, E., Lopes, M. L., and Duarte, R. S. (2021) Evaluation of drug susceptibility profile of Mycobacterium tuberculosis Lineage 1 from Brazil based on whole genome sequencing and phenotypic methods. *Memórias do Instituto Oswaldo Cruz* **115**
137. Rindi, L., Lari, N., and Garzelli, C. (2012) Large Sequence Polymorphisms of the Euro-American lineage of Mycobacterium tuberculosis: a phylogenetic reconstruction and evidence for convergent evolution in the DR locus. *Infection, Genetics and Evolution* **12**, 1551-1557
138. Merker, M., Blin, C., Mona, S., Duforet-Frebourg, N., Lecher, S., Willery, E., Blum, M. G., Rüsch-Gerdes, S., Mokrousov, I., and Aleksic, E. (2015) Evolutionary history and global spread of the Mycobacterium tuberculosis Beijing lineage. *Nature genetics* **47**, 242-249
139. Gupta, A., Sinha, P., Nema, V., Gupta, P. K., Chakraborty, P., Kulkarni, S., Rastogi, N., and Anupurba, S. (2020) Detection of Beijing strains of MDR M. tuberculosis and their association with drug resistance mutations in katG, rpoB, and embB genes. *BMC infectious diseases* **20**, 1-7
140. Hakamata, M., Takihara, H., Iwamoto, T., Tamaru, A., Hashimoto, A., Tanaka, T., Kaboso, S. A., Gebretsadik, G., Ilinov, A., and Yokoyama, A. (2020) Higher genome mutation rates of Beijing lineage of Mycobacterium tuberculosis during human infection. *Scientific reports* **10**, 1-11
141. Ribeiro, S. C., Gomes, L. L., Amaral, E. P., Andrade, M. R., Almeida, F. M., Rezende, A. L., Lanes, V. R., Carvalho, E. C., Suffys, P. N., and Mokrousov, I. (2014) Mycobacterium tuberculosis

© University of Pretoria

strains of the modern sublineage of the Beijing family are more likely to display increased virulence than strains of the ancient sublineage. *Journal of clinical microbiology* **52**, 2615-2624

142. Niemann, S., and Supply, P. (2014) Diversity and evolution of Mycobacterium tuberculosis: moving to whole-genome-based approaches. *Cold Spring Harbor perspectives in medicine* **4**, a021188
143. Wampande, E. M., Naniima, P., Mupere, E., Kateete, D. P., Malone, L. L., Stein, C. M., Mayanja-Kizza, H., Gagneux, S., Boom, W. H., and Joloba, M. L. (2019) Genetic variability and consequence of Mycobacterium tuberculosis lineage 3 in Kampala-Uganda. *PloS one* **14**, e0221644
144. Shah, Y., Maharjan, B., Thapa, J., Poudel, A., Diab, H. M., Pandey, B. D., Solo, E. S., Isoda, N., Suzuki, Y., and Nakajima, C. (2017) High diversity of multidrug-resistant Mycobacterium tuberculosis Central Asian Strain isolates in Nepal. *International Journal of Infectious Diseases* **63**, 13-20
145. Shuaib, Y. A., Utpatel, C., Kohl, T. A., Barilar, I., Diricks, M., Ashraf, N., Wieler, L. H., Kerubo, G., Mesfin, E. A., and Diallo, A. B. (2022) Origin and Global Expansion of Mycobacterium tuberculosis Complex Lineage 3. *Genes* **13**, 990
146. Yeboah-Manu, D., de Jong, B. C., and Gehre, F. (2017) The biology and epidemiology of Mycobacterium africanum. *Strain Variation in the Mycobacterium tuberculosis Complex: Its Role in Biology, Epidemiology and Control*, 117-133
147. Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K. E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., and Thwaites, G. (2013) Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. *Nature genetics* **45**, 1176-1182
148. Yimer, S. A., Namouchi, A., Zegeye, E. D., Holm-Hansen, C., Norheim, G., Abebe, M., Aseffa, A., and Tønjum, T. (2016) Deciphering the recent phylogenetic expansion of the originally deeply rooted Mycobacterium tuberculosis lineage 7. *BMC evolutionary biology* **16**, 1-10
149. Yimer, S. A., Norheim, G., Namouchi, A., Zegeye, E. D., Kinander, W., Tønjum, T., Bekele, S., Mannsåker, T., Bjune, G., and Aseffa, A. (2015) Mycobacterium tuberculosis lineage 7 strains are associated with prolonged patient delay in seeking treatment for pulmonary tuberculosis in Amhara Region, Ethiopia. *Journal of clinical microbiology* **53**, 1301-1309
150. Wilson, B. A., Garud, N. R., Feder, A. F., Assaf, Z. J., and Pennings, P. S. (2016) The population genetics of drug resistance evolution in natural populations of viral, bacterial and eukaryotic pathogens. *Molecular ecology* **25**, 42-66
151. Brown, T. S., Challagundla, L., Baugh, E. H., Omar, S. V., Mustaev, A., Auld, S. C., Shah, N. S., Kreiswirth, B. N., Brust, J., and Nelson, K. N. (2019) Pre-detection history of extensively drug-resistant tuberculosis in KwaZulu-Natal, South Africa. *Proceedings of the National Academy of Sciences* **116**, 23284-23291
152. da Silva, P. E. A., Von Groll, A., Martin, A., and Palomino, J. C. (2011) Efflux as a mechanism for drug resistance in Mycobacterium tuberculosis. *FEMS Immunology & Medical Microbiology* **63**, 1-9
153. Silhavy, T. J., Kahne, D., and Walker, S. (2010) The bacterial cell envelope. *Cold Spring Harbor perspectives in biology* **2**, a000414
154. Velayati, A. A., Farnia, P., Ibrahim, T. A., Haroun, R. Z., Kuan, H. O., Ghanavi, J., Farnia, P., Kabarei, A. N., Tabarsi, P., and Omar, A. R. (2009) Differences in cell wall thickness between resistant and nonresistant strains of Mycobacterium tuberculosis: using transmission electron microscopy. *Chemotherapy* **55**, 303-307
155. Pal, R., Hameed, S., Kumar, P., Singh, S., and Fatima, Z. (2017) Comparative lipidomics of drug sensitive and resistant Mycobacterium tuberculosis reveals altered lipid imprints. *3 Biotech* **7**, 1-10
156. Batt, S. M., Burke, C. E., Moorey, A. R., and Besra, G. S. (2020) Antibiotics and resistance: the two-sided coin of the mycobacterial cell wall. *The Cell Surface* **6**, 100044

© University of Pretoria

157. Ghazaei, C. (2018) Mycobacterium tuberculosis and lipids: Insights into molecular mechanisms from persistence to virulence. *Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences* **23**
158. Jackson, M. (2014) The mycobacterial cell envelope—lipids. *Cold Spring Harbor perspectives in medicine* **4**, a021105
159. Alderwick, L. J., Harrison, J., Lloyd, G. S., and Birch, H. L. (2015) The mycobacterial cell wall—peptidoglycan and arabinogalactan. *Cold Spring Harbor perspectives in medicine* **5**, a021113
160. Swain, S. S., Sharma, D., Hussain, T., and Pati, S. (2020) Molecular mechanisms of underlying genetic factors and associated mutations for drug resistance in Mycobacterium tuberculosis. *Emerging microbes & infections* **9**, 1651-1663
161. Ghajavand, H., Kargarpour Kamakoli, M., Khanipour, S., Pourazar Dizaji, S., Masoumi, M., Rahimi Jamnani, F., Fateh, A., Yaseri, M., Siadat, S. D., and Vaziri, F. (2019) Scrutinizing the drug resistance mechanism of multi-and extensively-drug resistant Mycobacterium tuberculosis: mutations versus efflux pumps. *Antimicrobial Resistance & Infection Control* **8**, 1-8
162. Kanji, A., Hasan, R., and Hasan, Z. (2019) Efflux pump as alternate mechanism for drug resistance in Mycobacterium tuberculosis. *Indian Journal of Tuberculosis* **66**, 20-25
163. Remm, S., Earp, J. C., Dick, T., Dartois, V., and Seeger, M. A. (2022) Critical discussion on drug efflux in Mycobacterium tuberculosis. *FEMS microbiology reviews* **46**, fuab050
164. Xu, G., Liu, H., Jia, X., Wang, X., and Xu, P. (2021) Mechanisms and detection methods of Mycobacterium tuberculosis rifampicin resistance: The phenomenon of drug resistance is complex. *Tuberculosis* **128**, 102083
165. Fisher, J. F., and Mobashery, S. (2016) β-Lactam resistance mechanisms: Gram-positive bacteria and Mycobacterium tuberculosis. *Cold Spring Harbor perspectives in medicine* **6**, a025221
166. Arun, K., Madhavan, A., Abraham, B., Balaji, M., Sivakumar, K., Nisha, P., and Kumar, R. A. (2020) Acetylation of isoniazid is a novel mechanism of isoniazid resistance in Mycobacterium tuberculosis. *Antimicrobial Agents and Chemotherapy* **65**, e00456-00420
167. Smith, T., Wolff, K. A., and Nguyen, L. (2012) Molecular biology of drug resistance in Mycobacterium tuberculosis. *Pathogenesis of Mycobacterium tuberculosis and its Interaction with the Host Organism*, 53-80
168. Hughes, D., and Andersson, D. I. (2017) Evolutionary trajectories to antibiotic resistance. *Annual Review of Microbiology* **71**, 579-596
169. Salverda, M. L., Dellus, E., Gorter, F. A., Debets, A. J., Van Der Oost, J., Hoekstra, R. F., Tawfik, D. S., and de Visser, J. A. G. (2011) Initial mutations direct alternative pathways of protein evolution. *PLoS genetics* **7**, e1001321
170. Kaiser, V. B., and Charlesworth, B. (2009) The effects of deleterious mutations on evolution in non-recombining genomes. *Trends in Genetics* **25**, 9-12
171. Trauner, A., Borrell, S., Reither, K., and Gagneux, S. (2014) Evolution of drug resistance in tuberculosis: recent progress and implications for diagnosis and therapy. *Drugs* **74**, 1063-1072
172. Li, Q.-j., Jiao, W.-w., Yin, Q.-q., Li, Y.-j., Li, J.-q., Xu, F., Sun, L., Xiao, J., Qi, H., and Wang, T. (2017) Positive epistasis of major low-cost drug resistance mutations rpoB531-TTG and katG315-ACC depends on the phylogenetic background of Mycobacterium tuberculosis strains. *International journal of antimicrobial agents* **49**, 757-762
173. Spies, F. S., von Groll, A., Ribeiro, A. W., Ramos, D. F., Ribeiro, M. O., Dalla Costa, E. R., Martin, A., Palomino, J. C., Rossetti, M. L., and Zaha, A. (2013) Biological cost in Mycobacterium tuberculosis with mutations in the rpsL, rrs, rpoB, and katG genes. *Tuberculosis* **93**, 150-154
174. Nghe, P., Kogenaru, M., and Tans, S. J. (2018) Sign epistasis caused by hierarchy within signalling cascades. *Nature communications* **9**, 1-9
175. Salvatore, P. P., Becerra, M. C., Abel zur Wiesch, P., Hinkley, T., Kaur, D., Sloutsky, A., and Cohen, T. (2016) Fitness costs of drug resistance mutations in multidrug-resistant

© University of Pretoria

Mycobacterium tuberculosis: a household-based case-control study. *The Journal of infectious diseases* **213**, 149-155

176. Cohen, K. A., Abeel, T., McGuire, A. M., Desjardins, C. A., Munsamy, V., Shea, T. P., Walker, B. J., Bantubani, N., Almeida, D. V., and Alvarado, L. (2015) Evolution of extensively drug-resistant tuberculosis over four decades: whole genome sequencing and dating analysis of Mycobacterium tuberculosis isolates from KwaZulu-Natal. *PLoS Med* **12**, e1001880
177. Zimic, M., Vargas, A., Kirwan, D., Rios, A., Gilman, R., Sheen, P., and Grandjean, L. (2020) Determination of potentially novel compensatory mutations in rpoc associated with rifampin resistance and rpob mutations in Mycobacterium tuberculosis Clinical isolates from peru.
178. Hughes, D., and Andersson, D. I. (2015) Evolutionary consequences of drug resistance: shared principles across diverse targets and organisms. *Nature Reviews Genetics* **16**, 459-471
179. Maddamsetti, R., Lenski, R. E., and Barrick, J. E. (2015) Adaptation, clonal interference, and frequency-dependent interactions in a long-term evolution experiment with Escherichia coli. *Genetics* **200**, 619-631
180. Gerrish, P. J., and Lenski, R. E. (1998) The fate of competing beneficial mutations in an asexual population. *Genetica* **102**, 127-144
181. Sun, G., Luo, T., Yang, C., Dong, X., Li, J., Zhu, Y., Zheng, H., Tian, W., Wang, S., and Barry III, C. E. (2012) Dynamic population changes in Mycobacterium tuberculosis during acquisition and fixation of drug resistance in patients. *The Journal of infectious diseases* **206**, 1724-1733
182. Eldholm, V., Norheim, G., von der Lippe, B., Kinander, W., Dahle, U. R., Caugant, D. A., Mannsåker, T., Mengshoel, A. T., Dyrhol-Riise, A. M., and Balloux, F. (2014) Evolution of extensively drug-resistant Mycobacterium tuberculosis from a susceptible ancestor in a single patient. *Genome biology* **15**, 1-11
183. Castro, R. A., Ross, A., Kamwela, L., Reinhard, M., Loiseau, C., Feldmann, J., Borrell, S., Trauner, A., and Gagneux, S. (2020) The genetic background modulates the evolution of fluoroquinolone-resistance in Mycobacterium tuberculosis. *Molecular biology and evolution* **37**, 195-207
184. Bergval, I., Kwok, B., Schuitema, A., Kremer, K., Van Soolingen, D., Klatser, P., and Anthony, R. (2012) Pre-existing isoniazid resistance, but not the genotype of Mycobacterium tuberculosis drives rifampicin resistance codon preference in vitro. *PLoS One* **7**, e29108
185. Van Niekerk, K., Pierneef, R., Reva, O. N., Korostetskiy, I. S., Ilin, A. I., and Akhmetova, G. K. (2018) Clade-specific distribution of antibiotic resistance mutations in the population of Mycobacterium tuberculosis. Prospects for drug resistance reversion. in *Basic Biology and Applications of Actinobacteria*, IntechOpen London, UK. pp 79-98
186. Consortium, C., and the 100, G. P. (2018) Prediction of susceptibility to first-line tuberculosis drugs by DNA sequencing. *New England Journal of Medicine* **379**, 1403-1415
187. Gillespie, J. J., Wattam, A. R., Cammer, S. A., Gabbard, J. L., Shukla, M. P., Dalay, O., Driscoll, T., Hix, D., Mane, S. P., and Mao, C. (2011) PATRIC: the comprehensive bacterial bioinformatics resource with a focus on human pathogenic species. *Infection and immunity* **79**, 4286-4298
188. Manson, A. L., Cohen, K. A., Abeel, T., Desjardins, C. A., Armstrong, D. T., Barry III, C. E., Brand, J., Jureen, P., Malinga, L., and Nordenberg, D. (2017) Genomic analysis of globally diverse Mycobacterium tuberculosis strains provides insights into the emergence and spread of multidrug resistance. *Nature genetics* **49**, 395
189. Chernyaeva, E. N., Shulgina, M. V., Rotkevich, M. S., Dobrynin, P. V., Simonov, S. A., Shitikov, E. A., Ischenko, D. S., Karpova, I. Y., Kostryukova, E. S., and Ilina, E. N. (2014) Genome-wide Mycobacterium tuberculosis variation (GMTV) database: a new tool for integrating sequence variations and epidemiology. *BMC genomics* **15**, 308
190. Organization, W. H. (2021) Catalogue of mutations in Mycobacterium tuberculosis complex and their association with drug resistance.

© University of Pretoria

191. Walter, K. S., Colijn, C., Cohen, T., Mathema, B., Liu, Q., Bowers, J., Engelthaler, D. M., Narechania, A., Lemmer, D., and Croda, J. (2020) Genomic variant-identification methods may alter Mycobacterium tuberculosis transmission inferences. *Microbial Genomics* **6**
192. Olson, N. D., Lund, S. P., Colman, R. E., Foster, J. T., Sahl, J. W., Schupp, J. M., Keim, P., Morrow, J. B., Salit, M. L., and Zook, J. M. (2015) Best practices for evaluating single nucleotide variant calling methods for microbial genomics. *Frontiers in genetics* **6**, 235
193. Langmead, B., and Salzberg, S. L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359
194. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, G. P. D. P. (2009) The sequence alignment/map format and SAMtools. *bioinformatics* **25**, 2078-2079
195. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J., and Prins, P. (2015) Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032-2034
196. McPherson, K. (2005) 1. Statistical methods for rates and proportions (3rd edn). Joseph L. Fleiss, Bruce Levin and Myunghee Cho Paik, Wiley, New Jersey, 2003. No. of pages: xxvii+ 760. Price: $99.95 (hardcover). ISBN: 0-471-52629-0. *Statistics in Medicine* **24**, 2744-2745
197. Qin, L., Wang, J., Lu, J., Yang, H., Zheng, R., Liu, Z., Huang, X., Feng, Y., Hu, Z., and Ge, B. (2019) A deletion in the RD105 region confers resistance to multiple drugs in Mycobacterium tuberculosis. *BMC biology* **17**, 1-12
198. Singh, P., Jamal, S., Ahmed, F., Saqib, N., Mehra, S., Ali, W., Roy, D., Ehtesham, N. Z., and Hasnain, S. E. (2021) Computational modeling and bioinformatic analyses of functional mutations in drug target genes in Mycobacterium tuberculosis. *Computational and Structural Biotechnology Journal* **19**, 2423-2446
199. Seifert, M., Catanzaro, D., Catanzaro, A., and Rodwell, T. C. (2015) Genetic mutations associated with isoniazid resistance in Mycobacterium tuberculosis: a systematic review. *PloS one* **10**, e0119628
200. Miotto, P., Tessema, B., Tagliani, E., Chindelevitch, L., Starks, A. M., Emerson, C., Hanna, D., Kim, P. S., Liwski, R., and Zignol, M. (2017) A standardised method for interpreting the association between mutations and phenotypic drug resistance in Mycobacterium tuberculosis. *European Respiratory Journal* **50**
201. Medley, J., Goff, A., Bettencourt, P. J., Dare, M., Cole, L., Cantillon, D., and Waddell, S. J. (2022) Dissecting the mycobacterium bovis BCG response to macrophage infection to help prioritize targets for anti-tuberculosis drug and vaccine discovery. *Vaccines* **10**, 113
202. Ye, X., Peng, T., Li, Y., Huang, T., Wang, H., and Hu, Z. (2022) Identification of an important function of CYP123: Role in the monooxygenase activity in a novel estradiol degradation pathway in bacteria. *The Journal of Steroid Biochemistry and Molecular Biology* **215**, 106025
203. Lagutkin, D., Panova, A., Vinokurov, A., Gracheva, A., Samoilova, A., and Vasilyeva, I. (2022) Genome-Wide study of drug resistant Mycobacterium tuberculosis and its intra-host evolution during treatment. *Microorganisms* **10**, 1440
204. Chen, X., Chen, J., Yan, B., Zhang, W., Guddat, L. W., Liu, X., and Rao, Z. (2020) Structural basis for the broad substrate specificity of two acyl-CoA dehydrogenases FadE5 from mycobacteria. *Proceedings of the National Academy of Sciences* **117**, 16324-16332
205. McNeil, M. B., O'Malley, T., Dennison, D., Shelton, C. D., Sunde, B., and Parish, T. (2020) Multiple mutations in Mycobacterium tuberculosis MmpL3 increase resistance to MmpL3 inhibitors. *MSphere* **5**, 10.1128/msphere. 00985-00920
206. Guo, J., Zheng, X., Xu, L., Liu, Z., Xu, K., Li, S., Wen, T., Liu, S., and Pang, H. (2010) Characterization of a novel esterase Rv0045c from Mycobacterium tuberculosis. *PLoS One* **5**, e13143
207. Sherlin, D., and Anishetty, S. (2015) Mechanistic insights from molecular dynamic simulation of Rv0045c esterase in Mycobacterium tuberculosis. *Journal of Molecular Modeling* **21**, 1-8

© University of Pretoria

208. Torres Ortiz, A., Coronel, J., Vidal, J. R., Bonilla, C., Moore, D. A., Gilman, R. H., Balloux, F., Kon, O. M., Didelot, X., and Grandjean, L. (2021) Genomic signatures of pre-resistance in Mycobacterium tuberculosis. *Nature communications* **12**, 7312
209. Raman, K., and Chandra, N. (2008) Mycobacterium tuberculosis interactome analysis unravels potential pathways to drug resistance. *BMC microbiology* **8**, 1-13
210. Jiang, X., Zhang, W., Gao, F., Huang, Y., Lv, C., and Wang, H. (2006) Comparison of the proteome of isoniazid-resistant and-susceptible strains of Mycobacterium tuberculosis. *Microbial drug resistance* **12**, 231-238
211. Jhingan, G. D., Kumari, S., Jamwal, S. V., Kalam, H., Arora, D., Jain, N., Kumaar, L. K., Samal, A., Rao, K. V., and Kumar, D. (2016) Comparative proteomic analyses of avirulent, virulent, and clinical strains of Mycobacterium tuberculosis identify strain-specific patterns. *Journal of Biological Chemistry* **291**, 14257-14273
212. Diamond, S., Rubin, B. E., Shultzaberger, R. K., Chen, Y., Barber, C. D., and Golden, S. S. (2017) Redox crisis underlies conditional light–dark lethality in cyanobacterial mutants that lack the circadian regulator, RpaA. *Proceedings of the National Academy of Sciences* **114**, E580-E589
213. Furió, V., Moreno-Molina, M., Chiner-Oms, Á., Villamayor, L. M., Torres-Puente, M., and Comas, I. (2021) An evolutionary functional genomics approach identifies novel candidate regions involved in isoniazid resistance in Mycobacterium tuberculosis. *Communications Biology* **4**, 1322
214. Wan, L., Hu, P., Zhang, L., Wang, Z.-X., Fleming, J., Ni, B., Luo, J., Guan, C.-X., Bai, L., and Tan, Y. (2023) Omics analysis of Mycobacterium tuberculosis isolates uncovers Rv3094c, an ethionamide metabolism-associated gene. *Communications Biology* **6**, 156
215. Tudó, G., Laing, K., Mitchison, D. A., Butcher, P. D., and Waddell, S. J. (2010) Examining the basis of isoniazid tolerance in nonreplicating Mycobacterium tuberculosis using transcriptional profiling. *Future medicinal chemistry* **2**, 1371-1383
216. Shekar, S., Yeo, Z. X., Wong, J. C., Chan, M. K., Ong, D. C., Tongyoo, P., Wong, S.-Y., and Lee, A. S. (2014) Detecting novel genetic variants associated with isoniazid-resistant Mycobacterium tuberculosis. *PloS one* **9**, e102383
217. Peng, Z., Chen, L., and Zhang, H. (2020) Serum proteomic analysis of Mycobacterium tuberculosis antigens for discriminating active tuberculosis from latent infection. *Journal of International Medical Research* **48**, 0300060520910042
218. Ates, L. S. (2020) New insights into the mycobacterial PE and PPE proteins provide a framework for future research. *Molecular microbiology* **113**, 4-21
219. Sharma, S., and Sharma, M. (2021) Proline-Glutamate/Proline-Proline-Glutamate (PE/PPE) proteins of Mycobacterium tuberculosis: the multifaceted immune-modulators. *Acta Tropica* **222**, 106035
220. Flores, J., and Espitia, C. (2003) Differential expression of PE and PE_PGRS genes in Mycobacterium tuberculosis strains. *Gene* **318**, 75-81
221. Mukhopadhyay, S., and Balaji, K. N. (2011) The PE and PPE proteins of Mycobacterium tuberculosis. *Tuberculosis* **91**, 441-447
222. Houben, E. N., Korotkov, K. V., and Bitter, W. (2014) Take five—Type VII secretion systems of Mycobacteria. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1843**, 1707-1716
223. Wang, L., Asare, E., Shetty, A. C., Sanchez-Tumbaco, F., Edwards, M. R., Saranathan, R., Weinrick, B., Xu, J., Chen, B., and Bénard, A. (2022) Multiple genetic paths including massive gene amplification allow Mycobacterium tuberculosis to overcome loss of ESX-3 secretion system substrates. *Proceedings of the National Academy of Sciences* **119**, e2112608119
224. Chen, J. M., Boy-Röttger, S., Dhar, N., Sweeney, N., Buxton, R. S., Pojer, F., Rosenkrands, I., and Cole, S. T. (2012) EspD is critical for the virulence-mediating ESX-1 secretion system in Mycobacterium tuberculosis. *Journal of bacteriology* **194**, 884-893

© University of Pretoria

225. Tiwari, S., Casey, R., Goulding, C. W., Hingley-Wilson, S., and Jacobs Jr, W. R. (2019) Infect and Inject: How Mycobacterium tuberculosis Exploits Its Major Virulence-Associated Type VII Secretion System, ESX-1. *Bacteria and Intracellularity*, 113-126
226. Bunduc, C. M., Bitter, W., and Houben, E. N. (2020) Structure and function of the mycobacterial type VII secretion systems. *Annual review of microbiology* **74**, 315-335
227. Gómez-González, P. J., Grabowska, A. D., Tientcheu, L. D., Tsolaki, A. G., Hibberd, M. L., Campino, S., Phelan, J. E., and Clark, T. G. (2023) Functional genetic variation in pe/ppe genes contributes to diversity in Mycobacterium tuberculosis lineages and potential interactions with the human host. *Frontiers in Microbiology* **14**
228. Safi, H., Lingaraju, S., Amin, A., Kim, S., Jones, M., Holmes, M., McNeil, M., Peterson, S. N., Chatterjee, D., and Fleischmann, R. (2013) Evolution of high-level ethambutol-resistant tuberculosis through interacting mutations in decaprenylphosphoryl-β-D-arabinose biosynthetic and utilization pathway genes. *Nature genetics* **45**, 1190-1197
229. Martini, M. C., Hicks, N. D., Xiao, J., Alonso, M. N., Barbier, T., Sixsmith, J., Fortune, S. M., and Shell, S. S. (2022) Loss of RNase J leads to multi-drug tolerance and accumulation of highly structured mRNA fragments in Mycobacterium tuberculosis. *PLoS pathogens* **18**, e1010705
230. Edwards, E. J. (2019) Evolutionary trajectories, accessibility and other metaphors: the case of C4 and CAM photosynthesis. *New Phytologist* **223**, 1742-1755
231. Baquero, F., Martinez, J. L., F. Lanza, V., Rodríguez-Beltrán, J., Galán, J. C., San Millán, A., Cantón, R., and Coque, T. M. (2021) Evolutionary pathways and trajectories in antibiotic resistance. *Clinical Microbiology Reviews* **34**, e00050-00019
232. Rindi, L., Lari, N., Cuccu, B., and Garzelli, C. (2009) Evolutionary pathway of the Beijing lineage of Mycobacterium tuberculosis based on genomic deletions and mutT genes polymorphisms. *Infection, Genetics and Evolution* **9**, 48-53
233. Lin, D., Wang, J., Cui, Z., Ou, J., Huang, L., and Wang, Y. (2021) A genome epidemiological study of mycobacterium tuberculosis in subpopulations with high and low incidence rate in Guangxi, South China. *BMC Infectious Diseases* **21**, 1-11
234. Pajuelo, D., Tak, U., Zhang, L., Danilchanka, O., Tischler, A. D., and Niederweis, M. (2021) Toxin secretion and trafficking by Mycobacterium tuberculosis. *Nature communications* **12**, 6592
235. Tak, U., Dokland, T., and Niederweis, M. (2021) Pore-forming Esx proteins mediate toxin secretion by Mycobacterium tuberculosis. *Nature Communications* **12**, 394
236. Das, C., Ghosh, T. S., and Mande, S. S. (2011) Computational analysis of the ESX-1 region of Mycobacterium tuberculosis: insights into the mechanism of type VII secretion system. *PLoS One* **6**, e27980
237. Gupta, A. K., Katoch, V. M., Chauhan, D. S., Sharma, R., Singh, M., Venkatesan, K., and Sharma, V. D. (2010) Microarray analysis of efflux pump genes in multidrug-resistant Mycobacterium tuberculosis during stress induced by common anti-tuberculous drugs. *Microbial drug resistance* **16**, 21-28
238. Lee, J., Kang, H., Lee, W., Cho, S., Kim, Y., and Lee, H. (2021) Efflux pump gene expression study using RNA-seq in multidrug-resistant TB. *The International Journal of Tuberculosis and Lung Disease* **25**, 974-981
239. Hicks, N. D., Giffen, S. R., Culviner, P. H., Chao, M. C., Dulberger, C. L., Liu, Q., Stanley, S., Brown, J., Sixsmith, J., and Wolf, I. D. (2020) Mutations in dnaA and a cryptic interaction site increase drug resistance in Mycobacterium tuberculosis. *PLoS pathogens* **16**, e1009063
240. Xu, J., Li, D., Shi, J., Wang, B., Ge, F., Guo, Z., Mu, X., Nuermberger, E., and Lu, Y. (2023) Bedquiline Resistance Mutations: Correlations with Drug Exposures and Impact on the Proteome in M. tuberculosis. *Antimicrobial Agents and Chemotherapy*, e01532-01522
241. Ramón-García, S., Martín, C., De Rossi, E., and Aínsa, J. A. (2007) Contribution of the Rv2333c efflux pump (the Stp protein) from Mycobacterium tuberculosis to intrinsic antibiotic resistance in Mycobacterium bovis BCG. *Journal of antimicrobial chemotherapy* **59**, 544-547

© University of Pretoria

242. Li, K.-K., Qu, D.-H., Zhang, H.-N., Chen, F.-Y., Xu, L., Wang, M.-Y., Su, H.-Y., Tao, S.-C., and Wu, F.-L. (2020) Global discovery the PstP interactions using Mtb proteome microarray and revealing novel connections with EthR. *Journal of proteomics* **215**, 103650
243. Mikheecheva, N. E., Zaychikova, M. V., Melerzanov, A. V., and Danilenko, V. N. (2017) A nonsynonymous SNP catalog of Mycobacterium tuberculosis virulence genes and its use for detecting new potentially virulent sublineages. *Genome biology and evolution* **9**, 887-899
244. Zhou, P., Wang, X., Zhao, Y., Yuan, W., and Xie, J. (2018) Sigma factors mediated signaling in Mycobacterium tuberculosis. *Future Microbiology* **13**, 231-240
245. Manganelli, R. (2014) Sigma factors: key molecules in Mycobacterium tuberculosis physiology and virulence. *Molecular genetics of mycobacteria*, 135-160
246. Farr, D. C., Haselhorst, T., Tan, L., Furness, J., Strong, E., Grice, I. D., West, N. P., and Houston, T. A. (2024) Reassessing the putative molecular Target (s) of potent antitubercular 2-(Alkylsulfonyl) acetamides. *European Journal of Medicinal Chemistry* **264**, 115983
247. Ganguli, G., Mukherjee, U., and Sonawane, A. (2019) Peroxisomes and oxidative stress: their implications in the modulation of cellular immunity during mycobacterial infection. *Frontiers in microbiology* **10**, 1121
248. Guo, Q., Bi, J., Lin, Q., Ye, T., Wang, Z., Wang, Z., Liu, L., and Zhang, G. (2022) Whole genome sequencing identifies novel mutations associated with bedaquiline resistance in mycobacterium tuberculosis. *Frontiers in cellular and infection microbiology* **12**, 807095
249. De Vos, M., Müller, B., Borrell, S., Black, P., Van Helden, P., Warren, R., Gagneux, S., and Victor, T. (2013) Putative compensatory mutations in the rpoC gene of rifampin-resistant Mycobacterium tuberculosis are associated with ongoing transmission. *Antimicrobial agents and chemotherapy* **57**, 827-832
250. Vargas, A. P., Rios, A. A., Grandjean, L., Kirwan, D. E., Gilman, R. H., Sheen, P., and Zimic, M. J. (2020) Determination of potentially novel compensatory mutations in rpoc associated with rifampin resistance and rpob mutations in Mycobacterium tuberculosis Clinical isolates from peru. *The International Journal of Mycobacteriology* **9**, 121-137
251. Solomonson, M., Huesgen, P. F., Wasney, G. A., Watanabe, N., Gruninger, R. J., Prehna, G., Overall, C. M., and Strynadka, N. C. (2013) Structure of the mycosin-1 protease from the mycobacterial ESX-1 protein type VII secretion system. *Journal of Biological Chemistry* **288**, 17782-17790
252. Peters, J. S., Ismail, N., Dippenaar, A., Ma, S., Sherman, D. R., Warren, R. M., and Kana, B. D. (2020) Genetic diversity in Mycobacterium tuberculosis clinical isolates and resulting outcomes of tuberculosis infection and disease. *Annual Review of Genetics* **54**, 511-537
253. Ghebremichael, S., Groenheit, R., Pennhag, A., Koivula, T., Andersson, E., Bruchfeld, J., Hoffner, S., Romanus, V., and Källenius, G. (2010) Drug resistant Mycobacterium tuberculosis of the Beijing genotype does not spread in Sweden. *PLoS One* **5**, e10893
254. de Souza, G. A., Fortuin, S., Aguilar, D., Pando, R. H., McEvoy, C. R., van Helden, P. D., Koehler, C. J., Thiede, B., Warren, R. M., and Wiker, H. G. (2010) Using a label-free proteomics method to identify differentially abundant proteins in closely related hypo-and hypervirulent clinical Mycobacterium tuberculosis Beijing isolates. *Molecular & Cellular Proteomics* **9**, 2414-2423
255. Hanekom, M., Van Pittius, N. G., McEvoy, C., Victor, T., Van Helden, P., and Warren, R. (2011) Mycobacterium tuberculosis Beijing genotype: a template for success. *Tuberculosis* **91**, 510-523
256. Wozniak, M., Tiuryn, J., and Wong, L. (2014) GWAMAR: genome-wide assessment of mutations associated with drug resistance in bacteria. *BMC genomics* **15**, 1-11

© University of Pretoria

Appendix

module load bowtie2-2.3.4.1

module load bcftools-1.7

module load samtools-1.10

module load sambamba

##bowtie build index

bowtie2-build GCF_000195955.2_ASM19595v2_genomic.fna Mtb_k12

mapping

bowtie2 –x Mtb_k12 -1 SRR6397327_1.fastq.gz -2 SRR6397327_2.fastq.gz | samtools view –b –o SRR6397327.bam

sort bam file

samtools sort –o SRR6397327.sort.bam SRR6397327.bam

index bam files

samtools index SRR6397327.sort.bam

remove duplicates

sambamba markdup –r SRR6397327.sort.bam SRR6397327.rdm.bam

variant calling added additional annotation for filtering

bcftools mpileup -Ou –f GCF_000195955.2_ASM19595v2_genomic.fna -a INFO/ADF,INFO/ADR,FORMAT/AD,DP,AF,SP SRR6397327.rdm.bam| bcftools call –ploidy 1 -mv -Ob -o SRR6397327.bcf

normalise

© University of Pretoria

bcftools norm –Ou –f GCF_000195955.2_ASM19595v2_genomic.fna –O z -o SRR6397327.norm.bcf

sort

bcftools convert -O v -o SRR6397327.vcf SRR6397327.bcf

tabix –f –p vcf SRR6397327.vcf

bcftools sort SRR6397327.vcf –O z –o SRR6397327.sort.vcf

tabix –f –p vcf SRR6397327.sort.vcf

bcftools filter -g3 -G10 -e '(TYPE="snp" && QUAL <= 150 && FORMAT/DP <= 20 && QUAL/(DP4[2]+DP4[3]) <= 1 && DP4[2]+DP4[3] <=5 && AF <= 0.95 && SP > 30) SRR6397327.sort.vcf -O z -o SRR6397327.sort.vcf.filtered_2.vcf.g

Appendix 1: The bash scripts used to do variant filtering of polymorphism in sequence files collected for analysis.

© University of Pretoria

Samplesn	Source	Country	Lineage	SRA/ENAa	INH	RIF	EMB	PZA
0163D	Casaliet.a	Russia	Lineage_2	ERR06757(	R	R	R	S
0166G	Casaliet.a	Russia	Lineage_4(Ural)	ERR06757	R	R	S	S
0175R	Casaliet.a	Russia	Lineage_2	ERR06757	R	R	S	R
0177T	Casaliet.a	Russia	Lineage_2	ERR06758(	S	S	S	S
0180X	Casaliet.a	Russia	Lineage_2	ERR06758	R	R	R	R
0182Z	Casaliet.a	Russia	Lineage_2	ERR06758	R	S	R	S
0217M	Casaliet.a	Russia	Lineage_2	ERR06758	R	R	R	S
0218N	Casaliet.a	Russia	Lineage_2	ERR06758	R	R	R	S
0219P	Casaliet.a	Russia	Lineage_4(LAM)	ERR06758	S	S	S	S
0233E	Casaliet.a	Russia	Lineage_2	ERR06758(	R	S	R	R
0345B	Casaliet.a	Russia	Lineage_2	ERR06758	R	R	S	S
0402N	Casaliet.a	Russia	Lineage_2	ERR06758	R	R	S	S
0403P	Casaliet.a	Russia	Lineage_2	ERR06758	S	S	S	S
0404Q	Casaliet.a	Russia	Lineage_2	ERR06759(	R	R	R	R
0459A	Casaliet.a	Russia	Lineage_4(LAM)	ERR06759	R	R	R	S
0514K	Casaliet.a	Russia	Lineage_2	ERR06759	R	R	S	R
0611Q	Casaliet.a	Russia	Lineage_2	ERR06759	S	S	S	S
0635R	Casaliet.a	Russia	Lineage_4(Ural)	ERR06759(	R	S	S	S
0640X	Casaliet.a	Russia	Lineage_4(LAM)	ERR06759	R	R	R	S
0032L	Casaliet.a	Russia	Lineage_2	ERR06760(	R	S	S	S
0101L	Casaliet.a	Russia	Lineage_4(LAM)	ERR06760	S	S	S	S
0113Z	Casaliet.a	Russia	Lineage_4	ERR06760	R	S	S	R
0309M	Casaliet.a	Russia	Lineage_4(Haarlem)	ERR06760	R	R	R	R
0584L	Casaliet.a	Russia	Lineage_2	ERR06760	S	S	S	R
0336R	Casaliet.a	Russia	Lineage_4(Ural)	ERR06760(	S	S	S	S
0096F	Casaliet.a	Russia	Lineage_2	ERR06760	R	R	S	S
0314S	Casaliet.a	Russia	Lineage_4(Ural)	ERR06760	R	R	S	S
0317W	Casaliet.a	Russia	Lineage_2	ERR06761(	R	S	S	S
0344A	Casaliet.a	Russia	Lineage_4(LAM)	ERR06761	R	S	S	R
0628J	Casaliet.a	Russia	Lineage_2	ERR06761	R	R	S	S
0028G	Casaliet.a	Russia	Lineage_4	ERR06761	S	S	S	S
0029H	Casaliet.a	Russia	Lineage_4(Ural)	ERR06761	R	R	S	R
0030J	Casaliet.a	Russia	Lineage_4(Haarlem)	ERR06761(	S	S	S	S
0043Y	Casaliet.a	Russia	Lineage_4(Haarlem)	ERR06761	S	S	S	S
0044Z	Casaliet.a	Russia	Lineage_2	ERR06761	R	R	S	R
0050F	Casaliet.a	Russia	Lineage_2	ERR06761	R	R	R	S
0053J	Casaliet.a	Russia	Lineage_2	ERR06762(	R	R	R	S
0061S	Casaliet.a	Russia	Lineage_2	ERR06762	R	R	R	S

Appendix 2: Metadata of *Mycobacterium tuberculosis* strains for each sequence file.

© University of Pretoria

'id	Position	Power	Alleles	Allelic_fre	Locus	Gene	Codon	Annotatio	Moltype
1	1	0.78	L|M,G,R	0.11|0.88	Rv0001	dnaA	1	Chromosc	T->A,T->G,
2	2	0.26	M|*,R,G	0.99|0.0	Rv0001	dnaA	1	Chromosc	T->A,T->C,
3	3	0.31	M|*,S,L	1.0|0.0	Rv0001	dnaA	1	Chromosc	G->C,G->G
4	4	0.33	M|*,W,C	1.0|0.0	Rv0001	dnaA	1	Chromosc	A->AAGAT
5	5	0.31	T|H,R,*,S,	1.0|0.0	Rv0001	dnaA	2	Chromosc	C->AGATA
6	6	0.0	T|P,R	1.0|0.0	Rv0001	dnaA	2	Chromosc	C->CAAGA
7	71	0.0	P|L	0.99|0.0	Rv0001	dnaA	24	Chromosc	C->T
8	237	0.01	R|E	0.99|0.0	Rv0001	dnaA	79	Chromosc	G->A
9	371	0.03	P|*	0.98|0.01	Rv0001	dnaA	124	Chromosc	C->T
10	645	0.0	T|P	1.0|0.0	Rv0001	dnaA	215	Chromosc	A->C,Not f
11	698	0.0	R|	1.0|0.0	Rv0001	dnaA	233	Chromosc	Not found
12	1126	0.0	A|R	0.99|0.0	Rv0001	dnaA	375	Chromosc	C->T
13	1131	0.02	I|Y	0.99|0.0	Rv0001	dnaA	377	Chromosc	C->A
14	1186	0.0	A|P	0.99|0.0	Rv0001	dnaA	395	Chromosc	G->A
15	1199	0.0	R|T	0.99|0.0	Rv0001	dnaA	400	Chromosc	G->A
16	1302	0.03	P|Q,L	0.99|0.0	Rv0001	dnaA	434	Chromosc	C->A,C->T
17	1326	0.01	Q|I,T	0.99|0.0	Rv0001	dnaA	442	Chromosc	G->C,G->T
18	1519	0.01	K|S	0.99|0.0	Rv0001	dnaA	506	Chromosc	C->T
19	2222	0.02	S|L	0.99|0.0	Rv0002	dnaN	57	DNA polyr	C->T
20	2357	0.0	V|L	0.99|0.0	Rv0002	dnaN	102	DNA polyr	C->T
21	2380	0.01	T|S	0.99|0.0	Rv0002	dnaN	110	DNA polyr	C->T
22	2422	0.0	E|R	0.99|0.0	Rv0002	dnaN	124	DNA polyr	A->C
23	2532	0.3	T|R	0.92|0.07	Rv0002	dnaN	160	DNA polyr	T->C
24	2539	0.0	M|R	0.99|0.0	Rv0002	dnaN	163	DNA polyr	T->C
25	2586	0.0	L|*,C	0.99|0.0	Rv0002	dnaN	178	DNA polyr	G->A,G->T
26	2745	0.0	G|	1.0|0.0	Rv0002	dnaN	231	DNA polyr	Not found
27	3143	0.01	L|Y,S	0.99|0.0	Rv0002	dnaN	364	DNA polyr	G->A,G->C
28	3170	0.0	R|D	0.99|0.0	Rv0002	dnaN	373	DNA polyr	C->A
29	3186	0.02	L|W	0.99|0.0	Rv0002	dnaN	378	DNA polyr	A->G
30	3192	0.0	G|A	0.99|0.0	Rv0002	dnaN	380	DNA polyr	A->G
31	3390	0.0	T|L,P	0.99|0.0	Rv0003	recF	37	DNA repli	G->C,G->T
32	3439	0.0	R|A	0.99|0.0	Rv0003	recF	53	DNA repli	G->A
33	3446	0.55	A|S	0.74|0.25	Rv0003	recF	56	DNA repli	C->T
34	3452	0.0	L|R	0.99|0.0	Rv0003	recF	58	DNA repli	T->C
35	3472	0.01	G|V	0.99|0.0	Rv0003	recF	64	DNA repli	A->G
36	3522	0.0	C|A	0.99|0.0	Rv0003	recF	81	DNA repli	T->C
37	3681	0.0	R|V	0.99|0.0	Rv0003	recF	134	DNA repli	C->T
38	3812	0.02	G|L	0.99|0.0	Rv0003	recF	178	DNA repli	G->C
39	4013	0.16	P|I	0.01|0.98	Rv0003	recF	245	DNA repli	T->C
40	4080	0.01	E|I,N	0.99|0.0	Rv0003	recF	267	DNA repli	G->A,G->T

Appendix 3: Annotated polymorphisms identified by the split program in 9388 sequence files of *Mycobacterium tuberculosis* strains.

© University of Pretoria

>H37Rv
MMMMTTPRPTRAIARPQKSVTETMLGLRLGTRALGCRGIEGADGVHGPIMASGTDGNNEH
KDSSPIEHAAYDSTNLTRVLMDAKAAILFRYIILGLPTYADLPRSRSTSAPGGAAITSRA
LPLVAGKTDVLYRIGDLAATRPSAANGYYGFLATTIHADLVVVAVSIPPQRGQRVVDTSS
TQQNGAAPTTVSVPGPAAPISLRTGVDPSGARKRGAENDARTDSLGRTLYPSYQPDPPGR
RQEGRRQEGGGQGIIDQDGERPVDGSGAEVLDIIELQWGGAPGAQAGLAIYVRGG*NEED
HSWAVLRSEAAFFQRAYEAAGSGHGDDPFTTGASSAALTRRADRAAAARAMMMVGLAFAK
TRLVFEWLLGVHHQQARRCLRRASVAPEFIYGTACAGSVTADTAVVVVSAPFVLDAGAGG
VASTLIRLLPLLYARMQLGPMVAAAELVPALIERRGAARAPTMWAAHSYVMLYLGVAAVS
TCYPAPFHAVTPAVVRNFEGTISPGDEFADGEEFAPTGLYRRRAEVRLDIGAWRGDEREG
SDPDHPLLLVVRVGHHGGGGIERSYGIHTALGYLVESLERSESSIALAYDDQAEALTSNA
TGVVVPPPPPPPPPPPPPPPPLQAASVLALTLGPRTYLDHVTRGRLSVMPTGATDVRYEM
VPDPNDRNRVAPTYPEIPNGVARPRLGYPPGWGDGQLPLRRLAWEGGGETEVLLGDRFCP
GSPSVGRVVMGWHPFLRKIMLDELHAGAPARDAKFERQPGVVYNFGNPSRPPAPPPRPPV
PPAAPLEELDSGAKAEVNGLKKFAAELEKVWDTHDMVELDEGDDPVGISEIEGTVQIEPV
MGNATVVVDDDPPVVVDDDPPPVVVDDDPPPVVVDDDPPPVVDDDAVVVAAA
GVLLIVWWIIELVPNAYLPNAEGEGGVGCHVVLLPVTPYSATIGPN*FYGVSFRRAVTAA
AACDRAANAVEARIDEEFAGRRDTAHPRSAVHVAVGRTKVHDRTRVDPEVASRIPGVHFA
AAAAAPLAGTAIAPDTTAVHMALLVILIDILRELFLDWGYFLALAALAHVDSTWPVAAAP
LDAYVRLFGVRRIDTDTINVGAMGAGLMVMLDLIIIRNALGGVATGRGCGQKDLLLFREE
PRRPDDHHHHNHTGLEGAPDCGRNALLLTIARLRSKR*ESRRGEILLTLTQQQAAAPSLS
DDADFSFGHGCHVTKYALPNPGDAVIILRAVLPATFVGTSGTKFPPDRLNGRDRHLRVTS
PRMPAELGPPAVGFPARVAVTYRKLLLAMGGSTAHPSFDPAVALDVVALVIIAPAAVNGL
TAISSRTGEPPLPGVGALLHRGGHTYHEYLLSSDLGALSAVGAARALTAAPSLASAALDA
SNAEQALSVAMRANPVSSGDSNIGGGGGGAGGGGGGGVGGGGGGGAGLADRALPTAGDGP
MSDINGSRARDLKDIQWAAAADNVYAFTIGQRPSVGAAGVFGSEVGHGVAATGVVVTIAR
RSYLGASQSGAIVRANGASAPGGDLRAWPIFPLSDLTATEALLFNARGGKTVQLDDTELG
WAVAAETRDARPSRAPAGGGNGAGGQAGGGAGAAGGGGGGGGGGDGAAAIGLGGGGLDG
VVGAASSGGGVGGGGGHGAGAGAGGGGANGGDGGAAGGTDRRYGSIATTADDSATVTNGL
GPSNRGPAWRTAVGTGGTEEKAEFAERSFMAMTVAADLATGGQGMTNGGDRS*WTGGKWG
DGVTRSAYPADPPAMGKLDPGRCMTVRPDKGDISLRRARGAVTTIRGIASITDIADLLQP
VEQPTVMLSQLWITGAFGGSYGFDDAARLFRRARLSRDLLGGALAGGDQQMAYPADKLAK
APTGIMGSAALTADAAAIQGPLGRYGIGEPNAAAVWLLQLGIGRDRTQETPSARAILPEA
GSPLGGPGGANRSLALEKIRIMPAAAMSVASYYGTTGLQAFDKSRHSGDGRLALVLVPYL
ASRGEDAMMGGLRAMFVVINDCNTFIHNLTVTGIIADSGLNSENDEEFKRAPGAYVGVAI
AQVGIAYEDVDAGYPTAPPIPPVAEVIADATLPVPIPPEVKVGRLGQNPKDVSPLPGPGA
GTGPRMVTRNVNPSPIIQQQAQPRRGTDTDVARLYLLARRVGEPLALSGLAALIKGRNNN
TTPPNNLIIGGARAQSRSGENSDAGRLQVRSATAGVIVHPQETDAARRAARQDQDDDNQV

Appendix 4: Pseudo sequences of amino acids for each of the 9388 *Mycobacterium tuberculosis* strains and reference genomes.

© University of Pretoria

ID	Location	Power	Allele stat	Allele fre	Locus tag	Gene nam	Codon	Annotatio
16025	1779274	0,95	G,L,P|H	0.99,0.0,0.	Rv1572c	hypotheti	9	Rv1572c, (
12802	1385255	0,96	G,P|G,R	0.99,0.0|0	Rv1243c	PE_PGRS2	475	PE-PGRS f.
14921	1636170	0,93	N,M,S|N,I	0.99,0.0,0.	Rv1452c	PE_PGRS2	687	PE-PGRS f.
9209	968434	0,92	Q,-,P|-,P	0.99,0.0,0.	Rv0872c	PE_PGRS1	604	PE-PGRS f.
13067	1418868	0,89	S,P,Q,R|L,	0.98,0.0,0.	Rv1269c	Conservec	29	Rv1269c, (
8248	859239	0,73	A,I,P|I,S	0.94,0.04,(	Rv0766c	cyp123	279	Probable (
11024	1191351	0,95	N,C|C,H	0.99,0.0|0	Rv1068c	PE_PGRS2	267	PE-PGRS f.
9204	968429	0,8	G,-,A,N|-,	0.97,0.0,0.	Rv0872c	PE_PGRS1	606	PE-PGRS f.
31031	3739978	0,99	N,Y|H,L	0.99,0.0|0	Rv3345c	PE_PGRS5	933	PE-PGRS f.
10911	1189139	0,97	G,A|A	0.99,0.0|1	Rv1067c	PE_PGRS1	429	PE-PGRS f.
31712	3766912	0,96	A,G,R,W|(	0.99,0.0,0.	Rv3350c	PPE56	64	PPE family
33168	3899451	0,95	H,P,R|P	0.99,0.0,0.	Rv3480c	Possible t	318	Rv3480c, (
13069	1418870	0,95	A,L|P,T	0.99,0.0|0	Rv1269c	Conservec	29	Rv1269c, (
18419	2061447	0,9	R,A,L|A,C,	0.99,0.0,0.	Rv1818c	PE_PGRS3	410	PE-PGRS f.
14920	1636166	0,8	G|G,N	1.0|0.01,0	Rv1452c	PE_PGRS2	689	PE-PGRS f.
22518	2564374	0,97	V,D|G	0.99,0.0|1	Rv2293c	hypotheti	220	Rv2293c, (
19478	2180650	0,99	S,G,R|H,R	0.99,0.0,0.	Rv1928c	Probable s	190	Rv1928c, (
3184	334657	0,97	L,A,G|A,G	0.99,0.0,0.	Rv0278c	PE_PGRS3	552	PE-PGRS f.
3266	335704	0,96	R,P|P	0.99,0.0|1	Rv0278c	PE_PGRS3	203	PE-PGRS f.
6567	673341	0,95	A|D,G	1.0|0.98,0	Rv0578c	PE_PGRS7	859	PE-PGRS f.
557	49692	0,96	G,A|G,E	0.99,0.0|0	Rv0045c	Possible h	83	Rv0045c, (
37405	4326464	0,95	N|I,T	1.0|0.98,0	Rv3854c	ethA	337	Monooxyg
18926	2133472	0,93	L,*,F,Y|Y	0.99,0.0,0.	Rv1883c	hypotheti	74	Rv1883c, (
8585	916684	0,85	R|R,L	1.0|0.01,0	Rv0823c	Possible t	322	Rv0823c, (
8249	859240	0,69	E,R,V|E,K,	0.99,0.0,0.	Rv0766c	cyp123	278	Probable (
38403	4408009	0,98	V,A|G	0.99,0.0|1	Rv3919c	gid	65	Probable g
3263	335701	0,98	R,A,H|A,G	0.99,0.0,0.	Rv0278c	PE_PGRS3	204	PE-PGRS f.
3259	335693	0,98	A,G|G,P,S	0.99,0.0|0	Rv0278c	PE_PGRS3	207	PE-PGRS f.
10920	1189147	0,97	T,A|A	0.99,0.0|1	Rv1067c	PE_PGRS1	427	PE-PGRS f.
17653	1986637	0,78	R,G|R,Q	0.99,0.0|0	Rv1754c	hypotheti	12	Rv1754c, (
12801	1385254	0,79	G,A|A	0.95,0.04|	Rv1243c	PE_PGRS2	475	PE-PGRS f.
30728	3737112	0,99	D,N|D,N	0.99,0.0|0	Rv3344c	PE_PGRS4	443	PE-PGRS f.
30794	3738028	0,99	H,G|A,G	0.99,0.0|0	Rv3344c	PE_PGRS4	138	PE-PGRS f.
31203	3743401	0,99	A,L,P|L	0.99,0.0,0.	Rv3346c	Conservec	19	Rv3346c, (
4046	373368	0,97	T,P|N	0.99,0.0|1	Rv0305c	PPE6	782	PPE family
30788	3738024	0,99	G,H,I,Q,R|	0.99,0.0,0.	Rv3344c	PE_PGRS4	139	PE-PGRS f.

Appendix 5: Universal makers of antibiotic resistance found in functional and non-functional genes in *M. tuberculosis* isolates.

© University of Pretoria

Position	Power	Alleles	Allelic_fre	Locus	Gene	Codon	Annotatio	MolType	Antibiotic
1280683	0.98	A,D|A,D	0.99,0.0|0	Rv1153c	omt	55	Probable (	G->C,G->G	EMB
1721425	0.97	G,R|G,R	0.99,0.0|0	Rv1526c	Probable g	213	Rv1526c, (	Not found	EMB
4363482	0.97	W,A|A,L	0.99,0.0|0	Rv3883c	mycP1	426	Membran	C->-,C->C,	EMB
4160535	0.96	A,-|A,M	0.99,0.0|0	Rv3716c	hypotheti	127	Rv3716c, (	C->-,C->C,	EMB
4407851	0.96	R,A|R,A	0.99,0.0|0	Rv3919c	gid	118	Probable g	G->-,G->G,	EMB
4385714	0.95	A,P|A,P	0.99,0.0|0	Rv3900c	Conserved	199	Rv3900c, (	C->-,C->C,	EMB
2550013	0.86	T,-,K,R,S|T	0.99,0.0,0.	Rv2277c	Possible g	6	Rv2277c, (	G->-,G->C,	EMB
4385712	0.84	L,H|L,P	0.99,0.0|0	Rv3900c	Conserved	200	Rv3900c, (	Not found	EMB
2604156	0.63	I,L|I,L	0.99,0.0|0	Rv2330c	lppP	23	Probable l	T->-,T->T,	EMB
2536625	0.99	R,G|R,G	0.99,0.0|0	Rv2264c	Conserved	576	Rv2264c, (	C->A,C->C	INH
1002282	0.98	A,G|A,G	0.99,0.0|0	Rv0897c	Probable (	45	Rv0897c, (	G->G,G->G	INH
1280683	0.98	A,D|A,D	0.99,0.0|0	Rv1153c	omt	55	Probable (	G->C,G->G	INH
1990551	0.98	A,G|A,G	0.99,0.0|0	Rv1759c	wag22	676	PE-PGRS f.	G->-,G->C,	INH
4363482	0.97	L,A|A,W	0.99,0.0|0	Rv3883c	mycP1	426	Membran	C->-,C->C,	INH
466846	0.96	S|S,R	1.0|0.01,0	Rv0387c	hypotheti	188	Rv0387c, (	Not found	INH
1468914	0.96	A,P,S|G,S	0.99,0.0,0.	Rv1313c	Possible t	198	Rv1313c, (	C->-,C->CT	INH
3100154	0.96	G,S|G,S	0.99,0.0|0	Rv2790c	ltp1	6	Probable l	A->-,A->A,	INH
4407851	0.96	R,A|R,A	0.99,0.0|0	Rv3919c	gid	118	Probable g	G->-,G->G,	INH
4160535	0.95	A,-|A,M	0.99,0.0|0	Rv3716c	hypotheti	127	Rv3716c, (	C->-,C->C,	INH
1721425	0.88	G,R|G,R	0.99,0.0|0	Rv1526c	Probable g	213	Rv1526c, (	Not found	INH
4385712	0.81	L,H|L,P	0.99,0.0|0	Rv3900c	Conserved	200	Rv3900c, (	Not found	INH
1002282	0.97	A,G|A,G	0.99,0.0|0	Rv0897c	Probable (	45	Rv0897c, (	G->G,G->G	PZA
1774033	0.97	A|A,Q	1.0|0.01,0	Rv1566c	Possible I	197	Rv1566c, (	T->-,T->C,	PZA
2455970	0.97	R,S|R,S	0.99,0.0|0	Rv2192c	trpD	259	Probable a	Not found	PZA
4407851	0.97	R,A|R,A	0.99,0.0|0	Rv3919c	gid	118	Probable g	G->-,G->G,	PZA
2789297	0.95	A,T|A,K	0.99,0.0|0	Rv2483c	plsC	576	Possible t	C->-,C->C,	PZA
4385714	0.93	A,P|A,P	0.99,0.0|0	Rv3900c	Conserved	199	Rv3900c, (	C->-,C->C,	PZA
4385712	0.88	L,H|L,P	0.99,0.0|0	Rv3900c	Conserved	200	Rv3900c, (	Not found	PZA
1280683	0.98	A,D|A,D	0.99,0.0|0	Rv1153c	omt	55	Probable (	G->C,G->G	RIF
1990551	0.98	A,G|A,G	0.99,0.0|0	Rv1759c	wag22	676	PE-PGRS f.	G->-,G->C,	RIF
2536625	0.98	R,G|R,G	0.99,0.0|0	Rv2264c	Conserved	576	Rv2264c, (	C->A,C->C	RIF
3100154	0.98	V,S|G,S	0.98,0.01|	Rv2790c	ltp1	6	Probable l	A->-,A->A,	RIF
1468914	0.97	A,P,S|G,S	0.98,0.0,0.	Rv1313c	Possible t	198	Rv1313c, (	C->-,C->CT	RIF
2302099	0.97	P,A|-,T	0.99,0.0|0	Rv2048c	pks12	1630	Polyketide	C->-,C->A/	RIF
4363482	0.97	L,A|A,W	0.99,0.0|0	Rv3883c	mycP1	426	Membran	C->-,C->C,	RIF

Appendix 6: The most discriminative polymorphisms to distinguish between resistant and susceptible strains in lineage 1.

© University of Pretoria

'id	Position	Power	Alleles	Allelic_fre	Locus	Gene	Codon	Annotatio	MolType
171	16119	1.0	R|L	1.0|1.0	Rv0014c	pknB	451	Transmem	C->A
183	17657	1.0	Q|R	1.0|1.0	Rv0015c	pknA	369	Transmem	T->C
503	44768	1.0	R|G	1.0|1.0	Rv0041	leuS	403	Probable l	A->G
756	70300	1.0	I|F	1.0|1.0	Rv0064	Probable	561	Rv0064, (N	A->-,A->T
927	82884	1.0	V|A	1.0|1.0	Rv0074	hypotheti	46	Rv0074, (N	T->-,T->C
1208	105139	1.0	G|V	1.0|1.0	Rv0095c	hypotheti	26	Rv0095c, (	C->A,C->T
1486	139297	1.0	G|V	1.0|1.0	Rv0115	hddA	262	Possible D	G->T
1504	141623	1.0	D|N	1.0|1.0	Rv0117	oxyS	142	Oxidative	G->A
1546	146236	1.0	K|N	1.0|1.0	Rv0120c	fusA2	513	Probable	Not found
1752	163705	1.0	P|S	1.0|1.0	Rv0136	cyp138	114	Probable	C->CT,C->
1753	163706	1.0	P|L	1.0|1.0	Rv0136	cyp138	114	Probable	C->T
2022	198720	1.0	K|E	1.0|1.0	Rv0169	mce1A	63	Mce-famil	A->G
2067	203038	1.0	I|T	1.0|1.0	Rv0172	mce1D	188	Mce-famil	T->C
2183	213147	1.0	D|Y	1.0|1.0	Rv0182c	sigG	332	Probable	C->A
2209	215977	1.0	A|V	1.0|1.0	Rv0185	hypotheti	88	Rv0185, (N	C->T
2256	222925	1.0	A|T	1.0|1.0	Rv0191	Probable	213	Rv0191, (N	G->A
2325	230170	1.0	P|L	1.0|1.0	Rv0194	Probable	1098	Rv0194, (N	C->CCC,C-
2566	264129	1.0	M|I	1.0|1.0	Rv0221	Possible t	21	Rv0221, (N	G->A
2596	267751	1.0	A|G	1.0|1.0	Rv0223c	Probable	5	Rv0223c, (	G->C
2603	270430	1.0	P|A	1.0|1.0	Rv0226c	Probable	379	Rv0226c, (	G->C
2804	296312	1.0	S|F	1.0|1.0	Rv0245	Possible o	103	Rv0245, (N	Not found
2858	305188	1.0	V|L	1.0|1.0	Rv0252	nirB	775	Probable	G->-,G->A
2910	311521	1.0	D|G	1.0|1.0	Rv0260c	Possible t	380	Rv0260c, (	T->-,T->C
4103	379528	1.0	E|D	1.0|1.0	Rv0311	hypotheti	119	Rv0311, (N	G->T
4128	383716	1.0	P|S	1.0|1.0	Rv0315	Possible b	39	Rv0315, (N	C->T
4921	456413	1.0	I|T	1.0|1.0	Rv0380c	Possible R	136	Rv0380c, (	A->G
5369	511754	1.0	R|S	1.0|1.0	Rv0425c	ctpH	1190	Possible n	G->T
5778	575907	1.0	A|V	1.0|1.0	Rv0486	mshA	187	Glycosyltr	C->T
6034	620625	1.0	I|M	1.0|1.0	Rv0529	ccsA	245	Possible c	A->G
6042	621598	1.0	P|L	1.0|1.0	Rv0530	hypotheti	231	Rv0530, (N	C->T
6704	686264	1.0	A|T	1.0|1.0	Rv0588	yrbE2B	113	Conserved	G->A
7181	754285	1.0	G|D	1.0|1.0	Rv0658c	Probable	42	Rv0658c, (	C->T
7246	761998	1.0	L|P	1.0|1.0	Rv0667	rpoB	731	DNA-direc	T->C
7351	773497	1.0	F|L	1.0|1.0	Rv0673	echA4	125	Possible e	C->A
7370	776182	1.0	D|N	1.0|1.0	Rv0676c	mmpL5	767	Probable	C->T
7490	798355	1.0	A|P	1.0|1.0	Rv0697	Probable	475	Rv0697, (N	G->C
7503	799666	1.0	D|G	1.0|1.0	Rv0699	Hypotheti	13	Rv0699, (N	A->G
7673	829698	1.0	D|E	1.0|1.0	Rv0737	Possible t	164	Rv0737, (N	C->A

Appendix 7: The most discriminative polymorphisms to distinguish between resistant and susceptible strains in lineage 1.2.

© University of Pretoria

'id	Position	Power	Alleles	Allelic_fre	Locus	Gene	Codon	Annotatio	MolType
16026	1779278	0.99	H,-,Q|-,Q	0.99,0.0,0.	Rv1572c	hypotheti	8	Rv1572c, (	Not found
18921	2133467	0.97	P,V|M,S,T	0.99,0.0|0	Rv1883c	hypotheti	76	Rv1883c, (	G->GATCG
13069	1418870	0.96	A,L,T|P,T	0.99,0.0,0.	Rv1269c	Conservec	29	Rv1269c, (	Not found
16025	1779274	0.96	G,L,P|H,P	0.98,0.0,0.	Rv1572c	hypotheti	9	Rv1572c, (	C->CGGGG
557	49692	0.95	G,A|G,E	0.99,0.0|0	Rv0045c	Possible h	83	Rv0045c, (	C->-,C->A,
28054	3311580	0.94	L,I,S|I,T	0.98,0.0,0.	Rv2958c	Possible g	141	Rv2958c, (	G->-,G->C
33168	3899451	0.94	H,P,R|P	0.97,0.01,(	Rv3480c	Possible t	318	Rv3480c, (	Not found
37405	4326464	0.94	N|N,I,T	1.0|0.01,0	Rv3854c	ethA	337	Monooxyg	A->-,A->A
28053	3311579	0.93	L,P|P	0.99,0.0|1	Rv2958c	Possible g	141	Rv2958c, (	A->-,A->A
13067	1418868	0.88	S,P,Q,R|L,	0.96,0.02,(	Rv1269c	Conservec	29	Rv1269c, (	A->AAGGA
8585	916684	0.86	R|R,L	1.0|0.03,0	Rv0823c	Possible t	322	Rv0823c, (	Not found
18925	2133468	0.84	P,*,C,R,W	0.91,0.0,0.	Rv1883c	hypotheti	76	Rv1883c, (	G->GCAGG
18923	2133469	0.83	E,A,D,G,V	0.98,0.0,0.	Rv1883c	hypotheti	75	Rv1883c, (	Not found
22313	2534564	0.8	P,L|L,R	0.9,0.09|0	Rv2262c	hypotheti	330	Rv2262c, (	A->-,A->A
17653	1986637	0.73	R,G|R,Q	0.99,0.0|0	Rv1754c	hypotheti	12	Rv1754c, (	T->-,T->TC
14664	1625327	0.72	N,G,L,R|R	0.98,0.0,0.	Rv1446c	opcA	14	Putative C	Not found
8248	859239	0.67	A,I,P|I,S	0.86,0.12,(	Rv0766c	cyp123	279	Probable (	C->CTCAG
8249	859240	0.64	E,R,V|E,K,	0.99,0.0,0.	Rv0766c	cyp123	278	Probable (	T->TAAGC
16025	1779274	0.96	G,L,P|H,P	0.98,0.0,0.	Rv1572c	hypotheti	9	Rv1572c, (	C->CGGGG
28054	3311580	0.96	L,I,S|I,T	0.98,0.0,0.	Rv2958c	Possible g	141	Rv2958c, (	G->-,G->C
13069	1418870	0.95	A,L,T|A,P,	0.99,0.0,0.	Rv1269c	Conservec	29	Rv1269c, (	Not found
33168	3899451	0.95	H,P,R|P	0.97,0.01,(	Rv3480c	Possible t	318	Rv3480c, (	Not found
36441	4206013	0.95	F,C|F,C	0.99,0.0|0	Rv3761c	fadE36	302	Possible a	A->-,A->A,
557	49692	0.94	G,A|G,E	0.99,0.0|0	Rv0045c	Possible h	83	Rv0045c, (	C->-,C->A,
28053	3311579	0.94	L,P|P	0.99,0.0|1	Rv2958c	Possible g	141	Rv2958c, (	A->-,A->A
37405	4326464	0.94	I,N,T|N	0.97,0.01,(	Rv3854c	ethA	337	Monooxyg	A->-,A->A
7511	802427	0.93	P,R|-	0.99,0.0|1	Rv0703	rplW	99	50S ribosc	C->ATCC,C
18924	2133470	0.93	E,M,Q|M	0.99,0.0,0.	Rv1883c	hypotheti	75	Rv1883c, (	C->CCCAT
10775	1168717	0.92	E,A,V|G	0.99,0.0,0.	Rv1046c	Hypotheti	171	Rv1046c, (	Not found
18926	2133472	0.9	L,F,Y|Y	0.99,0.0,0.	Rv1883c	hypotheti	74	Rv1883c, (	C->CAGGC
13067	1418868	0.88	S,P,Q,R|L,	0.96,0.02,(	Rv1269c	Conservec	29	Rv1269c, (	A->AAGGA
8585	916684	0.86	R|R,L	1.0|0.03,0	Rv0823c	Possible t	322	Rv0823c, (	Not found
18925	2133468	0.85	P,*,C,R,W	0.91,0.0,0.	Rv1883c	hypotheti	76	Rv1883c, (	G->GCAGG
22313	2534564	0.81	P,L|L,R	0.9,0.09|0	Rv2262c	hypotheti	330	Rv2262c, (	A->-,A->A,

Appendix 8: The most discriminative polymorphisms to distinguish between resistant and susceptible strains in lineage 2.

© University of Pretoria

'id	Position	Power	Alleles	Allelic_fre	Locus	Gene	Codon	Annotatio	MolType
9946	1082349	0.99	G,D|G,D	0.99,0.0|0	Rv0971c	echA7	79	Probable (	T->-,T->C,T
12689	1368322	0.99	G,P,R|P	0.99,0.0,0.	Rv1225c	hypotheti	134	Rv1225c, (	C->C,C->C(
17347	1955760	0.99	F,R,V|L,R	0.99,0.0,0.	Rv1730c	Possible p	496	Rv1730c, (	C->-,C->C,
17355	1955913	0.99	N,A|A,K	0.99,0.0|0	Rv1730c	Possible p	445	Rv1730c, (	C->-,C->C,
22338	2536688	0.99	S,Q|S,Q	0.99,0.0|0	Rv2264c	Conservec	555	Rv2264c, (	A->A,A->A
26530	3100148	0.99	N,S|N,S	0.99,0.0|0	Rv2790c	ltp1	8	Probable l	T->T,T->TC
26920	3156201	0.99	G,P,R|P	0.99,0.0,0.	Rv2848c	cobB	441	Probable (	C->CG,C->
10486	1121690	0.98	I|I,G	1.0|0.01,0	Rv1004c	Probable i	154	Rv1004c, (	Not found
20855	2325759	0.98	F,V|F,V	0.99,0.0|0	Rv2067c	hypotheti	38	Rv2067c, (	A->A,A->C
21931	2470342	0.98	Q,P|Q,P	0.99,0.0|0	Rv2205c	hypotheti	41	Rv2205c, (	T->C,T->T,
26358	3072705	0.98	I,D|I,D	0.99,0.0|0	Rv2762c	hypotheti	118	Rv2762c, (	T->T,T->TC
26534	3100154	0.98	S,G,R,V|G	0.99,0.0,0.	Rv2790c	ltp1	6	Probable l	A->-,A->A,
38408	4408087	0.97	R,P|R,P	0.99,0.0|0	Rv3919c	gid	39	Probable (	C->-,C->C,
38409	4408100	0.94	L,A|L,A	0.99,0.0|0	Rv3919c	gid	35	Probable (	G->-,G->G,
19116	2155168	0.63	S,T|S,I,N,T	0.99,0.0|0	Rv1908c	katG	315	Catalase-p	C->A,C->G
297	28366	0.99	I,-|-	0.99,0.0|1	Rv0023	Possible t	0	Rv0023, (N	A->-,A->A,
5114	477898	0.99	L,P|L,P	0.99,0.0|0	Rv0399c	lpqK	220	Possible c	A->A,A->A
9946	1082349	0.99	G,D|G,D	0.99,0.0|0	Rv0971c	echA7	79	Probable (	T->-,T->C,T
10486	1121690	0.99	I|I,G	1.0|0.0,0.9	Rv1004c	Probable i	154	Rv1004c, (	Not found
12689	1368322	0.99	P|P,G,R	1.0|0.0,0.9	Rv1225c	hypotheti	134	Rv1225c, (	C->C,C->C(
13935	1535147	0.99	D,A|D,A	0.99,0.0|0	Rv1363c	Possible n	83	Rv1363c, (	C->-,C->C,
26358	3072705	0.99	I,D|I,D	0.99,0.0|0	Rv2762c	hypotheti	118	Rv2762c, (	T->T,T->TC
26530	3100148	0.99	N,S|N,S	0.99,0.0|0	Rv2790c	ltp1	8	Probable l	T->T,T->TC
26920	3156201	0.99	P|P,G,R	1.0|0.0,0.9	Rv2848c	cobB	441	Probable (	C->CG,C->
37728	4359135	0.99	M,I|M,I	0.99,0.0|0	Rv3879c	espK	217	ESX-1 secr	Not found
8317	868157	0.98	Q,E|Q,E	0.99,0.0|0	Rv0774c	Probable (	66	Rv0774c, (	Not found
20855	2325759	0.98	F,V|F,V	0.99,0.0|0	Rv2067c	hypotheti	38	Rv2067c, (	A->A,A->C
37881	4370537	0.98	V,G|V,G	0.99,0.0|0	Rv3887c	eccD2	383	ESX conse	A->-,A->A,
38408	4408087	0.98	P,R|P,R	0.99,0.0|0	Rv3919c	gid	39	Probable (	C->-,C->C,
38409	4408100	0.93	A,L|A,L	0.99,0.0|0	Rv3919c	gid	35	Probable (	G->-,G->G,
19116	2155168	0.71	S|S,I,N,T	1.0|0.23,0	Rv1908c	katG	315	Catalase-p	C->A,C->G
20176	2263623	2.0	I|G,R	1.0|0.5,0.5	Rv2016	Hypotheti	66	Rv2016, (N	C->-,C->CC
22321	2535431	2.0	G|D,V	1.0|0.5,0.5	Rv2262c	hypotheti	41	Rv2262c, (	C->C,C->C,
22437	2550012	2.0	V|C,G	1.0|0.5,0.5	Rv2277c	Possible g	7	Rv2277c, (	G->-,G->G(
2	5	1.0	R|T	1.0|1.0	Rv0001	dnaA	2	Chromosc	T->A,T->C,
5	8	1.0	D|R	1.0|1.0	Rv0001	dnaA	3	Chromosc	C->AGATA
73	7585	1.0	S|T	1.0|1.0	Rv0006	gyrA	95	DNA gyras	G->C
95	9304	1.0	G|D	1.0|1.0	Rv0006	gyrA	668	DNA gyras	G->A

Appendix 9: The most discriminative polymorphisms to distinguish between resistant and susceptible strains in lineage 3.

© University of Pretoria

'id	Position	Power	Alleles	Allelic_fre	Locus	Gene	Codon	Annotatio	MolType
2929	313889	0.99	R,A|R,A	0.99,0.0|0	Rv0261c	narK3	94	Probable i	G->G,G->G
9905	1077624	0.99	A,V|A,G	0.99,0.0|0	Rv0966c	hypotheti	71	Rv0966c, (	G->-,G->A,
13750	1519822	0.99	S,-,W|-,W	0.99,0.0,0.	Rv1353c	Probable t	55	Rv1353c, (	C->C,C->C
13935	1535147	0.99	A,D|A,D	0.99,0.0|0	Rv1363c	Possible n	83	Rv1363c, (	C->-,C->C,
17662	1987018	0.99	H,-|-,L	0.99,0.0|0	Rv1755c	plcD	227	Probable p	G->-,G->A
18995	2141551	0.99	D,N|D,S	0.99,0.0|0	Rv1894c	hypotheti	107	Rv1894c, (	C->-,C->C,
30391	3709338	0.99	E,-,A|-	0.99,0.0,0.	Rv3323c	moaX	126	Probable I	T->G,T->T,
32287	3832353	0.99	A,P|A,P	0.99,0.0|0	Rv3413c	Unknown	232	Rv3413c, (	Not found
33213	3904957	0.99	H,-,D|-,G	0.99,0.0,0.	Rv3485c	Probable s	204	Rv3485c, (	C->C,C->C
37204	4303055	0.99	A,W|A,W	0.99,0.0|0	Rv3828c	Possible r	115	Rv3828c, (	C->C,C->C,
37733	4359174	0.99	P,T|P,T	0.99,0.0|0	Rv3879c	espK	204	ESX-1 secr	Not found
15634	1721425	0.98	G,R|G,R	0.99,0.0|0	Rv1526c	Probable g	213	Rv1526c, (	Not found
22401	2545196	0.98	-,Q,R|-,Q	0.99,0.0,0.	Rv2270	lppN	167	Probable l	C->C,C->C
26358	3072705	0.98	I,D|I,D	0.99,0.0|0	Rv2762c	hypotheti	118	Rv2762c, (	T->T,T->TC
37322	4314991	0.98	R,P|R,P	0.99,0.0|0	Rv3842c	glpQ1	191	Probable g	C->C,C->C
6844	706086	0.97	A,T|A,R	0.99,0.0|0	Rv0611c	Hypotheti	87	Rv0611c, (	C->C,C->T,
38387	4407851	0.96	R,A|R,A	0.99,0.0|0	Rv3919c	gid	118	Probable g	G->-,G->G,
9962	1083551	0.9	A,V|A,G	0.99,0.0|0	Rv0972c	fadE12	67	Acyl-CoA d	G->-,G->A,
2929	313889	0.99	A,R|A,R	0.99,0.0|0	Rv0261c	narK3	94	Probable i	G->G,G->G
9905	1077624	0.99	A,V|A,G	0.99,0.0|0	Rv0966c	hypotheti	71	Rv0966c, (	G->-,G->A,
13750	1519822	0.99	S,-,W|-,W	0.99,0.0,0.	Rv1353c	Probable t	55	Rv1353c, (	C->C,C->C
13935	1535147	0.99	A,D|A,D	0.99,0.0|0	Rv1363c	Possible n	83	Rv1363c, (	C->-,C->C,
14665	1625332	0.99	R,A|A,T	0.99,0.0|0	Rv1446c	opcA	12	Putative C	C->C,C->C
17662	1987018	0.99	H,-|-,L	0.99,0.0|0	Rv1755c	plcD	227	Probable p	G->-,G->A
18995	2141551	0.99	D,N|D,S	0.99,0.0|0	Rv1894c	hypotheti	107	Rv1894c, (	C->-,C->C,
26530	3100148	0.99	N,S|N,S	0.99,0.0|0	Rv2790c	ltp1	8	Probable l	T->T,T->TC
28211	3343384	0.99	K,R|K,R	0.99,0.0|0	Rv2986c	hupB	146	DNA-bind	T->C,T->T
30391	3709338	0.99	E,-,A|E,-,(	0.99,0.0,0.	Rv3323c	moaX	126	Probable I	T->G,T->T,
32287	3832353	0.99	A,P|A,P	0.99,0.0|0	Rv3413c	Unknown	232	Rv3413c, (	Not found
33213	3904957	0.99	H,-,D|-,G	0.99,0.0,0.	Rv3485c	Probable s	204	Rv3485c, (	C->C,C->C
37733	4359174	0.99	P,T|P,T	0.99,0.0|0	Rv3879c	espK	204	ESX-1 secr	Not found
6844	706086	0.98	A,T|A,R	0.99,0.0|0	Rv0611c	Hypotheti	87	Rv0611c, (	C->C,C->T,
9155	964771	0.98	V|V,P	1.0|0.0,0.9	Rv0867c	rpfA	256	Possible r	Not found
9464	1002282	0.98	A,G|A,G	0.99,0.0|0	Rv0897c	Probable o	45	Rv0897c, (	G->G,G->G
15634	1721425	0.98	G,R|G,R	0.99,0.0|0	Rv1526c	Probable g	213	Rv1526c, (	Not found
28018	3308313	0.98	H,P|H,P	0.99,0.0|0	Rv2955c	hypotheti	78	Rv2955c, (	T->-,T->T,T
37204	4303055	0.98	A,W|A,W	0.99,0.0|0	Rv3828c	Possible r	115	Rv3828c, (	C->C,C->C,
37322	4314991	0.98	R,P|R,P	0.99,0.0|0	Rv3842c	glpQ1	191	Probable g	C->C,C->C

Appendix 10: The most discriminative polymorphisms to distinguish between resistant and susceptible strains in lineage 4.

© University of Pretoria

'id	Position	Power	Alleles	Allelic_fre	Locus	Gene	Codon	Annotatio	MolType
69	7570	1.0	A|V	1.0|1.0	Rv0006	gyrA	90	DNA gyras	C->A,C->G
150	14861	1.0	G|V	1.0|1.0	Rv0012	Probable (	258	Rv0012, (N	G->A,G->G
156	15117	1.0	I|M	1.0|1.0	Rv0013	trpG	68	Possible a	C->G
171	16119	1.0	R|L	1.0|1.0	Rv0014c	pknB	451	Transmem	C->A
182	17654	1.0	Q|P	1.0|1.0	Rv0015c	pknA	370	Transmem	T->G
183	17657	1.0	Q|R	1.0|1.0	Rv0015c	pknA	369	Transmem	T->C
327	30943	1.0	S|P	1.0|1.0	Rv0026	hypotheti	408	Rv0026, (N	C->T
401	36008	1.0	D|H	1.0|1.0	Rv0032	bioF2	572	Possible 8	G->C
423	37305	1.0	S|W	1.0|1.0	Rv0035	fadD34	16	Probable f	C->G
426	37334	1.0	T|S	1.0|1.0	Rv0035	fadD34	26	Probable f	A->T
482	42281	1.0	C|F	1.0|1.0	Rv0039c	Possible c	24	Rv0039c, (	C->A
503	44768	1.0	R|G	1.0|1.0	Rv0041	leuS	403	Probable l	A->G
550	49360	1.0	V|I	1.0|1.0	Rv0045c	Possible h	194	Rv0045c, (	C->C,C->T
556	49690	1.0	G|H	1.0|1.0	Rv0045c	Possible h	84	Rv0045c, (	G->-,G->G,
743	69871	1.0	F|L	1.0|1.0	Rv0064	Probable (	418	Rv0064, (N	C->-,C->T
754	70267	1.0	V|F	1.0|1.0	Rv0064	Probable (	550	Rv0064, (N	G->-,G->T,
798	73200	1.0	N|D	1.0|1.0	Rv0066c	icd2	438	Probable i	T->C
893	79558	1.0	A|P	1.0|1.0	Rv0071	Possible n	25	Rv0071, (N	G->-,G->C,
899	79571	1.0	A|V	1.0|1.0	Rv0071	Possible n	29	Rv0071, (N	C->-,C->G,
1141	103756	1.0	P|H	1.0|1.0	Rv0094c	hypotheti	303	Rv0094c, (	G->A,G->G
1190	104936	1.0	T|A	1.0|1.0	Rv0095c	hypotheti	94	Rv0095c, (	G->-,G->A,
1191	104940	1.0	Q|H	1.0|1.0	Rv0095c	hypotheti	93	Rv0095c, (	C->-,C->A,
1192	104941	1.0	Q|R	1.0|1.0	Rv0095c	hypotheti	92	Rv0095c, (	C->-,C->TC
1193	104942	1.0	Q|E	1.0|1.0	Rv0095c	hypotheti	92	Rv0095c, (	G->-,G->C,
1194	104943	1.0	H|Q	1.0|1.0	Rv0095c	hypotheti	92	Rv0095c, (	Not found
1195	104944	1.0	A|V	1.0|1.0	Rv0095c	hypotheti	91	Rv0095c, (	G->-,G->A,
1196	104962	1.0	A|V	1.0|1.0	Rv0095c	hypotheti	85	Rv0095c, (	C->-,C->GC
1198	105007	1.0	S|T	1.0|1.0	Rv0095c	hypotheti	70	Rv0095c, (	C->-,C->CA
1208	105139	1.0	G|V	1.0|1.0	Rv0095c	hypotheti	26	Rv0095c, (	C->A,C->T
1332	123454	1.0	Q|*	1.0|1.0	Rv0104	hypotheti	380	Rv0104, (N	C->T
1336	123520	1.0	Y|H	1.0|1.0	Rv0104	hypotheti	402	Rv0104, (N	T->C
1486	139297	1.0	G|V	1.0|1.0	Rv0115	hddA	262	Possible D	G->T
1504	141623	1.0	D|N	1.0|1.0	Rv0117	oxyS	142	Oxidative	G->A
1519	143207	1.0	S|G	1.0|1.0	Rv0118c	oxcA	224	Probable (	T->C
1546	146236	1.0	K|N	1.0|1.0	Rv0120c	fusA2	513	Probable (	Not found
1555	147262	1.0	N|K	1.0|1.0	Rv0120c	fusA2	171	Probable (	G->T
1752	163705	1.0	P|S	1.0|1.0	Rv0136	cyp138	114	Probable (	C->CT,C->T
1753	163706	1.0	P|L	1.0|1.0	Rv0136	cyp138	114	Probable (	C->T
1786	166565	1.0	I|V	1.0|1.0	Rv0139	Possible o	247	Rv0139, (N	A->G

Appendix 11: The most discriminative polymorphisms to distinguish between resistant and susceptible strains in lineage 4.1.

© University of Pretoria

'id	Position	Power	Alleles	Allelic_fre	Locus	Gene	Codon	Annotatio	MolType
7235	761155	0.89	S,L,W|L	0.92,0.06,(	Rv0667	rpoB	450	DNA-direc	C->A,C->G
1180	104824	0.63	S,I|S,I	0.98,0.01|	Rv0095c	hypotheti	131	Rv0095c, (	C->-,C->A,
7235	761155	0.6	S,L,W|L	0.93,0.05,(	Rv0667	rpoB	450	DNA-direc	C->A,C->G

Appendix 12: The most discriminative polymorphisms to distinguish between resistant and susceptible strains in lineage 4 (Cameroon).

© University of Pretoria

'id	Position	Power	Alleles	Allelic_fre	Locus	Gene	Codon	Annotatio	MolType
7230	761109	1.0	D|Y	1.0|1.0	Rv0667	rpoB	435	DNA-direc	G->-,G->C,
7239	761277	1.0	I|L	1.0|1.0	Rv0667	rpoB	491	DNA-direc	A->C,A->G
20387	2288727	1.0	L|R	1.0|1.0	Rv2043c	pncA	172	Pyrazinam	A->-,A->C,
36777	4247469	1.0	Y|S	1.0|1.0	Rv3795	embB	319	Integral m	A->C,A->G
36943	4269124	1.0	A|V	1.0|1.0	Rv3806c	ubiA	237	Decapreny	G->A
38071	4383142	0.92	C,P|P	0.92,0.07|	Rv3897c	hypotheti	167	Rv3897c, (	A->-,A->A,
35656	4095000	0.71	D,T|T	0.83,0.16|	Rv3655c	hypotheti	101	Rv3655c, (	Not found
9145	964674	0.64	S,P|P	0.81,0.18|	Rv0867c	rpfA	288	Possible r	A->-,A->G,
12535	1341624	0.64	Q,K|K	0.81,0.18|	Rv1199c	Possible t	328	Rv1199c, (	G->T
16028	1779280	0.64	V,P|P	0.81,0.18|	Rv1572c	hypotheti	7	Rv1572c, (	C->-,C->C,

Appendix 13: The most discriminative polymorphisms to distinguish between resistant and susceptible strains in lineage 4 (Ghana).

© University of Pretoria

'id	Position	Power	Alleles	Allelic_fre	Locus	Gene	Codon	Annotatio	MolType
15572	1715967	1.6	L,*,P|*,P	0.8,0.06,0.	Rv1522c	mmpL12	549	Probable (	A->-,A->A,
19999	2239363	1.6	R,K,N|I,N	0.85,0.07,(	Rv1996	Universal	120	Rv1996, (N	C->-,C->CA
13750	1519822	1.06	L,W|F,W	0.8,0.2|0.5	Rv1353c	Probable t	55	Rv1353c, (	C->C,C->C
18927	2133473	1.0	I|R	1.0|1.0	Rv1883c	hypotheti	74	Rv1883c, (	A->ACGCC
21044	2339617	1.0	A|T	1.0|1.0	Rv2082	hypotheti	303	Rv2082, (N	G->-,G->C
21531	2413615	1.0	D|E	1.0|1.0	Rv2154c	ftsW	437	FtsW-like	Not found
22309	2534559	1.0	R|L	1.0|1.0	Rv2262c	hypotheti	332	Rv2262c, (	A->-,A->A
22380	2542964	1.0	N|T	1.0|1.0	Rv2268c	cyp128	438	Probable (	Not found
26733	3131468	1.0	L|Y	1.0|1.0	Rv2823c	hypotheti	103	Rv2823c, (	G->GACGG
36915	4263501	1.0	T|P	1.0|1.0	Rv3802c	Probable (	289	Rv3802c, (	Not found
37865	4369612	1.0	A|P	1.0|1.0	Rv3886c	mycP2	187	Probable a	G->-,G->C,
5	8	0.93	D,R,Y|R,Y	0.63,0.18,(	Rv0001	dnaA	3	Chromoso	C->AGATA
5634	552507	0.93	C|S	1.0|1.0	Rv0461	Probable t	161	Rv0461, (N	A->AATA,A
9838	1069151	0.93	V,G|G	0.93,0.06|	Rv0957	purH	316	Probable I	Not found
18925	2133468	0.93	C,*|W	0.93,0.06|	Rv1883c	hypotheti	76	Rv1883c, (	G->GCAGG
18931	2133479	0.93	E,G|G	0.93,0.06|	Rv1883c	hypotheti	72	Rv1883c, (	T->TCAC,T
19489	2180817	0.93	M,R|K	0.93,0.06|	Rv1928c	Probable s	134	Rv1928c, (	A->-,A->A
23545	2705439	0.93	A,G|G	0.93,0.06|	Rv2407	hypotheti	248	Rv2407, (N	C->G,Not f
23840	2751960	0.93	L,P|P	0.93,0.06|	Rv2450c	rpfE	74	Probable r	A->AG,A->
23841	2751963	0.93	N,T|T	0.93,0.06|	Rv2450c	rpfE	73	Probable r	T->G
25801	2993862	0.93	V,G|G	0.93,0.06|	Rv2677c	hemY	44	Probable p	A->C,Not f
27165	3190151	0.93	L,W|W	0.93,0.06|	Rv2879c	hypotheti	1	Rv2879c, (	Not found
27693	3262948	0.93	V,G|G	0.93,0.06|	Rv2934	ppsD	234	Phenolpth	Not found
27708	3266044	0.93	V,G|G	0.93,0.06|	Rv2934	ppsD	1266	Phenolpth	Not found
36943	4269124	0.93	A,V|V	0.93,0.06|	Rv3806c	ubiA	237	Decapreny	G->A
37867	4369618	0.93	A,P|P	0.93,0.06|	Rv3886c	mycP2	185	Probable a	G->A,Not f
1033	91805	0.86	A,G|G	0.86,0.13|	Rv0083	Probable (	469	Rv0083, (N	Not found
1244	109635	0.86	R,*|*	0.86,0.13|	Rv0099	fadD10	494	Possible f	Not found
1300	118429	0.86	P,Q|Q	0.86,0.13|	Rv0102	Probable (	239	Rv0102, (N	C->A,Not f
5633	552501	0.86	Y,C,S|S	0.86,0.06,(	Rv0461	Probable t	159	Rv0461, (N	A->-,A->A,
5821	584045	0.86	A,P|P	0.86,0.13|	Rv0493c	hypotheti	216	Rv0493c, (	C->G,C->T
6652	679865	0.86	F,V|V	0.86,0.13|	Rv0584	Possible c	213	Rv0584, (N	Not found
9248	970823	0.86	L,V|V	0.86,0.13|	Rv0873	fadE10	107	Probable a	Not found
13788	1524566	0.86	T,S|S	0.86,0.13|	Rv1356c	Hypotheti	86	Rv1356c, (	Not found
15574	1715969	0.86	Q,*|*	0.86,0.13|	Rv1522c	mmpL12	549	Probable (	Not found
17787	1990716	0.86	A,G|G	0.86,0.13|	Rv1759c	wag22	621	PE-PGRS f	A->-,C->-,(
18502	2064958	0.86	T,P|P	0.86,0.13|	Rv1820	ilvG	54	Probable a	A->C,Not f
19996	2239350	0.86	A,V|V	0.86,0.13|	Rv1996	Universal	116	Rv1996, (N	C->-,C->T
19997	2239352	0.86	N,D|D	0.86,0.13|	Rv1996	Universal	117	Rv1996, (N	A->-,A->G

Appendix 14: The most discriminative polymorphisms to distinguish between resistant and susceptible strains in lineage 4 (Haarlem).

© University of Pretoria

'id	Position	Power	Alleles	Allelic_fre	Locus	Gene	Codon	Annotatio	MolType
7625	820620	0.99	P,H|P,H	0.99,0.0|0	Rv0728c	serA2	286	Possible D	T->-,T->T,T
10866	1187298	0.99	L,P|L,P	0.99,0.0|0	Rv1064c	lpqV	9	Possible li	A->A,A->A
18931	2133479	0.99	E,G|P,T	0.99,0.0|0	Rv1883c	hypotheti	72	Rv1883c, (	T->TCAC,T
15995	1774033	0.93	A|A,Q	1.0|0.0,0.9	Rv1566c	Possible In	197	Rv1566c, (	T->-,T->C,T
10866	1187298	0.99	L,P|L,P	0.99,0.0|0	Rv1064c	lpqV	9	Possible li	A->A,A->A
18675	2094916	0.99	S,H|A,T	0.99,0.0|0	Rv1844c	gnd1	92	Probable (	Not found
18931	2133479	0.99	E,G|A,P,T	0.99,0.0|0	Rv1883c	hypotheti	72	Rv1883c, (	T->TCAC,T
37728	4359135	0.99	I,M|I,M	0.99,0.0|0	Rv3879c	espK	217	ESX-1 secr	Not found
7625	820620	0.98	P,H|P,H	0.99,0.0|0	Rv0728c	serA2	286	Possible D	T->-,T->T,T
22338	2536688	0.98	Q,S|Q,S	0.99,0.0|0	Rv2264c	Conserved	555	Rv2264c, (	A->A,A->A
15995	1774033	0.92	A|A,Q	1.0|0.0,0.9	Rv1566c	Possible In	197	Rv1566c, (	T->-,T->C,T
19116	2155168	0.67	S|S,I,N,T	1.0|0.23,0	Rv1908c	katG	315	Catalase-p	C->A,C->G
10866	1187298	0.99	L,P|L,P	0.99,0.0|0	Rv1064c	lpqV	9	Possible li	A->A,A->A
18931	2133479	0.99	E,G|P,T	0.99,0.0|0	Rv1883c	hypotheti	72	Rv1883c, (	T->TCAC,T
22515	2564368	0.97	G|G,A	1.0|0.01,0	Rv2293c	hypotheti	222	Rv2293c, (	G->-,G->C,
15995	1774033	0.94	A|A,Q	1.0|0.0,0.9	Rv1566c	Possible In	197	Rv1566c, (	T->-,T->C,T
10866	1187298	0.99	L,P|L,P	0.99,0.0|0	Rv1064c	lpqV	9	Possible li	A->A,A->A
7625	820620	0.98	P,H|P,H	0.99,0.0|0	Rv0728c	serA2	286	Possible D	T->-,T->T,T
22338	2536688	0.98	Q,S|Q,S	0.99,0.0|0	Rv2264c	Conserved	555	Rv2264c, (	A->A,A->A
15995	1774033	0.92	A|A,Q	1.0|0.0,0.9	Rv1566c	Possible In	197	Rv1566c, (	T->-,T->C,T
19116	2155168	0.62	S,T|S,I,N,T	0.99,0.0|0	Rv1908c	katG	315	Catalase-p	C->A,C->G

Appendix 15: The most discriminative polymorphisms to distinguish between resistant and susceptible strains in lineage 4 (LAM).

© University of Pretoria

'id	Position	Power	Alleles	Allelic_fre	Locus	Gene	Codon	Annotatio	MolType
8317	868157	0.97	Q,E|Q,E	0.99,0.0|0	Rv0774c	Probable	66	Rv0774c, (	Not found
13935	1535147	0.96	D,A|D,A	0.99,0.0|0	Rv1363c	Possible n	83	Rv1363c, (	C->-,C->C,
13935	1535147	0.98	D,A|D,A	0.99,0.0|0	Rv1363c	Possible n	83	Rv1363c, (	C->-,C->C,
8317	868157	0.96	Q,E|Q,E	0.99,0.0|0	Rv0774c	Probable	66	Rv0774c, (	Not found
8317	868157	0.95	Q,E|Q,E	0.99,0.0|0	Rv0774c	Probable	66	Rv0774c, (	Not found
13935	1535147	0.97	D,A|D,A	0.99,0.0|0	Rv1363c	Possible n	83	Rv1363c, (	C->-,C->C,
8317	868157	0.96	Q,E|Q,E	0.99,0.0|0	Rv0774c	Probable	66	Rv0774c, (	Not found

Appendix 16: The most discriminative polymorphisms to distinguish between resistant and susceptible strains in lineage 4 (S-type).

© University of Pretoria

<table>
<tr><th>'id</th><th>Position</th><th>Power</th><th>Alleles</th><th>Allelic_fre</th><th>Locus</th><th>Gene</th><th>Codon</th><th>Annotatio</th><th>MolType</th></tr>
<tr><td>22754</td><td>2604156</td><td>0.95</td><td>I,L|I,L</td><td>0.98,0.01|</td><td>Rv2330c</td><td>lppP</td><td>23</td><td>Probable l</td><td>T->-,T->T,T</td></tr>
<tr><td>22754</td><td>2604156</td><td>0.95</td><td>I,L|I,L</td><td>0.98,0.01|</td><td>Rv2330c</td><td>lppP</td><td>23</td><td>Probable l</td><td>T->-,T->T,T</td></tr>
<tr><td>22754</td><td>2604156</td><td>0.95</td><td>I,L|I,L</td><td>0.98,0.01|</td><td>Rv2330c</td><td>lppP</td><td>23</td><td>Probable l</td><td>T->-,T->T,T</td></tr>
</table>

Appendix 17: The most discriminative polymorphisms to distinguish between resistant and susceptible strains in lineage 4 (Tur).

© University of Pretoria

'id	Position	Power	Alleles	Allelic_fre	Locus	Gene	Codon	Annotatio	MolType
7235	761155	1.0	S|L	1.0|1.0	Rv0667	rpoB	450	DNA-direc	C->A,C->G
12520	1341099	1.0	I|V	1.0|1.0	Rv1198	esxL	32	Putative E	A->G
12523	1341114	1.0	T|A	1.0|1.0	Rv1198	esxL	37	Putative E	A->A,A->G
12524	1341120	1.0	S|G	1.0|1.0	Rv1198	esxL	39	Putative E	A->G
20435	2289213	1.0	Q|R	1.0|1.0	Rv2043c	pncA	10	Pyrazinam	T->C,T->G
36800	4249518	1.0	H|R	1.0|1.0	Rv3795	embB	1002	Integral m	A->G
37808	4363805	1.0	A|V	1.0|1.0	Rv3883c	mycP1	318	Membran	G->A,G->T
19116	2155168	0.95	S,T|T	0.95,0.04|	Rv1908c	katG	315	Catalase-p	C->A,C->G
21030	2339448	0.91	S,T|T	0.91,0.08|	Rv2082	hypotheti	247	Rv2082, (N	G->-,G->A
1001	89388	0.87	P,S|S	0.91,0.08|	Rv0080	hypotheti	123	Rv0080, (N	C->-,C->G,
2214	216586	0.87	Q,H|H	0.91,0.08|	Rv0186	bglS	106	Probable I	A->C,A->G
26305	3065920	0.87	P,L|L	0.91,0.08|	Rv2752c	hypotheti	91	Rv2752c, (	G->-,G->A,
16149	1789446	0.83	V,I|I	0.9,0.09|1	Rv1588c	Partial REI	131	Rv1588c, (	A->-,C->T,
1294	117403	0.75	A,V|V	0.9,0.1|1.0	Rv0101	nrp	2468	Probable p	C->T
2335	232110	0.75	T,M|M	0.9,0.1|1.0	Rv0196	Possible t	155	Rv0196, (N	C->T
5440	524095	0.75	S,P|P	0.9,0.1|1.0	Rv0435c	Putative c	147	Rv0435c, (	A->G
7609	818798	0.75	N,K|K	0.9,0.1|1.0	Rv0726c	Possible S	282	Rv0726c, (	Not found
10864	1186775	0.75	V,I|I	0.9,0.1|1.0	Rv1063c	hypotheti	17	Rv1063c, (	C->T
13043	1416410	0.75	L,R|R	0.9,0.1|1.0	Rv1267c	embR	313	Probable t	A->C
13117	1430158	0.75	A,T|T	0.9,0.1|1.0	Rv1279	Probable c	33	Rv1279, (N	G->A,G->C
13650	1502795	0.75	M,T|T	0.9,0.1|1.0	Rv1334	mec	52	Possible h	T->-,T->C
15494	1706034	0.75	C,R|R	0.9,0.1|1.0	Rv1514c	hypotheti	188	Rv1514c, (	A->G
15498	1706262	0.75	K,E|E	0.9,0.1|1.0	Rv1514c	hypotheti	112	Rv1514c, (	T->C
15880	1757105	0.75	Q,P|P	0.9,0.1|1.0	Rv1551	plsB1	554	Possible a	A->C
15931	1763855	0.75	D,G|G	0.9,0.1|1.0	Rv1559	ilvA	143	Probable t	A->G
20478	2294896	0.75	L,F|F	0.9,0.1|1.0	Rv2048c	pks12	4031	Polyketide	G->A
21787	2447282	0.75	E,D|D	0.9,0.1|1.0	Rv2185c	TB16.3	74	Conserved	T->G
22506	2562752	0.75	V,I|I	0.9,0.1|1.0	Rv2290	lppO	52	Probable c	G->A
24071	2794793	0.75	M,I|I	0.9,0.1|1.0	Rv2486	echA14	148	Probable c	G->A
25095	2899538	0.75	D,A|A	0.9,0.1|1.0	Rv2575	Possible c	67	Rv2575, (N	A->C
26337	3069778	0.75	F,L|L	0.9,0.1|1.0	Rv2756c	hsdM	103	Possible t	G->A,G->T
29053	3480435	0.75	H,R|R	0.9,0.1|1.0	Rv3113	Possible p	121	Rv3113, (N	A->-,A->G
30315	3696179	0.75	V,A|A	0.9,0.1|1.0	Rv3308	pmmB	439	Probable p	T->C
33155	3898637	0.75	V,I|I	0.9,0.1|1.0	Rv3479	Possible t	940	Rv3479, (N	G->A
34621	3958797	0.75	A,V|V	0.9,0.1|1.0	Rv3522	ltp4	117	Possible li	C->T
36229	4179179	0.75	I,V|V	0.9,0.1|1.0	Rv3729	Possible t	299	Rv3729, (N	A->G
36518	4214751	0.75	T,N|N	0.9,0.1|1.0	Rv3769	Hypotheti	46	Rv3769, (N	Not found
36758	4245055	0.75	T,N|N	0.9,0.1|1.0	Rv3794	embA	608	Integral m	Not found
37844	4367911	0.75	I,T|T	0.9,0.1|1.0	Rv3885c	eccE2	204	ESX conse	A->AC,A->

Appendix 18: The most discriminative polymorphisms to distinguish between resistant and susceptible strains in lineage 4 (Uganda).

© University of Pretoria

'id	Position	Power	Alleles	Allelic_fre	Locus	Gene	Codon	Annotatio	MolType
20966	2338200	0.78	V|V,G	1.0|0.2,0.8	Rv2081c	Conserved	103	Rv2081c, (	Not found
16498	1843567	0.66	I,P|A	0.94,0.05|	Rv1637c	hypotheti	43	Rv1637c, (	Not found
16502	1843636	0.87	H,*|D,I	0.94,0.05|	Rv1637c	hypotheti	20	Rv1637c, (	Not found
19478	2180650	0.85	S,G,R|H,R	0.85,0.02,(	Rv1928c	Probable s	190	Rv1928c, (	Not found
16922	1894296	0.8	G,P|P,R	0.86,0.13|	Rv1668c	Probable t	350	Rv1668c, (	Not found
20966	2338200	0.78	V|V,G	1.0|0.2,0.8	Rv2081c	Conserved	103	Rv2081c, (	Not found
16498	1843567	0.66	I,P|A	0.94,0.05|	Rv1637c	hypotheti	43	Rv1637c, (	Not found
1146	103793	0.99	D,Q,S|Q,R	0.99,0.0,0.	Rv0094c	hypotheti	291	Rv0094c, (	C->CTGCT,
15571	1715939	0.99	A,Q|F,Q	0.99,0.0|0	Rv1522c	mmpL12	559	Probable c	Not found
20966	2338200	0.97	V|V,G	1.0|0.02,0	Rv2081c	Conserved	103	Rv2081c, (	Not found
22754	2604156	0.96	I,L,W|I,L	0.98,0.0,0.	Rv2330c	lppP	23	Probable l	T->-,T->T,T
38403	4408009	0.96	V,G|G	0.99,0.0|1	Rv3919c	gid	65	Probable g	A->AC,A->
17822	1992320	0.95	Y,G,H|G	0.98,0.0,0.	Rv1759c	wag22	87	PE-PGRS f	Not found
18926	2133472	0.95	Y|*	1.0|1.0	Rv1883c	hypotheti	74	Rv1883c, (	C->CAGGC
7511	802427	0.93	P,-,L|-	0.99,0.0,0.	Rv0703	rplW	99	50S riboso	C->ATCC,C
16498	1843567	0.92	I,P|A	0.99,0.0|1	Rv1637c	hypotheti	43	Rv1637c, (	Not found
22313	2534564	0.91	L,P|L,R	0.99,0.0|0	Rv2262c	hypotheti	330	Rv2262c, (	A->-,A->A,
19487	2180815	0.79	V,C,L|C,W	0.93,0.05,(	Rv1928c	Probable s	135	Rv1928c, (	C->-,C->C,
18928	2133474	0.68	C|C,*	1.0|0.2,0.8	Rv1883c	hypotheti	74	Rv1883c, (	Not found
1146	103793	0.98	D,Q,S|Q,R	0.98,0.0,0.	Rv0094c	hypotheti	291	Rv0094c, (	C->CTGCT,
22754	2604156	0.97	I,L|I,L,W	0.98,0.01|	Rv2330c	lppP	23	Probable l	T->-,T->T,T
7511	802427	0.95	-,P|-,L,P	0.99,0.0|0	Rv0703	rplW	99	50S riboso	C->ATCC,C
17822	1992320	0.94	Y,G,H|G	0.97,0.0,0.	Rv1759c	wag22	87	PE-PGRS f	Not found
15152	1655176	0.93	P|P,A,L	1.0|0.03,0	Rv1467c	fadE15	110	Probable a	Not found
37709	4358495	0.93	G|G,E,T	1.0|0.03,0	Rv3879c	espK	430	ESX-1 secr	T->TC,T->T
22313	2534564	0.91	L,P|L,R	0.99,0.0|0	Rv2262c	hypotheti	330	Rv2262c, (	A->-,A->A,
19487	2180815	0.73	V,C,L|C,W	0.88,0.1,0.	Rv1928c	Probable s	135	Rv1928c, (	C->-,C->C,
8122	842057	0.9	V,F|F	0.9,0.09|1	Rv0750	hypotheti	9	Rv0750, (N	G->A,G->T
8123	842058	0.9	V,D|D	0.9,0.09|1	Rv0750	hypotheti	9	Rv0750, (N	T->A
22986	2626513	0.9	T,S|S	0.9,0.09|1	Rv2347c	esxP	3	Putative E	T->A,T->G
3574	341719	0.81	F,V|V	0.81,0.18|	Rv0281	Possible S	241	Rv0281, (N	C->CGG,C-
5475	529908	0.81	A,V|V	0.81,0.18|	Rv0440	groEL2	434	60 kDa cha	C->T
7272	764817	0.81	V,A,G|A,G	0.81,0.09,(	Rv0668	rpoC	483	DNA-direc	T->C,T->G
8120	842051	0.81	D,N|N	0.81,0.18|	Rv0750	hypotheti	7	Rv0750, (N	G->A
8125	842065	0.81	H,Q|Q	0.81,0.18|	Rv0750	hypotheti	11	Rv0750, (N	C->G
16024	1779243	0.81	H,P|P	0.81,0.18|	Rv1572c	hypotheti	19	Rv1572c, (	T->C,T->G,
18149	2030758	0.81	A,G|G	0.81,0.18|	Rv1793	esxN	22	Putative E	C->G
20213	2266660	0.81	R,C|C	0.81,0.18|	Rv2020c	hypotheti	21	Rv2020c, (	G->A
21008	2339276	0.81	V,L|L	0.81,0.18|	Rv2082	hypotheti	190	Rv2082, (N	G->-,G->C,
21017	2339424	0.81	M,T|T	0.81,0.18|	Rv2082	hypotheti	239	Rv2082, (N	G->-,T->-,T

Appendix 19: The most discriminative polymorphisms to distinguish between resistant and susceptible strains in lineage 4 (Ural).

© University of Pretoria

'id	Position	Power	Alleles	Allelic_fre	Locus	Gene	Codon	Annotatio	MolType
17280	1947061	0.99	G,E|G,V	0.99,0.0|0	Rv1720c	vapC12	120	Possible t	C->C,C->C
17698	1987453	0.99	S,N|A,N,T	0.99,0.0|0	Rv1755c	plcD	82	Probable	C->-,C->C
35941	4139180	0.99	G|H,Y	1.0|0.0,0.	Rv3696c	glpK	193	Probable	Not found
4408	424011	0.98	-,R|-,Q,R	0.99,0.0|0	Rv0353	hspR	125	Probable	C->C,C->C
6241	636921	0.98	S,P|A,P	0.99,0.0|0	Rv0545c	pitA	182	Probable	Not found
18676	2094921	0.98	H,R|A,T	0.99,0.0|0	Rv1844c	gnd1	90	Probable	C->-,C->C
26358	3072705	0.98	I,D,G|G	0.99,0.0,0.	Rv2762c	hypotheti	118	Rv2762c, (	T->T,T->T
26739	3131484	0.98	A,D,G,V|G	0.99,0.0,0.	Rv2823c	hypotheti	97	Rv2823c, (	G->GAGAT
37720	4358976	0.98	A,P|P	0.99,0.0|1	Rv3879c	espK	270	ESX-1 secr	Not found
37727	4359126	0.98	I,P|I,P	0.99,0.0|0	Rv3879c	espK	220	ESX-1 secr	Not found
38035	4380295	0.98	N,T|S,T	0.99,0.0|0	Rv3894c	eccC2	53	ESX conse	G->C,G->G
5162	483294	0.97	V,F|V,R	0.99,0.0|0	Rv0403c	mmpS1	122	Probable	Not found
19489	2180817	0.97	M,K,R|M,I	0.98,0.0,0.	Rv1928c	Probable	134	Rv1928c, (	A->-,A->A
24154	2798756	0.96	G,W|P,R	0.99,0.0|0	Rv2488c	Probable	709	Rv2488c, (	C->A,C->C
22756	2604165	0.9	L,*,I|*,C	0.99,0.0,0.	Rv2330c	lppP	20	Probable	A->-,A->A
16502	1843636	0.94	H,*|D,I	0.96,0.03|	Rv1637c	hypotheti	20	Rv1637c, (	Not found
16921	1894298	0.87	S,A|P	0.98,0.01|	Rv1668c	Probable	349	Rv1668c, (	A->-,A->A
21297	2372436	0.72	L|L,P,R	1.0|0.12,0	Rv2112c	dop	45	Deamidas	G->C,G->G
16922	1894296	0.7	G,P|P,R	0.9,0.09|0	Rv1668c	Probable	350	Rv1668c, (	Not found
25778	2990584	0.7	Q,-|-	0.96,0.03|	Rv2673	aftC	432	Possible a	C->CACCG
16502	1843636	0.98	H,*|D,I	0.99,0.0|0	Rv1637c	hypotheti	20	Rv1637c, (	Not found
16921	1894298	0.96	S,A|P	0.99,0.0|1	Rv1668c	Probable	349	Rv1668c, (	A->-,A->A
22313	2534564	0.94	L,P|L,R	0.99,0.0|0	Rv2262c	hypotheti	330	Rv2262c, (	A->-,A->A
25778	2990584	0.93	Q,-|-	0.99,0.0|1	Rv2673	aftC	432	Possible a	C->CACCG
16922	1894296	0.91	G,P|P,R	0.97,0.02|	Rv1668c	Probable	350	Rv1668c, (	Not found
38071	4383142	0.86	C,G,P|G	0.97,0.01,(	Rv3897c	hypotheti	167	Rv3897c, (	A->-,A->A
6241	636921	0.99	S,P|A,P	0.99,0.0|0	Rv0545c	pitA	182	Probable	Not found
35941	4139180	0.99	G|H,Y	1.0|0.0,0.	Rv3696c	glpK	193	Probable	Not found
38035	4380295	0.99	N,T|S,T	0.99,0.0|0	Rv3894c	eccC2	53	ESX conse	G->C,G->G
16921	1894298	0.98	S,A,G|P	0.99,0.0,0.	Rv1668c	Probable	349	Rv1668c, (	A->-,A->A
25465	2950630	0.98	F|P,R	1.0|0.99,0	Rv2624c	Universal	227	Rv2624c, (	Not found
37727	4359126	0.98	I,P|I,P	0.99,0.0|0	Rv3879c	espK	220	ESX-1 secr	Not found
9889	1075947	0.97	A,*,T|L,R,	0.99,0.0,0.	Rv0963c	hypotheti	51	Rv0963c, (	C->-,C->C
26732	3131470	0.97	H,R,T|T	0.99,0.0,0.	Rv2823c	hypotheti	102	Rv2823c, (	T->-,T->TG
22756	2604165	0.92	L,*,I|*,C	0.99,0.0,0.	Rv2330c	lppP	20	Probable	A->-,A->A
16028	1779280	0.93	S,T,V|P,T,	0.98,0.0,0.	Rv1572c	hypotheti	7	Rv1572c, (	C->-,C->C,
18922	2133465	0.85	C|C,G,V	1.0|0.02,0	Rv1883c	hypotheti	77	Rv1883c, (	Not found
16028	1779280	0.88	V,P,T|V,S	0.94,0.0,0.	Rv1572c	hypotheti	7	Rv1572c, (	C->-,C->C,
18922	2133465	0.83	C|C,G,V	1.0|0.04,0	Rv1883c	hypotheti	77	Rv1883c, (	Not found
5159	483287	0.92	E,P|P,R	0.97,0.02|	Rv0403c	mmpS1	124	Probable	C->AGGAC
20686	2306510	0.81	G,D,T|D,R	0.88,0.05,(	Rv2048c	pks12	160	Polyketid	G->C,G->C
22435	2550013	0.78	T,S|K,N	0.93,0.06|	Rv2277c	Possible g	6	Rv2277c, (	G->-,G->C,

Appendix 20: The most discriminative polymorphisms to distinguish between resistant and susceptible strains in lineage 4 (X-type).

© University of Pretoria

'id	Position	Power	Alleles	Allelic_fre	Locus	Gene	Codon	Annotatio	MolType
2510	257787	1.0	R|-	1.0|1.0	Rv0215c	fadE3	357	Probable a	C->CAAGG
7230	761109	1.0	D|Y	1.0|1.0	Rv0667	rpoB	435	DNA-direc	G->-,G->C,
12520	1341099	1.0	I|V	1.0|1.0	Rv1198	esxL	32	Putative E	A->G
12523	1341114	1.0	T|A	1.0|1.0	Rv1198	esxL	37	Putative E	A->A,A->G
12524	1341120	1.0	S|G	1.0|1.0	Rv1198	esxL	39	Putative E	A->G
12525	1341148	1.0	A|V	1.0|1.0	Rv1198	esxL	48	Putative E	C->T
22761	2604173	1.0	L|S	1.0|1.0	Rv2330c	lppP	17	Probable l	A->-,A->A,
35393	4060588	1.0	T|A	1.0|1.0	Rv3620c	esxW	2	Putative E	T->-,T->AT
36776	4247431	1.0	M|I	1.0|1.0	Rv3795	embB	306	Integral m	G->A,G->C
545	48957	0.94	P,H|H	0.94,0.05|	Rv0044c	Possible o	24	Rv0044c, (	T->G,T->T,
7402	781822	0.94	K,R|R	0.94,0.05|	Rv0682	rpsL	88	30S ribosc	A->C,A->G
10783	1169447	0.94	A,T|T	0.94,0.05|	Rv1047	Probable t	9	Rv1047, (N	C->ACGCC
16028	1779280	0.94	V,P|P	0.94,0.05|	Rv1572c	hypotheti	7	Rv1572c, (	C->-,C->C,
20889	2329747	0.94	G,R|R	0.94,0.05|	Rv2072c	cobL	134	Precorrin-	C->-,C->G
21125	2347464	0.94	L,P|P	0.94,0.05|	Rv2090	Probable 5	31	Rv2090, (N	T->-,T->C
27908	3296369	0.94	R,P|P	0.94,0.05|	Rv2947c	pks15	491	Probable p	C->CAGCC
16922	1894293	0.89	W,K|N	0.89,0.1|1	Rv1668c	Probable t	351	Rv1668c, (	Not found
36841	4253290	0.89	P,A|A	0.89,0.1|1	Rv3798	Probable t	100	Rv3798, (N	C->G
17822	1992320	0.84	Y,H|H	0.88,0.11|	Rv1759c	wag22	87	PE-PGRS f	Not found
22338	2536688	0.84	S,Q|Q	0.88,0.11|	Rv2264c	Conservec	555	Rv2264c, (	A->A,A->A
1180	104824	0.78	S,I|I	0.88,0.11|	Rv0095c	hypotheti	131	Rv0095c, (	C->-,C->A,
23644	2724179	0.78	I,T|T	0.88,0.11|	Rv2426c	hypotheti	2	Rv2426c, (	G->-,G->C,
37728	4359135	0.73	I,M|M	0.87,0.12|	Rv3879c	espK	217	ESX-1 secr	Not found
22754	2604156	0.63	I,L|L	0.85,0.14|	Rv2330c	lppP	23	Probable l	T->-,T->T,T
37534	4340994	0.63	F,C|C	0.85,0.14|	Rv3864	espE	242	ESX-1 secr	T->-,T->G,T
21125	2347464	0.66	L|L,P	1.0|0.33,0	Rv2090	Probable 5	31	Rv2090, (N	T->-,T->C
15577	1716472	0.72	S,P|P	0.86,0.13|	Rv1522c	mmpL12	381	Probable c	A->G,G->-
24502	2829779	0.61	T,A|A	0.84,0.15|	Rv2512c	Transposa	9	Rv2512c, (	T->C
7402	781822	1.0	K|R	1.0|1.0	Rv0682	rpsL	88	30S ribosc	A->C,A->G
20962	2338191	0.83	C,G|G	0.83,0.16|	Rv2081c	Conservec	106	Rv2081c, (	Not found
22436	2550014	0.83	T,A|A	0.83,0.16|	Rv2277c	Possible g	6	Rv2277c, (	C->-,T->-,T
23535	2704887	0.83	G,S|S	0.83,0.16|	Rv2407	hypotheti	64	Rv2407, (N	A->-,A->A,
1180	104824	0.66	S,I|I	0.66,0.33|	Rv0095c	hypotheti	131	Rv0095c, (	C->-,C->A,
37534	4340994	0.66	F,C|C	0.66,0.33|	Rv3864	espE	242	ESX-1 secr	T->-,T->G,T

Appendix 21: The most discriminative polymorphisms to distinguish between resistant and susceptible strains in lineage 5.

© University of Pretoria

'id	Position	Power	Alleles	Allelic_fre	Locus	Gene	Codon	Annotatio	MolType
19116	2155168	0.88	S,I,T|T	0.93,0.04,(	Rv1908c	katG	315	Catalase-p	C->A,C->G
2570	264754	0.68	H,R|-,R	0.94,0.05|	Rv0221	Possible t	230	Rv0221, (N	C->CGGAT
11885	1265912	0.6	S,P|P	0.93,0.06|	Rv1138c	Possible o	193	Rv1138c, (	Not found
17758	1989911	0.6	G,K|K	0.93,0.06|	Rv1759c	wag22	890	PE-PGRS f	A->-,A->A,
22435	2550013	0.96	-,T|R,T	0.96,0.03|	Rv2277c	Possible g	6	Rv2277c, (	G->-,G->C,
37735	4359182	0.72	R,G|P	0.96,0.03|	Rv3879c	espK	201	ESX-1 secr	G->-,G->G
2570	264754	0.61	H,R|-,R	0.9,0.09|0	Rv0221	Possible t	230	Rv0221, (N	C->CGGAT
2570	264754	0.7	H,R|-,R	0.91,0.08|	Rv0221	Possible t	230	Rv0221, (N	C->CGGAT

Appendix 22: The most discriminative polymorphisms to distinguish between resistant and susceptible strains in lineage 6.

© University of Pretoria

Query	Sbjct	[AB,aB,Ab	LD	Chi2	A->a|b	Err1	B->b|a	Err2	
Rv1313c|1	Pks12|163	[9206,85,2	0.71	3041.18	0.45	0.072	0.71	0.159	
Rv1313c|1	Rv3716c|1	[8685,25,5	0.82	1401.14	0.82	0.182	0.18	0.017	
Pks12|163	Rv3716c|1	[8676,34,6	0.63	518.68	0.63	0.138	0.08	0.011	
Rv3716c|1	Rv3900c|1	[8707,669,	0.78	118.21	0.01	0.004	0.78	0.516	

Appendix 23: Epistatic links between polymorphisms in lineage 1.

© University of Pretoria

Query	Sbjct	[AB,aB,Ab	LD	Chi2	A->a|b	Err1	B->b|a	Err2	
DnaA|1M	RecF|245I	[7433,155	0.72	881.02	0.12	0.008	0.72	0.092	
DnaA|1M	GyrA|20I(	[7503,158	0.92	1057.78	0.12	0.008	0.92	0.221	
DnaA|1M	Rv0008c|1	[7495,160	0.88	944.82	0.11	0.007	0.88	0.182	
DnaA|1M	Rv0012|2	[7504,133	0.95	2190.12	0.24	0.012	0.95	0.231	
DnaA|1M	Rv0064|4	[7392,139	0.73	1389.34	0.2	0.011	0.73	0.076	
DnaA|1M	Rv0064|7	[7502,162	0.9	913.03	0.1	0.007	0.9	0.214	
DnaA|1M	Rv0068|2	[7440,157	0.72	826.88	0.12	0.008	0.72	0.096	
DnaA|1M	Mak|17L(	[7506,163	0.91	866.67	0.1	0.007	0.91	0.241	
DnaA|1M	Mce1F|37	[7433,162	0.66	609.36	0.09	0.007	0.66	0.089	
DnaA|1M	Rv0193c|4	[7434,157	0.71	809.38	0.12	0.008	0.71	0.092	
DnaA|1M	Rv0194|7	[7436,155	0.73	899.93	0.13	0.008	0.73	0.094	
DnaA|1M	NrdB|32W	[7479,158	0.83	956.73	0.12	0.008	0.83	0.141	
DnaA|1M	Rv0259c|1	[7491,158	0.87	1009.21	0.12	0.008	0.87	0.169	
DnaA|1M	EccE3|216	[7502,163	0.9	875.21	0.1	0.007	0.9	0.214	
DnaA|1M	Rv0318c|2	[7438,162	0.68	636.8	0.09	0.007	0.68	0.093	
DnaA|1M	Rv0338c|6	[7503,158	0.92	1057.78	0.12	0.008	0.92	0.221	
DnaA|1M	Rv0395|8	[7425,160	0.66	667.26	0.1	0.007	0.66	0.085	
DnaA|1M	ThiG|75S(	[7491,162	0.85	825.2	0.1	0.007	0.85	0.168	
DnaA|1M	CtpH|689I	[7437,157	0.72	843.79	0.12	0.008	0.72	0.094	
DnaA|1M	MmpL2|4	[7433,158	0.7	766.71	0.11	0.008	0.7	0.091	
DnaA|1M	Rv0556|1	[7434,154	0.73	936.64	0.13	0.008	0.73	0.093	
DnaA|1M	Mce2A|51	[7424,154	0.7	892.65	0.13	0.008	0.7	0.087	
DnaA|1M	Rv0658c|7	[7507,164	0.92	859.38	0.09	0.007	0.92	0.249	
DnaA|1M	MmpL5|9	[7478,167	0.77	593.38	0.08	0.006	0.77	0.135	
DnaA|1M	PhoR|172	[7510,164	0.93	854.51	0.09	0.007	0.93	0.28	
DnaA|1M	Rv0812|1	[7510,166	0.93	791.56	0.09	0.006	0.93	0.28	
DnaA|1M	Rv0829|8	[7462,154	0.8	1027.4	0.13	0.008	0.8	0.117	
DnaA|1M	FadA|149	[7503,159	0.91	1040.8	0.12	0.007	0.91	0.221	
DnaA|1M	Rv0881|1	[7443,165	0.66	554.41	0.08	0.006	0.66	0.095	
DnaA|1M	Rv0891c|3	[7431,154	0.72	905.66	0.13	0.008	0.72	0.091	
DnaA|1M	PstA1|304	[7513,136	0.97	2084.81	0.22	0.011	0.97	0.33	
DnaA|1M	Rv0964c|1	[7435,157	0.71	801.1	0.11	0.008	0.71	0.093	
DnaA|1M	PepD|390	[7502,158	0.91	1086.49	0.12	0.008	0.91	0.215	
DnaA|1M	GalU|235	[7434,156	0.71	844.91	0.12	0.008	0.71	0.092	
DnaA|1M	MetS|39R	[7435,161	0.67	653.72	0.1	0.007	0.67	0.091	
DnaA|1M	GlyA1|36	[7503,159	0.91	1011.13	0.11	0.007	0.91	0.221	
DnaA|1M	Pks3|489*	[7511,138	0.97	2001.37	0.21	0.011	0.97	0.297	
DnaA|1M	Rv1320c|5	[7500,170	0.85	571.12	0.07	0.005	0.85	0.199	
DnaA|1M	Rv1321|1	[7502,160	0.91	1001.63	0.11	0.007	0.91	0.215	
DnaA|1M	Rv1364c|4	[7426,161	0.66	651.49	0.1	0.007	0.66	0.086	
DnaA|1M	Rv1374c|1	[7498,157	0.9	1086.8	0.12	0.008	0.9	0.195	

Appendix 24: Epistatic links between polymorphisms in lineage 1.2.

© University of Pretoria

Query	Sbjct	[AB,aB,Ab	LD	Chi2	A->a|b	Err1	B->b|a	Err2	
Rv0010c|1	Rv0823c|3	[8840,1,54	0.64	20.12	0.64	0.816	0.0	0.002	
Rv0010c|1	Rv1269c|2	[8876,1,51	0.64	21.72	0.64	0.816	0.0	0.002	
Rv0010c|1	Rv1435c|1	[8879,1,50	0.64	21.86	0.64	0.816	0.0	0.002	
Rv0010c|1	OpcA|9T(-	[8866,1,52	0.64	21.25	0.64	0.816	0.0	0.002	
Rv0010c|1	Rv1668c|3	[8496,1,89	0.63	11.37	0.63	0.816	0.0	0.001	
Rv0010c|1	Rv1883c|7	[8634,1,75	0.63	13.92	0.63	0.816	0.0	0.001	
Rv0010c|1	Rv1883c|7	[8629,1,75	0.63	13.82	0.63	0.816	0.0	0.001	
Rv0010c|1	Rv1883c|7	[8635,1,75	0.63	13.95	0.63	0.816	0.0	0.001	
Rv0010c|1	Rv1883c|7	[8629,1,75	0.63	13.82	0.63	0.816	0.0	0.001	
Rv0010c|1	Rv2823c|1	[8609,1,77	0.63	13.4	0.63	0.816	0.0	0.001	
Rv0010c|1	Rv3483c|1	[8838,1,55	0.64	20.04	0.64	0.816	0.0	0.002	
Rv0010c|1	Rv3897c|1	[8515,1,87	0.63	11.68	0.63	0.816	0.0	0.001	
Rv0036c|2	Rv0095c|9	[5860,11,3	0.59	25.14	0.59	0.259	0.0	0.001	
Rv0036c|2	Rv0095c|9	[4898,11,4	0.51	12.33	0.51	0.259	0.0	0.0	
Rv0036c|2	Rv0823c|3	[8832,9,51	0.77	419.95	0.77	0.296	0.05	0.01	
Rv0036c|2	Rv0823c|3	[7113,6,22	0.81	90.11	0.81	0.378	0.01	0.002	
Rv0036c|2	Rv1148c|2	[7411,16,1	0.52	45.8	0.52	0.197	0.0	0.002	
Rv0036c|2	Rv1269c|2	[8868,9,48	0.77	452.24	0.77	0.296	0.06	0.011	
Rv0036c|2	OpcA|9T(-	[8857,10,4	0.75	415.24	0.75	0.276	0.05	0.01	
Rv0036c|2	Rv1668c|3	[8490,7,85	0.82	276.12	0.82	0.345	0.03	0.006	
Rv0036c|2	Rv1883c|7	[8627,8,72	0.79	313.93	0.79	0.318	0.04	0.007	
Rv0036c|2	Rv1883c|7	[8622,8,72	0.79	311.6	0.79	0.318	0.04	0.007	
Rv0036c|2	Rv1883c|7	[8628,8,72	0.79	314.4	0.79	0.318	0.04	0.007	
Rv0036c|2	Rv1883c|7	[8623,7,72	0.82	331.69	0.82	0.345	0.04	0.007	
Rv0036c|2	Rv1945|24	[7291,13,2	0.61	56.49	0.61	0.231	0.0	0.002	
Rv0036c|2	LppA|92F(	[7201,16,2	0.51	38.15	0.51	0.197	0.0	0.001	
Rv0036c|2	LppA|93D	[7221,14,2	0.57	48.32	0.57	0.219	0.0	0.001	
Rv0036c|2	LppA|94D	[7216,16,2	0.51	38.65	0.51	0.197	0.0	0.001	
Rv0036c|2	LppB|36G	[5048,11,4	0.52	13.91	0.52	0.259	0.0	0.0	
Rv0036c|2	LppB|38N	[5089,11,4	0.52	14.36	0.52	0.259	0.0	0.0	
Rv0036c|2	LppB|43P(	[5079,11,4	0.52	14.25	0.52	0.259	0.0	0.0	
Rv0036c|2	LppB|44H	[5081,11,4	0.52	14.27	0.52	0.259	0.0	0.0	
Rv0036c|2	Rv2823c|1	[8604,6,74	0.84	342.3	0.84	0.378	0.04	0.007	
Rv0036c|2	Rv3483c|1	[8832,7,51	0.82	473.13	0.82	0.345	0.06	0.01	
Rv0036c|2	Rv3897c|1	[8508,8,84	0.79	265.61	0.79	0.318	0.03	0.006	
Rv0044c|2	Rv0095c|1	[5814,2,35	0.75	11.96	0.75	0.65	0.0	0.0	
Rv0044c|2	LppB|36G	[5057,2,43	0.71	7.75	0.71	0.65	0.0	0.0	
Rv0044c|2	LppB|38N	[5098,2,42	0.71	7.94	0.71	0.65	0.0	0.0	
Rv0044c|2	LppB|43P(	[5088,2,42	0.71	7.9	0.71	0.65	0.0	0.0	

Appendix 25: Epistatic links between polymorphisms in lineage 2.

© University of Pretoria

Query	Sbjct	[AB,aB,Ab	LD	Chi2	A->a|b	Err1	B->b|a	Err2	
DnaA|2T(:	FhaA|246I	[5206,211,	0.54	360.94	0.54	0.057	0.07	0.004	
DnaA|2T(:	FhaA|242I	[8004,183,	0.74	3293.73	0.74	0.063	0.47	0.028	
DnaA|2T(:	FhaA|240I	[8024,185,	0.73	3344.67	0.73	0.063	0.48	0.028	
DnaA|2T(:	Rv0095c|9	[4689,197,	0.53	270.74	0.53	0.06	0.05	0.003	
DnaA|2T(:	Rv0095c|9	[4689,198,	0.52	268.54	0.52	0.06	0.05	0.003	
DnaA|2T(:	Rv0095c|9	[4719,190,	0.55	293.07	0.55	0.061	0.05	0.003	
DnaA|2T(:	Cyp123|19	[8402,570,	0.52	1295.58	0.26	0.021	0.52	0.054	
DnaA|2T(:	CtpD|17R(	[8565,632,	0.89	1698.68	0.2	0.017	0.89	0.224	
DnaA|2T(:	CtpD|18V	[8515,488,	0.8	2808.46	0.37	0.026	0.8	0.109	
DnaA|2T(:	Rv1572c|7	[5628,173,	0.65	609.86	0.65	0.065	0.09	0.005	
DnaA|2T(:	Rv1730c|4	[8182,342,	0.53	2475.55	0.53	0.039	0.49	0.035	
DnaA|2T(:	Rv1730c|4	[8181,339,	0.53	2498.9	0.53	0.039	0.49	0.034	
DnaA|2T(:	Rv1754c|1	[8524,660,	0.68	1064.72	0.16	0.015	0.68	0.109	
DnaA|2T(:	Rv1907c|4	[8550,504,	0.89	2960.01	0.35	0.025	0.89	0.164	
DnaA|2T(:	Rv2016|65	[8527,456,	0.84	3272.17	0.41	0.029	0.84	0.123	
DnaA|2T(:	Rv2262c|3	[7566,111,	0.83	2740.7	0.83	0.087	0.35	0.017	
DnaA|2T(:	Rv2264c|5	[8572,752,	0.82	482.42	0.06	0.008	0.82	0.275	
DnaA|2T(:	Rv2407|63	[7937,157,	0.77	3309.89	0.77	0.07	0.45	0.025	
DnaA|2T(:	Rv2823c|1	[8576,708,	0.92	991.0	0.11	0.012	0.92	0.365	
DnaA|2T(:	Pks15|187	[8573,784,	0.67	166.72	0.02	0.005	0.67	0.265	
DnaA|2T(:	FadE36|30	[8523,626,	0.72	1401.23	0.2	0.017	0.72	0.111	
GyrA|95S(	TrpG|68I(:	[4908,4,33	0.99	1411.01	0.99	0.498	0.15	0.005	
GyrA|95S(	FhaA|246I	[4354,389:	0.83	660.23	0.11	0.003	0.6	0.01	
GyrA|95S(	Rv0039c|2	[4905,22,3	0.96	1335.75	0.96	0.21	0.14	0.005	
GyrA|95S(	Ino1|190R	[6542,8,17	0.98	2946.45	0.98	0.351	0.31	0.012	
GyrA|95S(	Rv0068|21	[8234,783,	0.96	2603.9	0.28	0.018	0.96	0.283	
GyrA|95S(	Rv0095c|9	[4221,261,	0.52	324.89	0.52	0.052	0.06	0.003	
GyrA|95S(	OxcA|224	[4897,6,33	0.98	1396.11	0.98	0.406	0.15	0.005	
GyrA|95S(	FusA2|56:	[8228,869,	0.93	1891.56	0.21	0.015	0.93	0.227	
GyrA|95S(	Rv0194|74	[8226,765,	0.94	2675.94	0.3	0.019	0.94	0.217	
GyrA|95S(	Rv0210|48	[4909,21,3	0.96	1342.14	0.96	0.215	0.14	0.005	
GyrA|95S(	EccE3|216	[8237,895,	0.96	1769.33	0.19	0.014	0.96	0.327	
GyrA|95S(	Rv0368c|2	[4911,20,3	0.96	1347.37	0.96	0.22	0.14	0.005	
GyrA|95S(	LpqM|296	[4910,40,3	0.93	1267.16	0.93	0.154	0.14	0.005	
GyrA|95S(	CtpH|689I	[8219,788,	0.91	2440.08	0.28	0.018	0.91	0.184	
GyrA|95S(	Rv0465c|1	[4921,75,3	0.87	1139.75	0.87	0.11	0.13	0.005	
GyrA|95S(	Rv0576|23	[4920,111,	0.81	1009.38	0.81	0.088	0.13	0.005	
GyrA|95S(	GalTb|173	[6639,54,1	0.93	2820.88	0.93	0.131	0.32	0.012	
GyrA|95S(	MmpL5|94	[8237,913,	0.95	1633.03	0.18	0.013	0.95	0.326	
GyrA|95S(	MmpL5|79	[4905,27,3	0.95	1315.7	0.95	0.189	0.14	0.005	
GyrA|95S(	PurF|476R	[4138,52,4	0.89	847.53	0.89	0.134	0.1	0.003	
GyrA|95S(	Rv0836c|2	[4908,39,3	0.93	1269.93	0.93	0.156	0.14	0.005	

Appendix 26: Epistatic links between polymorphisms in lineage 3.

© University of Pretoria

<table>
<tr><th>Query</th><th>Sbjct</th><th>[AB,aB,Ab</th><th>LD</th><th>Chi2</th><th>A->a|b</th><th>Err1</th><th>B->b|a</th><th>Err2</th><th></th></tr>
<tr><td>OpcA|12A</td><td>PncA|136</td><td>[7182,218</td><td>0.76</td><td>41.77</td><td>0.0</td><td>0.001</td><td>0.76</td><td>0.452</td><td></td></tr>
</table>

Appendix 27: Epistatic links between polymorphisms in lineage 4.

© University of Pretoria

Query	Sbjct	[AB,aB,Ab	LD	Chi2	A->a|b	Err1	B->b|a	Err2	
GyrA|90A	KatG|328\	[9152,229,	0.79	244.26	0.03	0.011	0.79	0.632	
GyrA|90A	SerA1|122	[9150,229,	0.65	200.95	0.03	0.011	0.65	0.408	
GyrA|90A	Rv3693|35	[9152,229,	0.79	244.26	0.03	0.011	0.79	0.632	
Rv0012|25	TrpG|68I(:	[4469,10,2	0.99	3117.47	0.99	0.315	0.33	0.01	
Rv0012|25	Rv0039c|2	[4457,7,23	0.99	3112.95	0.99	0.377	0.33	0.01	
Rv0012|25	Rv0045c|1	[5225,15,1	0.98	4295.5	0.98	0.256	0.46	0.015	
Rv0012|25	Rv0045c|8	[6793,1834	0.89	1782.12	0.21	0.009	0.89	0.121	
Rv0012|25	Rv0064|54	[4461,10,2	0.99	3106.75	0.99	0.315	0.33	0.01	
Rv0012|25	Rv0071|24	[6817,1504	0.95	2960.62	0.33	0.013	0.95	0.16	
Rv0012|25	Rv0071|29	[6851,1982	0.99	1579.2	0.16	0.008	0.99	0.575	
Rv0012|25	Rv0095c|9	[4797,89,2	0.93	3278.78	0.93	0.102	0.37	0.012	
Rv0012|25	Rv0095c|9	[4797,90,2	0.93	3274.95	0.93	0.102	0.37	0.012	
Rv0012|25	Rv0095c|9	[4813,96,2	0.92	3275.98	0.92	0.098	0.37	0.012	
Rv0012|25	Rv0095c|8	[5181,61,1	0.95	4021.58	0.95	0.125	0.44	0.014	
Rv0012|25	Rv0104|40	[4465,10,2	0.99	3112.1	0.99	0.315	0.33	0.01	
Rv0012|25	OxcA|224	[4471,17,2	0.98	3093.24	0.98	0.241	0.33	0.01	
Rv0012|25	Mce1R|17	[4563,177,	0.86	2630.92	0.86	0.07	0.32	0.01	
Rv0012|25	Mce1A|31	[4449,10,2	0.99	3090.73	0.99	0.315	0.33	0.01	
Rv0012|25	Rv0209|16	[4485,7,23	0.99	3150.61	0.99	0.377	0.33	0.01	
Rv0012|25	Rv0210|48	[4454,7,24	0.99	3108.94	0.99	0.377	0.33	0.01	
Rv0012|25	EccA3|6E(	[4422,10,2	0.99	3054.98	0.99	0.315	0.32	0.01	
Rv0012|25	Rv0325|74	[4463,12,2	0.99	3101.73	0.99	0.287	0.33	0.01	
Rv0012|25	Rv0368c|2	[4452,8,24	0.99	3102.42	0.99	0.352	0.33	0.01	
Rv0012|25	Rv0376c|1	[5166,11,1	0.99	4203.52	0.99	0.3	0.45	0.014	
Rv0012|25	Pta|122G(	[4465,9,23	0.99	3115.96	0.99	0.332	0.33	0.01	
Rv0012|25	LpqM|296	[4428,13,2	0.98	3051.41	0.98	0.276	0.32	0.01	
Rv0012|25	Rv0428c|1	[4403,11,2	0.99	3026.23	0.99	0.3	0.32	0.009	
Rv0012|25	Rv0465c|3	[6850,1511	0.99	3092.29	0.33	0.013	0.99	0.498	
Rv0012|25	Rv0465c|1	[4389,6,24	0.99	3027.08	0.99	0.407	0.32	0.009	
Rv0012|25	Rv0493c|1	[4448,8,24	0.99	3097.09	0.99	0.352	0.33	0.01	
Rv0012|25	Rv0576|23	[4354,6,25	0.99	2981.84	0.99	0.407	0.31	0.009	
Rv0012|25	GalK|198H	[4393,5,24	0.99	3036.11	0.99	0.446	0.32	0.009	
Rv0012|25	AtsD|349D	[4452,10,2	0.99	3094.73	0.99	0.315	0.33	0.01	
Rv0012|25	RpoC|109	[6851,1510	0.99	3100.56	0.33	0.013	0.99	0.576	
Rv0012|25	MmpL5|79	[4443,16,2	0.98	3059.76	0.98	0.248	0.33	0.01	
Rv0012|25	SppA|622	[6852,2470	0.96	173.17	0.01	0.002	0.96	0.696	
Rv0012|25	SerA2|242	[4440,9,24	0.99	3082.61	0.99	0.332	0.33	0.01	
Rv0012|25	Rv0797|54	[6482,1673	0.58	1327.79	0.24	0.011	0.58	0.041	
Rv0012|25	PurF|476F	[5187,14,1	0.99	4229.24	0.99	0.265	0.45	0.014	
Rv0012|25	Rv0823c|3	[6849,1992	0.98	1539.25	0.16	0.008	0.98	0.444	
Rv0012|25	Rv0836c|2	[4435,9,24	0.99	3075.98	0.99	0.332	0.33	0.01	
Rv0012|25	Rv0845|21	[4431,16,2	0.98	3043.9	0.98	0.248	0.32	0.01	
Rv0012|25	Rv0862c|7	[4438,7,24	0.99	3087.64	0.99	0.377	0.33	0.01	

Appendix 28: Epistatic links between polymorphisms in lineage 4.1.

© University of Pretoria

Query	Sbjct	[AB,aB,Ab	LD	Chi2	A->a|b	Err1	B->b|a	Err2	
Rv1572c|7	Rv3655c|1	[5794,3565	0.64	21.64	0.0	0.0	0.64	0.333	

Appendix 29: Epistatic links between polymorphisms in lineage 4 (Ghana).

© University of Pretoria

Query	Sbjct	[AB,aB,Ab	LD	Chi2	A->a|b	Err1	B->b|a	Err2	
RodA|461	HddA|192	[8861,11,4	0.87	1263.06	0.87	0.283	0.15	0.018	
RodA|461	Rv0210|15	[9022,33,2	0.63	1058.51	0.63	0.14	0.17	0.025	
RodA|461	Pnp|190R	[9061,37,2	0.59	1065.6	0.59	0.127	0.18	0.028	
RodA|461	Rv0612|11	[8936,22,3	0.75	1138.45	0.75	0.186	0.16	0.02	
RodA|461	Rv0686|22	[9076,36,2	0.6	1164.19	0.6	0.13	0.2	0.03	
RodA|461	RpsH|80V	[8994,30,3	0.67	1063.63	0.67	0.15	0.16	0.023	
RodA|461	Rv0744c|4	[8830,13,4	0.85	1131.26	0.85	0.257	0.14	0.017	
RodA|461	GgtA|211	[8944,40,3	0.55	664.03	0.55	0.119	0.12	0.018	
RodA|461	PurL|231V	[8845,22,4	0.75	925.07	0.75	0.186	0.13	0.016	
RodA|461	Rv1268c|2	[8947,26,3	0.71	1048.98	0.71	0.166	0.15	0.02	
RodA|461	Zwf2|300	[9032,30,2	0.67	1196.38	0.67	0.15	0.18	0.026	
RodA|461	UvrB|310	[9061,41,2	0.55	930.08	0.55	0.117	0.17	0.027	
RodA|461	Pks8|234	[8909,28,3	0.69	900.2	0.69	0.158	0.13	0.018	
RodA|461	PyrG|576	[8963,38,3	0.58	751.98	0.58	0.125	0.13	0.02	
RodA|461	Rv1823|22	[8901,19,3	0.78	1132.71	0.78	0.204	0.15	0.019	
RodA|461	Gnd1|90A	[8569,13,7	0.85	736.0	0.85	0.257	0.09	0.01	
RodA|461	Rv1883c|7	[8615,15,6	0.82	746.47	0.82	0.236	0.09	0.011	
RodA|461	Dgt|108H	[9004,32,2	0.64	1031.84	0.64	0.143	0.16	0.023	
RodA|461	SubI|258T	[8959,31,3	0.65	934.74	0.65	0.146	0.15	0.02	
RodA|461	ValS|697A	[8988,26,3	0.71	1172.65	0.71	0.166	0.17	0.023	
RodA|461	Rv2474c|4	[9023,39,2	0.57	872.67	0.57	0.122	0.16	0.024	
RodA|461	Gdh|118G	[9037,25,2	0.72	1398.23	0.72	0.171	0.2	0.027	
RodA|461	AspS|426	[9002,33,2	0.63	994.36	0.63	0.14	0.16	0.023	
RodA|461	FadD9|22	[9062,28,2	0.69	1401.95	0.69	0.158	0.21	0.03	
RodA|461	Rv2823c|1	[8597,13,6	0.85	765.75	0.85	0.257	0.09	0.011	
RodA|461	Rv2954c|2	[9070,33,2	0.64	1248.86	0.64	0.14	0.2	0.03	
RodA|461	LigA|528V	[9020,35,2	0.61	987.49	0.61	0.133	0.17	0.024	
RodA|461	Rv3060c|3	[8786,24,5	0.73	777.01	0.73	0.176	0.11	0.014	
RodA|461	Rv3201c|3	[9069,41,2	0.55	958.75	0.55	0.117	0.18	0.028	
RodA|461	Rv3272|92	[8945,31,3	0.65	900.35	0.65	0.146	0.14	0.02	
RodA|461	Rv3401|97	[9057,45,2	0.51	790.11	0.51	0.107	0.16	0.026	
RodA|461	IspD|81S(	[8968,23,3	0.74	1204.12	0.74	0.181	0.17	0.022	
RodA|461	Rv3729|70	[8927,29,3	0.67	913.41	0.67	0.154	0.14	0.019	
Rv0044c|1	HddA|192	[8832,40,3	0.75	1768.32	0.75	0.138	0.24	0.025	
Rv0044c|1	Rv0210|15	[8995,60,2	0.64	2019.92	0.64	0.104	0.33	0.038	
Rv0044c|1	Pnp|190R	[9025,73,1	0.57	1821.38	0.57	0.088	0.33	0.042	
Rv0044c|1	Rv0612|11	[8904,54,3	0.67	1707.5	0.67	0.112	0.26	0.029	
Rv0044c|1	Rv0686|22	[9044,68,1	0.6	2121.54	0.6	0.094	0.37	0.046	
Rv0044c|1	RpsH|80V	[8955,69,2	0.59	1545.68	0.59	0.093	0.27	0.032	
Rv0044c|1	Rv0744c|4	[8811,32,4	0.8	1884.53	0.8	0.159	0.24	0.024	
Rv0044c|1	GgtA|211	[8927,57,2	0.66	1732.27	0.66	0.108	0.27	0.03	
Rv0044c|1	PurL|231V	[8821,46,3	0.72	1587.37	0.72	0.126	0.23	0.024	
Rv0044c|1	Rv1268c|2	[8928,45,2	0.73	2065.36	0.73	0.127	0.3	0.032	

Appendix 30: Epistatic links between polymorphisms in lineage 4 (Haarlem).

© University of Pretoria

Query	Sbjct	[AB,aB,Ab	LD	Chi2	A->a|b	Err1	B->b|a	Err2	
Rv0080|12	Nrp|2468	[9369,9,4,	0.69	3243.45	0.49	0.235	0.69	0.415	
Rv0080|12	BglS|106C	[9371,9,2,	0.81	3835.62	0.49	0.235	0.81	0.639	
Rv0080|12	Rv0196|15	[9369,9,4,	0.69	3243.45	0.49	0.235	0.69	0.415	
Rv0080|12	Rv0435c|1	[9369,9,4,	0.69	3243.45	0.49	0.235	0.69	0.415	
Rv0080|12	Rv1063c|1	[9368,9,5,	0.64	3010.81	0.49	0.235	0.64	0.358	
Rv0080|12	EmbR|313	[9369,9,4,	0.69	3243.45	0.49	0.235	0.69	0.415	
Rv0080|12	Rv1279|32	[9367,9,6,	0.59	2809.19	0.49	0.235	0.59	0.316	
Rv0080|12	Mec|52M	[9367,9,6,	0.59	2809.19	0.49	0.235	0.59	0.316	
Rv0080|12	Rv1514c|1	[9367,9,6,	0.59	2809.19	0.49	0.235	0.59	0.316	
Rv0080|12	Rv1514c|1	[9369,9,4,	0.69	3243.45	0.49	0.235	0.69	0.415	
Rv0080|12	PlsB1|554	[9369,9,4,	0.69	3243.45	0.49	0.235	0.69	0.415	
Rv0080|12	IlvA|143D	[9369,9,4,	0.69	3243.45	0.49	0.235	0.69	0.415	
Rv0080|12	Pks12|403	[9367,9,6,	0.59	2809.19	0.49	0.235	0.59	0.316	
Rv0080|12	TB16.3|73	[9369,9,4,	0.69	3243.45	0.49	0.235	0.69	0.415	
Rv0080|12	LppO|51M	[9369,9,4,	0.69	3243.45	0.49	0.235	0.69	0.415	
Rv0080|12	EchA14|14	[9369,9,4,	0.69	3243.45	0.49	0.235	0.69	0.415	
Rv0080|12	Rv2575|67	[9369,9,4,	0.69	3243.45	0.49	0.235	0.69	0.415	
Rv0080|12	Rv2752c|9	[9371,9,2,	0.81	3835.62	0.49	0.235	0.81	0.639	
Rv0080|12	HsdM|102	[9369,9,4,	0.69	3243.45	0.49	0.235	0.69	0.415	
Rv0080|12	Rv3113|12	[9368,9,5,	0.64	3010.81	0.49	0.235	0.64	0.358	
Rv0080|12	PmmB|43	[9367,9,6,	0.59	2809.19	0.49	0.235	0.59	0.316	
Rv0080|12	Rv3479|93	[9369,9,4,	0.69	3243.45	0.49	0.235	0.69	0.415	
Rv0080|12	Ltp4|117A	[9369,9,4,	0.69	3243.45	0.49	0.235	0.69	0.415	
Rv0080|12	Rv3729|29	[9369,9,4,	0.69	3243.45	0.49	0.235	0.69	0.415	
Rv0080|12	EccE2|204	[9368,9,5,	0.64	3010.81	0.49	0.235	0.64	0.358	
Nrp|2468	Rv1063c|1	[9377,0,1,	1.0	8719.28	0.0	0.0	0.92	0.963	
Nrp|2468	Rv1148c|3	[9329,6,49	0.53	622.73	0.53	0.299	0.12	0.05	
Nrp|2468	Rv1279|32	[9376,0,2,	1.0	8137.13	0.0	0.0	0.86	0.658	
Nrp|2468	Mec|52M	[9376,0,2,	1.0	8137.13	0.0	0.0	0.86	0.658	
Nrp|2468	Rv1514c|1	[9376,0,2,	1.0	8137.13	0.0	0.0	0.86	0.658	
Nrp|2468	Pks12|403	[9376,0,2,	1.0	8137.13	0.0	0.0	0.86	0.658	
Nrp|2468	Rv2082|24	[9268,6,11	0.53	292.74	0.53	0.299	0.05	0.023	
Nrp|2468	Rv2277c|6	[6194,4,31	0.53	7.2	0.53	0.415	0.0	0.0	
Nrp|2468	Rv3113|12	[9377,0,1,	1.0	8719.28	0.0	0.0	0.92	0.963	
Nrp|2468	PmmB|43	[9376,0,2,	1.0	8137.13	0.0	0.0	0.86	0.658	
Nrp|2468	EccE2|204	[9377,0,1,	1.0	8719.28	0.0	0.0	0.92	0.963	
BglS|106C	EsxL|31I(3	[8794,5,58	0.51	43.39	0.51	0.33	0.0	0.003	
BglS|106C	EsxL|36L(3	[8822,5,55	0.51	45.96	0.51	0.33	0.0	0.004	
BglS|106C	EsxL|38A(	[8780,5,60	0.51	42.19	0.51	0.33	0.0	0.003	
BglS|106C	Rv2082|24	[9270,4,11	0.63	348.42	0.63	0.398	0.05	0.023	
BglS|106C	Rv2277c|6	[6195,3,31	0.58	7.36	0.58	0.492	0.0	0.0	
Rv0196|15	Rv1063c|1	[9377,0,1,	1.0	8719.28	0.0	0.0	0.92	0.963	

Appendix 31: Epistatic links between polymorphisms in lineage 4 (Uganda).

© University of Pretoria

Query	Sbjct	[AB,aB,Ab,ab]	LD	Chi2	A->a|b	Err1
FadD34|23L(1)	Rv0163|108E(3)	[9306,3,0,82]	0.99	9056.63	0.96	0.566
FadD34|23L(1)	Rv0188|129G(3)	[9306,27,0,58]	1.0	6389.43	0.68	0.158
FadD34|23L(1)	GroEL2|434A(1)	[9306,29,0,56]	1.0	6167.79	0.65	0.15
FadD34|23L(1)	CcsA|26V(3)	[9306,27,0,58]	1.0	6389.43	0.68	0.158
FadD34|23L(1)	MmaA4|232P(-1)	[9306,27,0,58]	1.0	6389.43	0.68	0.158
FadD34|23L(1)	RpoB|450S(1)	[7119,28,2187,57]	0.56	87.87	0.56	0.154
FadD34|23L(1)	RpsL|88K(1)	[9074,26,232,59]	0.68	1256.18	0.68	0.162
FadD34|23L(1)	Rv0750|26D(3)	[8912,27,394,58]	0.66	753.08	0.66	0.158
FadD34|23L(1)	Rv0794c|326A(-1)	[9306,27,0,58]	1.0	6389.43	0.68	0.158
FadD34|23L(1)	RpfA|196V(-1)	[9087,39,219,46]	0.52	823.05	0.52	0.117
FadD34|23L(1)	Rv1047|8D(3)	[8797,31,509,54]	0.61	503.81	0.61	0.142
FadD34|23L(1)	FadD11|470P(3)	[9306,27,0,58]	1.0	6389.43	0.68	0.158
FadD34|23L(1)	EsxN|22A(1)	[9238,32,68,53]	0.61	2514.8	0.61	0.139
FadD34|23L(1)	EsxN|23S(1)	[9241,31,65,54]	0.63	2657.78	0.63	0.142
FadD34|23L(1)	Rv1835c|420Y(-2)	[9306,28,0,57]	0.99	6278.6	0.66	0.154
FadD34|23L(1)	Rv2038c|54L(-2)	[9306,26,0,59]	0.99	6500.29	0.69	0.162
FadD34|23L(1)	CobL|298P(-1)	[9306,20,0,65]	0.99	7165.95	0.76	0.195
FadD34|23L(1)	LppB|175F(3)	[9134,7,172,78]	0.91	2628.07	0.91	0.361
FadD34|23L(1)	LppB|176N(1)	[9137,11,169,74]	0.86	2428.15	0.86	0.28
FadD34|23L(1)	LppB|177L(3)	[9130,10,176,75]	0.87	2414.0	0.87	0.296
FadD34|23L(1)	LppB|178G(1)	[9132,11,174,74]	0.86	2377.5	0.86	0.28
FadD34|23L(1)	Rv2694c|39I(-1)	[9306,27,0,58]	1.0	6389.43	0.68	0.158
FadD34|23L(1)	GlnA4|440A(-2)	[9306,26,0,59]	0.99	6500.29	0.69	0.162
FadD34|23L(1)	Rv2886c|184P(-2)	[9306,26,0,59]	0.99	6500.29	0.69	0.162
FadD34|23L(1)	Rv2913c|493P(-2)	[9306,21,0,64]	0.99	7054.95	0.75	0.188
FadD34|23L(1)	Rv3115|327K(3)	[7054,19,2252,66]	0.7	129.42	0.7	0.201
FadD34|23L(1)	NagA|3V(1)	[9306,26,0,59]	0.99	6500.29	0.69	0.162
FadD34|23L(1)	Rv3395A|1M(1)	[9306,27,0,58]	1.0	6389.43	0.68	0.158
FadD34|23L(1)	Cyp137|92A(-1)	[9306,20,0,65]	0.99	7165.95	0.76	0.195
Rv0163|108E(3)	Rv0188|129G(3)	[9309,24,0,58]	1.0	6625.33	0.7	0.171
Rv0163|108E(3)	Rv0378|22S(3)	[9306,0,3,82]	0.99	9056.63	0.0	0.0
Rv0163|108E(3)	GroEL2|434A(1)	[9309,26,0,56]	0.99	6395.5	0.68	0.161
Rv0163|108E(3)	CcsA|26V(3)	[9309,24,0,58]	1.0	6625.33	0.7	0.171
Rv0163|108E(3)	MmaA4|232P(-1)	[9309,24,0,58]	1.0	6625.33	0.7	0.171
Rv0163|108E(3)	RpoB|450S(1)	[7119,28,2190,54]	0.55	80.08	0.55	0.152
Rv0163|108E(3)	RpsL|88K(1)	[9076,24,233,58]	0.69	1260.17	0.69	0.171
Rv0163|108E(3)	Rv0750|26D(3)	[8915,24,394,58]	0.69	784.57	0.69	0.171
Rv0163|108E(3)	Rv0794c|326A(-1)	[9309,24,0,58]	1.0	6625.33	0.7	0.171
Rv0163|108E(3)	RpfA|196V(-1)	[9090,36,219,46]	0.54	856.2	0.54	0.124
Rv0163|108E(3)	Rv1047|8D(3)	[8800,28,509,54]	0.63	525.93	0.63	0.152
Rv0163|108E(3)	FadD11|470P(3)	[9309,24,0,58]	1.0	6625.33	0.7	0.171
Rv0163|108E(3)	EsxN|22A(1)	[9241,29,68,53]	0.64	2609.84	0.64	0.148
Rv0163|108E(3)	EsxN|23S(1)	[9244,28,65,54]	0.65	2758.09	0.65	0.152
Rv0163|108E(3)	Rv1835c|420Y(-2)	[9309,25,0,57]	0.99	6510.4	0.69	0.166

Appendix 32: Epistatic links between polymorphisms in lineage 4 (Ural).

© University of Pretoria

Query	Sbjct	[AB,aB,Ab,ab]	LD	Chi2	A->a|b	Err1
HspR|124P(3)	Gnd1|90A(-1)	[8579,3,800,9]	0.72	67.26	0.72	0.499
HspR|124P(3)	Rv1883c|75E(-2)	[8630,5,749,7]	0.54	41.04	0.54	0.341
HspR|124P(3)	Rv2823c|102N(-1)	[8606,4,773,8]	0.63	53.65	0.63	0.408
MmpS1|124E(-1)	MmpS1|120E(-2)	[9156,18,174,43]	0.69	1264.42	0.69	0.197
MmpS1|124E(-1)	Rv0963c|51A(-1)	[9259,18,71,43]	0.7	2457.31	0.7	0.197
MmpS1|124E(-1)	Gnd1|90A(-1)	[8566,16,764,45]	0.71	331.09	0.71	0.214
MmpS1|124E(-1)	Rv1883c|75E(-2)	[8617,18,713,43]	0.67	323.4	0.67	0.197
MmpS1|124E(-1)	Rv1928c|134M(-2)	[8164,12,1166,49]	0.77	247.54	0.77	0.258
MmpS1|124E(-1)	Rv2823c|102N(-1)	[8590,20,740,41]	0.64	279.32	0.64	0.182
MmpS1|124E(-1)	Rv3093c|203A(-2)	[9253,19,77,42]	0.68	2241.64	0.68	0.189
MmpS1|120E(-2)	Rv0963c|51A(-1)	[9120,157,54,60]	0.51	1294.52	0.26	0.041
MmpS1|120E(-2)	Rv1928c|134M(-2)	[8137,39,1037,17	0.79	941.33	0.79	0.144
Rv0963c|51A(-1)	Rv1928c|134M(-2)	[8152,24,1125,90]	0.75	446.4	0.75	0.18
Rv1046c|171E(-2)	Rv1572c|7V(-1)	[5790,11,3561,29]	0.55	19.98	0.55	0.256
Rv1046c|171E(-2)	Gnd1|90A(-1)	[8575,7,776,33]	0.8	278.55	0.8	0.343
Rv1046c|171E(-2)	Rv1883c|75E(-2)	[8628,7,723,33]	0.8	300.8	0.8	0.343
Rv1046c|171E(-2)	Rv2823c|102N(-1)	[8602,8,749,32]	0.78	270.72	0.78	0.315
Rv1572c|7V(-1)	PlcD|82A(-1)	[5785,3551,16,39]	0.52	25.02	0.0	0.001
Rv1572c|7V(-1)	Gnd1|90A(-1)	[5632,2950,169,6	0.66	626.55	0.1	0.005
Rv1572c|7V(-1)	Rv1883c|75E(-2)	[5661,2974,140,6	0.7	651.38	0.09	0.005
Rv1572c|7V(-1)	Pks12|179S(-1)	[5774,3492,27,98]	0.65	86.57	0.01	0.001
Rv1572c|7V(-1)	Rv2823c|102N(-1)	[5645,2965,156,6	0.67	630.21	0.09	0.005
Rv1572c|7V(-1)	FadE22|94S(-1)	[5765,3479,36,11	0.6	87.9	0.01	0.001
VapC12|120G(-2)	Rv1928c|134M(-2)	[8171,5,1167,48]	0.89	285.15	0.89	0.425
Gnd1|90A(-1)	Dop|45P(-2)	[8581,781,1,28]	0.96	285.74	0.03	0.006
Gnd1|90A(-1)	AftC|431G(3)	[8573,732,9,77]	0.88	721.92	0.08	0.01
Gnd1|90A(-1)	Rv2823c|97A(-2)	[8569,783,13,26]	0.63	167.64	0.02	0.005
Rv1883c|75E(-2)	Dop|45P(-2)	[8633,729,2,27]	0.92	284.29	0.03	0.006
Rv1883c|75E(-2)	AftC|431G(3)	[8622,683,13,73]	0.83	692.2	0.08	0.011
Rv1883c|75E(-2)	Rv2823c|97A(-2)	[8620,732,15,24]	0.58	151.36	0.02	0.006
Rv1928c|134M(-2)	Rv2488c|709G(-1)	[8162,1191,14,24]	0.57	85.42	0.01	0.003
Rv1928c|134M(-2)	Rv3093c|203A(-2)	[8141,1131,35,84]	0.66	355.62	0.05	0.006
Dop|45P(-2)	Rv2823c|102N(-1)	[8606,4,756,25]	0.84	231.46	0.84	0.464
LppP|20L(-1)	LppP|15A(-1)	[8105,1179,11,96]	0.88	534.79	0.06	0.006
AftC|431G(3)	Rv2823c|102N(-1)	[8599,11,706,75]	0.86	708.49	0.86	0.281
Rv2823c|102N(-1)	Rv2823c|97A(-2)	[8596,756,14,25]	0.6	159.84	0.02	0.005

Appendix 33: Epistatic links between polymorphisms in lineage 4 (X-type).

© University of Pretoria

<table>
<tr><td>Query</td><td>Sbjct</td><td>[AB,aB,Ab,ab]</td><td>LD</td><td>Chi2</td><td>A->a|b</td><td>Err1</td></tr>
<tr><td>Rv0044c|24H(-2)</td><td>Rv0095c|131S(-2)</td><td>[5814,2,3564,11]</td><td>0.75</td><td>11.96</td><td>0.75</td><td>0.65</td></tr>
<tr><td>Rv0044c|24H(-2)</td><td>MmpL12|381S(-1)</td><td>[4924,12,4454,1]</td><td>0.83</td><td>8.24</td><td>0.75</td><td>0.08</td></tr>
<tr><td>Rv0095c|131S(-2)</td><td>FadE3|357R(-2)</td><td>[5815,3565,1,10]</td><td>0.85</td><td>13.04</td><td>0.0</td><td>0.0</td></tr>
<tr><td>Rv0095c|131S(-2)</td><td>MmpL12|381S(-1)</td><td>[3939,516,1877,3059]</td><td>0.69</td><td>2521.94</td><td>0.69</td><td>0.037</td></tr>
<tr><td>Rv0095c|131S(-2)</td><td>Rv3798|99V(3)</td><td>[5814,3191,2,384]</td><td>0.99</td><td>643.97</td><td>0.06</td><td>0.003</td></tr>
<tr><td>EsxL|31I(3)</td><td>EsxL|48A(1)</td><td>[8783,480,16,112]</td><td>0.86</td><td>1448.46</td><td>0.17</td><td>0.018</td></tr>
<tr><td>EsxL|36L(3)</td><td>EsxL|48A(1)</td><td>[8805,458,22,106]</td><td>0.81</td><td>1356.12</td><td>0.17</td><td>0.019</td></tr>
<tr><td>EsxL|38A(3)</td><td>EsxL|48A(1)</td><td>[8771,492,14,114]</td><td>0.88</td><td>1467.03</td><td>0.17</td><td>0.018</td></tr>
<tr><td>MmpL12|381S(-1)</td><td>CobL|134G(-1)</td><td>[4922,14,4453,2]</td><td>0.73</td><td>7.84</td><td>0.66</td><td>0.094</td></tr>
<tr><td>MmpL12|381S(-1)</td><td>Pks15|491R(-2)</td><td>[4453,4289,2,647]</td><td>0.99</td><td>621.09</td><td>0.06</td><td>0.002</td></tr>
<tr><td>MmpL12|381S(-1)</td><td>EsxW|2T(-1)</td><td>[4356,2422,99,2514]</td><td>0.92</td><td>2766.46</td><td>0.32</td><td>0.01</td></tr>
<tr><td>MmpL12|381S(-1)</td><td>Rv3798|99V(3)</td><td>[4553,383,4452,3]</td><td>0.98</td><td>351.51</td><td>0.88</td><td>0.01</td></tr>
<tr><td>Rv1572c|7V(-1)</td><td>EsxW|2T(-1)</td><td>[5161,1617,640,1973]</td><td>0.6</td><td>2130.61</td><td>0.37</td><td>0.014</td></tr>
<tr><td>LppP|23I(-1)</td><td>LppP|17L(-2)</td><td>[8889,490,2,10]</td><td>0.82</td><td>145.05</td><td>0.01</td><td>0.006</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

Appendix 34: Epistatic links between polymorphisms in lineage 5.

© University of Pretoria

Appendix 35: The complete set of virulence factor genes found in each lineage of the *Mycobacterium tuberculosis* complex.

Lineage	Gene	Annotation	Virulence typ
1	None		
1.2	caeA dppA mce1A mce1C mce1D mce1F mce1R mce2A mce2B mce2F mce3A pstA1 yrbE1B yrbE2B	Probable carboxylesterase CaeA Probable periplasmic dipeptide-binding lipoprotein DppA Mce-family protein Mce1A Mce-family protein Mce1C Mce-family protein Mce1D Mce-family protein Mce1F Probable transcriptional regulatory protein Mce1R (probably GntR-family) Mce-family protein Mce2A Mce-family protein Mce2B Mce-family protein Mce2F Mce-family protein Mce3A Probable phosphate-transport integral membrane ABC transporter PstA1 Conserved integral membrane protein YrbE1B Conserved hypothetical integral membrane protein YrbE2B	Cell wall prot
	cyp125	Probable cytochrome P450 125 Cyp125	Cholesterol ca
	aceAa fadD28 mmpL7	Probable isocitrate lyase AceAa [first part] (isocitrase) (isocitratase) (Icl)	Lipid and metabolism

© University of Pretoria

	pks15 pks5 pks7 plcA plcC	Fatty-acid-AMP ligase FadD28 (fatty-acid-AMP synthetase) (fatty-acid-AMP synthase) Conserved transmembrane transport protein MmpL7 Probable polyketide synthase Pks15 Probable polyketide synthase Pks5 Probable polyketide synthase Pks7 Membrane-associated phospholipase C 1 PlcA (MTP40 antigen) Probable phospholipase C 3 PlcC	
	irtA mbtB	Iron-regulated transporter IrtA Phenyloxazoline synthase MbtB (phenyloxazoline synthetase)	Metal transpo
	acg sapM	Conserved protein Acg Acid phosphatase (acid phosphomonoesterase) (phosphomonoesterase) (glycerophosphatase)	Other
	pepD	Probable serine protease PepD (serine proteinase) (Mtb32B)	Proteases
	katG nuoG ponA2	Catalase-peroxidase-peroxynitritase T KatG	Proteins antimicrobial the macrophag

© University of Pretoria

		Probable NADH dehydrogenase I (chain G) NuoG (NADH-ubiquinone oxidoreductase chain G) Probable bifunctional membrane-associated penicillin-binding protein 1A/1B PonA2 (murein polymerase) [includes: penicillin-insensitive transglycosylase (peptidoglycan TGASE) + penicillin-sensitive transpeptidase (DD-transpeptidase)]	
4(Ghana)	pknG	Serine/threonine-protein kinase PknG (protein kinase G) (STPK G)	Serine-threoni proteinkinases
4(Haarlem)	sigG	Probable alternative RNA polymerase sigma factor SigG (RNA polymerase ECF type sigma factor)	Sigma factors
4(LAM)	mazF3 vapB22 vapB47 vapC13 vapC37 vapC38 vapC40 vapC46 vapC47 vapC6	Toxin MazF3 Possible antitoxin VapB22 Possible antitoxin VapB47 Possible toxin VapC13 Possible toxin VapC37 Contains PIN domain Possible toxin VapC38 Contains PIN domain Possible toxin VapC40 Contains PIN domain	TA systems

© University of Pretoria

		Possible toxin VapC46 Contains PIN domain Possible toxin VapC47 Contains PIN domain Possible toxin VapC6	
4(S-type)	dosT mprB phoR	Two component sensor histidine kinase DosT Two component sensor kinase MprB Possible two component system response sensor kinase membrane associated PhoR	Two-compone
4(Tur)	eccB1 eccD1 espA secA2	ESX conserved component EccB1 ESX-1 type VII secretion system protein Possible membrane protein ESX conserved component EccD1 ESX-1 type VII secretion system protein Probable transmembrane protein ESX-1 secretion-associated protein A, EspA Possible preprotein translocase ATPase SecA2	Type VII secr
2	mce3B	Mce-family protein Mce3B	Cell wall prot
	pks15	Probable polyketide synthase Pks15	Lipid and metabolism
	irtB	Iron-regulated transporter IrtB	Metal transpo

© University of Pretoria

	sodC	Periplasmic superoxide dismutase [Cu-Zn] SodC	Proteins antimicrobial the macrophag
	mazF2 vapC12 vapC46	Toxin MazF2 Possible toxin VapC12 Possible toxin VapC46 Contains PIN domain	TA systems
	eccCb1	ESX conserved component EccCb1 ESX-1 type VII secretion system protein	Type VII secr
3	lprL mce1R mce3A pstA1	Possible Mce-family lipoprotein LprL (Mce-family lipoprotein Mce2E) Probable transcriptional regulatory protein Mce1R (probably GntR-family) Mce-family protein Mce3A Probable phosphate-transport integral membrane ABC transporter PstA1	Cell wall prot
	ltp2	Probable lipid transfer protein or keto acyl-CoA thiolase Ltp2	Cholesterol ca
	pks15 pks7	Probable polyketide synthase Pks15 Probable polyketide synthase Pks7	Lipid and metabolism
	ctpV mbtB	Probable metal cation transporter P-type ATPase CtpV Phenyloxazoline synthase MbtB (phenyloxazoline synthetase)	Metal transpo

© University of Pretoria

	pepD rip	Probable serine protease PepD (serine proteinase) (Mtb32B) Membrane bound metalloprotease	Proteases
	katG	Catalase-peroxidase-peroxynitritase T KatG	Proteins antimicrobial the macrophag
	mazF3 vapB47 vapC37 vapC38	Toxin MazF3 Possible antitoxin VapB47 Possible toxin VapC37 Contains PIN domain Possible toxin VapC38 Contains PIN domain	TA systems
	dosT mprB	Two component sensor histidine kinase DosT Two component sensor kinase MprB	Two-compone
	secA2	Possible preprotein translocase ATPase SecA2	Type VII secr
4	None		
4.1	dppA mce1A mce1D mce1R mce2F mce3A mce4C yrbE2B	Probable periplasmic dipeptide-binding lipoprotein DppA Mce-family protein Mce1A Mce-family protein Mce1D Probable transcriptional regulatory protein Mce1R (probably GntR-family) Mce-family protein Mce2F Mce-family protein Mce3A Mce-family protein Mce4C	Cell wall prot

© University of Pretoria

		Conserved hypothetical integral membrane protein YrbE2B	
	ltp2	Probable lipid transfer protein or keto acyl-CoA thiolase Ltp2	Cholesterol ca
	fadD28 pks15 pks5 pks7 plcA plcC	Fatty-acid-AMP ligase FadD28 (fatty-acid-AMP synthetase) (fatty-acid-AMP synthase) Probable polyketide synthase Pks15 Probable polyketide synthase Pks5 Probable polyketide synthase Pks7 Membrane-associated phospholipase C 1 PlcA (MTP40 antigen)	Lipid and metabolism

© University of Pretoria

		Probable phospholipase C 3 PlcC	
	ctpV mbtB	Probable metal cation transporter P-type ATPase CtpV Phenyloxazoline synthase MbtB (phenyloxazoline synthetase)	Metal transpo
	rip	Membrane bound metalloprotease	Proteases
	katG ponA2	Catalase-peroxidase-peroxynitritase T KatG Probable bifunctional membrane-associated penicillin-binding protein 1A/1B PonA2 (murein polymerase) [includes: penicillin-insensitive transglycosylase (peptidoglycan TGASE) + penicillin-sensitive transpeptidase (DD-transpeptidase)]	Proteins antimicrobial the macropha
	sigG	Probable alternative RNA polymerase sigma factor SigG (RNA polymerase ECF type sigma factor)	Sigma factors
	mazF3 vapB47 vapC37 vapC38 vapC40	Toxin MazF3 Possible antitoxin VapB47 Possible toxin VapC37 Contains PIN domain Possible toxin VapC38 Contains PIN domain	TA systems

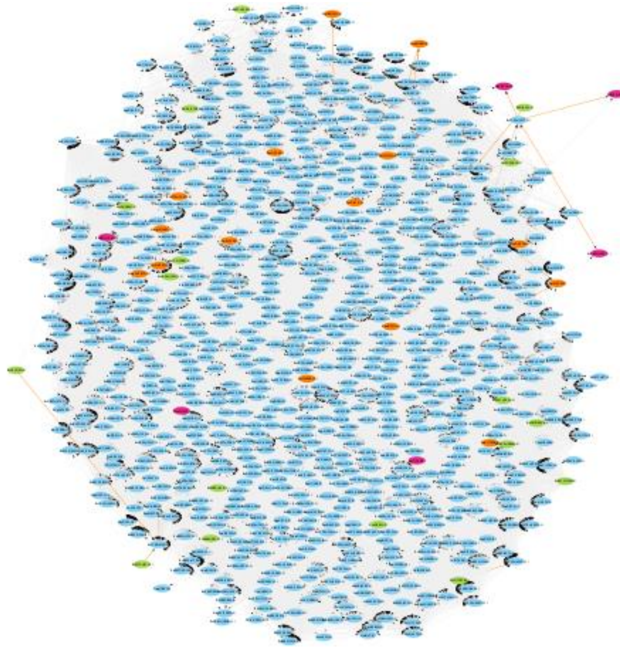
© University of Pretoria

		Possible toxin VapC40 Contains PIN domain	
	dosT mprB	Two component sensor histidine kinase DosT Two component sensor kinase MprB	Two-compone
	eccCb1 secA2	ESX conserved component EccCb1 ESX-1 type VII secretion system protein Possible preprotein translocase ATPase SecA2	Type VII secr
Lineage_4_Cameroon	None		
Lineage_4_Ghana	None		
Lineage_4_Haarlem	pks7	Probable polyketide synthase Pks7	Lipid and metabolism
	sigG	Probable alternative RNA polymerase sigma factor SigG (RNA polymerase ECF type sigma factor)	Sigma factors
	vapB16 vapB43 vapC31	Possible antitoxin VapB16 Possible antitoxin VapB43 Possible toxin VapC31 Contains PIN domain	TA systems
Lineage_4_LAM	katG	Catalase-peroxidase-peroxynitritase T KatG	Proteins antimicrobial the macropha
Lineage_4_S-type	katG	Catalase-peroxidase-peroxynitritase T KatG	Proteins antimicrobial the macropha
	mycP1	Membrane-anchored mycosin MycP1 (serine protease) (subtilisin-like	Type VII secr

© University of Pretoria

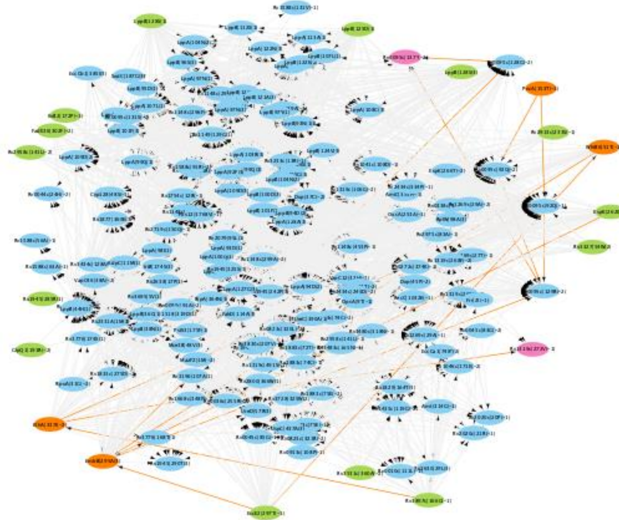
		protease) (subtilase-like) (mycosin-1)	
Lineage_4_Ural	mmaA4	Methoxy mycolic acid synthase 4 MmaA4 (methyl mycolic acid synthase 4) (MMA4) (hydroxy mycolic acid synthase)	Mycolic acid
Lineage_4_X-type	hspR	Probable heat shock protein transcriptional repressor HspR (MerR family)	Other
	vapC12	Possible toxin VapC12	TA systems
Lineage_5	pks15	Probable polyketide synthase Pks15	Lipid and metabolism
Lineage 6	katG	Catalase-peroxidase-peroxynitritase T KatG	Proteins antimicrobial the macropha

© University of Pretoria



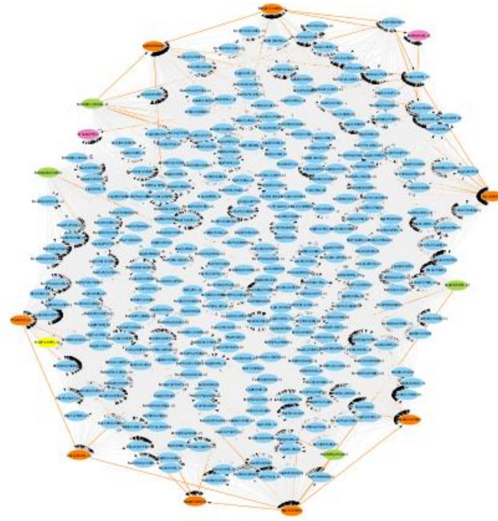
Appendix 36: Epistatic interactions showing various evolutionary paths in lineage 1.2. The green nodes represent a pre-requisite mutation, the blue nodes an intermediate mutation, the orange nodes represent DR mutations, and the pink nodes and final/compensatory mutation. Epistatic links to and from known DR mutations are highlighted by orange edges. Nodes are labelled by the respective gene name and codon position number following the horizontal line.

© University of Pretoria



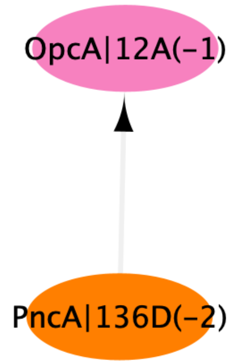
Appendix 37: Epistatic interactions showing various evolutionary paths lineage 2. The green nodes represent a pre-requisite mutation, the blue nodes an intermediate mutation, the orange nodes represent DR mutations, and the pink nodes and final/compensatory mutation. Epistatic links to and from known DR mutations are highlighted by orange edges. Nodes are labelled by the respective gene name and codon position number following the horizontal line.

© University of Pretoria



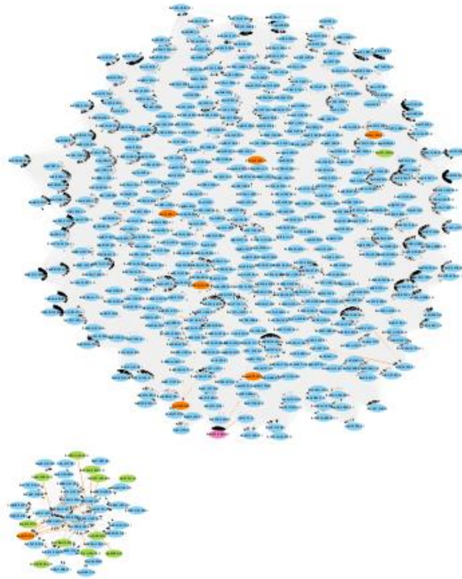
Appendix 38: Epistatic interactions showing various evolutionary paths towards drug resistance (DR) in lineage 3. The green nodes represent a pre-requisite mutation, the blue nodes an intermediate mutation, the orange nodes represent DR mutations, and the pink nodes and final/compensatory mutation. Epistatic links to and from known DR mutations are highlighted by orange edges. Nodes are labelled by the respective gene name and codon position number following the horizontal line.

© University of Pretoria



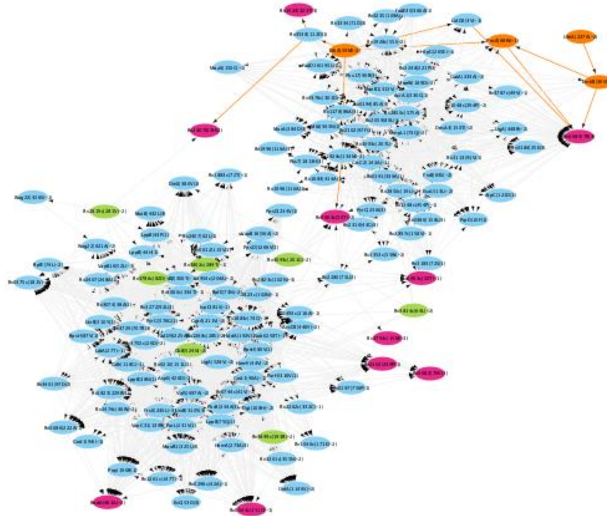
Appendix 39: Epistatic interactions showing the mutational path in lineage 4. The orange node represents a drug resistance (DR) mutation and the pink node a final/compensatory mutation. Epistatic links to and from known DR mutations are highlighted by orange edges. Nodes are labelled by the respective gene name and codon position number following the horizontal line.

© University of Pretoria



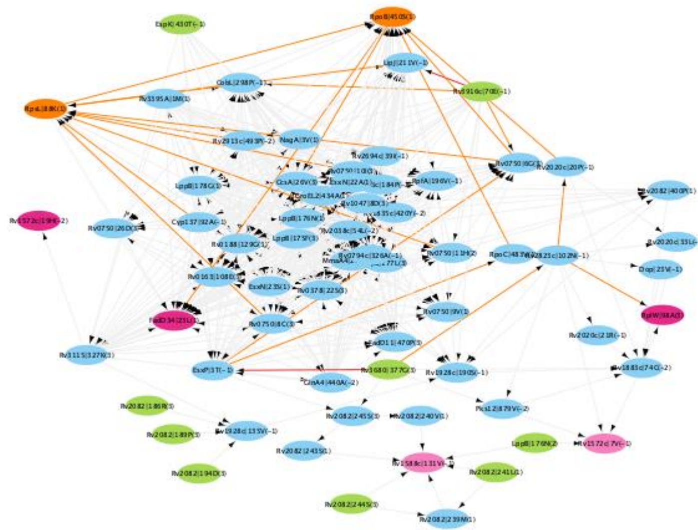
Appendix 40: Epistatic interactions showing various evolutionary paths in lineage 4.1. The green nodes represent a pre-requisite mutation, the blue nodes an intermediate mutation, the orange nodes represent DR mutations, and the pink nodes and final/compensatory mutation. Epistatic links to and from known DR mutations are highlighted by orange edges. Nodes are labelled by the respective gene name and codon position number following the horizontal line.

© University of Pretoria



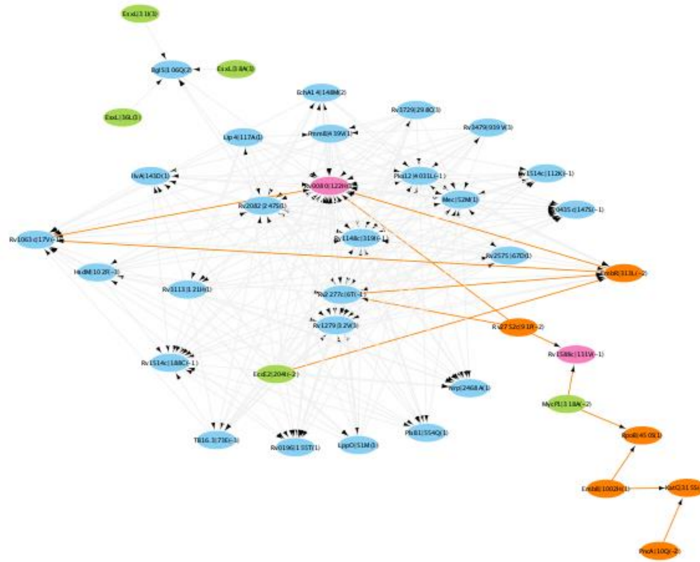
Appendix 41: Epistatic interactions showing various evolutionary paths in lineage 4(Haarlem). The green nodes represent a pre-requisite mutation, the blue nodes an intermediate mutation, the orange nodes represent DR mutations, and the pink nodes and final/compensatory mutation. Epistatic links to and from known DR mutations are highlighted by orange edges. Nodes are labelled by the respective gene name and codon position number following the horizontal line.

© University of Pretoria



Appendix 42: Epistatic interactions showing various evolutionary paths towards drug resistance (DR) in lineage 4(Ural). The green nodes represent a pre-requisite mutation, the blue nodes an intermediate mutation, the orange nodes represent DR mutations, and the pink nodes and final/compensatory mutation. Epistatic links to and from known DR mutations are highlighted by orange edges. Nodes are labelled by the respective gene name and codon position number following the horizontal line.

© University of Pretoria



Appendix 43: Epistatic interactions showing various evolutionary paths in lineage 4(Ural). The green nodes represent a pre-requisite mutation, the blue nodes an intermediate mutation, the orange nodes represent DR mutations, and the pink nodes and final/compensatory mutation. Epistatic links to and from known DR mutations are highlighted by orange edges. Nodes are labelled by the respective gene name and codon position number following the horizontal line.

© University of Pretoria