Exploring the evolution of drug resistance in *Mycobacterium tuberculosis* using Whole Genome Sequencing data

By

Dillon Muzondiwa

Submitted in partial fulfilment of the requirements for the degree
Master of Science (Bioinformatics)

in the

Centre for Bioinformatics and Computational Biology, Department of Biochemistry, Genetics and Microbiology
Faculty of Natural and Agricultural Sciences

UNIVERSITY OF PRETORIA

November 2019

SUMMARY

EXPLORING THE EVOLUTION OF DRUG RESISTANCE IN *MYCOBACTERIUM TUBERCULOSIS* USING WHOLE GENOME SEQUENCING DATA

by

Dillon Muzondiwa

Supervisor: Prof. O.N Reva

Department: Biochemistry, Genetics and Microbiology

University: University of Pretoria

Degree: Master of Science (Bioinformatics)

Keywords: *Mycobacterium tuberculosis*, Antibiotic Resistance, Whole-genome sequencing, Clade-specific, single nucleotide polymorphism

Mycobacterium tuberculosis (Mtb) remains a global challenge that has been worsened by the emergence of drug resistant strains of Mtb. We used publicly available Next Generation Sequencing (NGS) and drug susceptibility (DST) data to develop "Resistance sniffer", an online software program for the rapid prediction of lineage and Mtb drug resistance. Based on the distribution of polymorphisms in the genomes of Mtb, we calculated the power of association between the polymorphisms in different clades of Mtb and resistance to 13 anti-TB drugs. Our data suggests that the development of drug resistance in Mtb is a stepwise process that involves the accumulation of polymorphisms in the Mtb genome. We carefully curated the polymorphisms based on their association powers to create a diagnostic key that captures patterns of these polymorphisms that can be used to predict lineage and drug resistance in Mtb. This diagnosis key was incorporated into the Resistance Sniffer tool, an online software program that we developed for the rapid diagnosis of drug resistance in Mtb. The tool was tested using sequence data from the South Africa Medical Research Council (SA-MRC). Our data suggests that the majority of the

strains in SA may have been brought by the arrival of European settlers while the more resistant strains may have been introduced in the region by Asian travellers later on.

We next sought to determine non-random associations between polymorphic sites in genomes of Mtb. Using the attributable risk (R_a) statistical methods, we distinguished between functional associations and associations that may have been due to genetic drift events for different Mtb clades. We then integrated the (R_a) data with drug susceptibility and annotation data to generate networks in Cytoscape 3.71. These networks were then used to infer evolutionary trajectories that drive the emergence and fixation of the drug resistant phenotype in different clades of Mtb.

We demonstrate that strains from the Lineage 1.2 are associated with less complex functional associations compared to the strains from other clades such as the Asian and Euro-American clades. Our data also shows that the predisposition of strains from the Asian clades to develop multi-drug resistance may be attributed to a complex network of functional interactions of mutations in genes that are involved in several aspects of Mtb physiology such as cell wall modelling, lipid metabolism, stress response and DNA repair.

Declaration

1. I understand what plagiarism is and am aware of the University's policy in this regard.
2. I declare that this M.Sc. dissertation is my own original work. Where other people's work has been used (either from a printed source, Internet or any other source), this has been properly acknowledged and referenced in accordance with departmental requirements.
3. I have not used work previously produced by another student or any other person to hand in as my own.
4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

SIGNATURE

.....

Date:

Acknowledgements

I express my deepest gratitude to the following people and institutions:

- Prof. Oleg for his mentorship and guidance throughout this exciting project.

- Prof. Fourie Joubert, DR. Rian Pierneef for their unwavering support throughout my studies at the University of Pretoria.
- My mother for always being supportive unconditionally.
- My late Sister for always inspiring me through your memory.
- My niece Senzeni for always giving me the reason to work extra hard.
- My brothers for having my back throughout the final year of this project.
- My colleagues at the Centre for Bioinformatics and Computational Biology at the University of Pretoria.
- The National Research Foundation for funding this project.

Table of Contents

List of figures

List of Tables

Abbreviations

TB: Tuberculosis, the disease

MTBC: *Mycobacterium tuberculosis* complex

Mtb: *M. tuberculosis* species

DR: Drug resistant/drug resistance

DS: Drug susceptible

DST: Drug susceptibility testing

GWAS: Genome-wide association study

INDEL: Insertion or Deletion

MDR-TB: Multidrug-resistant tuberculosis

MIC: Minimum inhibitory concentration

NGS: Next-generation sequencing

PDIM: Pthiocerol dimycocerosate

pDST: Phenotypic drug susceptibility testing

SNP: Single Nucleotide Polymorphism

WGS: Whole Genome Sequencing

WHO: The World Health Organization

XDR-TB: extensively drug-resistant tuberculosis

Antibiotics:

AMK: amikacin

CAP: capreomycin

CS: cycloserin

EMB: ethambutol

ETH: ethionamide

INH: isoniazid

RIF: rifampicin

FLQ: any of the fluoroquinolone drugs

KAN: kanamycin

SM: streptomycin

PAS: para-aminosalicyclic acid

OFL: ofloxacin

Chapter 1-Introduction and Literature Review

1.1 Introduction

In 2017, there were an estimated 10 million new cases of tuberculosis (TB) and an estimated 1.6 million deaths attributed to the disease [1]. Despite the drop in global TB incidence and mortality rates in recent years, a lot of work still needs to be done if we are to attain the 2030 targets of the End TB Strategy: to reduce TB deaths by 90% and TB incidence by 80% [2]. TB epidemiological studies have shown that strains of varying genotypes are fuelling the current epidemic with the majority of TB patients having acquired the disease through recent infection (transmission) [3]. The variation in *Mycobacterium tuberculosis* (Mtb) allelic frequencies in different settings also suggests that the evolution of different bacterial mechanisms is at play [3]. This has been further supported by the observed differences in virulence in Mtb sub-lineages. Comparative genomic studies have led researchers to suggest that single nucleotide polymorphisms, insertions and deletions have led to the evolution of the Mtb genome [4]. The aetiological agent, Mtb is a Gram-positive bacilli characterized by a slow growth rate, dormancy, intracellular pathogenesis and genetic homogeneity [5]. The Mtb genome is made up of 4.4 mbp with a relatively constant G +C content of 65.6%, making it the second largest bacterial genome available after the *Escherichia coli* genome [5].

The emergence of drug-resistant TB (DR-TB) remains a major challenge in the war against TB. Since the introduction of antibiotics to the management of TB, drug resistance has always been a barrier to successful treatment [6]. The main reason for this can be attributed to the bacterial population's mutational capacity which is a function of both the mutation rate and the size of the bacterial population. In the early days clinicians adopted monotherapy but soon replaced it with a more effective approach - multi-drug therapy due to rapid development of resistance [7]. According to the World Health Organization (WHO) [1], over 558 000 people had developed TB that was resistant to rifampicin (RR-TB), the most potent of the first line drugs [1], 82% of these were classified as multi-drug resistant TB (MDR-TB) [1]. MDR-TB is referred to TB that has developed resistance to the two most powerful drugs, rifampicin (RIF) and isoniazid (INH). Extensive drug resistance TB (XDR-TB) refers to MDR-TB strains that have developed further

resistance to any of the second line injectable drugs and at least one fluoroquinolone. With 8.5% of the MDR-TB cases now classified as XDR-TB in 2019 there is need to prioritize on the development of tools for the rapid and accurate diagnosis of DR-TB. Treatment of DR-TB remains a challenge, Gandhi *et al.* [8] reported that rapid progression to death was recorded in 98% of XDR-TB patients during an outbreak in KZN South Africa [8]. Other experts have also coined the term totally drug resistant TB (TDR-TB) to describe strains that are resistant to all the currently available drugs although there is no agreed definition of TDR-TB yet [9]. Treatment of drug resistant TB is challenging compared to treatment of drug susceptible disease with global success rates of less than 50% for MDR-TB. The treatment process is costly and is often associated with poor outcomes. The drugs used are also highly toxic and can lead to severe side effects such as permanent deafness and psychiatric disorders [10]. These challenges can lead to poor compliance with the treatment regime and this in turn reduces the cure rates and can even lead to the amplification of resistance [9, 11]. Accurate drug susceptibility testing (DST) profiles are also crucial in the improvement of treatment outcomes as they ensure that only the effective anti-TB drugs are prescribed and reduces exposure to ineffective and toxic drugs [9].

Early diagnosis and correct treatment is the key to the control of MDR-TB and incorrect treatment of TB can be catastrophic both at patient and population level [12]. Misdiagnosis and inadequate treatment of MDR-TB or XDR-TB can lead to the positive selection of resistant subpopulations and hence the creation of resistant strains de novo. This can also lead to an increase to the number of ineffective drugs against an already M/XDR-TB strain [12]. Conventional phenotypic DST is still culture-based and slow due to the slow growth rate of Mtb. The method also requires expensive infrastructure (biosafety level 3 laboratories) which is not accessible in most low middle income which carries the highest burden of TB. This means that DST results are only available after weeks to months which is often too late for the TB patient. For complex drugs such as ethambutol (EMB) and pyrazinamide (PZA), phenotypic DST is often inaccurate and often lacks reproducibility [2, 13]. Several rapid molecular assays have been developed for the diagnosis of DR-TB. The global roll out of the WHO endorsed Cepheid Xpert MTB/RIF and Ultra assays (Cepheid, Sunnyvale, CA, USA) has led to an increase in the number of detected RR-TB cases [14]. However despite their success, these technologies are still limited in the number of loci they can examine, the number of drugs that can be tested and the inability to account for indels [9]. Major diagnostic gaps still remain to be covered, 558 000 people developed MDR/RR-TB in 2017 of which only 160 684 cases were detected and notified. Of these cases, only 25% of the patients were put on a treatment regimen that included a second line drug [15]. These challenges makes

WGS more important in the management of TB than any other infectious disease [16]. Unlike in most other bacterial pathogens, resistance plasmids and horizontal gene transfer play no role in the acquisition of drug resistance in *M. tuberculosis* [17]. The main cause of drug resistance in *M. tuberculosis* is the accumulation of point mutations and indels in genes coding for drug targets or converting enzymes [9]. Recent studies have shown that alterations to these genes and their interactions are just the first step in a longer and more complex process. These mutations are now known to interact with mutations in other genes. These interactions known as epistasis, play an important role in the development of the drug resistant phenotype as they are known to compensate for the fitness cost that is incurred when resistance is acquired. There is a need for more knowledge in the understanding and accurate detection of these processes that lead to the resistant phenotype. This can lead to the development of better diagnostics protocols as well as improved control strategies. Unfortunately the current body of knowledge does not allow for the prediction of epistatic interactions a priori so the only option we have is to detect them empirically by studying the genetics of drug resistance [17]. Unlike the currently available molecular assays, which can only examine limited mutations in specific gene targets, WGS-based assays can provide a near complete view of the whole resistome. This means using WGS assays, we have the potential to detect resistance for all available anti-TB drugs unlike current methods which are limited to only 5 drugs. WGS has the ability to detect rare mutations as well as indels that may not be detected by other molecular assays [2].

1.2 Mechanism of resistance to anti-TB drugs

Figure 1.1 below summarizes the mechanisms of action of some of the main anti-TB drugs currently in clinical use. These can be classified into four groups mainly: i) inhibition of cell wall biosynthesis; ii) disruption of cell membranes synthesis and energetics; iii) inhibition of RNA synthesis; and iv) inhibition of protein synthesis.

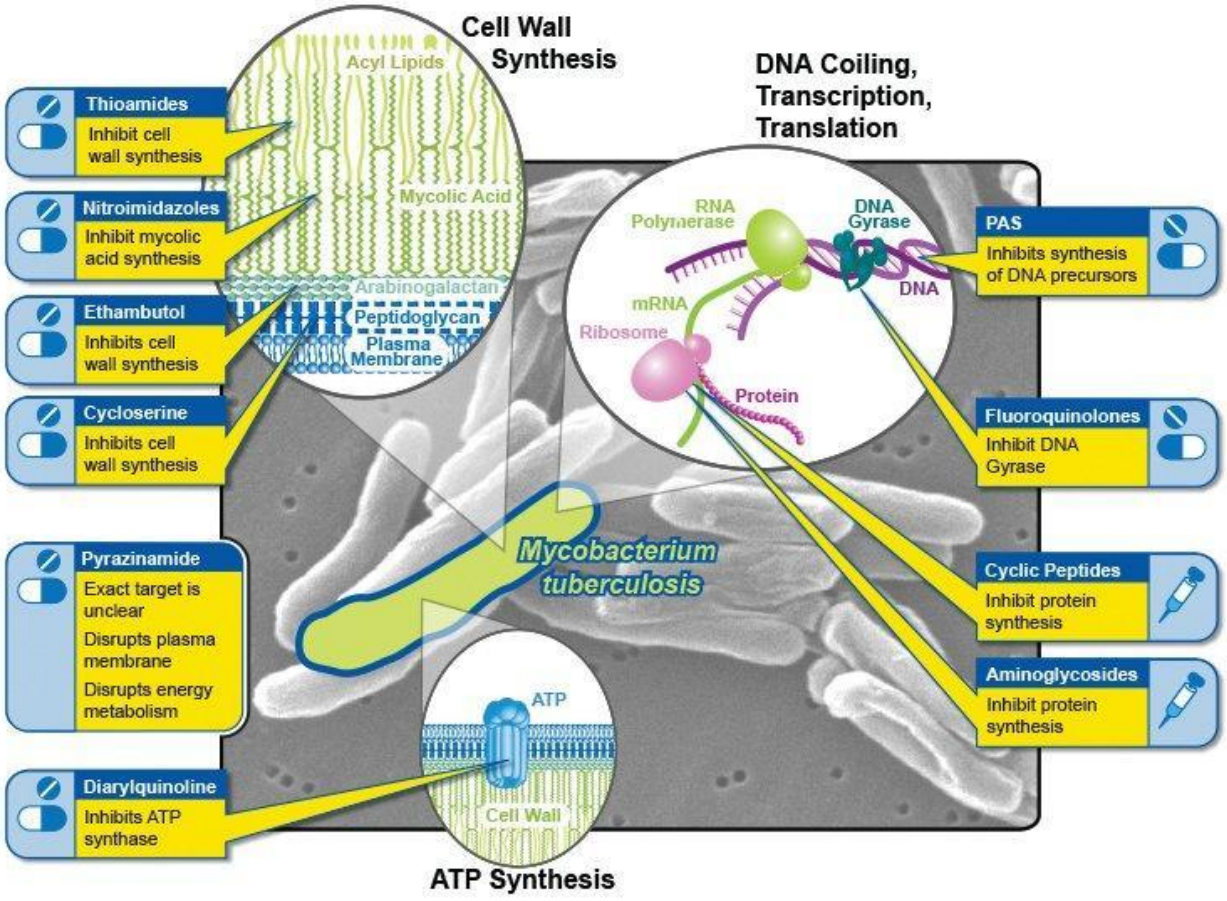

Figure 1. 1 Mechanisms of action of current TB drugs [18].

1.2.1 First line drugs

1.2.1.1 Rifampicin

MTB is associated with low metabolic activity as well as non-replication which are known factors as some of the driving factors of persistent infection [19-22]. Unlike most antibiotics currently in use which are only active against actively growing bacteria, rifampicin also known as rifampin is one of the most potent drugs in the fight against TB because of its effectiveness against actively

metabolizing and slow metabolizing bacilli [23-27]. The drug is known to disrupt the elongation of mRNA by binding to the beta subunit of RNA polymerase. Mutations clustered in codons 507-533 of the *rpoB* gene which codes for the beta subunit of RNA polymerase are known to mediate RIF resistance. The rifampicin resistance determining region (RRDR) is known to harbour mutations that contribute to 96% of rifampicin resistant cases and has been exploited in the development of modern molecular based diagnostic assays particularly codons 526 and 531 [23]. It must also be noted that rifampicin resistance has been detected in some strains that lack any *rpoB* alterations or that harbour mutations outside the RRDR [28]. Most genotypic assays for detecting RIF resistance have been focused on the RRDR such that the role of non-RRDR mutations in RIF resistance is poorly understood. Further work is also needed in understanding the role of the environmental conditions that favour the emergence of specific *rpoB* mutations in MTB [22]. Mono-resistance to rifampicin is a rare phenomenon and almost all rifampicin resistant Mtb strains are also resistant to at least one more antibiotic [24]. For this reason, clinicians have always used rifampicin resistance as a surrogate marker for MDR-TB [29].

1.2.1.2 Isoniazid

Isoniazid (INH) is one of the key components of current first-line regimens in the treatment of drug susceptible TB. Unlike rifampicin, INH is only potent against metabolically-active replicating bacilli [24]. INH is a prodrug which is activated by the catalase /peroxidase enzyme encoded by the *katG* gene [30]. The activated INH binds to the NADH-dependent enoyl-acyl carrier protein reductase which is encoded by the *inhA* gene. The binding is known to interfere with mycolic acid synthesis as well as disrupt multiple essential metabolic pathways. Mutations in the *katG*, *inhA* and its promoter regions have been implicated in INH resistance. The *kat*G S315I, S315N and S315T polymorphisms have been identified as the most common mechanisms of INH resistance. Mutations in the *inhA*/*mabA* promoter regions are also known to lead to drug titration and hence varying levels of resistance [31]. The most common of these mutations is found at position 15 of the *inhA* promoter region and has been linked with low levels of INH monoresistance as well as cross resistance with ethionamide (ETA) which is a structural analogue of INH [23, 24]. Mutations in the active site of *inhA* lead to a reduction in the affinity of the INH-NAD product [23, 32, 33]. A recent MIC/mutation data review by Miotto *et al.* also associated INH resistance with mutations in *kasA, furA, oxyR-ahpC* and *ndh* genes which are involved in mycolic acid synthesis pathways [31]. 84% of global phenotypic INH resistance can be accounted for using these mutations [23,

34]. There is a growing body of evidence that suggests that INH resistance mediated by the *katG* S315T mutation is a precursor of rifampicin resistance which makes the mutation an ideal marker in the diagnosis of MDR-TB [35]. These findings highlights the importance of early detection of antibiotic resistance in the management of TB.

1.2.1.3 Ethambutol

Ethambutol (EMB)'s mechanism of action makes it one of the most important first line drugs in the treatment and management of TB. The drug disrupts several pathways of actively multiplying bacilli, most importantly those that are involved in arabinogalactan biosynthesis in the cell wall [36, 37]. This inhibition of arabinan polymerization helps to facilitate the permeability of other drugs that are used in TB treatment regimens leading to improved treatment outcomes. EMB resistance has been linked to mutations in the embCAB operon which encodes the mycobacterial arabinosyl transferase enzyme [23, 36, 37]. Mutations in codon 306 of the *embB* have been initially associated with EMB resistance and several studies have implicated mutations in this particular gene as a stepping stone to poly-drug resistance in MTB [38]. However a study by Hazbon *et al.* concluded that mutations in *embB*306 were not necessarily determinants of EMB resistance but rather act as markers of strains that are predisposed to develop resistance to multiple drugs [24, 39]. Safi *et al.* performed several allelic exchange experiments that have demonstrated the complexity of the development of EMB resistance in MTB. Their study showed that contrary to earlier suggestions, acquisition of high level EMB resistance was a multi-step process [17]. It is also important to note that a third of EMB resistant isolates do not harbour any *embB* mutations, this suggests the existence of alternative mechanisms of EMB resistance [40]. The *ubiA* gene has also been linked to high levels of EMB resistance in certain African isolates [23]. The gene encodes decaprenyl-phosphate 5-phosphoribosyltransferase synthase, which is an important enzyme in cell wall synthesis pathways [41].

1.2.1.4 Pyrazinamide

The introduction of the nicotinamide analogue pyrazinamide (PZA) as one of the first line antibiotics in the treatment of DS-TB has resulted in the reduction of treatment durations to just six months [24]. The drug is also important in the development and evaluation of new anti-TB regimens as it is included in Phase II and III of all DS-TB and DR-TB clinical trials [24, 42]. TB

lesions are known to harbour semi-dormant bacilli which are difficult to control with most of the commonly used antibiotics due to the acidic conditions [43]. However PZA is active at low pH levels which makes it a vital drug in the effective treatment of TB [44]. The enzyme pyrazinamidase/nicotinamidase (PZase) activates PZA to pyrazinoic acid which in turn disrupts the proton motive force which is essential for membrane transport [23, 24, 44]. Recent studies have also linked pyrazinoic acid and its n-propyl ester to the inhibition of fatty acid synthase I in Mtb [24, 45]. Although traditional dogma has always been centred on the PZA being dependent on low pH environments, recent studies have shown low pH and intra-bacterial acidification is not necessarily a prerequisite for PZA activity [46, 47]. However this does not mean that PZA generally works at neutral pH, irrespective of the bacilli's metabolic state [48]. It rather shows that the bacterial metabolic state of the bacilli is a factor in determining the activity of the drug. PZA would still show activity at neutral pH in dormant persister Mtb even though it will be expected to show more activity at lower pH [48].

Although the mechanism of PZA resistance is poorly understood, mutations in the gene that encodes Pzase (*pncA*) and its promoter regions have been associated with PZA resistance. Mutations in this highly polymorphic gene have been attributed to 72% - 99% of PZA resistance [49, 50]. DST for PZA is highly inaccurate and challenging due to the low pH conditions required for the culture. The ribosomal protein I (*rpsA*) was initially suggested as a target of PZA, the drug was found to inhibit and trans-translation in Mtb [48]. However, this finding was recently disproved by Dillon *et al.* [47]. Recent studies have also shown that the aspartate decarboxylase (*panD*) is inhibited by PZA [51, 52]. The enzyme is involved in the synthesis of pantothenate and co-enzyme A, which is involved in energy production. Initially researchers were sceptical on *panD* as a drug target for pyrazinamide but a recent study by Gopal *et al.* confirmed the gene as a PZA drug target [53].

1.2.1.5 Streptomycin

Streptomycin (SM) was the first antibiotic treatment for TB in 1944 [24, 48]. The aminocyclitol drug inhibits protein synthesis in slow-growing bacteria by irreversibly binding to the ribosomal protein S12 and 16S rRNA and therefore interfering with the binding of formyl-methionyl-tRNA to the 30S subunit of the bacterial ribosome. The drug rapidly developed resistance due to its overuse in the previous century. SM resistance is known to be mediated by mutations in the *rpsL* and *rrs* genes. The *rpsL* gene encodes the ribosomal protein S12 while the 16S ribosomal subunit

is encoded by the *rrs* gene. Mutations in the 7-methyl guanosine methyltransferase gene (gidB) have been associated with low level SM resistance. *rpsL* mutations have been attributed to 50% of SM resistance while *rrs* mutations account for approximately 15% of resistance to the drug [24, 48]. Mutations in *gidB* gene which encodes 7-methylguanosine methyltransferase have also been associated with low level SM resistance. These mutations inhibit 16S ribosomal methylation leading to a reduction in the affinity between SM and the 16S ribosomal RNA (rRNA)-binding site [48, 54, 55].

1.2.2 Second line drugs

1.2.2.1 Injectable drugs

Kanamycin (KAN), amikacin (AMK) and capreomycin (CM) are the second line injectable drugs currently used in the management of TB [24]. Although the two aminoglycosides AMK and KAN and the cyclic polypeptide CM have different mechanisms of action, the three antibiotics exert their effect by inhibiting protein synthesis. The aminoglycosides are known to bind to the 16S rRNA in the 30S subunit of the MTB ribosome while CM disrupts translation and is understood to inhibit phenylalanine synthesis in MTB ribosomes [56]. Mutations in the *rrs* gene are understood to confer high level resistance to these drugs [57, 58]. Mutations in the rRNA methyltransferase *tlyA* have also been associated with CM resistance. These mutations are known to determine the absence of methylation activity [24, 59]. Mutations a1401g and g1484t in the *rrs* gene are known to be high confidence markers of resistance for the three drugs with the former associated with 70%-80% of CM and AMK resistance and 60% KAN resistance [23, 60]. Recent studies have also reported on cross resistance between the three drugs which is expected since the antibiotics share the same drug target. Low level resistance to kanamycin has been attributed to mutations at position -10 and -35 of the *eis* promoter [24]. Up to 80% of clinical isolates exhibiting low-level kanamycin resistance harbour these mutations [61, 62]. Cycloserine (CS) is another bacteriostatic agent that has been used six decades [48, 63, 64]. The drug, which is a cyclic analog to D-alanine inhibits peptidoglycan biosynthesis by blocking the conversion of L-alanine into D-alanine. Its use in TB treatment has been limited by its link to neurological toxicity. Currently the drug is only prescribed for only proven DR-TB cases. The genetic basis of CS resistance remains elusive despite recent studies that have linked mutations in the L-alanine dehydrogenase (*alr*), D-alanine (*ald*) and D-alanine ligase (*ddlA*) genes. Another challenge is that similar to PZA, the loss-

of-function mutations within the *ald* gene (1116 bp) are not confined to a specific region which makes it difficult to predict CS-resistance profiles using genotypic data [65].

1.2.2.2 Fluoroquinolones

Fluoroquinolones (FQ) are some of the most potent anti-TB currently prescribed as second-line treatment drugs for DR-TB. Newer generation FQs, such as levofloxacin and moxifloxacin are becoming important components of DR-TB treatment regimens [66]. FQs inhibit transcription during bacterial cell replication by binding to DNA gyrase. The tetramer type II topoisomerase consists of two subunits α and β which are encoded by the *gyrA* and *gyrB* genes which are responsible for the catalysis of DNA supercoiling [67]. Genetic mutations in those two genes are known to confer FQ resistance in MTB with *gyrA* gene harbouring most of these mutations. The quinolone resistance determining regions (QRDRs) are located between codons 74 and 113 and codons 500 and 538 of the *gyrA* gene [68, 69]. Codons 88, 90, 91 and 94 of the *gyrA* QRDR are known to harbour most of the common FQ mutations while often rare, mutations in *gyrB* have been linked with low levels of FQ resistance. It is important to note that there is also a mutation at position 95 in *gyrA* which is not associated with FQ resistance since it is also found in susceptible Mtb isolates [70]. An interesting finding by Aubry *et al.* showed that the simultaneous occurrence of T80A and A90G in *gyrA* led to hyper-susceptibility to a number of quinolones [24, 71]. Resistance to FLQ is one of the indicators of the development of XDR-TB. It is imperative to fully elucidate the full mechanism of FQ resistance because low level resistance in some of the new generation FQs has been recently reported.

1.2.2.3 Ethionamide

Ethionamide (ETA), derived from isonicotinic acid, is a structural analogue of INH [23]. The prodrug is activated by the mono-oxygenase enzyme to inhibit the binding of the enoyl-acyl carrier protein reductase and therefore inhibit cell wall synthesis. ETA resistance is believed to be caused by mutations in the *ethA* gene which encodes the mono-oxygenase enzyme [72, 73]. Mutations in the transcriptional repressor gene *ethR* as well as the *inhA*-gene and its promoter have also been associated with ETA resistance. Several studies have also revealed the role of inhA mutations in ETA/INH co-resistance. Mutations in these two genes are known to account for 70% of ETA

resistance [73]. Further investigations are required to fully elucidate the mechanism of resistance in ETA.

1.2.2.4 Para-aminosalicyclic acid (PAS)

PAS is one of the first antibiotics used in the treatment of TB together with streptomycin. The prodrug forms part of the second line regimens prescribed for DR-TB. Elucidating the mechanism of PAS resistance has been challenging until recently due to its gastrointestinal toxicity and low potency when compared to RIF and PZA [48]. PAS an analogue of para-amino benzoic acid, is a competitive inhibitor of dihyropteroate synthase which is involved in the folate synthesis pathway [24]. The prodrug is activated by thymidylate synthase and the activated drug competes for the enzyme with p-amino benzoic acid to inhibit iron uptake [74]. Mutations in the *thyA* gene have been identified as the main mechanism of resistance, accounting for 40% of PAS resistance [23, 75]. Mutations in the *folC* gene which encodes dihydrofolate synthase were also detected in clinically resistant strains [76]. The usage of PAS as an anti TB drug has been low due to its high toxicity levels and side effects and more work still needs to be done in fully decoding the mechanism of PAS resistance in MTB.

1.2.3 New and repurposed drugs

1.2.3.1 Bedaquiline

Bedaquiline (BDQ) is classified under the diarylquinolones, a new class of compounds used in the treatment of TB. The drug which has been a success in high burden countries exerts its action by targeting mycobacterial ATP synthase to inhibit bacterial respiration [77]. The drug is effective against dormant bacilli which makes it highly effective in treating both DS-TB and MDR-TB especially when prescribed together with PZA [77]. BDQ exerts its action by inhibiting mycobacterial ATP synthesis as it binds to the C subunit of the F0 complex of ATP synthase which is encoded by the *atpE* gene [48, 78]. In 2005 *in vitro* studies by Andries *et al.* implicated mutations in the *atpE* gene with BDQ resistance [77]. The gene encodes the bacterial F_1F_0 proton synthase which is involved in ATP synthesis as well as membrane energetics [23, 79]. Mutations in the Rv0678 gene have also been linked with BDQ resistance as well as cross resistance with clofazimine. *pepQ* mutations have also been linked with low level resistance to BDQ as well as clofazimine. Cross resistance of both drugs has been attributed to the upregulation of the *mmpL5*

which encodes the MmpL5 efflux pump [48]. The lack of DST data for new and repurposed antibiotics is still a challenge that needs to be addressed if we are to determine the complete mechanism of resistance to BDQ.

1.2.3.2 Clofazimine

Clofazimine (CFZ) was discovered in 1954. Initially the drug was used for the treatment of leprosy as it was ineffective against TB [80]. Studies have proved that CFZ is active against MDR-TB which led to the WHO recommending the drug as part of the new standardized short course regimen for DR-TB [81]. It must be noted that although the drug was approved by the FDA in 1986, its use in the treatment of DR-TB has not been approved by any stringent regulatory body and it is therefore prescribed "off-label" [82]. The exact mechanism of CFZ action is unknown, however studies performed in *M. smegtmatis* suggests that it is a prodrug which is activated by NAD dehydrogenase, to release reactive oxygen species upon re-oxidation by oxygen [23, 83]. Mutations in *Rv0678* have been linked to CFZ resistance together with mutations in *pepQ* and *Rv1979c* genes [84, 85]. Resistance to CFZ has also been associated with BDQ as explained above. The mechanism of action of CFZ is poorly understood due to its limited use as it is associated with adverse side effects. More studies still need to be done in order to fully exploit the capabilities of this drug in TB management.

1.2.3.3 Linezolid

Linezolid is the first oxazolidinone to be approved for the treatment of TB [86]. The drug is a protein synthesis inhibitor which exerts its effect by binding to the V domain of the 50S subunit of the mycobacterium ribosome [23, 87]. Linezolid resistance has been linked to mutations in the 23S (*rrL)* gene [22]. In vitro studies have associated mutations G2061T and G2572T in the *rrL* gene with high level linezolid resistance [88]. The study went on to report that no alterations to the *rrL* gene were found in isolates that bore low level linezolid resistance. The role of *rrL* mutations in linezolid resistance was further supported by Bloemberg *et al.* who detected A2572C and G2576T mutations in a patient with linezolid resistant TB [89]. Studies by Zimekov *et al.* have also implicated a mutation in the *rplC* gene in the development of linezolid [90]. The gene encodes the L3 protein on the 50S ribosome and the C154R mutation is the most frequent among linezolid resistant isolates [90]. Linezolid has been shown to improve treatment outcomes in complicated

MDR-TB as well as XDR-TB cases thus it will be of great benefit to elucidate its mechanisms of resistance [91].

1.2.3.4 Pretonamid and Delamid

Pretonamid and Delamid are classified under the nitroimidazole group of antibiotics [92, 93]. Both compounds are prodrugs that are activated by deazaflavin-dependent nitro-reductase which is encoded by the *ddn* gene [23]. The enzyme breaks down the prodrug into des-nitro-imidazole and two other unstable metabolites. Des-nitro-imidazole produces reactive nitrogen species as well as nitric oxide which may promote host-macrophages acting against MTB [94-96]. Delamid, formerly known as OPC-67683 is known to influence the synthesis of the bacterial cell walls by inhibiting methoxy-mycolic and keto-mycolic acids [97]. Mutations in *ddn* and *fgd1* genes which are involved in the activation of the prodrug as well as mutations in *fbiA*, *fbiB* and *fbiC* have been associated with resistance to both drugs [23, 95]. The *fbi* genes encode for proteins involved in the F420 biosynthetic pathway.

Table 1. 1 A summary of anti-TB antibiotics and possible mechanisms of drug resistance.

Antibiotic	**Abbreviation**	**Mechanism of action**	**Mutated genes associated with drug resistance**
Amikacin,Capreomycin,Kanamycin	AMK, CM, KAN	Inhibits protein synthesis through ribosomal binding.	*rrs, eis, tlyA* [57].
Cycloserine	CS	Cell wall biosynthesis inhibitor.	*alr, ddlA, cycA* [98, 99].
Ethambutol	EMB	The drug disrupts several pathways of actively multiplying bacilli, most	*embB, ubiA* [100].

		importantly those that are involved in arabinogalactan biosynthesis in the cell wall. This inhibition of arabinan polymerization helps to facilitate the permeability of other drugs that are used in TB treatment.	
Ethionamide	ETH	Structural analogue of INH (Dookie *et al.*,2018). The prodrug is activated by the mono-oxygenase enzyme to inhibit the binding of the enoyl-acyl carrier protein reductase and therefore inhibit cell wall synthesis.	*ethA, mshA, ndh, inhA, inhA* promoter [101].
Isoniazid	INH	INH is a prodrug which is activated by the catalase /peroxidase enzyme encoded by the katG gene [30]. The activated drug interferes with mycolic acid synthesis.	*katG, inhA, inhA-promoter* [102].

Fluoroquinolones	FLQ	Inhibit bacterial replication by blocking DNA gyrase important for the replication pathway.	*gyrA* and *gyrB* [67].
Para-amino salicyclic acid	PAS	A para- amino benzoic acid that inhibits folate synthesis.	*thyA* [23].
Pyrazinamide	PZA	Disrupts the proton motive force which is essential for membrane transport.	*pncA* [13].
Rifampicin	RIF	The drug is known to disrupt the elongation of mRNA by binding to the beta subunit of RNA polymerase.	*rpoB* [22].
Streptomycin	SM	Inhibits protein synthesis by irreversibly binding to the ribosomal protein S12 and 16S rRNA and therefore interfering with the binding of formyl-methionyl-tRNA to the 30S subunit of the bacterial ribosome.	*rrs* and *rpsL* [103], *gidB* [55],*whib7* [104].

1.3 The evolutionary path of drug resistance in *Mycobacterium tuberculosis*

1.3.1 Fitness and epistasis

The evolutionary trajectory which leads to drug resistance in MTB is heavily influenced by two factors: bacterial fitness and epistasis [17, 105-107] (see Figure 1.2 below). Epistasis is defined as a set of genetic interactions where the phenotypic effect of one mutation is dependent on the presence of one or more mutations [17]. Bacterial fitness is defined as the ability to adjust metabolism in order to adapt to a certain environment. Bacterial fitness can be summarised as a function of growth rate, transmissibility and virulence [17, 108, 109]. Any polymorphism that reduces bacterial fitness relative to the wild-type strain is said to carry a fitness cost [17]. Drug resistance mutations are generally known to carry a fitness cost in MTB [23]. However it is important to note that, this fitness cost is influenced by various factors such as strain genetic background and compensatory evolution. Certain MTB strains which harbour DR mutations with no fitness cost have been reported [108]. Although observations in clinical settings have also shown selection for minimum or no fitness cost mutations, data on the relative transmissibility of MDR-TB strains has been inconclusive when compared to DS strains [110]. Borrell *et al.* further suggests that fitness of drug resistant strains is a heterogeneous entity [111]. MDR-TB and XDR-TB is also rampant in regions in HIV pandemic regions, suggesting that DS strains might be more fit than fully DR strains [110]. Although this might be true for regions like sub-Saharan Africa, for countries in the eastern bloc of Europe and the former Soviet Union, prevalence rates of MDR-TB are still high despite their low HIV prevalence rates [111, 112]. The success of MDR strains in these regions can be explained by the role of compensatory evolution to mitigate the effects of fitness cost overtime. The frequency of the Beijing strains in these regions also supports the hypothesis that the strain genetic background plays a role in the development of drug resistance.

Despite the importance of epistasis in circumventing the fitness cost incurred by DR mutations in MTB, very few studies have been done until recently using WGS strategies. Epistasis can be classified as either positive epistasis when the mutations are beneficial or negative epistasis for deleterious mutations [111, 113, 114]. Several studies have reported on the occurrence of certain drug resistant mutations having a different fitness effects on strains with varying genetic backgrounds [115]. When the bacterial genome carries multiple DR mutations, the overall fitness cost of resistance will not only depend on the cumulative fitness costs of individual mutations but

on the epistatic interactions between those mutations as well. Positive epistasis is observed when the overall fitness cost of resistance is less than what would be observed when we add the effects of the individual DR mutations [111]. Subsequently, negative epistasis is observed when the fitness cost of carrying multiple DR mutations is more than what would be observed when we sum up the fitness effects of each individual mutation. Positive epistasis drives the evolution of MDR-TB by mitigating the fitness cost associated with it while negative epistasis decelerates the evolution of MDR-TB by inflating the fitness cost [111, 113]. Secondary mutations can also influence the evolution of MDR-TB by causing positive epistasis. Several studies on compensatory evolution have shown that certain non-DR mutations are involved in the evolution of many drug-resistant bacterial species [116]. Allelic exchange and directed mutagenesis experiments found that inserting compensatory mutations into wild-type DS strains led to a deleterious loss of fitness, thereby highlighting the effects of epistasis [111, 113, 117, 118].

The currently available body of knowledge does not allow us to theoretically predict epistatic interactions, the only possible way is to detect them empirically is through studying the genetics of drug resistance [17, 119]. The advent of WGS technologies resulted in a number of studies focusing on the evolution of drug resistant MTB. The use of anti-TB drugs imposes strong selective pressure on MTB populations [120] so despite the fact that these studies use different approaches which include phylogenetic, mutation frequency analysis and molecular epidemiology, the shared aim is to identify bacterial genes under positive evolutionary selection by drug pressure [17]. Trauner *et al.* reviewed some of these studies which were done recently [17]. In their review they identified not only known drug targets but other novel bacterial genes and intergenic regions whose function may be ancillary to DR mutations as shown in Figure 1.3 below. Genes that were found to be under positive selection in the presence of antibiotics were mainly involved in cell wall synthesis and homeostasis, transcriptional control, lipid metabolism and purine metabolisms. Further investigation of these findings can be beneficial in drug target discovery as well as broaden our understanding of MTB drug resistance [17, 121, 122]. Given the importance of RIF as an anti-microbial drug, it is not surprising that a number of studies on bacterial fitness have been focused on the *rpoB* gene [22]. A study by Gagneux *et al.* showed that mutation *rpoB* S450L which is one of the most frequent RIF resistant mutations in the clinic also carries the lowest fitness cost [123]. This mutation is known to be abundant among MDR-TB strains and has also been associated with a propensity to acquire compensatory mutations within the RNA polymerase genes (*rpoA, rpoB, rpoC*) [17]. Several studies on the transmissibility of MTB strains have associated these mutations with improved transmissibility [124-126]. One of the most important studies that confirms the role

of the mutations in restoring bacterial fitness were done by Brandis *et al.* Working on *Salmonella enterica*, the team demonstrated that the acquisition of these mutations improved the growth rate of slow growing *rpoB* mutants in RIF media [127, 128]. In 2013 Borrell *et al.* demonstrated positive epistasis between *gyrA* mutations and some *rpoB* alleles using *M. smegmatis* [129]. Their study showed that the mutants that carried specific combinations of *gyrA* and *rpoB* alleles were fitter than those strains harbouring single resistance determinants during competitive growth under standard conditions in vitro [22]. It is important to note that the same *gyrA/rpoB* SNP combinations were also identified in clinical XDR-TB isolates [129]. These findings suggests the need for further investigation on the link between transcription rate and DNA supercoiling in the evolution of MDR-TB. Certain INH mutations have also been linked with positive epistasis with the *rpoB* gene and thereby predisposing certain isolates to the development of MDR-TB. A further understanding of epistasis signals in MTB will also help in guiding the prescription of certain drugs in the treatment of TB. This will also prevent the fuelling of evolution into high level drug resistance and also to protect the effectiveness of new antibiotics.

1.3.2 The role of efflux pump systems in MTB drug resistance

The MTB genome is enriched with a large number of efflux pumps (148 genes coding for membrane transport peptides) which facilitates the expulsion of antibiotics from the bacterial cell [22]. These efflux pump belong to the major facilitator super family, the resistance nodulation division family and to the small drug resistance family [130]. Although wild type cells naturally express drug efflux pumps, the presence of mutations within regulatory genes as well as antibiotics is also known to induce drug efflux pump expression. The role of efflux pump systems in the development of MDR-TB has been a hot topic of research over the past few years [17, 131]. Several studies on clinical strains have reported on the overexpression of a number of efflux systems upon exposure to some of the widely used anti-TB drugs. Unlike drug target polymorphisms which usually cause high-level resistance, the overexpression of efflux pumps results in the reduction of intracellular levels of anti-TB drugs leading to the development of low-level resistance [130-132]. Studies have also reported on the vast overlap in substrate specificity among the efflux pumps present in Mtb [130, 133]. Work by Louw *et al.* and Ainsa *et al.* also showed that the same efflux pump can induce cross tolerance to structurally and mechanistically diverse substrates [130, 134, 135]. These findings supports the evidence that the rapid and non-specific ability of the efflux pumps to extrude highly noxious compounds can lead to the development of MDR phenotype [132, 136]. Strains utilizing these efflux systems might be

positively selected in the presence of suboptimal, low anti-TB drug concentrations. These low level resistant bacterial subpopulations may survive standardized treatment until a classical high level polymorphism emerges leading to the development of high level MDR [130, 136-139]. The differential expression of efflux pumps may also partly explain why some observed clinical strains harbouring the same resistance polymorphisms can have different phenotypic DST profiles [130, 140]. Efflux pump inhibitors are compounds capable of reducing the MICs of several anti-TB drugs in DR-TB strains [23, 130]. The introduction of efflux pump inhibitors as an adjunctive therapy in TB treatment regimens has the potential to reduce the duration of TB treatment [23, 130-132, 141, 142]. However it is important to note that exposing MDR-TB strains to efflux inhibitors does not simply translate to the restoration of full susceptibility, as resistance determining polymorphisms might have been acquired in other other genes that are not associated by these compounds [130].

1.3.3 The role of deficient DNA repair systems in the evolution of Mtb drug resistance

DNA repair systems are known to directly influence the type and frequency of mutations in bacteria [130]. For this reason, there has been growing interest in studying the role of DNA repair mechanisms in the evolution of drug resistance. Impaired DNA repair mechanisms are known to increase mutations rates and therefore positively select for DR bacterial strains. Ebrahimi-Rad et *al.* identified missense mutations in the putative anti-mutator (*mut*) genes in strains from the Beijing lineage [23, 143]. Strains from the Beijing lineage have been associated with a higher propensity to develop into MDR/XDR-TB strains when compared to strains from other lineages [144]. The team further hypothesized that increased mutation rates as a result of missense *mut* gene mutations could be linked to the prevalence of MDR in Beijing strains. These findings suggest the need for further studies in the link between these repair systems and drug resistance in Beijing strains. The combined use of modern tools such as WGS together with other traditional experimental methodologies such as Luria Delbruck fluctuation analysis [145, 146] has also enabled researchers to quantify the determinants of Mtb mutations in the setting of host infection [147]. A study by Ford *et al.* used Luria Delbruck fluctuation analysis to determine differences in mutation rates between strains of Mtb [147]. In their study, they reported that East Asian strains acquired resistance to multiple antibiotics at significantly higher rates (1.78 – 37.07 fold) compared to Euro-American strains [147]. Mutation rates have also been shown to be elevated in strains grown in environments containing sub-inhibitory concentrations of certain drugs, particularly

those whose primary mechanism is DNA damage such as fluoroquinolones [130, 148]. Recent transcriptional studies using GWAS in MTB have shown that DNA repair clusters are up-regulated in isolates that have been exposed to FLQ [130, 149, 150]. The use of quinolones has been linked with the development of resistance to other antibiotics of different classes [151, 152]. Although INH is not directly involved in DNA metabolic processes, the drug has also been linked to increased mutations rates in Mtb [153, 154]. Due to the importance of FLQs and INH in the management of TB, further interrogation of these findings is needed.

1.3.4 The role of strain genetic background

The global genetic diversity of Mtb is made up of seven distinct phylogenetic lineages [155]. Despite this genetic variation being small compared to other bacterial pathogens, the Mtb strain genetic background has been shown to influence bacterial fitness and drug susceptibility *in vitro* [111]. This means that strains that carry the same resistance mutation can have variable drug susceptibility profiles due to a difference in fitness costs. A study by Zaczeck *et al.* demonstrated that mutations in the *rpoB* genes resulted in different levels of rifampin resistance in strains from different phylogenetic lineages of Mtb [156]. Similar results were also observed in *katG* and *inhA* mutations that confer resistance to isoniazid. Certain lineages are also known to be highly associated with the MDR/XDR-TB phenotype and are highly transmissible [108, 110, 157-161]. Work by Gagneux *et al.* using strains from lineage 4 and lineage 2 demonstrated that the same histidine to aspartic acid mutation on codon 526 of the *rpoB* gene resulted in different fitness deficits in MTB strains from the two lineages [108, 130]. These studies were further supported by recent work by Castro *et al.* who investigated the role of strain genetic background in the evolution of FLQ resistance [162]. The team reported that the Mtb strain genetic background led to a significant variation in the frequency of resistance to ofloxacin [162]. The study also reported on a positive association between the resistance profiles of clinical isolates and those from *in vitro* isolates [155]. These findings highlights the importance of identifying the lineage of the infectious Mtb strain(s) in any diagnostic approach as some of the Mtb lineages might have acquired some

unique features prior to expanding [30].

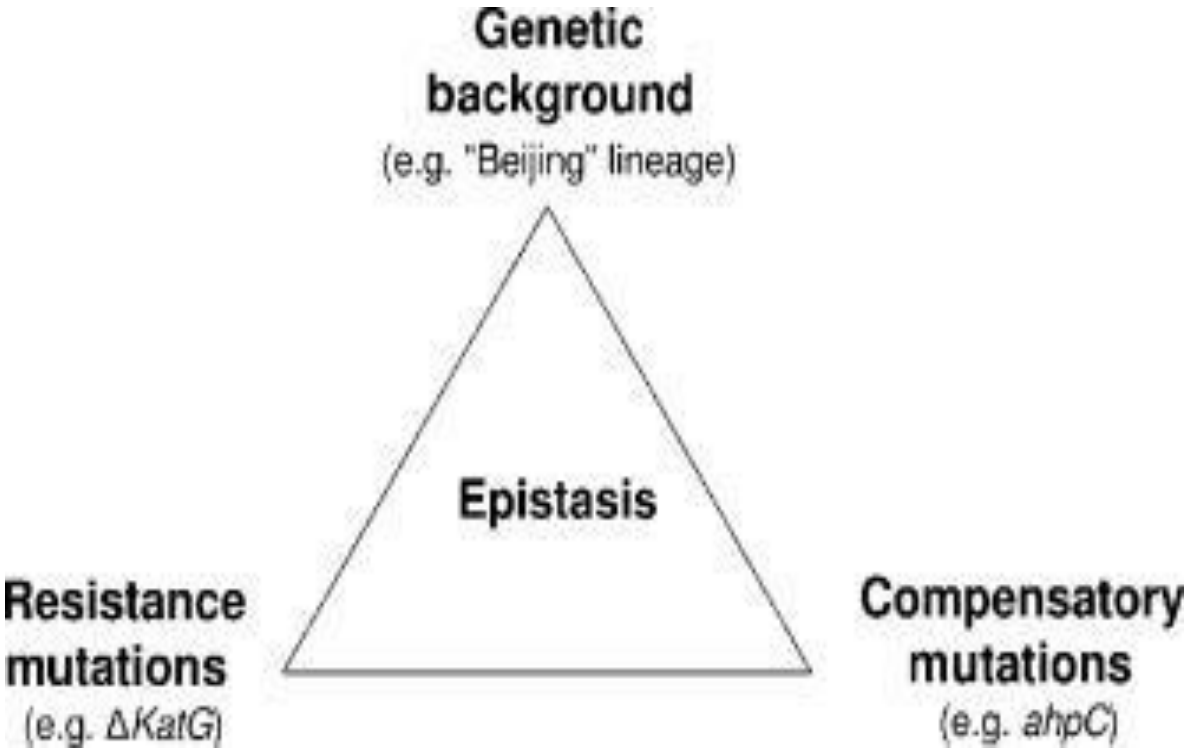

Figure 1. 2 Factors that determine the evolution of drug resistance in Mtb [111].

1.3.5 Cross resistance and hetero-resistance

Traditional dogma is rooted on the premise that Mtb infection is homogeneous, the use of WGS in recent studies has challenged this notion and suggested that the level of heterogeneity is much higher than previously thought. Several mechanisms by which Mtb heterogeneity in an individual arise have been suggested and these include (1) simultaneous infection by multiple strains (2) super infection or re-infection by a newer strains (3) the spontaneous emergence of genetic diversity during the course of infection [147, 163, 164]. The management of DR-TB is further complicated by hetero-resistance which occurs when both resistant and susceptible strains coexist within a specimen from a single patient. According to Cohen *et al.* [165], hetero-resistance is found in 5.38% of DR-TB and emerges either as a result of infection with different Mtb isolates or through the acquisition of mutations within a clonal Mtb population. Culturing of Mtb isolates often fails to account for hetero-resistance within samples which can lead to inaccurate results during pDST. This can have serious implications in patient management as a result of incorrect prescription of TB drug regimens. There is a need for diagnostic approaches that are sensitive enough to account for the role of low frequency variants in the emergence of DR-TB even before it is detectable by

conventional pDST. Several studies have reported on the variability that occurs at DR loci within an individual patient [155, 166-168]. Low frequency variants (<1-5% of the population) have been known to alternate their frequency as the infection progresses. This observed micro-heterogeneity has led to suggestions that there is a shuffling and sampling of the mutational landscape by Mtb until the fixation of one particular mutation eventually occurs [155, 165, 166]. The evolution of heterogeneity of drug resistance is beyond the scope of this review but for more information, readers are referred to a review by Gagneux *et al.*[155] which explains on the role of natural selection and genetic drift in the evolution of Mtb and the emergence of drug resistance.

1.4 Whole genome sequencing in TB management

The advent of NGS technologies was a game changer in the use of sequencing methodologies in clinical microbiology [31, 48, 169, 170]. Analysis of the high throughput data generated from WGS has led to the reconstruction of the Mtb phylogeny and thus expanding on our knowledge on the global distribution of Mtb [171, 172]. The high resolution of WGS has also been exploited in TB epidemiology as it allows us to track transmission dynamics through the analysis of SNPs in the Mtb genome. The use of WGS combined with social network analysis can be beneficial in managing outbreaks. A pilot study by Daum *et al.* in 2012 showed the potential of using NGS to detect Mtb drug resistance using the Ion Torrent Personal Genome Machine (Thermo Fisher Scientific Inc., Waltham, MA, USA) [170]. Their approach was superior to traditional molecular diagnostic methods in that, it allowed for the sequencing of the entire gene lengths, instead of specific loci of interests [48]. The work was subsequently followed by another study in 2013 which sequenced whole genomes of Mtb on Illumina (San Diego, CA, USA) NGS platforms for epidemiological and rapid molecular DST [48, 173, 174]. WGS has been shown to be superior when compared to conventional phenotypic diagnostic methods [48, 175]. WGS assays have been associated with high sensitivity for species identification (93%, 95% CI: 90-96%) and drug susceptibility (93%, 95% CI: 91-95 %) [48, 175]. The clinical turnaround time for WGS-based DST is twenty four hours which is an enormous improvement when compared to conventional pDST which takes at least takes 28 days to report. The rapidity of WGS DST offers a number of advantages in the management of TB as it allows for faster and more effective treatment and hence reduces onwards transmission by reducing the time of infectivity. The use of WGS is also cost effective compared to pDST as it produces DST results for the queried antibiotics simultaneously and thus eliminating the need for at least seven different molecular assays [48].

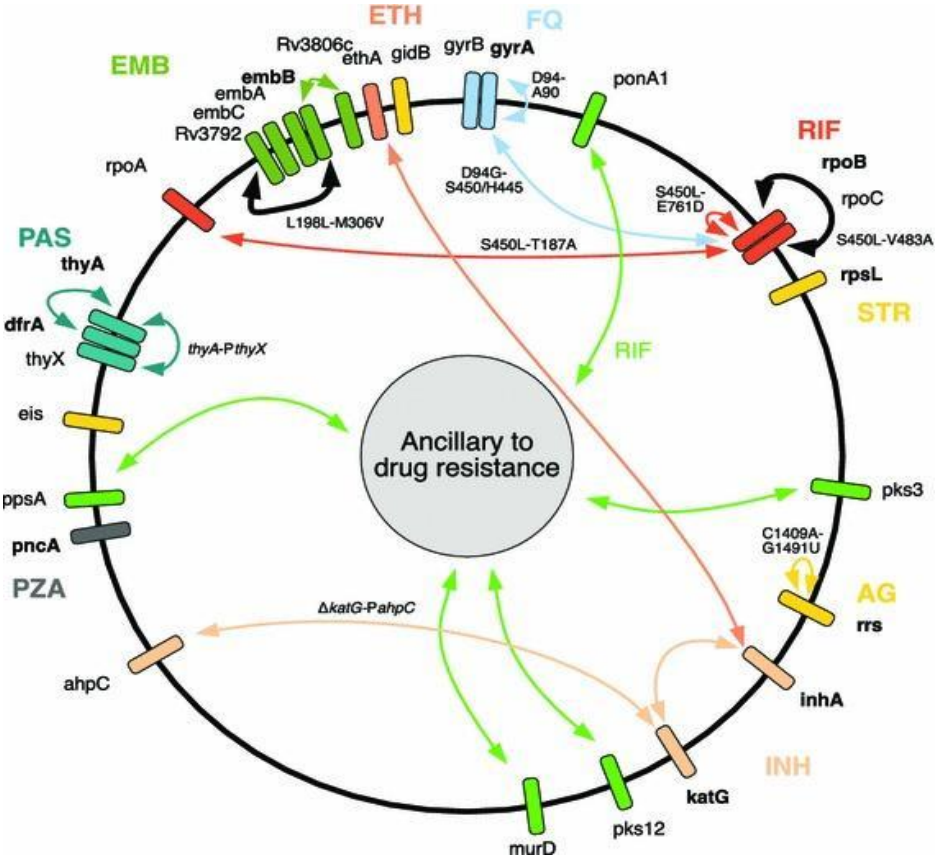

Figure 1. 3 Several gene interactions mediates drug resistance in Mtb. Genes are plotted according to their approximate genome position. Genes in bold are known to be directly associated with the DR phenotype. Lines represent putative epistatic interactions between DR genes and other secondary genes that play a role in drug resistance [17].

A number of different approaches in adopting WGS technologies in TB control have been formulated to date. Although these studies were from diverse research interests such as molecular epidemiology, phylogenetics, population genomics and pharmacology, the unifying aim of these studies was to have a deeper knowledge in the genetic mechanisms driving the evolution of drug resistance in Mtb [17]. In 2013, Köser *et al.* were one of the pioneers in the use of WGS in the rapid diagnosis of TB [174]. In their study they extracted and sequenced DNA directly from a positive MGIT tube. Their findings were not only consistent with pDST results (MDR-TB) but also went on further to detect mixed infection in the sample. The referral laboratory had reported resistance to nine antibiotics but WGS analysis further predicted resistance to five more antibiotics. This study highlighted the superiority of WGS in predicting antibiotic resistance in Mtb as it has potential to reduce the diagnostic time from weeks to only a few days, with the time taken to obtain positive culture being the only limiting factor [174]. Rodwell *et al.* from the Global Consortium for Drug–Resistant TB Diagnostics researched the feasibility of using mutations that were highly associated with the MDR/XDR-TB phenotype. In their study they analysed 417 Mtb isolates from

high burden regions in 4 countries. The team used Sanger sequencing to sequence eight genes (*katG, inhA, rpoB, gyrA, gyrB, rrs, eis, tlyA*) which are known to be highly associated with resistance to some of the commonly used TB antibiotics [176]. A review by Van Niekerk *et al.* highlighted some of the limitations of this approach such as the failure to account for the phylogenetic backgrounds of the investigated strains as well as the need for purifying the Mtb DNA and amplifying the genes of interests before identifying the presence of mutations [30]. Initial studies in the application of WGS in DR-TB management did not account for the role of interactions between mutations that may be associated with the DR-TB phenotype. A ground breaking study by Cui *et al.* used the bioinformatics software GBOOST to quantify and calculate the interactions of SNP pairs and identify gene pairs associated with drug resistance [177]. In their approach the team analysed two datasets that contained known DR strains as well as some pan-susceptible strains. A standard variant calling protocol was employed and the identified variants were filtered to remove phylogenetically related variants using the PLINK software [30, 177]. Using a *chi squared* test approach implemented in GBOOST, SNP-SNP interactions were identified. The output was filtered to select for non-synonymous mutations as well as gene pairs that contained at least one drug target gene. The study identified SNP pairs for INH, RIF, EMB, ETH but an interesting finding in this project was the abundance of gene pairs that consisted of the gene targets and the unique Pro-Pro-Glu (PPE) family of proteins which are abundant in the Mtb genome. Further research is needed to identify the possible role of these proteins in the development of DR-TB.

The plummeting cost of sequencing has also allowed research groups to study Mtb resistance on a much larger scale. One impactful study was done by Walker *et al.* [16] who did an analysis on 3651 drug resistant and susceptible Mtb genomes. Their study sought to predict resistance to eight first-line and second-line drugs using a compiled library of 232 genetic determinants of resistance in 23 candidate gene [165]. The results of this study showed high levels of accuracy with a mean specificity of 98% and 92% sensitivity. This suggests that our current knowledge on the genetic determinants of DR-TB can be reliably used to predict resistance to antibiotic drugs with great accuracy especially for the first-line drugs. A recent study by the 100,000 Genomes Project and Comprehensive Resistance Prediction for Tuberculosis (CRYPTIC) Consortium also supported these findings [165, 178]. Focusing on only first line drugs, the team analysed whole genome isolates of 10209 Mtb isolates using a comprehensive catalogue of mutations carefully curated from literature. Their results were consistent with the early findings of Walker *et al.* with a mean sensitivity of over 90%. One hallmark of this study was the improvement in the sensitivity of PZA predictions (91.3%) compared to 57% that was reported by Walker *et al.* [16, 165]. Despite these

high accuracy rates associated with using these known resistance mutations, our current catalogue of resistance determinants is still limited, up to 30% of INH resistant and 5% RIF resistant Mtb isolates do not harbour any mutations in the drug targets genes [130, 131]. Determinants of resistance still remain elusive for 10-40% of clinically resistant isolates [121]. Prediction of resistance to second line drugs and some complex drugs such as PZA is still challenging.

To address some of these challenges, genome-wide association study (GWAS) approaches have been implemented for Mtb [16, 165]. It is important to note that GWAS was initially designed to exploit the high rates of turnover and high throughput of human genomic data in order to identify variants in natural populations linked to phenotypic traits by statistical association [30]. The adoption of GWAS in bacterial studies has not been satisfactory due the nature of their population structures which reduces statistical power or produces false positives [179]. However the clonal nature of non-recombining microbes allows us to make associations between spurious variations and particular phenotypes. This has led to some researchers using GWAS approaches in order to identify novel resistance mutations. However, experimental validation of most of these GWAS predictions has not been performed [165]. Despite that, a number of new resistance determinants identified by these approaches have been successfully validated. The use of GWAS for antibiotic resistance in TB is further limited by the relatively low number of phenotypically resistant strains. Farhat *et al.* recently combined minimum inhibitory concentration (MIC) testing with GWAS to identify resistance genes in 1452 clinical Mtb isolates [180]. In their study they confirmed association at 13 non-canonical loci with two of them involving non-coding regions [180]. This novel approach was the first one that was able to quantify the proportion of the DR phenotype that is explained by genetic variation. The study further highlighted on how the complexity of the mechanisms driving the evolution of drug resistance limits the use of GWAS in drug resistance studies. For example, the step wise acquisition of DR mutations means that association conditioning is impossible in the absence of the primary gene mutation [180].

1.4.1 Challenges in using Whole Genome sequencing

Despite the huge strides that have been undertaken in adopting WGS in TB research, several limitations that can result in the misinterpretation of the data still exist within the current sequencing methodologies. Most of the WGS-based applications such as DR-TB diagnosis, population and phylogenetic analysis rely on the resolution of WGS in identifying SNPs. INDELs can be easily identified by the absence of expected reads or the presence of novel contigs relative

to the reference [147]. Several studies have also attributed the variability among mycobacteria to the loss of entire genes [181-183]. However, the identification of polymorphisms in repetitive regions, gene duplications chromosomal rearrangements and copy number changes of tandem repeats become challenging when using NGS methods and this can lead to a significant impact in the interpretation of the biological data especially when using single read technology [147]. The use of paired-end read technology addresses some of these limitations as it gives better resolution for these problematic regions.

Several repetitive regions are scattered within the Mtb genome, the sequencing of some of these regions is challenging when using short read sequencing technology [184]. This is problematic because several genes with significant biological meaning are located in some of these repeat regions. Some of these genes include the *pncA, esx,* PPE and PE whose involvement in Mtb drug resistance will be discussed in the following chapters. The detection of genomic duplications in Mtb is also challenging when using short read sequencing methods because of ambiguity in genome assembly [184]. Roberts *et al*. found out that some Beijing strains harbour a large scale duplication of ~350kb in the Rv3128-Rv3427 [147, 185]. This region is also the location of the transcriptional regulator *DosR* which regulates the response to hypoxia. Using short read sequencing technology for a region like this would result in several ways of building contigs making it difficult to identify the polymorphism. Advances in longer read sequencing methods as well as improved assembly technologies allows us to overcome some of these challenges [186].

1.4.2 WGS-based software tools for the prediction of drug resistance in Mtb

1.4.2.1 Mykrobe Predictor TB version 0.1.3

The software tool designed for Mtb as well as *Staphylococcus aureus* takes raw sequence data as input and generates a user friendly report within three minutes on a personal computer [187]. The tool is specifically designed to detect low-frequency resistance mutations and detected minor alleles in the DR loci are included in the output report even though they are not interpreted [188]. The program and its source code can be freely downloaded on Github (https:\\github.com\iqbal-lab\Mykrobe-predictor\release). Despite Mykrobe Predictor being an automatic tool, the end-user is still required to use the command line when using the program for batch uploads. The program also requires sequence files to be merged before they are analysed and this may cause some technical challenges for the end-user.

1.4.2.2 CASTB version 1.5

The comprehensive analysis server for the *Mycobacterium tuberculosis* complex is a web server (http://castb.ri.ncgm.go.jp/CASTB) that can analyse Mtb WGS data. The server uses a SNP catalogue to infer drug resistance to six drugs which are INH, EMB, RIF, PZA and ciprofloxacin (CFX) [188, 189]. The prediction result output (R=resistant; <blank>=not resistant) does not include a report of which variants were detected or how they were interpreted. Results are stored on the server for up to seven days. A recent review by Schleusener *et al.* on WGS-based tools criticized CASTB on its heavy reliance on automation and its lack of detailed output reports [188].

1.4.2.3 TBProfiler

TBProfiler (http:\\tbdr.lshtm.ac.uk) is another web-server for the prediction of antibiotic resistance in Mtb directly from raw sequence data [9]. The tool uses a catalogue of 1325 polymorphisms (including SNPs and indels) to infer the DR profiles for 11 anti-TB drugs [188]. The output also includes additional information on the phylogenetic background of the queried strains as well as further mutations in 22 candidate genes. However, TBProfiler does not offer any dedicated export and storage functionality [188]. An offline version of TBProfiler can be downloaded although data querying in batch mode requires proficiency in using the command-line. Updating the tool also requires advanced computational skills as it involves modifying the program's source code.

1.4.2.4 KvarQ

KvarQ (http://github.com/kvarq/kvarq/releases) is a software tool that scans fastq files for known variants [190]. The tool requires Python and a C compiler as dependencies [188]. Unlike other software that analyse raw sequence data as input, the tool extracts the required information directly from the sequencing reads without any need to align every read to a reference genome. Users can interact with the tool via the command-line or through the graphical user interface. KVarQ output is given in the 'JavaScript Object Notation' (json) format [188]. Updating the resistance catalogue also requires advanced skills as this involves changing the python files which make up the program's test-suites.

1.4.2.5 PhyResSE version 1.0

PhyResSE (http://phyresse.org) is a free to use web server that processes raw WGS data to predict for lineage and drug resistance profiles in Mtb [191]. PhyResSE integrates well established methods from FastQC, BWA, QualiMap, SAM tools and others with a broad spectrum of experimentally validated cases acquired at the Leibniz Lung Centre of Research Centre Borstel [191]. The tool‘s mutation library consists of 301 polymorphisms that are used to predict for drug resistance as well as 239 SNPs in 135 genes used for phylogenetic typing [188]. The tool offers a detailed output (batch mode) report of all mutations in csv file format. Unlike the other tools, PhyResSE allows the users to easily update its catalogue of variants [188]. This allows users to improve the accuracy of the program as they gain more understanding on the genetic basis of drug resistance in Mtb.

Table 1. 2 A summary of current WGS-based programs for the detection of DR in Mtb [188].

Feature	CASTB (Version 1.1)	KvarQ (Version 0.12.2)	Mykrobe Predictor TB (Version 0.1.3)	PhyResSE (Version 1.0)	TBProfiler
Web-based	Yes (registration needed)	No	No	Yes	Yes
Batchmode	No	Yes	In command-line version only	Yes	In command-line version only
Paired-end reads	Yes	Merged files only	Merged files only	Yes	Yes
Pipeline	Velvet de-Novo assembly; BIGSdb; MUMmer mapping; custom scripts	Python modules/ packages with C extensions (no mapping)	Stampy mapping (H37Rv Version 2); variant calling SAMtools	BWA mapping(H37Rv Version 3), pre-processing & variant calling with GATK	snap mapping (selected regions of H37Rv version 3); variant calling SAMtools
Plain language report	Yes	No	Yes	Yes, detailed	Yes
Data export	Yes	Yes	Yes	Yes	No
Basis of lineage prediction	Virtual LSP, *in silico* spoligotyping, RFLP or MIRU-VNTR	Comas *et al.* 2009, Stucki *et al.* 2012 and unpublished	Stucki et al. 2012	Homolka *et al.* 2012, Coll *et al.* 2014 and other published	Coll *et al.* 2014
Detection of NTM	Yes	No (MTBC only)	Yes	No (MTBC only)	No (MTBC only)
Variants reported	None	Only resistance mutations	Only resistance mutations	All mutations	All mutations in candidate genes
Detection of mixed infections	Only Beijing-non Beijing	Yes	No	Yes	Yes
Hetero resistance	No	Visual	Yes, without quality scores	>10%	Yes
Modification possible	No	Yes	Yes	Yes	Yes

1.5 Summary

Although it is one of the oldest diseases known to man, tuberculosis remains the leading global infectious killer. Despite a drop in the global incidence of the disease in the recent years, the persistence of drug resistant strains remains a stumbling block in achieving the goal of totally eradicating TB. Early diagnosis and correct treatment remain as limiting factors in TB control especially in low to middle income countries which carry a large portion of the disease burden. Referral laboratories in these settings still rely on traditional DST methods for DR-TB diagnosis. Rapid molecular assays have been successfully developed for DR-TB diagnosis and this has led to the WHO endorsing the Xpert-Ultra MTB assay. The global roll out of this technology has led to the improvement in DR-TB diagnosis, however, serious diagnosis gaps still remain due to the limitations of these technologies. The elucidation of the Mtb genome was a game changing milestone in TB research as it provided a more comprehensive view on Mtb physiology. Whole Genome sequencing has proved to be an attractive option in informing TB treatment decision as well as monitoring drug resistance patterns in high burden settings. However for us to be able to fully harness its potential, there is need for a thorough understanding of the genetic mechanisms

of resistance in Mtb. Several WGS-based assays for the diagnosis and management of DR-TB have been successfully developed over recent years. However, the uptake of these technologies in the referral laboratory setting has been hindered by the data complexities associated with WGS as well as a lack of bioinformatics skills among clinical microbiologists. The recent years have seen a number of WGS-based user-friendly platforms being developed. Despite an improvement in the user-experience, the success of these platforms depend on our understanding of the genetic basis of drug resistance in Mtb.

1.6 Project Overview

Traditional dogma has always assumed that resistance to an anti-TB drug can be attributed to a once-off acquisition of a resistance-conferring mutation. In this study we hypothesized that the antibiotic resistant phenotype in Mtb is associated with a specific pattern of multiple polymorphic sites in the Mtb genomes which can be used as reliable signatures of drug resistance even if only fragments of the genome are available for analysis. We further hypothesized that the step wise acquisition of drug resistance mutations follows an order of events whereby the acquisition of one mutation facilitates the acquisition of other additional mutations leading to the further development of drug resistance. The objective of this project is to develop an online tool for the prediction of drug resistance in *Mycobacterium tuberculosis* using NGS data and to develop an evolutionary model of the emergence of Mtb drug resistance. Our clade-based approach will employ a novel GWAS derivative to calculate the power of association between a clade and resistance to a specific anti-TB drug. This approach will allow for the identification of novel polymorphisms which may be associated with the DR-TB phenotype. Drug susceptibility is often reported as either "Resistant" or "Susceptible" and this often results in the incorrect classification of strains with low levels of resistance. Our program will report resistance as a probability and users will be able to identify strains that might develop full resistance in future. The program will allow users to upload files in different NGS file formats generated from different stages of the genome completion process. Program design and implementation are covered by Chapter 2.

In Chapter 3, we will use attributable risk statistics to further identify functional associations between mutations that were strongly associated with the DR phenotype. Attributable risk networking will then be used to infer the evolutionary pathways that drive drug resistance in different clades of Mtb.

The aim of this project is to elucidate the mechanisms of mutations in the Mtb genome with the bigger goal of gaining more understanding of the dynamics that drive the evolution of drug resistance in this pathogen. The main goals of this project include:

1. To optimize the accuracy of an alpha version of the program "Resistance Sniffer" so that it identifies the lineage as well as the drug susceptibility profile of Mtb strains and test its validity using real life Mtb sequences with well-characterized drug susceptibility and phylogenetic profiles provided by the partnering team from the TB platform of the SA Medical Research Council.
2. To expand the number of antibiotics that can be analysed by the Resistance Sniffer program and allow processing of genome sequence datasets in different file formats including raw NGS-generated fastq read files, genomic sequences in FASTA and GenBank formats as well as VCF variant calling files to cater for the needs of different potential users.
3. To identify clade-specific patterns of polymorphisms and confirm the role of compensatory mutations in drug resistance.
4. To provide the Resistance Sniffer program with a functionality of predicting the likelihood of a given TB strain to acquire drug resistance in future by identifying specific mutations preceding the development of antibiotic resistance.
5. Develop an evolutionary model for the emergence and fixation of Mtb drug resistance.

Chapter 2-Resistance Sniffer development and Implementation

2.1 Introduction

In the previous chapter, we highlighted on how the success of WGS sequencing in the control of TB is mainly dependent on our understanding of the genetic basis of drug resistance in Mtb. In order to fully harness the power of WGS, there is a need to set clear rules on the interpretation of variants detected using these approaches [192]. Ideally, the goal would be to assemble a high-quality library of genetic determinants of resistance which can then be implemented in DR-TB diagnostics platforms. For first line drugs such as INH and RIF whose mechanism of action have been extensively studied, the current body of knowledge is sufficient enough to distinguish between mutations that are associated with the drug-resistant phenotype and those that are not. However, for second line drugs and other drugs with complex mechanisms of action identification of such variants becomes challenging. Although *in vitro* allelic exchange experiments have been successfully used to identify significant resistance variations, this approach is not feasible for DR-TB diagnosis due to a number of limitations [193]. These methods are expensive, tedious and also limited to a small number of loci. *In silico* association studies have been successfully used in numerous settings to interrogate suspected genetic determinants of resistance particularly in nonessential genes, where numerous loss of function mutations can lead to the development of drug resistance [16, 192].

Despite the popularity of Genome-Wide Association Studies (GWAS) in identifying variants in natural populations linked to phenotypic features by statistical association, its use in bacterial populations has been minimal due to challenges that arise as a result of bacterial population structures [30]. Genomic diversity in bacteria may be moulded by population stratification, a phenomenon where there is an occurrence of subgroupings of strains that are on average more related to each other than other individuals in the broader population [179, 194]. This leads to spurious associations whereby associations are a result of genetic proximity rather than due to the phenotype of interest. Population stratification is widely rampant in highly clonal bacteria such as Mtb, and in other bacterial species with separate geographic or host-associated subpopulations [179, 195].

The use of traditional GWAS in Mtb studies is complicated by its high linkage disequilibrium (LD) as well as strong population structure due to its high clonality [196]. Population stratification

can lead to false associations while correcting for the problem can lead to a decrease in association power.

In this study we used a novel GWAS derivative as well as a larger sample size to control population structure as well as boosting association power. This approach allows for the detection of other novel resistance determinants which have not been reported yet.

2.2 Data download and preparation

Mutation data in the form of 2501 variant call format (VCF) files were downloaded from the GMTV database (https://mtb.dobzhanskycenter.org/cgi-bin/beta/main.py#custom/world). The database consists of data from *Mycobacterium tuberculosis* isolates sourced from different regions of the Russian Federation and worldwide. The database integrates drug resistance profiles, epidemiology, TB clinical outcome, year and place of isolation as well as molecular biology data [197]. The metadata includes information on drug resistance trials with respect to the following antibiotics: amikacin (AMK), capreomycin (CM), cycloserin (CS), ethambutol (EMB), ethionamide (ETH), isoniazid (INH), fluoroquinolones (FLQ), kanamycin (KAN), ofloxacin (OFL), para-amino salicyclic acid (PAS), pyrazinamide (PZA), rifampicin (RIF) and streptomycin (SM). The database also provides information on the phylogenetic clade of each sample. The quality of the microbiological, WGS and spoligotyping data is guaranteed by the institutions that provided the data to the database [197]. The dataset was further split to create a training dataset of 1300 samples. The validation dataset consisted of 1201 samples whose antibiotic phenotype data was available for all the drugs included in this study. An independent testing dataset of 742 Mtb genome sequences was obtained from the SA-MRC. We also obtained an additional testing dataset for lineage classification which consisted of 77 Mtb strains isolated in Sierra Leone (ENA accession number: PRJEB7727) from the PATRIC database [198]. Strains from this database have been described in previous studies [188, 191, 199].

2.3 Phylogenetic Lineage classification

M. tuberculosis H37Rv reference genome (NC_000962.3) was used to determine the polymorphisms. In addition to this data, discriminative single nucleotide polymorphisms (SNPs) were identified by whole genome alignment against reference genomes of *M. bovis* (NC_016804, NC_020245, NC_012207, NC_008769 and NC_002945) and *M. canettii* (NC_015848,

NC_019950, NC_019965, NC_019951 and NC_019952) available from the NCBI database. Variant calling was performed using Mauve 2.3.1 [200].

2.4 Construction of the diagnostic key

In total, 58,025 SNPs and indels leading to amino acid substitutions were selected for the analysis of their associations with Mtb clades and antibiotic resistance patterns. Allelic states and locations of SNPs associated with antibiotic resistance were obtained from the TB Drug Resistance Mutation Database (https://tbdreamdb.ki.se/Info/Default.aspx) [201].

Discriminative power of SNPs used for distinguishing between Mtb clades and/or drug sensitive versus drug resistant variants in the same clade was calculated by the following equation:

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

where A∩B is the number of strains in the clades *A* and *B* sharing the same allelic state of the locus *k*; N_A and N_B – sample sizes of the clades *A* and *B*, respectively. Power values were in the range from 0 to 1.

The output was filtered down to exclude polymorphisms in the PPE and PGRS gene as well as well-known lineage markers. SNPs with the highest discriminative power values were then selected to create the diagnostic key as explained below.

2.5 Resistance Sniffer algorithm

The program, Resistance Sniffer, was developed in Python 2.7 (also compatible with Python 2.5) and implemented as an on-line tool at http://resistance-sniffer.bi.up.ac.za/. The program is also available for download, with example input files, from http://resistance-sniffer.bi.up.ac.za/Mycobacterium_tuberculosis/help/ as a stand-alone tool. The accepted input includes complete sequences in Genbank or FASTA formats; sequences of predicted genes or proteins in FASTA format, uncompressed VCF files and raw Illumina *fastq* paired-end read files. The program maps raw sequences to the embedded reference genome sequence (*M. tuberculosis* H37Rv, NC_000962.3). The detected patterns of polymorphisms are then processed using the diagnosis key, which consists of a catalogue of clade-specific polymorphisms and genetic determinants of antibiotic resistance. The diagnosis key consists of bifurcating splits for each decision point (Figure 2.1). At each intermediate node, the program calculates normalized counts

of power values (Eq. 1) of diagnostic polymorphisms depending on the states of these sites in the given genome. As the program was designed to perform predictions based on partially sequenced genomes, the program does not expect to receive the states of all polymorphic sites assigned for a split and tries to make a decision based on the available sites. Optimally, the score for one bifurcating branch is expected to be 1.0, and for another branch – 0.0. If the maximal score is below 0.75, the program explores both alternative branches to avoid an erroneous decision on a top-level split. Moreover, reaching the leaf-node corresponding to an Mtb clade, the program tries a possibility that the strain may belong to a sister clade sharing similar polymorphisms. It must be emphasized that in this work we did not attempt to distinguish between phylogenetically significant traits and convergent polymorphisms. No conclusions regarding the phylogenetic relatedness between clades should be drawn from the neighbouring of the clades in the diagnostic key in Figure 2.1.

It should be noted that in many cases there are no clear borders between Mtb clades and intermediate strains do exist, hence in instances where the program cannot reach a confident conclusion with regards to strain affiliation, the program returns two top-scored clades.

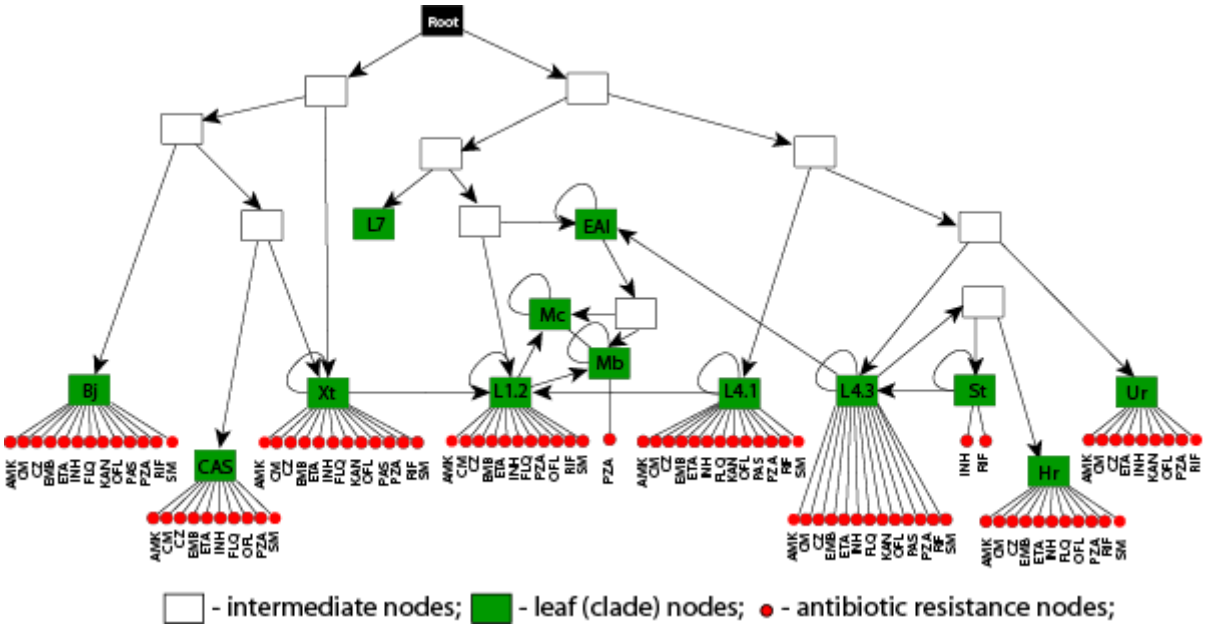

Figure 2. 1 The decision tree implemented in Resistance Sniffer. Leaf nodes shown in green denote Mtb clades: EAI – East African and Indian (Lineage 1.1); L1.2 – lineage 1.2; Bj – Beijing strains (lineage 2); CAS – Central Asian Strains (lineage 3); Xt – X-type strains; L4.1 – lineage 4.1 (H37Rv type strain); Ur – Ural strains (lineage 4.2); L4.3 – lineage 4.3; Hr – Haarlem strains (subtype of lineage 4.3); St – S-type (subtype of lineage 4.3); L7 – lineage 7; Mb – M. bovis and Mc – M. canettii (related to lineages 5 and 6). Intermediate nodes represent groups of clades. Antibiotic resistance nodes are: amikacin (AMK), capreomycin (CM), cycloserin (CS), ethambutol (EMB),

ethionamide (ETH), isoniazid (INH), fluoroquinolones (FLQ), kanamycin (KAN), ofloxacin (OFL), para-amino salicyclic acid (PAS), pyrazinamide (PZA), rifampicin (RIF) and streptomycin (SM).

Contrary to a general belief in the existence of DR mutations common for all Mtb strains, this approach proceeds from an assumption of parallel drug resistance evolution in Mtb clades which resulted in the creation of different clade-specific patterns of polymorphic sites associated with the antibiotic resistance phenotype [30]. Each clade node of the diagnostic key consists of associated sets of polymorphic sites which distinguish between antibiotic resistant and antibiotic sensitive variants for every Mtb clade. Using the same method described in the clade identification step above, the program calculates normalized counts of polymorphisms associated with the drug sensitivity (*SenCount*) and drug resistance (*ResCount*). In the next step, antibiotic resistance scores (*q* values) are calculated for every individual antibiotic by equation 2.

$$q = \frac{1+\left(\frac{1+ResCount}{1+SenCount}\right)}{2} \qquad (2)$$

In the following step, the resistance value (*R*) and the standard error (*Err*) are calculated by equations 3 and 4, respectively.

$$R = q \times 2^{(2\times ResCount-1)} \qquad (3)$$

$$Err = \frac{2q(1-q)\times 2^{(2\times ResCount-1)}}{\sqrt{N-1}} \qquad (4)$$

In Eq. 4, *N* is the number of diagnostic sites found in the given genome.

Figure 2.1 details the sequence of steps taken by the program in assigning the clade to the sample. This is followed by determining whether the strain is resistant or susceptible to each of the antibiotics. It must be noted that the number of antibiotics a strain may be resistant to depends on the clade affiliation of the given strain. East African and Indian (EAI) clade, Lineage 7 and *M. canettii* are considered to be sensitive to all antibiotics by default since the antibiotic metadata from GMTV and SA MRC indicated that none of the isolates belonging to these clades were resistant to any antibiotic. However, this does not rule out the possibility of the future discovery of antibiotic resistance in some of these strains. *M. bovis* isolates are set by default to be resistant

to PZA [202] while sensitive to all the other antibiotics [203]. Antibiotic resistance diagnostic keys will be added to these nodes when more data becomes available.

Program validation was performed on 1201 Mtb strains from GMTV and 742 strains from SA MRC, which were provided with antibiotic resistance/susceptibility patterns. The program performance was characterized by sensitivity (SENS), specificity (SPEC), Positive Predictive Value (PPV) and Negative Predictive Value (NPV) as shown in equations 5-8:

$$SENS = \frac{TP}{TP+FN} \quad (5)$$

$$SPEC = \frac{TN}{FP+TN} \quad (6)$$

$$PPV = \frac{TP}{TP+FP} \quad (7)$$

$$NPV = \frac{TN}{TN+FN} \quad (8)$$

In Eqs. 5-8, *TP* – number of true positive predictions; *FP* – false positives; *TN* – true negatives; and *FN* – false negatives.

2.5.1 Program Interface

The complexity of bioinformatics tools often deters would-be users from adopting WGS-based tools in the clinical setting. In this study we sought to develop a user friendly platform that can be easily accessed online at http://resistance-sniffer.bi.up.ac.za/Mycobacterium_tuberculosis. The link directs the users to the home page which contains information about the tool's usage instructions as well as links to the offline version as shown in Figure 2.2 below. The end-user interacts with the system by simply clicking on the browse button. This opens up a dialog box that allows the user to select the desired input file from the local hard drives. Once selected, the file is uploaded and processed by clicking on the upload and compare button as shown below. In this project we aimed to develop a tool that is applicable to different stages of the genome completion process and hence the tool is compatible with most of the standard NGS file types from raw reads to SNP level. The user interface provides information on the accepted input file formats. The interface also allows the user to provide their email address if they wish to be notified once the analysis has been completed. The *fastq* file upload menu allows the user the option to either upload single read files or use paired-end read files as input.

The interface also comes with a download menu which allows the users to download an offline version of the tool. This stand-alone version allows for the batch upload of genomic sequences. This improvement in automation can be useful in a large referral laboratory setting which has to process larger amounts of sequence data. Figure 2.3 shows the help and download button for the program where users can access a compressed version of the program for offline use. Figure 2.4 shows the structure of the subfolders that make up the offline version of the program. Assuming that the user of the offline version has basic knowledge of the command-line, the queried files are copied to the input folder as shown in Figure 2.5. The program can be run on any computer provided that Python 2.5 or 2.7 is installed by executing the *run.py* script shown in Figure 2.4. Assuming that there are no errors and the program is running well, the user's command line shell will display the progress of the analysis as shown in Figure 2.6. Upon completion, the program writes the output files in the output folder shown in Figure 2.4.

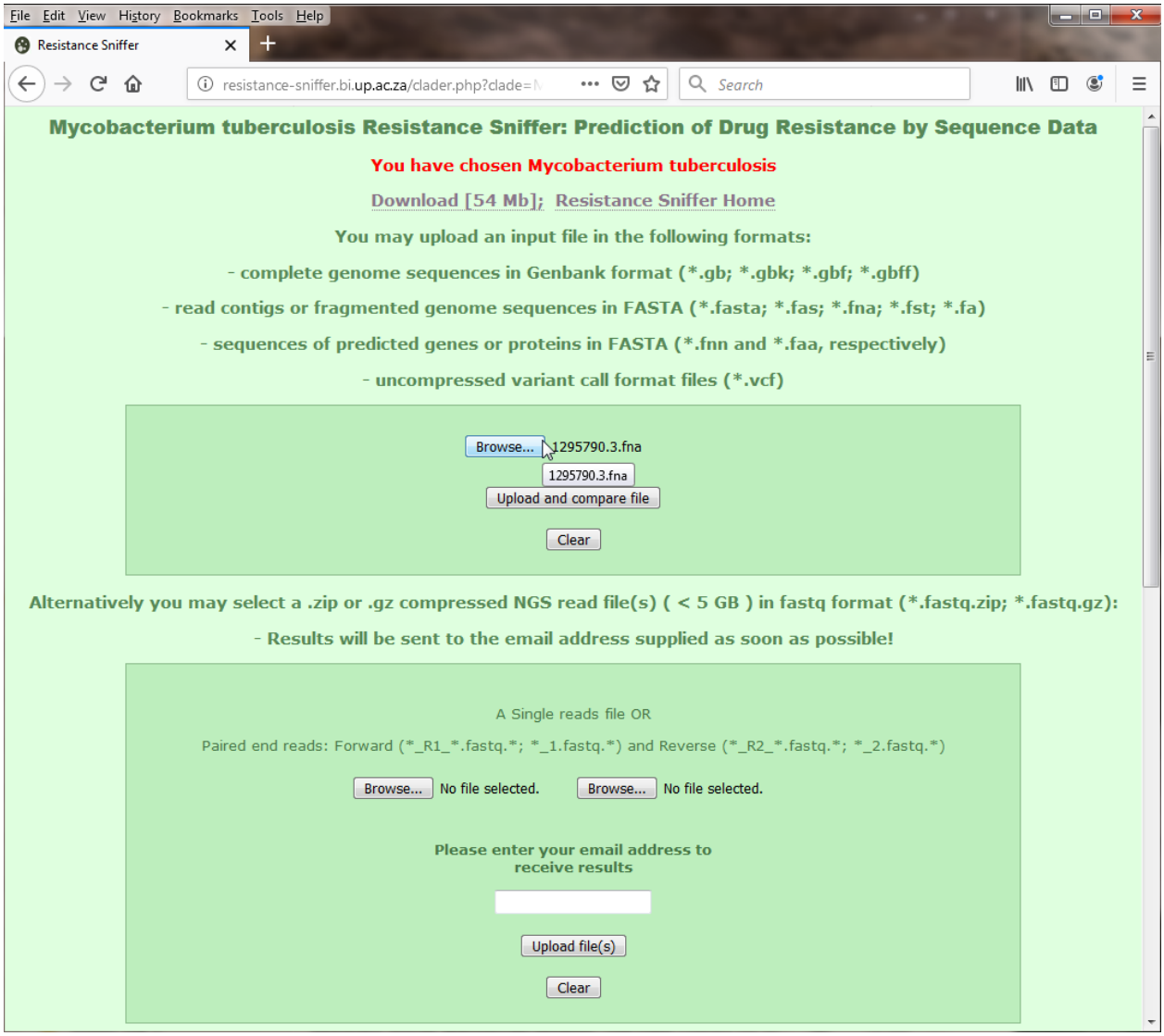

Figure 2. 2 Web user interface of Resistance Sniffer depicting the accepted input file types

.

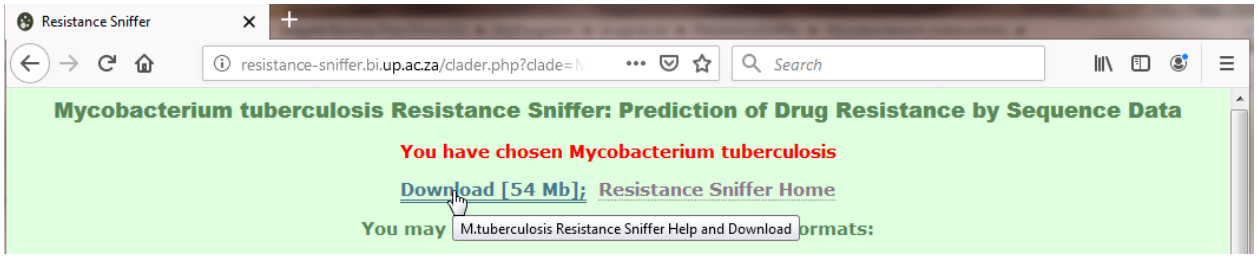

Figure 2. 3 Local version of the program is available from the Help and Download Web-site.

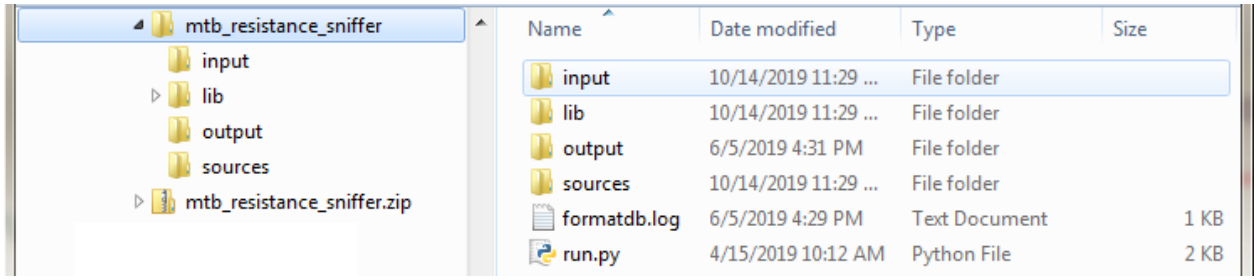

Figure 2. 4 Structure of subfolders of the local version of the program Mtb_resistance_sniffer.

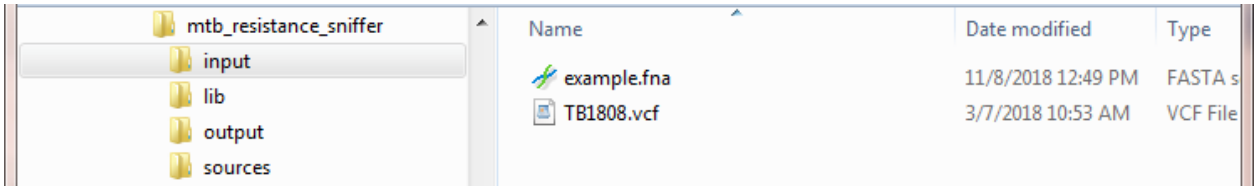

Figure 2. 5 Example input files in the folder "input".

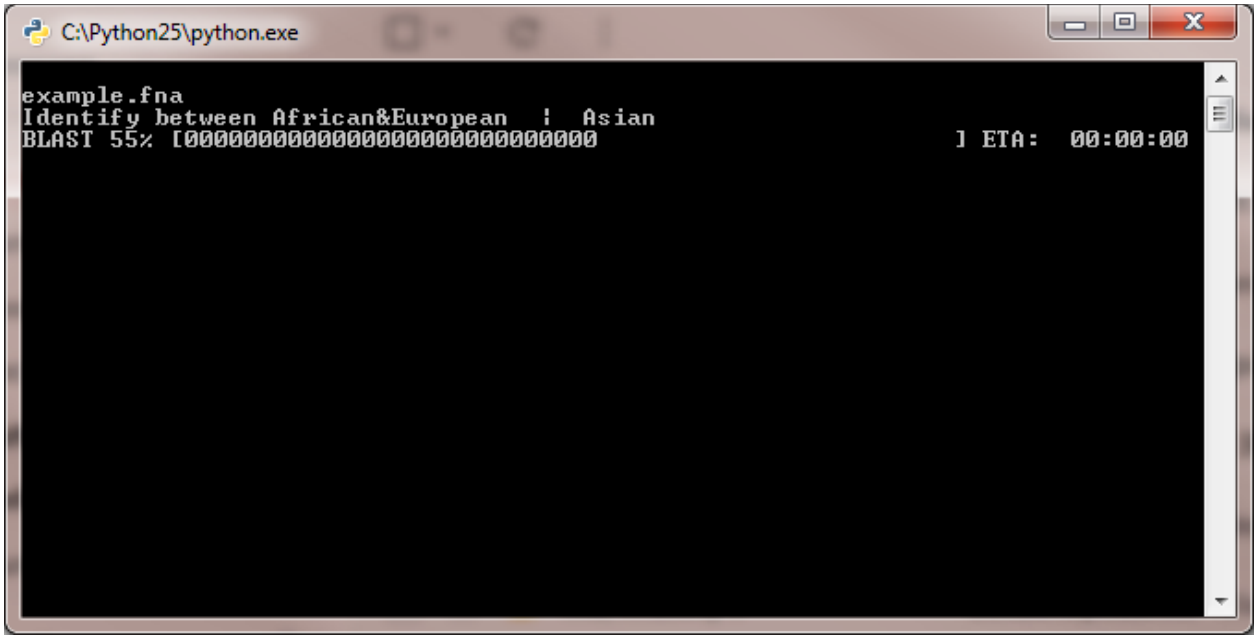

Figure 2. 6 Program run in the Command Prompt window.

Figure 2.7 below shows a schematic diagram of the Resistance Sniffer workflow. The program processes the input files depending on their formats and the identified variants in the genotyping module are compared against the identification table which makes up the diagnosis key. The data is then processed by the statistical evaluation module. It is important to note that the processing of raw *fastq* reads is not available on the offline version of the program because the Bowtie 2 aligner is embedded on the server side. The identification table is in simple text format as shown in Figure 2.8, this allows for flexibility in the application of the program meaning that the program can be easily updated to include newly discovered variants. The program can be easily adapted for other pathogens as well by simply creating an identification table applicable for that specific pathogen.

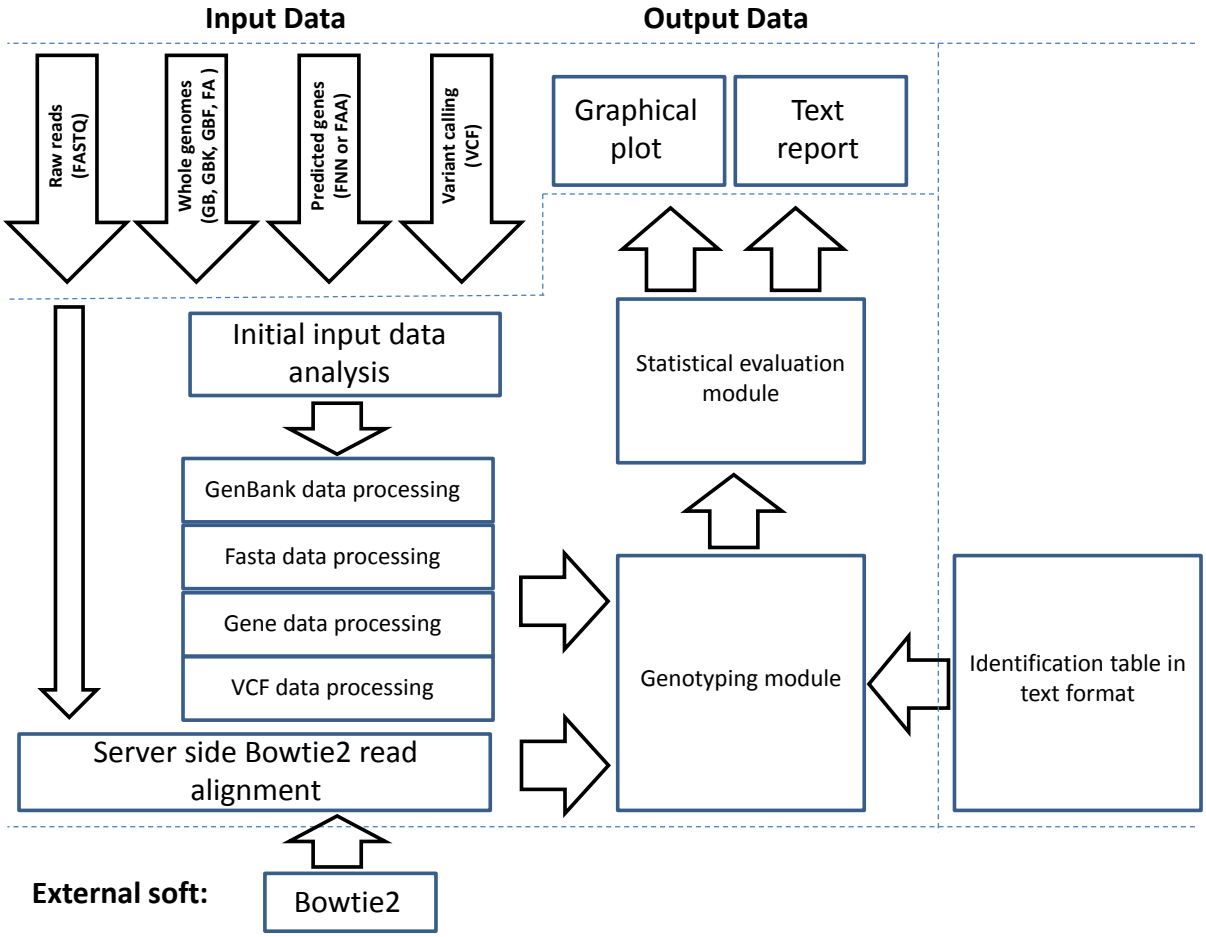

Figure 2. 7 Schematic diagram of the Resistance Sniffer workflow.

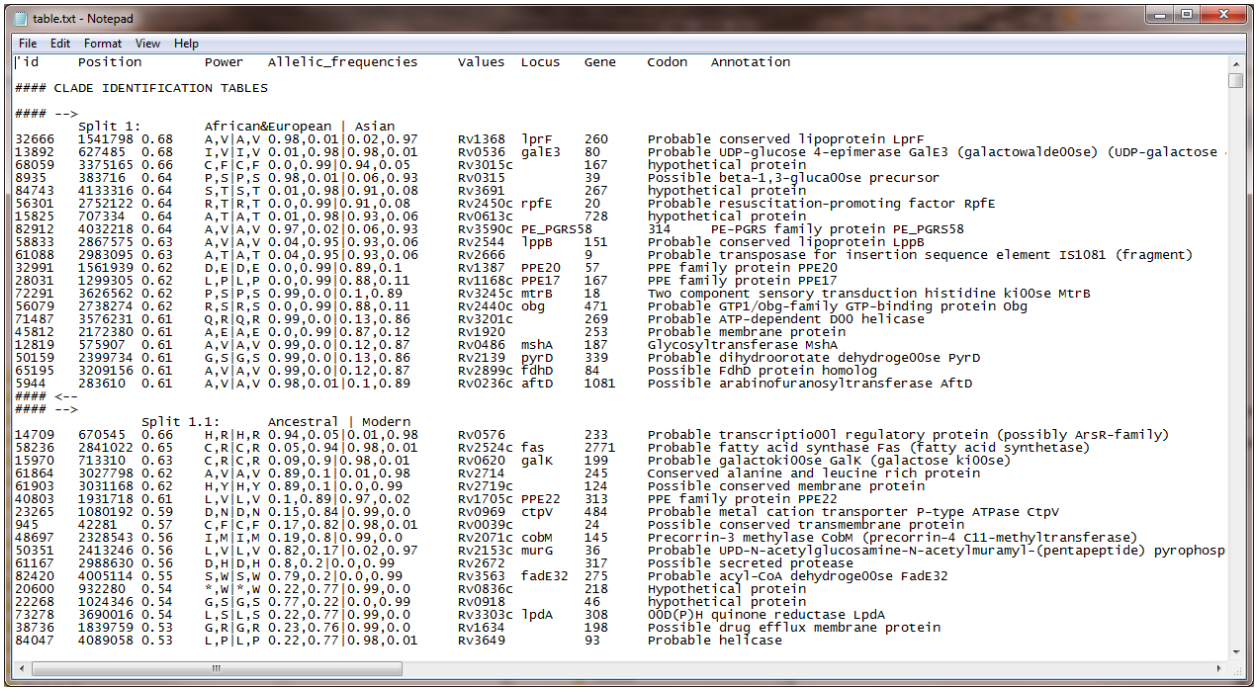

table.txt - Notepad

File Edit Format View Help

```
'id     Position        Power   Allelic_frequencies     Values  Locus   Gene    Codon   Annotation

#### CLADE IDENTIFICATION TABLES

#### -->
        Split 1:        African&European | Asian
32666   1541798 0.68    A,V|A,V 0.98,0.01|0.02,0.97     Rv1368  lprF    260     Probable conserved lipoprotein LprF
13892   627485  0.68    I,V|I,V 0.01,0.98|0.98,0.01     Rv0536  galE3   80      Probable UDP-glucose 4-epimerase GalE3 (galactowalde00se) (UDP-galactose
68059   3375165 0.66    C,F|C,F 0.0,0.99|0.94,0.05       Rv3015c         167     hypothetical protein
8935    383716  0.64    P,S|P,S 0.98,0.01|0.06,0.93      Rv0315          39      Possible beta-1,3-gluca00se precursor
84743   4133316 0.64    S,T|S,T 0.01,0.98|0.91,0.08      Rv3691          267     hypothetical protein
56301   2752122 0.64    R,T|R,T 0.0,0.99|0.91,0.08       Rv2450c rpfE    20      Probable resuscitation-promoting factor RpfE
15825   707334  0.64    A,T|A,T 0.01,0.98|0.93,0.06      Rv0613c         728     hypothetical protein
82912   4032218 0.64    A,V|A,V 0.97,0.02|0.06,0.93      Rv3590c PE_PGRS58       314     PE-PGRS family protein PE_PGRS58
58833   2867575 0.63    A,V|A,V 0.04,0.95|0.93,0.06      Rv2544  lppB    151     Probable conserved lipoprotein LppB
61088   2983095 0.63    A,T|A,T 0.04,0.95|0.93,0.06      Rv2666          9       Probable transposase for insertion sequence element IS1081 (fragment)
32991   1561939 0.62    D,E|D,E 0.0,0.99|0.89,0.1        Rv1387  PPE20   57      PPE family protein PPE20
28031   1299305 0.62    L,P|L,P 0.0,0.99|0.88,0.11       Rv1168c PPE17   167     PPE family protein PPE17
72291   3626562 0.62    P,S|P,S 0.99,0.0|0.1,0.89        Rv3245c mtrB    18      Two component sensory transduction histidine ki00se MtrB
56079   2738274 0.62    R,S|R,S 0.0,0.99|0.88,0.11       Rv2440c obg     471     Probable GTP1/Obg-family GTP-binding protein Obg
71487   3576231 0.61    Q,R|Q,R 0.99,0.0|0.13,0.86       Rv3201c         269     Probable ATP-dependent D00 helicase
45812   2172380 0.61    A,E|A,E 0.0,0.99|0.87,0.12       Rv1920          253     Probable membrane protein
12819   575907  0.61    A,V|A,V 0.99,0.0|0.12,0.87       Rv0486  mshA    187     Glycosyltransferase MshA
50159   2399734 0.61    G,S|G,S 0.99,0.0|0.13,0.86       Rv2139  pyrD    339     Probable dihydroorotate dehydroge00se PyrD
65195   3209156 0.61    A,V|A,V 0.99,0.0|0.12,0.87       Rv2899c fdhD    84      Possible FdhD protein homolog
5944    283610  0.61    A,V|A,V 0.98,0.01|0.1,0.89       Rv0236c aftD    1081    Possible arabinofuranosyltransferase AftD
#### <--
#### -->
                Split 1.1:      Ancestral | Modern
14709   670545  0.66    H,R|H,R 0.94,0.05|0.01,0.98      Rv0576          233     Probable transcriptio00l regulatory protein (possibly ArsR-family)
58236   2841022 0.65    C,R|C,R 0.05,0.94|0.98,0.01      Rv2524c fas     2771    Probable fatty acid synthase Fas (fatty acid synthetase)
15970   713310  0.63    C,R|C,R 0.09,0.9|0.98,0.01       Rv0620  galK    199     Probable galactoki00se GalK (galactose ki00se)
61864   3027798 0.62    A,V|A,V 0.89,0.1|0.01,0.98       Rv2714          245     Conserved alanine and leucine rich protein
61903   3031168 0.62    H,Y|H,Y 0.89,0.1|0.0,0.99        Rv2719c         124     Possible conserved membrane protein
40803   1931718 0.61    L,V|L,V 0.1,0.89|0.97,0.02       Rv1705c PPE22   313     PPE family protein PPE22
23265   1080192 0.59    D,N|D,N 0.15,0.84|0.99,0.0       Rv0969  ctpV    484     Probable metal cation transporter P-type ATPase CtpV
945     42281   0.57    C,F|C,F 0.17,0.82|0.98,0.01      Rv0039c         24      Possible conserved transmembrane protein
48697   2328543 0.56    I,M|I,M 0.19,0.8|0.99,0.0        Rv2071c cobM    145     Precorrin-3 methylase CobM (precorrin-4 C11-methyltransferase)
50351   2413246 0.56    L,V|L,V 0.82,0.17|0.02,0.97      Rv2153c murG    36      Probable UPD-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide) pyrophosp
61167   2988630 0.56    D,H|D,H 0.8,0.2|0.0,0.99         Rv2672          317     Possible secreted protease
82420   4005114 0.55    S,W|S,W 0.79,0.2|0.0,0.99        Rv3563  fadE32  275     Probable acyl-CoA dehydroge00se FadE32
20600   932280  0.54    *,W|*,W 0.22,0.77|0.99,0.0       Rv0836c         218     Hypothetical protein
22268   1024346 0.54    G,S|G,S 0.77,0.22|0.0,0.99       Rv0918          46      hypothetical protein
73278   3690016 0.54    L,S|L,S 0.22,0.77|0.99,0.0       Rv3303c lpdA    308     00D(P)H quinone reductase LpdA
38736   1839759 0.53    G,R|G,R 0.23,0.76|0.99,0.0       Rv1634          198     Possible drug efflux membrane protein
84047   4089058 0.53    L,P|L,P 0.22,0.77|0.98,0.01      Rv3649          93      Probable helicase
```

Figure 2. 8 The diagnosis key of the Resistance Sniffer

2.5.2 Output visualization

Most DR-TB databases approach the drug resistant phenotype as a binary entity which means that a strain is classified as either resistant or susceptible. However, our study suggests that the progression to the drug resistant phenotype is a stepwise process which highlights the need to develop ways to account for intermediate levels of drug resistance as well. Resistance Sniffer outputs the results as a bar plot of the probability that the strain is drug sensitive or drug resistant to the thirteen antibiotics. Figure 2.9 shows several examples of graphical outputs of the program. The drug susceptibility pattern estimated for the strain TB0775 from GMTV is demonstrated in Figure 2.9A. The strain was predicted to belong to the Beijing clade. The program prediction shows that this isolate has a high likelihood to be resistant to INH, KAN and SM, and may have an intermediate resistance towards FLQ, OFL, PAS and RIF. The experimentally detected profile of drug susceptibility available for this strain from GMTV confirms resistance to INH, SM and RIF, and sensitivity to EMB which agrees with the software prediction. This strain was not tested for other antibiotics.

Figure 2.9B shows the prediction of drug resistance for a highly fragmented assembly of a clinical isolate from the SAMRC. The strain was assigned with equal likelihoods of belonging to either CAS or to X-type. It may be possible that the queried sequence is from an intermediate variant; however, the ambiguity could be attributed to the quality of sequencing. Only small fractions of diagnostic sites were found in the sequences which resulted in an increased standard error of *R*-value estimation depicted on the plot by an increased length of black vertical whiskers. Nevertheless, the program predicted a high likelihood that this strain may be resistant to ETA, KAN and OFL, and may also show an intermediate resistance to CM, EMB, FLQ, INH, RIF and SM. Because the program could not distinguish between CAS and X-type, patterns of resistance were analysed for both these clades and the biggest *R*-values were selected. Resistance to PAS was not expected for either CAS or X-type isolates as there were no such isolates in the training dataset used for this program. This is why the program set this strain sensitive to PAS by default without any estimation. Setting of the drug sensitivity by default is depicted by short grey bar. Only a few diagnostic sites needed for PZA resistance prediction were found in this fragmented genome and they were uninformative. For example, in PZA sensitive CAS isolates, a *Met* residue is expected on the 134th codon of the Rv0040c gene while *Ile* is expected on the same locus for PZA resistant isolates. In the given strain, *Val* was found on this locus, a finding that does not fit with the

expectations. The program marked this antibiotic on the plot with a short red bar indicating an insufficiency in the information to make any decision.

In Figure 2.9C, isolate ERS458164 from Sierra Leone was predicted as *M. bovis* clade, which also includes predominantly drug susceptible Western African and *M. africanum* isolates. The current version of the program was not designed to analyses the drug resistance in this clade due to lack of published data. The program displays by default that the isolate is most likely susceptible to all antibiotics except for the vaccination *M. bovis* strains reported to be PZA resistant [202] that is indicated by highlighting the PZA resistance in the output file.

In Figure 2.9D, an example is given of the analysis of historical DNA, in NGS format, obtained from human remains of an individual who died from tuberculosis in XVIII century in Hungary [204]. The current analysis confirmed the affiliation of the Mtb strain with lineage 4 as reported in the original paper but with a better precision of the identification to the sub-lineage 4.3, which is common in Europe. This strain already possessed many mutations specific to future multidrug resistant Mtb variants of this lineage; however, this strain most likely was still susceptible to all antibiotics (sensitivity coef. 0.55).

Both the online and stand-alone program implementations also return a text output file listing the states of all diagnostic sites in addition to the graphical output in SVG file format. Resistance Sniffer predicts the resistance to antibiotics by analysing patterns of diagnostic polymorphisms in genome sequences. The actual resistance of a bacterium to antibiotics may be affected by some other factors that the program does not consider, particularly epigenetic modifications. For example, a recent study on the application of a new drug, FS-1, which causes the reversion of multidrug resistant bacteria back to the sensitive phenotype showed that the pattern of drug resistant mutations remained unchanged in the strains with reversed susceptibility to antibiotics [205]. Thus, the strains with reversed antibiotic sensitivity due to epigenetic modifications will be predicted as resistant by the pattern of diagnostic polymorphisms. Moreover, the same study showed that Mtb isolates from experimentally infected laboratory guinea pigs showed a range of antibiotic susceptibility values due to natural variations within the population even though all the animals were infected with the same MDR-TB strain. Hence the antibiotic resistance prediction is probabilistic by nature. The Resistance Sniffer program estimates a rough likelihood for an isolate to be resistant to one of thirteen antibiotics or to be sensitive to all of them. These estimations are

recorded in the text output file or may be displayed on the screen when the mouse pointer is placed over the bar on the plot (see Figure 2.9A).

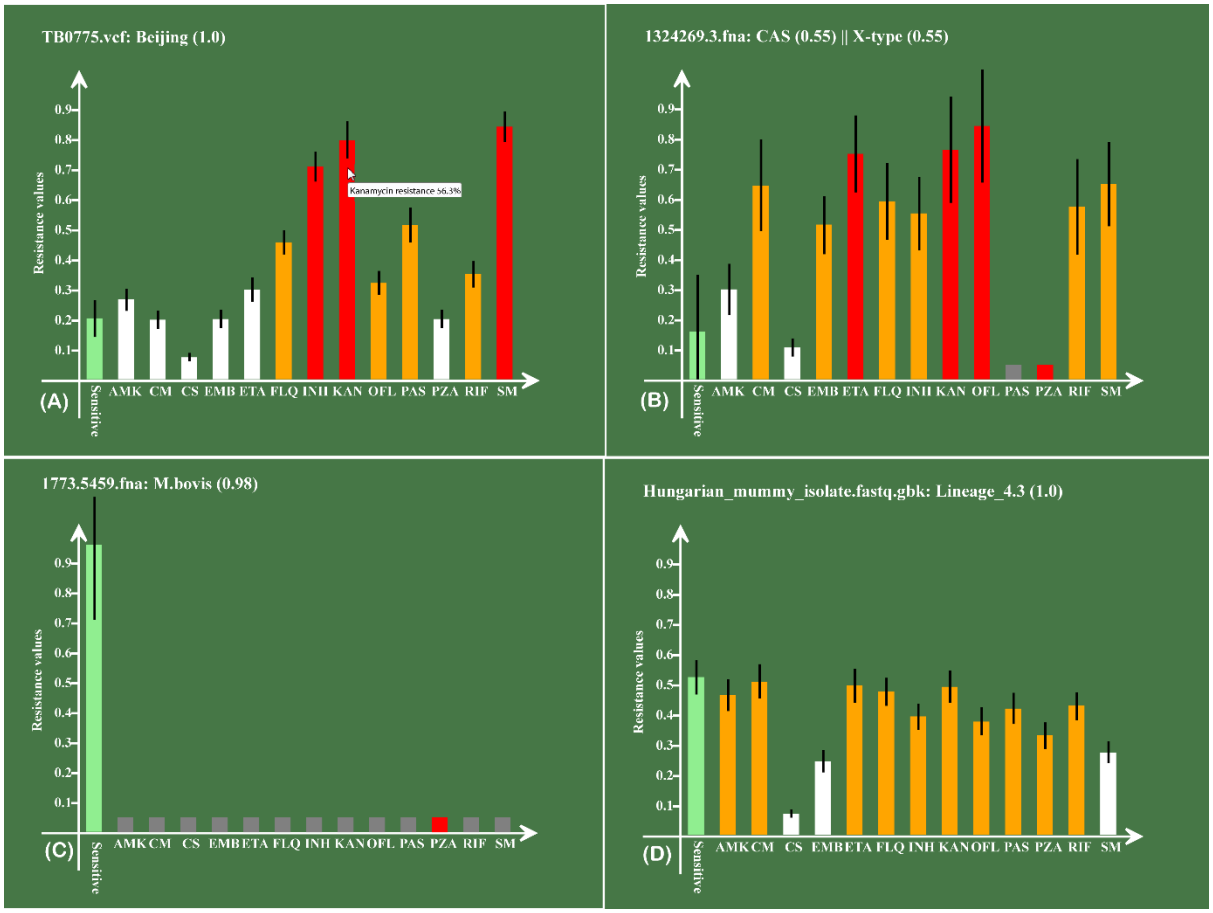

Figure 2. 9 Drug resistance predictions by Resistance Sniffer for the strains (A) TB0775; (B) 1324269.3; (C) 1773.5459 and (D) Hungarian mummy isolate. White columns show sensitivity to antibiotics with the confidence above 55%; red columns predict the resistance with the confidence above 55%; and orange columns show intermediate results. The green column depicts the likelihood for this strain to be sensitive to all 13 antibiotics. Estimated R-values are shown along the vertical axis. Standard errors of calculation are depicted by black vertical whiskers. Antibiotic resistance nodes are: amikacin (AMK), capreomycin (CM), cycloserin (CS), ethambutol (EMB), ethionamide (ETH), isoniazid (INH), fluoroquinolones (FLQ), kanamycin (KAN), ofloxacin (OFL), para-amino salicyclic acid (PAS), pyrazinamide (PZA), rifampicin (RIF) and streptomycin (SM).

2.6 Evaluation and Results

2.6.1 Phylogenetic clade classification

The accuracy of this tool depends on the correct classification of the phylogenetic lineage of the Mtb sequences. The precision of the program Resistance Sniffer for lineage classification was assessed using lineage information from the original GMTV test dataset, as well as lineage information provided for in the Sierra Leone dataset. It is important to note that lineage information

was not available for all the samples from the SA-MRC, 121 samples from the GMTV as well as for 11 of the Sierra Leone samples. The program was able to perform lineage classification for all samples in our test dataset with >95% accuracy. The main reason for discordance in classifying samples from the GMTV database can be attributed to the lack of the database's resolution in classifying sub-lineages of Mtb. Classification errors were also observed for 4 strains from the Sierra Leone dataset (ERS457923, ERS457211, ERS457423, and ERS457331) which were misclassified as either CAS, Beijing or Lineage 1.2. The predicted clades for the GMTV and SA-MRC strains whose lineage information was not available is shown in Figure 2.10.

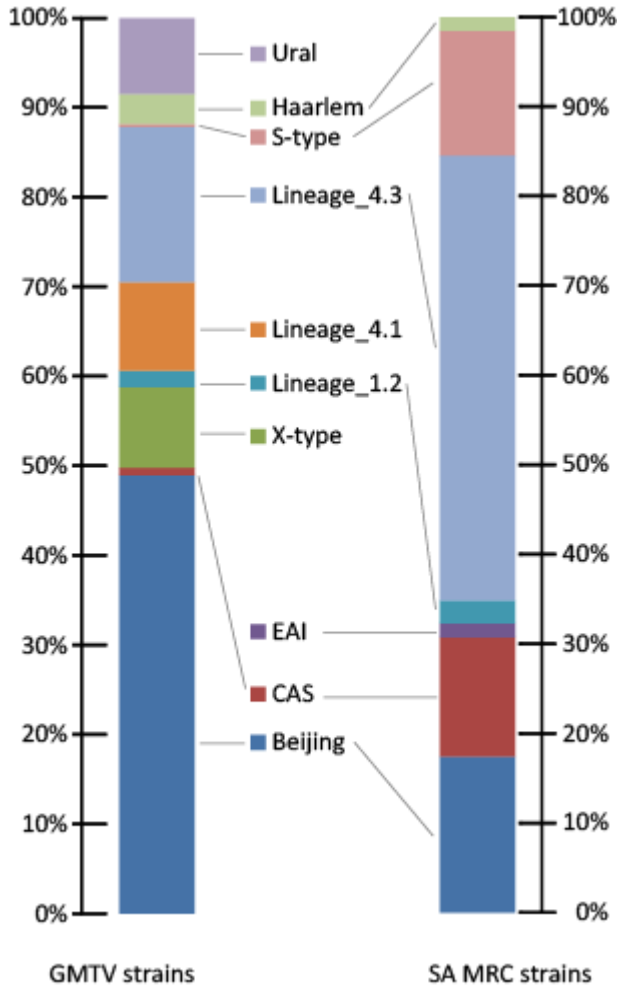

Figure 2. 10 Frequencies of clades assigned to Mtb strains from GMTV and SA MRC.

The GMTV database is predominantly constituted by Mtb strains isolated in Russia, while SA MRC presents clinical isolates from South Africa. The majority of the GMTV strains belong to the highly virulent Beijing clade, European lineage 4.3, Asian lineage 4.1 (type strain lineage from India) and Ural clade-specific for central Russia. European lineage 4.3 is even more prevalent

among South African isolates. It is followed by the S-type variants of this lineage, Beijing and CAS clades. The prevalence of the Euro-American strains among the South African isolates is an interesting finding which requires further research with emphasis on investigating the possible introduction of the disease as a result of colonial migration. EAI clade is present among South African isolates but not frequent, probably because of relatively lower virulence. Asian clades, Ural and lineage 4.1 were not found among the SA MRC isolates.

2.6.2 Power of association

In all the phylogenetic clades, the PE, PPE gene family were highly associated with resistance to antibiotics. This gene family consists of surface-exposed cell wall proteins which are known to affect cell wall structure and permeability and some have been shown to be antigens [121]. The association of these genes with the DR mutations is further complicated by homoplasy as these genes are known to be highly polymorphic, however the association of this family of genes with drug resistance is an important area of research. There was also a high association of mutations that code for genes that are involved in cell wall homeostasis such as *ppsA, murD, pks* and *ponA1*. Rifampicin resistance in Lineage 4.3 was highly associated with several mutations in the *ppsC, mas* and *ppsA* genes, a finding which may suggest that the PDIM biosynthesis system may play a leading role in the evolution of drug resistance in this clade.

The mutations in the *gidB* have been previously associated with the development of low level SM resistance in Mtb [206]. Our study found significant association between SM resistance in the Beijing clade and *gid* mutations at position 16, 48 and 210 of the Mtb genome. Mutations in the *Rv3908 mutT4* were also highly associated with DR in the Beijing, CAS, Lineage 1.2, X-type and lineage 4.3. However our method did not identify any association between this locus and lineage 4.1, S-type and Ural. This finding adds weight to the growing consensus that suggests that the high propensity of strains in some clades such as the Beijing clade to develop into MDR-TB strains can be attributed to elevated mutation rates as a result of missense mutations in the *mut* gene [143]. Contrary to what we observed in the *mut* gene, mutations in the Rv1316 *ogt* gene were highly associated with drug resistance in the Ural, Haarlem, lineage 4.3 and X-type clade but were absent in the Beijing, Lineage 4.1, CAS and S-type clades. The *ogt* is responsible for the removal of methyl groups from 06-methylguanine in the DNA repair process. Another interesting finding in our study was that unlike in the other clades, the X-type clade was not associated with any

mutations in the *inhA*, *eis* and *tlyA* mutations which have been previously associated with low level resistance and resistance to second-line aminoglycosides. As expected, mutations in the *embCAB* operon were highly associated with ethambutol resistance. We further investigated the role of these mutations by incorporating all the identified *embCAB* mutations in the Beijing clade into the Resistance Sniffer program and analysing the text output for the identified mutations in our Beijing samples. The only variation was observed in the *embB* gene at position 406, 354 and 306 which suggests that mutations in the *B* subunit of the operon may be the cause of ethambutol resistance in Mtb. Only two loci were highly associated with kanamycin resistance in both the CAS and Beijing strains. These were *tlyA* Q216R and Rv0007 G291D. There is a need to further investigate the possible role of these two mutations in the development of drug resistance in these two clades.

2.6.3 Accuracy of antibiotic resistance prediction

In total, 1,201 Mtb strains from GMTV and 742 strains from SA MRC were characterized by their sensitivity to one or several antibiotics resulting in 8,559 data entries. This information was used to validate and test the performance of the Resistance Sniffer program. Antibiotic resistance was predicted by *R*-values calculated by Eq. 3. This equation returns values in the range of 0 to 2; however, for the majority of tested strains *R*-values were below 1.0, and those strains showing higher *R*-values were antibiotic resistant. The program was set to reduce *R*-values to 1.0 if they were bigger.

Assignment of strains as sensitive or resistant with respect to a given antibiotic was performed by setting a cut-off *R*-value. If the cut-off value is 0, all the strains will be deemed resistant and fall either into true positive (TP) or false positive (FP) categories. Sensitivity of the program will be maximal (1.0) and specificity will be minimal (0.0). Conversely, all the strains will be deemed sensitive and fall into either the true negative (TN) or false negative (FN) categories with the cut-off value of 1.0. The *R* cut-off value was gradually changed from 0 to 1 with a step of 0.25 and the distribution of FP, FN, TP and TN strains was evaluated by Eqs.5-8. The calculated specificity and sensitivity of the program for different cut-off values are shown in Figure 2.11

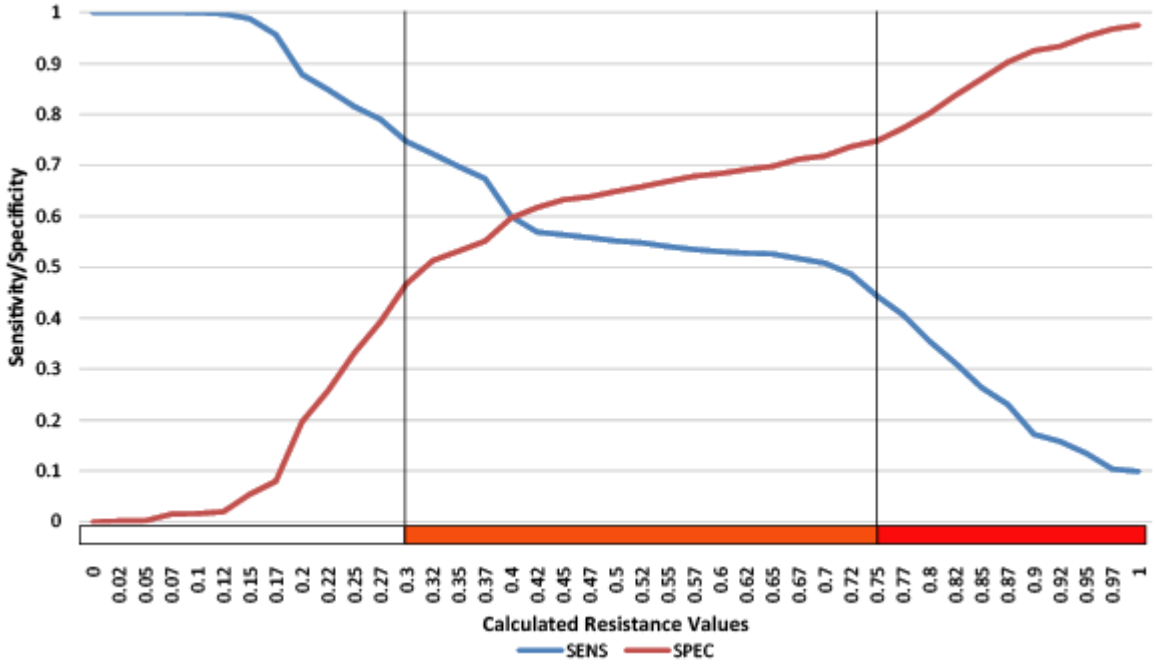

Figure 2. 11 Sensitivity and specificity of antibiotic resistance prediction with different R cutoff values. Vertical lines depict borders set in the program to distinguish between sensitive, potentially resistant and antibiotic resistant Mtb strains.

Resistance *R*-value above of 0.75 predicts the strain to be resistant against the given antibiotic with the likelihood of 55% or above. If the *R*-value is below 0.3, the strain is sensitive to the antibiotic with the likelihood 55% or above. The strains with intermediate R-values are either sensitive or resistant with equal likelihoods. This area of uncertainty has an important biological meaning as it depicts Mtb strains which may currently be sensitive but very close to gaining the resistance in the near future and should be marked as potentially dangerous. Using Resistance Sniffer on our test dataset we obtained higher sensitivity and specificity for both first line drugs INH and RIF (90-96%).

Testing of the program on 77 Mtb isolates from Sierra Leone characterized by sensitivity to EMB, INH, PZA, RIF and SM, or at least to one or several of these antibiotics (in total 285 strains per antibiotic measurements in Supplementary file 3) was performed with different cut-off *R*-values. Average values of sensitivity (0.5) and specificity (0.5) were achieved under an assumption that a strain is resistant to an antibiotic if the estimated *R*-value ≥ 0.25. An increase of the *R*-value cut-off led to a rapid increase in specificity and decrease in sensitivity. To improve the program performance, an additional parameter 'Sensitive' was introduced to reflect in the program output. This parameter was calculated as 1 ─ average of the top 6 *R*-values determined for different antibiotics. The rational of this approach was that a multidrug resistant Mtb strain may very likely

show some level of resistance to other antibiotics even if no specific genetic determinants of the specific antibiotic resistance were found. The optimal program performance was achieved with the *R*-value cut-off 0.3 under an assumption that the strain is resistant to all antibiotics if the Sensitivity coefficient is equal or lower than 0.2. With these settings, the susceptibility to antibiotics was correctly predicted in 184 cases and the resistance was correctly predicted in 19 cases. There were 41 false susceptibility predictions and 41 false resistance predictions. Calculated sensitivity and specificity were 0.32 and 0.82, respectively. As no data on the reliability of applied DST techniques was made available, it is not possible to judge whether the false negative and false positive predictions should be attributed to the experimental procedures of drug susceptibility testing or to the program algorithm. It should be noted that antibiotic resistance profiling of Mtb isolates in bacteriological laboratories is an error-prone procedure showing a relatively weak correlation with the clinical response due to slow growth rate of this bacterium and bad standardization of the procedures [207]. The accuracy is even worse when the data originates from different laboratories. It is expected that the sensitivity and specificity of the program cannot exceed the accuracy of the training dataset but it seems that the antibiotic resistance prediction by Resistance Sniffer does not add significantly to the expected level of errors of laboratory drug resistance trials. The meaning of antibiotic resistance likelihood predicted by Resistance Sniffer will be discussed in more detail below.

2.6.4 Case Studies

In this section we consider two examples of how a tool like Resistance Sniffer can be used in a futuristic practical setting. Figure 2.12 below shows the predicted resistance profile for strain 1397850 which was isolated in South Africa. From the output the program predicts probability of resistance to both INH and RIF for this strain to be slightly below 0.20. This indicates that the particular strain is non-MDR-TB. In an ideal clinical setting, treatment outcomes for this patient can be improved by carefully increasing the dosages of the first-line drugs. Although our tool predicts EMB, ETA, OFL and SM resistance, traditional DST results for this sample would indicate that the strain is susceptible to all drugs and this would be followed by a standardised anti-TB drugs regimen. However this is dangerous because a standardised regimen might result in inadequate dosing which might amplify resistance and risk the transmission of resistant strains at population level. In future we hope that a program like *Resistant Sniffer* will not only enable the clinician to identify isolates that are likely to become resistant but will also integrate other valuable information such as the extent to which mixed infection is affecting treatment outcome. This will

of course require the adoption of a universally accepted NGS-based-diagnostics reporting standard. Until pDST methods improves, prediction of EMB resistance must always be treated with caution due to the challenges associated with reproducing pDST results for the antibiotic. The high resistance value predicted for ETA in this sample can also alert clinicians to further investigate the possibility of the future development of cross-resistance with INH.

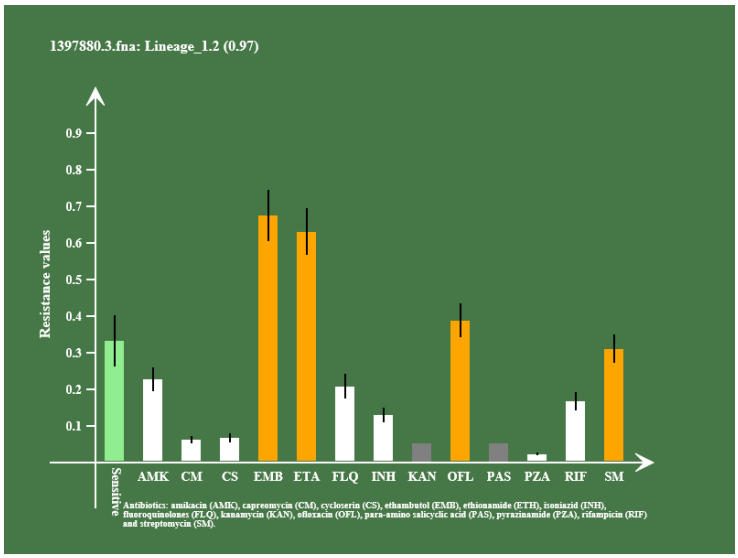

Figure 2. 12 Resistance Sniffer output for an isolate predicted to be a lineage 1.2 strain.

Figure 2.13 below shows a strain that is likely to belong to either Lineage 4.3 or Lineage 4.1. Our program predicts the sample to be MDR-TB which means that a more complicated drug regimen is needed for the patient. The necessary intervention for such a patient would require that the infected patient be quarantined until successful treatment to avoid transmission. The clinicians will also have to be cautious about introducing new or repurposed drugs such as bedaquiline in the treatment regimens as this might result in the development of resistance to these new generation antibiotics. The use of novel resistance reversion chemotherapy such as FS-1 [205] could also be considered. On a public health level, an adequate response will involve screening of individuals who had contact with the TB patient to ensure that the MDR-TB strains is not transmitted further. In the case of an outbreak caused by such a strain, the adequate response would involve prioritizing all resources towards curtailing infection.

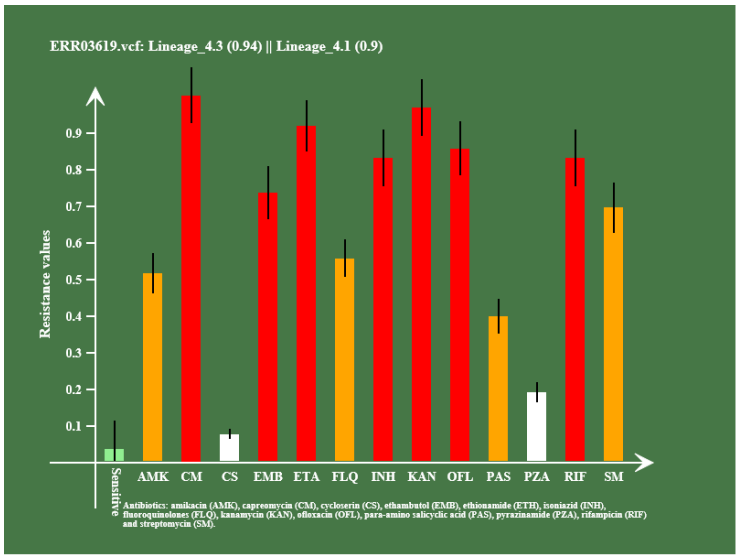

Figure 2. 13 Resistance Sniffer output for an isolate predicted to be a lineage_4.3 or lineage_4.1.

2.7 Discussion

Resistance polymorphisms appear to be clade-specific, which means that some mutations are more likely to be present in specific Mtb lineages [130, 201, 208]. There is also evidence suggesting that strains from certain lineages are predisposed to develop into MDR-TB strains [110, 157, 158]. In the present study we sought to identify clade-specific patterns of polymorphisms that may be associated with the drug resistant phenotype in Mtb. Multiple mutations associated with antibiotic resistance in Mtb are known from literature and are available from public databases (e.g. https://tbdreamdb.ki.se/Info/Default.aspx) [201]. However, for this study a decision was made to perform a *de novo* search for attributable mutations by calculating power coefficients (Eq. 1) of association between polymorphisms and the drug resistant phenotype for each antibiotic per clade. In many cases, but not always, highly scored polymorphisms were consistent with known drug resistance mutations in literature. Some of these highly scored polymorphisms were also located in genes associated with the antibiotic resistance and there is a need for further interrogation of these findings. Several examples are discussed below.

Mutations in the *emb*CAB operon have been known to be associated with resistance to EMB, particularly the substitutions in codons 306, 406 and 497 of *emb*B [38]. The current study confirmed that in the Beijing clade several polymorphisms were highly scored on these loci, particularly in codons 306, 354 and 406 of the *embB* gene. Strain TB0012 has mutations in codons

306 and 354 but shows a susceptible phenotype according to the GMTV database records. Strain TB0011 shows a resistant phenotype and possesses mutations in codons 354 and 406. Two other strains, TB0004 and TB0005, have mutations in all three codons although they both show a susceptible phenotype. These findings highlight the complexity of the development of antibiotic resistance in Mtb which requires accumulation of many other subordinate mutations acquired in a stepwise manner to develop sustainable drug resistance. This assumption was used as the basis of the Resistance Sniffer algorithm of estimation of the drug resistance likelihood by assessing the whole pool of genetic determinants biasedly distributed between antibiotic sensitive and resistant Mtb variants. This hypothesis was confirmed in numerous publications [100, 130, 209, 210]. However, it is also important to acknowledge that the discordance might also be a result of the challenges associated with the pDST for EMB. Mutations in *embB* only have been shown to result in slight increases in MIC values to EMB. This results in an overlap between the MIC distributions of wild-type strains and mutated strains [188]. However, secondary mutations are believed to increase the MIC further [210-212] hence identifying these secondary mutations is important in the diagnosis of EMB resistance. It was hypothesized that the *mut* genes may play a significant role in the acquisition of drug resistance in Mtb because missense mutations in these genes lead to higher mutations rates [130, 143]. Indeed, mutations in the *mutT4* gene *(Rv3908*) were associated with the drug resistance phenotype of Mtb strains of the Beijing clade, but no specific mutations were found in the strains of lineage 4.1 and S-type lineage, which, in contrast to Beijing strains, usually do not develop a wide spectrum of drug resistance (see Figure 1). The Beijing clade has been associated with high levels of drug resistance and a higher propensity to develop into MDR-TB and XDR-TB. The reason why Beijing isolates are often associated with MDR-TB remains elusive. Researchers have suggested that the strain background could be more efficient in mitigating the effects of fitness cost imposed by drug resistance [110, 159].

Another gene of interest in this study was *ogt (Rv1316),* which is known to remove methyl groups from O-6-methylguanine in DNA. Mutations in this gene were associated with SM resistance in Ural, Haarlem, lineage 4.3 and X-type clades. However, no significant associations with mutations in this gene and SM resistance were revealed in Beijing, CAS, S-type and lineage 4.1, which also have a high propensity to develop SM resistance but in a different way. Strangely enough, this study did not discover any significant correlation between mutations in genes *inhA, eis* and *tylA* and the drug resistance phenotype despite many publications linking these genes with drug resistance. Mutations in the *inhA* regulatory regions are known to confer low level INH and ETH resistance. The *tylA* and *eis* genes are both drug targets for the second-line injectable antibiotics.

This work discovered several mutations in the phenolpthiocerol synthesis polyketide gene (*ppsC)* and multifunctional mycoserosic acid synthase gene (*masA*) to be associated with rifampicin resistance in the lineage 4.3. A study by Bisson et al. [213] showed that the expression level of pps can be up to 10 fold higher in *rpoB* mutant strains relative to the RIF susceptible parent strain

Our study suggests that the progression to the drug resistant phenotype is a stepwise process involving the accumulation of multiple mutations contributing to the antibiotic resistant phenotype. Alternative hypotheses involving rare drug resistance mutations in Mtb populations were proposed by other authors. For example, in the publication by Carvalho *et al.* [214] it was suggested that rare cases of resistance to D-cycloserine are caused by low frequency mutations in target genes, *cycA*, *alr* and *ddlA*, rather than fitness cost reduction mediated by other compensatory mutations. Our study confirmed the importance of *cycA* V67C, L322R and M343T *alr*, and T365A *ddlA* substitutions in the development of CS resistance; however, significant associations with multiple compensatory mutations linked to other antibiotic resistance were also observed. It may explain the insignificant fitness cost of the CS resistance mutations as they occur only in organisms already possessing the compensatory mutations. The Resistance Sniffer program may identify Mtb strains on a trajectory to developing drug resistance by accumulation of pre-requisite mutations, even if phenotypic DST results for these strains do not show any evidence of antibiotic resistance yet.

The key to the total eradication of TB globally lies in early diagnosis and correct treatment. This has been hampered by the limitations in current laboratory methodologies in performing drug susceptibility testing. Current DST procedures for Mtb are time consuming, expensive and inaccurate, especially for the second line antibiotics. Horizontal gene transfer plays no role in the development of antibiotic resistance in Mtb. This makes WGS an attractive option in the diagnosis of TB as it has the potential to determine the full antibiogram provided we have detailed knowledge of all the genetic determinants of drug resistance [9]. In this study, we used a derivative of GWAS to identify clade-specific patterns of polymorphisms which showed a biased distribution regarding the drug resistant phenotype. The study was designed first of all as a proof of concept of the ability to predict drug resistance or the predisposition for drug resistance acquisition by Mtb isolates. However, the designed software tool, Resistance Sniffer, showed the sensitivity and specificity of the clade and the drug resistance identification similar to that of other available tools such as TBProfiler, MyKrobe, KvarQ and PhyResSE. A recent large-scale benchmarking comparison of the available tools on 6,746 Mtb isolated characterized by drug susceptibility patterns showed

applicability but also some limitations of the available tools (Ngo and Teo, 2019). According to this report, the specificity and sensitivity of the programs varied from 0.6 to 0.9 depending on which antibiotic was tested with the best results achieved when they are confirmed by more than one program. All the programs showed a much better ability to predict the absence of drug resistance rather than the specific drug resistance pattern. This is also true for the Resistance Sniffer program. It was not discussed in this review to which extent the performance of the program was affected by the level of fragmentation of the genome of interest as only whole genome sequences were used in the reported study. While the estimated sensitivity and specificity of Resistance Sniffer were lower than those of the above-mentioned programs, it should be noted that the current program was developed to analyse fragmented partial genome sequences including historical sequences (see Figure 2.9D) represented by different file formats. Particularly, unordered contigs of Mtb isolates from Sierra Leone in plain fasta format were used for the program evaluation The aim of the study was to estimate the propensity of a Mtb isolate to gain the antibiotic resistance rather than to delineate antibiotic resistant from antibiotic sensitive strains. The performance of the program may be improved in future studies by editing the diagnostic key table without the need to modify the program itself. A limiting factor was the size of the training dataset of Mtb strains with known drug susceptibility profiles. For certain clades the number of available records was not sufficient to boost the association power. Although there are currently more than 20 drugs that are used in the treatment of TB, this study was limited to only thirteen antibiotics due to the unavailability of phenotypic DST data for the omitted drugs. As more data becomes available, the diagnostic key table of the program will be updated.

It is expected that in the future NGS-based assays will replace phenotypic DST methods [215]. This study has demonstrated how whole or partial genome sequence data can be used to rapidly predict drug resistance in *M. tuberculosis*. However, it should be emphasized that the current version of the program was not designed for application in clinics or for assessing antibiotic treatment regiments. The major objective of the program was to provide scientists working in public health control institutes with reliable software to estimate the distribution of drug resistant infections by using NGS datasets in different stages of genome assembly including raw *fastq* files generated by sequencers. The study also acknowledges the recent efforts that has been made by several groups in harnessing the power of WGS in TB diagnosis as described in Chapter 1.4. However, a major concern that is often associated with WGS-based tools that are developed in academic settings is the lack of dedicated version control. Future work will involve ensuring the longevity of the analysis program constantly updating functionality according to the improved

understanding of the genetic basis of drug resistance in Mtb as well as improvements in sequencing technologies.

This work also added to the current body of knowledge a valid suggestion that drug resistant phenotype is associated not with individual mutations but with clade-specific patterns of polymorphisms. Effective prediction of drug resistance should start from a proper identification of clade affiliation of Mtb isolates.

Chapter 3-The evolution of drug resistance in *Mycobacterium tuberculosis*

3.1 Introduction

In the previous chapter we used a clade based approach to determine the strength of association between each mutation and phenotypic resistance to 13 anti-TB drugs. We then created a catalogue of genetic determinants of resistance by manually curating polymorphisms that were highly associated with drug resistance in TB. The catalogue was then used in the creation of an online tool that predicts antibiotic resistance in Mtb using NGS data. However, our diagnostic approach is limited to our current understanding of the genetic basis of resistance. As our understanding of the mechanisms of resistance increasingly improves, it is becoming more clear that the mechanisms that lead to the emergence and success of resistant Mtb strains is much more complex than we previously anticipated. This has led to an increasing interest in Mtb resistance research thus helping to fill the gaps, inspiring new research and innovation in public health care. With the data we had available and previous work done in our group by Van Nierkerk *et al.* [30] we saw the potential to further broaden our research to determine non-random associations between polymorphic sites in genomes of Mtb. This will not only allow us to identify harbinger mutations that can allows us to predict resistance as early as possible [23], it will also allow us to minimise the discordance between genotype and phenotype in Mtb. In this chapter, several statistics that were introduced in the review by Van Nierkerk *et al.* [30] will be used, the most important being the Levin's attributable risk statistic [216] (R_a), which will be used to identify functional and/or genetic drift associations between mutations in the Mtb genome by using equation (1) and equation (2) for a statistical validation of attributable risk values. Allele combination frequencies are denoted as P_{AB}, P_{Ab}, P_{aB} and P_{ab} , while N represents the total number of the analysed Mtb strains.

$$R_{A\to a|b} = \frac{P_{AB}P_{ab} - P_{aB}P_{Ab}}{(P_{AB}+P_{aB})(P_{aB}+P_{ab})} \qquad (1)$$

$$StdErr = \sqrt{\frac{P_{Ab}+R_{aA\to a|b}(P_{AB}+P_{ab})}{N\times P_{aB}}} \qquad (2)$$

In the case of co-dependence between a primary mutation and another secondary mutation, DR mutations will be denoted as mutations from initial allele *A* to allele *a*. The secondary site mutations will have the most frequent allele represented by *B*, while *b* represents all the other

alternative variants at that particular locus in the Mtb population. In cases where we have to estimate the risk of a drug resistance mutation from *A* to *a* in a subpopulation of organisms possessing the *b* allele at the secondary polymorphic site, we calculate the $R_{a\ parameter}$ using the Eq. (1). The Fleiss' standard error parameter *StdErr* is calculated by Eq. (2) [30]. Alternatively, in cases where the genotype is *a* meaning the drug resistance mutation is already present, we may estimate the risk of a subordinate compensatory mutation from *B* to *b* by Eq. (3) and Eq. (4) will be used to calculate the Fleiss' standard error.

$$R_{B\to b|a} = \frac{P_{AB}P_{ab} - P_{aB}P_{Ab}}{(P_{AB}+P_{Ab})(P_{Ab}+P_{ab})} \qquad (3)$$

$$StdErr = \sqrt{\frac{P_{aB}+R_{B\to b|a}(P_{AB}+P_{ab})}{N\times P_{Ab}}} \qquad (4)$$

Finally, Eqn. (5) will be used to calculate the range of confidence values for the calculated attributable risks.

$$[1 - EXP(ln(1 - R_a) - 1.96 \times StdErr)]\ to\ [1 - EXP(ln(1 - R_a) + 1.96 \times StdErr)] \qquad (5)$$

The rationale of this approach was first introduced by Nierkerk *et al.* [30], who compared the functional associations between mutations in the Mtb genome. Considering the co-distribution between the *katG* S315T and *stp* D69Y mutations first, both mutation pairs are associated with high linkage disequilibrium values. Using the equations described above, we can see that the emergence of the *katG*S315T is significantly dependent on the presence of the *stp* D69Y ($A\rightarrow a_{|b}$ = 91-99% CI). When we use the same approach to calculate the co-distribution of the same mutation in the *katG* gene and another concomitant mutation ppe35 L896S, this pair also shows a strong association >90%. At first glance one would be tempted to assume that both of the secondary mutations are essential for the emergence of the *katG* S315T mutations. However, stark differences are observed when we calculate $B\rightarrow b_{|a}$, which represents the dependence in the opposite direction. Here we can see that the distribution of the mutation *stp* D69Y is independent of the state of the *katG* locus with the attributable risk in the range 21% - 27%. In contrast, the distribution of the PPE35 gene polymorphism depends as much on the state of the *katG* locus as the *katG* polymorphism depends on the state of PPE35 that is over 90%. The high level of bi-directional symmetry exhibited by the pair of *katG* and *ppe35* polymorphisms suggests a strong

association between these mutations due to genetic drift. In other words, the specific mutation in *ppe35* may be a genetic marker of the antibiotic resistant sub-population of Mtb but hardly any functional associations between these genes. By contrast, an apparent one-directional dependence of the *katG* mutation associated with drug resistance on the polymorphic state of *stp* mutation suggests that the *stp* mutation is a prerequisite mutation preceding the isoniazid resistance acquisition by Mtb conferred by the *katG* S315T mutation. Prerequisite mutations make it possible for bacteria to acquire the mutations necessary for the antibiotic mutations. Another important scenario to imagine is when a concomitant mutation in the Mtb population shows a one-directional dependence to a drug resistance mutation. This implies that the secondary mutation is acquired by the population as a compensatory mechanism to mitigate the fitness cost imposed by the drug resistance mutation.

3.2 Materials and methods

3.2.1 Data sourcing

Data for this project was obtained from the GMTV database as described in the previous chapter. The collated data of all 2501 Mtb strains were then used to calculate the mutation frequencies for all the clades in our analysis. Our initial analysis on the strength of association between polymorphisms in the Mtb genome and phylogenetic clade has shown strong associations between some well-known drug resistance mutations with specific Mtb clades. The analysis also identified some novel possible determinants of resistance. This analysis was further confirmed by the validation dataset which we used on our online prediction program. Strains from the Beijing, Haarlem and Lineage 4.3 lineages harboured polymorphisms that were highly associated with drug resistance while Lineage 1.2, Ural and X-type strains harboured a few of these mutations. These findings were consistent with the phenotypic data which showed that strains from the former clades were highly associated with MDR/XDR-TB while the latter clades appeared to harbour more susceptible strains. The Fisher's exact test with Bonferroni adjustment was used to confirm these findings.

3.2.2 Functional associations between Mtb mutations

As in the previous chapter, we took a lineage-based approach to classify mutations which have been associated with DR according to clade affiliation. We then applied the R_a methods described

above to generate co-dependence data between drug resistant mutation pairs for each clade. We further selected only those associations where there is no overlap between $A \rightarrow a_{|b}$ and $B \rightarrow b_{|a}$ indicating one-directional dependence between polymorphisms. The co-dependence information was integrated with the pDST metadata from GMTV and annotation information from the Tuberculosis Drug Resistance Mutation Database (TBDream) database as input for the creation of directed DR functional association networks in Cytoscape 3.7. Using Cytoscape 3.7, it was possible to create a mapping from $A \rightarrow a_{|b}$ to $B \rightarrow b_{|a,}$ where the emergence of the primary mutation depended on the presence of a mutation on the secondary locus. The reverse mapping $B \rightarrow b_{|a}$ to $A \rightarrow a_{|b}$ represents the dependence in the opposite direction. In our approach, we define the mutations as nodes which are linked by edges where a dependence between two mutations exists. To infer evolutionary trajectories, we used three different node colours, green to indicate a mutation where the evolutionary path started, blue for nodes where evolution is still in progress and orange to represent mutations where evolution has ended and that mutation is fixed.

3.4 Results

3.4.1 Lineage 1.2

For Lineage 1.2, the network is small with only a few paths connecting the mutations. Strains in Lineage 1.2 are generally known to be drug susceptible [217]. The only mutations that were found in known drug targets were in the *gyrA* gene, which is known to be associated with resistance to fluoroquinolones. Our study suggests that mutations in the *frdC* (*Rv1554)* gene may be prerequisites to the acquisition of mutations in *gyrA.* The gene *frdC* is known to affect quinol binding [218] although there is a need to investigate its possible role in Mtb control. Our findings also suggest that mutations in *pepD* also precede the acquisition of *gyrA* mutations in this Mtb clade. The gene *pepD* is a stress response protein whose loss of function subsequently affects the expression of other stress response determinants in Mtb [219]. Exposure to antibiotics is known to induce a complex stress response in Mtb, which results in changes in metabolic activity leading to the development of drug resistance. Our study also suggests that the well documented *stp* (*Rv2333*) [220] efflux pump may compensate for the loss of fitness that might result from the mutations in *gyrA.* The role of the hypothetical conserved protein Rv1378 in compensating for FLQ resistance also needs to be investigated.

Mtb is enriched with lipoproteins that have been linked to its virulence as well as drug resistance [221]. In this study we found a link to the possible role of mutations in the *lppB* genes which are likely compensated by several mutations in the *lppA* gene, developing multidrug resistance along the way as evidenced by the increase in the number of resistant drugs as the evolution progresses and terminates with *lprf* gene mutations. The *recX* gene is a negative regulator of *recA* which is a central protein in the bacterial response to DNA damage [222]. The acquisition of the *recX* mutation is likely as a response to DNA damage imposed by the action of fluoroquinolones. A study by Gillespie *et al.* [148] reported that the exposure of Mtb isolates to sub-inhibitory concentrations of fluoroquinolones leads to an elevation in mutation rates.

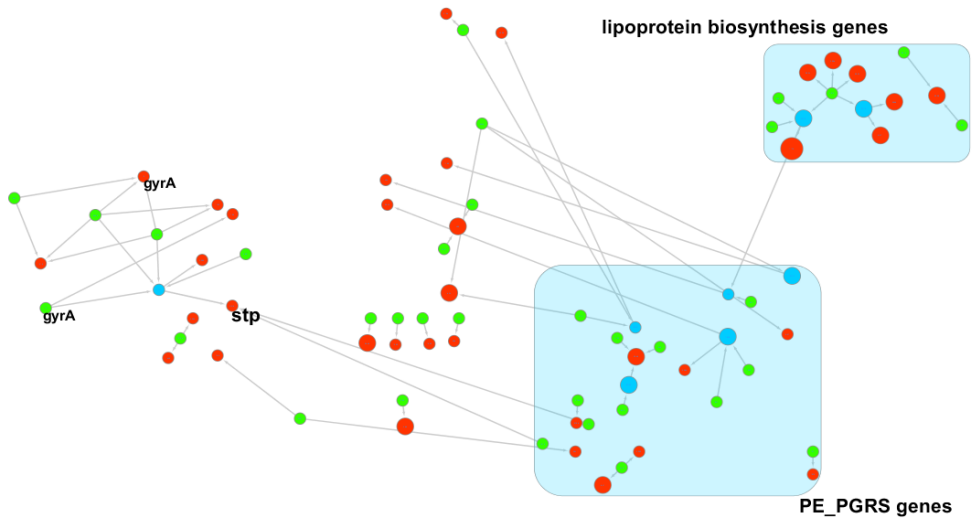

Figure 3. 1 Evolutionary network of the DR mutations for the Lineage 1.2 clade. Strains from this clade have been associated with better treatment outcomes and high drug susceptibility. The green-coloured nodes represent mutations where the evolutionary path has emerged, the blue nodes represent mutations where evolution is still in progress and orange nodes represent mutations where evolution has ended. The size of the nodes are proportional to the number of drugs associated with the mutation.

3.4.2 Haarlem lineage

Our findings in this clade suggest a complex form of interaction which involves some of the main drug targets of Mtb leading to the development of MDR-TB. Our analysis suggests a possible central role of mutations in a putative dehydrogenase /reductase gene (*Rv1928*) as a starting point of the path to resistance to a number of important TB drugs. According to our analysis, polymorphisms in this locus facilitate the development of further mutations in *embB*, which are believed to be associated with the resistance to ethambutol [223]. Interestingly they in turn provoke accumulation of mutations in *rpoB* followed by mutations in *rpsL* with this evolutionary path ending with mutations in *katG*, all of them rendering the resistance to rifampicin [224],

streptomycin and isoniazid [103]. These findings suggests that the development of MDR-TB in this clade starts with the development of low level resistance to ethambutol, which facilitates further resistance to rifampicin followed by resistance to streptomycin and ending with the fixation of high level resistance to isoniazid. Our data also suggested an alternative path, which also bypasses the development of mutations in *embB* and rather follows a trajectory that involves mutations in *ethA* facilitating the acquisition of *rpoB*, which then follows the same path that involves the acquisition of mutations in *embB* and finally in *katG*. It is also important to note that the *ethA* gene encodes a structural analogue of isoniazid, which is activated by the catalase/peroxidase enzyme encoded by the *katG* gene. Moreover, mutations in *ethA* and *katG* have been associated with ethionamide and isoniazid cross-resistance [225]. The involvement of both *ethA* and *embB* in the initial stages of the evolutionary path of DR-TB in Haarlem also adds weight to our hypothesis which suggests that the development of MDR-TB is a stepwise process whereby the initial acquisition of a single mutation facilitates the further acquisition of additional mutations resulting in a path that starts with an intermediate level of resistance and ends with the MDR-TB phenotype. Both *ethA and embB* have been linked with low-level ethionamide/isoniazid and ethambutol resistance respectively while *katG* mutations are known to confer high-level resistance to isoniazid [226]. Another possible evolutionary path suggested by our analysis involves Haarlem strains acquiring mutations in the *pncA* gene which encodes the drug target for pyrazinamide [227] which in turn facilitates the accumulation of *rpoB* mutations before ending the evolution with the fixation of katG mutations. This evolutionary pathway suggests that the order of resistance in some Haarlem strains is as follows PZA, RIF, SM, and INH. Another locus of interest in this evolutionary trajectory is the gene that encodes lactate dehydrogenase *lldD*2 (*Rv1872c*), which seems to play a role as an alternative termination point of the drug resistance evolution. It must be noted that pathways which involve the encoded enzyme may be responsible for the development of a mono-drug-resistance in Haarlem strains [228].

Several efflux pumps mutations such as those in *mmpl* and *fadD* may also play compensatory roles in the development of MDR-TB in Mtb strains of the Haarlem lineage as evidenced by their intermediary role in this complex system. Another finding of our analysis that might need further investigation is the link between mutations in the *Rv2319c* stress protector, which precede the acquisition of mutations in the error prone *Rv2839* initiation factor (*infB*), which has been shown to reduce fitness in Salmonella [229]. Mutations in the *infB* have also been associated with the development of clarithromycin resistance in *H. pylori* [229]. It is also important to note that the

same stress protector (*Rv2319*) is also linked to the development of mutations in the *Rv1047* gene, which has been associated with drug tolerance and the development of drug resistance [230]. The loss of functionality of the DNA repair machinery may also play a role in the evolution of DR-TB in the Haarlem clade.

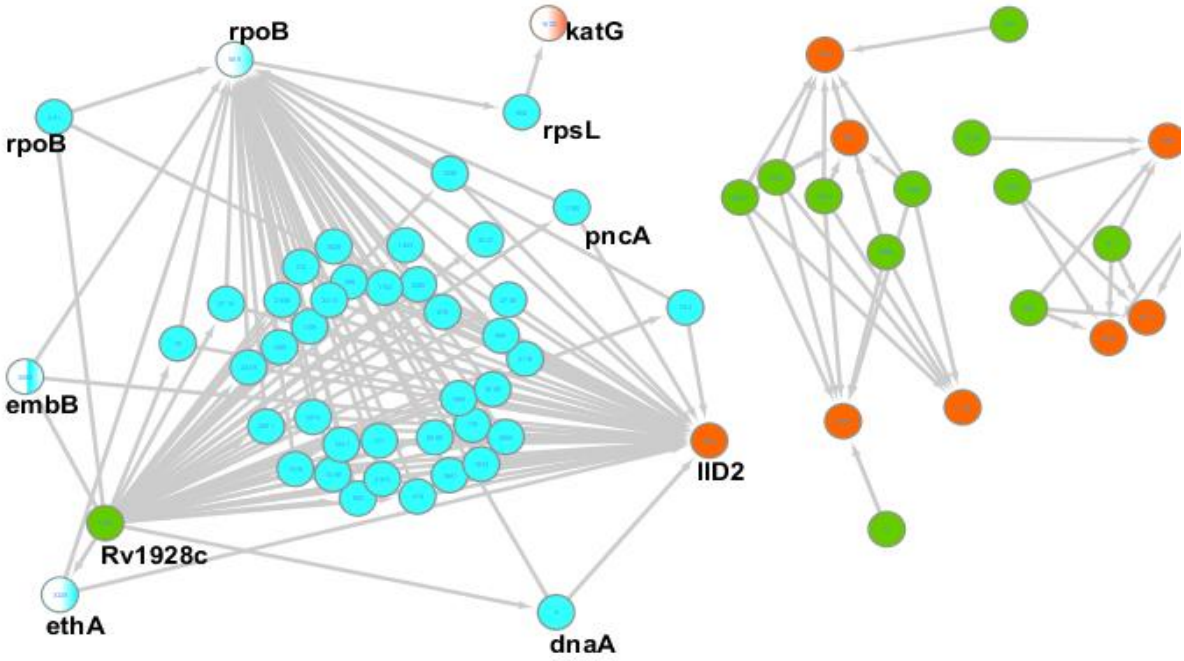

Figure 3. 2 Evolutionary network of DR for the Haarlem clade. The green-coloured nodes represent mutations where the evolutionary path has emerged, the blue nodes represent mutations where evolution is still in progress and orange nodes represent mutations where evolution has ended. The size of the nodes are proportional to the number of drugs associated with the mutation.

3.4.3 Ural lineage

The Central Asian clade Ural, also known as the lineage 4.2, is a sister clade of the North American Haarlem clade also cited as a sub-lineage of the lineage 4.3 [208]. Compared to the Haarlem clade, the network generated from the Ural strains was less complex. An interesting hub in this network involves the *rpsL* gene as an intermediate for various pathways. One such pathway terminates at the *groEL* gene (*Rv0440*), which encodes a chaperonin that has been implicated in the development of resistance to aminoglycoside [231, 232]. Aminoglycosides, such as streptomycin, are known to exert their effect by causing a translational misreading [233]. From the created network of attributable risks (Figure 3.3) we can infer that the development of resistance to aminoglycosides may require some compensatory mutations in this chaperonin. In this pathway, the drug resistance

evolution starts with mutations in *Rv0516*, which is associated with the responses to osmotic stress [234]. Disruption of this gene function has been shown to modify peptidoglycan thickness as well as enhancing drug resistance [234]. This is followed by acquisition of mutations in *rpsL*, which is involved in streptomycin resistance as described in Chapter 1. This leads to acquisition of mutations in *murA*, which has been attributed with the development of resistance to fosfomycin [235], an antibiotic which is no longer in use. The evolutionary path then terminates with mutations in the *groEL* gene. This pathway suggest that the resistance to aminoglycosides is facilitated by the loss of cell wall remodelling function in Mtb and is further compensated for by the over-expression of chaperonin genes [231, 232]. It is important to note that this evolutionary path also contains alternative starting loci such as the *fadD34* gene, which encodes an efflux pump. An interplay of defective stress response systems coupled with an over expression of efflux pumps may be the gateway to aminoglycoside resistance.

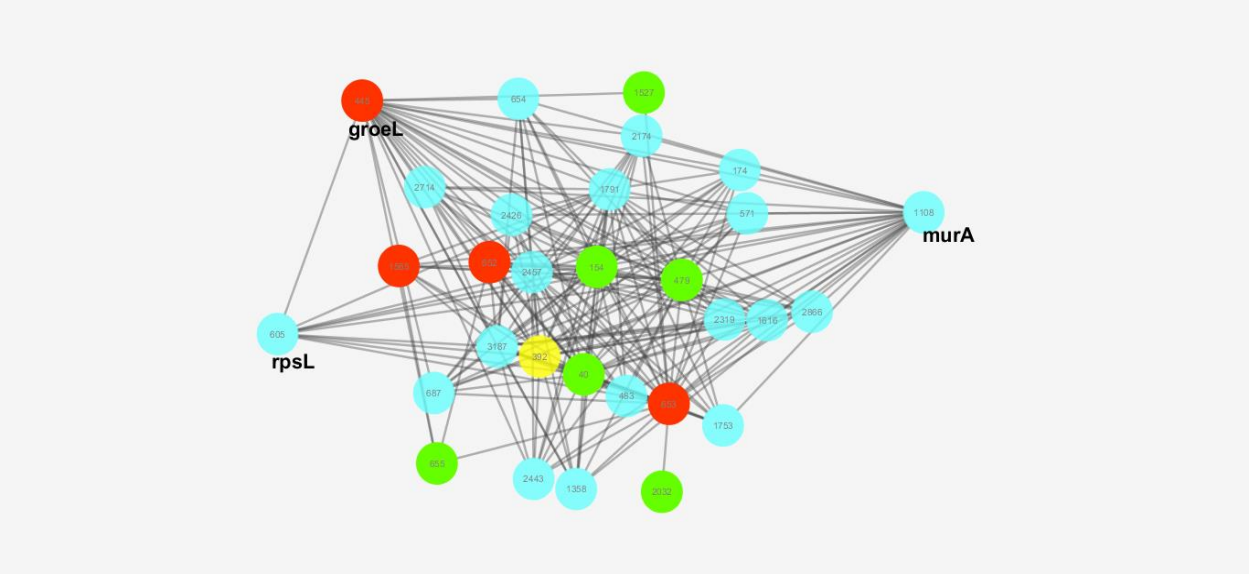

Figure 3. 3 Evolutionary network of DR mutations for the Ural clade. The green-coloured nodes represent mutations where the evolutionary path has emerged, the blue nodes represent mutations where evolution is still in progress and orange nodes represent mutations where evolution has ended. The size of the nodes are proportional to the number of drugs associated with the mutation.

3.4.4 Asian (Lineage 2) clades

The Beijing clade has been associated with hyper-virulence and propensity to develop into MDR-TB strains. In our analysis, we further grouped the Beijing, CAS, and X-type clades under one group. Figure 3.4 shows a complex evolutionary network that involves mutations in cell wall

permeability and drug efflux pump systems. These mutations may speed up the evolutionary clock and facilitate the emergence of other secondary mutations.

In this network, a mutation involving the *katG* gene was identified at the 315 codon location. This mutation is an intermediary in several evolution pathways of drug resistance development which includes other genes involved in DNA repair (*dnaQ*), transport and efflux pump systems (*tatD*), and cell wall biosynthesis (*ppsA*) as the final points of these evolutionary pathways. One pathway of this network involves mutations in the efflux pump proteins *esx* and whiB6 preceding the acquisition of drug resistance mutations in *katG*. This path ends with acquisition of mutations in the DNA repair gene *dnaQ*. This may suggest that changes in the efflux pump systems facilitate the initial development of the resistance to isoniazid, which in turn is compensated by several other mutations which arise as a result of elevated mutation rates. Several studies have reported on the elevated mutation rate in Mtb of the Beijing clade [236]. The increase in the mutation rate can be attributed to the mutation in the DNA proof-reading protein [236, 237]. This inference can be further supported by established links between *embB* mutations and mutations in the efflux pump genes. Ethambutol is included in TB first line regimens because its mechanism of action disrupts mycobacterial cell permeability and thus allows for the improved uptake of other TB drugs. From the network of attributable risks, we can further hypothesize that the low level resistance to ethambutol, coupled with the increased activity of efflux pump proteins results in an environment associated with sub-inhibitory concentration of antibiotics, which is conducive for the selection of DR strains. Our data supports the findings of several studies that have shown that rifampicin resistance emerges after the strain has already developed isoniazid resistance [23, 35, 238]. We can further hypothesize the hyper-mutator phenotype due to disrupted DNA repair machinery further facilitates the rapid acquisition of multiple compensatory mutations which restores the fitness of the now MDR-TB strain.

In our analysis, we identified a single *rpoB* mutation that is followed by mutations in several genes responsible for transmembrane transport as well as cell wall maintenance systems. It is also important to note that mutations in *rpoB* also preceded mutations in the dnaQ gene, although mutations in this gene have been linked with elevated mutation rates in *E. coli* [239], recent studies have shown that Mtb does uses a different DNA proof-reading machinery which involves the DNA polymerase DNAE1 exonucleases instead of the canonical dnaQ exonuclease [240, 241].

Our study identified an interesting short evolutionary path starting with an acquisition of *rpoB* mutations followed by mutations in *Rv3711c*, which encodes a possible DNA polymerase III. There is need to investigate on how the loss of *rpoB* function may be compensated by mutations in this gene. DnaQ is a homolog of gram-negative 3'-5' exonuclease subunit, which is responsible for DNA proof-reading. This compromised fidelity system may be a key to the acceleration of mutation rate, which further leads to the development of resistance to other drugs. For example, mutations or deletions in the *dnaQ* gene in *E. coli* have been attributed to raising the mutation rate 100- 1000 fold [242, 243].

One mutation in the *pncA* gene is followed by three terminal mutations, 2 of which are located in the *ppsA* gene as shown in the network. A possible role of this gene in Mtb drug resistance was explained earlier on. Another locus of interest is *tatD* gene (*Rv1008*), which encodes a probable deoxy-ribonuclease. This mutation was also a terminal point in several evolutionary pathways which started with the acquisition of *rpoB* mutations.

Our attributable risk networking data also revealed a direct link between a mutation in the *cycA* gene and the *rpsL* 43 mutation. This may suggest a possible role of cycloserine resistance conferring mutations in alleviating the fitness cost associated with the aminoglycoside resistance mutations.

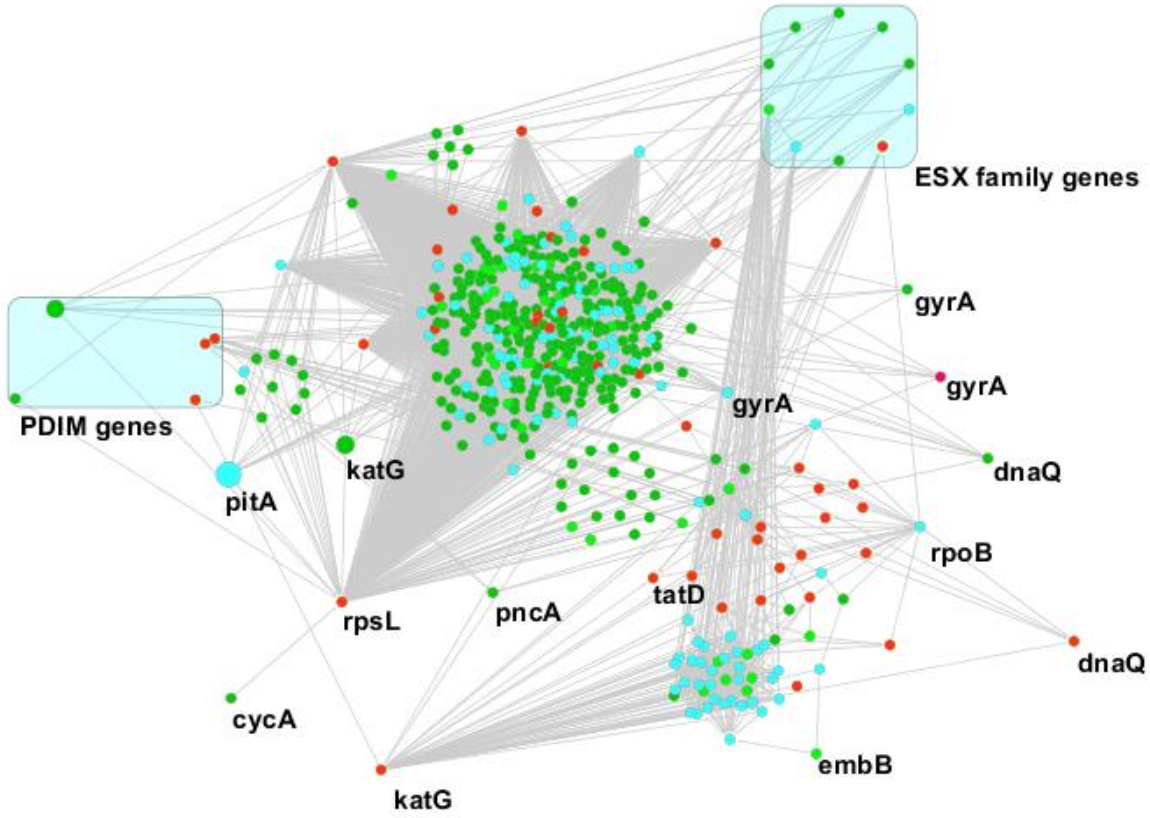

Figure 3. 4 Evolutionary network of DR in the combined Asian clade. The node sizes are proportional to the number of drugs that specific mutations was linked to. The green-coloured nodes represents mutations where the evolutionary path has emerged, the blue nodes represent mutations where evolution is still in progress and orange nodes represent mutations where evolution has ended.

Figure 3.5 shows the previous network on Figure 3.4 with the node sizes representing node degrees, i.e. the numbers of linked neighbour nodes. This plot can give us an indication of which mutations play a central role in the evolution of drug resistance in Asian Mtb strains. The *rpsL* gene seems to play a central role as an evolution hub of the Asian strain. Interestingly, this mutation seems to directly interact with all of mutations in the drug target genes except for *rpoB, pncA* and *embB.* This finding may suggest that the introduction of aminoglycosides in the treatment schemes triggered the MDR-TB evolution in this clade. Figure 3.5 also highlights several main loci playing the central role as compensatory mutations. These include two genes involved in membrane transport (*Rv1258c* and *Rv2434*), a transcriptional regulator (*Rv0823c*) and the well-known gene *rpoC*, which has been previously associated with the restored fitness and improved transmission in MDR strains [125]. Although not shown in the network, several PPE and PE-PGRS family gene mutations also seem to play an important role in fitness restoration, the notable being the ones occurring in the PPE13 protein.

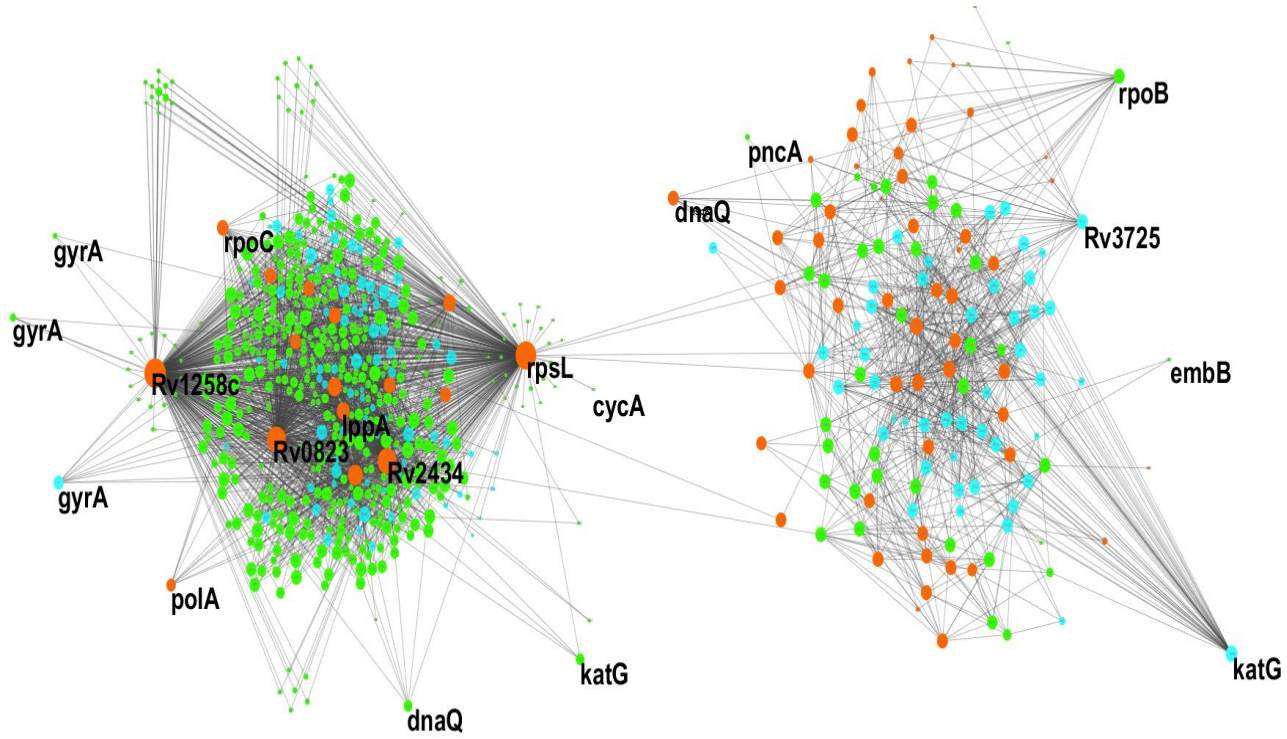

Figure 3. 5 Evolutionary network of DR mutations for the combined Asian clades. In this visualization, the size of the nodes are proportional to the node degree. The green-coloured nodes represent mutations where the evolutionary path has emerged, the blue nodes represent mutations where evolution is still in progress and orange nodes represent mutations where evolution has ended.

3.4.5 Lineage 4

In our analysis we further combined the European, American and Central Asian strains (Lineage 4.3, Haarlem and Ural clades) into one group Lineage 4. As expected, the combined data produced a more complex network as shown in Figure 3.6. From this network we identified some interesting mutations which were initially missed when the networks were created for individual clades. One such mutation is in *cycA*, which is linked to thirty five other mutations forming a sub-network that starts with mutations in genes that are involved in the pthiocerol dimycocerosate (PDIM) pathway, membrane transport and lipid metabolism. The *cycA* mutation forms an intermediary hub for several drug resistance development paths ending with an acquisition of 2 mutations in the hypothetical protein *Rv2323c*. There is a need to investigate the possible role of this gene in compensating for cycloserine resistance. Interestingly, the *cycA* mutation is also directly linked to another mutation in *rpoC*. Mutations in the *rpoC* gene have been known to play a role in mitigating fitness costs that are associated with MDR-TB [125]. Another mutation of interest that was identified in this network is located in the *embC* gene. The *emb* operon is a point of active research on the evolution of MDR-TB. Our previous networks have identified mutations only in the *B*

subunit of this operon. In this network, a mutation in *embC* preceded the acquisition of another mutation in *birA*, which is a biotin operon repressor. This evolutionary path ends up with an acquisition of 2 mutations in the membrane protein genes *mmpS1* (*Rv0403*). Interestingly, this pathway starts with resistance to AMK but terminates with further resistance to several other drugs leading to the XDR-TB phenotype. Interactions between these genes are likely linked to adaptive changes of the mycobacterial cell wall taking place during the evolution of drug resistance. Another mutation of interest in this clade is in *ddlA*, which encodes D-alanine-D-alanine ligase. This gene has been previously linked with cycloserine resistance [244]. In our analysis, *ddlA* mutations are followed by three mutations in other genes leading to the development of further resistance to other additional drugs. Two of these possible compensatory mutations are of unknown function while one is in the *isdB* gene, which is involved in the isoprenoid biosynthesis pathway. The current network also identified two important *gyrA* mutations, the first one was a terminal point for several drug resistance pathways that started with mutations in the following genes: transcriptional regulatory protein Rv3833, the phosphate transporter *pitA*, which have been linked to the compensatory evolution [245] and also in the membrane transport protein *Rv3447c*. The second *gyrA* mutation is an intermediate step in a pathway that ends with an acquisition of a mutation in *Rv3383c* gene. The possible role of this gene in compensating for fluoroquinolone resistance fitness cost requires further investigation. Another path includes initial mutations in *rpsL* followed by mutations in several transporter genes and in *groEL* as it was in the Asian Mtb strains.

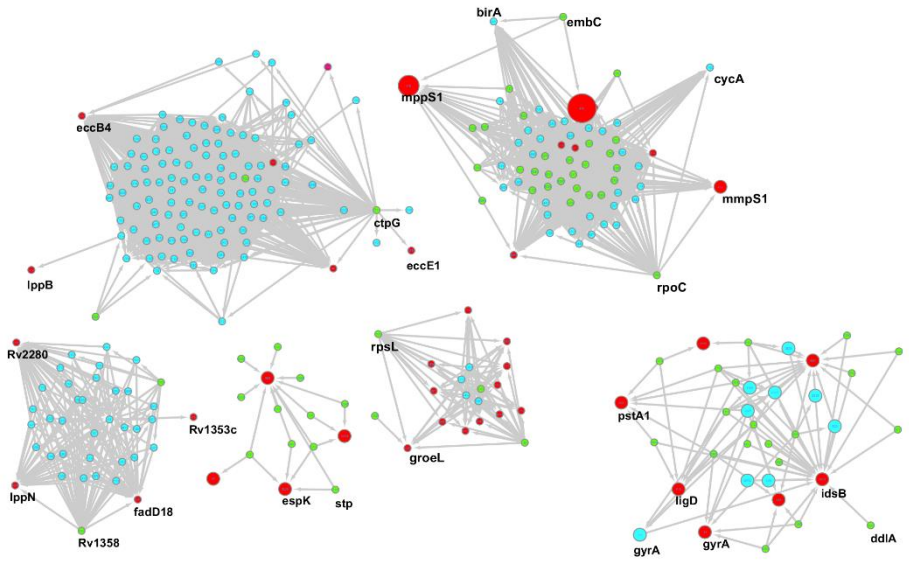

Figure 3. 6 Evolutionary network of DR mutations for the combined Lineage 4 clades, the green-coloured nodes represent mutations where the evolutionary path has emerged, the blue nodes represent mutations where evolution is

still in progress and orange nodes represent mutations where evolution has ended. The size of the nodes are proportional to the number of drugs associated with the mutation.

3.5 Discussion

The evolution of drug resistance from first unrecognized mutations to the fixation of high-level drug resistance in Mtb populations remains elusive. Although earlier studies have identified several genetic determinants of antibiotic resistance, especially for the first-line drugs, the use of WGS has shown us that the paths leading to drug resistance are much more complex than previously anticipated. Mtb employs several strategies as a response to exposure to anti-TB drugs. Research has shown that several factors, which include the clade-specific genetic background, fitness cost and compensatory mechanisms play a significant role in the evolution of drug resistance in Mtb. Unfortunately, we cannot directly measure some of these factors experimentally. Mathematical modelling can be used to elucidate the interplay between these factors and how they influence each other and the DR-TB resistance. In this study we used a statistical model to identify functional associations between mutations which were highly associated with the DR phenotype in Mtb. We did this by determining co-dependencies of pairs of mutations and using the results to infer the order of acquisition of these mutations. We then combined this information together with pDST information as an input for attributable risk networks, which were visualized using the Cytoscape software. In this context, we defined a new approach of inferring and analysis of consequent mutations associated with DR development via their attributable risk values.

Through the use of the attributable risk networks, we inferred the evolutionary trajectories of drug resistance development in different clades of Mtb. As we expected, the mechanisms governing the acquisition and fixation of drug resistance varied significantly between the clades. The analysis from these networks showed statistically reliable interactions between mutations in genes involving DNA replication-reparation systems, stress response, lipid metabolism and membrane transport genes as possible drivers of the evolution of drug resistance in Mtb. The Asian clade, particularly the Beijing lineage, was associated with the most complex network compared to other clades. A propensity to MDR-TB development in the Beijing lineage compared to other lineages can be linked to a partial disruption of DNA repair genes as a mechanism of increasing the genetic variability of the population. This may suggest that the evolution of drug resistance in this clade is fuelled by the elevated mutation rate as a result of altered DNA repair processes. This may also explain why the strains of this clade are more prone to developing into highly virulent and transmissible DR strains compared to strains from other lineages. This is further supported by our

finding in the Haarlem clade network, which also showed an involvement in DR development of mutations in DNA repair genes that was not reported before from the available literature. Compared to other strains of the Lineage 4, the strains of the Haarlem clade were more prone to develop MDR-TB. The Lineage 1.2 network was rather simple, involving just a few mutations. This observation explains why the strains of this clade are known to be mostly susceptible to antibiotics. The occurrence of the gyrase mutation pair in this clade may also explain the hyper-susceptibility of strains belonging in this clade. Aubry *et al.* also did a functional analysis study that reported on the link between novel gyrase mutations and hyper-susceptibility in Mtb strains [71].

In our study, we also paid a special attention to mutations in well-characterized drug target genes. Interestingly enough, most of the pathways that linked these DR mutations have terminated with the acquisition of mutations in the *katG* gene. This finding may suggest that the resistance to isoniazid emerges first before the bacteria acquires additional resistance to other drugs. Interestingly, isoniazid was one of the first drugs prescribed for TB and it may be possible that some *katG* mutations resulted from diversifying selection rather than purifying selection. This suggestion can also be strengthened when we consider the role of *rpsL* mutations as a central hub of DR in networks generated from both the Asian and Lineage 4 strains. The early use of the aminoglycoside streptomycin in TB mono-therapy might have also resulted in the fixation of the *rpsL* mutations.

The involvement of the mutations PPE and PE_PGRS genes was strongly suggested in our networks leading to a conclusion that some of these proteins may play a role in the evolution of DR in Mtb. In this study, a mutation in the PPE13 gene was the terminal point of several evolutionary pathways involving other MDR mutations. Previous studies have often discarded or ignored these abundant proteins due to their highly polymorphic nature, but as sequencing technologies improve, the value of investigating their role is immense.

From this study we can propose a DR evolutionary model in Mtb that starts with mechanisms that ensure that the bacterium is exposed only to sub-inhibitory concentrations of antibiotics. As a survival mechanism, Mtb activates stress response proteins that in turn affect the fidelity of the DNA processes. This leads to the acquisition of DR mutations, which are further compensated for by additional mutations in secondary genes. The interaction of several lipid biosynthesis genes and membrane transport genes also suggests that the evolution of DR in Mtb involves a reconstitution

of the mycobacterial cell wall in order to maximize the efficiency in systems that control cell wall permeability.

A study like this paves the way for future research that can be beneficial in the management of TB. Through the attributable risk network analysis we can enable identification of potential targets for DR-TB reversion drugs. There is also potential in identifying additional determinants of DR, which can be used as markers of intermediate steps of DR development in Mtb. Current pDST methods cannot detect the low level resistance and this often leads to inadequate dosage of anti-TB drugs and this in turn fuels the trajectory towards increased DR. Incorporating these in our catalogue of genetic determinants can further improve the performance of current WGS-based diagnostic tools. The information derived from analysing these networks can also be used in influencing therapeutic treatment regiments. For example, the low complexity of Lineage 1.2 networks may suggest that the treatment outcomes of patients infected with strains of this clade be improved with increased dosages of the standard regimen drugs, with a minimum risk of further fuelling drug resistance. On the other hand, the direct interaction between mutations in *katG* and *rpsL* genes in both the Asian and the Lineage 4 clades may suggest that clinicians prescribe a combination of isoniazid and second-line aminoglycosides with extreme caution to avoid fuelling the resistance of the antibiotics. The same is also true for regimens that require the use of fluoroquinolones and rifampicin simultaneously.

We acknowledge a number of limitations in our study. First of all, there are no existing pathway models for DR interactions in Mtb, which can be used to benchmark our findings. As data on DR studies in Mtb continues to accumulate, there is a need for an integration of the information that can be used to build pathway models that can capture the complexity of the evolution of DR in Mtb. This "resistome" knowledgebase will need a careful curation as well as the ability to be integrated with other tools that are used in Mtb research. Another limitation of this study was that no functional predictions were done to determine the effects of the identified variants on protein structure. The use of strain H37Rv as an alignment reference in this study may also have resulted in some lineages markers to be incorrectly identified as functional variants.

Future work will require multi-disciplinary approaches where the use of such models will guide experimentalists, while the data they generate will be used to improve the quality of the output from our models. The unavailability of pDST data especially for the second line drugs also was a limitation in the construction of our statistical models as well as in selecting input parameters for these models. The quality of the output from mathematical models is largely dependent on the

quality of data that is used to formulate the model structure and to inform model parameters [130]. Another usual pitfall associated with mathematical models is the desire to accurately integrate the multiple facets of biological complexity, without becoming intensively technical and too tedious to parameterize.

Chapter 4-Concluding Remarks

6.1 Major findings

Here we have used WGS and statistical models to show that the antibiotic resistant phenotype in Mtb is associated with a specific pattern of multiple polymorphic sites in the Mtb genome. This is contrary to traditional dogma that was based on the once-off acquisition of a drug-resistance-conferring mutation as the primary cause of drug resistance in Mtb. We further showed that these patterns of polymorphisms which are acquired in a stepwise manner can be used as reliable signatures of drug resistance even when only fragments of the genome are available for analysis. We demonstrated the potential use of these findings in further improving the rapidity of using WGS-based DR-TB diagnostic tools by cutting the turnaround time required to diagnose DR-TB. This was done by developing the Resistance Sniffer tool which can be used to predict drug resistance in Mtb using partial and complete Mtb sequences in a variety of file formats that are produced in different stages of the genome completion process. By integrating all the polymorphisms associated with resistance to 13 drugs in different clades of Mtb into a single diagnosis key, it was possible to simultaneously predict the lineage as well as the drug resistance profile of an Mtb sample.

In this study we also identified clade-specific patterns of polymorphisms that are associated with the drug resistant phenotype. By comparing the identified patterns of polymorphisms, our data indicates that the emergence of a DR-associated mutation facilitates the acquisition of other additional mutations resulting in the further development of drug resistance. This finding indicates that drug resistance is rather continuous and not binary and highlights a major limitation in the way drug resistance is often reported (resistant or susceptible) with no indication of strains that might have developed intermediate resistance. This is problematic as it often results in clinicians oversimplifying resistance leading to the prescription of inadequate TB regimens that have the potential to select for DR strains. We addressed this problem when we were developing our Resistance Sniffer program by ensuring that the end user has an understanding of the level of resistance that is being reported in the output.

6.1 The role of compensatory evolution

Strains belonging to lineage 1.2 have been associated with high drug susceptibility and improved treatment outcomes when compared to strains from the East Asian, Beijing and Euro-American sublineages. Our data indicates that epistasis may be a major factor in the evolution of drug resistance in different clades of Mtb. In order to interpret the differences in the role of epistasis in

the different clades of Mtb, we used Levin's attributable risk statistics to identify functional associations between DR affiliated mutations. From these data, we generated co-dependence networks for the different Mtb clades. From these networks we determined that not only compensatory mutations are responsible for the evolution of drug resistance in Mtb, instead we also identified some prerequisite mutations mainly in genes that are responsible for cell wall physiology and efflux pump systems. These mutations are likely to be responsible for ensuring that only sub-inhibitory concentrations of the antibiotics are absorbed by the bacterial cell. This suggestion is further cemented by the observed functional associations between mutations in the *emb* operon and several of these efflux pump mutations. Accurate dosage of ethambutol is notoriously difficult as it does not distribute easily in adipose tissue [246]. Our data indicates that the beginning of the evolution of drug resistance in Mtb is often associated with the emergence of low-level ethambutol resistance.

Networks from the Asian clades also consisted of several mutations in genes that are responsible for DNA fidelity. These mutations though to a lesser extent, were also identified in the Euro-American networks. This finding indicates that the association between MDR-TB and the Asian clades can be attributed to increased mutation rates in this clade. Our data also showed the involvement of several mutations in stress response as well as lipid metabolism genes. Finally, we developed a model of the evolution of Mtb drug resistance based on our observations of the mutation functional associations in the different clades. From this model we can suggest an evolutionary path that starts by the exposure of the bacterial cell to subinhibitory concentrations of antibiotics, this elicits the stress response mechanism most likely as a result of exposure to INH. This leads to the activation of error-prone DNA fidelity mechanisms resulting in the elevation of the basal mutation rates. The elevated mutation rates facilitates for the acquisition of additional mutations eventually leading to the development of high level MDR-TB. Our model emphasizes on the importance of adequate antibiotic dosage in the prevention of drug resistance and this has been supported by several studies in clinical acquired resistance [247-249]. Based on our findings, elucidating the mechanisms that link the DNA fidelity systems to fluctuations in mutation rates in Mtb which remain generally unknown would be invaluable.

6.2 Concluding remarks

The Resistant Sniffer program is an online tool for the rapid prediction of antibiotic resistance in Mtb using complete genome as well as partial genome sequences. This project resulted in the following achievements:

1. We identified patterns of polymorphisms in Mtb genomes which are associated with the drug resistant phenotype. We investigated how these patterns can be used as markers of drug resistance in a lineage specific way and estimated the reliability of these diagnostic polymorphisms for the lineage and drug resistance prediction. The hypothesis was that contrary to the traditional dogma of a once off acquisition of a DR mutation leading to DR-TB, the antibiotic resistance phenotype is associated with clade-specific patterns of multiple polymorphic sites in Mtb genomes. The study demonstrated that these polymorphic sites can be used to identify the drug susceptibility profile as well as the lineages of Mtb strains even when only fragments of the genome were available for analysis. Using a GWAS derivative, we took a clade-based approach to identify patterns of polymorphic patterns, which were associated with resistance to 13 anti-TB drugs. These patterns were carefully curated to develop a diagnosis key for the Resistance Sniffer program
2. A software tool for the rapid prediction of antibiotic resistance to 13 anti-TB drugs was successfully developed. The web address for the Resistant Sniffer program is http://resistance-sniffer.bi.up.ac.za/Mycobacterium_tuberculosis. The interface allows the users to upload NGS sequences in different widely used formats. For larger files, the user has the option to receive their results via email. The site also offers the user to download a compressed version of the program which can be used on a personal computer provided that Python 2.5 or 2.7 is installed. The site also comes with a help link to guide the users on how to run the program offline as outlined in Chapter 2.
3. The Resistance Sniffer tool was validated using our validation dataset from the GMTV database and testing was done on our independent testing dataset from the South African Medical Research Council. We showed that the TB pandemic in South Africa may have been brought by European migration during colonial times while the resistance-prone Asian strains may have been brought to the country much later as a result of improved travel due to globalization. Our study also showed that the first-line drug predictions were associated with higher specificity and sensitivity compared to the second-line drugs whose mechanisms of action is poorly understood.

4. The use of NGS in the management of DR-TB is limited by our understanding of the genetic mechanisms that drive the evolution of drug resistance in MTB. As set out by the final goal of this project, we attempted to develop an evolutionary model for the emergence of MDR-TB. Using Levin's attributable risk statistics, we identified functional associations between mutations that had been linked with drug resistance in the different clades of Mtb. We filtered out those associations that may have arisen due to genetic events and integrated the remaining data with annotation information as well as drug susceptibility information from the GMTV and TBDream databases to infer evolutionary networks based on the co-dependencies of the mutations. Additionally we analysed these clade-specific networks to identify harbinger as well as compensatory mutations that are involved in the evolution of MDR-TB. We confirmed the role of epistasis in the development of drug resistance in Mtb. We also identified some polymorphisms in secondary sites that can be incorporated into the diagnostic key of our Resistance Sniffer program in order to improve its accuracy.
5. The outcomes of this projects highlights the attractiveness of using NGS in the fight against DR-TB. The approach we took when we developed our online prediction not only addresses the challenge of the bioinformatics skills deficiency among clinicians but also provides a broader range in terms of the variety of NGS file formats that can be analysed on our platform. The output from our program also captures strains that may exhibit low to medium levels of resistance thus allowing clinicians to carefully monitor patients who are at risk of developing MDR-TB in future. In developing this program we were also aware of the rapid expansion in the knowledge base surrounding the genetic mechanisms of Mtb resistance, for this reason we ensured that our system can be easily updated by modifying the text based diagnosis key. This flexibility is not only limited to Mtb only, our system can be easily modified to analyse for other pathogens as well.
6. In line with the final goal this project we successfully created a model of the evolutionary trajectory of MDR-TB and assisted in understanding why some strains from certain lineages tend to be highly associated with MDR-TB.

A manuscript was prepared by summarizing the results of this study and submitted for review to the International Journal of Medical Microbiology.

6.3 Future perspectives

The magnitude of the global Mtb drug resistance burden remains a cause for concern with successful treatment outcomes estimated at 60% for MDR-TB and 35% for XDR-TB respectively [250, 251]. Given the well-documented association between DR-TB and mortality, tackling the drug resistance challenge is essential to curbing the high morbidity and mortality rates associated with TB. As global cases of MDR-TB and XDR-TB continue to rise, the need for new innovative diagnostic and treatment strategies becomes imperative. Fortunately, the advances in NGS technologies are proving to be a game changer in DR-TB detection. However, in order to fully exploit the potential of this platform in DR-TB diagnosis, there is a need to clarify the relationship between the genotype, phenotype and clinical outcomes. It is encouraging to note that initiatives such as the ReSeqTB [252] and the Comprehensive Resistance Prediction for Tuberculosis (CRypTIC) [253] have been set up to standardize the use of NGS in DR-TB control. The WHO has already implemented WGS in drug resistance surveillance and is planning to evaluate sequencing technologies for routine genotypic DST in 2019 [1, 254]. Several countries (for example, the United Kingdom and the Netherlands) have also adopted WGS-guided solutions in their public health systems with more countries expected to follow suit [254]. Our partnering team at the SA-MRC (Pretoria) is planning to create a *M. tuberculosis* genome variation database that will integrate clinical, epidemiological and microbiological data with genome variations based on WGS data. We expect that such a project will benefit immensely from our predictive tool. We also expect the project to generate more Mtb sequence data which in turn will be used to update the Resistance Sniffer program and therefore improve the sensitivity and specificity of the program. The flexibility of our approach can also be exploited by end-users for other applications such as routine molecular epidemiology investigations, laboratory cross contamination assessment and the diagnosis of other pathogens [31, 255]. Performing NGS-based diagnosis directly from sputum would improve the turnaround time for DR-TB diagnosis even further as it bypasses the time-consuming process of obtaining Mtb cultures before diagnosis. However, this is challenging due to the contamination of Mtb DNA by human DNA [256].

There is also a need to reconsider the “One size fits all” approach that has traditionally governed TB treatment. Several studies have recently recommended the need for individualized therapy for TB patients as a way for improving patient outcomes [215, 256]. Already rapid diagnosis platforms like Resistance Sniffer are able to simultaneously report on a wide range of drugs which makes it easier for clinicians to make well-informed decisions when crafting individualized treatment regimens based on patient outcomes. The recent interest in WGS-based DR-TB research has been

encouraging as witnessed by the number of publications that have been generated recently. In future, the TB research community will benefit immensely from the implementation of a "gold standard" to WGS data analysis in DR-TB diagnosis as it will ensure reproducibility and comparability of the different approaches to diagnosis.

References

1. World Health Organization, *Global tuberculosis report 2018*. 2018, World Health Organization,: Geneva. p. 1 - 8.
2. Georghiou, S., et al., *The use of next-generation sequencing technologies for the detection of mutations associated with drug resistance in Mycobacterium tuberculosis complex: technical guide*. 2018.
3. Van der Spuy, G., et al., *Changing Mycobacterium tuberculosis population highlights clade-specific pathogenic characteristics.* Tuberculosis, 2009. **89**(2): p. 120-125.
4. Filliol, I., et al., *Global phylogeny of Mycobacterium tuberculosis based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set.* J Bacteriol, 2006. **188**(2): p. 759-72.
5. Cole, S.T., et al., *Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.* Nature, 1998. **393**(6685): p. 537-44.
6. Youmans, G.P. and A.G. Karlson, *Streptomycin sensitivity of tubercle bacilli: Studies on recently isolated tubercle bacilli and the development of resistance to streptomycin in vivo.* American review of tuberculosis, 1947. **55**(6): p. 529-535.
7. DUNNER, E., W.B. BROWN, and J. WALLACE, *The effect of streptomycin with para-amino salicylic acid on the emergence of resistant strains of tubercle bacilli.* Diseases of the chest, 1949. **16**(6): p. 661-666.
8. Gandhi, N.R., et al., *Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa.* Lancet, 2006. **368**(9547): p. 1575-80.
9. Coll, F., et al., *Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences.* Genome Med, 2015. **7**(1): p. 51.
10. Yang, T.W., et al., *Side effects associated with the treatment of multidrug-resistant tuberculosis at a tuberculosis referral hospital in South Korea: A retrospective study.* Medicine, 2017. **96**(28): p. e7482.
11. Shean, K., et al., *Drug-associated adverse events and their relationship with outcomes in patients receiving treatment for extensively drug-resistant tuberculosis in South Africa.* PLoS One, 2013. **8**(5): p. e63057.
12. Walker, T.M., et al., *Whole genome sequencing for M/XDR tuberculosis surveillance and for resistance testing.* Clin Microbiol Infect, 2017. **23**(3): p. 161-166.
13. Demers, A.M., et al., *Direct Susceptibility Testing of Mycobacterium tuberculosis for Pyrazinamide by Use of the Bactec MGIT 960 System.* J Clin Microbiol, 2016. **54**(5): p. 1276-81.
14. Theron, G., et al., *Accuracy and impact of Xpert MTB/RIF for the diagnosis of smear-negative or sputum-scarce tuberculosis using bronchoalveolar lavage fluid.* Thorax, 2013. **68**(11): p. 1043-51.
15. Organization, W.H., *Global Tuberculosis Report 2018*. 2018.
16. Walker, T.M., et al., *Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study.* Lancet Infect Dis, 2015. **15**(10): p. 1193-1202.
17. Trauner, A., et al., *Evolution of drug resistance in tuberculosis: recent progress and implications for diagnosis and therapy.* Drugs, 2014. **74**(10): p. 1063-72.
18. Group, N.T.W., *Multidrug-resistant and extensively drug-resistant tuberculosis: the National Institute of Allergy and Infectious Diseases Research agenda and*

recommendations for priority research. The Journal of infectious diseases, 2008. **197**(11): p. 1493-1498.
19. Connolly, L.E., P.H. Edelstein, and L. Ramakrishnan, *Why is long-term therapy required to cure tuberculosis?* PLoS medicine, 2007. **4**(3): p. e120.
20. Barry 3rd, C.E., et al., *The spectrum of latent tuberculosis: rethinking the biology and intervention strategies.* Nature Reviews Microbiology, 2009. **7**(12): p. 845.
21. Rittershaus, E.S., S.-H. Baek, and C.M. Sassetti, *The normalcy of dormancy: common themes in microbial quiescence.* Cell host & microbe, 2013. **13**(6): p. 643-651.
22. Koch, A., V. Mizrahi, and D.F. Warner, *The impact of drug resistance on Mycobacterium tuberculosis physiology: what can we learn from rifampicin?* Emerging microbes & infections, 2014. **3**(1): p. 1-11.
23. Dookie, N., et al., *Evolution of drug resistance in Mycobacterium tuberculosis: a review on the molecular determinants of resistance and implications for personalized care.* Journal of Antimicrobial Chemotherapy, 2018. **73**(5): p. 1138-1151.
24. Palomino, J. and A. Martin, *Drug resistance mechanisms in Mycobacterium tuberculosis.* Antibiotics, 2014. **3**(3): p. 317-340.
25. Almeida Da Silva, P.E. and J.C. Palomino, *Molecular basis and mechanisms of drug resistance in Mycobacterium tuberculosis: classical and new drugs.* Journal of antimicrobial chemotherapy, 2011. **66**(7): p. 1417-1430.
26. Rattan, A., A. Kalia, and N. Ahmad, *Multidrug-resistant Mycobacterium tuberculosis: molecular perspectives.* Emerging infectious diseases, 1998. **4**(2): p. 195.
27. Mitchison, D.A., *Basic mechanisms of chemotherapy.* Chest, 1979. **76**(6): p. 771-780.
28. Yuen, L.K., D. Leslie, and P.J. Coloe, *Bacteriological and molecular analysis of rifampin-resistant Mycobacterium tuberculosis strains isolated in Australia.* Journal of Clinical Microbiology, 1999. **37**(12): p. 3844-3850.
29. Traore, H., et al., *Detection of rifampicin resistance in Mycobacterium tuberculosis isolates from diverse countries by a commercial line probe assay as an initial indicator of multidrug resistance.* The international journal of tuberculosis and lung disease, 2000. **4**(5): p. 481-484.
30. Van Niekerk, K., et al., *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*. 2018, IntechOpen: London. p. 79 - 98.
31. World Health Organization, *The use of next-generation sequencing technologies for the detection of mutations associated with drug resistance in Mycobacterium tuberculosis complex: technical guide*. 2018, World Health Organization: Geneva. p. 1 -10.
32. Banerjee, A., et al., *inhA, a gene encoding a target for isoniazid and ethionamide in Mycobacterium tuberculosis.* Science, 1994. **263**(5144): p. 227-230.
33. Larsson, P., et al., *The complete genome sequence of Francisella tularensis, the causative agent of tularemia.* Nat Genet, 2005. **37**.
34. Gegia, M., et al., *Treatment of isoniazid-resistant tuberculosis with first-line drugs: a systematic review and meta-analysis.* The Lancet Infectious Diseases, 2017. **17**(2): p. 223-234.
35. Manson, A.L., et al., *Genomic analysis of globally diverse Mycobacterium tuberculosis strains provides insights into the emergence and spread of multidrug resistance.* Nat Genet, 2017. **49**(3): p. 395-402.
36. Takayama, K. and J.O. Kilburn, *Inhibition of synthesis of arabinogalactan by ethambutol in Mycobacterium smegmatis.* Antimicrobial agents and chemotherapy, 1989. **33**(9): p. 1493-1499.

37. Mikusova, K., et al., *Biogenesis of the mycobacterial cell wall and the site of action of ethambutol.* Antimicrobial agents and chemotherapy, 1995. **39**(11): p. 2484-2489.
38. Zhao, L.-l., et al., *Analysis of embCAB mutations associated with ethambutol resistance in multidrug-resistant Mycobacterium tuberculosis isolates from China.* Antimicrobial agents and chemotherapy, 2015. **59**(4): p. 2045-2050.
39. Hazbón, M.H., et al., *Role of embB codon 306 mutations in Mycobacterium tuberculosis revisited: a novel association with broad drug resistance and IS6110 clustering rather than ethambutol resistance.* Antimicrobial agents and chemotherapy, 2005. **49**(9): p. 3794-3802.
40. Ahmad, S., A.-A. Jaber, and E. Mokaddas, *Frequency of embB codon 306 mutations in ethambutol-susceptible and-resistant clinical Mycobacterium tuberculosis isolates in Kuwait.* Tuberculosis, 2007. **87**(2): p. 123-129.
41. Giri, A., et al., *Polymorphisms in Rv3806c (ubiA) and the upstream region of embA in relation to ethambutol resistance in clinical isolates of Mycobacterium tuberculosis from North India.* Tuberculosis (Edinb), 2018. **108**: p. 41-46.
42. Whitfield, M.G., et al., *A Global Perspective on Pyrazinamide Resistance: Systematic Review and Meta-Analysis.* PLoS One, 2015. **10**(7): p. e0133869.
43. Shehzad, A., et al., *Challenges in the development of drugs for the treatment of tuberculosis.* The Brazilian journal of infectious diseases, 2013. **17**(1): p. 74-81.
44. Zhang, Y. and D. Mitchison, *The curious characteristics of pyrazinamide: a review.* The international journal of tuberculosis and lung disease, 2003. **7**(1): p. 6-21.
45. Zimhony, O., et al., *Pyrazinoic acid and its n-propyl ester inhibit fatty acid synthase type I in replicating tubercle bacilli.* Antimicrobial agents and chemotherapy, 2007. **51**(2): p. 752-754.
46. Peterson, N.D., et al., *Uncoupling environmental pH and intrabacterial acidification from pyrazinamide susceptibility in Mycobacterium tuberculosis.* Antimicrobial agents and chemotherapy, 2015. **59**(12): p. 7320-7326.
47. Dillon, N.A., et al., *Anti-tubercular Activity of Pyrazinamide is Independent of trans-Translation and RpsA.* Scientific reports, 2017. **7**(1): p. 6135.
48. Miotto, P., et al., *Drug resistance mechanisms and drug susceptibility testing for tuberculosis.* Respirology, 2018. **23**(12): p. 1098-1113.
49. Miotto, P., et al., *Mycobacterium tuberculosis pyrazinamide resistance determinants: a multicenter study.* MBio, 2014. **5**(5): p. e01819-14.
50. Whitfield, M.G., et al., *Mycobacterium tuberculosis pncA polymorphisms that do not confer pyrazinamide resistance at a breakpoint concentration of 100 micrograms per milliliter in MGIT.* Journal of clinical microbiology, 2015. **53**(11): p. 3633-3635.
51. Zhang, S., et al., *Mutations in panD encoding aspartate decarboxylase are associated with pyrazinamide resistance in Mycobacterium tuberculosis.* Emerging microbes & infections, 2013. **2**(6): p. e34.
52. Shi, W., et al., *Aspartate decarboxylase (PanD) as a new target of pyrazinamide in Mycobacterium tuberculosis.* Emerging microbes & infections, 2014. **3**(1): p. 1-8.
53. Gopal, P., et al., *Pyrazinoic acid inhibits mycobacterial coenzyme a biosynthesis by binding to aspartate decarboxylase PanD.* ACS infectious diseases, 2017. **3**(11): p. 807-819.
54. Spies, F.S., et al., *Identification of mutations related to streptomycin resistance in clinical isolates of Mycobacterium tuberculosis and possible involvement of efflux mechanism.* Antimicrobial agents and chemotherapy, 2008. **52**(8): p. 2947-2949.
55. Okamoto, S., et al., *Loss of a conserved 7-methylguanosine modification in 16S rRNA confers low-level streptomycin resistance in bacteria.* Molecular microbiology, 2007. **63**(4): p. 1096-1106.

56. Maus, C.E., B.B. Plikaytis, and T.M. Shinnick, *Mutation of tlyA confers capreomycin resistance in Mycobacterium tuberculosis.* Antimicrobial agents and chemotherapy, 2005. **49**(2): p. 571-577.
57. Alangaden, G.J., et al., *Mechanism of resistance to amikacin and kanamycin in Mycobacterium tuberculosis.* Antimicrobial agents and chemotherapy, 1998. **42**(5): p. 1295-1297.
58. Suzuki, Y., et al., *Detection of Kanamycin-ResistantMycobacterium tuberculosis by Identifying Mutations in the 16S rRNA Gene.* Journal of clinical microbiology, 1998. **36**(5): p. 1220-1225.
59. Johansen, S.K., et al., *Capreomycin binds across the ribosomal subunit interface using tlyA-encoded 2'-O-methylations in 16S and 23S rRNAs.* Molecular cell, 2006. **23**(2): p. 173-182.
60. Georghiou, S.B., et al., *Evaluation of genetic mutations associated with Mycobacterium tuberculosis resistance to amikacin, kanamycin and capreomycin: a systematic review.* PloS one, 2012. **7**(3): p. e33275.
61. Zaunbrecher, M.A., et al., *Overexpression of the chromosomally encoded aminoglycoside acetyltransferase eis confers kanamycin resistance in Mycobacterium tuberculosis.* Proceedings of the National Academy of Sciences, 2009. **106**(47): p. 20004-20009.
62. Campbell, P.J., et al., *Molecular detection of mutations associated with first-and second-line drug resistance compared with conventional drug susceptibility testing of Mycobacterium tuberculosis.* Antimicrobial agents and chemotherapy, 2011. **55**(5): p. 2032-2041.
63. Li, Y., et al., *Cycloserine for treatment of multidrug-resistant tuberculosis: a retrospective cohort study in China.* Infection and drug resistance, 2019. **12**: p. 721.
64. Epstein, I., K. Nair, and L. Boyd, *Cycloserine, a new antibiotic, in the treatment of human pulmonary tuberculosis: a preliminary report.* Antibiotic Med., 1955. **1**(2): p. 80-93.
65. Desjardins, C.A., et al., *Genomic and functional analyses of Mycobacterium tuberculosis strains implicate ald in D-cycloserine resistance.* Nature genetics, 2016. **48**(5): p. 544.
66. Dheda, K., et al., *The epidemiology, pathogenesis, transmission, diagnosis, and management of multidrug-resistant, extensively drug-resistant, and incurable tuberculosis.* The lancet Respiratory medicine, 2017. **5**(4): p. 291-360.
67. Takiff, H.E., et al., *Cloning and nucleotide sequence of Mycobacterium tuberculosis gyrA and gyrB genes and detection of quinolone resistance mutations.* Antimicrobial agents and chemotherapy, 1994. **38**(4): p. 773-780.
68. Cheng, A.F., et al., *Multiplex PCR amplimer conformation analysis for rapid detection of gyrA mutations in fluoroquinolone-resistant Mycobacterium tuberculosis clinical isolates.* Antimicrobial agents and chemotherapy, 2004. **48**(2): p. 596-601.
69. Maruri, F., et al., *A systematic review of gyrase mutations associated with fluoroquinolone-resistant Mycobacterium tuberculosis and a proposed gyrase numbering system.* Journal of Antimicrobial Chemotherapy, 2012. **67**(4): p. 819-831.
70. Musser, J.M., *Antimicrobial agent resistance in mycobacteria: molecular genetic insights.* Clinical microbiology reviews, 1995. **8**(4): p. 496-514.
71. Aubry, A., et al., *Novel gyrase mutations in quinolone-resistant and-hypersusceptible clinical isolates of Mycobacterium tuberculosis: functional analysis of mutant enzymes.* Antimicrobial Agents and Chemotherapy, 2006. **50**(1): p. 104-112.
72. DeBarber, A.E., et al., *Ethionamide activation and sensitivity in multidrug-resistant Mycobacterium tuberculosis.* Proceedings of the National Academy of Sciences, 2000. **97**(17): p. 9677-9682.

73. Brossier, F., et al., *Molecular investigation of resistance to the antituberculous drug ethionamide in multidrug-resistant clinical isolates of Mycobacterium tuberculosis.* Antimicrobial agents and chemotherapy, 2011. **55**(1): p. 355-360.
74. Axelsen, P.H., *Essentials of antimicrobial pharmacology: a guide to fundamentals for practice*. 2002: Springer Science & Business Media.
75. Rengarajan, J., et al., *The folate pathway is a target for resistance to the drug para-aminosalicylic acid (PAS) in mycobacteria.* Molecular microbiology, 2004. **53**(1): p. 275-282.
76. Zhao, F., et al., *Binding pocket alterations in dihydrofolate synthase confer resistance to para-aminosalicylic acid in clinical isolates of Mycobacterium tuberculosis.* Antimicrobial agents and chemotherapy, 2014. **58**(3): p. 1479-1487.
77. Andries, K., et al., *A diarylquinoline drug active on the ATP synthase of Mycobacterium tuberculosis.* Science, 2005. **307**(5707): p. 223-227.
78. Field, S.K., *Bedaquiline for the treatment of multidrug-resistant tuberculosis: great promise or disappointment?* Therapeutic advances in chronic disease, 2015. **6**(4): p. 170-184.
79. Ismail, N., et al., *Stepwise acquisition of rv0678 and atpE mutations conferring bedaquiline resistance: an in vitro study.* Antimicrobial agents and chemotherapy, 2019: p. AAC. 00292-19.
80. BROWNE, S.G. and L. Hogerzeil, *" B 663" in the treatment of leprosy. Preliminary report of a pilot trial.* Leprosy review, 1962. **33**(1): p. 6-10.
81. Barry, V.C., et al., *A new series of phenazines (rimino-compounds) with high antituberculosis activity.* Nature, 1957. **179**(4568): p. 1013-1015.
82. Lu, Y., et al., *Activities of clofazimine against Mycobacterium tuberculosis in vitro and in vivo.* Zhonghua jie he he hu xi za zhi= Zhonghua jiehe he huxi zazhi= Chinese journal of tuberculosis and respiratory diseases, 2008. **31**(10): p. 752-755.
83. Lechartier, B. and S.T. Cole, *Mode of action of clofazimine and combination therapy with benzothiazinones against Mycobacterium tuberculosis.* Antimicrobial agents and chemotherapy, 2015. **59**(8): p. 4457-4463.
84. Xu, J., et al., *Primary clofazimine and bedaquiline resistance among isolates from patients with multidrug-resistant tuberculosis.* Antimicrobial agents and chemotherapy, 2017. **61**(6): p. e00239-17.
85. Zhang, S., et al., *Identification of novel mutations associated with clofazimine resistance in Mycobacterium tuberculosis.* Journal of Antimicrobial Chemotherapy, 2015. **70**(9): p. 2507-2510.
86. Leach, K.L., et al., *Linezolid, the first oxazolidinone antibacterial agent.* Annals of the New York Academy of Sciences, 2011. **1222**(1): p. 49-54.
87. Zhang, Y., *The magic bullets and tuberculosis drug targets.* Annu. Rev. Pharmacol. Toxicol., 2005. **45**: p. 529-564.
88. Hillemann, D., S. Rüsch-Gerdes, and E. Richter, *In vitro-selected linezolid-resistant Mycobacterium tuberculosis mutants.* Antimicrobial agents and chemotherapy, 2008. **52**(2): p. 800-801.
89. Bloemberg, G.V., et al., *Acquired resistance to bedaquiline and delamanid in therapy for tuberculosis.* New England Journal of Medicine, 2015. **373**(20): p. 1986-1988.
90. Zimenkov, D.V., et al., *Examination of bedaquiline-and linezolid-resistant Mycobacterium tuberculosis isolates from the Moscow region.* Journal of Antimicrobial Chemotherapy, 2017. **72**(7): p. 1901-1906.
91. Sun, Z., et al., *Comparison of gyrA gene mutations between laboratory-selected ofloxacin-resistant Mycobacterium tuberculosis strains and clinical isolates.* International journal of antimicrobial agents, 2008. **31**(2): p. 115-121.

92. Stover, C.K., et al., *A small-molecule nitroimidazopyran drug candidate for the treatment of tuberculosis.* Nature, 2000. **405**(6789): p. 962.
93. Carlos Palomino, J. and A. Martin, *Tuberculosis clinical trial update and the current anti-tuberculosis drug portfolio.* Current medicinal chemistry, 2013. **20**(30): p. 3785-3796.
94. Laskin, D.L., et al., *Macrophages, reactive nitrogen species, and lung injury.* Annals of the New York Academy of Sciences, 2010. **1203**: p. 60-65.
95. Manjunatha, U.H., et al., *Identification of a nitroimidazo-oxazine-specific protein involved in PA-824 resistance in Mycobacterium tuberculosis.* Proceedings of the National Academy of Sciences, 2006. **103**(2): p. 431-436.
96. Singh, R., et al., *PA-824 kills nonreplicating Mycobacterium tuberculosis by intracellular NO release.* Science, 2008. **322**(5906): p. 1392-1395.
97. Matsumoto, M., et al., *OPC-67683, a nitro-dihydro-imidazooxazole derivative with promising action against tuberculosis in vitro and in mice.* PLoS medicine, 2006. **3**(11): p. e466.
98. Oppong, Y.E., et al., *Genome-wide analysis of Mycobacterium tuberculosis polymorphisms reveals lineage-specific associations with drug resistance.* BMC genomics, 2019. **20**(1): p. 252.
99. Chen, J., et al., *Identification of novel mutations associated with cycloserine resistance in Mycobacterium tuberculosis.* Journal of Antimicrobial Chemotherapy, 2017. **72**(12): p. 3272-3276.
100. Telenti, A., et al., *The emb operon, a gene cluster of Mycobacterium tuberculosis involved in resistance to ethambutol.* Nature medicine, 1997. **3**(5): p. 567.
101. Hicks, N.D., et al., *Bacterial Genome-Wide Association Identifies Novel Factors That Contribute to Ethionamide and Prothionamide Susceptibility in Mycobacterium tuberculosis.* mBio, 2019. **10**(2): p. e00616-19.
102. Ramaswamy, S.V., et al., *Single nucleotide polymorphisms in genes associated with isoniazid resistance in Mycobacterium tuberculosis.* Antimicrobial agents and chemotherapy, 2003. **47**(4): p. 1241-1250.
103. Brzostek, A., et al., *Molecular characterisation of streptomycin-resistant Mycobacterium tuberculosis strains isolated in Poland.* The International Journal of Tuberculosis and Lung Disease, 2004. **8**(8): p. 1032-1035.
104. Reeves, A.Z., et al., *Aminoglycoside cross-resistance in Mycobacterium tuberculosis due to mutations in the 5' untranslated region of whiB7.* Antimicrobial agents and chemotherapy, 2013. **57**(4): p. 1857-1865.
105. Weinreich, D.M., et al., *Darwinian evolution can follow only very few mutational paths to fitter proteins.* science, 2006. **312**(5770): p. 111-114.
106. Thomas, V.L., A.C. McReynolds, and B.K. Shoichet, *Structural bases for stability–function tradeoffs in antibiotic resistance.* Journal of molecular biology, 2010. **396**(1): p. 47-59.
107. Lozovsky, E.R., et al., *Stepwise acquisition of pyrimethamine resistance in the malaria parasite.* Proceedings of the National Academy of Sciences, 2009. **106**(29): p. 12025-12030.
108. Gagneux, S., et al., *Impact of bacterial genetics on the transmission of isoniazid-resistant Mycobacterium tuberculosis.* PLoS Pathog, 2006. **2**(6): p. e61.
109. Müller, B., et al., *The heterogeneous evolution of multidrug-resistant Mycobacterium tuberculosis.* Trends in Genetics, 2013. **29**(3): p. 160-169.
110. Borrell, S. and S. Gagneux, *Infectiousness, reproductive fitness and evolution of drug-resistant Mycobacterium tuberculosis.* Int J Tuberc Lung Dis, 2009. **13**(12): p. 1456-66.

111. Borrell, S. and S. Gagneux, *Strain diversity, epistasis and the evolution of drug resistance in Mycobacterium tuberculosis.* Clinical Microbiology and Infection, 2011. **17**(6): p. 815-820.
112. Organization, W.H., *Global tuberculosis control: epidemiology, strategy, financing: WHO report 2009*. 2009: World Health Organization.
113. Trindade, S., et al., *Positive epistasis drives the acquisition of multidrug resistance.* PLoS genetics, 2009. **5**(7): p. e1000578.
114. Cordell, H.J., *Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans.* Human molecular genetics, 2002. **11**(20): p. 2463-2468.
115. Melnyk, A.H., A. Wong, and R. Kassen, *The fitness costs of antibiotic resistance mutations.* Evolutionary applications, 2015. **8**(3): p. 273-283.
116. Andersson, D.I. and D. Hughes, *Antibiotic resistance and its cost: is it possible to reverse resistance?* Nature Reviews Microbiology, 2010. **8**(4): p. 260.
117. zur Wiesch, P.A., et al., *Population biological principles of drug-resistance evolution in infectious diseases.* The Lancet infectious diseases, 2011. **11**(3): p. 236-247.
118. Shcherbakov, D., et al., *Directed mutagenesis of Mycobacterium smegmatis 16S rRNA to reconstruct the in vivo evolution of aminoglycoside resistance in Mycobacterium tuberculosis.* Molecular microbiology, 2010. **77**(4): p. 830-840.
119. Lehner, B., *Molecular mechanisms of epistasis within and between genes.* Trends in Genetics, 2011. **27**(8): p. 323-331.
120. Mortimer, T.D., A.M. Weber, and C.S. Pepperell, *Signatures of Selection at Drug Resistance Loci in Mycobacterium tuberculosis.* mSystems, 2018. **3**(1): p. e00108-17.
121. Farhat, M.R., et al., *Genomic analysis identifies targets of convergent positive selection in drug-resistant Mycobacterium tuberculosis.* Nature genetics, 2013. **45**(10): p. 1183.
122. Zhang, H., et al., *Genome sequencing of 161 Mycobacterium tuberculosis isolates from China identifies genes and intergenic regions associated with drug resistance.* Nature genetics, 2013. **45**(10): p. 1255.
123. Gagneux, S., et al., *The competitive cost of antibiotic resistance in Mycobacterium tuberculosis.* Science, 2006. **312**(5782): p. 1944-1946.
124. Lanzas, F., et al., *Multidrug-resistant tuberculosis in panama is driven by clonal expansion of a multidrug-resistant Mycobacterium tuberculosis strain related to the KZN extensively drug-resistant M. tuberculosis strain from South Africa.* Journal of clinical microbiology, 2013. **51**(10): p. 3277-3285.
125. De Vos, M., et al., *Putative compensatory mutations in the rpoC gene of rifampin-resistant Mycobacterium tuberculosis are associated with ongoing transmission.* Antimicrobial agents and chemotherapy, 2013. **57**(2): p. 827-832.
126. Casali, N., et al., *Microevolution of extensively drug-resistant tuberculosis in Russia.* Genome research, 2012. **22**(4): p. 735-745.
127. Brandis, G. and D. Hughes, *Genetic characterization of compensatory evolution in strains carrying rpoB Ser531Leu, the rifampicin resistance mutation most frequently found in clinical isolates.* Journal of Antimicrobial Chemotherapy, 2013. **68**(11): p. 2493-2497.
128. Brandis, G., et al., *Fitness-compensatory mutations in rifampicin-resistant RNA polymerase.* Molecular microbiology, 2012. **85**(1): p. 142-151.
129. Borrell, S., et al., *Epistasis between antibiotic resistance mutations drives the evolution of extensively drug-resistant tuberculosis.* Evolution, medicine, and public health, 2013. **2013**(1): p. 65-74.
130. Fonseca, J., G. Knight, and T. McHugh, *The complex evolution of antibiotic resistance in Mycobacterium tuberculosis.* International journal of infectious diseases, 2015. **32**: p. 94-100.

131. Louw, G., et al., *A balancing act: efflux/influx in mycobacterial drug resistance.* Antimicrobial agents and chemotherapy, 2009. **53**(8): p. 3181-3189.
132. Machado, D., et al., *Contribution of efflux to the emergence of isoniazid and multidrug resistance in Mycobacterium tuberculosis.* PloS one, 2012. **7**(4): p. e34538.
133. Balganesh, M., et al., *Rv1218c, an ABC transporter of Mycobacterium tuberculosis with implications in drug discovery.* Antimicrobial agents and chemotherapy, 2010. **54**(12): p. 5167-5172.
134. Louw, G.E., et al., *Rifampicin reduces susceptibility to ofloxacin in rifampicin-resistant Mycobacterium tuberculosis through efflux.* American journal of respiratory and critical care medicine, 2011. **184**(2): p. 269-276.
135. Aínsa, J.A., et al., *Molecular cloning and characterization of Tap, a putative multidrug efflux pump present in Mycobacterium fortuitum and Mycobacterium tuberculosis.* Journal of bacteriology, 1998. **180**(22): p. 5836-5843.
136. Srivastava, S., et al., *Efflux-Pump—Derived Multiple Drug Resistance to Ethambutol Monotherapy in Mycobacterium tuberculosis and the Pharmacokinetics and Pharmacodynamics of Ethambutol.* The Journal of infectious diseases, 2010. **201**(8): p. 1225-1231.
137. Martinez, J. and F. Baquero, *Mutation frequencies and antibiotic resistance.* Antimicrobial agents and chemotherapy, 2000. **44**(7): p. 1771-1777.
138. Webber, M. and L. Piddock, *The importance of efflux pumps in bacterial antibiotic resistance.* Journal of Antimicrobial Chemotherapy, 2003. **51**(1): p. 9-11.
139. Viveiros, M., et al., *Isoniazid-induced transient high-level resistance in Mycobacterium tuberculosis.* Antimicrobial agents and chemotherapy, 2002. **46**(9): p. 2804-2810.
140. Huitric, E., et al., *Resistance levels and rpoB gene mutations among in vitro-selected rifampin-resistant Mycobacterium tuberculosis mutants.* Antimicrobial agents and chemotherapy, 2006. **50**(8): p. 2860-2862.
141. Andries, K., et al., *Acquired resistance of Mycobacterium tuberculosis to bedaquiline.* PLOS one, 2014. **9**(7): p. e102135.
142. Dutta, N.K., M.L. Pinn, and P.C. Karakousis, *Reduced emergence of isoniazid resistance with concurrent use of thioridazine against acute murine tuberculosis.* Antimicrobial agents and chemotherapy, 2014. **58**(7): p. 4048-4053.
143. Rad, M.E., et al., *Mutations in putative mutator genes of Mycobacterium tuberculosis strains of the W-Beijing family.* Emerging infectious diseases, 2003. **9**(7): p. 838.
144. Stagg, H.R., et al., *A tale of two settings: the role of the Beijing genotype in the epidemiology of multidrug-resistant tuberculosis.* European Respiratory Journal, 2014. **43**(2): p. 632-635.
145. Jones, M., S. Thomas, and A. Rogers, *Luria-Delbrück fluctuation experiments: design and analysis.* Genetics, 1994. **136**(3): p. 1209-1216.
146. Luria, S.E. and M. Delbrück, *Mutations of bacteria from virus sensitivity to virus resistance.* Genetics, 1943. **28**(6): p. 491.
147. Ford, C.B., *The Evolution of Drug Resistant Mycobacterium Tuberculosis*. 2012, Havard.
148. Gillespie, S.H., et al., *Effect of subinhibitory concentrations of ciprofloxacin on Mycobacterium fortuitum mutation rates.* Journal of Antimicrobial Chemotherapy, 2005. **56**(2): p. 344-348.
149. O'sullivan, D., et al., *Mycobacterium tuberculosis DNA repair in response to subinhibitory concentrations of ciprofloxacin.* Journal of antimicrobial chemotherapy, 2008. **62**(6): p. 1199-1202.
150. Boshoff, H.I., et al., *The transcriptional responses of Mycobacterium tuberculosis to inhibitors of metabolism novel insights into drug mechanisms of action.* Journal of Biological Chemistry, 2004. **279**(38): p. 40174-40184.

151. Long, R., et al., *Empirical treatment of community-acquired pneumonia and the development of fluoroquinolone-resistant tuberculosis.* Clinical infectious diseases, 2009. **48**(10): p. 1354-1360.
152. Deutschendorf, C., L.Z. Goldani, and R.P. dos Santos, *Previous use of quinolones: a surrogate marker for first line anti-tuberculosis drugs resistance in HIV-infected patients?* The Brazilian Journal of Infectious Diseases, 2012. **16**(2): p. 142-145.
153. Chen, L.-C., et al., *Identifying co-targets to fight drug resistance based on a random walk model.* BMC systems biology, 2012. **6**(1): p. 5.
154. Waddell, S.J., et al., *The use of microarray analysis to determine the gene expression profiles of Mycobacterium tuberculosis in response to anti-bacterial compounds.* Tuberculosis, 2004. **84**(3-4): p. 263-274.
155. Gagneux, S., *Ecology and evolution of Mycobacterium tuberculosis.* Nature Reviews Microbiology, 2018. **16**(4): p. 202.
156. Zaczek, A., et al., *Genetic evaluation of relationship between mutations in rpoB and resistance of Mycobacterium tuberculosis to rifampin.* BMC microbiology, 2009. **9**(1): p. 10.
157. Bifani, P.J., et al., *Global dissemination of the Mycobacterium tuberculosis W-Beijing family strains.* Trends in microbiology, 2002. **10**(1): p. 45-52.
158. Drobniewski, F., et al., *Drug-resistant tuberculosis, clinical virulence, and the dominance of the Beijing strain family in Russia.* Jama, 2005. **293**(22): p. 2726-2731.
159. Fenner, L., et al., *Effect of mutation and genetic background on drug resistance in Mycobacterium tuberculosis.* Antimicrobial agents and chemotherapy, 2012. **56**(6): p. 3047-3053.
160. Baker, L., et al., *Molecular analysis of isoniazid-resistant Mycobacterium tuberculosis isolates from England and Wales reveals the phylogenetic significance of the ahpC− 46A polymorphism.* Antimicrobial agents and chemotherapy, 2005. **49**(4): p. 1455-1464.
161. Van Doorn, H., et al., *Public health impact of isoniazid-resistant Mycobacterium tuberculosis strains with a mutation at amino-acid position 315 of katG: a decade of experience in The Netherlands.* Clinical microbiology and infection, 2006. **12**(8): p. 769-775.
162. Castro, R.A.D., et al., *The Evolution of Fluoroquinolone-Resistance in Mycobacterium tuberculosis is Modulated by the Genetic Background.* bioRxiv, 2019: p. 659045.
163. Bandera, A., et al., *Molecular epidemiology study of exogenous reinfection in an area with a low incidence of tuberculosis.* Journal of clinical microbiology, 2001. **39**(6): p. 2213-2218.
164. de Viedma, D.G., et al., *Analysis of clonal composition of Mycobacterium tuberculosis isolates in primary infections in children.* Journal of clinical microbiology, 2004. **42**(8): p. 3415-3418.
165. Cohen, K.A., et al., *Deciphering drug resistance in Mycobacterium tuberculosis using whole-genome sequencing: progress, promise, and challenges.* Genome Med, 2019. **11**(1): p. 45.
166. Trauner, A., et al., *The within-host population dynamics of Mycobacterium tuberculosis vary with treatment efficacy.* Genome biology, 2017. **18**(1): p. 71.
167. Eldholm, V., et al., *Evolution of extensively drug-resistant Mycobacterium tuberculosis from a susceptible ancestor in a single patient.* Genome biology, 2014. **15**(11): p. 490.
168. Didelot, X., et al., *Within-host evolution of bacterial pathogens.* Nature Reviews Microbiology, 2016. **14**(3): p. 150.
169. Deurenberg, R.H., et al., *Application of next generation sequencing in clinical microbiology and infection prevention.* Journal of biotechnology, 2017. **243**: p. 16-24.

170. Daum, L.T., et al., *Next-generation ion torrent sequencing of drug resistance mutations in Mycobacterium tuberculosis strains.* Journal of clinical microbiology, 2012. **50**(12): p. 3831-3837.
171. Ford, C., et al., *Mycobacterium tuberculosis–heterogeneity revealed through whole genome sequencing.* Tuberculosis, 2012. **92**(3): p. 194-201.
172. Iketleng, T., et al., *Mycobacterium tuberculosis Next-Generation Whole Genome Sequencing: Opportunities and Challenges.* Tuberculosis research and treatment, 2018. **2018**.
173. Török, M.E., et al., *Rapid whole-genome sequencing for investigation of a suspected tuberculosis outbreak.* Journal of clinical microbiology, 2013. **51**(2): p. 611-614.
174. Koser, C.U., et al., *Whole-genome sequencing for rapid susceptibility testing of M. tuberculosis.* N Engl J Med, 2013. **369**(3): p. 290-2.
175. Witney, A.A., et al., *Clinical application of whole-genome sequencing to inform treatment for multidrug-resistant tuberculosis cases.* Journal of clinical microbiology, 2015. **53**(5): p. 1473-1483.
176. Rodwell, T.C., et al., *Predicting extensively drug-resistant Mycobacterium tuberculosis phenotypes with genetic mutations.* J Clin Microbiol, 2014. **52**(3): p. 781-9.
177. Cui, Z.-J., et al., *Bioinformatics Identification of Drug Resistance-Associated Gene Pairs in Mycobacterium tuberculosis.* International journal of molecular sciences, 2016. **17**(9): p. 1417.
178. Allix-Beguec, C., et al., *Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing.* N Engl J Med, 2018. **379**(15): p. 1403-1415.
179. Chen, P.E. and B.J. Shapiro, *The advent of genome-wide association studies for bacteria.* Current opinion in microbiology, 2015. **25**: p. 17-24.
180. Farhat, M.R., et al., *GWAS for quantitative resistance phenotypes in Mycobacterium tuberculosis reveals resistance genes and regulatory regions.* Nature communications, 2019. **10**(1): p. 2128.
181. Kato-Maeda, M., et al., *The nature and consequence of genetic variability within Mycobacterium tuberculosis.* The Journal of clinical investigation, 2001. **107**(5): p. 533-537.
182. Brosch, R., et al., *A new evolutionary scenario for the Mycobacterium tuberculosis complex.* Proceedings of the national academy of Sciences, 2002. **99**(6): p. 3684-3689.
183. Tsolaki, A.G., et al., *Functional and evolutionary genomics of Mycobacterium tuberculosis: insights from genomic deletions in 100 strains.* Proceedings of the National Academy of Sciences, 2004. **101**(14): p. 4865-4870.
184. Collins, A., *The Challenge of Genome Sequence Assembly.* The Open Bioinformatics Journal, 2018. **11**(1).
185. Roberts, D.M., et al., *Two sensor kinases contribute to the hypoxic response of Mycobacterium tuberculosis.* Journal of Biological Chemistry, 2004. **279**(22): p. 23082-23087.
186. Pevzner, P.A., H. Tang, and M.S. Waterman, *An Eulerian path approach to DNA fragment assembly.* Proceedings of the national academy of sciences, 2001. **98**(17): p. 9748-9753.
187. Bradley, P., et al., *Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis.* Nature communications, 2015. **6**: p. 10063.
188. Schleusener, V., et al., *Mycobacterium tuberculosis resistance prediction and lineage classification from genome sequencing: comparison of automated analysis tools.* Scientific Reports, 2017. **7**: p. 46327.

189. Iwai, H., et al., *CASTB (the comprehensive analysis server for the Mycobacterium tuberculosis complex): A publicly accessible web server for epidemiological analyses, drug-resistance prediction and phylogenetic comparison of clinical isolates.* Tuberculosis (Edinb), 2015. **95**(6): p. 843-844.
190. Steiner, A., et al., *KvarQ: targeted and direct variant calling from fastq reads of bacterial genomes.* BMC Genomics, 2014. **15**: p. 881.
191. Feuerriegel, S., et al., *Sequence analysis for detection of first-line drug resistance in Mycobacterium tuberculosis strains from a high-incidence setting.* BMC microbiology, 2012. **12**(1): p. 90.
192. Miotto, P., et al., *A standardised method for interpreting the association between mutations and phenotypic drug resistance in Mycobacterium tuberculosis.* European Respiratory Journal, 2017. **50**(6): p. 1701354.
193. Nebenzahl-Guimaraes, H., et al., *Systematic review of allelic exchange experiments aimed at identifying mutations that confer drug resistance in Mycobacterium tuberculosis.* Journal of antimicrobial chemotherapy, 2013. **69**(2): p. 331-342.
194. Balding, D.J., *A tutorial on statistical methods for population association studies.* Nature reviews genetics, 2006. **7**(10): p. 781.
195. Falush, D. and R. Bowden, *Genome-wide association mapping in bacteria?* Trends in microbiology, 2006. **14**(8): p. 353-355.
196. Smith, N.H., et al., *Bottlenecks and broomsticks: the molecular evolution of Mycobacterium bovis.* Nature Reviews Microbiology, 2006. **4**(9): p. 670.
197. Chernyaeva, E.N., et al., *Genome-wide Mycobacterium tuberculosis variation (GMTV) database: a new tool for integrating sequence variations and epidemiology.* BMC Genomics, 2014. **15**(1): p. 308.
198. Wattam, A.R., et al., *PATRIC, the bacterial bioinformatics database and analysis resource.* Nucleic acids research, 2013. **42**(D1): p. D581-D591.
199. Homolka, S., et al., *High genetic diversity among Mycobacterium tuberculosis complex strains from Sierra Leone.* BMC microbiology, 2008. **8**(1): p. 103.
200. Darling, A.C.E., et al., *Mauve: multiple alignment of conserved genomic sequence with rearrangements.* Genome research, 2004. **14**(7): p. 1394-1403.
201. Sandgren, A., et al., *Tuberculosis drug resistance mutation database.* PLoS medicine, 2009. **6**(2): p. e1000002.
202. Ritz, N., et al., *Susceptibility of Mycobacterium bovis BCG vaccine strains to antituberculous antibiotics.* Antimicrobial agents and chemotherapy, 2009. **53**(1): p. 316-318.
203. Rousseau, P. and M. Dupuis, *Antituberculous drug susceptibility testing of Mycobacterium bovis BCG strain Montreal.* Canadian journal of microbiology, 1990. **36**(10): p. 735-737.
204. Kay, G.L., et al., *Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe.* Nature communications, 2015. **6**: p. 6717.
205. Ilin, A.I., et al., *Genomic insight into mechanisms of reversion of antibiotic resistance in multidrug resistant Mycobacterium tuberculosis induced by a nanomolecular iodine-containing complex FS-1.* Frontiers in cellular and infection microbiology, 2017. **7**: p. 151.
206. Wong, S.Y., et al., *Mutations in gidB confer low-level streptomycin resistance in Mycobacterium tuberculosis.* Antimicrobial agents and chemotherapy, 2011. **55**(6): p. 2515-2522.
207. Kim, S., *Drug-susceptibility testing in tuberculosis: methods and reliability of results.* European Respiratory Journal, 2005. **25**(3): p. 564-569.

208. Comas, I., et al., *Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved.* Nature genetics, 2010. **42**(6): p. 498.
209. Safi, H., et al., *Allelic exchange and mutant selection demonstrate that common clinical embCAB gene mutations only modestly increase resistance to ethambutol in Mycobacterium tuberculosis.* Antimicrobial agents and chemotherapy, 2010. **54**(1): p. 103-108.
210. Safi, H., et al., *Evolution of high-level ethambutol-resistant tuberculosis through interacting mutations in decaprenylphosphoryl-β-D-arabinose biosynthetic and utilization pathway genes.* Nature genetics, 2013. **45**(10): p. 1190.
211. Böttger, E., *The ins and outs of Mycobacterium tuberculosis drug susceptibility testing.* Clinical microbiology and infection, 2011. **17**(8): p. 1128-1134.
212. Yakrus, M.A., et al., *Molecular and growth-based drug susceptibility testing of Mycobacterium tuberculosis complex for ethambutol resistance in the United States.* Tuberculosis research and treatment, 2016. **2016**.
213. Bisson, G.P., et al., *Upregulation of the phthiocerol dimycocerosate biosynthetic pathway by rifampin-resistant, rpoB mutant Mycobacterium tuberculosis.* Journal of bacteriology, 2012. **194**(23): p. 6441-6452.
214. Carvalho, L.P., et al., *Antibiotic resistance evasion is explained by rare mutation frequency and not by lack of compensatory mechanisms.* BioRxiv, 2018: p. 374215.
215. Gröschel, M.I., et al., *Pathogen-based precision medicine for drug-resistant tuberculosis.* PLoS pathogens, 2018. **14**(10): p. e1007297.
216. Fleiss, J.L., B. Levin, and M.C. Paik, *Statistical methods for rates and proportions*. 2013: John Wiley & Sons.
217. Wiens, K.E., et al., *Global variation in bacterial strains that cause tuberculosis disease: a systematic review and meta-analysis.* BMC medicine, 2018. **16**(1): p. 196.
218. Dash, P., et al., *Three-dimensional models of Mycobacterium tuberculosis proteins Rv1555, Rv1554 and their docking analyses with sildenafil, tadalafil, vardenafil drugs, suggest interference with quinol binding likely to affect protein's function.* BMC structural biology, 2018. **18**(1): p. 5.
219. White, M.J., et al., *PepD participates in the mycobacterial stress response mediated through MprAB and SigE.* Journal of bacteriology, 2010. **192**(6): p. 1498-1510.
220. Ramón-García, S., et al., *Contribution of the Rv2333c efflux pump (the Stp protein) from Mycobacterium tuberculosis to intrinsic antibiotic resistance in Mycobacterium bovis BCG.* Journal of antimicrobial chemotherapy, 2007. **59**(3): p. 544-547.
221. Becker, K. and P. Sander, *Mycobacterium tuberculosis lipoproteins in virulence and immunity–fighting with a double-edged sword.* FEBS letters, 2016. **590**(21): p. 3800-3819.
222. Venkatesh, R., et al., *RecX protein abrogates ATP hydrolysis and strand exchange promoted by RecA: insights into negative regulation of homologous recombination.* Proceedings of the National Academy of Sciences, 2002. **99**(19): p. 12091-12096.
223. Srivastava, S., et al., *Nucleotide polymorphism associated with ethambutol resistance in clinical isolates of Mycobacterium tuberculosis.* Current microbiology, 2006. **53**(5): p. 401-405.
224. Bahrmand, A.R., et al., *High-level rifampin resistance correlates with multiple mutations in the rpoB gene of pulmonary tuberculosis isolates from the Afghanistan border of Iran.* Journal of clinical microbiology, 2009. **47**(9): p. 2744-2750.
225. Morlock, G.P., et al., *ethA, inhA, and katG loci of ethionamide-resistant clinical Mycobacterium tuberculosis isolates.* Antimicrobial agents and chemotherapy, 2003. **47**(12): p. 3799-3805.

226. Zhang, Y. and W. Yew, *Mechanisms of drug resistance in Mycobacterium tuberculosis [State of the art series. Drug-resistant tuberculosis. Edited by CY. Chiang. Number 1 in the series]*. The International Journal of Tuberculosis and Lung Disease, 2009. **13**(11): p. 1320-1330.
227. Lee, K.W., J.M. Lee, and K.S. Jung, *Characterization of pncA mutations of pyrazinamide-resistant Mycobacterium tuberculosis in Korea.* Journal of Korean medical science, 2001. **16**(5): p. 537.
228. Grandjean, L., et al., *The association between Mycobacterium tuberculosis genotype and drug resistance in Peru.* PLoS One, 2015. **10**(5): p. e0126271.
229. Zorzet, A., et al., *Error-prone initiation factor 2 mutations reduce the fitness cost of antibiotic resistance.* Molecular microbiology, 2010. **75**(5): p. 1299-1313.
230. Vilchèze, C., et al., *Enhanced respiration prevents drug tolerance and drug resistance in Mycobacterium tuberculosis.* Proceedings of the National Academy of Sciences, 2017. **114**(17): p. 4495-4500.
231. Ojha, A., et al., *GroEL1: a dedicated chaperone involved in mycolic acid biosynthesis during biofilm formation in mycobacteria.* Cell, 2005. **123**(5): p. 861-873.
232. Goltermann, L., M.V. Sarusie, and T. Bentin, *Chaperonin GroEL/GroES Over-expression promotes aminoglycoside resistance and reduces drug susceptibilities in Escherichia coli following exposure to sublethal aminoglycoside doses.* Frontiers in microbiology, 2016. **6**: p. 1572.
233. Goltermann, L., L. Good, and T. Bentin, *Chaperonins fight aminoglycoside-induced protein misfolding and promote short-term tolerance in Escherichia coli.* Journal of Biological Chemistry, 2013. **288**(15): p. 10483-10489.
234. Hatzios, S.K., et al., *Osmosensory signaling in Mycobacterium tuberculosis mediated by a eukaryotic-like Ser/Thr protein kinase.* Proceedings of the National Academy of Sciences, 2013. **110**(52): p. E5069-E5077.
235. De Smet, K.A., et al., *Alteration of a single amino acid residue reverses fosfomycin resistance of recombinant MurA from Mycobacterium tuberculosis.* Microbiology, 1999. **145**(11): p. 3177-3184.
236. Ford, C.B., et al., *Mycobacterium tuberculosis mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis.* Nature genetics, 2013. **45**(7): p. 784.
237. Gumbo, T., *Biological variability and the emergence of multidrug-resistant tuberculosis.* Nature genetics, 2013. **45**(7): p. 720.
238. Meacci, F., et al., *Drug resistance evolution of a Mycobacterium tuberculosis strain from a noncompliant patient.* Journal of clinical microbiology, 2005. **43**(7): p. 3114-3120.
239. Echols, H., C. Lu, and P. Burgers, *Mutator strains of Escherichia coli, mutD and dnaQ, with defective exonucleolytic editing by DNA polymerase III holoenzyme.* Proceedings of the National Academy of Sciences, 1983. **80**(8): p. 2189-2192.
240. Baños-Mateos, S., et al., *High-fidelity DNA replication in Mycobacterium tuberculosis relies on a trinuclear zinc center.* Nature communications, 2017. **8**(1): p. 855.
241. Rock, J.M., et al., *DNA replication fidelity in Mycobacterium tuberculosis is mediated by an ancestral prokaryotic proofreader.* Nature genetics, 2015. **47**(6): p. 677.
242. Fijalkowska, I.J. and R.M. Schaaper, *Mutants in the Exo I motif of Escherichia coli dnaQ: defective proofreading and inviability due to error catastrophe.* Proc Natl Acad Sci U S A, 1996. **93**(7): p. 2856-61.
243. Degnen, G.E. and E.C. Cox, *Conditional mutator gene in Escherichia coli: isolation, mapping, and effector studies.* J Bacteriol, 1974. **117**(2): p. 477-87.
244. Regmi, S.M., et al., *Polymorphisms in drug-resistant-related genes shared among drug-resistant and pan-susceptible strains of sequence type 10, Beijing family of*

Mycobacterium tuberculosis. International journal of mycobacteriology, 2015. **4**(1): p. 67-72.
245. Moura de Sousa, J., et al., *Potential for adaptation overrides cost of resistance.* Future microbiology, 2015. **10**(9): p. 1415-1431.
246. Hasenbosch, R., et al., *Ethambutol-induced optical neuropathy: risk of overdosing in obese subjects.* The International Journal of Tuberculosis and Lung Disease, 2008. **12**(8): p. 967-971.
247. Van Tongeren, L., et al., *Therapeutic drug monitoring in the treatment of tuberculosis: a retrospective analysis.* The International Journal of Tuberculosis and Lung Disease, 2013. **17**(2): p. 221-224.
248. Pasipanodya, J.G., et al., *Serum drug concentrations predictive of pulmonary tuberculosis outcomes.* The Journal of infectious diseases, 2013. **208**(9): p. 1464-1473.
249. Weiner, M., et al., *Association between acquired rifamycin resistance and the pharmacokinetics of rifabutin and isoniazid among patients with HIV and tuberculosis.* Clinical infectious diseases, 2005. **40**(10): p. 1481-1491.
250. Ahuja, S.D., et al., *Multidrug resistant pulmonary tuberculosis treatment regimens and patient outcomes: an individual patient data meta-analysis of 9,153 patients.* PLoS medicine, 2012. **9**(8): p. e1001300.
251. Guenther, G., et al., *Treatment outcomes in multidrug-resistant tuberculosis.* New England Journal of Medicine, 2016. **375**(11): p. 1103-1105.
252. Cirillo, D.M., et al., *Reaching consensus on drug resistance conferring mutations (Part 1).* International journal of mycobacteriology, 2016. **5**: p. S31-S32.
253. Consortium, C. and G.P. the 100, *Prediction of susceptibility to first-line tuberculosis drugs by DNA sequencing.* New England Journal of Medicine, 2018. **379**(15): p. 1403-1415.
254. Meehan, C.J., et al., *Whole genome sequencing of Mycobacterium tuberculosis: current standards and open issues.* Nature Reviews Microbiology, 2019.
255. Organisation, W.H., *Global Priority List of Antibiotic-Resistant Bacteria to Guide Research, Discovery and Development of New Antiniotics.* . WHO, 2017.
256. Council, N.R., *Toward precision medicine: building a knowledge network for biomedical research and a new taxonomy of disease*. 2011: National Academies Press.