TÉCNICO LISBOA

A genome-wide perspective on local and global genetic determinants of drug resistant tuberculosis

Pedro Gomes Duarte

Thesis to obtain the Master of Science Degree in

Microbiology

Supervisors: Dr. João Ruben Lucas Mota Perdigão and Professor Isabel Maria de Sá Correia Leite de Almeida

Examination Committee

Chairperson: Professor Leonilde de Fátima Morais Moreira
Supervisor: Dr. João Ruben Lucas Mota Perdigão
Members of the Committee: Dr. Diana Isabel Oliveira Machado

October 2019

Preface

The work presented in this thesis was performed at the Laboratory of Molecular Mycobacteriology, Research Institute for Medicines, University of Lisbon (Lisbon, Portugal), during the period September 2018-August 2019 under the supervision of Dr. João Perdigão and Professor Isabel Portugal. The thesis was co-supervised at Instituto Superior Técnico by Professor Isabel Sá Correia.

Declaration

I declare that this document is an original work of my own authorship and that it fulfils all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Acknowledgements

I would like to thank my supervisors, Dr. João Perdigão and Professor Isabel Portugal for their guidance throughout the course of my M.Sc. dissertation.

Abstract

Tuberculosis (TB) is the leading infectious disease caused by a single agent, and multidrug-resistant (MDR) TB poses a serious health hazard, given its low treatment success rate. The advent of whole genome sequencing (WGS) has widened the analytical spectrum of the *Mycobacterium tuberculosis* genome, enabling a broader characterization of genetic determinants of drug-resistant TB. A dedicated WGS pipeline was employed to uncover the relationship between drug resistance-associated genotypes of *M. tuberculosis* clinical isolates from Portugal and resistance levels of first- and second-line drugs. Differing resistance levels were found among same-class drugs harbouring identical drug resistance-associated genotypes. Novel *gid* Ala167Asp, *rrs* 1076insT, *ethA* Met1Leu, *alr* Met343Thr and Phe4Leu mutations were characterized as associated with drug resistance, while the previously reported association between *embB* Asp354Ala and ethambutol resistance was discredited. A global analysis of the diversity of drug resistance associated mutations and their association with the human-adapted *M. tuberculosis* complex lineages has demonstrated a significant association between prevalent *fabG1* C-15T, *rpoB* Ser450Leu, *embB* Met306Val, *rpsL* Lys43Arg and *rrs* A1401G mutations and lineages 2 (L2) and 4 (L4). A genome-wide single nucleotide polymorphism analysis used to determine pairwise differences between strains, revealed higher proportions of clustered strains within L2 and L4, which coupled with their increased representation in transnational genomic clusters elucidate the genetic characteristics behind the proclivity of these two lineages to disseminate globally and quickly develop MDR.

Keywords

Tuberculosis, whole genome sequencing, antituberculosis drugs, drug resistance levels, genetic background

Resumo

Tuberculose (TB) é a principal doença infeciosa provocada por um único agente infecioso, e TB multirresistente (MDR) representa um grave risco para a saúde humana, dada a sua baixa taxa de sucesso terapêutico. A sequenciação completa de genomas (WGS) ampliou o espetro analítico do genoma do *Mycobacterium tuberculosis*, permitindo assim uma caracterização mais abrangente de determinantes genéticos de TB resistente. Um *pipeline* de WGS foi implementado para desvendar a relação entre genótipos de resistência de isolados clínicos de Portugal e os níveis de resistência a fármacos de primeira e segunda linha. Níveis de resistência discordantes foram identificados entre fármacos da mesma classe contendo genótipos de resistência idênticos. Novas mutações: *gid* Ala167Asp, *rrs* 1076insT, *ethA* Met1Leu, *alr* Met343Thr e Phe4Leu foram descritas como estando associadas à resistência a fármacos antibacilares, enquanto a associação entre a mutação *embB* Asp354Ala e resistência ao etambutol foi refutada. Uma análise global da diversidade de mutações associadas à resistência e a sua associação com as linhagens do complexo *M. tuberculosis* adaptadas ao ser humano demonstraram uma associação significativa entres as mutações *fabG1* C-15T, *rpoB* Ser450Leu, *embB* Met306Val, *rpsL* Lys43Arg e *rrs* A1401G, e as linhagens 2 (L2) e 4 (L4). Uma análise genómica de polimorfismos de nucleótido único, usada para determinar diferenças entre estirpes emparelhadas, revelou maiores proporções de estirpes das L2 e L4 agrupadas, que, juntamente com as suas sobre-representações em grupos genómicos, elucida as características genéticas por detrás da tendência destas linhagens de se disseminarem a nível global e de desenvolverem MDR.

Palavras-Chave

Tuberculose, sequenciação completa de genomas, fármacos antibacilares, níveis de resistência a antibióticos, fundo genético

List of abbreviations

ACP: acyl carrier protein reductase, AMK: amikacin, BCG: bacillus Calmette-Guérin, CAP: capreomycin, CAS: Central Asian Strain, CC: critical concentration, CFUs: colony forming units, CFZ: clofazimine, DCs: dendritic cells, DCS: D-cycloserine, DHF: dihydrofolate, DHFR: dihydrofolate reductase, DHFS: dihydrofolate synthase enzyme, DHPS: dihydropteroate synthase, DPA: decaprenylphosphoryl-*β*-D-arabinose, DST: drug susceptibility testing, EAI: East-African Indian, EMB: ethambutol, ENA: European Nucleotide Archive, ETH: ethionamide, FLQ: fluoroquinolone, GCs: genomic clusters, GWAS: genome-wide association study, HIV: human immunodeficiency virus, H_2Pte: dihydropteroate, H_2PtePAS: hydroxy dihydropteroate, H_2PtePAS-Glu: hydroxy dihydrofolate, ICDSs: interrupted coding sequences, IGRAs: interferon-*γ* release assays, INH: isoniazid, L1: lineage 1, L2: lineage 2, L3: lineage 3, L4: lineage 4, L5: lineage 5, L6: lineage 6, L7: lineage 7, LJ: Löwenstein-Jensen, LAM: Latin American-Mediterranean, LZD: linezolid, KAN: kanamycin, MDR: multidrug-resistant, MIC: minimum inhibitory concentration, MIRU: mycobacterial interspersed repetitive units, MTBC: *Mycobacterium tuberculosis* complex, NOD2: nucleotide-binding oligomerization domain protein 2, NPV: negative predictive value, NTM: nontuberculous mycobacteria, ORF: open reading frame, PAS: para-aminosalicylic acid, PCR: polymerase chain reaction, PPV: positive predictive value, PZA: pyrazinamide, QC: quality control, RD: region of difference, RFB: rifabutin, RIF: rifampin, RFLP: restriction fragment length polymorphism, RRDR: rifampin resistance determining region, SIT: Shared International Types, SLID: second-line injectable drug, SNPs: single nucleotide polymorphisms, STB: smooth tubercle bacilli, STR: streptomycin, TAG: triacylglycerol, TB: tuberculosis, THF: tetrahydrofolate, TST: tuberculin skin test, VCF: variant call format, VNTR: variable number tandem repeats, XDR: extensively drug-resistant

Index

List of figures

List of tables

Chapter 1 – Introduction

Epidemiology and pathobiology of tuberculosis

Tuberculosis (TB) or more precisely, pulmonary TB is caused by the bacterial pathogen *Mycobacterium tuberculosis*. More than 90 % of people infected with this pathogen have an asymptomatic latent infection, which accounts for an estimated 2 billion people worldwide (Zumla et al., 2013). The risk of active disease is approximately 5 % within the first 18 months of infection and another 5 % throughout the remaining lifetime (Andrews et al., 2012). The risk of active disease can be aggravated to 15 % in people infected with human immunodeficiency virus (HIV) (Aaron et al., 2004). The risk of death in HIV-infected people with TB is twice than that of HIV-infected people without TB, which is attributed to the progressive nature of the HIV infection rather than that of TB (Havlir & Barnes, 1999). HIV-infected people are, therefore, a risk group with regards to TB disease, for which special care and prevention is required.

Definitive awareness of the infectious nature of TB goes back to 1882, when Robert Koch first publicly described its etiological agent, *M. tuberculosis*, which by then was known to have caused one fourth of all deaths in Europe (Holmberg, 1990). This finding granted Koch his 1905 Nobel Prize and was the catalyst for the subsequent scientific research that lead to the TB control efforts being implemented today. Despite the magnitude of Koch's discovery, therapy only arrived 60 years later, with the introduction of the first effective antituberculosis agent, streptomycin (STR), in 1943. An increase in drug discovery occurred throughout the following years, with the discovery of para-aminosalicylic acid (PAS), in 1945; isonicotinic acid hydrazide or isoniazid (INH), in 1951; pyrazinamide (PZA) and cycloserine (DCS), in 1952; ethionamide (ETH), in 1956; rifampin (RIF), in 1957; and ethambutol (EMB), in 1962 (Keshavjee & Farmer, 2012).

During this period of drug development, the first case of drug resistance in *M. tuberculosis* was documented, in the late 1940s. This prompted the use of combined chemotherapy to prevent the emergence of drug resistance, although patients infected with drug-resistant strains were less likely to achieve bacteriological cure (Mitchison & Nunn, 1986). It was only in the early 1990s that drug-resistant TB received global attention, and with it the need to develop an initiative that would monitor the emergence and spread of resistance to antituberculosis drugs worldwide, with the Global Tuberculosis Program of the World Health Organization, in 1994 (World Health Organization, 2014). The first outbreaks of multidrug-resistant (MDR) TB, which are defined as having resistance to at least INH and RIF, were also detected at this time. Soon after, a higher degree of resistance, to a wider category of antituberculosis agents was emerging, leading to the recognition of extensively drug-resistant (XDR) TB in 2006, which is defined as MDR TB plus resistance to a fluoroquinolone (FLQ) and at least one of the following second-line injectable drugs (SLIDs): amikacin (AMK), kanamycin (KAN) or capreomycin (CAP) (Center for Disease Control and Prevention, 2006).

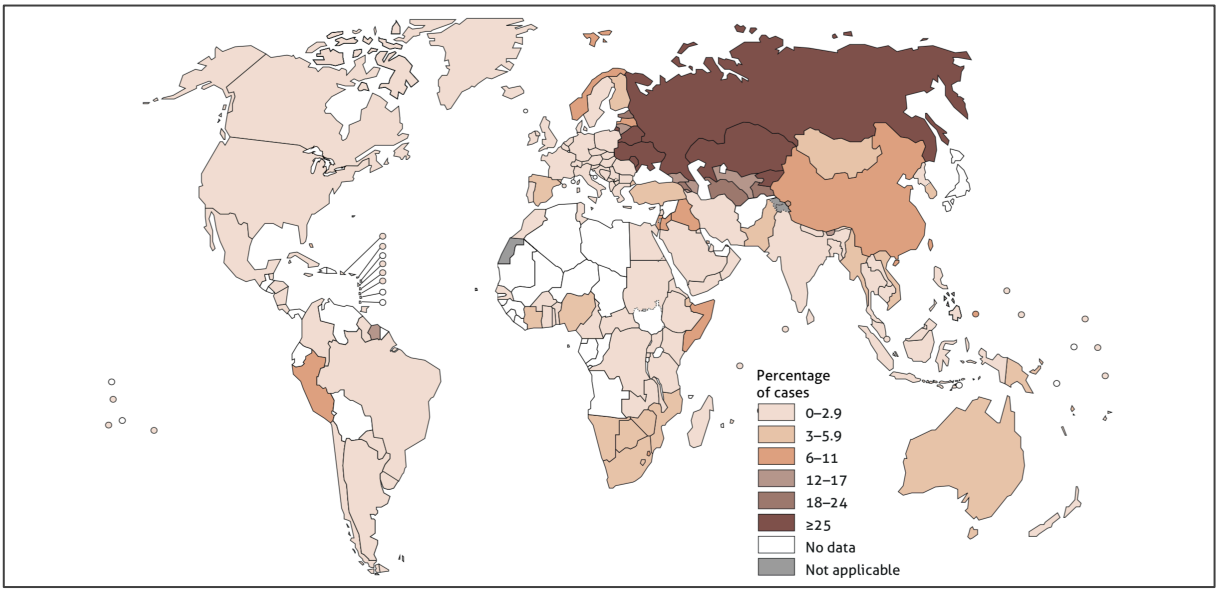

Figure 1.1. Percentage of new MDR/RIF-resistant TB cases. Data between 2004 and 2019. Adapted from "Global tuberculosis report 2019" by World Health Organization, 2019, p. 59. Copyright 2019 by the World Health Organization.

Recent data shows that in most parts of the world, the percentage of patients with newly diagnosed TB who have MDR TB remains below 5.9 % (Figure 1.1). The same cannot be said for countries in eastern Europe and central Asia, considering these have serious MDR TB epidemics. Such high levels of MDR TB in these regions may be attributed to the mismanagement of patient care and the disruption of drug supply that ensued after the collapse of the Soviet Union. Significantly high percentages occur in Belarus (34.1 %), Estonia (19.5 %), Kazakhstan (25.2 %), Kyrgyzstan (26.4 %), Moldova (23.7 %), Russia Federation (19.3 %), Ukraine (24.0 %), and Uzbekistan (23.2 %). Regarding the notification rate of new cases of TB, as well the rate of MDR TB, Portugal has seen a decreasing trend, of -6 % per year and of -9 % per year, respectively, since 2000 (Zignol et al., 2016).

The infectious life cycle of *M. tuberculosis* begins with the inhalation of contaminated droplets from the atmosphere by the human host. Once inside the lungs the bacilli are phagocyted by alveolar macrophages, mononuclear cells specialized in phagocytosis and deployed for nonspecific immune reactions. This in turn induces a proinflammatory response that consist in the recruitment of a wide variety of mononuclear cells from neighbouring blood vessels, which then will form the granuloma or tubercle. Bacilli usually remain contained within the granuloma and this particular stage of the infection does not display any signs of active disease. Although, containment can fail if the immune system is unable to continuously halt the progression of infection, and this often occurs in cases of old age, malnutrition or HIV-coinfection. In these cases, the granuloma releases infectious bacilli which spread through the airways, enabling the dissemination of *M. tuberculosis* by aerosols (Russel, 2001).

The immune response to *M. tuberculosis* can be divided in four distinct stages: innate immune response, immunological equilibrium, TB reactivation, and transmission. The first stage involves the recruitment of phagocytic cells to the site of infection, which include macrophages, neutrophils, monocytes and dendritic cells (DCs). These get infected by the progressive expansion of the bacilli population, forming an early granuloma. Unlike other infectious diseases, for which the recruitment of phagocytic cells is capable of restricting and eliminating the invading pathogen, the mycobacterial infection takes advantage of this, using them as cellular niches through which it can expand its bacterial population (Davis & Ramakrishnan, 2009). There are evidences that show that *M. tuberculosis* is able to modulate the trafficking and maturation of phagosomes, intracellular vesicles which house mycobacteria once it has been phagocyted, allowing for lysosomal evasion (Chackerian et al., 2002). Another unique feature of the *M. tuberculosis* is its ability to spread cell-to-cell using the ESX1 type VII secretion system, responsible for promoting necrotic death of infected cells, allowing for an effectively escape of its current cellular host to infect other cells and expand its population (Davis & Ramakrishnan, 2009). The absence of this virulence mechanism is a characteristic of the *M. bovis* Calmette-Guérin (BCG) strain, which attenuates it and allows its use as a vaccine (Ernst, 2012). The lack of an ESX1 secretion system, which is encoded by the mycobacterial region of difference 1 (RD1) virulence locus causes an attenuated infection and poor granuloma formation, where macrophages display a more rounded shape, move slowly which in turn cover less territory. A $RD1^+$ phenotype is associated with an induced rapid macrophage migration which ironically is more beneficial to *M. tuberculosis* proliferation within the granuloma (Ramakrishnan, 2012). Alongside inducing cell death, *M. tuberculosis* is capable of inhibiting host cell apoptosis, allowing for a prolonged survival and the accumulation of infectious bacilli before being released by an induced cell death. The use of innate immune cells for a nonspecific response to mycobacterial infection is accompanied with an innate mechanism of *M. tuberculosis* identification by multiple pattern-recognition receptors (Ernst, 2012). Of these, Toll-like receptors are able to recognize several mycobacterial agonists, such as lipoproteins, phosphatidylinositol mannans, lipomannan and mycobacterial DNA. The nucleotide-binding oligomerization domain protein 2 (NOD2), as well as other receptors are able to recognize specific mycobacterial peptidoglycan subunits. Stimulation of these recognition receptors induces a proinflammatory response by the host that promote local and systemic mobilization and activation of immune cells to combat infection (Cooper, 2009).

The second stage of the immune response culminates in an immunological equilibrium were *M. tuberculosis* bacilli suffers a growth arrest, through which the infection enters in a latent phase. This process is mediated by an adaptive immune response to *M. tuberculosis*, which is known to have a delayed onset (Ernst, 2012). Studies in mice show that this delayed initiation only comes into effect 11-14 days post-aerosol infection. This is likely due to the required lengthy transport of live bacteria from the lungs to the draining lymph nodes by myeloid DCs where the earliest activation of *M. tuberculosis* antigen-specific naïve $CD4^+$ T cells occurs (Reiley et al.,

2008). Considering that the development of antigen-specific regulatory T cells also occurs during the course of infection, a delayed priming of $CD4^+$ and $CD8^+$ T cells is to be expected. The accumulation of effector of $CD4^+$ and $CD8^+$ T cells in the lungs leads to bacterial population stasis (Shafiani et al., 2010; Mogues et al., 2001). Alongside the delayed adaptive immune response to mycobacterial infection, other causes may be behind the inability of the host to successfully eliminate infectious *M. tuberculosis* bacilli. Antigen-specific effector $CD4^+$ T cells may be unable to identify and stimulate infected macrophages so to increase their bactericidal activity. This is likely due to the downregulation of necessary mycobacterial antigens during infection which in turn decrease the rate of $CD4^+$ T cell activation. Another possible cause of decreased effector $CD4^+$ T cell's ability to stimulate macrophage microbicidal capacity might be due to a *M. tuberculosis*-mediated inhibition of interferon-γ signalling, where mycobacterial cell wall components and lipoproteins play a part (Ramakrishnan, 2012).

The third stage of the immunological life cycle is TB reactivation. A considerable amount of medical conditions and risk factors are linked with TB reactivation. Such are HIV-coinfection, diabetes mellitus, glucocorticoid treatments and malnutrition (Ernst, 2012). Concomitant HIV and TB infection drives TB reactivation from its latent phase, as it is demonstrated that HIV targets and depletes *M. tuberculosis* antigen-specific $CD4^+$ T cells with greater frequency than $CD4^+$ T cells specific to other antigens. This results in a weakened immune response that potentiates *M. tuberculosis* proliferation and transition to an active and symptomatic infection (Geldmacher et al., 2010). Diabetes mellitus is a prevalent disease in developing countries and its increase in prevalence has amplified the severity of TB epidemics. This may be due to an adverse putative delay in the adaptive immune response to *M. tuberculosis* in which trafficking of DCs from the lungs to the lymph nodes is delayed (Vallerskog et al., 2010). Malnutrition is linked with TB reactivation in the sense that malnourished patients have lower levels of leptin, a regulator of energy expenditure and appetite, which also modulates the development and function of T helper cells (Schaible & Kaufmann, 2007; Faggioni et al., 2001).

The final and fourth stage is transmission of infectious mycobacteria, usually by coughing to new susceptible hosts. The capacity to transmit TB varies between infected individuals, as some are more infectious than others. People with cavitary TB – a severe form of TB where lung tissue is destroyed to the point where macroscopic open spaces arise containing numerous mycobacterial bacilli – are especially capable of transmitting increased amounts of airborne bacilli (Ernst, 2012). There is a direct correlation between the degree of host immune response and the likelihood of developing cavitary TB. The probability of developing cavitary TB is fourfold in patients with higher concentrations of circulating $CD4^+$ T cells than those with lower concentrations. In contrast, HIV-coinfected people are less likely to develop cavitary TB than HIV-uninfected people (Kwan & Ernst, 2011). This is likely due to the aforementioned depletion of *M. tuberculosis* antigen-specific $CD4^+$ T cells amongst HIV-coinfected people. In all, a robust understanding of the immunological mechanisms in light of TB disease is known, although further research is needed to fully uncovered the dynamics between the human host and *M. tuberculosis*.

Molecular typing and diagnosis of *M. tuberculosis* infections

A major component in epidemiological studies of TB outbreaks is the use of genotyping methods to characterize and differentiate between *M. tuberculosis* isolates. Genotyping, which has been successfully used since the early 1990s, is useful for primarily differentiating between recently transmitted and reactivation disease. Cases with isolates that share the same genotype are considered clustered and therefore epidemiological linked. The opposite can also occur, where isolates have unique genotypes that are not shared with other isolates, within that same population, and may have resulted from reactivation of latent TB infection that was acquired either outside the population of interest or within that same population albeit outside the time period of the respective epidemiological study. Recurrent TB episodes can also be differentiated between relapse from latent infection, which have the same isolate, and re-infection, which typically display different isolates. Specific molecular markers are used to characterize the genotype of *M. tuberculosis* isolates, which ideally have a rate of change rapid enough to show variation in a local community but slow enough so that case-specific isolates do not dramatically change in a short period of time (Kato-Maeda et al., 2011).

The standard genotyping technique for molecular epidemiological studies has initially relied upon IS*6110* DNA fingerprinting, which is based on the detection of variable copy number of the insertion element *6110* in the genomes of clinical *M. tuberculosis* isolates (Thierry et al., 1990; van Embden et al., 1993; Cave et al., 1994). Differences in restriction fragment length polymorphism (RFLP) patterns of clinical isolates depend on the number of IS*6110* insertion sites present in the *M. tuberculosis* genome. This is determined by the amount of mycobacterial DNA fragments containing IS*6110* after digestion with the *Pvu*II restriction enzyme. Such DNA fragments are separated by gel electrophoresis followed by hybridization with IS*6110* probes through Southern blotting. Differing patterns are obtained not only by the copy number of IS*6110* but also by their genomic location resulting in size polymorphic DNA fragments on which these insertion elements are found (Barnes & Cave, 2003). The main advantages of using IS*6110*-RFLP method is its high discriminatory power and availability of studies for comparison. The main limitations of using this technique is its low discriminatory power provided by clinical isolates with five or fewer IS*6110* bands and the fact that it needs 2-3 µg of high-quality DNA for reliable results and its difficult inter-laboratory comparability (Kato-Maeda et al., 2011).

To circumvent these limitations two polymerase chain reaction (PCR)-based techniques were favoured globally. One of which is spacer oligonucleotide typing (spoligotyping), and it involves the amplification of the direct-repeat locus – a chromosomal region containing 10 to 50 copies of a 36-bp direct-repeat separated by spacer DNA which vary in sequence and size (37 to 41-bp) – by PCR, followed by the hybridization of the labelled PCR products to a membrane which has covalently-bound oligonucleotides corresponding to 43 unique spacers from the direct-repeat locus (Kamerbeek et al., 1997). *M. tuberculosis* strains are differentiated by the intrinsic presence and absence of specific spacers within their genomes. The simplicity of the

resulting genotype as positive or negative for each of the 43 spacers allows for it to be expressed and stored in digital format (Demay et al., 2012). Alongside the small amounts of DNA required for this technique, spoligotyping has significant advantages in terms of its applicability. One major disadvantage is its lower discriminatory power when compared with IS*6110*-based genotyping (Barnes & Cave, 2003).

The second PCR-based genotyping technique is the mycobacterial interspersed repetitive units (MIRU)-variable number tandem repeats (VNTR) typing. MIRU are composed of 40 to 100-bp repetitive sequences of which 41 have already been found scattered within the *M. tuberculosis* H37Rv genome (Mazars et al., 2001). VNTR sequences which were first found in eukaryotes, vary in number and their instability can be attributed to DNA polymerase slippage and recombination events. In *M. tuberculosis* the presence of VNTR is likely due to its lack of a DNA mismatch repair system, which is encoded by the highly conserved *mutS* and *mutL* genes, commonly present in other organisms (Sun et al., 2015). MIRU-VNTR genotyping involves the amplification of specific repetitive loci and the visualization of its respective amplicons by gel electrophoresis. Considering the length of each repetitive unit is known, their variable copies can be discerned through the amplicon's molecular weight, allowing for the differentiation of *M. tuberculosis* isolates. This technique has its own advantages, most notably automatization which increases its inter-laboratory reproducibility, the ability to digitize the results and its similar or higher discriminatory power than IS*6110*-RFLP genotyping, depending on the number and combination of MIRU-VNTR loci being used (Supply et al., 2000). For these reasons, MIRU-VNTR typing is considered the new gold standard for molecular epidemiological studies (Kato-Maeda et al., 2011).

The abovementioned molecular techniques may be useful to study TB outbreaks, but those do not cover the full spectrum of the TB epidemic. The greater majority of people infected with *M. tuberculosis* have a latent TB infection, for which no direct diagnosis is available. Presently, there are only two acceptable indirect diagnostics tests for latent infection: the *in vivo* tuberculin skin test (TST) and the *ex vivo* interferon-*γ* release assays (IGRAs) (Borkowska et al., 2017). The TST has been in use for 100 years and consists in an intradermal injection of a purified protein derivative of tuberculin to the forearm from which an induration – a localized hardening of the skin – forms as a consequence of an immune response. After 2-3 days, the diameter of the induration is measured and the result is given to the patient (Mack et al., 2009; Kruczak et al., 2014). Despite the simplicity of the test, TST has a high probability of generating false positives, due to the fact that tuberculin contains over 200 different antigens nonspecific to *M. tuberculosis*. Two main causes of this issue are the patients vaccinated with BCG, which can create a positive skin reaction even after 15 years succeeding vaccination, and people infected with nontuberculous mycobacteria (NTM) (Borkowska et al., 2017). On the other hand, IGRAs measure the levels of interferon-*γ* produced by $CD4^+$ T cells in the blood, after stimulation with *M. tuberculosis*-specific antigens encoded by the RD1 locus, such as the culture filtrate protein 10 and the early secretory antigen-6 (Lawn, 2015). Because the antigens used in IGRAs do not

cross-react with the BCG vaccine and NTM, this test has a better sensitivity (76-88 %) than that of TST (70 %) The specificity is also higher for IGRAs (86-99 %) than for TST (66 %) (Borkowska et al., 2017).

The advent of drug-resistant TB poses a threat to the efficacy of available antituberculosis therapies. Reliable diagnostics tests are key to ascertain drug susceptibility of TB before administering appropriate treatment regiments. Conventional drug susceptibility testing (DST) remains the standard of care for many reference laboratories worldwide. Classically, DST fall upon three distinct methodologies: the proportion method, the resistance ratio or the absolute concentration method, where the former has been adopted as the reference by the World Health Organization (Canetti et al., 1963; Canetti et al., 1969). In the proportion method, differing amounts of *M. tuberculosis* are inoculated in solid agar or egg-based Löwenstein-Jensen (LJ) medium with and without a critical concentration (CC) of the drug of interest, separately. Mycobacterial colonies are then counted at predetermined growth intervals and if the number of colonies in the drug-containing plate is greater than a critical proportion (usually >1 %) of that of the drug-free plate, the clinical isolate is classified as resistant to the respective drug. Although, newer methods have been adopted for faster reading times, like automated liquid media preparations with radiometric or non-radiometric systems for detecting mycobacterial growth in which DST is also carried out using the proportion method rationale (Heysell & Houpt, 2012). The BACTEC MGIT 960 is one such systems which uses mycobacteria growth indicator tubes containing oxygen-labile sensors that fluoresce under ultraviolet light as the mycobacteria consume oxygen, reducing growth detection in matter of days when compared with solid agar preparations (Pfyffer et al., 1999). Despite its time-reducing advantage, liquid media preparations are more prone to a higher contamination rate than solid agar-based approaches (Muyoyeta et al., 2009).

Drug resistance mechanisms of *M. tuberculosis*

Antimicrobial resistance mechanisms can be divided into two categories: intrinsic resistance and acquired resistance. In *M. tuberculosis*, intrinsic resistance pertains to inherent cellular and physiological characteristics such as cell wall impermeability, dormancy, porin channels, efflux pumps, modification of drug targets, enzymatic degradation or modification of drugs, enzymatic modification of drugs and activation of transcription regulators associated with drug resistance genes (Nasiri et al., 2017). The mycobacterial cell wall is especially hydrophobic which greatly contributes to its innate ability to oppose to drug entry. This is due to its lipid-rich composition in peptidoglycan, arabinogalactan polysaccharides and long-chain mycolic acids. Aside from that, various enzymes involved in cell wall biosynthesis, which aren't solely unique to mycobacteria, present intrinsic structural changes which render certain drugs ineffective, like amoxicillin, carbapenems (β-lactams) and fosfomycin (Alderwick et al., 2015; Gupta et al., 2010). Dormancy – a state in which an organism doesn't self-replicate and has low-to-absent metabolic

activity – is characteristic of *M. tuberculosis* in latent infection which allows it to persist throughout drug treatment despite even being susceptible. This is a clear case of phenotypic drug persistence in the absence of a corresponding genotype. Because *M. tuberculosis* is in a dormancy state, it produces less drug targets proteins and even shuts down DNA and RNA synthesis, making it phenotypically resistant to first- and second-line antituberculosis drugs that target such physiological functions (Smith et al., 2012; Wayne & Hayes, 1996). Porins, which are important in *M. tuberculosis* for the flow of hydrophilic solutes across its highly hydrophobic cell wall are associated with intrinsic drug resistance, demonstrated by knock-out experiments in *M. smegmatis* and in heterologous expression of porin-encoding genes in *M. tuberculosis*, which greatly increase their susceptibility to previously tolerated drugs (Stephan et al., 2004). Active efflux systems play a relevant role in antimicrobial resistance in *M. tuberculosis* and their activity can be attributed to five different structural families: the major facilitator superfamily, the small multidrug resistance family, the resistance-nodulation-cell division superfamily, the adenosine triphosphate-binding cassette superfamily, and the multidrug and toxic compound extrusion family (Nasiri et al., 2017). Several natural occurring genes in *M. tuberculosis* can modify specific drug targets and protect the bacilli against their bactericidal effects. Such genes are the *erythromycin resistance methylase* which is capable of altering the drug-binding site of the 23S rRNA, preventing the binding of macrolides; and the *mfpA* genes which encode pentapeptide repeat proteins that bind and protect DNA gyrase proteins from the lethal action of quinolones (Buriánková et al., 2004; Hegde et al., 2005). Drug-degrading enzymes also play a part, with the most notably example in *M. tuberculosis* being the Amber class-A β-lactamase encoded by the *blaC* gene, that catalyses the hydrolysis of the β-lactam ring of β-lactam and carbapenem drugs (Hugonnet & Blanchard, 2007; Tremblay et al., 2010). Transcription regulator-mediated expression of genes can provide increased resistance to antituberculosis drugs. The WhiB7 is a transcriptional activator of drug resistance-associated genes that is induced by exposure to sub-inhibitory concentrations of antibiotics and various stress conditions. The latter controls the expression of the *eis* gene which is important for the *M. tuberculosis* survival within the macrophage (Wei et al., 2000). The DosR or dormancy transcriptional regulator, a two-component system that increases *M. tuberculosis* survivability, is induced under hypoxia, multiple physiological stress and also antibiotic exposure (Nasiri et al., 2017).

Acquired antimicrobial resistance in bacteria is usually obtained through the acquisition of drug resistance-associated genes by horizontal gene transfer and spontaneous mutations in chromosomal genes normally associated with drug targets. The former mechanism is rare or completely absent in *M. tuberculosis*, leaving its ability to acquire resistance mechanisms exclusively through *de novo* mutations, which typically produces four distinct resistance mechanisms in *M. tuberculosis*: drug target modification, decreased prodrug activation, drug target overexpression and overexpression of drug modifying enzymes (Perdigão & Portugal, 2019). The abovementioned mechanisms will be discussed in regards to first- and second-line antituberculosis drugs currently being used in TB chemotherapy.

INH is the most efficient drug for the treatment of *M. tuberculosis* and is part of the first-line antituberculosis therapy. It is active exclusively against mycobacteria, especially slow-growing mycobacteria. This drug enters the cell by passive diffusion and kills only dividing mycobacteria under aerobic conditions. It is initially bacteriostatic during the first 24 hours, after which it becomes bactericidal (Barclay et al., 1953). INH is also a prodrug, as it needs to first be converted into a range of activated species in order to exert its bactericidal function. It is converted by the mycobacterial catalase-peroxidase KatG enzyme encoded by the *katG* gene (Johnsson & Schultz, 1994). The mechanism of action of INH involves the inhibition of mycolic acid biosynthesis, more specifically by forming an adduct with NAD^+, interfering with the NADH-dependent enoyl-acyl carrier protein reductase (ACP) of the fatty acid synthase type II system, encoded by the *inhA* gene (Vilchèze & Jacobs, 2007). INH also appears to target the β-ketoacyl ACP synthase, KasA, encoded by the *kasA* gene, and a subunit of alkyl hydroperoxide reductase, AhpC, encoded by the *ahpC* gene (Mdluli et al., 1998; Wilson & Collins, 1996), though contrary evidence by Hazbón and co-workers (2006) suggest *kasA* is associated with INH susceptibility. Resistance to INH mainly occurs when there are non-synonymous mutations in the *katG* gene, leading to a decreased or loss of KatG activity, which in turn disables activation of the prodrug (Heym et al., 1995). Mutations in the open reading frame (ORF) of the *inhA* gene also confers INH resistance by modifying the aminoacidic composition of InhA. A Ser94Ala substitution in the NADH binding pocket of InhA reduces the enzyme's affinity to the NAD^+ cofactor and found to contribute to higher resistance levels to INH and ETH when co-occurring with *inhA* promoter mutations (Machado et al., 2013). Hypermorphic mutations in the promoter region of *inhA* lead to target overexpression of InhA, another prevalent mechanism of INH resistance (Larsen et al., 2002). Mutations in the promoter region of the *ahpC* gene have also been found in INH-resistant *M. tuberculosis* strains, and is thus linked with INH resistance (Wilson & Collins, 1996), although Baker and co-workers (2005) have found no evidence that would support this, suggesting instead that *ahpC* promoter mutations likely play a compensatory role in the absence of catalase-peroxidase activity.

RIF is the second most used first-line drug in TB chemotherapy, and it is part of the rifamycin antibiotic class, alongside rifabutin (RFB) and rifapentine. RIF prevents elongation of RNA, a critical step of RNA synthesis by binding itself to the *β* subunit of the RNA polymerase, encoded by the *rpoB* gene. This drug is still active against quiescent *M. tuberculosis*, considering basal transcription continues during its stationary growth phase. Despite the fact that RNA polymerase is composed of four polypeptides: α, β, β′ and σ, with the first three being encoded by the *rpoA*, *rpoB* and *rpoC*, respectively, over 95 % of RIF-resistant *M. tuberculosis* strains harbour non-synonymous mutations in the *rpoB* gene, which usually cluster around the 81-bp rifampin resistance determining region (RRDR) (Cohen et al., 2014). Within those, the Ser450Leu (~41 %), His445Tyr (~36 %), and Asp435Val (~9 %) mutations are the most prevalent. The Ser450Leu substitution induces structural changes in the RNA polymerase which disorders the *β* subunit fork loop 2 amidst binding of RIF. The His445Tyr substitution also induces structural

changes although these elicit a conformational change to the RIF binding pocket, decreasing the RNA polymerase's affinity to RIF. While the Asp435Val substitution does not contribute to structural changes, it does alter the electrostatic distribution of the RIF binding pocket which makes binding of RIF less favourable (Molodtsov et al., 2017). It has been demonstrated that *M. tuberculosis* strains with laboratory-generated mutations in the RRDR have reduced fitness compared to their drug-susceptible ancestors when grown in the absence of RIF. In contrast, clinical isolates with RIF-resistance associated mutations don't show differences in fitness costs between RIF-susceptible strains, despite having the same mutations as the RIF-resistant laboratory-derived strains (Gagneux et al., 2006). Comparing the genomes of both RIF-resistant and RIF-susceptible clinical strains show that the RIF-resistant strains harbour significant amounts of non-synonymous mutations in the *rpoA* and *rpoC* genes. These are actually compensatory mutations for the fitness cost created by RIF resistance-conferring mutations in the *rpoB* gene. Also, an increased frequency of compensatory mutations occurs in high burden MDR TB settings, like eastern Europe and central Asia, which may explain the success of MDR TB in those regions (Comas et al., 2012).

PZA is one of last remaining antituberculosis drugs whose mechanism of action is yet to be properly characterized. Up until recently the prevalent theory was that the active form of the prodrug, pyrazinoic acid acted as a proton ionophore under extracellular acid conditions which results in a loss of cellular membrane potential (Zhang et al., 2003a; Zhang et al., 2003b). PZA is converted to pyrazinoic acid by the mycobacterial pyrazinamidase encoded by the *pncA* gene, which seems likely considering the vast majority of PZA-resistant clinical isolates harbour inactivating mutations in *pncA*. However, the supposed requirement of low extracellular pH to activate PZA and the theorized ionophore function of pyrazinoic acid no longer constitutes as the closest acceptable definition of the mechanism of action of PZA. It is believed that extracellular acidification is one of the possible growth stress conditions that promotes the activity of PZA, with other stresses possibly involved. A more plausible explanation is that pyrazinoic acid has a single target under all stress conditions but this target is only essential to *M. tuberculosis* in specific conditions (Anthony et al., 2018). The mycobacterial aspartate decarboxylase, encoded by the *panD* gene, has been found to be a target of pyrazinoic acid, corroborated by the presence of mutations in *panD* in PZA-resistant clinical isolates (Gopal et al., 2017). Besides *panD*, *rpsA*, which is involved in trans-translation – a rescue mechanism for stalled ribosomes –, has been found to also harbor mutations in PZA-resistant clinical isolates (Shi et al., 2011). Although this may elude to the possibility of *rpsA* being a second target of pyrazinoic acid, conflicting results have been found that disprove the effect of pyrazinoic acid on the expression of *rpsA*, suggesting that *rpsA* and trans-translation are not part of pyrazinoic acid's mode of action, but rather that mutations in *rpsA* are involved in reducing conversion to a PZA-susceptible phenotype (Anthony et al., 2018).

EMB is responsible for inhibiting cell wall biosynthesis by disrupting the arabinosylation of the cell wall arabinogalactans and lipoarabinomannans. EMB inhibits arabinosyltransferases

encoded by the *embCAB* but has a higher affinity towards EmbB, which is specifically responsible for catalysing arabinosylation of arabinogalactans (Takayama & Kilburn, 1989). The most prevalent mutations in the *embB* gene occur in the 306 codon which encodes a methionine, conferring varying degrees of EMB resistance depending on which specific mutation is present. EMB resistance by *embB* Met306 mutations is not straightforward, as these can be found in EMB-susceptible isolates or fail to confer a level of phenotypical tolerance to EMB that supersedes the CC. Nevertheless, high-level resistance to EMB is dependent on *embB* Met306 mutations, suggesting that multigenic mutational events are implicated. Such genes include the less prevalent mutation-liable *embA* and *embC*, but also *ubiA*, which is involved in the decaprenylphosphoryl-*β*-D-arabinose (DPA) pathway and thought to increase intracellular levels of DPA that may compete with EMB for the EMB binding site of Emb enzymes (Perdigão & Portugal, 2019). The association between the latter gene and EMB resistance has also been established by a genome-wide association study (GWAS) (Coll et al., 2018).

STR is a streptidine aminoglycoside that binds to the 30S ribosomal subunit, causing inhibition of translation initiation and misreading of mRNA. In STR resistance develops through target modification mutations in the *rrs*, *rpsL* and *gid* genes. The *rrs* is an essential gene that encodes the 16S rRNA, which has key structures that interact with aminoglycosides, such as the 530 stem loop and the 915 turn. Mutations in *rrs* are found in ~15 % of STR-resistant strains, where the A514C and A908C are commonly associated with STR resistance. *rspL* encodes the S12 ribosomal protein, a structural component of the ribosome that serves to stabilize the pseudo-knot that is formed by the 16S rRNA component of the 30S ribosome. Despite being an essential gene, non-synonymous mutations are tolerated as they don't interfere with ribosomal function but compromise aminoglycoside binding. Mutations in the *rpsL* gene are the most prevalent (~50 %) among STR-resistant strains, with the Lys43Arg and Lys88Arg being the most common. *gid* encodes a 7-methylguanosine methyltransferase that modify specific residues on the 16S rRNA gene. Decreased ribosomal methylation reduces the affinity of 16S rRNA binding site to STR. Mutations in *gid* are found among ~20 % of STR-resistant strains but only confer low-level resistance to STR (Cohen et al., 2014).

AMK and KAN, much like STR, are part of the aminoglycoside family of antituberculosis drugs, although these are SLIDs primarily used to treat MDR TB. AMK and KAN share the same target, the 16S rRNA. Unlike STR, AMK- and KAN-resistance associated mutations in the *rrs* gene are usually found further downstream on the *rrs* gene, where the most prevalent mutation A1401G is 100 % specific to both AMK and KAN resistance, occurring in ~85 % of AMK- or KAN-resistant strains (Cohen et al., 2014). The *eis* gene encodes an aminoglycoside acetyltransferase which acetylates KAN and AMK, decreasing their ability to bind to the 30S ribosomal subunit, although this acetyltransferase activity demonstrates a higher affinity towards KAN (Zaunbrecher et al., 2009). Mutations in the *eis* gene ORF contribute to both AMK and KAN resistance, however hypermorphic mutations in the promotor region of *eis* seem to be exclusively linked with KAN resistance (Cohen et al., 2014).

CAP is a cyclic peptide, part of the SLIDs and has the same mechanism of action as AMK and KAN, despite being structurally different. Like AMK and KAN, the A1401G mutation in the *rrs* gene is the most prevalent among CAP-resistant isolates, with C1402T and G1484T mutations being also associated with CAP resistance although these occur less frequently. Resistance to CAP is often more distinctly associated with mutations in the *tlyA* gene, which encodes a methyltransferase that methylates the 1402 nucleotide position of the *rrs* gene enhancing antimicrobial activity of CAP. Mutations in *tlyA* result in a loss-of-function of TlyA, leading to an unmethylated *rrs* gene that decreases the ribosome's sensitivity to CAP (Engström, 2016). Mutations in the promoter region of the *idsA2* gene have also been statistically proven to be significantly associated with CAP resistance by a GWAS approach (Coll et al., 2018).

FLQs are a major component of second-line antituberculosis therapy, of which ofloxacin (OFX) and moxifloxacin (MFX) are commonly used, with MFX having a higher antimicrobial activity than OFX. FLQs act by inhibiting the topoisomerase II or DNA gyrase, critical for maintaining mycobacterial viability under DNA replication by regulating the degree of DNA supercoiling. DNA gyrase is tetramer formed by two *α* and *β* subunits, encoded by the *gyrA* and *gyrB* genes respectively. Resistance to FLQs is generated through mutations in the quinolone resistance determining regions of *gyrA* and *gyrB*. The most prevalent mutations occur in codons 90 and 94 of the *gyrA* gene (Palomino & Martin, 2014).

ETH is a prodrug and a structural analogue prodrug of INH, meaning they have the same mechanism of action, sharing the same target, InhA. Contrary to INH, ETH is not activated by KatG, but by the monooxygenase EthA. Such diverging activating pathways lead to incomplete cross-resistance between INH and ETH as cross-resistance between these drugs can still be mediated by *inhA* mutations. Interestingly, it appears that mutations in the promoter region of the *fabG1-inhA* operon are frequent among INH-resistance isolates (~35 %) but not so much among ETH-resistant isolates. Mutations in the ORF of *inhA* are substantially more prevalent in ETH-resistant isolates (~68 %), than in INH-resistant isolates, where these appear to be rare. Resistance mechanisms unique to ETH arise from mutations in the *ethA* and *ethR* genes. Unlike the Ser315Thr mutation in the *katG* gene, which is present in 94 % of INH-resistant isolates, there is no particular mutation in the *ethA* gene that is overrepresented among ETH-resistant isolates. *ethR* encodes a transcriptional repressor which downregulates *ethA* by binding to the *ethA* promoter. Missense mutations in *ethR* lead to a loss-of-function of EthR which increases *M. tuberculosis*'s susceptibility to ETH. Resistance to ETH mediated by *ethR* develops through mutations in its promoter region that overexpress EthR, which negatively regulates the expression of *ethA*, thereby abrogating the enzymatic activation of ETH (Vilchèze & Jacobs, 2014).

PAS is a second-line antituberculosis drug that interferes with folate biosynthesis by inserting itself as an alternate precursor. As a structural analogue of the folate precursor para-aminobenzoic acid, PAS is converted into hydroxy dihydropteroate (H_2PtePAS), an analogue of dihydropteroate (H_2Pte), catalysed by the dihydropteroate synthase (DHPS). H_2PtePAS is then

glutaminated by the dihydrofolate synthase enzyme (DHFS or FolC) to form hydroxy dihydrofolate (H_2PtePAS-Glu). H_2PtePAS-Glu, unlike its natural-occurring analogue dihydrofolate (DHF), inhibits the activity of dihydrofolate reductase (DHFR), resulting in mycobacterial growth arrest. Mutations in the *folC* gene cause H_2Pte binding pocket alterations in FolC that reduce H_2PtePAS substrate specificity to FolC, thus decreasing its inhibitory activity over DHFR (Zhao et al., 2014). *thyA* and *thyX* also appear to be linked with PAS resistance. *thyA* encodes a thymidylate synthase that yields DHF, while *thyX*, which is similar to *thyA*, encodes a flavin-dependent thymidylate synthase, that yields tetrahydrofolate (THF), instead of DHF. Expression levels of *thyA* appear to be much higher than those of *thyX* indicating that ThyA is preferred in *M. tuberculosis*. Inactivation of ThyA by an in-frame deletion in the *thyA* gene seems to confer PAS resistance. *thyX* has been demonstrated by mutational experiments to be an essential gene for *M. tuberculosis*, and considering that its function is very similar to that of *thyA*, mutations in *thyX* may be associated with PAS resistance (Fivian-Hughes et al., 2012). *ribD* encodes a bifunction enzyme involved in riboflavin synthesis that also has DHFR activity. Clinical isolates that overexpressed RibD appear to increase PAS resistance, suggesting hypermorphic mutations in *ribD* are associated with PAS resistance (Cohen et al., 2014).

DCS is an analogue of the amino acid D-alanine that inhibits two essential enzymes fundamental for peptidoglycan formation. Alanine racemase, encoded by the *alr* gene converts L-alanine into D-alanine and is a target for DCS (Cohen et al., 2014). DCS inhibits Alr by irreversibly covalently bonding to pyridoxal 5′-phosphate, a cofactor that aids the function of Alr. Mutations in codons located directly in the enzyme's active site show increased levels of DCS resistance which suggest that *alr* missense mutations prevent DCS-mediated inhibition of Alr (Nakatani et al., 2017). Mutations in the *iniA* gene are also associated with DCS resistance (Coll et al., 2018).

Linezolid (LZD) is an oxazolidinone that binds to the 23S portion of the 50S subunit of rRNA, inhibiting the formation of the initiation complex and consequently disabling protein synthesis. The 23S rRNA is encoded by the *rrl* gene, and mutations in this gene are associated with LZD resistance. Mutations in the *rplC* gene, responsible for encoding the L3 protein of the 50S ribosomal subunit have also been found among LZD-resistant clinical isolates (Perdigão et al., 2016; Cohen et al., 2014).

Clofazimine (CFZ) is a riminophenazine drug that is commonly used against leprosy. With the rise of MDR TB, clofazimine has been increasingly used to treat pulmonary TB and is known to shorten MDR TB treatment to 9 months. Resistance to CFZ occurs through mutations in the *rv0678* gene, which encodes a transcriptional repressor of efflux pump MmpS5-MmpL5, causing overexpression of the efflux pump. Mutations in *rv0678* are the main resistance mechanism against CFZ, as these are found among ~96 % of CFZ-resistant isolates (Zhang et al., 2015).

Evolution of the *M. tuberculosis* complex

Human TB is caused by members of the *Mycobacterium tuberculosis* complex (MTBC), a group of closely related species (>99 % nucleotide sequence identity). It is believed that the MTBC, an obligate pathogen, has evolved from an environmental *Mycobacterium* ancestor with genome downsizing and deletion of non-essential genes playing a major part in the evolutionary process as to adapt to an intracellular pathogenic lifestyle. This is apparent considering that the MTBC has a genome size of 4.4 million bp (4.4 Mbp), significantly smaller than those of its closest NTM relatives, *M. marinum* and *M. kansaii* (6.6 Mbp and 6.4 Mbp in size, respectively). Another event that may have contributed to the evolutionary divergence of the MTBC from other mycobacteria is the acquisition of new genes through horizontal gene transfer (Gagneux, 2018). The smooth tubercle bacilli (STB), which include *M. canettii* and *M. prototuberculosis*, share 89.3 % of their genome with those of the remaining MTBC members. It is hypothesized that the ancestor of the MTBC shared similarities with the modern STB. Supporting this notion, the majority of 81 interrupted coding sequences (ICDSs) found in MTBC genomes were also present and interrupted in STB genomes, with four of the ICDSs in the MTBC being present and intact in STB genomes. However, none of the ICDSs shared between STB genomes were observed in other MTBC genomes, which further reinforces the hypothesis that STB were ancestors of the MTBC, considering that the aforementioned molecular scars found in MTBC genomes mostly likely occurred in its most recent common ancestor after it had diverged from STB-like progenitors. Pairwise comparison of genome-wide single nucleotide polymorphisms (SNPs) shows that the MTBC form a single compact cluster within a much larger reticulated network of STB, further supporting this theory (Supply et al., 2013). Besides the absent reciprocity of shared intra-species molecular scars between different mycobacterial species as a marker for evolutionary dating, the ratio of the rates of nonsynonymous and synonymous changes (dN/dS) also highlight diverging evolutionary trends between STB and the MTBC. Comparing a set of protein coding sequences of *M. canettii* with the majority-rule consensus set of the remaining MTBC strains, Hershberg and co-workers (2008) found that the dN/dS was 0.18, implying that *M. canettii* have a reduced rate of nonsynonymous mutations than synonymous mutations. This is a consequence of purifying selection, in which slightly deleterious nonsynonymous mutations are eliminated from the gene pool through natural (negative) selection over extended periods of time. Contrarily, the dN/dS between the remaining MTBC strains (excluding *M. canettii*) was considerably higher, at 0.57. A high dN/dS may indicate that slightly deleterious nonsynonymous mutations that are destined to be eliminated by purifying selection have not yet been removed, a hallmark of closely related strains, or that a reduced selective constraint is at play. In this case, the latter seems more likely as nonsynonymous mutations between the remaining MTBC don't appear to be slightly deleterious, but rather behave as selectively neutral, contributing to a reduced purifying selection. Therefore, the high dN/dS between the remaining MTBC strains indicates that, despite

their closely genomic relatedness, intra-specific diversification occurred, likely due to selective adaptations driven by host immunity (Hershberg et al., 2008).

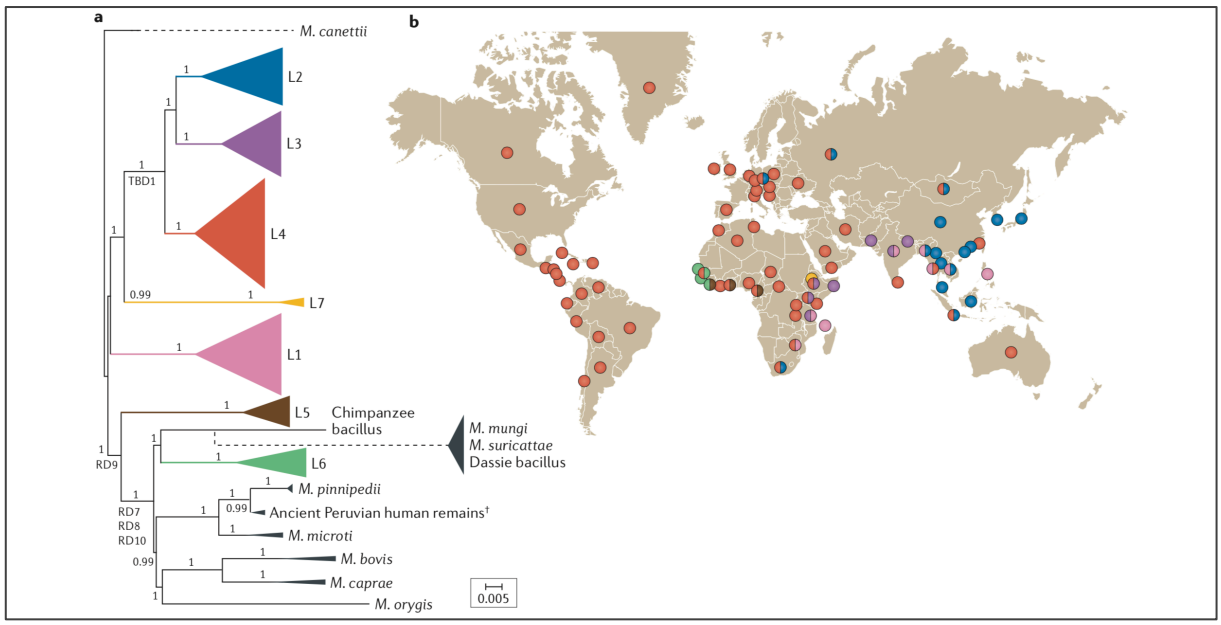

Figure 1.2. Global phylogeography of the human-adapted MTBC. *a*. Genome-based phylogeny of the MTBC rooted with *M. canetti*. MTBC comprises of several human-adapted lineages (in colour) and wild- and domestic-adapted strains (in grey). *M. tuberculosis*-specific deletion 1 and RD deletions are characteristic mutational events of the evolutionary paths of different MTBC lineages. *b*. Global distribution of the human-adapted MTBC lineages. Adapted from "Ecology and evolution of *Mycobacterium tuberculosis*" by S. Gagneux, 2018, *Nature Reviews Microbiology*, *16*, p. 204. Copyright 2018 by the Macmillan Publishers Limited, part of Springer Nature.

After the establishment of the MTBC ancestor, the MTBC evolved into human-adapted strains and strains adapted to wild and domestic mammals. Human TB is mainly caused by *M. tuberculosis sensu stricto* and *M. africanum*. The human-adapted MTBC strains can still be categorized into seven different lineages, lineage 1 (L1), lineage 2 (L2), lineage 3 (L3), lineage 4 (L4), lineage 7 (L7), which correspond to *M. tuberculosis sensu stricto* strains. Lineage 5 (L5) and lineage 6 (L6) correspond to *M. africanum* (Figure 1.2). Human-adapted MTBC lineages differ in their geographic distribution, with some occurring globally and others restricted to specific geographical regions. L2 and L4 are the most widespread, with L2 occurring predominantly in East Asia. L1 and L3 are usually found within regions around the Indian Ocean. L5 and L6 are heavily restricted to West Africa, whereas L7 is exclusively found in Ethiopia (Gagneux, 2018).

The effect of the *M. tuberculosis* genetic background on virulence and drug resistance

Acquired drug resistance, bacterial fitness and virulence in the MTBC seems to be associated with specific genetic backgrounds. The Beijing genotype, which belongs to L2 has been extensively studied since it was first identified in 1995 from patients infected with *M. tuberculosis* in China and Mongolia and initially suggested to have emerged 1000 years ago in Northern-Central China (East Asia), with subsequent introduction in other parts of the world by human migration (Mokrousov, 2008; van Soolingen et al., 1995). The Beijing genotype gained worldwide notoriety for being associated with the MDR TB outbreak in New York, in 1990, caused by a MDR W-Beijing strain in a HIV-infected population (Bifani et al., 1996). It has been demonstrated that W-Beijing strains display increased biological fitness after acquiring drug resistance, which may explain their prevalence in the abovementioned MDR TB outbreak (Toungoussova et al., 2004). The Beijing genotype has since then been responsible for copious TB outbreaks throughout the world, many of which encompassing drug-resistant strains. Despite belonging to well-characterized clade, the Beijing genotype is diverse, displaying a variety of IS*6110*-RFLP patterns, and can be further divided into two major groups, depending on whether an IS*6110* insertion is present or absent in the NTF locus. Strains with an IS*6110* in the NTF region are designated as typical, while strains without an IS*6110* in the NTF region are designated as atypical, and are thought to be closely related to the progenitor of the Beijing clade (Hanekom et al., 2007). Typical Beijing strains are considered to be more virulent than their atypical counterparts, a characteristic responsible for their worldwide dissemination and high frequency in TB outbreaks. The hypervirulence in Beijing strains is attributed to its interactions with host immunity, as it has been found that Beijing strains grow faster and promote higher death rates among *M. tuberculosis* infected-macrophages than non-Beijing strains (Zhang et al., 1999). Beijing strains reduce the level of macrophage apoptosis allowing for an increased intracellular bacilli growth. This hypervirulence has also been demonstrated in *in vivo* studies, where mice infected with Beijing strains died more rapidly than mice infected with non-Beijing strains (Manca et al., 2001). However, increased virulence is not a trait common to all Beijing strains, as some sub-lineages seem to be more virulent than others (Theus et al., 2007). Beijing strains employ genotype-specific immunological characteristics that increase their virulence. Beijing strains ability to synthetize phenolic glycolipids catalysed by the polyketide synthase, reduces type 1 T helper cell cytokine production, limiting the host immune response (Reed et al., 2004). Beijing strains also have a reduced antigen expression when compared to non-Beijing strains which allows them minimize recognition by the host immune response and increase their prevalence (RIndi et al., 2007). Beijing strains have increased levels of sigma factor SigA which binds to the promoter region of the *eis* gene and upregulates it. Increased expression of *eis* induces a higher interleukin-10/tumour necrosis factor *α* secretion ratio and consequently increases host immunosuppression, which is advantageous for the survival of Beijing strains (Wu et al., 2009). Beijing strains also display increased dormancy regulon gene expression than non-Beijing strains,

with an overexpression of the *dosR* gene which enables adaptation to hypoxia and reactive nitrogen species by *M. tuberculosis*, and increased expression of *rv3130c* that encodes for a triacylglycerol (TAG) synthase, increasing production of TAG, a readily-available energy source that increases survivability under anaerobic conditions (Reed et al., 2007). Increased host immune response evasion, decreased host immune response and increased survivability during latency phase are characteristics specific to the Beijing genotype that have allowed it to thrive in humans and prevail in multiple geographical regions (Hanekom et al., 2011). It is hypothesized that the emergence of the Beijing genotype is linked to BCG vaccination. A review of scientific literature indicates that mass BCG vaccination in South-East Asia has been a selective force for the emergence of the Beijing family, and that BCG vaccination constitutes as a risk factor rather a preventive agent against TB in populations with a high incidence of Beijing strain infections (Abebe & Bjune, 2006).

The modern MTBC lineages, L2, L3 and L4 seem to have increased virulence when compared to the ancient lineages, L1 and L6. The modern lineages induce lower levels of inflammatory cytokines in the early phase of infection, delaying pro-inflammatory response (Saelens et al., 2019; Chae & Shin, 2018). Modern MTBC lineages have been demonstrated to undergo fast replication in human macrophages and in aerosol-infected mice (Chae & Shin, 2018). A study where the intracellular growth, inflammatory response and autophagy were tested between strains of the EAI1_SOM (L1), EAI2_Manilla (L1) and *M. africanum* (L5), and strains from the H3, T1 (L4) and CAS1_Kili (L3) clades, found that 7 days post-infection of human monocyte-derived macrophages, strains from the modern lineages L3 and L4 had a higher capacity of surviving and replicating intracellularly, displaying higher colony forming units (CFUs), than strains from the ancient lineages L1 and L5, which had lower CFUs (Romagnoli et al., 2018). Within the modern MTBC lineages, mice infected with HN878 and HN60 strains (L2) die earlier compared with mice infected the CDC1551 strain (L4), which can be attributed to a lower production of pro-inflammatory cytokines (Saelens et al., 2019).

A WGS study which measured the cluster size, mutation rates, early transmission rates and transmissibility between L1-4 *M. tuberculosis* infections, discovered that L2 and L3 strains seem more likely to be clustered and be found in large clusters compared to L4, while L1 strains are less likely to be clustered and found in large clusters. Moreover, L2 seems to harbour less SNPs per transmission (mutation rate), with L1 having the highest mutation rate. L2 *M. tuberculosis* strains also appear to occur in recent infections more often than strains belonging to L1. Lastly, the authors found L2 and L3 strains to transmit more likely than L4 strains (Guerra-Assunção et al., 2015).

Despite its association with hypervirulence, several studies also indicate an association between the Beijing genotype and MDR and XDR profiles (Dymova et al., 2011; Ghebremichael et al., 2010; Kozińska & Augustynowicz-Kopeć, 2015; Park et al., 2005; Tan et al., 2017; Mai et al., 2017). This notorious association between the Beijing genotype and drug resistance shows however differing association strengths at the sub-lineage level: Central Asia Outbreak Clade and

European-Russian W148 clades show a higher degree of association with MDR TB compared with other Beijing clades (Merker et al., 2015). *M. tuberculosis* strains belonging to the ancient or atypical sub-lineages SIT26 and SIT3 have also been described to occur as MDR/XDR strains with significantly higher frequency than their modern or typical counterparts (Iwamoto et al., 2008). In Southern China, frequencies in INH and STR resistance were significantly higher in ancient Beijing sub-lineages than in modern sub-lineages, reinforcing the prevailing association between emergence of drug resistance and the ancient Beijing genotype (Tan et al., 2017).

Several studies have examined the occurrence and distribution of drug resistance across the different MTBC lineages. Particularly, within regions of L1 presence, the Central Asian Strain (CAS) (L3) lineage shows significantly higher proportions of MDR and XDR strains than the East-African Indian (EAI) (L1) sublineage (Couvin et al., 2019). Also, across reference hospitals in Vietnam, Beijing strains are more likely to be drug-resistant than EAI (L1) strains and in South India, MDR is more associated with CAS (L3), Beijing (L2) and T (L4) genotypes than with the EAI (L1) genotype (Nguyen et al., 2016; Shanmugam et al., 2011). In Beijing, China, however, there is no significant association between drug resistance and the Beijing genotype (Liu et al., 2017). Together, these results reinforce the hypothesis that modern MTBC lineages are more likely to acquire drug resistance than ancient lineages.

Beijing strains' proclivity to develop MDR is linked with its robust association with acquiring INH and RIF resistance, considering *katG* Ser315Thr and *rpoB* Ser450Leu mutations are, albeit not exclusively, highly associated with the Beijing genotype, particularly with the ancient Beijing group (Dymova et al., 2011; Gagneux et al., 2006; Ghebremichael et al., 2010; Hillemann et al., 2005; Park et al., 2005; Li et al., 2017). Although, there is a significantly higher proportion of *katG* mutations among ancient Beijing strains, these lack in mutational diversity, while the modern Beijing strains have a higher diversity of *katG* mutations and a significantly higher proportion of *inhA* promoter mutations (Li et al., 2017). A high prevalence of *katG* mutations among ancient Beijing lineages contributes to a stronger association between INH resistance when compared to modern strains, and even EAI (L1) strains (Maeda et al., 2014). A significantly high proportion of INH resistant strains are associated with the Beijing genotype, when compared to non-Beijing strains (Casali et al., 2014; Tan et al., 2017). However, contrary evidence in China found a negative correlation between INH resistance and the W-Beijing genotype (Wang et al., 2015). The *inhA* promoter C-15T mutation has been found to be significantly associated with L1 *M. tuberculosis* strains (Fenner et al., 2012; Gagneux et al., 2006). Another study also found *inhA* promoter mutations to be significantly more associated with non-Beijing strains than with the Beijing genotype (Park et al., 2005).

RIF resistance has been shown to be significantly associated ancient Beijing strains, specifically (Mokrousov et al., 2006). In Korea, the *rpoB* Ser450Leu mutation has been demonstrated to be significantly more associated with non-Beijing strains than with Beijing strains (Park et al., 2005), and in Russia a significant association of the *rpoB* Ser450Leu mutation was not only found in Beijing strains but also in L4 strains (Casali et al., 2014), which contradicts

the previously stated data in this regard. Mutations in the codon 445 of the *rpoB* gene appear to be more associated with non-Beijing strains than with Beijing strains (Hillemann et al., 2005). In contrast, an absence of association between mutations in the RRDR and the Beijing and non-Beijing genotypes has been reported in Sweden (Huitric et al., 2006). The role of the genetic background in the dynamics of RIF resistance emergence is also illustrated by the higher mutational frequency associated with RIF resistance among Beijing strains in comparison with the EAI (L1) genotype and the higher concentrations of RIF necessary to kill high-density Beijing cultures (de Steenwinkel et al., 2012).

PZA resistance has been demonstrated to be statistically significantly associated with the ancient Beijing genotype (Mokrousov et al., 2006). Contrarily, a study found no significant association between PZA monoresistance or PZA-MDR and the L1, L2 and L4 (Budzik etal., 2014). An enlightening WGS study by Baddam and co-workers (2018) has revealed a L3-specific mutation which may be associated with PZA resistance. This particular mutation was single nucleotide deletion in the *rv2044c* gene which resulted in a frameshift disruption of the stop codon of the *rv2044c* gene, causing a fusion between the *rv2044c* gene and the downstream *pncA* gene. This digenic fusion completely disables the ability of the *pncA* gene to code for an active pyrazinamidase, preventing activation of the prodrug, and therefore lead to PZA resistance.

The drug resistance rate of EMB was found to be significantly higher in non-W-Beijing strains than in the Beijing family (Wang et al., 2015). Contrarily, EMB-resistant isolates had a significantly higher proportion in the Beijing genotype than in non-Beijing genotypes (Tan et al., 2017). Conflicting reports of lineage-specific associations with EMB resistance, hinder the ability to paint clear correlation between EMB resistance and the genetic background of the MTBC. Despite this, it appears that mutations in the codon 306 of the *embB* gene are important for acquisition of MDR in L4 strains, but not in L2 strains (Mortimer et al., 2018).

The Beijing genotype is significantly associated with high-level STR resistance, particularly with the *rpsL* Lys43Arg mutation when compared with non-Beijing strains (Sun et al., 2016; Sun et al., 2010), which is known to confer high-level STR resistance, although this association did hold in comparison between Beijing isolates and L4 isolates (Smittipat et al., 2016). The Beijing genotype was also found to be significantly associated with the Glu92Asp Ala205Ala double mutation in the *gid* gene (Sun et al., 2016). STR resistance seems to be significantly more associated with the ancient Beijing genotype than with the modern Beijing genotype, and even more than with the EAI (L1) genotype (Smittipat et al., 2016; Maeda et al., 2014). Overall, STR resistance seems to be significantly associated with the Beijing genotype, when compared with non-Beijing genotypes (Tan et al., 2017). STR resistance, in conjunction with the Beijing genotype appears to increase transmissibility fitness of Beijing strains when compared to EAI (L1) strains (Buu et al., 2012). In absence of RIF resistance, STR resistance appears to be more significantly associated with L4 strains than with Beijing strains, although both STR and RIF resistance seems to be more associated with Beijing strains (Zhou et al., 2017).

The Beijing genotype, and more broadly the L2 is significantly associated with FLQ resistance (Yuen et al., 2013). FLQ-resistant Beijing strains are more likely to have *gyrA* mutations and high-level resistance than FLQ-resistant non-Beijing strains (Duong et al., 2009). The Beijing genotype has also been found to be significantly associated with LZD resistance in MDR and XDR TB (Zhang et al., 2014). Despite numerous reports on the association between the MTBC genetic background and drug resistance, the molecular mechanisms behind these associations are not fully understood.

A comprehensive study by Ford and co-workers (2013), which measured the rate of acquisition of RIF resistance of drug susceptible *M. tuberculosis* isolates from L2 and L4, found that every isolate from L2 acquired RIF resistance at a significantly higher rate than all the L4 isolates, with a near ten-fold difference between the means of the two lineages. An in-depth analysis of the mechanisms driving lineage-associated differences in the rates of RIF resistance acquisition revealed that L2 strains present a higher basal mutation rate than L4 strains, which increases the probability of acquiring a higher number of RIF resistance-associated mutations. Furthermore, the authors also found a three-fold increase in the rate at which L2 isolates acquired resistance to INH and EMB, in comparison with L4 isolates. Similar mechanisms to those behind the increased rate of acquisition of RIF resistance may also play a role in the aforementioned associations between L2, and more specifically the Beijing family, and their resistance to multiple first-line and second-line antituberculosis drugs.

This thesis will explore, on a local scale, the relationship between a large set of drug resistance-associated mutations, obtained through a dedicated genome-wide analysis of *M. tuberculosis* clinical isolates from Portugal, and their respective levels of resistance, for 12 first- and second-line antituberculosis drugs. The diagnostic capability of the present whole genome sequencing (WGS) pipeline for detecting accurate drug resistance profiles will be evaluated using phenotypic data collected from previous standard DST assays. On a global scale, a similar WGS pipeline will be used to predict the genotypic drug resistance profiles, using a large set of publicly available *M. tuberculosis* genome sequences, to study their worldwide distribution and correlate with the population structure of *M. tuberculosis*. Moreover, the global diversity of drug resistance-associated mutations, as well as their statistical association with the genetic background of the MTBC will be investigated to understand the implications of the genetic background on drug resistance acquisition. Lastly, the plausible association between the genetic background of the MTBC and respective *M. tuberculosis* strains capacity to cluster together will be explored in order to better understand lineage differences in drug resistance acquisition, transmissibility and global dissemination.

Chapter 2 – A genome-wide perspective on the molecular basis of drug resistance and its correlation with resistance levels of *M. tuberculosis* in Portugal

Europe has seen an average annual decline of 4.5 % in the notification rate for TB, during the period between 2013 and 2017. Portugal has had, in the same time period, an above-average annual decline in the notification rate of 6.6 %. Nonetheless, Portugal, along with Eastern European countries like Bulgaria, Estonia, Latvia, Lithuania, Poland and Romania, has the one the highest notification rates in Europe (European Center for Disease Prevention and Control, 2019).

Considering that, in recent years, the MDR and XDR TB rates in Portugal were particularly high, varying between 9.9 and 12.5 %, and 44.3 and 66.1 %, respectively, ascertaining drug resistance profiles through DST is crucial to the implementation of an adequate and effective treatment program that can achieve both bacteriological cure and prevent drug-resistant TB relapse.

Throughout the years, molecular studies have linked drug-resistant *M. tuberculosis* clinical isolates with drug resistance-associated genotypes specific to Portugal, enabling the improvement of current molecular diagnostic tools to help identify drug-resistant TB (Portugal et al., 2004; Perdigão et al., 2008; Perdigão et al., 2010; Perdigão et al., 2013; Perdigão et al., 2014). However, qualitative studies of the genetic determinants of drug resistance are often not sufficient to accurately predict drug resistance, for which quantitative analyses of drug resistance-associated mutations on their contributions to varying degrees of drug resistance levels is needed. Several studies have already been conducted in Portugal, that have clarified differences between *M. tuberculosis* drug resistance-associated genotypes and the degree of resistance level to INH, ETH and EMB (Machado et al., 2013; Perdigão et al., 2009).

Knowing the association between drug resistance levels and drug resistance genotypes of clinical isolates provides additional information that could help devise a better therapeutic regiment, on a patient-to-patient basis. Applying WGS allows to extend the spectrum of identified drug resistance associated mutations, characterize the allelic profile linked with drug resistance and attempt to categorize these according to the resulting drug resistance level.

The work described in the present chapter seeks to characterize the mutational diversity and prevalence of drug resistance associated mutations at a nationwide and genome-wide level in Portugal while attempting to establish links with drug resistance levels. For this purpose, a herein established pipeline enabled the complete allelic characterization associated with drug resistance for which a correlation with respective drug resistance levels for 12 first- and second-line antituberculosis drugs is explored.

Materials and Methods

Clinical isolates

A total of 207 *M. tuberculosis* clinical isolates, sourced nationwide from 15 districts in Portugal, including the Madeira and Azores archipelagos, were selected for the present study. The study sample is composed of 191 isolates retrospectively selected between 2008-2016 as part of the TB National Laboratory Surveillance VigLab program, across the National Institute for Hygiene and Tropical Medicine of the NOVA University of Lisbon and Faculty of Pharmacy of the University of Lisbon. This sample includes 71.1 % of the MDR TB cases notified by national public health authorities in the same period and comprises an excess of 31 XDR TB isolates to those notified in the same period. An additional 16 historical strains isolated between 1995-2007 were also included.

Genomic variants pipeline

All clinical isolates had been previously subjected to WGS using an Illumina NGS sequencing platform producing 100/150-bp paired-end reads. Raw sequence data was initially subjected to quality control (QC) using Trimmomatic with default settings which removes and trims low-quality reads. QC-passed reads were subsequently mapped to the reference genome of the *M. tuberculosis* H37Rv (NCBI reference sequence: NC_000962.3), using the Burrows-Wheeler aligner tool with the BWA-MEM algorithm. Variant calling of genome-wide SNPs and insertions and deletions (indels) was performed using both Samtools and Genome Analysis ToolKit, and only concordant variants between both tools were retained for downstream analysis.

Resistance predicting pipeline

Genotypic profiles of drug resistance were inferred through a revised mutation database comprised of 1,290 validated drug resistance-associated mutations, related to 37 different drug resistance-associated genes (Table A.1), collected from recent WGS and molecular studies (Coll et al., 2018; Coll et al., 2015; Perdigão et al., 2010). Briefly, this in-house pipeline screens variant call format (VCF) files obtained upon mapping and variant calling for variants falling within pre-specified target regions comprising drug resistance associated genes and, when applicable, respective promoter regions. The initial set of variants detected in this way is afterwards compared with the abovementioned mutation database, and isolates harbouring mutations present in this database were classed as genotypically resistant to the respective drugs to which the mutations are associated, while retaining information regarding other mutations present in all drug resistance associated genes for each drug which enables the characterization of the complete allelic configuration associated with drug resistance.

Phenotypic DST

All clinical isolates had been previously subjected to phenotypic DST as part of routine TB laboratory diagnosis DST against first-line drugs was performed using the standardized procedure of fluorometric BACTEC MGIT 960 system (Becton Dickinson, Sparks, MD, USA) for INH, RIF, STR, EMB, PZA. For clinical isolates dating before 2000, DST had been carried out through the radiometric BACTEC 460 (Becton Dickinson, Sparks, MD, USA) for susceptibility to all first-line drugs except PZA which was not determined.

Minimum inhibitory concentration (MIC) assay

A subset of 40 *M. tuberculosis* clinical isolates, available at the Faculty of Pharmacy of the University of Lisbon biobank was subjected to MIC determination for 12 antituberculosis drugs. All clinical isolates were previously grown in LJ medium during 2–3 weeks at 37 °C, after which a bacterial suspension was prepared by homogenization in a glass tube with glass beads, and resuspended in sterile bidistilled water. Mycobacterial cell suspension turbidity was adjusted to 0.5 MacFarland units using a densitometer (Grant Instruments, Cambridge, UK). Mycobacterial cell suspensions were then diluted to 1:100 in Middlebrook 7H9 broth (BD Difco, Berkshire, UK) supplemented with 10 % Middlebrook OADC. 96-well microplates with prepared with serial dilutions of each 12 drugs in a scheme identical to the Sensititre *M. tuberculosis* MYCOTB (Thermo Scientific, Waltham, MD, USA) MIC plates. The following drugs at the specified concentration intervals were included in each plate: INH (0.03–4.00 mg/L), RIF (0.12–16.00 mg/L), RFB (0.12–16.00 mg/L), STR (0.25–32.00 mg/L), AMK (0.12–16.00 mg/L), KAN (0.6–40.0 mg/L), OFX (0.25–32.00 mg/L), MFX (0.06–8.00 mg/L), ETH (0.3–40.0 mg/L), PAS (0.5–64.0 mg/L) and DCS (2–256 mg/L). Each well was inoculated with 0.1 mL of the prepared bacterial suspension and incubated at 37 °C with growth assessments at 10 and 14 days post inoculation by microscopic observation in an inverted microscope. MICs were determined as the lowest drug concentration preventing visible growth.

Statistical analysis

Performance of the resistance prediction pipeline as evaluated through the calculation of sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) using phenotypic data as the reference (gold standard) method.

Results and Discussion

To initially identify, on a genome-wide scale, the mutational diversity associated with drug-resistant *M. tuberculosis* in Portugal, the present study assembled the largest *M. tuberculosis* genomic dataset available. The study comprises of a total of 207 *M. tuberculosis* clinical isolates, 102 (49.3 %) of which were classified as MDR isolates, 49 (32.5 %) as XDR, 37 (17.9 %) as drug-resistant (DR) isolates that do not constitute as MDR or XDR, and 19 isolates susceptible to every antituberculosis drugs. The latter sub-set of 19 isolates was added for comparative purposes and enables the verification of the genetic background of susceptible strains co-circulating with other drug-resistant strains. The genome of all strains had been previously subjected to WGS and genome-wide variants called as described herein. To compile a list of variants associated with drug resistance, a bioinformatic pipeline was simultaneously developed to screen a list of genes and mutations with putative association to drug resistance from VCF files and, predict individual drug resistance while inferring on the prevalence of specific drug resistance-associated mutations, allelic configurations and subsequent epidemiological impact. Moreover, an analysis of the relationship between drug resistance-associated genotypes and their respective MIC was done for a subset of 40 *M. tuberculosis* to further clarify the association between mutations and drug resistance, as well as the degree to which individual mutations contribute to high-level drug resistance.

Isoniazid

The most prevalent mutations in INH-resistant isolates were the C-15T nucleotide substitution in the promoter region of the *fabG1-inhA* operon (63.0 %), the Ser94Ala (38.6 %) and Ile194Thr (19.6 %) amino acid substitutions in the *inhA* gene, and the Ser315Thr (29.9 %) and Arg463Leu (10.9 %) substitutions in the *katG* gene (Figure 2.1). Curiously, mutations in the promoter region and ORF of the *inhA* gene appear to be more prevalent than *katG* mutations in INH-resistant *M. tuberculosis* clinical isolates from Portugal, contrary to the global trend (Figure 3.1). A previous study that investigated the molecular epidemiology of MDR TB in Lisbon, Portugal, found a similar mutation distribution tendency, although having reported a much higher proportion of *fabG1* C-15T promoter mutations (91.4 %) than *katG* Ser315Thr mutations (12.1 %), compared to this study (Perdigão et al., 2008). The aforementioned four most prevalent INH resistance-associated mutations were found exclusively among INH-resistance isolates, which reinforces their association with INH resistance. The same does not apply for the *katG*_Arg463Leu mutation, as it was exclusively found among INH-susceptible isolates in its single allelic configuration, as this is a phylogenetic polymorphism that can be found on isolates belonging to the Principal Genetic Group 1 (Sreevatsan et al., 1997a). The mutational diversity within INH-resistant isolates is scarce, with only 21 distinct mutations having been reported,

among 4 different INH resistance-associated genes: *inhA*, *katG*, *ahpC* and *kasA*, with *katG* having the highest mutational diversity (Figure 2.2).

The INH MIC analysis has revealed 14 unique genotypes with varying resistance levels (Figure 2.2). Isolates with single *fabG1* C-15T promoter mutations show a minor resistance level (0.50 mg/L), but isolates with double *inhA* mutations, in particular the *fabG1* C-15T promoter mutations complemented with *inhA* Ile194Thr or Ser94Ala mutations show a high-level resistance to INH (3.71 mg/L and 3.75 mg/L, respectively). These results are concordant with previous findings on INH-resistant *M. tuberculosis* clinical isolates from Portugal where double *inhA* mutations appear to confer higher levels of INH resistance as opposed to single *fabG1* C-15T promoter mutations (Machado et al., 2013). The *kasA* Gly269Ser mutation, unlike the *kasA* Asn400Ser mutation, has an associated MIC of 0.5 mg/L, above the CC for INH, justifying its association with INH resistance and constituting a valid resistance-associated molecular maker for the detection of INH resistance. The addition of the *fabG1* C-15T promoter mutation along with the *kasA* Gly269Ser mutation does not show multiplicative effects on INH resistance. The *katG* Arg463Leu mutation alone is not significantly associated with INH resistance as it only presents a 1.2 increase in the MIC (0.12 mg/L) compared to the CC for INH.

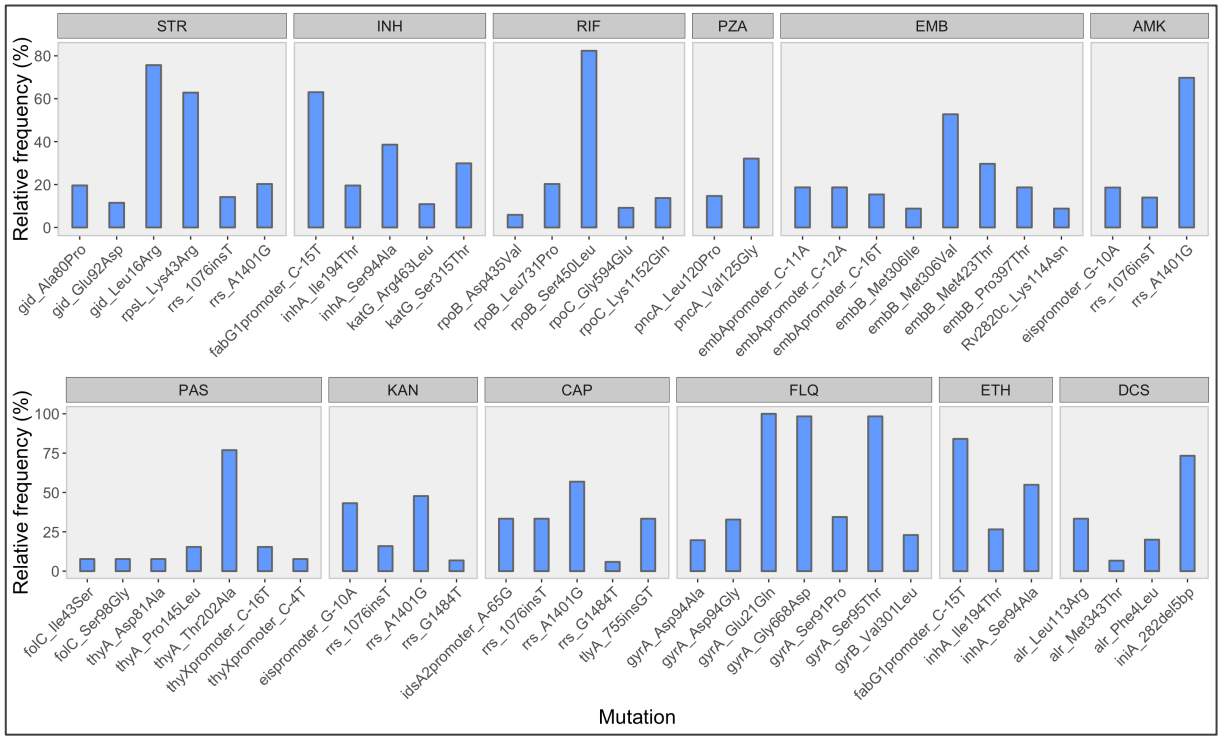

Figure 2.1. Relative frequencies of mutations within respective drug-resistant *M. tuberculosis* clinical isolates from the local dataset. Drug-resistant isolates were diagnosed by DST. Only mutations with a relative frequency of 5 % or higher are shown.

The *katG* Ser315Thr is associated with high-level INH resistance (4.0 mg/L), and its association with the *fabG1* T-8C promoter mutation does not appear to increase the level of INH

resistance. Interestingly, the double *fabG1* C-15T promoter *katG* Thr380Ile mutations have a higher MIC (2 mg/L) than the single *fabG1* C-15T promoter mutation (0.5 mg/L) which suggests that the *katG* Thr380Ile mutation plays a role in conferring multiplicative effects on INH resistance, unlike other *fabG1* promoter/*katG* double mutations. The double *ahpC* G-48A promoter *katG* Gly494Ala mutations show a MIC (0.5 mg/L) 5-fold of the CC for INH. The *ahpC* G-48A promoter mutation has already been included in the revised mutation database as a valid INH resistance-associated mutation, although it is unclear whether the uncharacterized *katG* Gly494Ala mutation could play a role in conferring INH resistance. The triple *ahpC* G-88A promoter *katG* Arg463Leu *katG* Ser315Thr genotype showed high-level INH resistance (4 mg/L). This may be attributed to the *katG* Ser315Thr mutation, as the *katG* Arg463Leu mutation does not seem associated with INH resistance. It is unclear whether the *ahpC* G-88A promoter mutations is associated with INH resistance.

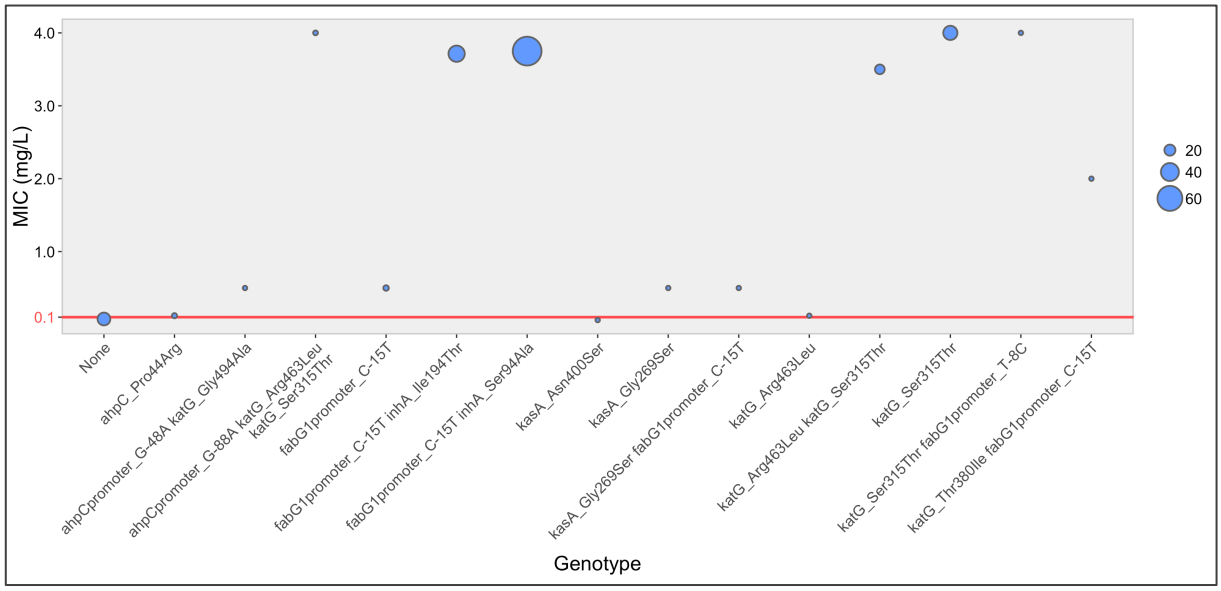

Figure 2.2. Mean MIC of INH relative to INH associated genotypes. Genotypes pertain to a subset of 40 *M. tuberculosis* clinical isolates from Portugal. Circle radius is proportional to the number of isolates that have the associated genotype. The red line indicates the CC of INH for MGIT 960 (World Health Organization, 2018a).

Regarding the diagnostic power of the revised mutation database to detect INH resistance, both the sensitivity and specificity values were significantly high (Table 2.1). Curiously, the NPV for INH was particularly low (0.688), meaning there was a considerable number of false negative genomic profiles (10) within the WGS-predicted negative results (32). A closer look of the genetic make-up of these clinical isolates, did not reveal any mutation within the INH resistance-associated loci present in this analysis, suggesting other genes may be involved in INH resistance acquisition. One false positive genomic profile was reported

harbouring the *kasA* Gly269Ser mutation (Table A.2), which contradicts its already proven association with INH resistance.

Table 2.1. Validity of drug resistance results to twelve first- and second-line antituberculosis drugs obtained by WGS against DST (reference diagnostic test).

Drug	Sensitivity[1]	Specificity[2]	PPV[3]	NPV[4]
INH	0.946	0.957	0.994	0.688
RIF	0.974	0.981	0.993	0.930
PZA	0.826	0.979	0.978	0.829
EMB	0.967	0.724	0.733	0.966
STR	0.905	0.780	0.912	0.767
AMK	0.953	0.692	0.562	0.973
KAN	1.000	0.879	0.863	1.000
CAP	0.961	0.927	0.875	0.978
FLQ	0.951	0.977	0.967	0.966
ETH	0.876	0.971	0.990	0.702
PAS	0.308	0.966	0.500	0.927
DCS	0.333	0.980	0.714	0.906

[1] Sensitivity is the probability of identifying a true positive result. [2] Specificity is the probability of identifying a true negative result. [3] Positive predictive value (PPV) is the probability that a positive result is a true positive. [4] Negative predictive value (NPV) is the probability that a negative result is a true negative.

Rifamycins

The most predominant mutations in RIF-resistant isolates were the Ser450Leu (82.4 %), Leu731Pro (20.3 %) and Asp435Val (5.9 %) mutations in the *rpoB* gene, and the Lys1152Gln (13.7 %) and Gly594Glu (9.2 %) mutations in the *rpoC* gene. Unsurprisingly, the Ser450Leu mutation was by far the most predominant in the *rpoB* gene, and has been reported as a very prevalent mutation (91.4 %) among RIF-resistant *M. tuberculosis*. The Asp435Val mutations as also been previously reported in the aforementioned study, with similar frequencies (6.9 %) (Perdigão et al., 2008). The Leu731Pro mutation is quite predominant despite being outside the 81-bp RRDR. The Leu731Pro mutation has been found to be associated with microevolution towards MDR of *M. tuberculosis* clinical isolates from pertaining to the Q1 clade, from Portugal (Perdigão et al., 2014b). The *rpoC* mutations are likely compensatory mutations that emerged as consequence of the reduced fitness induced by *rpoB* mutations (Comas et al., 2012), particularly with the Ser450Leu substitution. This is evident, as 100.0 and 44.4 % of *M. tuberculosis* isolates from Portugal with the Lys1152Gln and Gly594Glu *rpoC* mutations, respectively, are strictly associated with the *rpoB* Ser450Leu mutation. In all, the present dataset displayed a significant mutational

diversity across the two RIF resistance-associated loci, with 29 unique *rpoB* mutations and 24 unique *rpoC* mutations (Figure 2.14).

Regarding RIF associated genotypes, the great majority showed to be associated with high-level RIF resistance (Figure 2.3). Only the *rpoC* Glu1092Asp and Gly594Glu single mutations are not associated with RIF resistance, which can be attributed to their role has compensatory mutations (Comas et al., 2012).

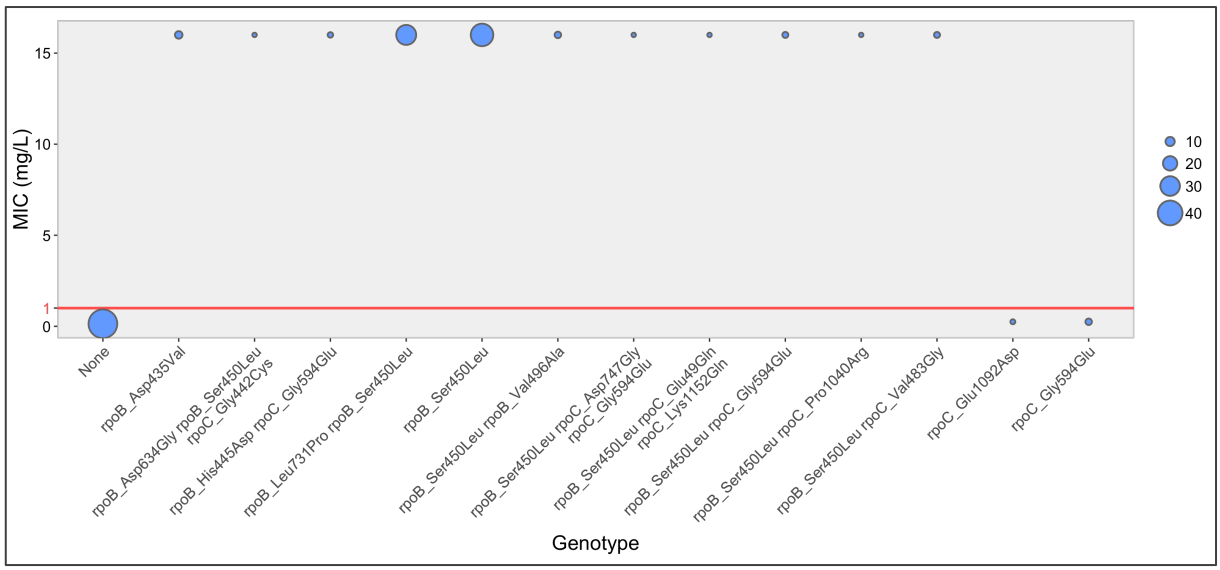

Figure 2.3. Mean MIC of RIF relative to RIF associated genotypes. Genotypes pertain to a subset of 40 *M. tuberculosis* clinical isolates from Portugal. Circle radius is proportional to the number of isolates that have the associated genotype. The red line indicates the CC of RIF for MGIT 960 (World Health Organization, 2018a).

Although RFB shares identical resistance-associated genotypes with RIF, the degree of resistance level varies between the two drugs (Figure 2.4). *rpoB* Ser450Leu mutations shows greater variability in conferring high-level RFB resistance, depending on which additional mutations are present. A noteworthy difference between the two drugs seems to be the *rpoB* Asp435Val mutation, which in the case of RIF is associated with high-level resistance (16.0 mg/L) but in the case of RFB its associated MIC (0.12 mg/L) does not exceed the CC for RFB. Previously studies have shown the disparity of the effects of the *rpoB* Asp435Val mutation between RIF and RFB susceptibility, corroborating the mutation's contribution to RIF resistance but not to RFB resistance (Schön et al., 2013; Cavusoglu et al., 2004).

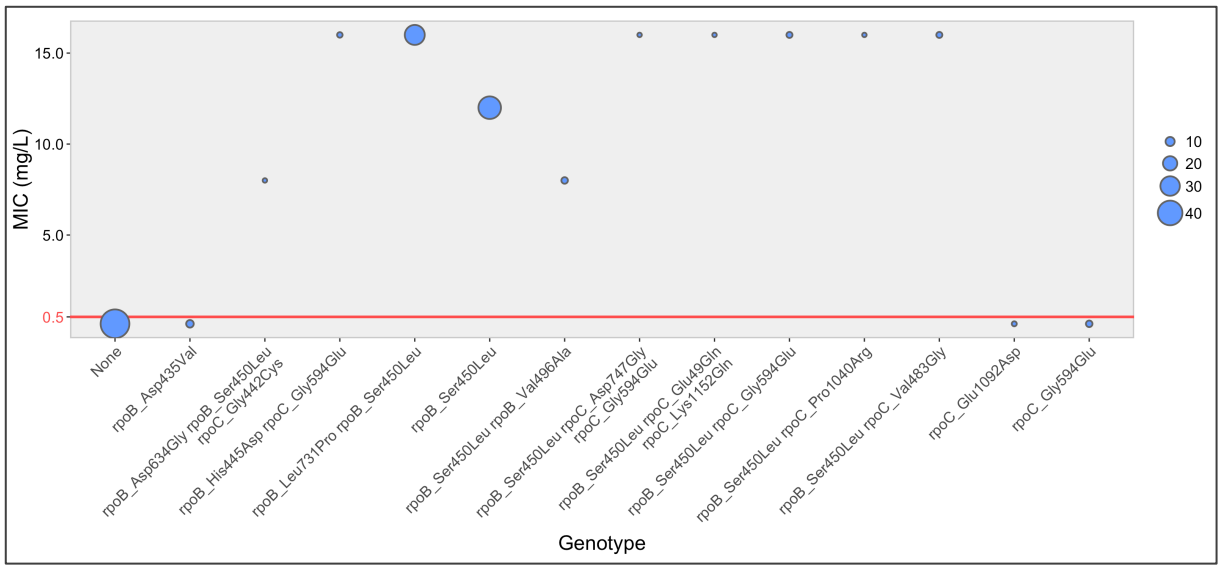

Figure 2.4. Mean MIC of RFB relative to RFB associated genotypes. Genotypes pertain to a subset of 40 *M. tuberculosis* clinical isolates from Portugal. Circle radius is proportional to the number of isolates that have the associated genotype. The red line indicates the CC of RFB for 7H10 agar (Center for Disease Control and Prevention, 2018).

RIF showed one of the highest sensitivity, specificity, PPV and NPV values, while presenting the highest consistency compared to other drugs, which confirms RIF-associated mutations as valid genetic markers for predicting RIF resistance. One exception was the detection of the *rpoB* His445Asn mutation in one false positive genomic profile (Table A.2).

Pyrazinamide

The most prevalent mutations among PZA-resistant isolates were the Val125Gly (32.1 %) and Leu120Pro (14.7 %) substitutions in the *pncA* gene. Two previous molecular studies have also reported these mutations to be the most predominant among *M. tuberculosis* clinical isolates from Portugal (Perdigão et al., 2008; Portugal et al., 2004). PZA-resistant isolates harboured exclusively *pncA* mutations, with a total of 36 unique mutations (Figure 2.14). No mutations were found in the *rpsA* and *panD* genes, even though there are studies that have already demonstrated an association between these two gene and PZA resistance (Kan et al., 2019; Ramirez-Busby et al., 2017; Zhang et al., 2013). Concerning the diagnosis power of the present WGS approach, the sensitivity value for PZA was not as high (0.826) compared to other drugs (Table 2.1). This might indicate a substantial number of false negative genomic profiles (19) within PZA-resistant isolates. A more likely scenario that justifies such a low sensitivity value may be due to incorrect phenotypic data, considering DST for PZA is usually very susceptible to protocol variations and human error, unlike for other antituberculosis drugs. These isolates harboured eight unique uncharacterized *pncA* mutations, which could be associated with PZA resistance. Another hypothesis that might explain the relatively low sensitivity is that there could be other genes

involved in acquired PZA resistance, like the *clpC* and *gpsI* genes, which were not included in the present analysis (Anthony et al., 2018). Consequently, the substantial number of false negative genomic profiles contributed to a low NPV value. Closer analysis of genomic variant data has revealed 80.7 and 4.6 % of PZA-resistant isolates have at least one mutation in the *clpC* and *gpsI* genes, respectively. Despite the absence of validated PZA resistance-associated mutations within these two genes, *clpC* might be involved in PZA resistance acquisition. Future research is encouraged to ascertain the validity of the *clpC* gene as a novel genetic determinant of PZA resistance.

Ethambutol

The most prevalent mutations in EMB-resistant isolates were the Met306Val (52.7 %), Met423Thr (29.7 %), Pro397Thr (18.7 %) and Met306Ile (8.8 %) substitutions in the *embB* gene, the C-11A (18.7 %), C-12A (18.7 %) and C-16T (15.4%) mutations in promoter region of the *embA* gene, and the Lys114Asn substitution (8.8 %) in the *rv2820c* putative gene. The *embB* Met306Val mutation is the most predominant EMB resistance-associated mutation, having been previously reported among 53.3 % of *M. tuberculosis* clinical isolates from Portugal (Perdigão et al., 2009). The *embB* Met423Thr and Pro397Thr mutations have been demonstrated to be involved in microevolution events towards acquisition of EMB resistance within *M. tuberculosis* isolates from the Q1 and Lisboa3 clades, respectively (Perdigão et al., 2014b). The C-11A and C-12A *embA* promoter mutations were found to be exclusively associated with *embB* Pro397Thr mutations, while the C-16T *embA* promoter mutation is exclusively associated with *embB* Met306Val and Met423Thr mutations, suggesting these only emerge after main mechanisms of EMB resistance have been acquired. In all, 31 unique mutations were found among EMB-resistant isolates, with *embB* mutations being the most diverse, accounting for almost half of all unique mutations within the 7 distinct EMB resistance-associated loci: *embB*, *embA*, *embC*, *embR*, *ubiA*, *Rv2820c* and *Rv3300c* (Figure 2.14).

The EMB MIC analysis (Figure 2.5) has revealed that single *embB* Met306Val mutations only confer intermediate-level resistance to EMB (8.0 mg/L), and that high-level resistance to EMB is dependent on multigenic mutations, particularly the combination of *embB* and *embA* promoter mutations. The C-11A, C-12A *embA* promoter mutations, in conjunction with *embB* Pro397Thr mutation, as well as the C-16T *embA* promoter mutation in combination with the Met306Val and Met423Thr *embB* mutations have shown to confer high-level EMB resistance (16.0 mg/L). The single C-16T *embA* promoter mutation has revealed a MIC under the CC for EMB, indicating that individual *embA* promoter mutations are not associated with EMB resistance. Considering single *embB* Met423Thr mutations are not associated with EMB resistance, their abovementioned association with microevolutionary steps towards EMB resistance among Q1 *M. tuberculosis* isolates may be unfounded. The *embB* Met306Ile mutation, which has been previously demonstrated to be associated with EMB resistance, albeit with lower MICs than the

embB Met306Val (Sreevatsan et al., 1997b), does not appear to be associated with EMB resistance (4.0 mg/L). The *embB* Asp354Ala, which is present in the current database, does not seem to be associated with EMB resistance and therefore should not be included for future genomic analysis of EMB resistance. The single *rv2820c* Lys114Asn has showed to be associated with EMB resistance, although this association is unsubstantiated, as there is a *rv2820c* Lys114Asn *embB* Met306Ile double mutation with a MIC under the CC for EMB. Further research is needed to verify the validity of *rv2820c* Lys114Asn as a future EMB resistance-associated mutation.

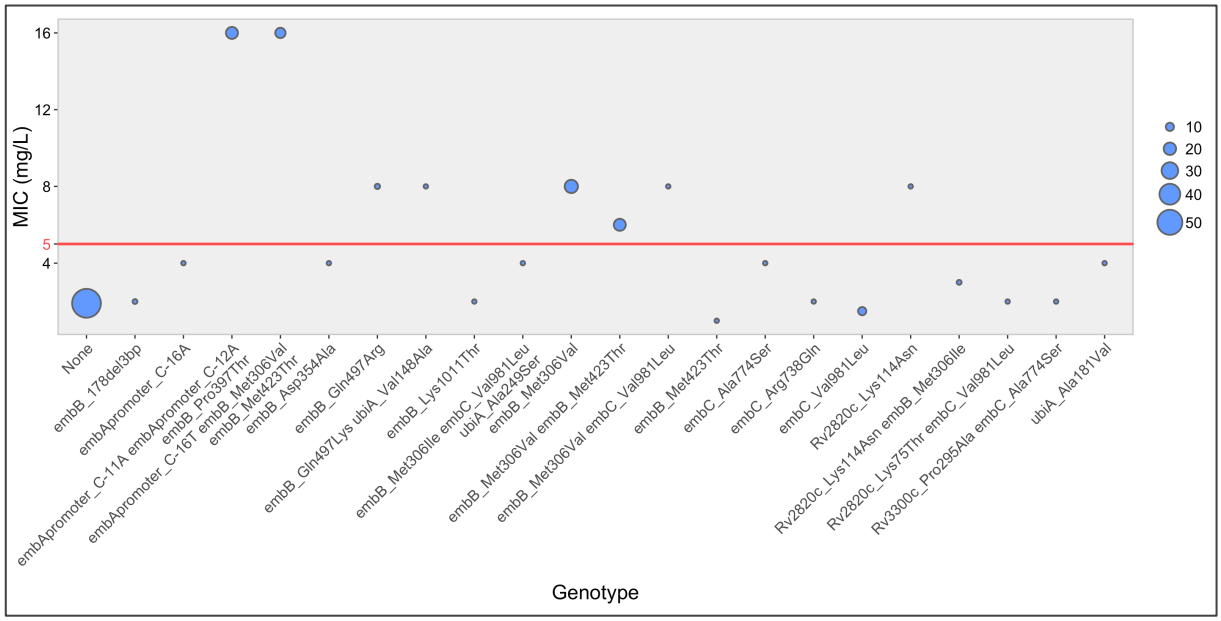

Figure 2.5. Mean MIC of EMB relative to EMB associated genotypes. Genotypes pertain to a subset of 40 *M. tuberculosis* clinical isolates from Portugal. Circle radius is proportional to the number of isolates that have the associated genotype. The red line indicates the CC of EMB for MGIT 960 (World Health Organization, 2018b).

The WGS approach reported a considerably low specificity value for EMB (Table 2.1), which is attributed to the high number of false positive genomic profiles (Table A.2). A total of 13 unique EMB resistance-associated mutations were found among the false positives, which highlights the lack of congruency between EMB resistance-associated mutations and resistance levels to EMB, as it has been demonstrated that mutations in the 306 codon of the *embB* gene, commonly associated with high-level EMB resistance, fail to deliver a phenotypic resistance level that supersedes the CC for EMB (Perdigão & Portugal, 2019). Considering *embB* mutations were by far the most prevalent, being almost ubiquitous among EMB-resistant M. tuberculosis clinical isolates (94.5 %), detection of isolated mutations within the remaining *embA*, *embC*, *embR* and *ubiA* genes may give false positives, as these alone are not associated with EMB resistance, but rather associated with conferring multiplicative effects along with *embB* mutations which lead to high-level resistance to EMB (Perdigão & Portugal, 2019).

Streptomycin

The most prevalent mutations within STR-resistant isolates were Lys43Arg substitution in the *rpsL* gene (62.8 %), the Leu16Arg (75.7 %), Ala80Pro (19.6 %) and Glu92Asp (11.5 %) substitutions in the *gid* gene, and the A1401G (20.3 %) nucleotide change as well as the 1076insT (14.2 %) insertion in the *rrs* gene. The high proportion of *rpsL* Lys43Arg mutations has been previously demonstrated to be highly frequent among *M. tuberculosis* clinical isolates from Portugal (66.7 %) (Perdigão et al., 2008). The *gid* Leu16Arg mutation is not associated with STR resistance, but is rather exclusively associated with the Latin American-Mediterranean (LAM) lineage (L4) (Spies et al., 2011). The *gid* Ala80Pro mutation is a phylogenetic marker for the Q1 clade which is also associated with STR resistance (Perdigão et al., 2014a). A total of 30 unique mutations were found among STR-resistant isolates, with the great majority attributed to the *gid* gene (Figure 2.14).

The majority of STR resistance-associated genotypes either showed a MIC below the CC or are associated with low-level STR resistance (Figure 2.6). *gid* mutations-containing *M. tuberculosis* isolates have been previously found to only moderately increase STR resistance and could be found among both STR susceptible and STR resistance isolates (Wong et al., 2011). Concerning isolates with double *gid* mutations that have the *gid* Leu16Arg mutation and are STR-resistant, their acquired resistance can be solely attributed to the remaining *gid* mutation, since the *gid* Leu16Arg mutation is a natural polymorphism that is not associated with STR resistance (Spies et al., 2011). As previously reported, the *gid* Ala80Pro was found to be associated with STR resistance (Perdigão et al., 2014a), although this study has revealed a much lower level of resistance (1.4 mg/L), only slightly higher than the CC for STR. The double *gid* 104delC Gly130Ala mutations genotype showed an intermediate-level resistance to STR (4 mg/L) although it is unclear whether this is attributed to a synergistic interaction between the two mutations or solely due one individual *gid* mutation. Considering both mutations are not included in the present database, further research should be conducted to validate their individual contributions to STR resistance. The analysis revealed a novel mutation, the *gid* Ala167Asp substitution to be associated with low-level STR resistance (2 mg/L), and should be incorporated into the present mutation database to improve genome-based detection of STR-resistant isolates. The *rpsL* Lys43Arg and Lys88Arg substitutions are associated with high-level STR resistance (32 mg/L). The *rrs* A1401G nucleotide substitution although found among *rpsL* Lys88Arg high-level resistance genotypes, is more likely associated with low-intermediate-level STR resistance and therefore does not constitute a major STR resistance mechanism despite its predominance among STR-resistant isolates (Figure 2.1).

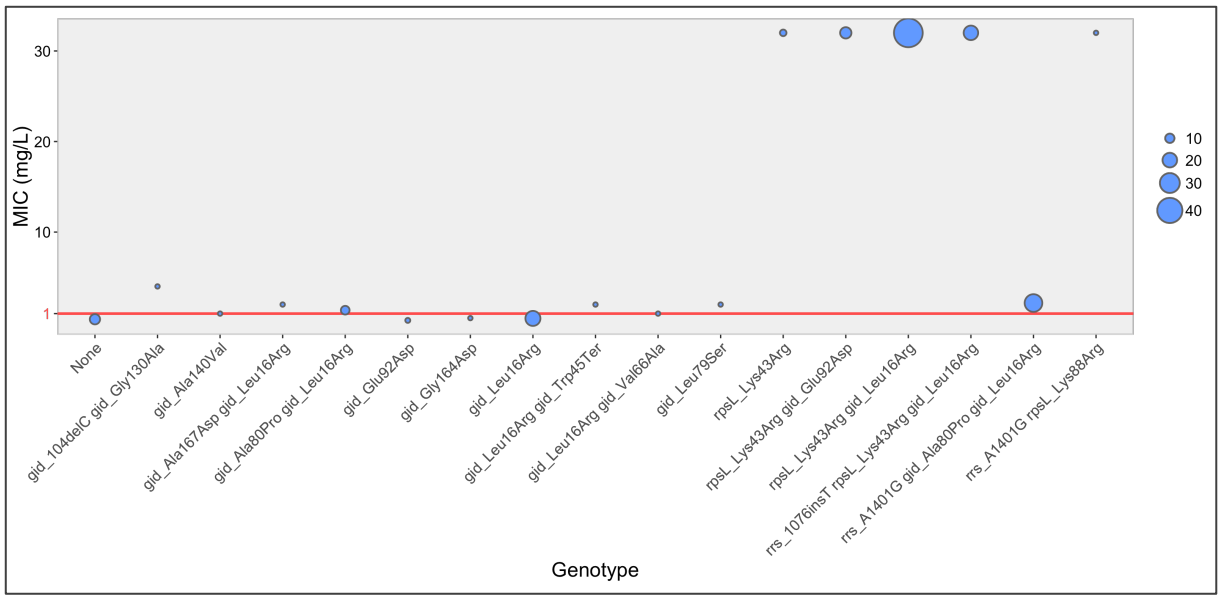

Figure 2.6. Mean MIC of STR relative to STR associated genotypes. Genotypes pertain to a subset of 40 *M. tuberculosis* clinical isolates from Portugal. Circle radius is proportional to the number of isolates that have the associated genotype. The red line indicates the CC of STR for MGIT 960 (World Health Organization, 2018b).

The sensitivity value for STR was only 0.905, which can be attributed to the detection of 14 false negative *M. tuberculosis* genomic profiles (Table A.3). These harboured 11 unique *gid* and *rrs* uncharacterized mutations that could be associated with STR resistance. Worryingly, the specificity value was even lower (Table 2.1), which is due to the presence of 13 false positive genomic profiles, harbouring 4 different STR resistance-associated mutations, including the prevalent *rpsL* Lys43Arg mutation (Table A.2). Given that these mutations were already described as reliable markers of STR resistance, it is unclear why these were found in DST-derived susceptible isolates.

Second-line injectable agents

Among the SLIDs, the most predominant mutations shared by AMK, KAN and CAP were the A1401G (69.8, 47.7 and 56.9 %, respectively) nucleotide substitution and the 1076insT (14.0, 15.9 and 33.3 %, respectively) insertion in the *rrs* gene. The G-10A mutation in the promoter of the *eis* gene is frequently found only among AMK- (18.6 %) and KAN-resistant isolates (43.2 %), albeit with a higher proportion within the latter, which could be due to the Eis higher affinity towards KAN (Cohen et al., 2014). The *rrs* G1484T mutation is only slightly frequent among KAN- (6.8 %) and CAP-resistant isolates (5.9 %), although it is not a resistance mechanism exclusive to these drugs, as there is a residual presence of this mutation among AMK-resistant isolates, which was already previously reported by a molecular characterization of XDR TB in Portugal (Perdigão et al., 2010). The GT dinucleotide insertion in the 755 position of the *tlyA* gene was

considerably frequent among CAP-resistant isolates (33.3 %) (Perdigão et al., 2010). The A-65G nucleotide substitution in the promoter region of the *idsA2* gene was found to be present among 33.3 % of CAP-resistant isolates, being the sole CAP resistance-associated mutation pertaining to the *idsA2* gene, discovered in a recent GWAS (Coll et al., 2018).

Despite the low variety in genotypic results, the AMK and KAN MIC analyses have revealed novel drug resistance-associated mutations (Figures 2.7 and 2.8). The *rrs* A1401G mutations is associated with both high-level AMK and KAN resistance (16.0 and 40.0 mg/L, respectively). The G-10A *eis* promoter mutation is associated low-level AMK resistance, as it only contributed to a 2-fold increase (2 mg/L) from the CC for AMK. In contrast, the same mutation contributes to intermediate-level KAN resistance, with an 8-fold increase (20 mg/L) from the CC for KAN. These results are concordant with the prevailing notion that *eis* promoter mutations contribute to a higher level of resistance to KAN compared to AMK, given that the Eis protein has more affinity towards KAN (Cohen et al., 2014). The *rrs* 1076insT is a novel mutation is associated with both high-level resistance to AMK and KAN, for which it should be added to the present database to improve detection of AMK- and KAN-resistant isolates.

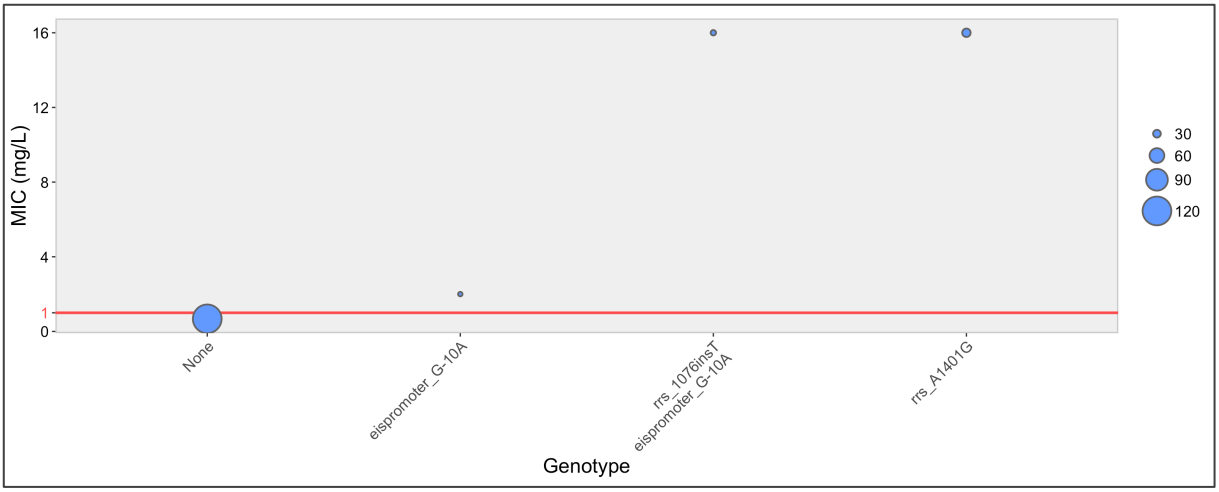

Figure 2.7. Mean MIC of AMK relative to AMK associated genotypes. Genotypes pertain to a subset of 40 *M. tuberculosis* clinical isolates from Portugal. Circle radius is proportional to the number of isolates that have the associated genotype. The red line indicates the CC of AMK for MGIT 960 (World Health Organization, 2018b).

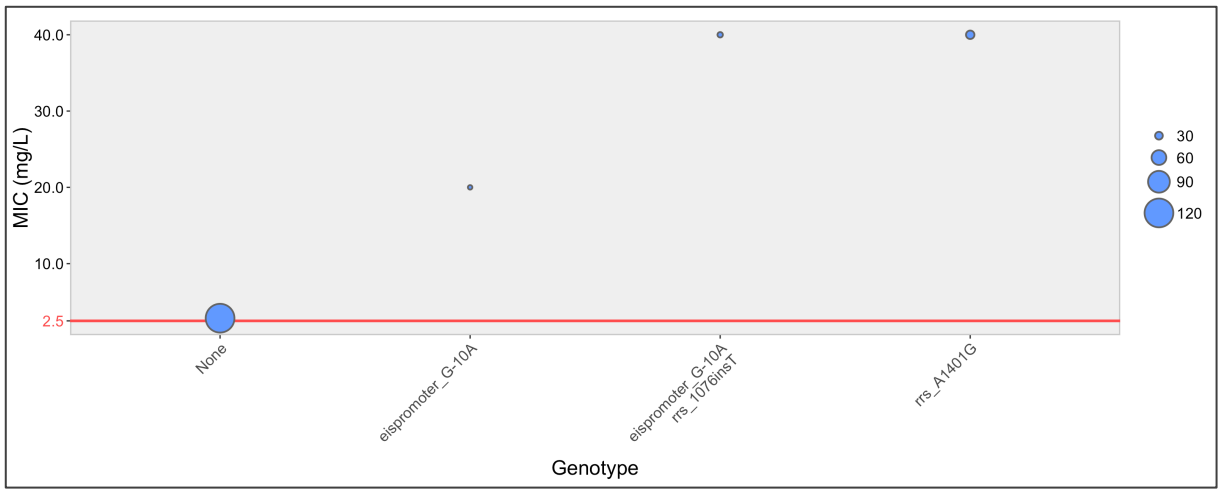

Figure 2.8. Mean MIC of KAN relative to KAN associated genotypes. Genotypes pertain to a subset of 40 *M. tuberculosis* clinical isolates from Portugal. Circle radius is proportional to the number of isolates that have the associated genotype. The red line indicates the CC of KAN for MGIT 960 (World Health Organization, 2018b).

The specificity value for AMK was notably low (Table 2.1) indicating a moderate probability of the current WGS approach to diagnose false positive results. This suggests that there is a considerable amount of supposed AMK resistance-associated mutations in the database that could erroneously classify negative results as positive ones. The inclusion of *eis* promoter mutations may be the cause for such a decreased rate of detection of true negative results, considering *eis* mutations associated with AMK resistance were found among 30 of the total 32 false positives (Table A2). This is concordant with AMK lower affinity to the encoded aminoglycoside acetyltransferase compared to KAN, suggesting hypermorphic mutations in *eis* can in fact yield decreased susceptibility to AMK to a near-borderline resistance level which can lead to inconsistent phenotypic results (Cohen et al., 2014).

Fluoroquinolones

The most prevalent mutations within FLQ-resistant isolates were the Glu21Gln (100.0 %), Gly668Asp (98.4 %), Ser95Thr (98.4 %), Ser91Pro (34.4 %), Asp94Gly (32.8 %), Asp94Ala (19.7 %) substitutions in the *gyrA* gene, and the Val301Leu substitution in the *gyrB* gene (23.0 %). Despite their high frequencies, the *gyrA* Glu21Gln, Gly668Asp, Ser95Thr mutations constitute natural polymorphisms and are not associated with FLQ resistance (Lau et al., 2011).

The OFX and MFX MIC analyses shared the same genotypes (Figures 2.9 and 2.10). As predicted, the *gyrA* Glu21Gln Gly668Asp Ser95Thr genotype is not associated with FLQ resistance, considering these three mutations are natural polymorphisms exclusively associated with the L4 genetic background (Lau et al., 2011). Isolates containing this genotype are the closest approximation to a wild-type susceptible genotype. The *gyrA* Asp94Ala, Asp94Gly,

Ser91Pro mutations and the *gyrB* Asp461His mutation appear to contribute to greater levels of MFX resistance than OFX resistance. The MIC for MFX associated with *gyrA* Asp94Ala mutation is 4-fold the CC for MFX, but the MIC for OFX associated with same mutation displays only a 2-fold increase from the CC for OFX. Similar trends occur to the remaining aforementioned mutations. *gyrA* Asp94Gly mutations appear to confer higher levels of resistance to both OFX and MFX in comparison to the *gyrA* Asp94Ala mutation, an observation which already has been previously demonstrated (Rigouts et al., 2016). A study by Sirgel and co-workers (2012) has showed that mutations in the codon 94 of the *gyrA* gene are linked with higher levels of OFX and MFX resistance than mutations in the codon 91 of the same gene. This study contradicts the previous assumption as the *gyrA* Ser91Pro mutation displayed high MICs for both OFX and MFX when compared with this previous reporting (Sirgel et al., 2012). Furthermore, there is no clear evidence that any *gyrA* Asp94 mutation outcompetes *gyrA* Ser91 mutations in terms of conferring higher levels of OFX and MFX resistance, considering that the *gyrA* the Ser91Pro mutation displays lower levels of FLQ resistance than the *gyrA* Asp94Gly mutation, while conferring higher FLQ resistance levels when compared with the *gyrA* Asp94Ala mutation.

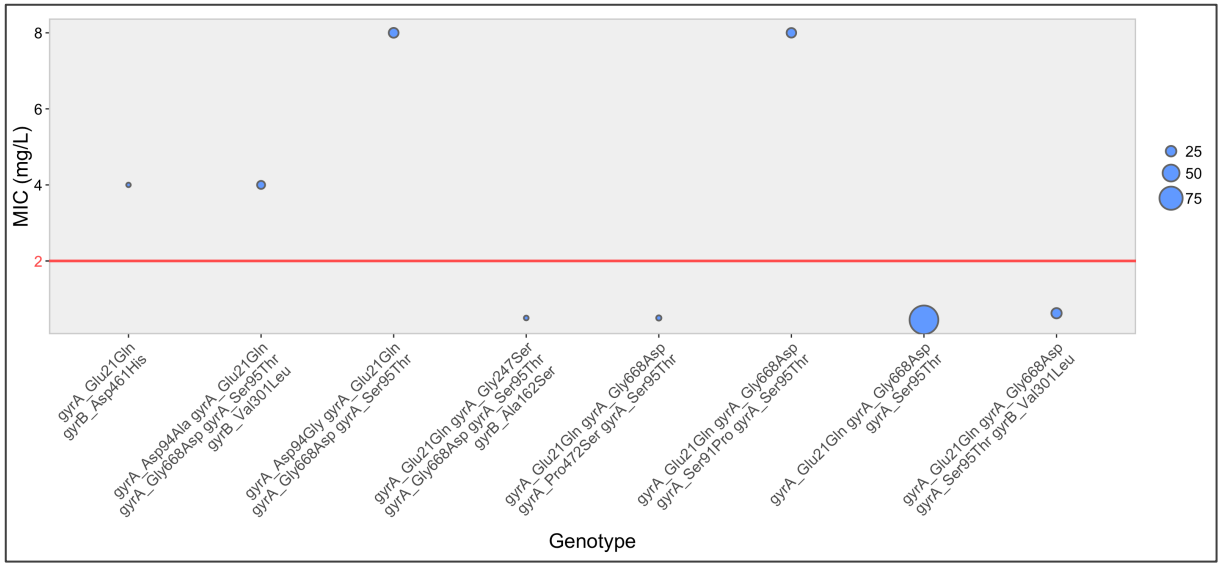

Figure 2.9. Mean MIC of OFX relative to OFX associated genotypes. Genotypes pertain to a subset of 40 *M. tuberculosis* clinical isolates from Portugal. Circle radius is proportional to the number of isolates that have the associated genotype. The red line indicates de CC of OFX for MGIT 960 (World Health Organization, 2018b).

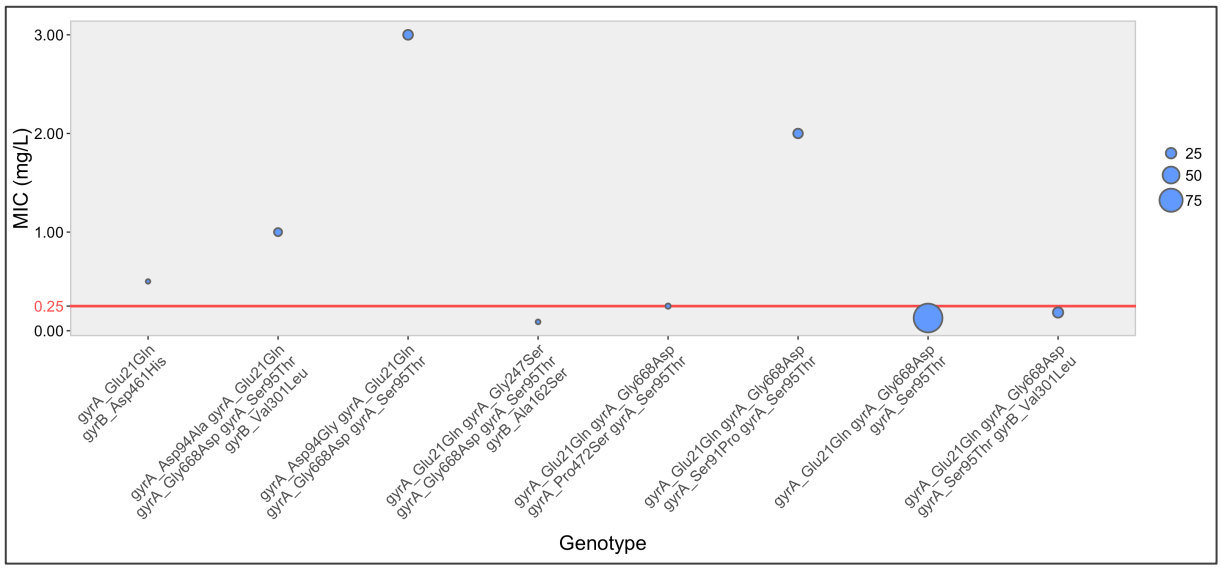

Figure 2.10. Mean MIC of MFX relative to MFX associated genotypes. Genotypes pertain to a subset of 40 *M. tuberculosis* clinical isolates from Portugal. Circle radius is proportional to the number of isolates that have the associated genotype. The red line indicates the CC of MFX for MGIT 960 (World Health Organization, 2018b).

The sensitivity and specificity values for FLQ were considerably high (0.951 and 0.977, respectively) (Table 2.1). Only 2 false positives were detected, harbouring the FLQ resistance-associated mutations: *gyrA* His70Arg and *gyrA* Thr80Ala (Table A.2), which could question their validity as drug resistance genetic markers.

Ethionamide

The most prevalent mutations in ETH-resistant isolates were the *fabG1* C-15T promoter mutation (84.1 %), and the Ser94Ala (54.9 %) and Ile194Thr (26.5 %) substitutions in the *inhA* gene. Unlike INH, resistance to ETH in *M. tuberculosis* clinical isolates from Portugal occurs almost exclusively through drug target mutations, and mechanisms of prodrug inactivation usually play a minor role in acquisition of ETH resistance (Vilchèze & Jacobs, 2014). Despite *inhA* mutations being the most prevalent in the present dataset, the mutational diversity is below that of *ethA*, as 14 unique mutations were found for this gene, compared to only 5 unique *inhA* mutations (Figure 2.14).

ETH resistance in *M. tuberculosis* clinical isolates from Portugal, is mainly attributed to *fabG1-inhA* promoter mutations (Figure 2.11), which are involved in cross-resistance to INH. Like in INH-resistant isolates, *inhA* double mutations provide high-level ETH resistance when compared with single *inhA* mutations (Machado et al., 2013). This is the case with the C-15T *fabG1* promoter mutation, which only has a MIC of 15 mg/L, 3 times higher than the CC for ETH, unlike the C-15T *fabG1* promoter *inhA* Ser194Thr and C-15T *fabG1* promoter *inhA* Ser94Ala double mutations (40 mg/L), which both show an 8-fold increase. The T-8C *fabG1* promoter *ethA*

Met1Leu double mutation also displayed a high MIC to ETH (40 mg/L). Following the reasoning that *fabG1* promoter mutations only confer intermediate-level resistance to ETH, there is plausible evidence that the *ethA* Met1Leu mutation alone is associated with ETH resistance. The latter is not currently included in the present mutation database, and its addition may improve detection of ETH-resistant *M. tuberculosis* isolates.

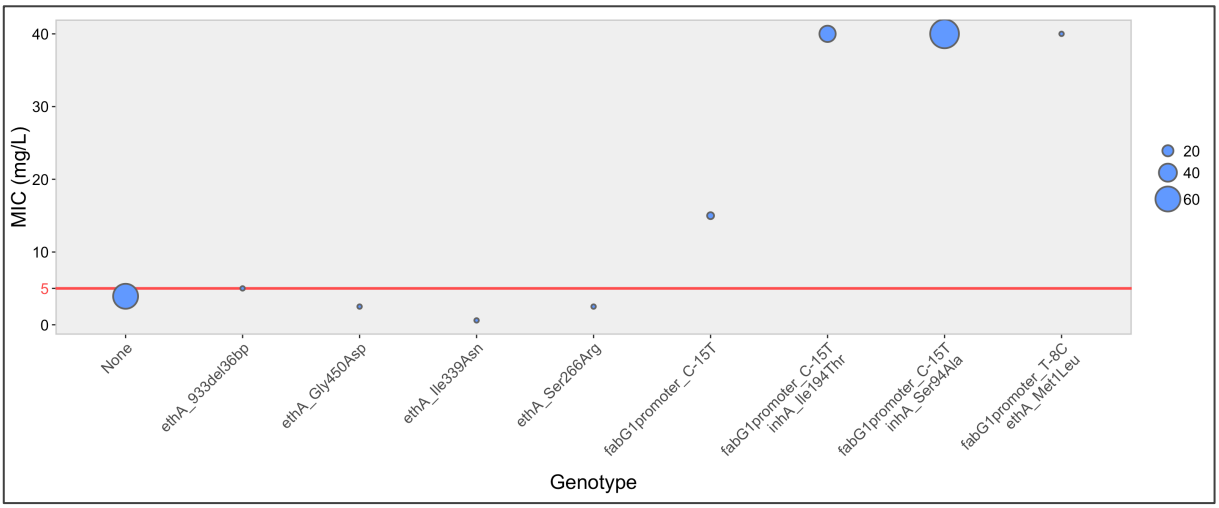

Figure 2.11. Mean MIC of ETH relative to ETH associated genotypes. Genotypes pertain to a subset of 40 *M. tuberculosis* clinical isolates from Portugal. Circle radius is proportional to the number of isolates that have the associated genotype. The red line indicates the CC of ETH for MGIT 960 (World Health Organization, 2018b).

The sensitivity for ETH was relatively low (0.876), which can be attributed the 14 false negative genomic profiles, detected by the WGS approach. Along with these, 11 unique mutations pertaining to the *ethA* and *inhA* genes were detected, which may be associated with ETH resistance (Table A.3). Given that there is a reduced difference between the number of false negative and true negative genomic profiles, the NPV is also low (Table 2.1)

Para-aminosalicylic acid

The most predominant mutations in PAS-resistant isolates were the Thr202Ala (76.9 %), Pro145Leu (15.4 %), Asp81Ala (7.7 %) amino acid substitutions in the *thyA* gene, the C-16T (15.4 %) and C-4T (7.7 %) *thyX* promoter mutations, and the Ile43Ser (7.7 %) and Ser98Gly (7.7 %) substitutions in the *folC* gene. The *thyA* Thr202Ala mutation has been previously found to be a phylogenetic marker for the LAM lineage and is not associated with PAS resistance (Feuerriegel et al., 2010).

The PAS MIC analysis did not reveal as much information as previous analyses, considering it only provides data on one genotype, the *thyA* Thr202Ala mutation (Figure 2.12). This mutation, which has already been referred to be solely associated with L4 of the MTBC

(Feuerriegel et al., 2010), is shown to be primarily linked with PAS-susceptible isolates. Interestingly, 49 isolates, without any PAS resistance-associated mutations displayed an average MIC of 3.2 mg/L, above the CC for PAS. The presence of PAS-resistant isolates devoid of drug resistance-associated genotypes is likely due to the reduced diversity of PAS resistance-associated loci incorporated in the present study. Considering only the *thyA*, *thyX*, *folC* and *ribD* genes were included, mutations in additional PAS resistance-associated loci could be driving PAS resistance acquisition.

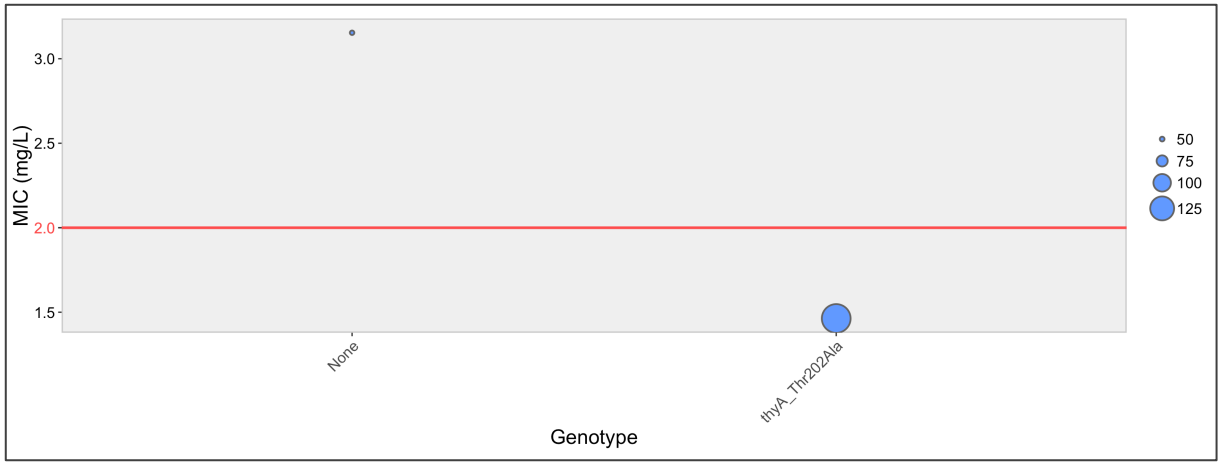

Figure 2.12. Mean MIC of PAS relative to PAS associated genotypes. Genotypes pertain to a subset of 40 *M. tuberculosis* clinical isolates from Portugal. Circle radius is proportional to the number of isolates that have the associated genotype. The red line indicates the CC of PAS for 7H10 agar (Center for Disease Control and Prevention, 2018).

Cycloserine

The most prevalent mutations among DCS-resistant isolates were a 5-bp deletion in the 282 position in the *iniA* gene (73.3 %), and the Leu113Arg (33.3 %), Phe4Leu (20.0 %) and Met343Thr (6.7 %) substitutions in the *alr* gene.

The 5-bp deletion in position 282 of the *iniA* gene, found in 73.3 % of DCS-resistant isolates, is not associated with DCS resistance, displaying a MIC (16.0 mg/L) below the CC for DCS (Figure 2.13). Despite the near-breakpoint level MIC values, the *alr* Leu113Arg, Met343Thr and Phe4Leu mutations seem to be associated with DCS resistance. Excluding the *alr* Leu113Arg mutation, the remaining mutations are novel polymorphisms associated DCS resistance that should be included in the present database to improve detection of DCS-resistant isolates.

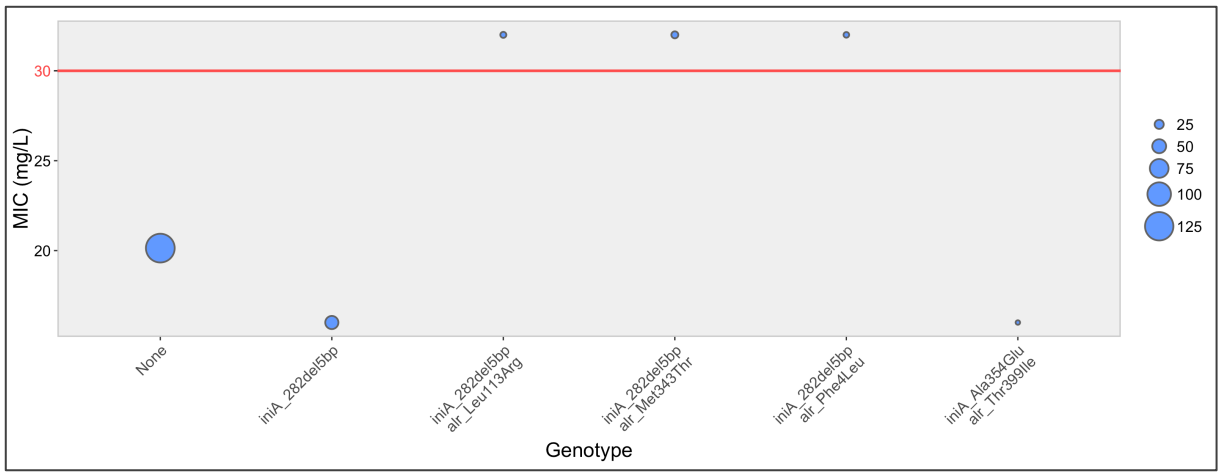

Figure 2.13. Mean MIC of DCS relative to DCS associated genotypes. Genotypes pertain to a subset of 40 *M. tuberculosis* clinical isolates from Portugal. Circle radius is proportional to the number of isolates that have the associated genotype. The red line indicates the CC of DCS for LJ (World Health Organization, 2018b).

The sensitivity values for PAS and DCS were very low (0.308 and 0.333, respectively) (Table 2.1). Such low values indicate a high probability of diagnosing false negative genomic profiles for PAS and DCS, suggesting a lack of PAS and DCS resistance-associated mutations in the present database that could correctly identify PAS and DCS resistance. This is especially true for DCS, as it there are only 2 validated DCS resistance-associated mutations (Table A.1). Drug resistance mechanisms against PAS and DCS involve multiple genes and proteins, but only a few validated mutations and loci were included in the genomic analysis employed for this study. Genes such as *dfrA*, which has been linked with PAS resistance (Moradigaravand et al., 2016), and *ald*, *ddl* and *cycA*, which are known DCS resistance-associated genes (Desjardins et al., 2016; Chen et al., 2017), were not included in the database. A review of genomic variant data has revealed no mutations within the *dfrA*, among PAS-resistant isolates. Notwithstanding, 26.7 % of DCS-resistant showed to harbour at least one mutation within the *ald* gene. Most interesting of all, were the *ddl* and *cycA* genes, for which all DCS-resistance isolates showed at least on mutation pertaining to these genes. Further research should be considered to assess the validity of the abovementioned DCS resistance-associated genes, particularly, the *ddl* and *cycA* as future molecular determinants of DCS resistance.

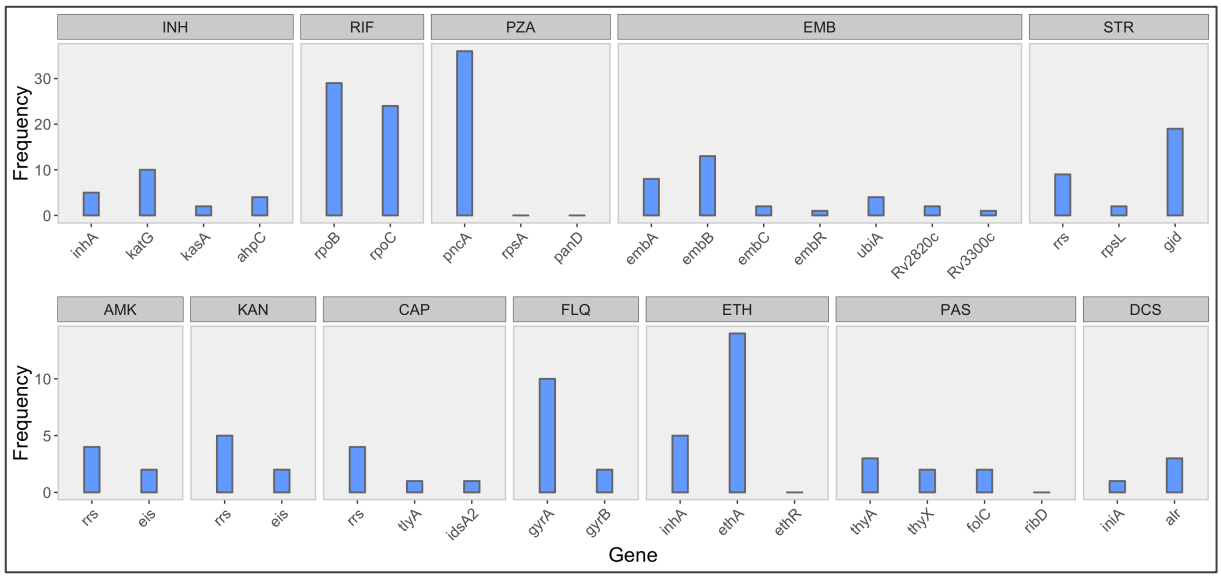

Figure 2.14. Mutational diversity of drug resistance-associated genes within respective drug-resistant *M. tuberculosis* clinical isolates from the local dataset. Drug-resistant isolates were diagnosed by DST.

Conclusion

The present study has revealed valuable insight on the association between drug resistance-associated mutation and respective drug resistance levels from a local subset of *M. tuberculosis* clinical isolates, from Portugal. Higher INH resistance levels were associated with single or double *katG* Ser315Thr mutations or with double *inhA* promoter-ORF mutations, like the *fabG1* C-15T promoter-*inhA* Ile194Thr and *fabG1* C-15T promoter-*inhA* Ser94Ala double mutations. High RIF resistance levels are strictly associated with *rpoB* mutations, where the Ser450Leu mutation is the most prevalent among local and global *M. tuberculosis* isolates. *rpoC* mutations are not associated with RIF resistance, but rather as compensatory mechanisms for the fitness cost imposed by RIF exposure. High EMB resistance levels are dependent on multigenic mutations, particularly a combination of *embB* ORF and *embA* promoter mutations, while single *embB* Met306Val mutations only conferring intermediate resistance levels. Among the second-line aminoglycoside drugs, the highest resistance levels for STR were caused by the *rpsL* Lys43Arg mutation, but for AMK and KAN resistance the well characterized *rrs* A1401G was the one of the highest contributors, along with the novel *rrs* 1076insT insertion. Regarding FLQ resistance, high levels were associated with mutations in the 91 and 94 codons of the *gyrA* gene. High ETH resistance levels, much like INH, were associated with double *inhA* promoter-ORF mutations. The MIC analysis of genomic variant profiles also revealed novel mutations associated with drug resistance, like the *gid* Ala167Asp, *rrs* 1076insT, *ethA* Met1Leu, *alr* Met343Thr and Phe4Leu mutations, and its introduction in resistance-predicting WGS pipelines may improve detection of genomic resistance profiles to second-line drugs. Contrarily, the *embB* Asp354Ala,

which is currently considered to be associated with EMB resistance, did not exhibit an associated MIC above the CC for EMB resistance, and it should be disregarded in future analyses.

The diagnostic power of the employed WGS pipeline was high for most first-line antituberculosis drugs. A higher proportion of false negatives genomic profiles was found for PZA, STR, PAS and DCS, suggesting other mutations or genes may be contributing the drug resistance. EMB and AMK showed higher numbers of associated false positives, which suggest a substantial number of previously validated resistance-associated mutations are not sufficient to solely contribute to drug resistance.

Chapter 3 – Global genome-wide molecular snapshot on the molecular basis of drug resistance and transmission dynamics of *Mycobacterium tuberculosis*

The World Health Organization, as part of their Sustainable Development Goals, has set the ambitious goal of eliminating TB by aiming at a 95 % reduction of its global incidence and mortality rates by 2035 (World Health Organization, 2019). At the core of this ambitious strategy, three pillars encompass patient-centred care, bold policies, supportive systems and intensified research and innovation. It is this very context that WGS is emerging as a technology with the capability of improving some of the existing diagnostic tools and promote surveillance of TB prevalence and *M. tuberculosis* genotypes, detection of drug resistance-associated mutations, and identification of transmission clusters and outbreaks (Meehan et al., 2019).

The standard workflow of WGS of *M. tuberculosis* involves culturing of clinical isolates on solid or liquid media, followed by the extraction of DNA from cells, library preparation and sequencing using short read technologies such as Illumina platforms. Several dry-lab steps are needed once sequencing of reads is complete in order to detect and extract genomic variants such as SNPs, indels, in a high throughput manner, like read quality control, mapping and excision of mobile genetic elements, depending on the nature of the downstream analyses (Cole et al., 1998). Genomic variants can then be compared with a list of high-confidence resistance associated variants enabling WGS to predict monodrug resistance, multidrug resistance or even pan-susceptibility, replacing the need for phenotypic testing, which requires more time and might be more expensive (Coll et al., 2015; The CRyPTIC Consortium and the 100,000 Genomes Project, 2018). Besides drug resistance profiling, WGS has the added advantage to classify strains to one of the 7 human-adapted MTBC lineages, and their sub-lineages, directly derived from variant calls using a list of lineage-defining SNPs (Lipworth et al., 2019; Coll et al., 2014). This is important for understanding population structure and potential phenotypic differences between lineages and for comparing isolates on the global level. Genomic variant data can also be used monitor clinical isolates and study their transmission dynamics. The most common approach involves the use of SNP cut-off-based clustering methods, having comparable results with the widely employed multilocus sequence typing (Meehan et al., 2018). This approach consists of assembling a list of high-confidence, unambiguous SNPs found in each isolate, so that predefined SNP distance thresholds can be used to confidently cluster strains and define recent transmission chains. Considering the MTBC has a very low genetic diversity, thresholds of 5 or 12 SNPs are frequently used to delineate epidemiological links (Walker et al., 2013). The described WGS-based approaches have been shown to perform better than contact tracing and portray a higher resolution than classic approaches such as MIRU-VNTR (Jajou et al., 2018; Wyllie et al., 2018).

The present chapter of this thesis seeks to provide a global high-resolution molecular snapshot of the *M. tuberculosis* population structure along with the geographical dispersion of

known strain lineages and sub-lineages. Moreover, using known resistance markers, this chapter also aims to characterize the geographical distribution of MDR TB and XDR TB isolates, identify transnational genomic clusters (GCs) and, explore the association between strain lineage and resistance associated mutations.

Materials and Methods

Clinical isolates and raw sequence data

A total of 28,385 *M. tuberculosis* genome sequences alongside its respective metadata (date and location of clinical strain's isolation) was extracted from XML files associated with sample IDs deposited at the European Nucleotide Archive (ENA).

Genomic variants pipeline

The study includes an initial set of 28,385 *M. tuberculosis* isolates for which publicly available raw sequence data was retrieved from the ENA. This dataset includes publicly available sequence data made available until the 31st of July, 2018, composed of whole genome sequencing libraries generated by an Illumina platform under a paired-end sequencing mode.

Read mapping, variant calling and strain typing

Upon raw read trimming and filtering, using Trimmomatic, initial mapping and variant calling was carried out using the Snippy pipeline and the reference genome of *M. tuberculosis* H37Rv (NCBI reference sequence: NC_000962.3). Snippy implements read mapping using the Burrows-Wheeler Aligner tool with the BWA-MEM algorithm and, variant calling using Freebayes. Based on variants called, strains were classified according to its (sub)lineage using a SNP barcode (Coll et al., 2014). *In silico*, spoligotyping was carried out using SpoTyping and the respective Shared International Types (SIT)/Clades by querying the SITVIT (Demay et al., 2012).

Resistance predicting pipeline

See the corresponding section from Chapter 2 (p. 22).

Genomic clustering of *M. tuberculosis* clinical isolates

Based on the entire breadth of genomic SNPs identified across the complete dataset of 28,385 available records at ENA, a SNP table was constructed where SNP positions falling in the proline-glutamate/proline-glutamate-glutamate genes and low mappability regions assessed

using a 50-bp-based K-mer score have been removed (Coll et al., 2018). Further SNP calling quality was refined by reclassifying SNPs, whose read coverage and proportion of variant-displaying reads were below the respective 20 and 90 % cut-offs, as missing calls. SNP positions and strains showing a 10 % excess of missing calls were removed from the analysis. The post-filtered dataset was composed by 25,729 *M. tuberculosis* sequence data records comprising of a total of 399,242 SNPs. DNA pseudomolecules were assembled from variant SNP positions and SNP-based pairwise distance calculated between strains using a model which measures the proportion or the number of sites that differ between each pair of sequences. The *M. tuberculosis* strains were clustered using the complete-linkage method. Multi-strain clusters were examined for the presence of identical strains (strains whose pairwise distance is 0). Only one strain, from each identical strain group was preserved based on the availability of associated geographical data and on having the earliest isolation record. This latter step was included to remove existing duplicate records for the same isolate and to avoid statistical bias introduced by strain over-clustering and clonality. The final dataset included 18,445 unique *M. tuberculosis* strains. A hierarchical single-linkage clustering was employed on this final dataset with GCs being set using a 5-SNP maximum distance cut-off.

Statistical analysis of the association between drug resistance-associated genomic variants and the MTBC genetic background

The lineage profiles were determined for the global dataset of *M. tuberculosis* clinical isolates, using a dedicated SNP barcode (Coll et al., 2014). A χ^2 test was employed to determine the association between the genetic background, particularly the L1, L2, L3 and L4, and genomic variants, normalized for drug-resistant isolates. Only mutations with expected frequencies, among lineage-specific associated *M. tuberculosis* isolates of 5 or more were considered suitable for the aforementioned statistical analysis. Mutations significantly associated with the MTBC genetic background were subjected to a second set of four χ^2 tests to determine the single-association with each of the abovementioned lineages. The latter analyses were performed using the Yates correction term. The level of significance was 0.05 (Jones, 2002).

Results and Discussion

Global population structure and geographical dispersion

With the main goal of providing a global molecular snapshot from publicly available *M. tuberculosis* WGS data, the present study includes an initial dataset composed of publicly available WGS raw sequence data for 28,385 strains. Bioinformatic analysis of the raw sequence dataset by mapping against *M. tuberculosis* H37Rv enabled the identification of genome-wide variants along with *in silico* based spoligotyping. Upon quality control and deduplication of the

dataset by removal of isolates showing identical profiles, a final dataset composed of 18,445 *M. tuberculosis* clinical isolates. Moreover, it was possible to extract and assemble metadata on the country of origin for 11,095 isolates and the year of isolation for 10,751 out of the 18,445 isolates. Based on genomic variants detected for individual strains and through *in silico* spoligotyping it was possible to classify each individual strain by (sub)lineage and spoligotyping-based SIT/Clade.

Six human-adapted lineages were detected in the global dataset, accounting for a total of 17,850 *M. tuberculosis* isolates, distributed across L1 (2,196), L2 (5,283), L3 (1,864), L4 (8,440), L5 (34) and L7 (33). L1 and L3 strains were found evenly distributed across the 5 continents (Table 3.1), but occur with greater prevalence in countries around the Indian Ocean, like Djibouti, Malawi, Zimbabwe, Bangladesh, India, Thailand and Vietnam (Table A.4). The L2 and L4, which are known for their ubiquity, are found in each continent with a sizable proportion. Nonetheless, L2 strains were found predominantly in countries from East Asia and Oceania, while L4 strains were predominantly found in African, American and European countries (Table 3.1). L5 and L7 strains had a negligible presence in the global dataset, with L5 strains having a minor presence in Côte D'Ivoire, Italy and United Kingdom, while L7 strains were found exclusively in Ethiopia (Table A.4). The geographical distribution of the aforementioned lineages follows a previously described pattern (Gagneux, 2018).

Table 3.1. Distribution of *M. tuberculosis* clinical isolates from the global dataset, according to the human-adapted MTBC lineages, across 5 continents.

Continent	Relative frequency (%)					
	L1	L2	L3	L4	L5	L7
Africa	11.6	16.5	9.6	61.1	0.0	1.2
America	14.0	16.3	5.5	64.2	0.0	0.0
Asia	17.3	63.8	2.0	16.9	0.0	0.0
Europe	3.9	34.7	13.2	48.1	0.1	0.0
Oceania	4.2	76.9	2.8	16.0	0.0	0.0

A global perspective on drug resistance and diversity of drug resistance-associated mutations

Following the characterization of the populational structure of the global dataset, the mutational diversity associated with drug resistance-associated loci was assessed to infer drug susceptibility profiles using genotypic data. This approach was carried out with a validated database of drug resistance-associated mutations, from which low confidence genes associated with drug resistance were removed to implement a more conservative, and confident, approach to resistance detection. In total, 8,235 unique mutations were identified, with only 624 (7.6 %)

known to be associated with drug resistance. Mutations were screen for their prevalence among respective drug-resistant isolates (Figure 3.1).

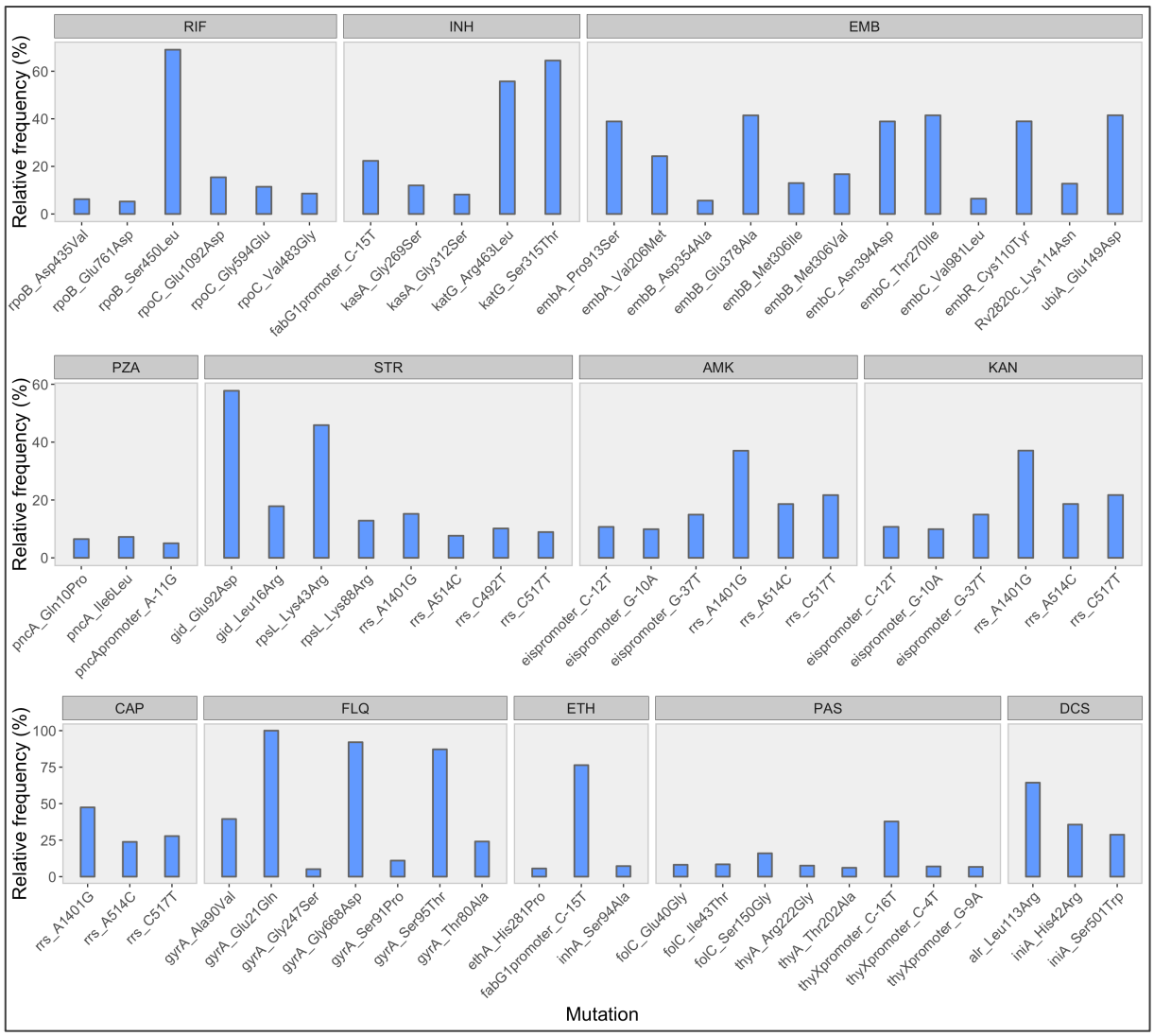

Figure 3.1. Relative frequencies of mutations within respective drug-resistant *M. tuberculosis* clinical isolates from the global dataset. Drug-resistant isolates were diagnosed by WGS. Only mutations with a relative frequency of 5 % or higher are shown.

As expected, the most predominant mutations within INH- and RIF-resistance isolates were the *katG* Ser315Thr and *rpoB* Ser450Leu missense mutations. Mutations within EMB resistance-associated loci did not reveal any singly-predominant mutation, like due to the multigenic dependency required for the acquisition of EMB resistance (Perdigão & Portugal, 2018). PZA-resistant isolates did not reveal any predominant PZA resistance-associated mutation, alluding to the fact that too little is known regarding the molecular determinants of PZA resistance. Omitting the *gid* Glu92Asp mutation, which is a known phylogenetic marker for L2 (Coll et al., 2014), the most prevalent mutation within STR-resistant isolates is the *rpsL* Lys43Arg mutation, associated with high-level STR resistance. SLID-resistant isolates shared the *rrs*

A1401G mutations as their most frequent mutation, although its respective proportion for each drug is lower than previous estimations (Cohen et al., 2014). The most prevalent mutations within FLQ-resistant isolates were the *gyrA* Glu21Gln, Gly668Asp and Ser95Thr substitutions, albeit these are natural polymorphism and don't contribute to FLQ resistance (Lau et al., 2011). Aside these, the second most frequent and FLQ resistance-associated mutations were the *gyrA* Ala90Val and Thr80Ala mutations, where the former is known to contribute to a higher level in a single allelic configuration (Aubry et al., 2006). Regarding ETH resistance, the *fabG1* C-15T promoter mutation was by far the most prevalent contributor of drug resistance. Curiously, *inhA* mutations seem to more frequently associated with ETH resistance than with INH resistance, while prodrug inactivation plays a greater role in the latter (Vilchèze & Jacobs, 2014). The *thyX* C-16T promoter mutations was the most frequent among PAS-resistant isolates, suggesting the thymidylate synthase, encoded by the *thyX* gene as a plausible drug target. Considering ThyX is essential for DNA replication and repair, drug target overexpression is a viable PAS resistance mechanism (Fivian-Hughes et al., 2012). The most prevalent mutation among DCS-resistant isolates was the *alr* Leu113Arg substitution. Alongside the *iniA* His42Arg mutation, these two mutations are the only DCS resistance-associated mutations currently present in the revised mutation database (Table A.1). The *iniA* Ser501Trp mutation also displayed a significant proportion within DCS-resistance, although it is unclear whether it is associated with DCS resistance.

The global dataset comprised of 9,053 (49.1 %) pan-susceptible isolates, 5,688 (30.8 %) DR isolates, 3,393 (18.4 %) MDR isolates and 311 (8.4 %) XDR isolates. A global snapshot of MDR TB was assessed by distributing the respective genomic profiles to each of the available 54 countries (Figure 3.2). Apart from the Democratic Republic of The Congo, Estonia, Ireland, Kazakhstan, Mali, Mozambique, Myanmar/Burma and South Korea, whose respective number of *M. tuberculosis* isolates are too low (<20) to draw significant conclusions, countries like Argentina, Azerbaijan, Bangladesh, Belarus, Brazil, China, Georgia, India, Italy, Moldova, Papua New Guinea, Peru, Portugal, Romania, Russian Federation, Switzerland, Tajikistan, Thailand and Uganda are highly represented regarding the prevalence of MDR TB in this sample. The high prevalence of MDR TB in the aforementioned regions is concordant with previous reports (World Health Organization, 2019), although it is expectedly overinflated in this sample as MDR TB samples are expected to be sequenced more frequently than susceptible isolates or even isolates showing DR profiles. Nevertheless, the data herein conveyed a worrisome picture of the epidemiological impact of MDR TB as a health hazard, especially within high-burden developing countries, where preventive measures and access to effective antituberculosis chemotherapy is usually scarce. Moreover, the fact that new antituberculosis drugs haven't been introduced for over 50 years (apart from a few exceptions), and considering new bactericidal mechanisms haven't been developed, has allowed for the widespread dissemination of drug-resistant *M. tuberculosis* isolates, rendering available antibiotics inadequate for the treatment of drug-resistant TB, specially MDR TB.

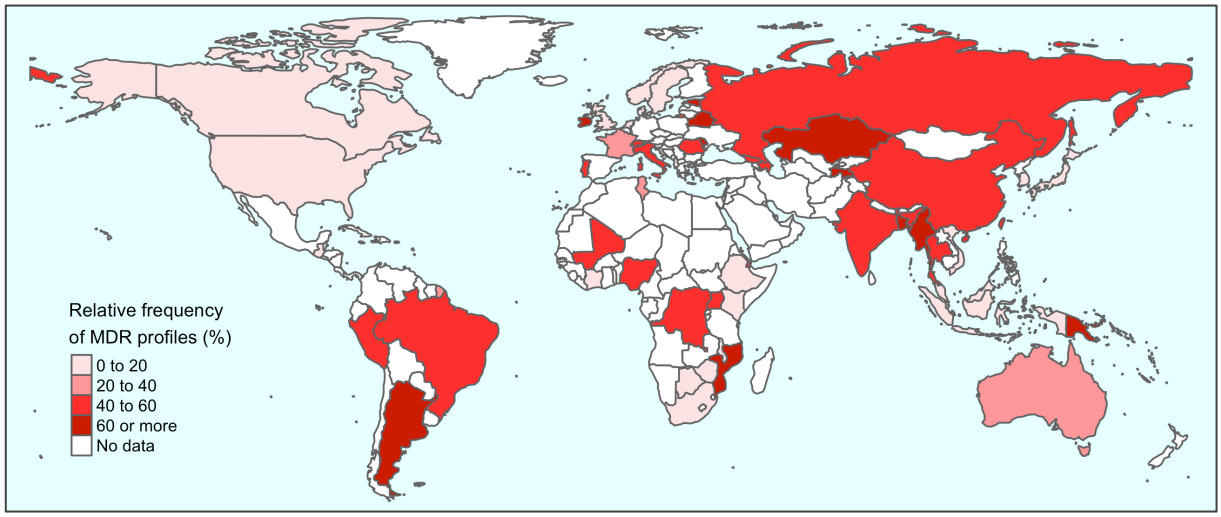

Figure 3.2. Geographical distribution of WGS-derived MDR profiles from the global dataset of *M. tuberculosis* clinical isolates.

Contrarily to the bias towards MDR TB representation in the study sample, the XDR TB rate is less likely to show such sampling bias as it is normalized against MDR profiles. Significantly high proportions of XDR TB can be found in Azerbaijan, Belarus, China, Georgia, Moldova, Portugal, South Africa and United Kingdom (Figure 3.3). Despite its low representation of MDR TB in the present sample (15.8 %), South Africa has one of the highest proportions of XDR TB (20.7 %). A recent prospective study as revealed that the XDR TB epidemic in South Africa is likely due to an increased transmission rate of resistant *M. tuberculosis*, instead of an inadequate treatment of MDR TB (Shah et al., 2017), which is typically attributed to the emergence of MDR TB within high-burden TB countries. This is substantiated by an WGS study which was able to date the emergence of MDR and XDR TB by estimating the mutation rate within this region. Considering that the mutation rate is particularly high in South Africa, it suggested that acquisition of drug resistance-associated mutations occurred soon after antituberculosis chemotherapy was discovered, with the emergence of MDR and XDR possibly having occurred in 1984 and 1995, respectively. A premature occurrence of XDR TB, before its previous reported emergence in 2005, may constitute a plausible justification for the current XDR epidemic crisis occurring in South Africa (Cohen et al., 2015).

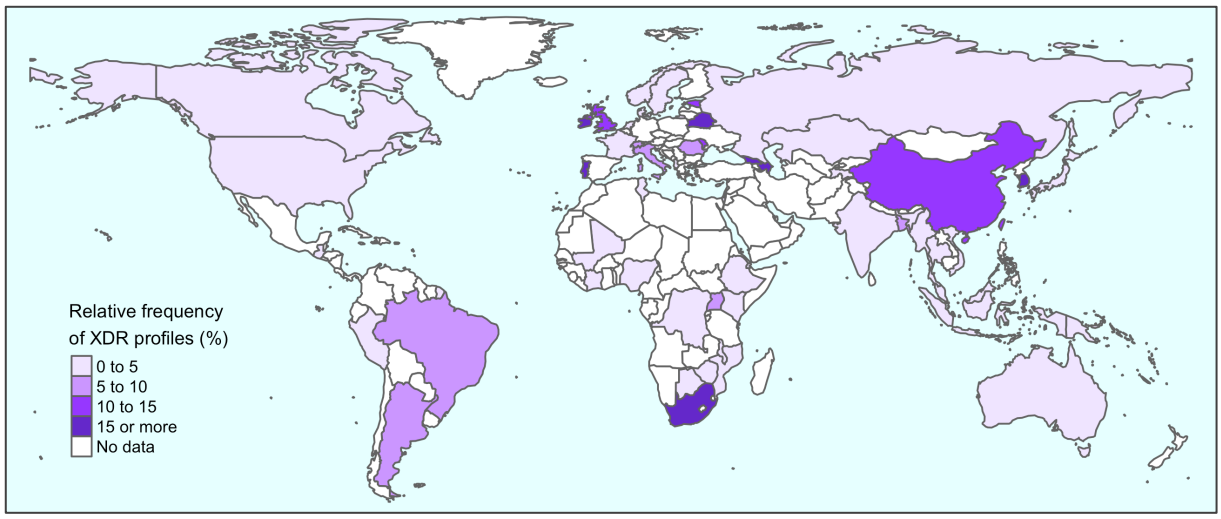

Figure 3.3. Geographical distribution of WGS-derived XDR profiles from the global dataset of *M. tuberculosis* clinical isolates.

An analysis of the temporal distribution of MDR and XDR profiles was conducted using the dates of isolation within the metadata associated to the present study (Figure 3.4). Accompanying the introduction of NGS technologies, an increased efforted for sequencing *M. tuberculosis* isolates has been done, with an exponential increase in sequenced isolates by 2008. Although there is noticeably overaccumulation of MDR and XDR profiles within the last decade, this does not mean that MDR and XDR TB is a recently occurring phenomenon. As the number of sequenced genomes of old *M. tuberculosis* clinical isolates increases, a clearer picture of MDR and XDR emergence will be obtained. Considering 3 WGS-derived XDR isolates were characterized dating between 2004 and 2005, the present data suggests XDR TB may have occurred earlier than its international recognition back in 2006 (Center for Disease Control and Prevention, 2006).

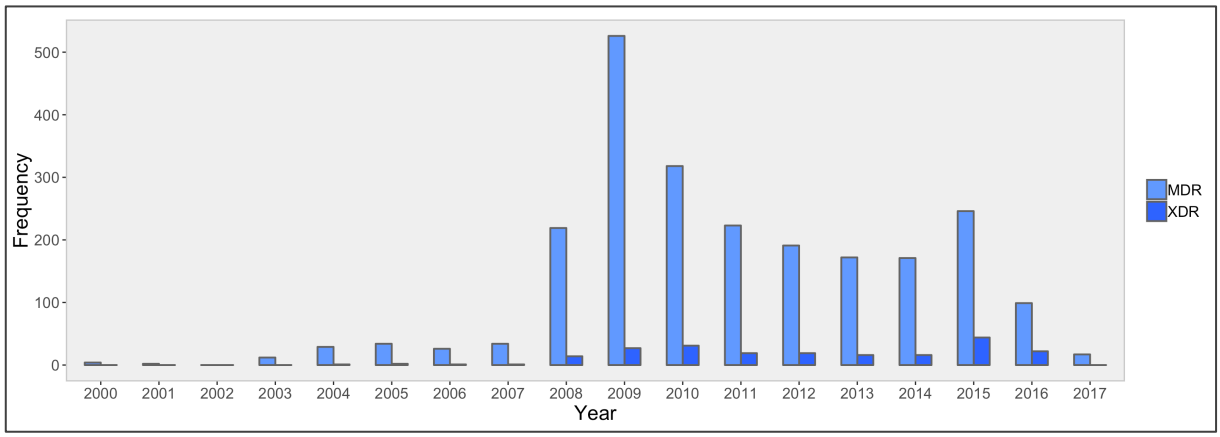

Figure 3.4. Temporal distribution of WGS-derived MDR and XDR profiles from the global dataset of *M. tuberculosis* clinical isolates.

Linking resistance development and molecular basis of resistance with the MTBC genetic background

Further investigation was conducted to assess the association between drug resistance-associated mutations and the MTBC genetic background, by examining their distribution across different lineages. Within the 8,235 identified mutations, only 31 showed a significant association with any of the of the 4 *M. tuberculosis sensu stricto* lineages: L1, L2, L3 or L4. Within these 31 mutations, only 15 were associated with drug resistance (Table 3.2).

Regarding INH resistance-associated mutations, the C-15T *fabG1* promoter mutation and the *kasA* Gly269Ser mutations were found to be significantly associated with L4 (Table 3.2). Contradicting evidence suggests the C-15T *fabG1* promoter mutation is primarily linked with L1 strains. Despite this, the majority of isolates with low-level INH resistance that harboured the C-15T *fabG1* promoter mutation belonged to L4 (Fenner et al., 2012). Considering that in the previous chapter the C-15T *fabG1* promoter mutation was proven to be associated with low-level INH resistance, its association with L4 strains is not unfounded and could be primarily driven by the high association with the Lisboa3 and Q1 strains in Portugal.

Similarly, with INH-resistant isolates, the C-15T *fabG1* promoter mutation, which confers cross-resistance to ETH, was also found to be associated with the L4 among ETH-resistant isolates.

Two RIF resistance-associated mutations were found to be associated with the MTBC genetic background. The Asp435Val substitution in the *rpoB* gene was found to be significantly associated with L4 while the Ser450Leu substitution is associated with L2. The *rpoB* Asp435Val mutation has been previously associated with L4-specific WGS analyses (Mortimer et al., 2018), suggesting that it is positively selected under this phylogenetic subgroup of the MTBC. There is ambiguity on which lineage the prevalent *rpoB* Ser450Leu mutation is associated to. A study has shown that the Ser450Leu mutation is less likely to be associated with the Beijing genotype – the main L2 sub-lineage – than with non-Beijing strains (Park et al., 2005), while another study revealed the opposite (Dymova et al., 2011). The data herein obtained demonstrates its association with the L2 lineage probably owing to the highly dynamic spread of the Beijing family (L2) across Asia towards Europe. Specific MDR TB sub-clades of the Europe-Russia W148 and Central Asia Outbreak clades presently comprise MDR TB superclusters spanning mainly across Europe and Asia with RIF resistance mainly driven by the *rpoB* Ser450Leu mutation (Merker et al., 2015).

Table 3.2. Drug resistance-associated mutations with a significant association with one of the 4 main human-adapted MTBC lineages, normalized for respective drug-resistant *M. tuberculosis* clinical isolates. P-values <0.05 imply statistical significance.

Mutation	Drug	P-value				Relative frequency (%)			
		L1	L2	L3	L4	L1	L2	L3	L4
fabG1	INH	0.021	<0.001	<0.001	<0.001	11.3	32.8	3.2	52.6[1]
promoter_C-15T	ETH	0.001	<0.001	0.174	<0.001	11.3	32.8	3.2	52.6[1]
*kasA*_Gly269Ser	INH	<0.001	<0.001	<0.001	<0.001	0.1	0.0	0.0	99.9[1]
*rpoB*_Asp435Val	RIF	0.016	<0.001	0.001	<0.001	0.8	33.9	9.2	56.2[1]
*rpoB*_Ser450Leu	RIF	<0.001	<0.001	<0.001	<0.001	1.8	53.9[1]	3.6	40.7
*embB*_Asp354Ala	EMB	<0.001	<0.001	0.113	<0.001	0.3	87.0[1]	1.6	11.1
*embB*_Glu378Ala	EMB	<0.001	<0.001	<0.001	<0.001	99.8[1]	0.0	0.0	0.2
*embB*_Gly406Ala	EMB	<0.001	0.110	0.147	<0.001	2.0	28.5	1.6	67.9[1]
*embB*_Met306Val	EMB	<0.001	<0.001	0.208	<0.001	2.9	63.5[1]	3.7	29.9
*embC*_Thr270Ile	EMB	<0.001	<0.001	<0.001	<0.001	99.8[1]	0.0	0.0	0.2
*rpsL*_Lys43Arg	STR	<0.001	<0.001	0.182	<0.001	1.9	80.6[1]	3.4	14.1
*rrs*_A1401G	AMK	0.002	0.043	0.398	<0.001	2.3	57.3[1]	3.1	37.3
	KAN	0.003	0.038	0.394	<0.001	2.3	57.3[1]	3.1	37.3
	CAP	<0.001	0.009	0.018	<0.001	2.3	57.3[1]	3.1	37.3
*rrs*_C517T	STR	<0.001	<0.001	0.797	<0.001	7.1	83.1[1]	3.7	6.1
	AMK	0.002	<0.001	0.885	<0.001	7.1	83.1[1]	3.7	6.1
	KAN	0.001	<0.001	0.887	<0.001	7.1	83.1[1]	3.7	6.1
	CAP	0.060	<0.001	0.334	<0.001	7.1	83.1[1]	3.7	6.1
eis promoter_C-12T	AMK	0.005	<0.001	0.003	<0.001	0.5	21.6	0.0	77.8[1]
	KAN	0.006	<0.001	0.003	<0.001	0.5	21.6	0.0	77.8[1]
eis promoter_G-10A	AMK	0.002	<0.001	0.016	<0.001	0.0	74.0[1]	0.6	25.4
	KAN	0.002	<0.001	0.016	<0.001	0.0	74.0[1]	0.6	25.4
eis promoter_G-37T	AMK	<0.001	<0.001	0.001	<0.001	0.0	98.4[1]	0.0	1.6
	KAN	<0.001	<0.001	<0.001	<0.001	0.0	98.4[1]	0.0	1.6

[1] Proportions >50 % imply a significant association with the lineage in question.

EMB resistance is attributed to wide variety of mutations, from multiple loci. The lineage-association analysis has revealed 5 EMB resistance-associated mutations with significant association with the MTBC genetic background (Table 3.2). The Asp354Ala and Gly406Ala mutations in the *embB* gene are associated with the L2 and L4, respectively. Lineage associations for these mutations were only found through single-patient reports (Aung et al., 2016; Ali et al., 2015) which are most likely attributed to the genetic makeup of the *M. tuberculosis* population from which the patient got infected rather than a controlled unbiased analysis. Notwithstanding, the *embB* Glu378Ala and the *embC* Thr270Ile mutations were found to be exclusively associated

with L1 strains. Concordant evidence confirms these findings while also revealing their added association with the remaining ancestral L5 and L6 (Brossier et al., 2015). While Mortimer and co-workers (2018) have found that the *embB* Met306Val mutation is more associated with L4 strains, the present study contradicts this, suggesting that the abovementioned mutation to be significantly more associated with L2.

Prevalent mutations associated with intermediate to high-level drug resistance such as the *rpoB* Ser450Leu and *embB* Met306Val substitutions were the ones that shared ambivalent results when compared with available literature. This is likely due to their prevalence among several MTBC lineages across multiple geographical locations, which depending on the nature of the clinical sample could provide conflicting association results.

The *rpsL* Lys43Arg mutation, which confers high-level STR resistance was found to be associated primarily with L2 (Table 3.2). This mutation has been previously described having a significant association with Beijing strains, when compared with non-Beijing strains (Smittipat et al., 2016; Sun et al., 2010). The *rrs* A1401G and C517T mutations were found to be significantly associated with L2 among STR, AMK, KAN and CAP resistant *M. tuberculosis* isolates, with the exception of the *rrs* A1401G mutation which had no association with any of the 4 main human-adapted lineages, within STR-resistant isolates. Altogether, *rrs* mutations have been previously described to be significantly more associated with the Beijing genotype than with non-Beijing strains, among SLID-resistant strains (Miotto et al., 2012). Three *eis* promoter mutations were found to be significantly associated with the MTBC genetic background. Apart from the C-12T *eis* promoter mutation, all *eis* promoter mutations are associated with L2, which has been corroborated by a previous study (Casali et al., 2014).

Global diversity of GCs and transnational clone dispersal

To investigate the degree of clonality within the study sample a hierarchical clustering analysis based upon 399,242 high-quality SNPs was undertaken, with the main goal of identifying GCs associated with recent transmission of the disease. A cut-off distance of 5 SNPs was applied, enabling the identification of a total of 1,791 GCs encompassing 7,373 isolates. The proportion of clustered and non-clustered strains was ascertained for each of the human-adapted MTBC lineages, as well as their distribution across different resistance profiles (Table 3.3). L4 strains were found to be more clustered than strains from other lineages. This contradicts the findings of Guerra-Assunção and co-workers (2015), which found L2 and L3 strains to be more clustered than L4 strains. Nevertheless, this study corroborates the finding that L1 is the least likely to be clustered. The propensity for strains to cluster together suggests a higher likelihood of recent infection, as their close relatedness imply the occurrence of an outbreak of a single clone which disseminated throughout an immunosuppressed population. The fact that L2 and L4 strains, were more likely to be clustered than other lineages, elucidates their increased transmission

capabilities and worldwide dissemination, as geographically distant strains from these lineages have a higher probability of being closely related.

High proportions of MDR TB were found among L2, L3 and L4 strains, but especially greater within L3 strains, although the causes for this are unclear, given the previously mentioned lack of statistically significant association between that lineage and prevalent drug resistance-associated mutations essential for the microevolutionary process towards MDR.

A greater diversity of sub-lineages or clades was found within L4 (Table A.5), accounting for 49 unique clades. Fifteen of these belonged to the ill-defined T family, the most prevalent in the L4 sample (34.5 %) of this study. Following that, the second most prevalent strain family was the LAM, encompassing 11 different clades and found in 24.0 % of L4 strains. Clades belonging to the T family were found to harbour a higher proportion of MDR strains, compared to clades belonging to the LAM family (Table A.5).

The second most diverse lineage was the L1, with 15 unique clades, almost all of which were from the EAI family (11), which represented 73.2 % of all L1 strains. As stated earlier and summarily, L1 strains are usually found non-clustered, a pattern found in all clades belonging to L1. Furthermore, MDR profiles were not present in nearly all the L1 clades, with each of them harbouring a great majority of pan-susceptible isolates.

The great majority of L3 strains belonged to the CAS family (78.5), within 4 associated clades. Curiously, the MDR TB rates in these clades were especially high, and considering that most L3 strains belong to the CAS family, it is not surprising that L3 strains displayed higher rates of MDR TB than other lineages (Table 3.3).

Concerning L2, the majority of its strains belonged to the Beijing family (88,6 %) and associated rates MDR TB were also high (45.3 %).

Altogether, the rates of MDR TB within strains belonging to the T, LAM, CAS and Beijing genotypes were all undisputedly higher compared to that of the EAI genotype, which has also been previously demonstrated in previous studies (Couvin et al., 2019; Nguyen et al., 2016). This study has revealed CAS strains to have a higher propensity of developing MDR, when compared with the T, LAM and Beijing genotypes, although it is unclear whether sampling bias played a role.

Table 3.3. Distribution of GCs, clustered and non-clustered associated *M. tuberculosis* isolates, and of drug resistance profiles across the human-adapted MTBC lineages.

Lineage	GCs	Relative frequency (%)					
		Clustered	Non-clustered	S[1]	DR	MDR	XDR
L1	163	20.4	79.6	94.5	5.1	0.1	0.2
L2	510	39.9	60.1	20.3	32.2	44.3	3.2
L3	221	37.6	62.4	15.0	7.5	77.1	0.4
L4	842	45.5	54.5	23.6	16.4	58.5	1.5
L5	0	0.0	100.0	97.1	2.9	0.0	0.0
L7	5	36.4	63.6	100.0	0.0	0.0	0.0

[1] Pan-susceptible.

Of the total 1,791 GCs, 68 (3.8 %) harboured transnational *M. tuberculosis* strains displaying varying resistance profiles. Six of these transnational GCs were found to harbour at least one MDR *M. tuberculosis* strain from each encompassing country, across several geographical regions like, Europe, Central Asia, South America and Southeast Asia (Figure 3.5). The abovementioned GCs demonstrate an overrepresentation of the modern Beijing genotype (L2.2.1.1) which are characterized for having a genetic deletion in the RD207 locus unlike its ancestral counterpart, the proto-Beijing sub-lineage (L2.1) (Luo et al., 2015). Given that modern Beijing strains are reported to have a higher proclivity to be hypervirulent, ubiquitous and acquire MDR it is not uncommon to find the majority of transnational clusters to have a modern Beijing genetic constitution distributed across various geographical regions (Ribeiro et al., 2014; Luo et al., 2015; Ford et al., 2013). The remaining non-Beijing GCs harboured strains belonging to L4, all representing different L4-affiliated clades. The GC1314 harboured transnational strains belonging to the L4.8 sub-lineage, which is mainly comprised of strains from the T clade (Coll et al., 2014). According to Stucki and co-workers (2016) the L4.8 is a generalist sub-lineage, meaning it can be found globally. Nonetheless, the results conveyed by this study have revealed a higher tendency for L4.8 strains to occur in Europe. The GC219 is comprised of strains belonging to the L4.2.1 sub-lineage, also known as the Ural clade (Coll et al., 2014). Considering that the Ural family does not disseminate as quickly as other L2 sub-lineages, and that its point of origin occurred in present day South Ukraine and Georgia, it is not surprising to find that the strains pertaining to this GC are circumscribed within Eastern European countries, such as Moldova and the Russian Federation (Mokrousov, 2012). The GC286 harboured strains belonging to the L4.3.3 sub-lineage, which is part of the LAM family (Coll et al., 2014). The LAM family is the most frequent and widespread sub-lineage within L4 and has been predicted to have originated in Europe. It is hypothesized that its ubiquity is likely due to European migration and colonization, which justifies its prevalence in South America (Stucki et al., 2016).

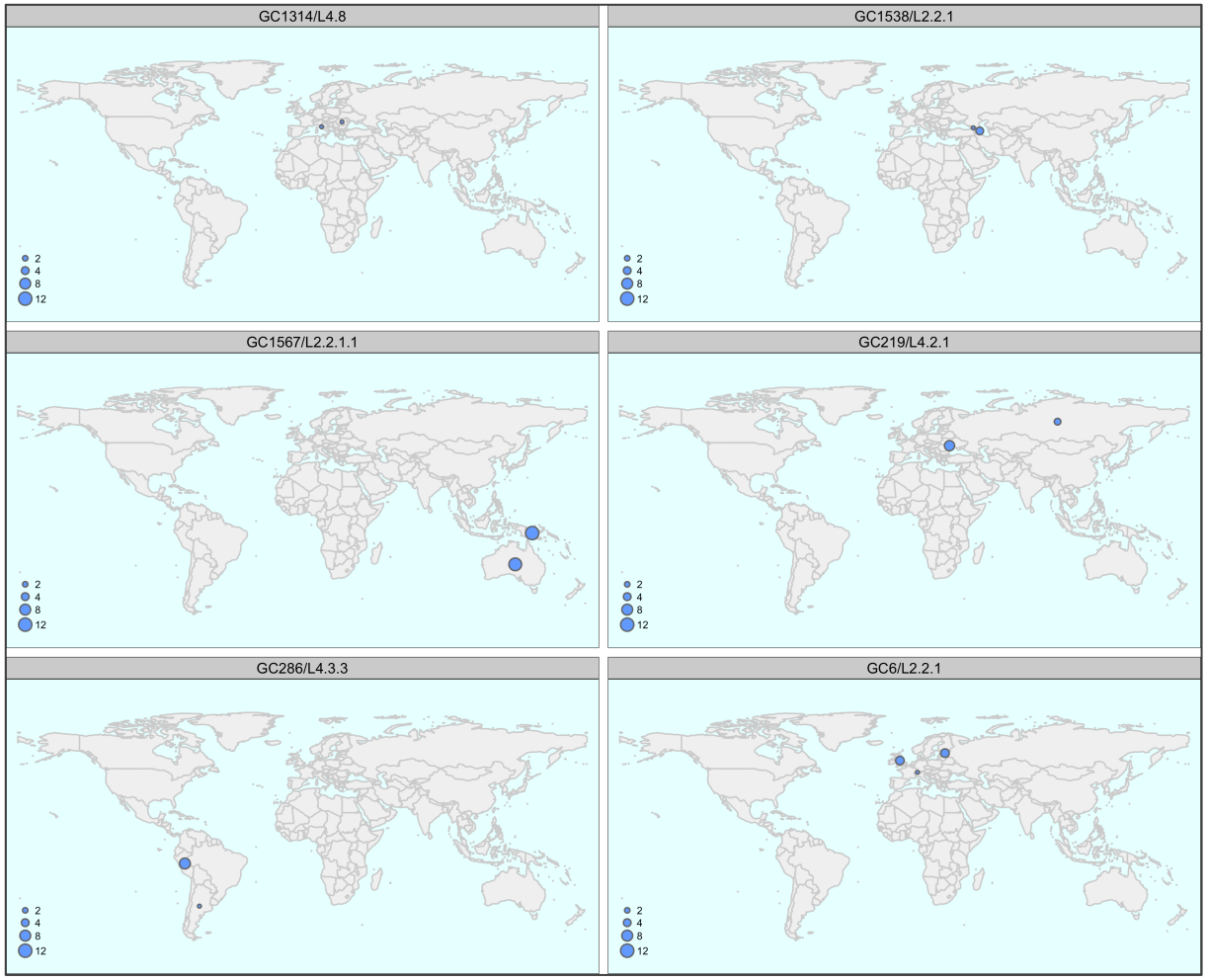

Figure 3.5. Geographical distribution of 6 transnational GCs harbouring closely related *M. tuberculosis* isolates. Circle radius is proportional to the number of *M. tuberculosis* isolates specific to the respective country.

In light of these results, it is clear that L2 *M. tuberculosis* strains, particularly those belonging to the modern Beijing sub-lineage, are specially dangerous, as these are more prone to be involved in recent infections and increased efforts must be taken to prevent its further worldwide dissemination and acquisition of MDR and even XDR, considering the majority of lineage-associated drug resistance associated mutations are more likely to arise within *M. tuberculosis* strains belonging to the L2 genotype, including the prevalent *rpoB* Ser450Leu mutation, which constitutes as one of the first steps in the microevolution path towards MDR. Added research should be carried out to assess whether the remaining mutations identified by this study arise within monophyletic groups or through a phylogenetic-independent manner, as homoplasies, which are more likely to result from common selective pressures like antituberculosis chemotherapy.

Conclusion

The genome-wide analysis of the global dataset of m tuberculosis clinical isolates gave an insightful perspective of the worldwide distribution of the human-adapted MTBC lineages, revealing the ubiquitous nature of L2 and L4, where human migration is hypothesized to have played an important role. The geographical distribution of WGS-derived MDR and XDR profiles accurately mirrored previous descriptions of reference reports, indicating WGS pipelines could become a standard method for monitoring drug-resistant TB.

This study also characterized the mutational diversity of the global clinical dataset, identifying the most prevalent mutations among drug-resistant isolates, which revealed *rpoB* Ser450Leu, *katG* Ser315Thr, *embB* Glu378Ala, *embB* Met306Val/Ile, *embC* Thr270Ile, *Rv2820c* Lys114Asn, *rpsL* Lys43Arg, *rrs* A1401G, *gyrA* Ala90Val, *fabG1* promoter C-15T, *thyX* promoter C-16T and *alr* Leu113Arg as predominant drug resistance-associated mutations, while highlighting other prevalent uncharacterized mutations which could be associated with drug resistance.

Significant associations between drug resistance-associated mutations and the MTBC genetic background were found among L2 and L4 primarily, with L2 being significantly associated with the majority of prevalent *rpoB* Ser450Leu, *fabG1* promoter C-15T, *embB* Met306Val, *rpsL* Lys43Arg and *rrs* A1401G mutations, suggesting an increased propensity for developing MDR.

Higher proportions of MDR TB were found to be strictly linked with modern L2, L3 and L4, specifically with their respective Beijing, CAS and LAM sub-lineages. L2 and L4 displayed high proportions of affiliated strains within GCs when compared with the remaining MTBC lineages, implying increased transmissibility, which explains their overrepresentation in transnational clusters.

The fact that *M. tuberculosis* isolates belonging to L2 and L4 have a higher tendency for being clustered, displaying a greater association with drug resistance-associated mutations, and having been found in the majority of transnational GCs, goes to show how far these strains are capable of disseminating and how fast these can acquire drug resistance. Understanding the dynamics of the MTBC genetic background on drug resistance acquisition can help improve current efforts in eliminating TB worldwide.

Chapter 4 – Closing considerations and future perspectives

This thesis allowed for the characterization, on a genome-wide scale, of drug resistance-associated mutations of two distinct sets of *M. tuberculosis* clinical isolates, each intend for a local and global study of the MTBC resistome diversity, respectively. The local *M. tuberculosis* analysis focused on how different allelic configurations, pertaining to drug resistance-associated loci, contributed to differing levels of drug resistance, while concomitantly allowing for the validation of the diagnostic features of the employed WGS pipeline for detecting genomic drug resistance profiles, using drug resistance phenotypic profiles. Accordingly, novel interactions between drug resistance-associated genotypes and respective drug resistance levels were described, as well as novel individual drug resistance-associated mutations. The validation of the aforementioned WGS pipeline as a diagnostic tool also clarified the accuracy of known drug resistance-associated mutation for detecting true drug resistance genotypes, while emphasizing the need to incorporate additional mutations to improve the detection of second-line drug resistance genotypes, primarily.

The global analysis sought to characterize the diversity of drug resistance-associated mutations and genomic drug resistance profiles across multiple geographical regions, and throughout the different lineages and sub-lineages that comprise the genetic background of the MTBC. Moreover, strains associated within these same lineages and sub-lineages were examined for their intrinsic aptitude to cluster together and therefore extrapolate their innate transmissibility and capability to disseminate cross-nationally. A statistical analysis of the association between drug resistance associated mutations and the lineage diversity of the MTBC, revealed that the majority of first- and second-line drug resistance-associated mutations were linked with the prolific L2 and L4, which justifies their previously reported increased proclivity to develop MDR and XDR. Additionally, strains pertaining to these same lineages were found ubiquitously distributed which correlates with the aforementioned tendency for strains belonging to these lineages to form high-density clustering patterns.

For future analyses, a phylogenomic analysis of the global dataset of *M. tuberculosis* clinical isolates is advised to better understand the evolutionary mechanisms behind the phylogenetic distribution of drug resistance-associated mutations, while expanding the current knowledge on the microevolutionary dynamics of drug resistance-associated mutations acquisition across different lineages and sub-lineages of the MTBC.

References

Aaron, L., Saadoun, D., Calatroni, I., Launay, O., Mémain, N., Vincent, V., … Lortholary, O. (2004). Tuberculosis in HIV-infected patients: A comprehensive review. *Clinical Microbiology and Infection*, *10*(5), 388–398. doi: 10.1111/j.1469-0691.2004.00758.x

Abebe, F., & Bjune, G. (2006). The emergence of Beijing family genotypes of *Mycobacterium tuberculosis* and low-level protection by bacille Calmette-Guérin (BCG) vaccines: Is there a link? *Clinical and Experimental Immunology*, *145*(3), 389–397. doi: 10.1111/j.1365-2249.2006.03162.x

Alderwick, L. J., Harrison, J., Lloyd, G. S., & Birch, H. L. (2015). The Mycobacterial Cell Wall – Peptidoglycan and Arabinogalactan. *Cold Spring Harbor Perspectives in Medicine*, *5*. doi: 10.1101/cshperspect.a021113

Ali, A., Hasan, Z., McNerney, R., Mallard, K., Hill-Cawthorne, G., Coll, F., … Hasan, R. (2015). Whole genome sequencing based characterization of extensively drug-resistant *Mycobacterium tuberculosis* isolates from pakistan. *Plos One*, *10*(2), 1–17. doi: 10.1371/journal.pone.0117771

Andrews, J. R., Noubary, F., Walensky, R. P., Cerda, R., Losina, E., & Horsburgh, C. R. (2012). Risk of Progression to Active Tuberculosis Following Reinfection with *Mycobacterium tuberculosis*. *Clinical Infectious Diseases*, *54*. doi: 10.1093/cid/cir951

Anthony, R. M., den Hertog, A. L., & van Soolingen, D. (2018). 'Happy the man, who, studying nature's laws, Thro' known effects can trace the secret cause.' Do we have enough pieces to solve the pyrazinamide puzzle? *Journal of Antimicrobial Chemotherapy*, *73*(7), 1750–1754. doi: 10.1093/jac/dky060

Aubry, A., Veziris, N., Cambau, E., Jarlier, V., & Fisher, L. M. (2006). Novel Gyrase Mutations in Quinolone-Resistant and -Hypersusceptible Clinical Isolates of *Mycobacterium tuberculosis*: Functional Analysis of Mutant Enzymes. *Antimicrobial Agents and Chemotherapy*, *50*(1), 104–112. doi: 10.1128/AAC.50.1.104

Aung, H. L., Tun, T., Moradigaravand, D., Köser, C. U., Nyunt, W. W., Aung, S. T., … Hill, P. C. (2016). Whole genome sequencing of multidrug-resistant *Mycobacterium tuberculosis* isolates from Myanmar. *Journal of Global Antimicrobial Resistance*, *6*, 113–117. doi: 10.1016/j.jgar.2016.04.008

Baddam, R., Kumar, N., Wieler, L. H., Lankapalli, A. K., Ahmed, N., Peacock, S. J., & Semmler, T. (2018). Analysis of mutations in pncA reveals non-overlapping patterns among various lineages of *Mycobacterium tuberculosis*. *Scientific Reports*, *8*(1), 1–9. doi: 10.1038/s41598-018-22883-9

Baker, L. V., Brown, T. J., Maxwell, O., Gibson, A. L., Fang, Z., Yates, M. D., & Drobniewski, F. A. (2005). 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, *49*(4), 1455–1464. doi: 10.1128/AAC.49.4.1455-1464.2005

Barclay, W. R., Ebert, R. H., & Kochweser, D. (1953). Mode of action of isoniazid. *American Review of Tuberculosis*, *67*, 490–496.

Barnes, P. F., & Cave, M. D. (2003). Molecular Epidemiology of Tuberculosis. *New England Journal of Medicine*, *349*(12), 1149–1156. doi: 10.1056/NEJMra021964

Borkowska, D. I., Napiórkowska, A. M., Brzezińska, S. A., Kozińska, M., Zabost, A. T., & Augustynowicz-Kopeć, E. M. (2017). From Latent Tuberculosis Infection to Tuberculosis. News in Diagnostics (QuantiFERON-Plus). *Polish Journal of Microbiology*, *66*(1), 5–8. doi: 10.5604/17331331.1234987

Brossier, F., Sougakoff, W., Bernard, C., Petrou, M., Adeyema, K., Pham, A., … Veziris, N. (2015). Molecular analysis of the *embCAB* locus and *embR* gene involved in ethambutol resistance in clinical isolates of *Mycobacterium tuberculosis* in France. *Antimicrobial Agents and Chemotherapy*, *59*(8), 4800–4808. doi: 10.1128/AAC.00150-15

Budzik, J. M., Jarlsberg, L. G., Higashi, J., Grinsdale, J., Hopewell, P. C., Kato-maeda, M., & Nahid, P. (2014). Pyrazinamide Resistance, *Mycobacterium tuberculosis* Lineage and Treatment Outcomes in San Francisco, California. *Plos One*, *9*(4), 1–5. doi: 10.1371/journal.pone.0095645

Buriánková, K., Doucet-Populaire, F., Dorson, O., Gondran, A., Ghnassia, J. C., Weiser, J., & Pernodet, J. L. (2004). Molecular Basis of Intrinsic Macrolide Resistance in the *Mycobacterium tuberculosis* Complex. *Antimicrobial Agents and Chemotherapy*, *48*(1), 143–150. doi: 10.1128/AAC.48.1.143-150.2004

Buu, T. N., van Soolingen, D., Huyen, M. N. T., Lan, N. T. N., Quy, H. T., Tiemersma, E. W., … Cobelens, F. G. J. (2012). Increased transmission of Mycobacterium tuberculosis Beijing genotype strains associated with resistance to streptomycin: A population-based study. *Plos One*, *7*(8), 1–10. doi: 10.1371/journal.pone.0042323

Canetti, G., Froman, S., Grosset, J., Hauduroy, P., Langerova, M., Mahler, H. T., … Sula, L. (1963). Mycobacteria: Laboratory Methods for Testing Drug Sensitivity and Resistance. *Bulletin of the World Health Organization*, *29*(12), 565–578.

Canetti, G., Fox, W., Khomenko, A., Mahler, H. T., Menon, N. K., Mitchison, D. A., … Smelev, N. A. (1969). Advances in techniques of testing mycobacterial drug sensitivity, and the use of sensitivity tests in tuberculosis control programmes. *Bulletin of the World Health Organization*, *41*(1), 21–43.

Casali, N., Nikolayevskyy, V., Balabanova, Y., Harris, S. R., Ignatyeva, O., Kontsevaya, I., … Drobniewski, F. (2014). Evolution and transmission of drug-resistant tuberculosis in a Russian population. *Nature Genetics*, *46*(3), 279–286. doi: 10.1038/ng.2878

Cave, M. D., Eisenach, K. D., Templeton, G., Salfinger, M., Mazurek, G., Bates, J. H., & Crawford, J. T. (1994). Stability of DNA fingerprint pattern produced with IS6110 in strains of Mycobacterium tuberculosis. *Journal of Clinical Microbiology*, *32*(1), 262–266.

Cavusoglu, C., Karaca-Derici, Y., & Bilgic, A. (2004). In-vitro activity of rifabutin against rifampicin-resistant *Mycobacterium tuberculosis* isolates with known *rpoB* mutations. *Clinical Microbiology and Infection*, *10*(7), 662–665. doi: 10.1111/j.1469-0691.2004.00917.x

Chackerian, A. A., Alt, J. M., Perera, T. V., Dascher, C. C., & Behar, S. M. (2002). Dissemination of Mycobacterium tuberculosis is influenced by host factors and precedes the initiation of T-cell immunity. *Infection and Immunity*, *70*(8), 4501–4509. doi: 10.1128/IAI.70.8.4501-4509.2002

Center for Disease Control and Prevention. (2006). Emergence of Mycobacterium tuberculosis with extensive resistance to second-line drugs – worldwide, 2000-2004. *Morbidity and Mortality Weekly Report*, *55*(11), 301-305.

Center for Disease Control and Prevention. (2018). Mycobacterium tuberculosis Complex Drug Susceptibility Testing Report for August 2018 Survey.

Chae, H., & Shin, S. J. (2018). Importance of differential identification of *Mycobacterium tuberculosis* strains for understanding differences in their prevalence, treatment efficacy, and vaccine development. *Journal of Microbiology*, *56*(5), 300–311. doi: 10.1007/s12275-018-8041-3

Chen, J., Zhang, S., Cui, P., Shi, W., Zhang, W., & Zhang, Y. (2017). Identification of novel mutations associated with cycloserine resistance in *Mycobacterium tuberculosis*. *Journal of Antimicrobial Chemotherapy*, *72*(12), 3272–3276. doi: 10.1093/jac/dkx316

Cohen, K. A., Bishai, W. R., & Pym, A. S. (2014). Molecular Basis of Drug Resistance in Mycobacterium tuberculosis. *Molecular Genetics of Mycobacteria*, *2*(3), 413–429. doi: 10.1128/microbiolspec.mgm2-0036-2013

Cohen, K. A., Abeel, T., Manson McGuire, A., Desjardins, C. A., Munsamy, V., Shea, T. P., … Earl, A. M. (2015). Evolution of Extensively Drug-Resistant Tuberculosis over Four Decades: Whole Genome Sequencing and Dating Analysis of *Mycobacterium tuberculosis* Isolates from KwaZulu-Natal. *Plos Medicine*, *12*(9), 1–22. doi: 10.1371/journal.pmed.1001880

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., … Barrel, B. G. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, *393*, 537–544. doi: 10.1038/31159

Coll, F., McNerney, R., Guerra-Assunção, J. A., Glynn, J. R., Perdigão, J., Viveiros, M., … Clark, T. G. (2014). A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nature Communications*, *5*, 4–8. doi: 10.1038/ncomms5812

Coll, F., McNerney, R., Preston, M. D., Guerra-Assunção, J. A., Warry, A., Hill-Cawthorne, G., … Clark, T. G. (2015). Rapid determination of anti-tuberculosis drug resistance from whole genome sequences. *Genome Medicine*, *7*(1), 1–10. doi: 10.1186/s13073-015-0164-0

Coll, F., Phelan, J., Hill-Cawthorne, G. A., Nair, M. B., Mallard, K., Ali, S., … Matsumoto, T. (2018). Genome-wide analysis of multi- and extensively drug-resistant. *Nature Genetics*, *50*, 307–316 doi: 10.1038/s41588-017-0029-0

Comas, I., Borrell, S., Roetzer, A., Rose, G., Malla, B., Kato-Maeda, M., … Gagneux, S. (2012). Whole genome sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. *Nature Genetics*, *44*(1), 106–110. doi: 10.1038/ng.1038

Cooper, A. M. (2009). Cell-Mediated Immune Responses in Tuberculosis. *Annual Review of Immunology*, *27*(1), 393–422. doi: 10.1146/annurev.immunol.021908.132703

Couvin, D., Reynaud, Y., & Rastogi, N. (2019). Two tales: Worldwide distribution of Central Asian (CAS) versus ancestral East-African Indian (EAI) lineages of *Mycobacterium tuberculosis* underlines a remarkable cleavage for phylogeographical, epidemiological and demographical characteristics. *Plos One*, *14*(7), 1–20. doi: 10.1371/journal.pone.0219706

Davis, J. M., & Ramakrishnan, L. (2009). The Role of the Granuloma in Expansion and Dissemination of Early Tuberculous Infection. *Cell*, *136*(1), 37–49. doi: 10.1016/j.cell.2008.11.014

Demay, C., Liens, B., Burguière, T., Hill, V., Couvin, D., Millet, J., … Rastogi, N. (2012). SITVITWEB - A publicly available international multimarker database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. *Infection, Genetics and Evolution*, *12*(4), 755–766. doi: 10.1016/j.meegid.2012.02.004

Desjardins, C. A., Cohen, K. A., Munsamy, V., Abeel, T., Maharaj, K., Walker, B. J., … Pym, A. S. (2016). Genomic and functional analyses of *Mycobacterium tuberculosis* strains implicate ald in D-cycloserine resistance. *Nature Genetics*, *48*(5), 544–551. doi: 10.1038/ng.3548

Dymova, M. A., Kinsht, V. N., Cherednichenko, A. G., Khrapov, E. A., Svistelnik, A. V., & Filipenko, M. L. (2011). Highest prevalence of the *Mycobacterium tuberculosis* Beijing genotype isolates in patients newly diagnosed with tuberculosis in the Novosibirsk oblast, Russian Federation. *Journal of Medical Microbiology*, *60*(7), 1003–1009. doi: 10.1099/jmm.0.027995-0

Duong, D. A., Duyen, N. T. H., Lan, N. T. N., Dai, V. H., Ha, D. T. M., Vo, S. K., … Caws, M. (2009). Beijing genotype of *Mycobacterium tuberculosis* is significantly associated with high-level fluoroquinolone resistance in Vietnam. *Antimicrobial Agents and Chemotherapy*, *53*(11), 4835–4839. doi: 10.1128/AAC.00541-09

Engström, A. (2016). Fighting an old disease with modern tools: Characteristics and molecular detection methods of drug-resistant Mycobacterium tuberculosis. *Infectious Diseases*, *48*(1), 1–17. doi: 10.3109/23744235.2015.1061205

Ernst, J. D. (2012). The immunological life cycle of tuberculosis. *Nature Reviews Immunology*, *12*, 581–591. doi: 10.1038/nri3259

European Center for Disease Prevention and Control. (2019). Tuberculosis surveillance and monitoring in Europe 2019. doi: 10.2900/512553

Faggioni, R., Feingold, K. R., & Grunfeld, C. (2001). Leptin regulation of the immune response and the immunodeficiency of malnutrition. *FASEB Journal*, *15*(14), 2565–2571. doi: 10.1096/fj.01-0431rev

Fenner, L., Egger, M., Bodmer, T., Altpeter, E., Zwahlen, M., Jaton, K., … Gagneux, S. (2012). Effect of mutation and genetic background on drug resistance in *Mycobacterium*

tuberculosis. *Antimicrobial Agents and Chemotherapy*, *56*(6), 3047–3053. doi: 10.1128/AAC.06460-11

Feuerriegel, S., Köser, C., Trübe, L., Archer, J., Gerdes, S. R., Richter, E., & Niemann, S. (2010). Thr202Ala in thyA is a marker for the Latin American mediterranean lineage of the *Mycobacterium tuberculosis* complex rather than Para-aminosalicylic acid resistance. *Antimicrobial Agents and Chemotherapy*, *54*(11), 4794–4798. doi: 10.1128/AAC.00738-10

Guerra-Assunção, J. A., P. E. M., Crampin, A. C., Houben, R. M. G. J., Mzembe, T., Mallard, K., Coll, F., … Glynn, J. R. (2015). Large-scale whole genome sequencing of *M. tuberculosis* provides insights into transmission in a high prevalence area. *eLife*, *4*, 1–17. doi: 10.7554/eLife.05166

Fivian-Hughes, A. S., Houghton, J., & Davis, E. O. (2012). Mycobacterium tuberculosis thymidylate synthase gene thyX is essential and potentially bifunctional, while thyA deletion confers resistance to p-aminosalicylic acid. *Microbiology*, *158*(2), 308–318. doi: 10.1099/mic.0.053983-0

Ford, C. B., Shah, R. R., Maeda, M. K., Gagneux, S., Murray, M. B., Cohen, T., … Fortune, S. M. (2013). *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nature Genetics*, *45*(7), 784–790. doi: 10.1038/ng.2656

Gagneux, S. (2018). Ecology and evolution of Mycobacterium tuberculosis. *Nature Reviews Microbiology*, *16*(4), 202–213. doi: 10.1038/nrmicro.2018.8

Gagneux, S., Burgos, M. V., DeRiemer, K., Enciso, A., Muñoz, S., Hopewell, P. C., … Pym, A. S. (2006). Impact of bacterial genetics on the transmission of isoniazid-resistant *Mycobacterium tuberculosis*. *PLoS Pathogens*, *2*(6), 0603–0610. doi: 10.1371/journal.ppat.0020061

Gagneux, S., Long, C. D., Small, P. M., Van, T., Schoolnik, G. K., & Bohannan, B. J. M. (2006). The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. *Science*, *312*(5782), 1944–1946. doi: 10.1126/science.1124410

Geldmacher, C., Ngwenyama, N., Schuetz, A., Petrovas, C., Reither, K., Heeregrave, E. J., … Koup, R. A. (2010). Preferential infection and depletion of Mycobacterium tuberculosis-specific CD4 T cells after HIV-1 infection. *Journal of Experimental Medicine*, *207*(13), 2869–2881. doi: 10.1084/jem.20100090

Ghebremichael, S., Groenheit, R., Pennhag, A., Koivula, T., Andersson, E., Bruchfeld, J., … Källenius, G. (2010). Drug-resistant *Mycobacterium tuberculosis* of the beijing genotype does not spread in Sweden. *Plos One*, *5*(5). doi: 10.1371/journal.pone.0010893

Gopal, P., Nartey, W., Ragunathan, P., Sarathy, J., Kaya, F., Yee, M., … Dick, T. (2017). Pyrazinoic Acid Inhibits Mycobacterial Coenzyme A Biosynthesis by Binding to Aspartate Decarboxylase PanD. *ACS Infectious Diseases*, *3*(11), 807–819. doi: 10.1021/acsinfecdis.7b00079

Gupta, R., Lavollay, M., Mainardi, J. L., Arthur, M., Bishai, W. R., & Lamichhane, G. (2010). The *Mycobacterium tuberculosis* protein Ldt Mt2 is a nonclassical transpeptidase required for virulence and resistance to amoxicillin. *Nature Medicine*, *16*(4), 466–469. doi: 10.1038/nm.2120

Hanekom, M., Van Der Spuy, G. D., Streicher, E., Ndabambi, S. L., McEvoy, C. R. E., Kidd, M., … Warren, R. M. (2007). A recently evolved sub-lineage of the *Mycobacterium tuberculosis* Beijing strain family is associated with an increased ability to spread and cause disease. *Journal of Clinical Microbiology*, *45*(5), 1483–1490. doi: 10.1128/JCM.02191-06

Hanekom, M., Pittius, N. C. G. Van, Mcevoy, C., Victor, T. C., Helden, P. D. Van, & Warren, R. M. (2011). *Mycobacterium tuberculosis* Beijing genotype: A template for success. *Tuberculosis*, *91*(6), 510–523. doi: 10.1016/j.tube.2011.07.005

Havlir, D. V., & Barnes, P. F. (1999). Tuberculosis in Patients with Human Immunodeficiency Virus Infection. *New England Journal of Medicine*, *340*(5), 367–373. doi: 10.1056/NEJM199902043400507

Hazbón, M. H., Brimacombe, M., Del Valle, M. B., Cavatore, M., Guerrero, M. I., Varma-Basil, M., … Alland, D. (2006). Population genetics study of isoniazid resistance mutations and evolution of multidrug-resistant *Mycobacterium tuberculosis*. *Antimicrobial Agents and Chemotherapy*, *50*(8), 2640–2649. doi: 10.1128/AAC.00112-06

Hegde, S. S., Vetting, M. W., Roderick, S. L., Mitchenall, L. A., Maxwell, A., Takiff, H. E., & Blanchard, J. S. (2005). A fluoroquinolone resistance protein from Mycobacterium tuberculosis that mimics DNA. *Science*, *308*, 1480–1483. doi: 10.1126/science.1110699

Heym, B., Alzari, P. M., Honore, N., & Cole, S. T. (1995). Missense mutations in the catalase-peroxidase gene, *katG*, are associated with isoniazid resistance in *Mycobacterium tuberculosis*. *Molecular Microbiology*, *15*(2), 235–245. doi: 10.1111/j.1365-2958.1995.tb02238.x

Heysell, S. K., & Houpt, E. R. (2012). The future of molecular diagnostics for drug-resistant tuberculosis. *Expert Review of Molecular Diagnostics*, *12*(4), 395–405. doi: 10.1586/erm.12.25

Hershberg, R., Lipatov, M., Small, P. M., Sheffer, H., Niemann, S., Homolka, S., … Gagneux, S. (2008). High functional diversity in Mycobacterium tuberculosis driven by genetic drift and human demography. *PLoS Biology*, *6*(12), 2658–2671. doi: 10.1371/journal.pbio.0060311

Hillemann, D., Kubica, T., Rüsch-Gerdes, S., & Niemann, S. (2005). Disequilibrium in distribution of resistance mutations among *Mycobacterium tuberculosis* Beijing and non-Beijing strains isolated from patients in Germany. *Antimicrobial Agents and Chemotherapy*, *49*(3), 1229–1231. doi: 10.1128/AAC.49.3.1229-1231.2005

Holmberg, S. D. (1990). The rise of tuberculosis in America before 1820. *American Review of Respiratory Disease*, *142*(5), 1228–1232. doi: 10.1164/ajrccm/142.5.1228

Hugonnet, J. E., & Blanchard, J. S. (2007). Irreversible inhibition of the *Mycobacterium tuberculosis* β-lactamase by clavulanate. *Biochemistry*, *46*(43), 11998–12004. doi: 10.1021/bi701506h

Huitric, E., Werngren, J., Juréen, P., & Hoffner, S. (2006). Resistance levels and *rpoB* gene mutations among in vitro-selected rifampin-resistant *Mycobacterium tuberculosis* mutants. *Antimicrobial Agents and Chemotherapy*, *50*(8), 2860–2862. doi: 10.1128/AAC.00303-06

Iwamoto, T., Yoshida, S., Suzuki, K., & Wada, T. (2008). Population structure analysis of the Mycobacterium tuberculosis Beijing family indicates an association between certain sub-lineages and multidrug resistance. *Antimicrobial Agents and Chemotherapy*, *52*(10), 3805–3809. doi: 10.1128/AAC.00579-08

Jajou, R., de Neeling, A., van Hunen, R., de Vries, G., Schimmel, H., Mulder, A., … van Soolingen, D. (2018). Epidemiological links between tuberculosis cases identified twice as efficiently by whole genome sequencing than conventional molecular typing: A population-based study. *Plos One*, 13(4). doi: 10.1371/journal. pone.0195413

Johnsson, K., & Schultz, P. G. (1994). Mechanistic Studies of the Oxidation of Isoniazid by the Catalase Peroxidase from *Mycobacterium tuberculosis*. *Journal of the American Chemical Society*, *116*(16), 7425–7426. doi: 10.1021/ja00095a063

Jones, D. (2002). *Pharmaceutical Statistics*. London, United Kingdom: Pharmaceutical Press

Kamerbeek, J., Schouls, L., Kolk, A., Van Agterveld, M., Van Soolingen, D., Kuijper, S., … Van Embden, J. (1997). Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. *Journal of Clinical Microbiology*, *35*(4), 907–914.

Kato-Maeda, M., Metcalfe, J. Z., & Flores, L. (2011). Genotyping of *Mycobacterium tuberculosis*: application in epidemiologic studies. *Future Microbiology*, *6*(2), 203–216. doi: 10.2217/fmb.10.165

Keshavjee, S., & Farmer, P. E. (2012). Tuberculosis, Drug Resistance, and the History of Modern Medicine. *New England Journal of Medicine*, *367*(10), 931–936. doi: 10.1056/NEJMra1205429

Khan, M. T., Khan, A., Rehman, A. U., Wang, Y., Akhtar, K., Malik, S. I., & Wei, D. Q. (2019). Structural and free energy landscape of novel mutations in ribosomal protein S1 (rpsA) associated with pyrazinamide resistance. *Scientific Reports*, *9*(1), 1–12. doi: 10.1038/s41598-019-44013-9

Kozińska, M., & Augustynowicz-Kopeć, E. (2015). Drug resistance and population structure of *Mycobacterium tuberculosis* Beijing strains isolated in Poland. *Polish Journal of Microbiology*, *64*(4), 399–401. doi: 10.5604/17331331.1185243

Kruczak, K., Duplaga, M., Sanak, M., Cmiel, A., Mastalerz, L., Sladek, K., & Nizankowska-Mogilnicka, E. (2014). Comparison of IGRA tests and TST in the diagnosis of latent tuberculosis infection and predicting tuberculosis in risk groups in Krakow, Poland. *Scandinavian Journal of Infectious Diseases*, *46*(9), 649–655. doi: 10.3109/00365548.2014.927955

Kwan, C., & Ernst, J. D. (2011). HIV and tuberculosis: A deadly human syndemic. *Clinical Microbiology Reviews*, *24*(2), 351–376. doi: 10.1128/CMR.00042-10

Larsen, M. H., Vilchèze, C., Kremer, L., Besra, G. S., Parsons, L., Salfinger, M., … Jacobs, W. R. (2002). Overexpression of inhA, but not kasA, confers resistance to isoniazid and

ethionamide in *Mycobacterium smegmatis*, *M. bovis* BCG and *M. tuberculosis*. *Molecular Microbiology*, *46*(2), 453–466. doi: 10.1046/j.1365-2958.2002.03162.x

Lau, R. W. T., Ho, P. L., Kao, R. Y. T., Yew, W. W., Lau, T. C. K., Cheng, V. C. C., … Yam, W. C. (2011). Molecular characterization of fluoroquinolone resistance in *Mycobacterium tuberculosis*: Functional analysis of *gyrA* mutation at position 74. *Antimicrobial Agents and Chemotherapy*, *55*(2), 608–614. doi: 10.1128/AAC.00920-10

Lawn, S. D. (2015). Advances in diagnostic assays for tuberculosis. *Cold Spring Harbor Perspectives in Medicine*, *5*(12), 1–18. doi: 10.1101/cshperspect.a017806

Li, Q. jing, Jiao, W. wei, Yin, Q. qin, Li, Y. jia, Li, J. qiong, Xu, F., … Shen, A. dong. (2017). Positive epistasis of major low-cost drug resistance mutations rpoB531-TTG and katG315-ACC depends on the phylogenetic background of Mycobacterium tuberculosis strains. *International Journal of Antimicrobial Agents*, *49*(6), 757–762. doi: 10.1016/j.ijantimicag.2017.02.009

Lipworth, S., Jajou, R., De Neeling, A., Bradley, P., Van Der Hoek, W., Maphalala, G., … Van Soolingen, D. (2019). SNP-IT tool for identifying subspecies and associated lineages of Mycobacterium tuberculosis complex. *Emerging Infectious Diseases*, *25*(3), 482–488. doi: 10.3201/eid2503.180894

Liu, Y., Jiang, X., Li, W., Zhang, X., Wang, W., & Li, C. (2017). The study on the association between Beijing genotype family and drug susceptibility phenotypes of *Mycobacterium tuberculosis* in Beijing. *Scientific Reports*, *7*(1), 1–7. doi: 10.1038/s41598-017-14119-z

Luo, T., Comas, I., Luo, D., Lu, B., Wu, J., Wei, L., … Gao, Q. (2015). Southern East Asian origin and coexpansion of *Mycobacterium tuberculosis* Beijing family with Han Chinese. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(26), 8136–8141. doi: 10.1073/pnas.1424063112

Manca, C., Tsenova, L., Bergtold, A., Freeman, S., Tovey, M., Musser, J. M., … Kaplan, G. (2001). Virulence of a Mycobacterium tuberculosis clinical isolate in mice is determined by failure to induce Th1 type immunity and is associated with induction of IFN-α/β. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(10), 5752–5757. doi: 10.1073/pnas.091096998

Macedo, R., Silva, C., Pinto, C., Furtado, C., & Brum, L. (2011). Tuberculosis drug-resistance in Lisbon, Portugal: a 6-year overview. *Clinical Microbiology and Infection*, *17*(9), 1397–1402. doi: 10.1111/j.1469-0691.2010.03351.x

Machado, D., Perdigão, J., Ramos, J., Couto, I., Portugal, I., Ritter, C., … Viveiros, M. (2013). High-level resistance to isoniazid and ethionamide in multidrug-resistant *Mycobacterium tuberculosis* of the Lisboa family is associated with *inhA* double mutations. *Journal of Antimicrobial Chemotherapy*, *68*(8), 1728–1732. doi: 10.1093/jac/dkt090

Maeda, S., Hang, N. T. L., Lien, L. T., Thuong, P. H., Hung, N. V., Hoang, N. P., … Keicho, N. (2014). Mycobacterium tuberculosis strains spreading in Hanoi, Vietnam: Beijing sub-lineages, genotypes, drug susceptibility patterns, and host factors. *Tuberculosis*, *94*(6), 649–656. doi: 10.1016/j.tube.2014.09.005

Mai, T. Q., Van Anh, N. T., Hien, N. T., Lan, N. H., Giang, D. C., Hang, P. T. T., … Sintchenko, V. (2017). Drug resistance and *Mycobacterium tuberculosis* strain diversity in TB/HIV co-infected patients in Ho Chi Minh city, Vietnam. *Journal of Global Antimicrobial Resistance*, *10*, 154–160. doi: 10.1016/j.jgar.2017.07.003

Mazars, E., Lesjean, S., Banuls, A. L., Gilbert, M., Vincent, V., Gicquel, B., … Supply, P. (2001). High-resolution minisatellite-based typing as a portable approach to global analysis of Mycobacterium tuberculosis molecular epidemiology. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(4), 1901–1906. doi: 10.1073/pnas.98.4.1901

Meehan, C. J., Goig, G. A., Kohl, T. A., Verboven, L., Dippenaar, A., Ezewudo, M., … Van Rie, A. (2019). Whole genome sequencing of *Mycobacterium tuberculosis*: current standards and open issues. *Nature Reviews Microbiology*, *17*(9), 533–545. doi: 10.1038/s41579-019-0214-5

Merker, M., Blin, C., Mona, S., Duforet-frebourg, N., Lecher, S., Willery, E., … Ballif, M. (2015). Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nature Genetics*, *47*(3), 242–249. doi: 10.1038/ng.3195

Miotto, P., Cabibbe, A. M., Mantegani, P., Borroni, E., Fattorini, L., Tortoli, E., … Cirillo, D. M. (2012). GenoType MTBDRsl performance on clinical samples with diverse genetic background. *European Respiratory Journal*, *40*(3), 690–698. doi: 10.1183/09031936.00164111

Mitchison, D. A., & Nunn, A. J. (1986). Influence of initial drug resistance on the response to short-course chemotherapy of pulmonary tuberculosis. *American Review of Respiratory Disease*, *133*(3), 423–430. doi: 10.1164/arrd.1986.133.3.423

Mogues, T., Goodrich, M. E., Ryan, L., LaCourse, R., & North, R. J. (2001). The relative importance of T cell subsets in immunity and immunopathology of airborne Mycobacterium tuberculosis infection in mice. *Journal of Experimental Medicine*, *193*(3), 271–280. doi: 10.1084/jem.193.3.271

Molodtsov, V., Scharf, N. T., Stefan, M. A., Garcia, G. A., & Murakami, K. S. (2017). Structural basis for rifamycin resistance of bacterial RNA polymerase by the three most clinically important RpoB mutations found in *Mycobacterium tuberculosis*. *Molecular Microbiology*, *103*(6), 1034–1045. doi: 10.1111/mmi.13606

Mokrousov, I. (2012). The quiet and controversial: Ural family of *Mycobacterium tuberculosis*. *Infection, Genetics and Evolution*, *12*(4), 619–629. doi: 10.1016/j.meegid.2011.09.026

Mokrousov, I., Jiao, W. W., Sun, G. Z., Liu, J. W., Valcheva, V., Li, M., … Dong Shen, A. (2006). Evolution of drug resistance in different sub-lineages of *Mycobacterium tuberculosis* Beijing genotype. *Antimicrobial Agents and Chemotherapy*, *50*(8), 2820–2823. doi: 10.1128/AAC.00324-06

Moradigaravand, D., Grandjean, L., Martinez, E., Li, H., Zheng, J., Coronel, J., & Moore, D. (2016). *dfrA thyA* Double Deletion in para-Aminosalicylic Acid-Resistant. *Antimicrobial Agents and Chemotherapy*, *60*(6), 4–7. doi: 10.1128/AAC.00253-16

Mortimer, T. D., Weber, A. M., & Pepperell, C. S. (2018). Signatures of Selection at Drug Resistance Loci in *Mycobacterium tuberculosis*. *mSystems*, *3*(1), 1–16. doi: 10.1128/mSystems.00108-17

Mdluli, K., Slayden, R. A., Zhu, Y. Q., Ramaswamy, S., Pan, X., Mead, D., … Barry, C. E. (1998). Inhibition of a *Mycobacterium tuberculosis* β-Ketoacyl ACP synthase by isoniazid. *Science*, *280*(5369), 1607–1610. doi: 10.1126/science.280.5369.1607

Muyoyeta, M., Schaap, J. A., De Haas, P., Mwanza, W., Muvwimi, M. W., Godfrey-Faussett, P., & Ayles, H. (2009). Comparison of four culture systems for Mycobacterium tuberculosis in the zambian national reference laboratory. *International Journal of Tuberculosis and Lung Disease*, *13*(4), 460–465.

Nakatani, Y., Opel-Reading, H. K., Merker, M., Machado, D., Andres, S., Kumar, S. S., … Köser, C. U. (2017). Role of Alanine Racemase Mutations in Mycobacterium tuberculosis D-Cycloserine Resistance. *Antimicrobial Agents and Chemotherapy*, *61*(12), 1–5. doi: 10.1128/AAC.01575-17

Nasiri, M. J., Haeili, M., Ghazi, M., Goudarzi, H., Pormohammad, A., Fooladi, A. A. I., & Feizabadi, M. M. (2017). New insights in to the intrinsic and acquired drug resistance mechanisms in mycobacteria. *Frontiers in Microbiology*, *8*. doi: 10.3389/fmicb.2017.00681

Nguyen, V. A. T., Bañuls, A. L., Tran, T. H. T., Pham, K. L. T., Nguyen, T. S., Nguyen, H. Van, … Choisy, M. (2016). *Mycobacterium tuberculosis* lineages and anti-tuberculosis drug resistance in reference hospitals across Viet Nam. *BMC Microbiology*, *16*(1), 1–9. doi: 10.1186/s12866-016-0784-6

Niemann, S., & Supply, P. (2014). Diversity and evolution of *Mycobacterium tuberculosis*: Moving to whole genome-based approaches. *Cold Spring Harbor Perspectives in Medicine*, *4*(12). doi: 10.1101/cshperspect.a021188

Palomino, J. C., & Martin, A. (2014). Drug Resistance Mechanisms in *Mycobacterium tuberculosis*. *Antibiotics*, *3*, 317–340. doi: 10.3390/antibiotics3030317

Park, Y. K., Shin, S., Ryu, S., Sang, N. C., Koh, W. J., O, J. K., … Gill, H. B. (2005). Comparison of drug resistance genotypes between Beijing and non-Beijing family strains of *Mycobacterium tuberculosis* in Korea. *Journal of Microbiological Methods*, *63*(2), 165–172. doi: 10.1016/j.mimet.2005.03.002

Perdigão, J., Macedo, R., Fernandes, E., Brum, L. & Portugal, I. (2008). Multidrug-Resistant Tuberculosis in Lisbon, Portugal: A Molecular Epidemiological Perspective *Microbial Drug Resistance*, *14*(2), 133–143. doi: 10.1089/mdr.2008.0798

Perdigão, J., Macedo, R., Machado, D., Silva, C., Jordão, L., Couto, I., … Portugal, I. (2014). GidB mutation as a phylogenetic marker for Q1 cluster *Mycobacterium tuberculosis* isolates and intermediate-level streptomycin resistance determinant in Lisbon, Portugal. *Clinical Microbiology and Infection*, *20*(5), 278–284. doi: 10.1111/1469-0691.12392

Perdigão, J., Macedo, R., Malaquias, A., Ferreira, A., Brum, L., & Portugal, I. (2010). Genetic analysis of extensively drug-resistant Mycobacterium tuberculosis strains in Lisbon, Portugal. *Journal of Antimicrobial Chemotherapy*, *65*(2), 224–227. doi: 10.1093/jac/dkp452

Perdigão, J., Macedo, R., Ribeiro, A., & Brum, L. (2009). Genetic characterization of the ethambutol resistance-determining region in Mycobacterium tuberculosis: prevalence and significance of *embB*306 mutations. *International Journal of Antimicrobial Agents*, *33*, 334–338. doi: 10.1016/j.ijantimicag.2008.09.021

Perdigão, J., Macedo, R., Silva, C., Machado, D., Couto, I., Viveiros, M., … Portugal, I. (2013). From multidrug-resistant to extensively drug-resistant tuberculosis in Lisbon, Portugal: The stepwise mode of resistance acquisition. *Journal of Antimicrobial Chemotherapy*, *68*(1), 27–33. doi: 10.1093/jac/dks371

Perdigão, J., Maltez, F., Machado, D., Silva, H., Pereira, C., Silva, C., … Portugal, I. (2016). Beyond extensively drug-resistant tuberculosis in Lisbon, Portugal: a case of linezolid resistance acquisition presenting as an iliopsoas abscess. *International Journal of Antimicrobial Agents*, *48*(5), 569–570. doi: 10.1016/j.ijantimicag.2016.07.026

Perdigão, J., & Portugal, I. (2019). Genetics and roadblocks of drug-resistant tuberculosis. *Infection, Genetics and Evolution*, *72*, 113–130. doi: 10.1016/j.meegid.2018.09.023

Perdigão, J., Silva, H., Machado, D., Macedo, R., Maltez, F., Silva, C., … Viveiros, M. (2014). Unraveling *Mycobacterium tuberculosis* genomic diversity and evolution in Lisbon, Portugal, a highly drug-resistant setting. *BMC Genomics*, *15*, 1–20. doi: 10.1186/1471-2164-15-991

Pfyffer, G. E., Bonato, D. A., Ebrahimzadeh, A., Gross, W., Hotaling, J., Kornblum, J., … Siddiqi, S. (1999). Multicenter laboratory validation of susceptibility testing of Mycobacterium tuberculosis against classical second-line and newer antimicrobial drugs by using the radiometric BACTEC 460 technique and the proportion method with solid media. *Journal of Clinical Microbiology*, *37*(10), 3179–3186.

Portugal, I., Barreiro, L., Moniz-Pereira, J., & Brum, L. (2004). *pncA* Mutations in Pyrazinamide-Resistant *Mycobacterium tuberculosis* Isolates in Portugal. *Antimicrobial Agents and Chemotherapy*, *48*(7), 2736–2738. doi: 10.1128/AAC.48.7.2736

Ramakrishnan, L. (2012). Revisiting the role of the granuloma in tuberculosis. *Nature Reviews Immunology*, *12*(5), 352–366. doi: 10.1038/nri3211

Ramirez-Busby, S. M., Rodwell, T. C., Fink, L., Catanzaro, D., Jackson, R. L., Pettigrove, M., … Valafar, F. (2017). A Multinational Analysis of Mutations and Heterogeneity in PZase, RpsA, and PanD Associated with Pyrazinamide Resistance in M/XDR *Mycobacterium tuberculosis*. *Scientific Reports*, *7*(1), 1–9. doi: 10.1038/s41598-017-03452-y

Reed, M. B., Domenech, P., Manca, C., Su, H., Barczak, A. K., Kreiswirth, B. N., & Kaplan, G. (2004). A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. *Nature*, *431*(7004), 84–87. doi: 10.1038/nature02837

Reed, M. B., Gagneux, S., DeRiemer, K., Small, P. M., & Barry, C. E. (2007). The W-Beijing lineage of *Mycobacterium tuberculosis* overproduces triglycerides and has the DosR dormancy regulon constitutively upregulated. *Journal of Bacteriology*, *189*(7), 2583–2589. doi: 10.1128/JB.01670-06

Ribeiro, S. C. M., Gomes, L. L., Amaral, E. P., Andrade, M. R. M., Almeida, F. M., Rezende, A. L., … Lasunskaia, E. B. (2014). *Mycobacterium tuberculosis* strains of the modern sub-lineage of the beijing family are more likely to display increased virulence than strains of the ancient sub-lineage. *Journal of Clinical Microbiology*, *52*(7), 2615–2624. doi: 10.1128/JCM.00498-14

Rigouts, L., Coeck, N., Gumusboga, M., Rijk, W. B. De, Aung, K. J. M., Hossain, M. A., … Van Deun, A. (2016). Specific *gyrA* gene mutations predict poor treatment outcome in MDR-TB. *Journal of Antimicrobial Chemotherapy*, *71*(2), 314–323. doi: 10.1093/jac/dkv360

Rindi, L., Peroni, I., Lari, N., Bonanni, D., Tortoli, E., & Garzelli, C. (2007). Variation of the expression of *Mycobacterium tuberculosis* ppe44 gene among clinical isolates. *FEMS Immunology and Medical Microbiology*, *51*(2), 381–387. doi: 10.1111/j.1574-695X.2007.00315.x

Reiley, W. W., Calayag, M. D., Wittmer, S. T., Huntington, J. L., Pearl, J. E., Fountain, J. J., … Woodland, D. L. (2008). ESAT-6-specific CD4 T cell responses to aerosol Mycobacterium tuberculosis infection are initiated in the mediastinal lymph nodes. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(31), 10961–10966. doi: 10.1073/pnas.0801496105

Romagnoli, A., Petruccioli, E., Palucci, I., Camassa, S., Carata, E., Petrone, L., … Fimia, G. M. (2018). Clinical isolates of the modern *Mycobacterium tuberculosis* lineage 4 evade host defense in human macrophages through eluding IL-1β-induced autophagy article. *Cell Death and Disease*, *9*(6), 1–12. doi: 10.1038/s41419-018-0640-8

Russel, D. G. (2001). *Mycobacterium tuberculosis*: here today, and here tomorrow. *Nature Reviews Molecular Cell Biology*, *2*, 569–578. doi: 10.1038/35085034

Saelens, J. W., Viswanathan, G., & Tobin, D. M. (2019). Mycobacterial evolution intersects with host tolerance. *Frontiers in Immunology*, *10*, 1–14. doi: 10.3389/fimmu.2019.00528

Schaible, U. E., & Kaufmann, S. H. E. (2007). Malnutrition and infection: Complex mechanisms and global impacts. *Plos Medicine*, *4*(5), 0806–0812. doi: 10.1371/journal.pmed.0040115

Schön, T., Juréen, P., Chryssanthou, E., Giske, C. G., Kahlmeter, G., Hoffner, S., & Ängeby, K. (2013). Rifampicin-resistant and rifabutin-susceptible *Mycobacterium tuberculosis* strains: A breakpoint artefact? *Journal of Antimicrobial Chemotherapy*, *68*(9), 2074–2077. doi: 10.1093/jac/dkt150

Shafiani, S., Tucker-Heard, G., Kariyone, A., Takatsu, K., & Urdahl, K. B. (2010). Pathogen-specific regulatory T cells delay the arrival of effector T cells in the lung during early tuberculosis. *Journal of Experimental Medicine*, *207*(7), 1409–1420. doi: 10.1084/jem.20091885

Shah, N. S., Auld, S. C., Brust, J. C. M., Mathema, B., Ismail, N., Moodley, P., … Gandhi, N. R. (2017). Transmission of extensively drug-resistant tuberculosis in South Africa. *New England Journal of Medicine*, *376*(3), 243–253. doi: 10.1056/NEJMoa1604544

Shanmugam, S., Selvakumar, N., & Narayanan, S. (2011). Drug resistance among different genotypes of Mycobacterium tuberculosis isolated from patients from Tiruvallur, South India. *Infection, Genetics and Evolution*, *11*(5), 980–986. doi: 10.1016/j.meegid.2011.03.011

Shi, W., Zhang, X., Jiang, X., Yuan, H., Lee, J. S., Barry, C. E., … Zhang, Y. (2011). Pyrazinamide inhibits trans-translation in *Mycobacterium tuberculosis*. *Science*, *333*(6049), 1630–1632. doi: 10.1126/science.1208813

Sirgel, F. A., Warren, R. M., Streicher, E. M., Victor, T. C., Van helden, P. D., & Böttger, E. C. (2012). *gyrA* mutations and phenotypic susceptibility levels to OFXoxacin and moxifloxacin in clinical isolates of *Mycobacterium tuberculosis*. *Journal of Antimicrobial Chemotherapy*, *67*(5), 1088–1093. doi: 10.1093/jac/dks033

Smith, T., Wolff, K. A., & Nguyen, L. (2012). Molecular Biology of Drug Resistance in *Mycobacterium tuberculosis*. *Current Topics in Microbiology and Immunology*, *374*, 53–80. doi: 10.1007/82_2012_279

Smittipat, N., Juthayothin, T., Billamas, P., Jaitrong, S., Rukseree, K., Dokladda, K., … Palittapongarnpim, P. (2016). Mutations in *rrs*, *rpsL* and *gidB* in streptomycin-resistant *Mycobacterium tuberculosis* isolates from Thailand. *Journal of Global Antimicrobial Resistance*, *4*, 5–10. doi: 10.1016/j.jgar.2015.11.009

Spies, F. S., Ribeiro, A. W., Ramos, D. F., Ribeiro, M. O., Martin, A., Palomino, J. C., … Zaha, A. (2011). Streptomycin resistance and lineage-specific polymorphisms in *Mycobacterium tuberculosis gidB* gene. *Journal of Clinical Microbiology*, *49*(7), 2625–2630. doi: 10.1128/JCM.00168-11

Sreevatsan, S., Pan, X., Stockbauer, K. E., Connell, N. D., Kreiswirth, B. N., Whittam, T. S., & Musser, J. M. (1997). Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(18), 9869–9874. doi: 10.1073/pnas.94.18.9869

Sreevatsan, S., Stockbauer, K. E., Pan, X., Kreiswirth, B. N., Moghazeh, S. L., Jacobs, W. R., … Musser, J. M. (1997). Ethambutol resistance in Mycobacterium tuberculosis: Critical role of embB mutations. *Antimicrobial Agents and Chemotherapy*, *41*(8), 1677–1681. doi: 10.1128/AAC.41.8.1677

de Steenwinkel, J. E. M., ten Kate, M. T., de Knegt, G. J., Kremer, K., Aarnoutse, R. E., Boeree, M. J., … Bakker-Woudenberg, I. A. J. M. 2012. Drug Susceptibility of *Mycobacterium tuberculosis* Beijing Genotype and Association with MDR TB. *Emerging Infectious Diseases*, *18*(4), 660–663. doi: 1 0.3201/eid1804.110912

Stephan, J., Mailaender, C., Etienne, G., Daffe, M., & Niederweis, M. (2004). Multidrug Resistance of a Porin Deletion Mutant of *Mycobacterium smegmatis*. *Antimicrobial Agents and Chemotherapy*, *48*(11), 4163–4170. doi: 10.1128/AAC.48.11.4163

Stucki, D., Brites, D., Jeljeli, L., Coscolla, M., Liu, Q., Trauner, A., … Gagneux, S. (2016). *Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sub-lineages. *Nature Genetics*, *48*(12), 1535–1543. doi: 10.1038/ng.3704

Sun, Z., Li, W., Xu, S., & Huang, H. (2015). The discovery, function and development of the variable number tandem repeats in different Mycobacterium species. *Critical Reviews in Microbiology*, *42*(5), 738–758. doi: 10.3109/1040841X.2015.1022506

Sun, Y. J., Luo, J. T., Wong, S. Y., & Lee, A. S. G. (2010). Analysis of *rpsL* and *rrs* mutations in Beijing and non-Beijing streptomycin-resistant *Mycobacterium tuberculosis* isolates from Singapore. *Clinical Microbiology and Infection*, *16*(3), 287–289. doi: 10.1111/j.1469-0691.2009.02800.x

Sun, H., Zhang, C., Xiang, L., Pi, R., Guo, Z., Zheng, C., … Sun, Q. (2016). Characterization of mutations in streptomycin-resistant *Mycobacterium tuberculosis* isolates in Sichuan, China and the association between Beijing-lineage and dual-mutation in *gidB*. *Tuberculosis*, *96*, 102–106. doi: 10.1016/j.tube.2015.09.004

Supply, P., Mazars, E., Lesjean, S., Vincent, V., Gicquel, B., & Locht, C. (2000). Variable human minisatellite-like regions in the Mycobacterium tuberculosis genome. *Molecular Microbiology*, *36*(3), 762–771. doi: 10.1046/j.1365-2958.2000.01905.x

Supply, P., Marceau, M., Mangenot, S., Roche, D., Rouanet, C., Khanna, V., … Brosch, R. (2013). Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of Mycobacterium tuberculosis. *Nature Genetics*, *45*(2), 172–179. doi: 10.1038/ng.2517

Takayama, K., & Kilburn, J. O. (1989). Inhibition of synthesis of arabinogalactan by ethambutol in *Mycobacterium smegmatis*. *Antimicrobial Agents and Chemotherapy*, *33*(9), 1493–1499. doi: 10.1128/AAC.33.9.1493

Tan, Y., Su, B., Zheng, H., Song, Y., Wang, Y., & Pang, Y. (2017). Molecular characterization of prothionamide-resistant *Mycobacterium tuberculosis* isolates in Southern China. *Frontiers in Microbiology*, *8*, 1–8. doi: 10.3389/fmicb.2017.02358

The CRyPTIC Consortium and the 100,000 Genomes Project (2018). Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. *New England Journal of Medicine*, *379*, 1403-1415. doi: 10.1056/NEJMoa1800474

Theus, S., Eisenach, K., Fomukong, N., Silver, R. F., & Cave, M. D. (2007). Beijing family *Mycobacterium tuberculosis* strains differ in their intracellular growth in THP-1 macrophages. *International Journal of Tuberculosis and Lung Disease*, *11*(10), 1087–1093.

Thierry, D., Cave, M. D., Eisenach, K. D., Crawford, J. T., Bates, J. H., Gicquel, B., … Guesdon, J. L. (1990). IS6110, an IS-like element of *Mycobacterium tuberculosis* complex. *Nucleic Acids Research*, *18*(1), 188. doi: 10.1093/nar/18.1.188

Toungoussova, O. S., Caugant, D. A., Sandven, P., Mariandyshev, A. O., & Bjune, G. (2004). Impact of drug resistance on fitness of *Mycobacterium tuberculosis* strains of the W-Beijing genotype. *FEMS Immunology and Medical Microbiology*, *42*(3), 281–290. doi: 10.1016/j.femsim.2004.05.012

Tremblay, L. W., Fan, F., & Blanchard, J. S. (2010). Biochemical and structural characterization of *Mycobacterium tuberculosis* β-lactamase with the carbapenems ertapenem and doripenem. *Biochemistry*, *49*(17), 3766–3773. doi: 10.1021/bi100232q

Vallerskog, T., Martens, G. W., & Kornfeld, H. (2010). Diabetic Mice Display a Delayed Adaptive Immune Response to Mycobacterium tuberculosis. *The Journal of Immunology*, *184*(11), 6275–6282. doi: 10.4049/jimmunol.1000304

van Embden, J. D. A., Cave, M. D., Crawford, J. T., Dale, J. W., Eisenach, K. D., Gicquel, B., … Small, P. M. (1993). Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: Recommendations for a standardized methodology. *Journal of Clinical Microbiology*, *31*(2), 406–409.

Vilchèze, C., & Jacobs JR., W. R. (2014). Resistance to Isoniazid and Ethionamide in *Mycobacterium tuberculosis*: Genes, Mutations, and Causalities. *Microbiology Spectrum*, *2*(4), 1–21. doi: 10.1128/microbiolspec.mgm2-0014-2013

Vilchèze, C., & Jacobs, W. R. Jr. (2007). The Mechanism of Isoniazid Killing: Clarity Through the Scope of Genetics. *Annual Review of Microbiology*, *61*(1), 35–50. doi: 10.1146/annurev.micro.61.111606.122346

Walker, T. M., Ip, C. L. C., Harrell, R. H., Evans, J. T., Kapatai, G., Dedicoat, M. J., … Peto, T. E. A. (2013). Whole genome sequencing to delineate Mycobacterium tuberculosis outbreaks: A retrospective observational study. *The Lancet Infectious Diseases*, *13*(2), 137–146. doi: 10.1016/S1473-3099(12)70277-3

Wang, X. H., Ma, A. G., Han, X. X., Gu, X. M., Fu, L. P., Li, P. G., … Wang, L. J. (2015). Correlations between drug resistance of beijing/W lineage clinical isolates of *Mycobacterium tuberculosis* and sub-lineages: A 2009–2013 prospective study in Xinjiang province, China. *Medical Science Monitor*, *21*, 1313–1318. doi: 10.12659/MSM.892951

Wayne, L. G., & Hayes, L. G. (1996). An in vitro model for sequential study of shiftdown of Mycobacterium tuberculosis through two stages of nonreplicating persistence. *Infection and Immunity*, *64*(6), 2062–2069.

Wei, J., Dahl, J. L., Moulder, J. W., Roberts, E. A., O'Gaora, P., Young, D. B., & Friedman, R. L. (2000). Identification of a Mycobacterium tuberculosis gene that enhances mycobacterial survival in macrophages. *Journal of Bacteriology*, *182*(2), 377–384. doi: 10.1128/JB.182.2.377-384.2000

Wyllie, D. H., Davidson, J. A., Grace Smith, E., Rathod, P., Crook, D. W., Peto, T. E. A., … Campbell, C. (2018). A Quantitative Evaluation of MIRU-VNTR Typing Against Whole genome Sequencing for Identifying *Mycobacterium tuberculosis* Transmission: A Prospective Observational Cohort Study. *EBioMedicine*, *34*, 122–130. doi: 10.1016/j.ebiom.2018.07.019

Wilson, T. M., & Collins, D. M. (1996). ahpC, a gene involved in isoniazid resistance of the Mycobacterium tuberculosis complex. *Molecular Microbiology*, *19*(5), 1025–1034. doi: 10.1046/j.1365-2958.1996.449980.x

Wong, S. Y., Lee, J. S., Kwak, H. K., Via, L. E., Boshoff, H. I. M., & Barry, C. E. (2011). Mutations in *gidB* confer low-level streptomycin resistance in *Mycobacterium tuberculosis*. *Antimicrobial Agents and Chemotherapy*, *55*(6), 2515–2522. doi: 10.1128/AAC.01814-10

World Health Organization. (2014). Antimicrobial resistance: global report on surveillance.

World Health Organization. (2018). The use of next-generation sequencing technologies for the detection of mutations associated with drug resistance in *Mycobacterium tuberculosis* complex: technical guide.

World Health Organization. (2018). Technical Report on critical concentrations for drug susceptibility testing of medicines used in the treatment of drug-resistant tuberculosis.

World Health Organization. (2019). Global tuberculosis report 2019.

Yuen, C. M., Kurbatova, E. V., Click, E. S., Cavanaugh, J. S., & Cegielski, J. P. (2013). Association between *Mycobacterium tuberculosis* complex phylogenetic lineage and acquired drug resistance. *Plos One*, *8*(12), 8–16. doi: 10.1371/journal.pone.0083006

Zaunbrecher, M. A., Sikes, R. D., Metchock, B., Shinnick, T. M., & Posey, J. E. (2009). Overexpression of the chromosomally encoded aminoglycoside acetyltransferase eis confers kanamycin resistance in Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(47), 20004–20009. doi: 10.1073/pnas.0907925106

Zhang, Y., & Mitchison, D. (2003). The curious characteristics of pyrazinamide: A review. *International Journal of Tuberculosis and Lung Disease*, *7*(1), 6–21.

Zhang, Y., Wade, M. M., Scorpio, A., Zhang, H., & Sun, Z. (2003). Mode of action of pyrazinamide: Disruption of Mycobacterium tuberculosis membrane transport and energetics by pyrazinoic acid. *Journal of Antimicrobial Chemotherapy*, *52*(5), 790–795. doi: 10.1093/jac/dkg446

Zhang, S., Chen, J., Shi, W., Liu, W., Zhang, W., & Zhang, Y. (2013). Mutations in panD encoding aspartate decarboxylase are associated with pyrazinamide resistance in *Mycobacterium tuberculosis*. *Emerging Microbes and Infections*, *2*. doi: 10.1038/emi.2013.38

Zhang, S., Chen, J., Cui, P., Shi, W., Zhang, W., & Zhang, Y. (2015). Identification of novel mutations associated with clofazimine resistance in *Mycobacterium tuberculosis*. *Journal of Antimicrobial Chemotherapy*, *70*, 2507–2510. doi: 10.1093/jac/dkv150

Zhang, Z., Pang, Y., Wang, Y., Liu, C., & Zhao, Y. (2014). Beijing genotype of Mycobacterium tuberculosis is significantly associated with linezolid resistance in multidrug-resistant and extensively drug-resistant tuberculosis in China. *International Journal of Antimicrobial Agents*, *43*(3), 231–235. doi: 10.1016/j.ijantimicag.2013.12.007

Zhang, M., Gong, J., Yang, Z., Samten, B., Cave, M. D., & Barnes, P. F. (1999). Enhanced Capacity of a Widespread Strain of Mycobacterium tuberculosis to Grow in Human Macrophages. *The Journal of Infectious Diseases*, *179*(5), 1213–1217. doi: 10.1086/314738

Zhao, F., Wang, X., Erber, L. N., Luo, M., Guo, A., Yang, S., … Gao, Y. (2014). Binding Pocket Alterations in Dihydrofolate Synthase Confer Resistance to para-Aminosalicylic Acid in Clinical Isolates of *Mycobacterium tuberculosis*. *Antimicrobial Agents and Chemotherapy 58*(3), 1479–1487. doi: 10.1128/AAC.01775-13

Zhou, Y., Van Den Hof, S., Wang, S., Pang, Y., Zhao, B., Xia, H., … Van Soolingen, D. (2017). Association between genotype and drug resistance profiles of *Mycobacterium tuberculosis* strains circulating in China in a national drug resistance survey. *Plos One*, *12*(3), 1–13. doi: 10.1371/journal.pone.0174197

Zignol, M., Dean, A. S., Falzon, D., van Gemert, W., Wright, A., van Deun, A., Portaels, F., Laszlo, A., Espinal, M. A., Pablos-Méndez, A., Bloom, A., Aziz, M. A., Weyer, K., Jaramillo, E., Nunn, P., Floyd, K., & Raviglione, M. C. (2016). Twenty Years of Global Surveillance of Antituberculosis-Drug Resistance. *New England Journal of Medicine*, *375*(11), 1081–1089. doi: 10.1056/NEJMsr1512438

Zumla, A., Raviglione, M., Hafner, R., & Fordham von Reyn, C. (2013). Tuberculosis. *New England Journal of Medicine*, *368*(8), 745–755. doi: 10.1056/NEJMra1200894

Annexes

Table A.1. Revised mutation database.

Drug	Gene	Mutation
INH	*ahpC*	C-81T, G-74A, C-72T, G-66A, C-57T, C-54T, C-52A, C-52T, G-51T, T-49G, G-48A, G-48T, T-44A, C-39T, C-20T, G-5A, A-4G, Pro2Ser, Thr5Ile, Phe10Ile, Asp33Asn, Asp73His, Glu76Lys, Leu191Arg
	inhA	A-92T, G-67C, C-34T, G-24T, G-17T, A-16G, C-15T, A-11T, T-8A, T-8C, T-8G, T-5A, Lys8Asn, Leu11Val, Ile16Thr, Ile21Val, Ile21Asn, Ile21Thr, Gly40Trp, Ile47Thr, Val78Ala, Ser94Ala, Ser94Leu, Ser94Trp, Ile95Thr, Ile194Thr
	kasA	Asp66Asn, Met77Ile, Arg121Lys, Gly269Ser, Gly312Ser, Gly387Asp, Phe413Leu
	katG	G-12A, A-10C, G-7A, Met1Ala, Met1Leu, Pro2Ser, Thr11Ala, Thr12Pro, Ser17Asn, Gly19Asp, Asn35Asp, 37del8bp, Trp38Ter, Leu48Gln, Ala61Thr, Asp63Glu, Ala65Thr, Ala66Pro, Ile71Asn, Asp72Gly, Asp74Gly, Asp74Tyr, Met84Ile, Thr85Pro, Gln88Arg, Trp90Ter, Trp91Arg, Asp94Gly, Asp94Ala, Gly96Cys, Gly99Glu, 100insT, Arg104Leu, Arg104Gln, Ala106Val, Trp107Ter, Trp107Arg, His108Gln, His108Asp, Ala109Val, Ala110Val, 111delC, Asp117Ala, Gly121Val, Gly121Cys, Gly125Cys, Met126Ile, Gln127Pro, Arg128Pro, Asn138Ser, Asn138Thr, Asn138Asp, Asn138His, Ala139Pro, Ser140Arg, Ser140Asn, Ser140Gly, Leu141Phe, Asp142Ala, Lys143Thr, Leu148Arg, 151delT, Tyr155Cys, Tyr155Ser, Ser160Leu, Ala162Thr, 168del15bp, Gly169Ala, Ala172Val, Ala172Thr, Met176Ile, Thr180Lys, Gly186Val, 187insG, Asp189Ala, Asp189Gly, Trp191Gly, Trp191Arg, Glu195Lys, 197ins4bp, Trp198Ter, Lys200Ter, Asn218Lys, Gln224Glu, Tyr229Phe, Val230Ala, Gly234Glu, Gly234Arg, Asn236Thr, Asn238Lys, Pro241Pro, Ala243Ser, Arg249Cys, Thr251Met, Phe252Leu, Met257Ile, Met257Thr, Asn258Ser, Asp259Glu, Glu261Lys, Thr262Arg, Ala264Thr, Thr275Ser, Thr275Ala, Thr275Pro, Gly279Asp, Ala281Val, Gly285Asp, Glu289Asp, Ala291Pro, Gln295Pro, Gly297Val, Gly299Ala, Gly299Cys, Trp300Ter, Trp300Gly, Ser302Arg, Tyr304Ser, Gly305Ala, Gly307Glu, Gly307Ala, Gly307Arg, Thr308Pro, Gly309Asp, Gly309Cys, Gly309Ser, Asp311Glu, Asp311Gly, Asp311Tyr, Ser315Arg, Ser315Ile, Ser315Asn, Ser315Thr, Ser315Gly, Gly316Asp, Gly316Ser, Ile317Val, Ile317Leu, Trp321Cys, Trp321Ter, Trp321Leu, Trp321Ser, Trp321Arg, Thr324Pro, Thr326Met, Trp328Cys, Trp328Leu, Trp328Ser, Trp328Gly, Ser331Cys, Ile335Thr, Ile335Val, Leu336Arg

Drug	Gene	Mutation (Cont.)
INH	*katG*	Tyr337Phe, Tyr337Cys, Trp341Ser, Lys345Thr, Ala350Ser, Ala350Thr, Gln352Ter, Asp357Asn, Asp357His, Ala361Asp, Pro365Leu, 371del12bp, 374insC, 375ins3bp, Leu378Pro, Ala379Val, Thr380Ile, Asp381Gly, Leu384Arg, Ile393Asn, Thr394Ala, Trp397Ter, Ala409Asp, Tyr413His, Lys414Asn, Arg418Leu, Arg418Gln, Asp419His, Ala424Val, Ala424Glu, Ala424Gly, Gly428Arg, Pro429Ser, Gln434Ter, Thr435Ile, Ala444Thr, Leu449Phe, Ser457Ile, Lys459Ter, Gln461Pro, Ile462Thr, Arg463Trp, Ala464Ser, Val473Phe, Trp477Ter, Arg484Ser, Gly485Val, Lys488Asn, Gly491Cys, Arg496Leu, Gln502His, Trp505Ser, Trp505Arg, Arg515Cys, Gln525Pro, Asn529Asp, Ala550Asp, Ala551Gly, 563delC, Phe567Ser, Asp573Asn, Ala574Val, 574delC, Ala574Glu, Leu587Met, Leu587Pro, Pro589Thr, Gly593Asp, Glu607Lys, Met609Ile, Leu619Pro, Gly629Ser, Leu634Phe, Ala636Glu, Leu653Pro, Gly685Arg, Asp695Ala, Ser700Pro, Val710Ala, Ala713Pro, Ala716Pro, Gln717Pro, Ala727Asp, Trp728Cys, Asp735Ala, Asp735Asn, 797delC, 805delC, 903insA, 930delT, 1039delC, 1093delA, 1172ins3bp, 1205delG, 1331insT, 1561ins64bp, 1669delG, 1851insGT, 2207del3bp
RIF	*rpoB*	Pro45Ser, Pro45Ala, Pro45Leu, Val170Leu, Val170Phe, Ala286Val, Thr400Ala, Gln409His, Glu423Ala, Phe424Leu, Phe424Val, Phe424Ser, Gly426Ser, Gly426Asp, Thr427Ser, Thr427Pro, Thr427Ala, Ser428Arg, Ser428Ile, Ser428Thr, Gln429Lys, Gln429His, Leu430Val, Leu430Pro, Leu430Arg, Ser431Arg, Ser431Gly, Ser431Ile, Ser431Thr, Gln432Lys, Gln432Ter, Gln432Glu, 1295del9bp, Gln432Leu, Gln432Pro, Gln432Arg, Gln432His, Phe433Leu, Phe433Val, Met434Thr, 1303del6bp, Met434Ile, Asp435Asn, Asp435Tyr, Asp435His, Asp435Val, Asp435Ala, Asp435Gly, Asp435Glu, Gln436Leu, Gln436His, Asn437Tyr, Asn437His, Asn437Asp, Asn437Ile, Asn437Thr, Asn438Lys, Leu440Met, Leu440Leu, Leu440Pro, Ser441Leu, Ser441Trp, Gly442Trp, Gly442Ala, Leu443Ser, Leu443Trp, Thr444Pro, Thr444Ile, His445Asn, His445Tyr, His445Asp, His445Leu, His445Pro, His445Arg, His445Gln, Lys446Gln, Lys446Asn, Arg447His, Arg447Pro, Ser450Leu, Ser450Trp, Leu452Pro, Pro454Ser, Pro454His, Pro454Leu, Pro454Arg, Leu457Pro, Glu460Gly, Leu464Met, Ser472Ala, Ile480Val, Ile480Thr, Glu481Gly, Thr482Pro, Pro483Leu, Leu490Val, Ile491Phe, Ile491Leu, Ile491Val, Ser493Leu, Arg552Cys, Glu592Asp, Leu731Pro, Glu761Asp, Ser874Tyr, Ser874Phe, 1275del6bp, 1276del6bp, 1294ins6bp, 1294ins3bp, 1294del6bp, 1297ins3bp, 1297del9bp, 1297del3bp, 1299del3bp, 1299del12bp, 1300ins3bp, 1300del12bp

Drug	Gene	Mutation (Cont.)
RIF	*rpoB*	1306del6bp, 1306del3bp, 1308del6bp, 1312del3bp, 1336del3bp
	rpoC	Gln329Lys, Gly332Ser, Gly332Cys, Gly332Arg, Asn416Ser, Gly433Ser, Gly433Cys, Pro434Gln, Pro434Leu, Pro434Arg, Lys445Arg, Phe452Ser, Val483Ala, Val483Gly, Trp484Gly, Asp485Asn, Ile491Thr, Ile491Val, Leu516Pro, Gly519Ser, Gly519Arg, Gly519Asp, Leu527Val, Asn698Ser, Asp735Asn, Asp747Ala, Glu750Asp, Thr825Ala, Ile885Val, Val1039Ala, Val1039Gly, Pro1040Thr, Pro1040Ala, Pro1040Leu, Pro1040Ser, Pro1040Arg, Gln1125His, Lys1152Gln, Val1252Leu, Val1252Met
PZA	*pncA*	T-12G, T-12C, A-11G, A-11C, T-10C, T-7C, Met1Ile, Met1Thr, Ala3Glu, Ala3Pro, Leu4Trp, Leu4Ser, Ile5Ser, Ile6Ser, Ile6Leu, Val7Asp, Val7Gly, Val7Phe, Val7Ile, Asp8Glu, Asp8Gly, Asp8Ala, Asp8Tyr, Asp8Asn, Val9Gly, Val9Ala, Val9Leu, Gln10Ter, Gln10Lys, Gln10Arg, Gln10Pro, 10del8bp, Asp12Ala, Asp12Asn, Asp12Gly, Phe13Leu, Phe13Ser, 13del11bp, Cys14Ter, Cys14Trp, Cys14Tyr, Cys14Gly, Cys14Arg, Gly17Val, Gly17Asp, Ser18Pro, Leu19Arg, Leu19Pro, Val21Gly, Val21Ala, Gly23Val, Gly24Asp, Gly24Arg, Ala26Gly, Leu27Pro, Ala28Asp, 30delG, Ile31Ser, Tyr34Ser, Tyr34Asp, Tyr34Ter, Leu35Arg, Leu35Pro, Glu37Ter, Tyr41Ter, Tyr41His, His43Pro, Val44Gly, Val44Ala, Val45Gly, Ala46Pro, Ala46Val, Ala46Glu, Thr47Ser, Thr47Ala, Thr47Pro, Asp49Val, Asp49Gly, Asp49Ala, Asp49Asn, Asp49His, 49delC, His51Tyr, His51Asn, His51Gln, His51Arg, His51Pro, Asp53Ala, Asp53Asn, Pro54Leu, Pro54Ala, Pro54Ser, Pro54Gln, Pro54Thr, 54insCC, His57Leu, His57Arg, His57Pro, His57Tyr, His57Asn, His57Asp, Phe58Leu, Phe58Ser, Ser59Pro, Thr61Pro, 61delG, Pro62Leu, Pro62Arg, Asp63Gly, Asp63Ala, Asp63His, Tyr64Asp, Ser66Pro, 66delT, Ser67Pro, Trp68Ter, Trp68Leu, Trp68Ser, Trp68Arg, Trp68Gly, Pro69Leu, Pro69Arg, His71Arg, His71Tyr, His71Asn, His71Asp, His71Gln, Cys72Ter, Cys72Trp, Cys72Tyr, Cys72Arg, 72delC, 73delC, Thr76Ile, Thr76Pro, Gly78Asp, Gly78Ser, Gly78Cys, Ala79Gly, Phe81Val, His82Leu, His82Arg, His82Tyr, His82Asp, Leu85Arg, Leu85Pro, 86delG, Thr87Met, Ser88Ter, Ile90Ser, Phe94Leu, Phe94Ser, Tyr95Ter, Lys96Thr, Lys96Glu, Lys96Gln, Lys96Asn, Lys96Arg, Gly97Arg, Gly97Cys, Gly97Ser, Gly97Asp, Tyr99Ter, Ala102Val, Ala102Thr, 102insA, Tyr103His, Tyr103Ter, Tyr103Cys, Tyr103Ser, Ser104Arg, Glu107Ter, Gly108Arg, 113del24bp, Thr114Ala, Thr114Pro, Leu116Arg, Leu116Pro, Leu116Val, Asn118Thr, Trp119Cys, Trp119Ter, Trp119Arg, Leu120Pro, Arg121Pro, Val125Asp, Val125Gly, Val128Gly, Val130Gly, Val130Ala, Val131Phe, Gly132Val, Gly132Asp, Gly132Ala

Drug	Gene	Mutation (Cont.)
PZA	*pncA*	Gly132Ser, Ile133Asn, Ile133Thr, Ala134Val, Ala134Ser, Thr135Pro, Asp136Gly, Asp136Tyr, Asp136Asn, Asp136His, His137Arg, His137Pro, Cys138Ter, Cys138Tyr, Cys138Ser, 138insC, Cys138Arg, Val139Met, Val139Gly, Val139Ala, Val139Leu, Arg140Ser, Arg140Gly, Arg140His, Gln141Pro, Gln141Ter, Thr142Met, Thr142Lys, Thr142Ala, Thr142Pro, Ala143Thr, Ala143Pro, Ala146Thr, Ala146Val, Arg148Ser, Leu151Ser, 152del33bp, Thr153Asn, Arg154Met, Arg154Gly, Val155Gly, Val155Ala, Val155Leu, Leu159Arg, Leu159Pro, Thr160Pro, 160delT, Ala161Pro, Gly162Asp, 163delG, 164delC, 164insA, Ala165Thr, Thr168Asn, Thr168Pro, Ala171Val, Ala171Glu, Ala171Thr, Ala171Pro, Leu172Arg, Leu172Pro, Glu174Gly, 174delA, Met175Ile, Met175Arg, Met175Val, Ser179Arg, Val180Phe, Glu181Asp, Leu182Ser, Leu182Trp, Ser185Ile, Ser185Thr, 194insT, 220ins10bp, 223insC, 241delT, 289insA, 290delC, 303insC, 303insT, 308insG, 313del5bp, 314delG, 343delG, 370ins18bp, 370insCT, 384delCC, 384insCT, 390ins9bp, 391del11bp, 391del4bp, 393delC, 393insCC, 394insC, 399insA, 405insG, 407del11bp, 408delC, 409insG, 416insC, 419delCA, 422insCC, 422insC, 430insCC, 438delC, 445delC, 454delA, 455del8bp, 467insA, 477delG, 482ins4bp, 503insGC, 514delG, 520ins5bp, 534insG
	rpsA	Glu433Asp, Arg474Trp, Arg474Leu
	panD	His21Arg, Ile49Val, Met117Ile, Glu126Ter, Ala128Ser, Glu130Gly, Leu136Arg, Val138Ala, Val138Gly, Val138Glu, 395delC
EMB	*embA*	G-43C, C-16A, C-16G, C-16T, C-15G, C-12A, C-12T, C-11T, C-11A, C-8A, C-8T, Asp4Asn, Gly321Ser, Gly350Asp, Ala462Val, Asp833Ala
	embB	Ala221Gly, Ala225Ser, Val227Gly, Leu239Pro, Asp240His, Arg257Trp, Ala271Val, Ser272Cys, Ser272Ile, Arg274Pro, Ala281Pro, Val282Leu, Val282Gly, Phe285Leu, Phe287Val, Phe287Cys, Ile293Asn, Gly294Ser, Asn296His, Asn296Lys, Ser297Ala, Ser297Leu, Ser298Trp, Asp299Glu, Met306Leu, Met306Val, Met306Thr, Met306Ile, Val309Gly, Asp311His, Asp311Gly, His312Arg, Ala313Val, Gly314Ala, Tyr315Cys, Ser317Thr, Asn318His, Asn318Ser, Asn318Lys, Tyr319Asn, Tyr319Ser, Tyr319Cys, Phe320Leu, 983delA, Asp328Tyr, Asp328Val, Asp328Gly, Asp328His, Phe330Val, Gly331Arg, Trp332Leu, Tyr333His, Tyr334His, Ser347Cys, Ser347Ile, Ser347Thr, Asp354Ala, Ala356Ser, Ala356Val, Ala357Ser, Ala357Val, Gly358Val, Leu359Ile, Val360Met, Ser366Pro, Arg367Pro, Glu368Gln, Glu368Ala, Glu368Asp, Val369Leu, Val369Ala, Leu370Arg, Pro371Arg, Gly374Val, Pro375Ala, Val377Met, Val377Gly, Glu378Lys

Drug	Gene	Mutation (Cont.)
EMB	*embB*	Glu378Ala, Ala379Thr, Ala379Asp, Ser380Arg, Ser380Asn, Tyr384Asn, Ala388Gly, Trp395Arg, Trp395Cys, Pro397Thr, Pro397Gln, Phe398Tyr, Asn399His, Asn399Asp, Asn400Lys, Gly401Ser, Pro404Ser, Glu405Asp, Gly406Ser, Gly406Cys, Gly406Asp, Gly406Ala, Ser412Leu, Met423Ile, Met423Thr, Pro430Leu, Ala431Thr, Val435Gly, Val436Gly, Thr437Ala, Pro446His, Gly448Val, Val452Leu, Gly459Ala, Arg460Cys, Arg460Leu, Pro461Ser, Arg469Pro, Arg471Pro, Met482Ile, Gln497Lys, Gln497Pro, Gln497Arg, Gln497His, Glu504Asp, Arg507Lys, Ala659Thr, Gly745Asp, Gln853Pro, Asp959Ala, Met1000Arg, His1002Arg, Asp1024Asn, Thr1082Ala
	embC	Glu187Glu, Ala244Thr, Ala247Pro, Leu251Arg, Ala254Gly, Thr270Ile, Gly272Ser, His285Tyr, Val287Phe, Gly288Trp, Gly288Val, Tyr296His, Tyr296Ser, Ile297Leu, Ile297Thr, Met300Arg, Arg302Gly, Val303Gly, Ala307Thr, Gly308Asp, Tyr309Asn, Met310Lys, Gly325Ser, Trp326Arg, Tyr327Asn, Ala779Val
	embR	Val7Gly, Thr32Pro, Asp53His, Leu172Arg, Ala176Val, Arg186Gly, Ala187Val, Ser188Arg, Glu200Ala, Leu211Met, Asp223Glu, Arg230Trp, Gly240Arg, Leu292Met, Leu320Val, Ser326Asn, Val350Gly, Arg356Gly, Ile369Thr, Gln379Arg
	ubiA	Ala38Thr, Ala38Ser, Ala38Val, Ser173Ala, Lys174Thr, Lys174Arg, Phe176Leu, Met180Val, Val188Ala, Val229Ala, Gly234Val, Ala249Thr
STR	*rpsL*	Arg9His, Thr40Ile, Thr41Ser, Lys43Thr, Lys43Arg, Lys51Asn, Val52Gly, Gly84Val, Arg86Trp, Arg86Pro, Val87Leu, Lys88Gln, Lys88Met, Lys88Thr, Lys88Arg, Val93Met
	rrs	G190A, G277C, G427T, G427C, C462T, C492T, C513T, A514T, A514C, C516T, C517T, G628C, C799T, G888A, C905A, C905G, A906G, A907T, A907C, A908G, C1229T, T1239C, A1401G, C1402T, C1402A, G1484T
	gid	Ala8Val, Val65Ala, Val65Gly, Pro75Arg, Pro75Leu, Val77Ala, Val77Gly, Leu79Trp, Leu79Ser, Ala80Pro, Val139Leu, Ser149Arg
AMK	*rrs*	G-7T, C513T, A514T, A514C, C517T, C905A, A906G, G1158T, A1338C, A1401G, C1402A, C1402T, G1484T
	eis	G-54T, T-53G, G-37T, C-15G, C-14T, A-13G, C-12T, G-10C, G-10A, C-8A
KAN	*rrs*	C513T, A514C, A514T, C517T, C905A, A906G, A1401G, C1402A, C1402T, G1484T
	eis	G-54T, T-53G, G-37T, C-15G, C-14T, A-13G, C-12T, G-10C, G-10A, C-8A
CAP	*rrs*	C513T, A514C, A514T, C517T, C905A, A906G, A1401G, C1402A, C1402T, G1484T

Drug	Gene	Mutation (Cont.)
CAP	*tlyA*	Arg3Ter, Arg14Trp, Arg18Ter, Gln22Ter, 23delA, 26delC, 67del3bp, Ala67Glu, Lys69Glu, Ala91Glu, Leu118Pro, Val128Glu, Leu150Pro, Pro183Leu, Gln184Ter, Phe185Leu, Asn236Lys, Glu238Lys, 310delG, 400delA, 477delG, 586delG, 653delT, 673delGT, 758delC, 755insGT
	idsA2	A-65G
FLQ	*gyrA*	Arg68Gly, Ser69Thr, His70Arg, Ala74Ser, Thr80Ala, Gly88Cys, Gly88Ala, Asp89Asn, Ala90Glu, Ala90Val, Ala90Gly, Ser91Pro, Ser91Ala, Ile92Met, Asp94Asn, Asp94Tyr, Asp94His, Asp94Val, Asp94Ala, Asp94Gly, Pro102His, Leu109Val, Leu105Arg, Asn193Ser
	gyrB	Ala90Gly, Arg446Cys, Arg446His, Arg446Leu, Ser447Phe, Asp461Asn, Asp461His, Asp461Ala, Gly470Cys, Gly470Ala, Ile486Leu, Asp494Ala, Asn499Tyr, Asn499Asp, Asn499Thr, Asn499Lys, Thr500Pro, Thr500Asn, Thr500Ile, Glu501Val, Glu501Asp, Ala504Thr, Ala504Val, Thr507Met, Gln538His, Val670Phe
ETH	*inhA*	A-92T, G-67C, C-34T, G-24T, G-17T, A-16G, C-15T, A-11T, T-8A, T-8G, T-8C, T-5A, Ile21Thr, Ser94Ala, Ile194Thr
	ethA	Met1Arg, Gly11Ala, His22Pro, Gln24Ter, Gly43Cys, Gly43Ser, Pro51Leu, Asp58Ala, Thr61Met, Tyr84Asp, 112delT, Gln165Pro, Thr186Lys, Glu223Lys, Thr232Ala, Gln246Ter, Gln269Ter, Leu272Pro, Ser329Leu, Leu333Arg, Ile338Ser, Thr342Lys, Gln347Ter, Ala381Pro, Gly385Asp, Tyr386Cys, Thr392Ala, Leu397Arg, Cys403Gly, Gly413Asp, Tyr461His, Arg463Ser, 705delA, 770delC, 1292delG, 1325delGC
	ethR	T-65C, Ala95Thr, Phe110Leu
PAS	*thyA*	Thr22Pro, Thr22Ala, Tyr36Cys, His75Asn, Gly76Ter, Val77Phe, Trp83Cys, Trp83Ter, Gly91Arg, Gln97Arg, Trp98Ter, Trp101Arg, Ser105Pro, 113delA, Arg126Gln, Phe152Val, Cys161Tyr, His207Arg, 219ins9bp, Arg222Gly, Pro224Leu, Arg235Pro, 262delG, 374insA, 474delG
	thyX	C-16T, G-9A, C-4A, C-4T
	folC	Thr20Pro, Glu40Gln, Ile43Thr, Ile43Ser, Arg49Pro, Glu40Gly, Arg49Trp, Leu56Val, Asn73Ser, Arg91Trp, Ser150Cys, Ser150Arg, Ser150Gly, Phe152Ser, Phe152Leu, Glu153Gly, Glu153Ala, Ala183Pro
	ribD	Gly8Arg
DCS	*alr*	Leu113Arg
	iniA	His42Arg

Table A.2. Diversity and frequency of WGS-derived drug resistance-associated mutations among false positive genomic profiles from the local dataset.

Drug	False Positives	Mutation (Frequency)
INH	1	*kasA*_Gly269Ser (1)
RIF	1	*rpoB*_His445Asn (1)
PZA	2	*pncA*_Gly97Cys (1), *pncA*_Val155Leu (1)
EMB	32	*embA* promoter_C-11A (4), *embA* promoter_C-12A (4), *embA* promoter_C-16G (1), *embA* promoter_C-16T (1), *embA* promoter_C-8A (1), *embB*_Gln497Arg (2), *embB*_Gly406Ala (1), *embB*_Gly406Asp (1), *embB*_Met306Ile (5), *embB*_Met306Val (15), *embB*_Met423Thr (8), *embB*_Pro397Thr (4), *embB*_Ser297Ala (1)
STR	13	*gid*_Ala80Pro (7), *rpsL*_Lys43Arg (1), *rrs*_A1401G (4), *rrs*_C492T (4)
AMK	32	*eis* promoter_C-12T (2), *eis* promoter_G-10A (28), *rrs*_A906G (1), *rrs*_C1402A (1), *rrs*_G1484T (1)
KAN	7	*eis* promoter_C-12T (1), *eis* promoter_G-10A (4), *rrs*_A1401G (1), *rrs*_A906G (1)
CAP	7	*idsA2* promoter_A-65G (4), *rrs*_A1401G (1), *rrs*_A906G (1), *rrs*_C517T (1), *tlyA*_755insGT (1)
FLQ	2	*gyrA*_His70Arg (1), *gyrA*_Thr80Ala (1)
ETH	1	*fabG1* promoter_C-15T (1)
PAS	4	*thyX* promoter_C-16T (3), *thyX* promoter_C-4T (1)
DCS	2	*alr*_Leu113Arg (2)

Table A.3. Diversity and frequency of WGS-derived mutations among false negative genomic profiles from the local dataset.

Drug	False Negatives	Mutation (Frequency)
INH	10	*ahpC*_Pro44Arg (2), *katG*_1947insC (1), *katG*_270insA (1)
RIF	4	*rpoB*_1282del9bp (1), *rpoC*_Gly594Glu (1)
PZA	19	*pncA*_251insG (5), *pncA*_283insA (1), *pncA*_416del3bp (2), *pncA*_452del13bp (1), *pncA*_485ins10bp (1), *pncA*_Glu15Ter (1), *pncA*_Ile133Ser (1), *pncA*_Phe13Ile (1)
EMB	3	*Rv2820c*_Leu212Val (1)
STR	14	*gid*_104delC (1), *gid*_117delG (1), *gid*_Arg21Trp (1), *gid*_Glu92Gln (2), *gid*_Gly130Ala (1), *gid*_Gly34Glu (2), *gid*_Leu16Arg (6), *gid*_Ser136Ter (1), *gid*_Trp45Ter (1), *gid*_Val112Gly (1), *rrs*_C924T (1)
AMK	2	N/A
KAN	0	N/A
CAP	2	N/A
FLQ	3	*gyrA*_Glu21Gln (3), *gyrA*_Gly247Ser (1), *gyrA*_Gly668Asp (3), *gyrA*_Ser95Thr (3), *gyrB*_Val301Leu (1)
ETH	14	*ethA*_182delG (1), *ethA*_343delT (1), *ethA*_886delA (1), *ethA*_Ala185Thr (1), *ethA*_Cys403Arg (2), *ethA*_Glu400Asp (1), *ethA*_Gly450Asp (1), *ethA*_Leu194Pro (1), *ethA*_Lys224Ter (1), *ethA*_Ser266Arg (1), *inhA*_Ile21Val (1)
PAS	9	*folC*_Ser98Gly (1), *thyA*_Asp81Ala (1), *thyA*_Thr202Ala (9)
DCS	10	*alr*_Met343Thr (1), *alr*_Phe4Leu (3), *iniA*_282del5bp (6)

Table A.4. Distribution of *M. tuberculosis* clinical isolates from the global dataset, according to the human-adapted MTBC lineages, across 53 countries. [1] Countries with a sample size <20.

Continent	Country	Relative frequency (%)					
		L1	L2	L3	L4	L5	L7
Africa	Botswana[1]	0.0	23.1	0.0	76.9	0.0	0.0
	Côte D'Ivoire	0.0	0.0	0.0	96.6	3.4	0.0
	Democratic Republic of The Congo[1]	0.0	0.0	0.0	100	0.0	0.0
	Djibouti	33.9	5.1	20.3	40.7	0.0	0.0
	Ethiopia	0.0	0.0	0.0	0.0	0.0	100
	Kenya[1]	0.0	0.0	0.0	100	0.0	0.0
	Malawi	18.8	4.7	12.6	63.9	0.0	0.0
	Mali[1]	0.0	33.3	0.0	66.7	0.0	0.0
	Mozambique[1]	0.0	0.0	0.0	100	0.0	0.0

Continent	Country	Relative frequency (%) (Cont.)					
		L1	L2	L3	L4	L5	L7
Africa	Nigeria	0.0	0.0	0.0	100	0.0	0.0
	South Africa	0.1	43.9	1.9	54.0	0.0	0.0
	Tunisia[1]	0.0	0.0	0.0	100	0.0	0.0
	Uganda	0.8	5.0	24.0	70.2	0.0	0.0
	Zimbabwe[1]	0.0	27.8	22.2	50.0	0.0	0.0
America	Argentina	0.0	0.0	0.0	100	0.0	0.0
	Brazil	0.8	0.8	0.4	98.0	0.0	0.0
	Canada	20.5	22.0	8.8	48.7	0.0	0.0
	Guatemala	0.0	25.0	0.0	75.0	0.0	0.0
	Peru	0.3	9.3	0.0	90.4	0.0	0.0
	United States	36.0	27.0	1.1	36.0	0.0	0.0
Asia	Azerbaijan	0.0	73.2	0.0	26.8	0.0	0.0
	Bangladesh	9.8	41.5	22.0	26.8	0.0	0.0
	China	0.0	94.9	0.5	4.6	0.0	0.0
	India	14.9	23.8	44.6	16.8	0.0	0.0
	Indonesia	2.3	36.9	0.0	60.8	0.0	0.0
	Israel[1]	16.7	16.7	0.0	66.7	0.0	0.0
	Japan[1]	0.0	100	0.0	0.0	0.0	0.0
	Kazakhstan[1]	0.0	100	0.0	0.0	0.0	0.0
	Malaysia[1]	40.0	40.0	0.0	20.0	0.0	0.0
	Myanmar/Burma[1]	6.7	80.0	0.0	13.3	0.0	0.0
	Singapore[1]	0.0	100	0.0	0.0	0.0	0.0
	South Korea[1]	0.0	100	0.0	0.0	0.0	0.0
	Tajikistan	0.0	83.7	0.0	16.3	0.0	0.0
	Thailand	21.3	70.0	0.0	8.6	0.0	0.0
	Vietnam	24.0	64.2	0.0	11.7	0.0	0.0
Europe	Albania[1]	0.0	0.0	0.0	100	0.0	0.0
	Belarus	0.0	74.6	0.0	25.4	0.0	0.0
	Denmark	0.0	0.0	0.0	100	0.0	0.0
	Estonia[1]	0.0	100	0.0	0.0	0.0	0.0
	France[1]	16.7	50.0	0.0	33.3	0.0	0.0
	Georgia	0.0	72.0	0.0	28.0	0.0	0.0
	Ireland[1]	50.0	25.0	0.0	25.0	0.0	0.0
	Italy	3.7	29.6	0.0	64.8	1.9	0.0
	Moldova	0.0	38.9	0.0	61.1	0.0	0.0
	Netherlands	6.7	20.0	0.0	73.3	0.0	0.0

Continent	Country	Relative frequency (%) (Cont.)					
		L1	L2	L3	L4	L5	L7
Europe	Norway	0.0	4.8	88.9	6.3	0.0	0.0
	Romania	0.0	2.6	0.0	97.4	0.0	0.0
	Russian Federation	0.0	65.4	0.2	34.4	0.0	0.0
	Sweden[1]	100	0.0	0.0	0.0	0.0	0.0
	Switzerland	8.6	34.5	17.2	39.7	0.0	0.0
	United Kingdom	10.6	10.6	30.1	48.4	0.3	0.0
Oceania	Australia	6.6	66.2	4.4	22.8	0.0	0.0
	Papua New Guinea	0.0	96.1	0.0	3.9	0.0	0.0

Table A.5. Distribution of GCs, clustered and non-clustered associated *M. tuberculosis* isolates, and of drug resistance profiles across lineage-associated clades.

Lineage	Clade (%)	GCs	Relative frequency (%)					
			Clustered	Non-clustered	S[1]	DR	MDR	XDR
L1	EAI5 (17.35)	42	24.9	75.1	95.3	4.7	0.0	0.0
	EAI2-Manila (15.12)	10	6.3	93.7	97.6	2.4	0.0	0.0
	EAI4-VNM (10.75)	12	10.6	89.4	97.9	2.1	0.0	0.0
	EAI1-SOM (8.93)	22	33.2	66.8	88.8	9.2	0.0	2.0
	EAI6-BGD1 (8.29)	28	33.0	67.0	99.5	0.5	0.0	0.0
	EAI3-IND (5.92)	11	19.2	80.8	91.5	7.7	0.0	0.8
	ATYPIC (5.24)	28	34.8	65.2	88.7	10.4	0.9	0.0
	EAI2-nonthaburi (4.96)	5	14.7	85.3	98.2	1.8	0.0	0.0
	Unknown (1.41)	1	6.5	93.5	100.0	0.0	0.0	0.0
	EAI8-MDG (0.96)	1	23.8	76.2	61.9	38.1	0.0	0.0
	EAI7-BGD2 (0.59)	2	30.8	69.2	92.3	7.7	0.0	0.0
	EAI (0.23)	1	20.0	80.0	80.0	0.0	20.0	0.0
	AFRI (0.18)	1	50.0	50.0	100.0	0.0	0.0	0.0
	Manu1 (0.09)	0	0.0	100.0	50.0	0.0	50.0	0.0
	EAI2 (0.05)	0	0.0	100.0	100.0	0.0	0.0	0.0
L2	Beijing (88.62)	474	40.3	59.7	20.9	30.9	45.3	2.9
	ATYPIC (4.33)	71	52.0	48.0	12.7	50.2	28.8	8.3
	Unknown (1.14)	6	18.3	81.7	20.0	10.0	70.0	0.0
	Manu_ancestor (0.59)	3	29.0	71.0	25.8	35.5	38.7	0.0

Lineage	Clade (%)	GCs	Relative frequency (%) (Cont.)					
			Clustered	Non-clustered	S[1]	DR	MDR	XDR
L2	Manu3 (0.08)	1	25.0	75.0	25.0	0.0	75.0	0.0
	H (0.02)	0	0.0	100.0	0.0	100.0	0.0	0.0
	EAI5 (0.02)	0	0.0	100.0	0.0	0.0	100.0	0.0
L3	CAS1-Delhi (50.11)	117	30.0	70.0	16.3	9.1	74.2	0.4
	CAS1-Kili (19.8)	53	53.7	46.3	10.0	4.1	85.9	0.0
	ATYPIC (7.4)	36	45.7	54.3	11.6	2.2	86.2	0.0
	CAS (6.33)	16	41.5	58.5	11.9	6.8	79.7	1.7
	CAS2 (2.25)	8	38.1	61.9	31.0	16.7	52.4	0.0
	Manu3 (2.25)	7	47.6	52.4	31.0	0.0	69.0	0.0
	Beijing (0.8)	3	33.3	66.7	6.7	20.0	73.3	0.0
	Unknown (0.75)	2	21.4	78.6	28.6	14.3	57.1	0.0
	Manu2 (0.05)	0	0.0	100.0	0.0	0.0	100.0	0.0
L4	T1 (21.94)	174	44.8	55.2	14.4	11.2	73.4	0.9
	LAM11-ZWE (7.52)	115	55.4	44.6	6.6	3.3	90.1	0.0
	Unknown (4.85)	37	33.3	66.7	29.1	12.7	57.7	0.5
	LAM3 (4.67)	41	41.9	58.1	76.1	20.1	3.6	0.3
	ATYPIC (4.48)	109	50.5	49.5	18.5	23.5	54.2	3.7
	LAM9 (4.3)	44	30.0	70.0	36.1	30.9	32.2	0.8
	S (3.57)	33	43.9	56.1	10.0	18.3	70.4	1.3
	H1 (3.38)	37	35.8	64.2	19.3	11.6	68.1	1.1
	Cameroon (3.2)	13	70.7	29.3	67.8	5.9	26.3	0.0
	T2 (3.03)	26	27.3	72.7	23.0	24.6	49.2	3.1
	X1 (3.01)	28	62.6	37.4	16.1	9.1	73.6	1.2
	Ural-1 (2.95)	34	46.6	53.4	14.1	33.3	47.4	5.2
	LAM1 (2.56)	27	46.8	53.2	14.4	25.9	50.5	9.3
	X2 (2.29)	18	52.8	47.2	16.1	11.4	72.5	0.0
	T4-CEU1 (2.24)	15	83.1	16.9	1.1	0.5	98.4	0.0
	T5 (1.79)	11	70.2	29.8	3.3	3.3	91.4	2.0
	T (1.71)	21	45.1	54.9	25.7	4.9	68.8	0.7
	H2 (1.54)	6	85.4	14.6	4.6	83.1	6.2	6.2
	X3 (1.49)	16	29.4	70.6	13.5	15.1	69.8	1.6
	LAM (1.45)	16	64.8	35.2	43.4	46.7	9.8	0.0
	LAM4 (1.35)	18	57.9	42.1	36.8	38.6	20.2	4.4

Lineage	Clade (%)	GCs	Relative frequency (%) (Cont.)					
			Clustered	Non-clustered	S[1]	DR	MDR	XDR
L4	T3 (0.97)	10	23.2	76.8	22.0	11.0	64.6	2.4
	H3 (0.86)	6	26.0	74.0	20.5	15.1	64.4	0.0
	T5-RUS1 (0.77)	9	27.7	72.3	84.6	12.3	1.5	1.5
	LAM6 (0.76)	7	34.4	65.6	57.8	35.9	1.6	4.7
	LAM5 (0.63)	8	43.4	56.6	37.7	49.1	11.3	1.9
	LAM2 (0.58)	5	26.5	73.5	12.2	42.9	44.9	0.0
	T2-Uganda (0.57)	6	33.3	66.7	91.7	0.0	4.2	4.2
	T3-ETH (0.53)	6	26.7	73.3	37.8	22.2	40.0	0.0
	Blank (0.45)	3	71.1	28.9	23.7	50.0	0.0	26.3
	Ural-2 (0.44)	4	24.3	75.7	8.1	2.7	89.2	0.0
	T4 (0.43)	2	16.7	83.3	5.6	11.1	83.3	0.0
	H (0.27)	4	39.1	60.9	13.0	0.0	87.0	0.0
	Turkey (0.26)	2	54.5	45.5	9.1	27.3	59.1	4.5
	T1-RUS2 (0.24)	0	0.0	100.0	10.0	10.0	80.0	0.0
	Manu2 (0.21)	1	5.6	94.4	11.1	5.6	83.3	0.0
	T5-Madrid2 (0.17)	2	50.0	50.0	50.0	7.1	42.9	0.0
	LAM8 (0.14)	0	0.0	100.0	16.7	0.0	83.3	0.0
	T2/LAM3 (0.07)	0	0.0	100.0	50.0	33.3	16.7	0.0
	PINI2 (0.06)	1	60.0	40.0	80.0	20.0	0.0	0.0
	T-H37Rv (0.04)	0	0.0	100.0	0.0	33.3	66.7	0.0
	EAI5 (0.02)	0	0.0	100.0	100.0	0.0	0.0	0.0
	T3-OSA (0.02)	0	0.0	100.0	50.0	0.0	50.0	0.0
	AFRI_2 (0.02)	0	0.0	100.0	50.0	0.0	50.0	0.0
	Manu1 (0.02)	0	0.0	100.0	0.0	0.0	100	0.0
	T-tuscany (0.01)	0	0.0	100.0	100.0	0.0	0.0	0.0
	EAI8-MDG (0.01)	0	0.0	100.0	100.0	0.0	0.0	0.0
	PINI (0.01)	0	0.0	100.0	0.0	0.0	100.0	0.0
	CAS1-Delhi (0.01)	0	0.0	100.0	0.0	0.0	100.0	0.0
L5	AFRI_2 (52.94)	0	0.0	100.0	100.0	0.0	0.0	0.0
	AFRI_3 (8.82)	0	0.0	100.0	100.0	0.0	0.0	0.0
	CAS (5.88)	0	0.0	100.0	100.0	0.0	0.0	0.0
	Unknown (2.94)	0	0.0	100.0	100.0	0.0	0.0	0.0
	ATYPIC (2.94)	0	0.0	100.0	100.0	0.0	0.0	0.0
L7	Unknown (93.94)	5	35.5	64.5	100.0	0.0	0.0	0.0
	ATYPIC (3.03)	0	100.0	0.0	100.0	0.0	0.0	0.0

[1] Pan-susceptible.