UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

INSTITUTO DE
BIOMEDICINA DE
VALENCIA CSIC

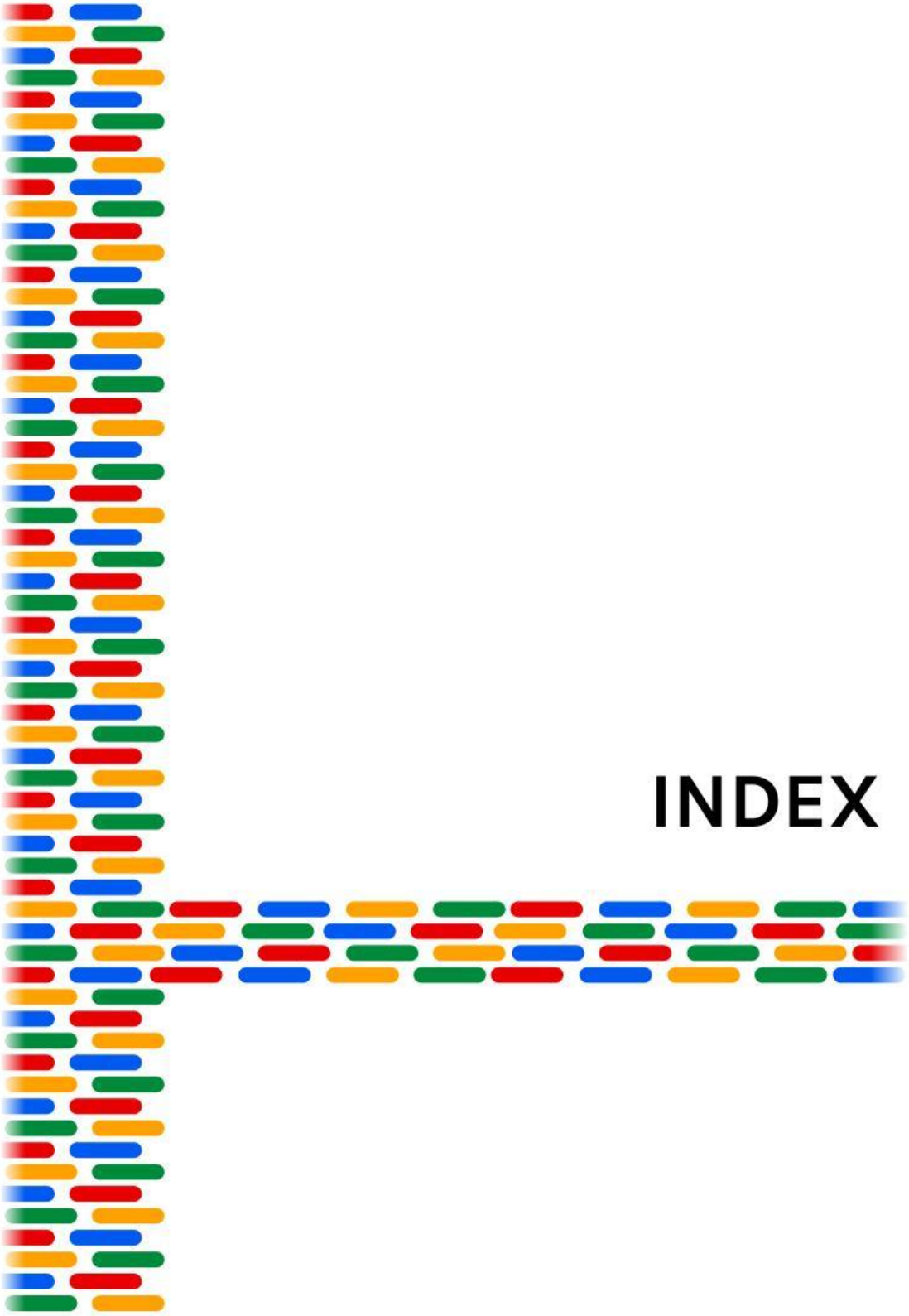
Functional and evolutionary genomics of antibiotic resistance in tuberculosis: from biology to global DR-TB control

PhD Thesis

Miguel Ángel Moreno Molina

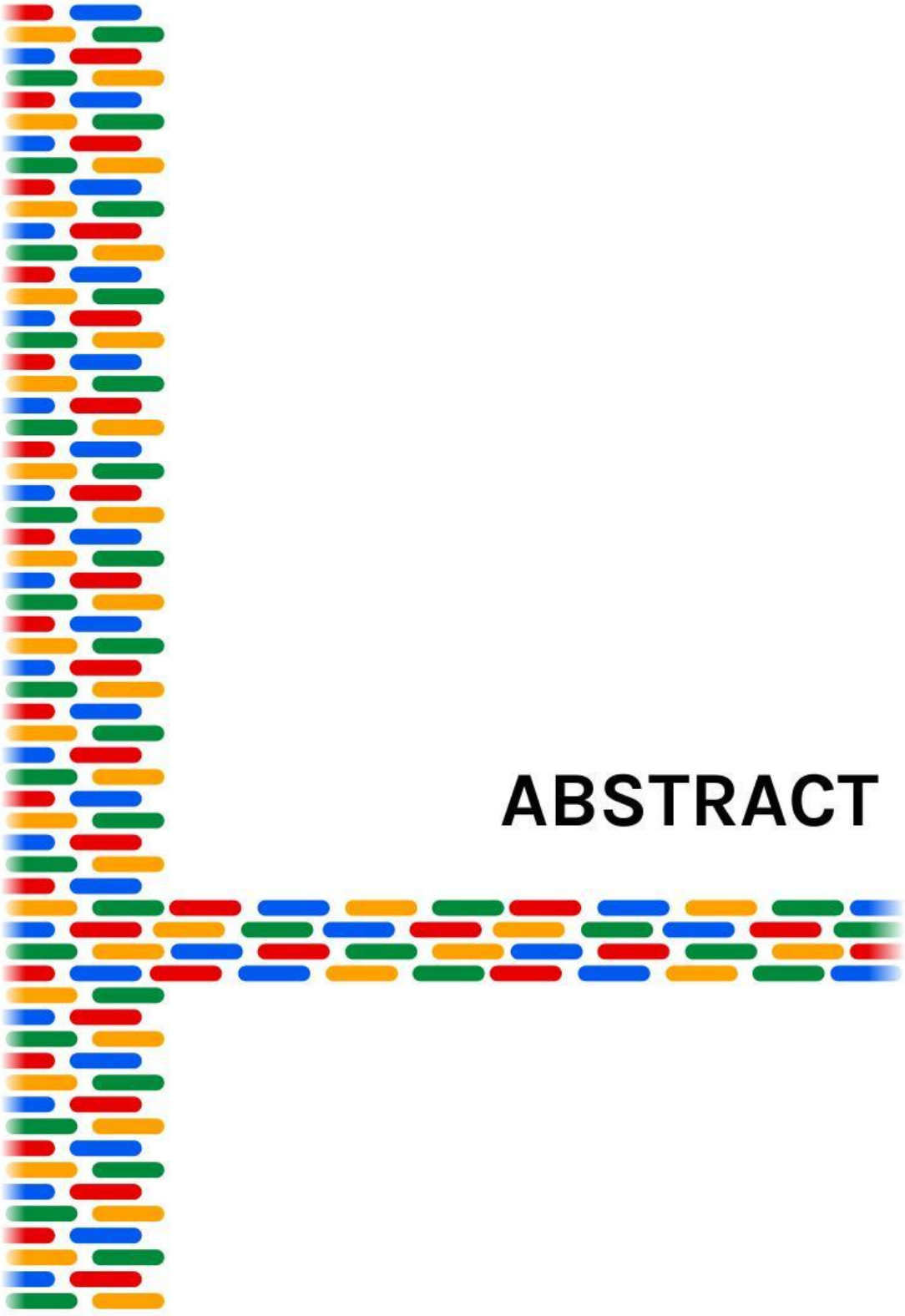
Supervisor: Iñaki Comas Espadas

Doctoral Programme in Biotechnology
Universitat Politècnica de València
July 2023



INDEX

INDEX



ABSTRACT

ABSTRACT

Antibiotic resistance in tuberculosis is a serious global health problem and a growing obstacle to the control of the disease. In 2021, 12% of TB deaths were due to resistant strains of the bacterium, with this percentage being much higher in some countries where, in addition, a large proportion of new cases are already resistant. These data imply a great economic burden for public health services and threaten to eliminate the progress in tuberculosis control that has been made in the last 30 years. In this thesis we use genomics as a tool to study different aspects of the biology of the bacterium in relation to the development of antibiotic resistance and provide new knowledge to tackle the challenge of its eradication.

First, we characterize new genomic determinants of resistance to isoniazid, one of the most widely-used first-line antibiotics. To do so, we used an approach based on functional genomics and phylogenetic association. The method consists of transposon sequencing of bacterial populations exposed to the antibiotic to determine which genes are involved in both sensitivity and resistance, and then filtering with genomic data from a global collection of clinical strains. We have verified the importance of the metabolic pathways of bacterial wall synthesis in the mechanism of action of isoniazid, and also discovered new genes involved in cellular redox balance that confer low-level resistance to this antibiotic. These results can be used to develop new diagnostic techniques or therapeutic

targets, and the method is applicable to new antibiotics to predict future resistance determinants.

Second, we explored the bacterial diversity of tuberculosis at its natural site of infection and the population dynamics of antibiotic resistance. We have analyzed samples from different parts of the lung lesion or granuloma from patients in Georgia, a country with a high incidence of drug-resistant TB, and detected a significant number of polyclonal infections, i.e. infections caused by more than one different strain of the bacterium. These different strains can in turn be resistant to different antibiotics, which makes it essential to detect them in time to offer the patient the most appropriate treatment possible. The surgical specimens gave us a more complete picture of bacterial diversity than sputum, the routine clinical specimen, and allowed us to detect polyclonal infections more accurately. The data show that we are underestimating these types of infections in high-burden TB countries and they may adversely affect treatment outcome, so in these settings a second sampling during the course of the antibiotic regimen would be advisable.

Third, we evaluated the role of one of the most important comorbidities of tuberculosis, HIV, in the development of resistance during the first weeks of treatment. To this end, we have sampled patients from Mozambique with and without HIV on a serial basis during the first month to analyze the impact of co-infection on both immune and antibiotic selective pressures. We have detected a higher total diversity in seronegative patients and

furthermore, in comparison, HIV+ patients present difficulties in eliminating this diversity during the early stages of treatment that could affect their success. Thanks to deep sequencing, we have also been able to observe an accumulation of variants in resistance-related genes, and we have associated some of them with changes in the MICs (minimum inhibitory concentrations) of the samples. These small changes during the first month may serve as predictors of a reduced ability to eliminate bacterial diversity, and could indicate future treatment failure.

In summary, in this thesis we have focused on better understanding the bacterial and host factors that have an impact on the development of antibiotic-resistant tuberculosis and as a result will allow the development of new diagnostics and epidemiological models to control it. Genomic sequencing is a great tool to study this and other pathogens in real time, and offers the possibility of detecting resistance quickly and with high certainty in the clinical setting. A correct diagnosis ensures adequate, shorter and more successful treatment, and also prevents transmission to new hosts. The techniques and results presented here contribute to achieve these objectives and improve the global control of the drug-resistant tuberculosis epidemic.

RESUMEN

La resistencia a antibióticos en tuberculosis es un grave problema de salud global y un obstáculo cada vez mayor para el control de la enfermedad. En 2021, el 12% de las muertes por tuberculosis fueron debidas a cepas resistentes de la bacteria, siendo este porcentaje mucho más elevado en algunos países donde además una gran parte de los nuevos casos son ya resistentes. Estos datos implican una gran carga económica para los servicios de salud pública y amenazan con eliminar los progresos en el control de la tuberculosis que se han hecho en los últimos 30 años. En esta tesis usamos la genómica como una herramienta para estudiar diferentes aspectos de la biología de la bacteria en relación al desarrollo de resistencia a antibióticos y aportamos nuevo conocimiento para afrontar el reto de su erradicación.

En primer lugar, caracterizamos nuevos determinantes genómicos de la resistencia a isoniazida, uno de los antibióticos de primera línea más ampliamente usados. Para ello utilizamos una aproximación basada en genómica funcional y asociación filogenética. El método consiste en secuenciación de transposones de poblaciones bacterianas expuestas al antibiótico para determinar los genes implicados tanto en sensibilidad como en resistencia, y su filtrado con datos genómicos de una colección global de cepas clínicas. Hemos comprobado la importancia de las vías metabólicas de síntesis de la pared bacteriana en el mecanismo de acción del antibiótico, y además descubierto nuevos genes implicados en el balance redox celular que

otorgan resistencia de bajo nivel a la bacteria. Estos resultados pueden emplearse para desarrollar nuevas técnicas diagnósticas o dianas terapéuticas, y el método es aplicable a nuevos antibióticos para predecir futuros determinantes de la resistencia.

En segundo lugar, exploramos la diversidad bacteriana de la tuberculosis en su lugar natural de infección y las dinámicas poblacionales de la resistencia a antibióticos. Hemos analizado muestras de diferentes partes de la lesión pulmonar o granuloma de pacientes de Georgia, un país con alta incidencia de tuberculosis resistente, y detectamos una importante cantidad de infecciones policlonales, es decir, infecciones causadas por más de una cepa diferente de la bacteria. Estas diferentes cepas pueden a su vez ser resistentes a antibióticos distintos, lo que hace esencial poder detectarlas a tiempo para ofrecer al paciente un tratamiento lo más adecuado posible. Las muestras de cirujía nos otorgaron una visión más completa de la diversidad bacteriana que el esputo, la muestra clínica que se toma de rutina, y nos permitieron detectar las infecciones policlonales con mayor precisión. Los datos muestran que estamos subestimando este tipo de infecciones en los países de alta carga de tuberculosis y pueden afectar negativamente al resultado del tratamiento, por lo que en estos entornos sería recomendable un segundo muestreo durante el curso del régimen antibiótico.

Y en tercer lugar, evaluamos el papel de una de las comorbilidades más importantes de la tuberculosis, el VIH, en el desarrollo de resistencias durante las primeras

semanas de tratamiento. Para ello, hemos tomado muestras de pacientes de Mozambique con y sin VIH de forma seriada durante el primer mes para analizar cuál es el impacto de la coinfección en las presiones selectivas tanto inmunes como antibióticas. Hemos detectado una mayor diversidad total en los pacientes seronegativos y además, en comparación, los pacientes VIH+ presentan dificultades para eliminar esta diversidad durante las etapas tempranas del tratamiento que podrían afectar a su éxito. Gracias a la secuenciación profunda, también hemos podido observar una acumulación de variantes en genes relacionados con resistencia, y hemos asociado algunas de ellas a cambios en las CMI (concentraciones mínimas inhibitorias) de las muestras. Estos pequeños cambios durante el primer mes pueden servir como predictores de una menor capacidad para eliminar la diversidad bacteriana, y podrían llegar a indicar un futuro fallo en el tratamiento.

En resumen, en esta tesis nos hemos centrado en entender mejor los factores bacterianos y del hospedador que tienen un impacto en el desarrollo de la tuberculosis resistente a antibióticos y como resultado permitirán desarrollar nuevos diagnósticos y modelos epidemiológicos para controlarla. La secuenciación genómica es una gran herramienta para estudiar éste y otros patógenos en tiempo real, y ofrece la posibilidad de detectar resistencias rápidamente y con una alta certeza en el ámbito clínico. Un correcto diagnóstico asegura un tratamiento adecuado, más corto y exitoso, y además previene la transmisión a nuevos huéspedes. Las técnicas y resultados expuestos aquí

contribuyen a lograr estos objetivos y mejorar el control global de la epidemia de tuberculosis resistente.

RESUM

La resistència a antibiòtics en tuberculosi és un greu problema de salut global i un obstacle cada vegada més gran per al control de la malaltia. El 2021, el 12% de les morts per tuberculosi van ser degudes a soques resistents del bacteri, sent aquest percentatge molt més elevat en alguns països on a més una gran part dels nous casos són ja resistents. Aquestes dades impliquen una gran càrrega econòmica per als serveis de salut pública i amenacen amb eliminar els progressos en el control de la tuberculosi que s'han fet els darrers 30 anys. En aquesta tesi fem servir la genòmica com una eina per estudiar diferents aspectes de la biologia del bacteri en relació amb el desenvolupament de resistència a antibiòtics i aportem nou coneixement per afrontar el repte de la seua eradicació.

En primer lloc, caracteritzem nous determinants genòmics de la resistència a isoniazida, un dels antibiòtics de primera línia més àmpliament utilitzats. Per això utilitzem una aproximació basada en genòmica funcional i associació filogenètica. El mètode consisteix en seqüenciació de transposons de poblacions bacterianes exposades a l'antibiòtic per determinar els gens implicats tant en sensibilitat com en resistència, i el filtratge amb dades genòmiques d'una col·lecció global de soques clíniques. Hem comprovat la importància de les vies metabòliques de síntesi de la paret bacteriana al mecanisme d'acció de l'antibiòtic, i a més descobert nous gens implicats en el balanç redox cel·lular que atorguen resistència de baix nivell al bacteri. Aquests resultats es poden fer servir per

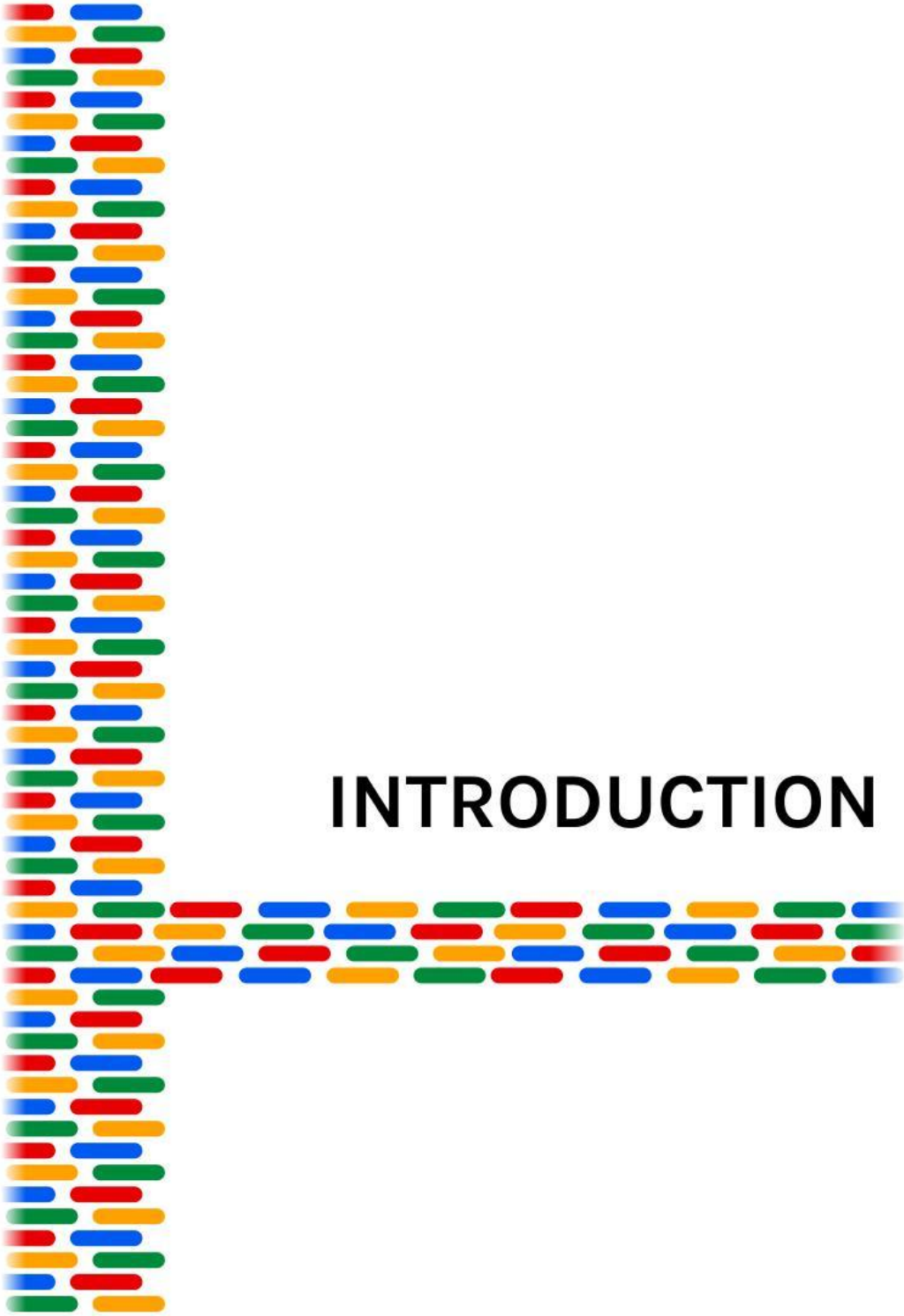
desenvolupar noves tècniques diagnòstiques o dianes terapèutiques, i el mètode és aplicable a nous antibiòtics per predir futurs determinants de la resistència.

En segon lloc, explorem la diversitat bacteriana de la tuberculosi al lloc natural d'infecció i les dinàmiques poblacionals de la resistència a antibiòtics. Hem analitzat mostres de diferents parts de la lesió pulmonar o granuloma de pacients de Geòrgia, un país amb alta incidència de tuberculosi resistent, i detectem una quantitat important d'infeccions policlonals, és a dir, infeccions causades per més d'una soca diferent del bacteri. Aquestes diferents soques poden ser resistents a antibiòtics diferents, cosa que fa essencial poder detectar-les a temps per oferir al pacient un tractament el més adequat possible. Les mostres de cirurgia ens van atorgar una visió més completa de la diversitat bacteriana que l'esput, la mostra clínica que es pren de rutina, i ens van permetre detectar les infeccions policlonals amb més precisió. Les dades mostren que estem subestimant aquest tipus d'infeccions als països d'alta càrrega de tuberculosi i poden afectar negativament el resultat del tractament, per això en aquests entorns seria recomanable un segon mostreig durant el curs del règim antibiòtic.

I en tercer lloc, avaluem el paper d'una de les comorbiditats més importants de la tuberculosi, el VIH, en el desenvolupament de resistències durant les primeres setmanes de tractament. Per això, hem pres mostres de pacients de Moçambic amb i sense VIH de forma seriada durant el primer mes per analitzar quin és l'impacte de la

coinfecció a les pressions selectives tant immunes com antibiòtiques. Hem detectat una diversitat total més gran en els pacients seronegatius i a més, en comparació, els pacients VIH+ presenten dificultats per eliminar aquesta diversitat durant les etapes primerenques del tractament que podrien afectar el seu èxit. Gràcies a la seqüenciació profunda, també hem pogut observar una acumulació de variants en gens relacionats amb resistència, i hem associat algunes a canvis en les CMI (concentracions mínimes inhibitòries) de les mostres. Aquests subtils canvis durant el primer mes poden servir com a predictors d'una capacitat menor per eliminar la diversitat bacteriana, i podrien arribar a indicar un futur fracàs en el tractament.

En resum, en aquesta tesi ens hem centrat a entendre millor els factors bacterians i de l'hoste que tenen un impacte en el desenvolupament de la tuberculosi resistent a antibiòtics i com a resultat permetran desenvolupar nous diagnòstics i models epidemiològics per controlar-la. La seqüenciació genòmica és una gran eina per estudiar aquest i altres patògens en temps real, i ofereix la possibilitat de detectar resistències ràpidament i amb una certesa alta en l'àmbit clínic. Un diagnòstic correcte assegura un tractament adequat, més curt i exitós, i a més prevé la transmissió a nous hostes. Les tècniques i els resultats exposats aquí contribueixen a assolir aquests objectius i millorar el control global de l'epidèmia de tuberculosi resistent.



INTRODUCTION

INTRODUCTION

Antibiotic resistance and tuberculosis

Antibiotic resistance is a major global health problem and an increasingly serious threat to public health [1]. Antibiotic resistance occurs when bacteria change in a way that reduces or eliminates the effectiveness of the chemical compounds designed to cure or prevent infections [2]. This can occur through a process of natural selection, where bacteria with a genetic mutation that allows them to survive antibiotic treatment are more likely to persist and reproduce [3]. It is becoming a major public health issue as the development of antibiotic-resistant bacteria is outpacing the development of new antibiotics [4].

The consequences of antibiotic resistance are widespread and can have serious implications for public health. Antibiotic-resistant bacteria can cause more severe and longer-lasting infections, resulting in higher healthcare costs, increased morbidity and mortality, and a decreased quality of life [5]. Antibiotic resistance is also associated with increased length of hospital stay and increased risk of hospital-acquired infections [5]. Thus the problem of antibiotic resistance is quite complex and requires a multifaceted approach to address it. Some strategies to tackle antibiotic resistance include the judicious use of antibiotics, the development of new antibiotics, the development of new diagnostic and laboratory tests, the

implementation of infection prevention and transmission control measures, or the use of vaccines. These strategies must be implemented in combination in order to effectively address the problem and prevent both the emergence and spread of drug-resistant bacterial strains [6].

Tuberculosis (TB) is an airborne infectious disease mainly caused by *Mycobacterium tuberculosis (*MTB*)*, a human obligate pathogen belonging to the *Mycobacterium tuberculosis* complex (MTBC). It is one of the most common and deadliest diseases in human history, and it is estimated that up to one-fourth of the world's population has been exposed to it [7]. TB is actually an ancient disease that has been around for a very long time, and it has adapted to the human host over many centuries. Despite mycobacteria in general and MTB in particular show a low genetic diversity compared to other bacteria, it still translates into a sizable amount of phenotypic diversity. MTB has evolved into eight distinct lineages that are classified according to their genetic makeup and are used to identify different strains of the bacteria [8,9]. Also, they are thought to cause different levels of disease severity. Many of them are restricted to some geographic areas, while others like Lineage 4 or 2 are globally distributed and more successful [10].

Regarding the natural history of TB, it has been described that after infection, bacteria are phagocytized by macrophages in the lower sections of the lungs and internalized into deeper tissues where they are kept contained in granulomas [7]. These structures are the perfect example of the duality of MTB infection: they keep bacteria

confined from the rest of the body to protect from further infection, but at the same time provide the perfect environment for them to thrive, recruiting phagocytic cells in which they can replicate. At this stage, infection can remain latent or reactivate -usually within two years- and progress to active disease [7]. The mechanisms of TB latency and reactivation are not well understood yet, but evidence suggests that the host immune response plays an important role. The necrotizing granuloma eventually cavitates, producing coughing and completing the infective cycle via small bacteria-containing aerosols that can transmit to the next host.

Although TB can be treated with antibiotics, it is still a major problem due to its ability to become resistant over time [11]. The first-line treatment for tuberculosis is a combination of four different drugs: isoniazid (INH), rifampicin (RIF), ethambutol (EMB), and pyrazinamide (PZA). Treatment with this combination of antibiotics is generally recommended for anyone with active TB disease, regardless of their age, sex, or underlying medical conditions and is typically continued for six months or longer, depending on the individual's response to the medications. In cases where the infecting strain is resistant to first-line drugs, second-line drugs may be used, but they are more costly and often more toxic. This second-line treatment is recommended when a patient fails to respond to, or is unable to take, first-line TB drugs and it must be tailored to the specific needs of the patient and the type of TB infection. Second-line treatment of TB typically includes the use of injectable agents such as amikacin (AMK) or capreomycin (CPR), and oral medications

such as fluoroquinolones, ethionamide or cycloserine. Other drugs, such as linezolid and clofazimine, may also be used as part of a second-line regimen. Two new TB drugs, bedaquiline and delamanid, were approved for use in 2012 and 2014, respectively. These drugs have recently been included in the WHO treatment guidelines and offer hope for individuals with drug-resistant TB, with a new all-oral 6-month regimen consisting of bedaquiline, pretomanid, linezolid and moxifloxacin (BPaLM) [12,13].

Unsurprisingly, TB is still one of the top ten causes of death in the world, and was responsible for an estimated 1.6 million deaths in 2021, with 191,000 of those due to DR-TB [14]. The challenge of eradicating drug resistance in tuberculosis has been the focus of public health efforts for decades. Drug-resistant TB (DR-TB) has emerged as a growing public health challenge in many countries, and is a major cause of morbidity and mortality in many regions. Traditionally, DR-TB has been broadly classified into two main categories: multidrug resistant tuberculosis (MDR-TB), caused by bacteria that is resistant to two of the first-line treatment drugs, isoniazid (INH) and rifampicin (RIF), and extensively-drug resistant tuberculosis (XDR-TB), a form of TB caused by bacteria that is resistant to at least four of the most commonly used antitubercular drugs, including isoniazid and rifampicin but also second-line antibiotics, any fluoroquinolone (FQ) and at least one of the injectable drugs line kanamycin (KAN) [11]. However, definitions are changing as the first options for treating MDR-TB are now all-oral drugs. The importance of both MDR-TB and XDR-TB resides in their higher severity, with an increased mortality rate and

greater difficulty to treat. They are particularly concerning as they are also more difficult to control and are most common in low- and middle-income countries, where access to treatment is limited and the risk of transmission is high.

Mechanisms and evolution of drug resistance in *M. tuberculosis*

Mycobacteria possess a complex, thick and hydrophobic cell wall that confers them intrinsic resistance to a wide range of antimicrobials. The envelope of *Mycobacterium tuberculosis* has been described to incorporate a wide variety of lipids -like mycolic acids, unique to mycobacteria- and very few porins, making it scarcely permeable to external molecules [15]. This renders many hydrophobic compounds useless to treat tuberculosis, like most macrolides and tetracyclines or some rifamycins and fluoroquinolones. We also have experimental evidence that blocking lipid synthesis renders the bacteria susceptible to many of these drugs. However, antibiotics that manage to surpass the cell wall can be enzymatically degraded, like the cleavage of the penicillin β-lactam ring by mycobacterial β-lactamases [16], or modified like in the case of aminoglycosides acetylation to render them inactive [17]. Also, a wide range of efflux systems are thought to contribute to drug resistance in various levels of effectiveness. This is supported by the fact that many of

these efflux pumps are non-specific and thus highly permissive towards structurally-diverse molecules [18,19].

We have seen how enzymatic action and efflux systems, together with the particular structure and composition of the mycobacterial cell wall, make MTB intrinsically resistant to many antibiotics. However, antimicrobial resistance can also be evolved during the standard therapy to both first- and second-line drugs, and through various different mechanisms [20]. As with many other pathogens, the most common acquired resistance mechanism is drug target alteration, which in MTB mostly happens through chromosomal mutations in the form of single nucleotide polymorphisms (SNPs). The specificity of antibiotics to their targets is usually so high that small alterations in the target protein can diminish their affinity or even render them completely useless. A classic example is rifampicin resistance by means of point mutations in the RNA polymerase subunit *rpoB* [21]. Another possibility is the prevention of prodrug activation in the cell, as some antitubercular drugs are actually prodrugs. This is the case of isoniazid, which needs to be activated via the catalase peroxidase *katG*, and resistance is achieved by mutations that strongly reduce this enzyme activity [22]. Finally, resistance can arise as a result of an imbalance of the drug and its target, via the overexpression of the latter. An example of this mechanism is also illustrated by isoniazid resistance due to the overexpression of its target, the reductase *inhA* involved in mycolic acid biosynthesis [23]. Finally, it is important to note that in addition to these mechanisms, variants may arise that confer low-level

resistance, which in turn provides the necessary evolutionary space for the bacteria to develop high-level resistance through a canonical mechanism [24,25].

The evolution of *de novo* drug resistance at the population scale has classically been attributed to suboptimal control of transmission, poor adherence to treatment or irregular drug supply in low-income settings. However, we see drug resistance arise in all kinds of settings, even under optimal treatment conditions, which suggests that both pathogen and host determinants play a role in its development [26]. Both patient heterogeneity in the form of immune status or suboptimal drug penetration can create opportunities for resistance to arise in subinhibitory antibiotic concentrations [27,28]. On the pathogen side, emergence of drug resistance is heavily influenced by other factors such as the effective population size during infection, the particular mutation rate of MTB or the fitness cost associated with many resistance-conferring mutations. Population size is in turn influenced by the infection bottleneck -the number of bacteria received by the host upon transmission- and the ability to control the infection once it has been established. Higher bacterial loads statistically increase the chances for resistance to develop, even under optimal drug concentrations. MTB mutation rate during infection is nonetheless low compared to other pathogens and it has been estimated to be around 0.5 SNPs/genome/year [29].

It is important to understand that, canonically, drug resistance comes with a fitness cost. This is, the ability of a

particular bacterium to grow and thrive in the context of a competitive environment, and those with drug-resistance mutations that suffer a lower or even null fitness cost are more likely to persist even in the absence of antibiotic pressure. Compensation mechanisms have been described in MTB, for example mutations in *rpoC* that arise to make up for the loss of fitness that comes with rifampicin resistance through *rpoB* mutations [30]. And finally, the genetic background of the infecting strain is also relevant for both the evolution of drug resistance, its levels and its associated fitness cost [31], as there is evidence that some lineages like Lineage 2 have a higher tendency to develop drug resistance [32].

All these factors combined, we have been observing a rise of drug-resistant tuberculosis in the last years. The WHO has collected data and monitored the levels of drug-resistant TB since 1994 and the last available report shows a total of 450,000 DR-TB cases in 2021, out of a total of 10.6 million new TB cases [14]. Just 3.6% of all new cases were DR-TB in 2021, with 18% being previously treated patients. However, in certain high-burden DR-TB settings, acquisition of drug resistance is majoritarily due to transmission and not treatment failure. DR-TB is most prevalent in low- and middle-income countries, where access to healthcare is more limited and diagnosis is often delayed. In 2021, nearly 90% of the estimated DR-TB cases occurred in these settings. Some of the highest proportions of DR-TB cases are found in Russia, South Africa, Eastern Europe or Central Asia. In many of these regions, treatment success rates are alarmingly low when DR-TB strains are

involved. This scenario shows that the intrinsic resistance of *Mycobacterium tuberculosis* to numerous classes of antibiotics, the increasing number of drug-resistant strains, and the scarcity of newly-developed drugs are endangering the success of controlling the disease in the coming years.

Genomics and drug resistance detection in tuberculosis

With the advent of next-generation sequencing, bacterial genomics research received an important boost in terms of affordability, scale and resolution. Having access to the complete genome meant we could associate genotypes to clinically-relevant phenotypes like drug resistance. Traditionally, accurate resistance detection has required dedicated expertise and infrastructure. Drug susceptibility testing (DST) is the standard diagnostic method, but it is slow and can often delay adequate treatment when the infecting strain is not susceptible. The most successful commercial platform for this purpose is the Bactec MGIT system [33], which can detect isoniazid and rifampicin resistance, yet it can take up to 3 weeks to provide a result. There are also molecular tests that help improve this diagnostic gap, like the widely-known GeneXpert, a standardized PCR-based method that can detect rifampicin resistance from a clinical isolate in a cost-effective manner that is very convenient in low-income settings [34,35]. Tests like this, despite being fast, are very limited in the range of variants that they can detect and thus lack flexibility, heavily

relying on our previous knowledge of resistance-conferring genotypes [36].

In this regard, whole-genome sequencing (WGS) has seen increasing use as a tool to predict drug resistance in tuberculosis, vastly improving our ability to provide an accurate diagnosis in a short time window [37]. Having the ability to correctly detect mutations that confer antibiotic resistance is vital to avoid starting a patient in a suboptimal therapy, and WGS can offer results in 48h with an adequate setup. Illumina platforms that utilize short-read sequencing technology have been established as the standard both in research and in the clinic, although long-read technologies such as PacBio and Nanopore are also gaining traction after years of development [38]. TB whole-genome sequencing routinely aims for at least ~60X of read depth and a ~98% genome coverage, which is sufficient for most transmission and drug resistance analyses. However, deep sequencing at 500X or even 1000X has proved to be useful for many research applications, despite its higher cost preventing it from reaching the clinic. Other types of sequencing strategies to detect drug resistance in TB also exist, like targeted sequencing in the form of amplicon panels that cover a range of resistance-related genes [39,40]. This approach is both faster and cheaper at the expense of losing information of the rest of the genome, but in turn it can achieve extremely high read depths. That is critical to identify small fractions of a sample -bacterial subpopulations- that have evolved to be drug-resistant during the course of treatment under the selective pressure of antibiotics, a phenomenon known as heteroresistance

[41,42]. Both deep sequencing and amplicon sequencing offer the chance to detect these resistant subpopulations that may have evolved *de novo* during treatment or are present in the sample due to polyclonal infections, which appear to be more frequent in high-burden TB settings than previously thought. Nonetheless, with the decreasing trend in sequencing costs, achieving a higher read depth -and resolution- in clinical settings may become a reality sooner than we think, as detecting these two phenomenons is imperative for a successful treatment outcome.

As we have already stated, MTB is a particularly clonal organism that has no evidence of horizontal gene transfer, so it is predominantly through single nucleotide mutations (SNPs) and small insertions and deletions (INDELs) that drug resistance is acquired [43]. This narrow range of variability facilitates the interpretation of genomic data, and as more MTB sequences started to be available, researchers quickly noted a pattern by which some variants appeared to always happen alongside the same phenotypes [44]. This realization was the start of the genomic prediction of drug resistance in tuberculosis and, succeeding this finding, many studies started to associate WGS data to DST results to infer causal relationships and test these variants in the lab. After this, with the early small lists of resistance-related variants, the first big study using genomic sequencing to predict drug resistance in TB was published in 2015 [45], and it accomplished both high sensitivity and specificity values for isoniazid and rifampicin. Resistance to other antibiotics, however, appeared to be more complex to predict, pointing out that we were still missing many

genomic resistance determinants. And so the need emerged to start developing a robust, evidence-based catalog to make accurate drug resistance predictions using WGS data. Several databases and tools were developed, for example PhyResSE [46], TBProfiler [47] or genTB [48], and eventually large-scale consortiums like CRyPTIC or ReSeqTB were established, gathering TB experts to build and analyze huge global strain collections coupled with DST data. This proved to be the optimal path in systematically identifying new variants that confer antibiotic resistance in MTB.

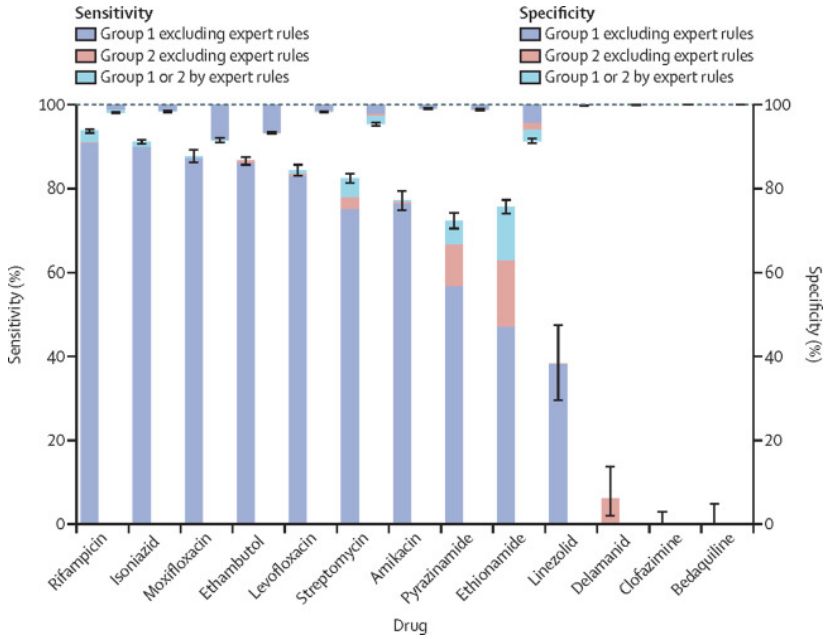


Figure 1. Sensitivity and specificity for all drugs included in the 2021 WHO catalog for TB drug resistance diagnosis. Source: Walker *et al*. 2021 [49].

In 2021, these endeavors finally achieved the milestone of publishing the first WHO-sponsored drug resistance catalog for tuberculosis, which describes more than 17,000 algorithm and manually-curated variants to predict resistance to first-, second-line and new drugs [49]. It was built from the analysis of more than 38,000 samples and achieved a pooled sensitivity of more than 80% for the most widely-used antibiotics in TB (Figure 1). The catalog strives to be the groundwork of whole-genome sequencing as a reliable diagnostic tool for tuberculosis, and ongoing updates are now focusing on improving geographical representation of diverse strains, newer drugs like bedaquiline, epistatic interactions and recently-described mechanisms like drug tolerance. Antibiotic resistance diagnosis through genomics is a reality that seems closer each day and will allow us to offer real-time personalized care to every TB patient.

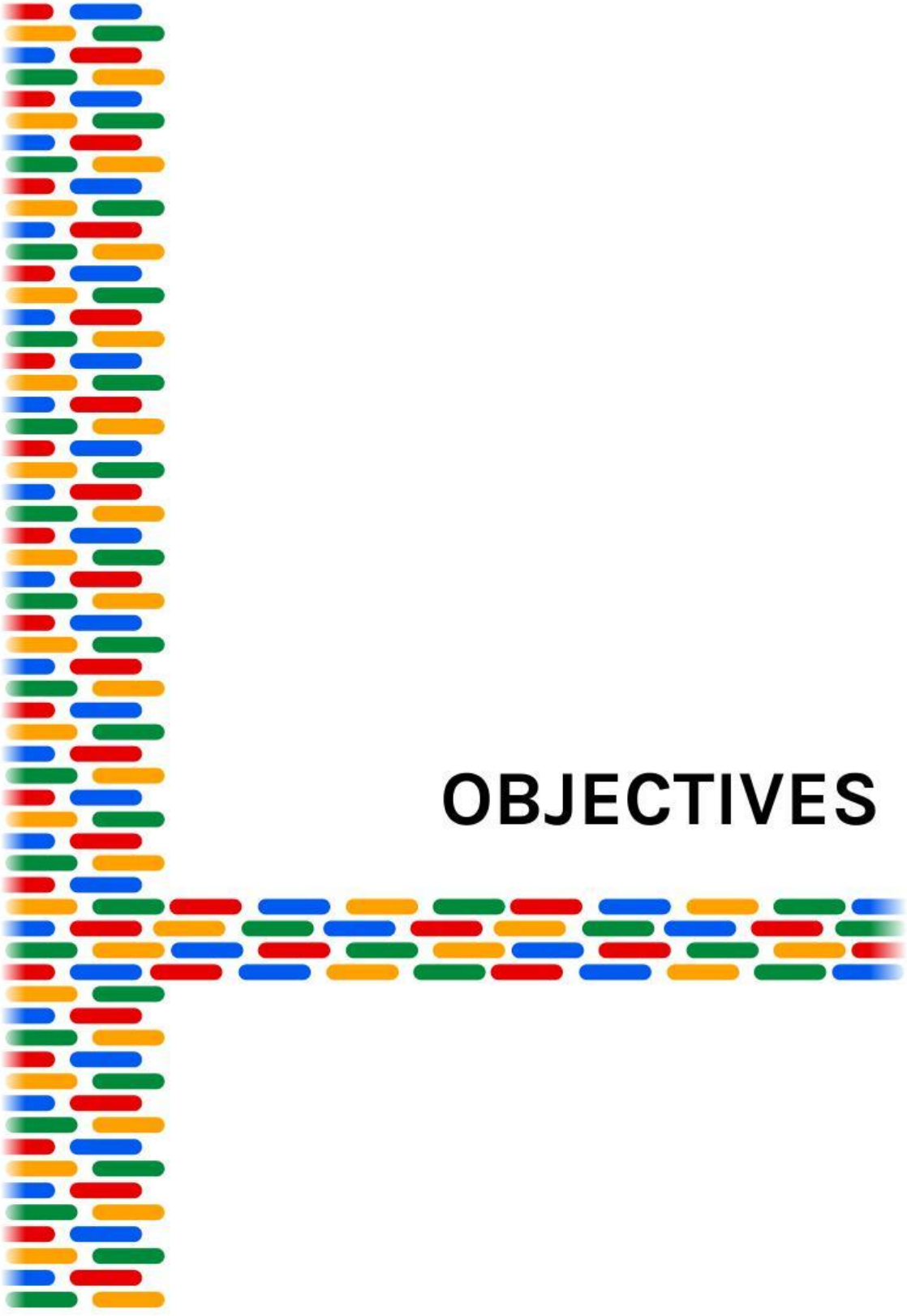
Tuberculosis comorbidities and TB-HIV coinfection

Comorbidities associated with TB are conditions that occur alongside the TB infection, and can complicate treatment, cause poorer outcomes, or increase the risk of death [50]. The most common comorbidities associated with TB include HIV, diabetes, cardiovascular disease, mental health conditions, and substance use disorders. HIV is the most frequent comorbidity, and is associated with more severe TB, more frequent and longer-lasting symptoms, and

decreased treatment success. Diabetes is another common comorbidity for TB, and it is associated with an increased risk of TB infection and lower treatment success rates. Cardiovascular disease is also associated with an increased risk of TB infection and poorer treatment outcomes. In addition, mental health conditions or substance use disorders can impact the management of TB, as they can lead to poor adherence to treatment, and they can also worsen the physical symptoms of TB, making them harder to manage.

HIV is however the most common comorbidity associated with TB and also a major risk factor for infection, with people living with HIV being up to 20 times more likely to develop TB than those without HIV [51]. The clinical outcomes of TB-HIV coinfection are worsened due to the presence of HIV, as several studies have observed that with HIV patients have a higher risk of TB mortality and drug-resistant TB than those without HIV, although the mechanisms by which this occurs are poorly understood [52]. The observation is that HIV patients develop TB with a subclinical manifestation, have lower bacterial loads, but usually progress faster than non-coinfected patients. Some argue that it is probably due to HIV-associated immunosuppression, which reduces the body's ability to fight off TB. Also, it has been shown that drug bioavailability is different in HIV patients due to them not forming granulomas as in regular TB infection. Finally, there is evidence to suggest that HIV-associated stigma and discrimination may also contribute to poorer clinical outcomes of TB-HIV coinfection [52]. But perhaps the key

factor is our lack of understanding of the pathogen-host interactions that drive the development and reactivation of TB in immunosuppressed individuals, with lots of work still to be done in this regard.



OBJECTIVES

OBJECTIVES OF THIS DISSERTATION

The challenge of eradicating DR-TB requires a comprehensive approach and thus the World Health Organization has identified a series of urgent problems to tackle in the End TB Strategy. These include the accurate detection of drug resistance, improved diagnostic techniques, a robust control of the transmission of resistant strains or the evaluation of comorbidities like HIV in the acquisition and spread of resistance. This dissertation revolves around the use of bacterial genomics as a tool to study different aspects of MTB biology and undertake the former challenges. Each chapter constitutes a step forward in the eradication of DR-TB in accordance with the following objectives:

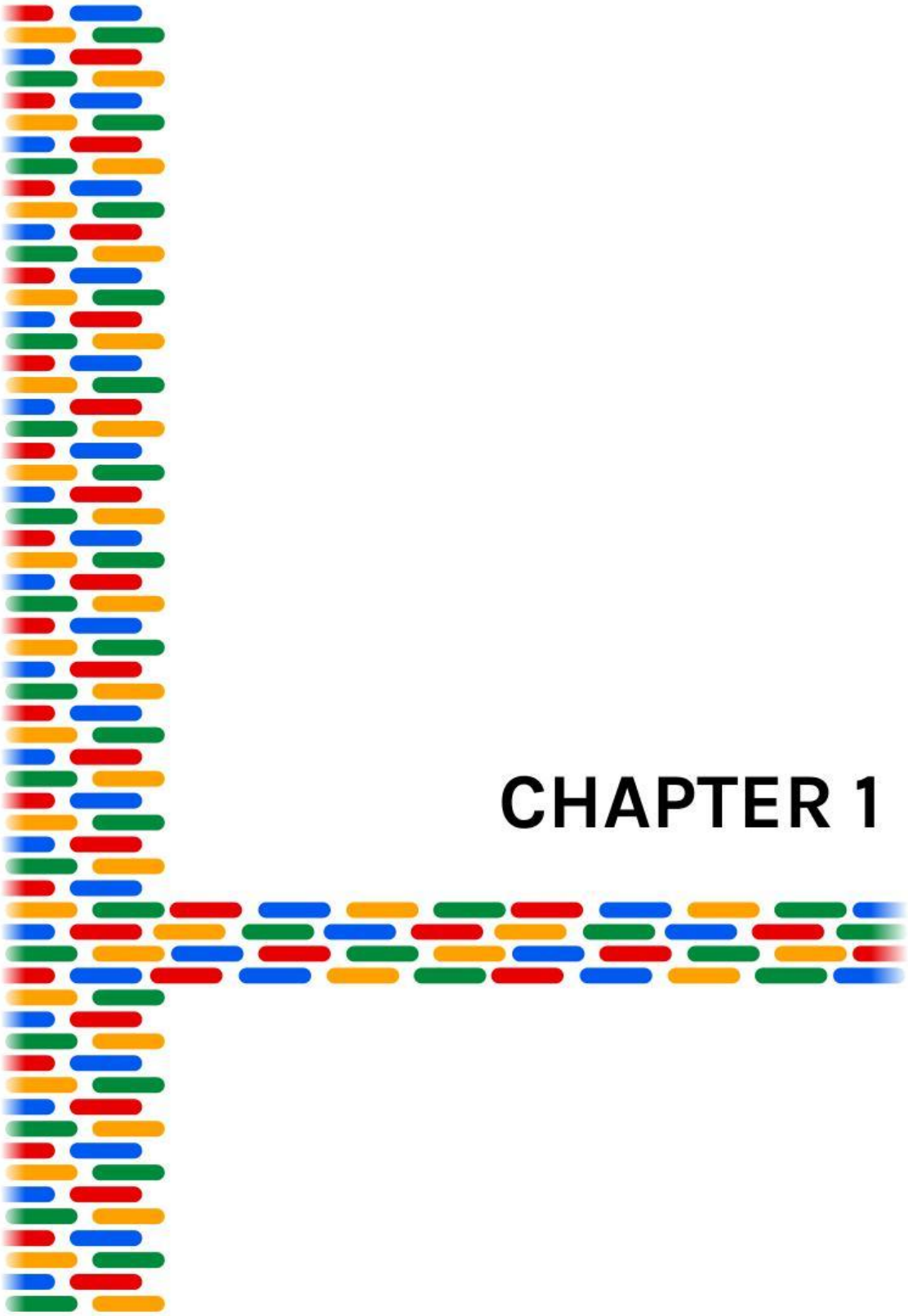
1. Characterize the genomic architecture and determinants of drug resistance in *Mycobacterium tuberculosis*. This objective corresponds with Chapter 1, in which we use functional genomics and phylogenetic association to unravel the architecture of isoniazid resistance, expanding the current knowledge of this antibiotic's mechanism of action and providing evidence of new genes and pathways involved in low-level resistance.

2. Improve diagnosis and treatment of DR-TB in high-burden settings. This objective corresponds with Chapter 2, in which we explore MTB diversity in the site of

infection using lung surgical samples from MDR and XDR-TB patients and discover a high rate of polyclonal infections. Different strains involved in these types of infection often have different drug resistance profiles, and we highlight how access to accurate and timely diagnosis of DR-TB is essential for successful treatment in Georgia, a high-burden DR-TB setting.

3. Evaluate the impact of TB-HIV coinfection in drug resistance evolution. This objective corresponds with Chapter 3, in which we delve into how HIV affects the antibiotic selective forces that shape MTB subpopulations during early treatment. We used deep-sequencing in two cohorts from Mozambique, both a high-burden TB and HIV country, to study the dynamics of early drug resistance acquisition. Results show how HIV impacts the rate at which TB subpopulations are eliminated and how we can link low-frequency variants to MIC shifts that can end up developing high-level drug resistance and eventual treatment failure.

In the following chapters, we will approach the study of DR-TB from both the bacterial and the host's point of view using functional and evolutionary genomics and providing valuable insights about its biology that will help us advance in the global control of this epidemic.



CHAPTER 1

CHAPTER 1

An evolutionary functional genomics approach identifies novel candidate regions involved in isoniazid resistance in *Mycobacterium tuberculosis*

Publication reference

Furió, V., Moreno-Molina, M., Chiner-Oms, Á. et al. An evolutionary functional genomics approach identifies novel candidate regions involved in isoniazid resistance in *Mycobacterium tuberculosis*. Commun Biol 4, 1322 (2021). https://doi.org/10.1038/s42003-021-02846-z

Victoria Furió[1*], Miguel Moreno-Molina[1*], Álvaro Chiner-Oms[1], Luis M Villamayor[2], Manuela Torres-Puente[1], Iñaki Comas[1,3]

* Equal contribution

[1] Institute of Biomedicine of Valencia (IBV-CSIC), Valencia (46020), Spain.
[2] FISABIO Public Health (CSISP), Valencia (46010), Spain.
[3] CIBER in Epidemiology and Public Health, Madrid (28029), Spain.

Chapter summary

Efforts to eradicate tuberculosis are hampered by the rise and spread of antibiotic resistance. Several large-scale projects have aimed to specifically link clinical mutations to resistance phenotypes, but they were limited in both their explanatory and predictive powers. Here, we combine functional genomics and phylogenetic associations using clinical strain genomes to decipher the architecture of isoniazid resistance and search for new resistance determinants. This approach has allowed us to confirm the main target route of the antibiotic, determine the clinical relevance of redox metabolism as an isoniazid resistance mechanism and identify novel candidate genes harboring resistance mutations in strains with previously unexplained isoniazid resistance. This approach can be useful for characterizing how the tuberculosis bacilli acquire resistance to new antibiotics and how to forestall them.

INTRODUCTION

In 2018, an estimated 484,000 people contracted drug-resistant tuberculosis and a further 214,000 people died from it [53]. Resistance to antitubercular drugs has been present ever since their introduction decades ago but it is now becoming a pressing problem as it hampers our ability to control and eradicate the disease. Drug-resistant tuberculosis requires longer treatments, has lower cure rates and spreads in the population, particularly in high-burden countries [53]. Licensing new antibiotics is not a definitive solution as the bacteria can develop resistance to those antibiotics as well [54,55]. A new approach is needed in which a thorough understanding of the evolutionary forces shaping resistance helps us understand how it is acquired and how it can be reversed.

Most of what we know of tuberculosis drug resistance comes from genetic association studies in which a particular mutation is associated with a specific resistance phenotype [56]. We now have large databases of diagnostic mutations with which we can reliably predict the resistance phenotype of our strain when we determine its genomic sequence [57]. For instance, we can detect rifampicin resistance with a 92% sensitivity, but the figure drops to 87% for isoniazid and 58% for ethambutol [58]. However, there is still a knowledge gap as the catalog of mutations is incomplete and we do not know most of the resistance-causing mutations and mechanisms for some antibiotics. To close this gap, there are a series of ongoing efforts by consortiums like ReSeqTB and CRyPTIC where

tens of thousands of isolates are being phenotyped and genotyped in order to obtain a comprehensive mutation database with the overarching aim to develop new diagnostic assays with maximum specificity and sensitivity. However, we still need more than mutation databases to effectively combat drug resistance. Firstly, it is impossible to predict the phenotype for a mutation never seen before. For this reason, it is very difficult to accurately predict resistance to newly-licensed antibiotics. Additionally, an approach that prioritizes diagnostic mutations generally provides very little information on other mutations that contribute to the resistant phenotype but are normally overlooked because their clinical effect is small or they are in genes not known to be associated with resistance. Finally, we need extensive insight on the genetic architecture of resistance, and especially on any changes that can increase sensitivity to the antibiotic. This is important as this information could be used to find companion drugs that potentiate the action of antibiotics or that prevent or even reverse resistance [59].

One way to unveil the genetic basis of resistance is by means of functional genomics, like transposon mutagenesis approaches. This technique involves the genetic alteration of every gene in the genome for explicit genotype-phenotype associations [60,61], thus revealing more genetic determinants than regular association studies do. This approach successfully overcomes the shortcomings of genetic association studies: it can be used in a prospective way, as it involves the systematic generation and testing of resistant mutants; it can detect both genes with large and small effects on resistance; and it explicitly

detects genes that increase sensitivity when disrupted, thus indicating which genes are most promising for treatments to prevent or reverse the evolution of resistance. However, transposon mutagenesis alters the gene by disrupting it, highly informative about the biology of resistance but limited in clinical explanation potential as most types of mutations found in clinical resistance of *M. tuberculosis* are single nucleotide polymorphisms. Conversely, the low diversity of the MTBC, its clonality and the fact that clinical resistance is encoded in the chromosome makes *Mycobacterium tuberculosis* amenable for phylogenetic association tests [62], which can determine which mutations are associated with resistance in the bacterial phylogeny and are thus clinically relevant.

In this chapter, we provide a synergistic approach that uses functional genomics and phylogenetic inference from clinical data to provide an in-depth picture of resistance to the first-line antibiotic isoniazid. Isoniazid is a well-studied drug, yet we are still unable to determine the causal mutation in around 6% of resistant strains [62], although some researchers have reported up to 25% in certain settings [63]. Here, we systematically determine the effect on isoniazid resistance of every non-essential gene in the tuberculosis genome using transposon sequencing and afterwards we use clinical data to find out which of those genes are more likely to harbor resistance mutations. We successfully find novel regions associated with increased resistance *in vitro*, determine two major resistance pathways for the mode of action of the antibiotic and identify novel associated regions to clinical resistance not described

before. We believe this approach will help uncover the resistance determinants for poorly-studied antibiotics, as well as deepen our understanding of resistance emergence, spread and evolution.

METHODS

Strains, media and culture conditions

We used *Mycobacterium tuberculosis* strain H37Rv, kindly supplied by Darío García de Viedma. Bacteria were cultured at 37ºC in Middlebrook 7H9 supplemented with 10% ADC (both from BD) and 0.05% Tween 80 (Difco) for liquid cultures and in Middlebrook 7H10 supplemented with 10% OADC (both from BD) for solid cultures. All experiments were conducted in a BSL3 laboratory using a biosafety cabinet.

Mutant pool generation and selection experiment

We generated a mutant pool using the protocol by Long *et al* [60]. Briefly, we harvested 100 mL *Mycobacterium tuberculosis* H37Rv culture and washed twice with MP buffer to remove the Tween. We then transduced the bacteria with 10^{11} pfu phiMycomarT7 for 20 hours in a total volume of 10 mL. Afterwards, we pelleted the cells and washed away excess phage twice with PBS-Tween 80. Finally, we plated the transduced bacteria in three 25x25 square plates (Corning) containing Middlebrook 7H10 media supplemented with OADC, 0.05% Tween 80 and 20 ug/mL kanamycin (Panreac). After three weeks, mutant bacteria

were scraped from the agar, homogenized in liquid media and stored at -80ºC. For each experiment, we used a starter culture of the pool at OD 0.8 - 1.0 to inoculate two 100-mL roller bottles in parallel with approximately 10^7 bacteria each. One of the bottles contained either 0.18 or 0.20 ug/mL isoniazid (Panreac) while the other contained just plain media and served as a control. We allowed the bacteria to grow for about 13 generations and stored final populations at -80ºC.

DNA extraction

Mycobacterial cultures were pelleted by centrifugation and resuspended in 500 µL TE buffer. Following inactivation by heat at 80ºC for 1 hour, lysozyme (50 µL of a 10 mg/mL stock) was added and samples were incubated overnight at 37ºC. Then 50 µL of proteinase K (10 mg/mL stock solution) were added, incubating for 1 h at 60°C with shaking in a thermomixer. After this, 100 µL 5 M NaCl and 100 µL 10% CTAB were mixed by inverting, samples were frozen for 15 min at –80°C and re-incubated at 60°C for 15 min with shaking. Once cooled, 700 µL chloroform–isoamyl alcohol (24:1) were mixed in, yielding a white, homogenous solution. Samples were centrifuged, transferred into 700 µL cold isopropanol and left at -20ºC overnight. Then, they were pelleted and washed with 70% ethanol and dried in a speed vacuum concentrator for 10 min. Finally, DNA was resuspended in 50 µL TE and its concentration was determined with a QuBit 3.0 Fluorometer (Thermo Fisher Scientific). Also, the amount of

contaminating phage DNA was estimated by PCR to ensure it was low and would not affect sequencing results.

Library preparation and sequencing

We followed the protocol described by Long *et al [60]* with some modifications. After extraction, samples were quantified using a QuBit 3.0 Fluorometer (Thermo Fisher Scientific). Then, 50 µl of each DNA sample were transferred to Covaris tubes, centrifuged and fragmented to an approximate size of 550 bp using standard settings (Illumina TruSeq Library Prep Reference Guide). Samples were size-selected using NucleoMag NGS Clean-up and Size Select (Macherey-Nagel) to 50 uL final volume. The DNA end-repair was performed using NEBNext End Repair Module (New England BioLabs), and dA-tailing was achieved with NEBNext dA-tailing Module (New England BioLabs), both as per the manufacturer's instructions. Next, a stock of barcoded adapters was prepared by mixing 20 µl of 50 µM oligonucleotides (detailed in Supplementary Table 1) in a final concentration of 2 µM $MgCl_2$, heating to 93ºC for 10 min and reducing the temperature by 3ºC/cycle over 2 hours until reaching 20ºC. These double-stranded adapters were then ligated to the purified dA-tailed DNA using T4 DNA ligase from NEBNext Quick Ligation Module (New England BioLabs).

After another purification step, we performed a PCR to selectively amplify the transposon-chromosomal junctions using a pair of primers specific to the end of the transposon

and ligated adapter (Supplementary Table 1) with the following parameters: initial denaturation at 98ºC for 5 min, twenty cycles of denaturation at 98ºC for 20 sec, annealing at 65ºC for 15 sec and extension at 72ºC for 30 sec, with a final extension at 72ºC for 3 min. We size-selected ~500bp products using magnetics beads, and a standard 8-cycle indexing PCR introduced the Illumina indexes required for sequencing. Final libraries were validated on a Bioanalyzer DNA chip (Agilent Technologies) to verify size and then quantified again using Qubit. Libraries were sequenced on the Illumina NextSeq 550 platform using the High Output v2 kit (150 cycles), producing an average of 35 million raw paired reads per sample with a good quality distribution.

Bioinformatic analysis

Quality control of sequencing files was performed using FASTQC, after which they underwent quality trimming by PRINSEQ. The selected criteria for keeping sequences was a mean Phred quality score of 20 in a 20 bp sliding window. Next we processed the cleaned sequences by means of a custom Python script that served two purposes: for every read pair, it first scanned the beginning of the forward read looking for a 'TGTTA' motif that marked the start of the transposon insertion and cut the sequence at the TA site; second, it looked for the random barcode in the reverse read, cut it from the sequence, and appended it to the header as a comment if it passed a structure check.

We then mapped the reads to the *M. tuberculosis* H37Rv reference genome (NC_000962.3) using BWA with default parameters but keeping header comments in the resulting SAM files. It was important to detect PCR duplicates and remove them to correctly estimate the proportion of each mutant in the original pool. This step was again performed by a custom Python script that used the barcode, strand, mapping coordinate and fragment length information to define unique reads. After removing PCR duplicates, the insertion count final list for each sample was generated, with coordinates determined by the mapping point of the 'TA' site. At this point we were able to determine a ~78% insertion density and thus the high quality of our mutant pool.

We developed our own pipeline to analyze the Tn-seq data. Our idea was to determine which genes consistently alter isoniazid resistance and current approaches work better when the size of the effect is large. First, we normalized insertion counts by 40% trimmed mean and generated a T0 library that was sampled 1000 times to obtain z-scores that represented the standardized deviation of each site from its expected insertion count. We also generated two new libraries, R1 and R2, by subtracting z-scores from controls to their treatment's counterparts. A sliding window analysis was designed to evaluate the significance of inserted genomic features with the aim of detecting zones with equally consistent insertion changes. Each window containing between 6 and 10 'TA' sites underwent a Wilcoxon test that determined if the region was inserted more or less than expected by chance. We only

considered windows that fell within the limits of annotated genomic features and eliminated the ones that spanned more than one. In terms of fitness, a window above the expected insertion level was defined as an increased resistance window, and one below the expected insertion level was considered an increased sensitivity window. A multitest correction was applied to all p-values after the analysis, obtaining a final q-value for each window. Then we established a call for each 'TA' site by judging the windows in which it appeared. If half of its windows got a significant call, that call was applied to the 'TA' site, otherwise it was left as non-significant. Non-inserted 'TA' sites in the T0 libraries were not called and were annotated as 'not evaluated'. Finally, 'TA' sites were assigned to genomic features to give these a call. Features containing at least 70% of a certain significant call were given that same call if their median z-score was either positive or negative and their ranking among all features was above or below 50% respectively. We selected all significant calls from either of the two concentrations tested and a final call for each gene and intergenic region was obtained following this pipeline. Scripts necessary to perform the analyses are available at https://gitlab.com/tbgenomicsunit.

To determine how reproducible our results were, we compared them with published data from Xu *et al [64]* and we found approximately five times more resistance-altering features than they did, suggesting that our results might provide a more detailed picture of the genetic architecture of isoniazid resistance. This is probably due to a combination of stronger selective conditions, an increased number of

inserted sites and the fact that our analytical method tests for consistency independently of the size of the effect. Coincidence between our and their sets of resistance-altering genes was significantly higher than expected (chi-squared test, chi = 297.28, p-value < 0.0001), with particularly strong association in the direction of the effect (Supplementary Figure 1A). Finally, we checked if an effect in the same direction was found for all our resistance-altering regions even if it was not significant. Indeed, our resistance-increasing features had significantly higher resistance than non-significant ones (Wilcoxon test, z = 8.93, p-value < 0.0001), while sensitivity-increasing features showed lower resistance (Wilcoxon test, z = 11.54, p-value < 0.0001; Supplementary Figure 1B).

Clinical dataset phylogeny and ancestral state reconstruction

A 4,763-strain dataset consisting of different worldwide clinical isolates was constructed from publicly available databases. We downloaded all FASTQ files from NCBI using their fastq-dump tool, mapped them to a predicted *M. tuberculosis* ancestor reference and called SNPs using VarScan 2. An alignment of all homozygous variable positions among isolates was generated and a phylogenetic tree was constructed using FastTree 2.1.

We then proceeded to reconstruct the ancestral state of every polymorphism using PAUP 4.0a158 with a custom weight matrix that punished reversions with a 10X multiplier

(see Supplementary Text 1 for the assumptions block including this matrix). As we had more positions than the program could compute at a time, we had to split the alignment into four 60K-SNP pieces before building the NEXUS files as input for the program. We obtained a list of changes associated with the tree's nodes and this output was parsed using a custom Python script, yielding a summary table of changes that was further processed to add more information about each change. The final table contains for each SNP event the tree node in which it occurs, associated genomic feature, translational impact, homoplasy and antibiotic resistance details.

Phylogenetic association test

We used R to perform an association test linking particular SNPs occurring in clinical settings to antibiotic resistance. We began by defining susceptible and resistant tree branches according to the absence or presence, respectively, of an antibiotic-resistance associated mutation from a high-confidence resistance mutations list based on PhyResSE [46] and ReSeqTB. Along with isoniazid we considered all first-line antibiotics, because although isoniazid resistance tends to appear first [65], resistance mutations are not always detected. Furthermore, resistant strains can accumulate non-specific or low-level rare mutations and we are also interested in those mutations. In any case, we repeated the analysis only with isoniazid resistance mutations and the results were very similar. After we determined "resistant" and "sensitive" subtrees of the

phylogeny, we eliminated all diagnostic mutations and calculated the number of all mutations, non-synonymous mutations and homoplasies for each genomic feature in resistant subtrees. We then determined the expected number of mutations in each category by resampling the assigned branch for each mutation 10,000 times and recalculating the numbers for each category. We determined that one particular genomic feature was phylogenetically associated with resistance if it ranked higher than 9,500 of the 10,000 samplings in any of the three categories.

We used the entire list of mutations for all first and second-line antibiotics for several reasons: i) we expect most strains resistant to other antibiotics to be resistant to isoniazid as well since not all isoniazid resistance mutations are known and resistance mutations in tuberculosis tend to appear in a stepwise fashion with isoniazid resistance mutations being one of the first; ii) even in the cases where no proper isoniazid resistance mutation has occurred, other low-level resistance mutations may have been acquired and they are also relevant to the evolution of resistance; and iii) as mutations that confer resistance to different antibiotics are highly correlated due to the nature of the treatment, it is very difficult to disentangle one from the other and it is better to study resistance as a whole independently of the specific antibiotic. All diagnostic mutations used to mark the onset of resistance in the phylogeny were subsequently eliminated from analysis.

Resistance prediction in the clinical dataset

We used 444 isoniazid-resistant strains from the Cryptic dataset [58]: 362 with typical isoniazid resistance mutations and 82 without any known isoniazid resistance mutations. We analyzed the raw sequence data using our lab's validated pipeline (available at https://gitlab.com/tbgenomicsunit/ThePipeline) and determined all the single nucleotide mutations above a 10% frequency for each strain in the sample. We centered our analysis on "rare" mutations (mutations appearing in only 1 to 3 strains) and all diagnostic resistance mutations were excluded. We first determined the number of strains in each group (resistant with typical mutations or resistant with no clear mutations) that contained at least one mutation in any of the candidate genes and then we compared this number to the results we obtained using 1000 random subsets of genes with the same number of features as our candidates list (PE/PPE, phage and repetitive sequences excluded [66]). The relative position of the sensitivity obtained with our candidate list revealed how relevant those genes are to clinical resistance. To build the ROC curve, we similarly analyzed 188 randomly sampled pan-susceptible strains from the CRypTIC consortium and added each gene sequentially in descending order of specificity. When two genes had the same specificity, the one with the highest sensitivity took precedence.

Candidates validation using BCG mutants

We selected 30 BCG Danish mutants (24 candidate genes plus 6 controls) from the BCCM/ITM Mycobacteria Collection, which were regrown and shipped in 7H11 solid medium. Upon reception, bacteria were collected and further amplified in 7H9 liquid medium to generate stocks for MIC determination experiments. We set up 96-well microtiter plates (using 7H9-OADC) with two-fold serial dilutions of isoniazid, with concentrations ranging from 0.015 to 8 µg/mL and leaving two wells without antibiotic to have a growth control. BCG mutants were inoculated by rows, adding 10^4 bacteria per well, along with a wild-type BCG Danish strain as internal control in each plate. After a 7-day incubation at 37ºC, 20 µL of 0.02% resazurin were added to each well, and plates were allowed to incubate for 2 more days. At the 24h and 48h time points, resazurin color change was visually assessed and a 50 µL aliquot was inactivated with 50 µL of 4% paraformaldehyde and placed in a black plate for fluorescence reading. Resazurin reduction was assessed by measuring fluorescence in a Tecan Infinite M Plex plate reader, allowing for a precise estimation of IC50 and IC90 values. Briefly, we subtracted the negative controls and re-calculated the fluorescence values as relatives to their respective growth control. We determined the slope of the steepest part of the inhibition curve and used that estimation to determine the concentrations at which inhibition was exactly 50% and 90% relative to the growth control.

Statistics and Reproducibility

We performed all statistical analyses using R v3.3.3 and publicly available databases (cited in the main text). The specific test performed in each case (Fisher's exact test, Wilcoxon test or custom) is indicated throughout the text. The TnSeq selection experiment was performed twice and the same pool was tested in two independent blocks, in each block we seeded one control and one experimental culture from the same starter culture. Depth of sequencing for the TnSeq experiment was designed so that there would be around 50 reads per insertion to ensure adequate coverage of all insertion sites. Experimental cultures were inoculated with 10 million bacteria, ensuring that on average each insertion mutant would have 100 copies. We allowed bacteria to grow for 13 generations because that is enough to see differences in growth for the different mutants. We compared with similar results from other groups and found good agreement. We tested 24 insertion mutants for candidate genes for confirmation, comparing them with 6 control mutants. Insertion mutants were tested in randomized blocks and with a quality control strain.

All analyses can be reproduced using datasets and code available at https://gitlab.com/tbgenomicsunit.

RESULTS

Functional genomics allows for detection of resistance-associated genomic regions

We generated a highly saturated *M. tuberculosis* H37Rv pool with over 100,000 different transposon-insertion mutants following the protocol by Long *et al [60].* Using transposon sequencing (TnSeq), we found 58,389 out of 74,603 possible insertion sites had at least one read, meaning a saturation of 78%. For comparison, a systematic study with 14 independent pools found saturations in the range of 42% to 64% and a combined saturation of 84.3% [67], meaning that our pool is highly saturated. Additionally, as many as 43% of our non-inserted sites and only 0.03% of our inserted sites were in regions described as essential in that study. The pool also showed a 10-fold increase in the frequency of bacteria resistant to isoniazid compared to the original clone (Supplementary Figure 2A).

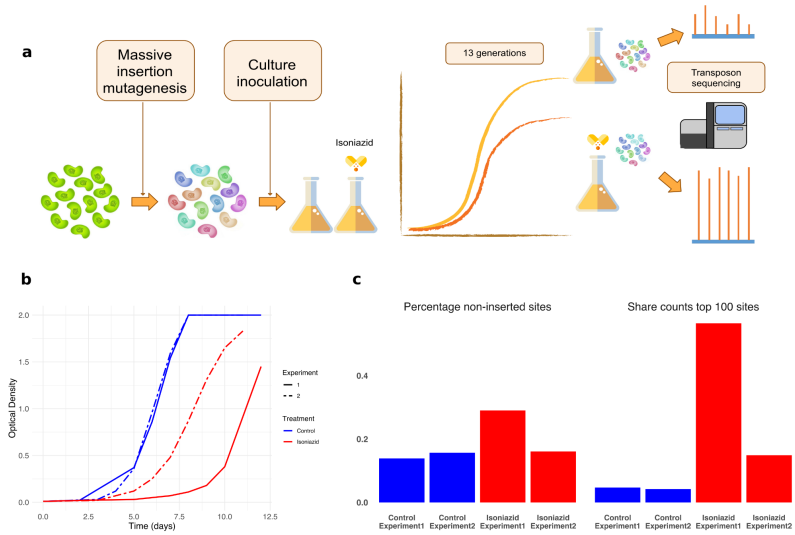


Figure 2. Transposon sequencing detects changes in response to isoniazid. a) Design of the experiment. Parallel antibiotic-containing and antibiotic-free cultures were inoculated with a saturated insertion mutant pool. After ~13 generations, bacterial DNA was extracted and sequenced to determine the relative abundance of each mutant. **b**) Optical density of the different cultures throughout the experiment. Graph shows that isoniazid partially inhibits bacterial growth. **c**) Isoniazid-containing cultures show strong enrichment of a fraction of insertions indicating a selective advantage relative to the bulk of the population, showing that those cultures experienced higher levels of selection (blue = control, red = isoniazid experiment).

The pool was tested in duplicate with a subinhibitory dose of isoniazid close to the IC50 for 13 generations (Figure 2A). We expected this specific dose of isoniazid to provide intermediate levels of selection and to maximize the number of genomic features detected. Optical density

measurements showed that isoniazid was partially inhibiting bacterial growth (Figure 2B). In the presence of isoniazid the proportion of isoniazid-resistant bacteria increased 100-fold to 1000-fold, while control cultures showed no change (Supplementary Figure 2B).

We determined the frequencies of the different insertion mutants in all four experimental populations using TnSeq. Isoniazid-treated populations had a higher proportion of sites with null frequency and the top 100 sites comprised a larger share of the total counts (Figure 2C). We transformed the normalized data into standardized fitness measurements, which can be directly compared between populations. We defined resistance as the net change in fitness in the presence of the antibiotic and calculated it as the difference between fitness in the presence and absence of the antibiotic for each insertion site (Supplementary Data 1). Insertion mutants for *katG*, the gene most frequently involved in isoniazid resistance, were disproportionately overrepresented in antibiotic-treated populations and thus displayed very high resistance values. All these results show that the selection step had the intended effect.

It is important to note that transposon libraries have limitations as they only allow us to study the effect of gene disruptions. This has two main consequences: i) we cannot study essential genes because they cannot tolerate insertion, and ii) we cannot observe the effect of more subtle genetic changes such as single nucleotide mutations. To overcome these limitations, we used two main approaches: first, we used functional and pathway analysis to understand

which portions of bacterial metabolism were involved in isoniazid resistance and second, we used phylogenetic association to determine which genes were accumulating mutations in clinical settings.

We analyzed all insertion sites with an annotation-aware sliding window approach to find changes in resistance that were consistent over stretches of the genome independent of the size of the effect. We detected a total of 555 genes and intergenic regions that alter isoniazid sensitivity when disrupted (resistance-altering genomic features). Of those regions, 411 were associated with increased resistance while 144 were associated with increased sensitivity (resistance-increasing and sensitivity-increasing features, respectively). Figure 3 depicts these regions ordered by their fitness in the presence of the antibiotic. Given that fitness in the presence of the antibiotic is the primary driver of the resistance phenotype, we observed resistance was split into two groups according to whether the genes conferred increased sensitivity or resistance when disrupted. Multiple features showed a significant change in resistance, implying that they could, in theory, confer clinically relevant resistance in vivo when mutated. Features that could be tested but showed no significant effect were considered non-associated features. Our method did not allow us to test regions that had fewer than 6 inserted TA sites, although not all of these regions were essential. Intergenic regions tend to be small and often harbor regulatory sequences for the genes they precede but were massively overrepresented in non-evaluated features. Thus, we can assume that intergenic regions preceding candidate genes and with resistance scores that show the

same sign as those in the gene are probably associated with resistance. Using this approach, we found 126 additional probable resistance-altering features, 82 of which were associated with increased resistance (Figure 4A). Among resistance-altering features we found several regions known to be associated with clinical isoniazid resistance, such as *katG*, *ahpC* and its promoter region, and *fabG1* and its promoter region [68]. Our results were also consistent with similar data from Xu *et al* [64], further confirming that resistance-altering features are associated with isoniazid resistance.

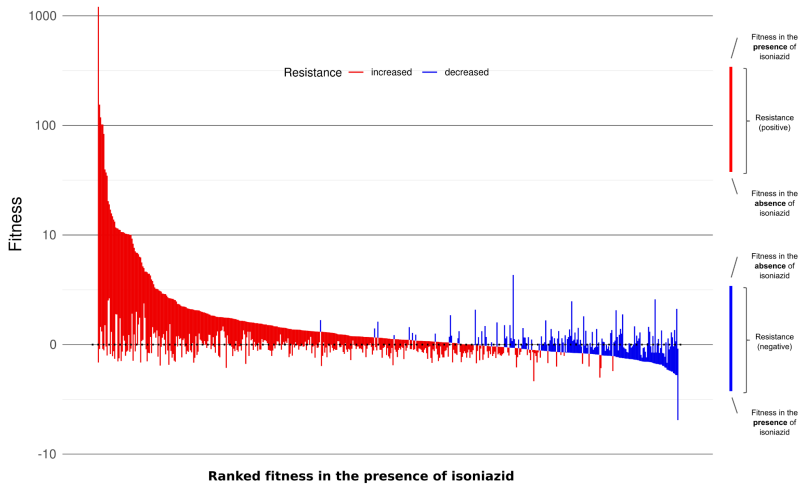


Figure 3. Isoniazid resistance is associated with multiple genomic regions. Median fitness in the presence and absence of isoniazid and median resistance for all resistance-altering genes, ordered in decreasing fitness in the presence of the antibiotic (red = increased resistance, blue = decreased resistance). Most genes that increased resistance when disrupted also had a higher fitness in the presence of the antibiotic.

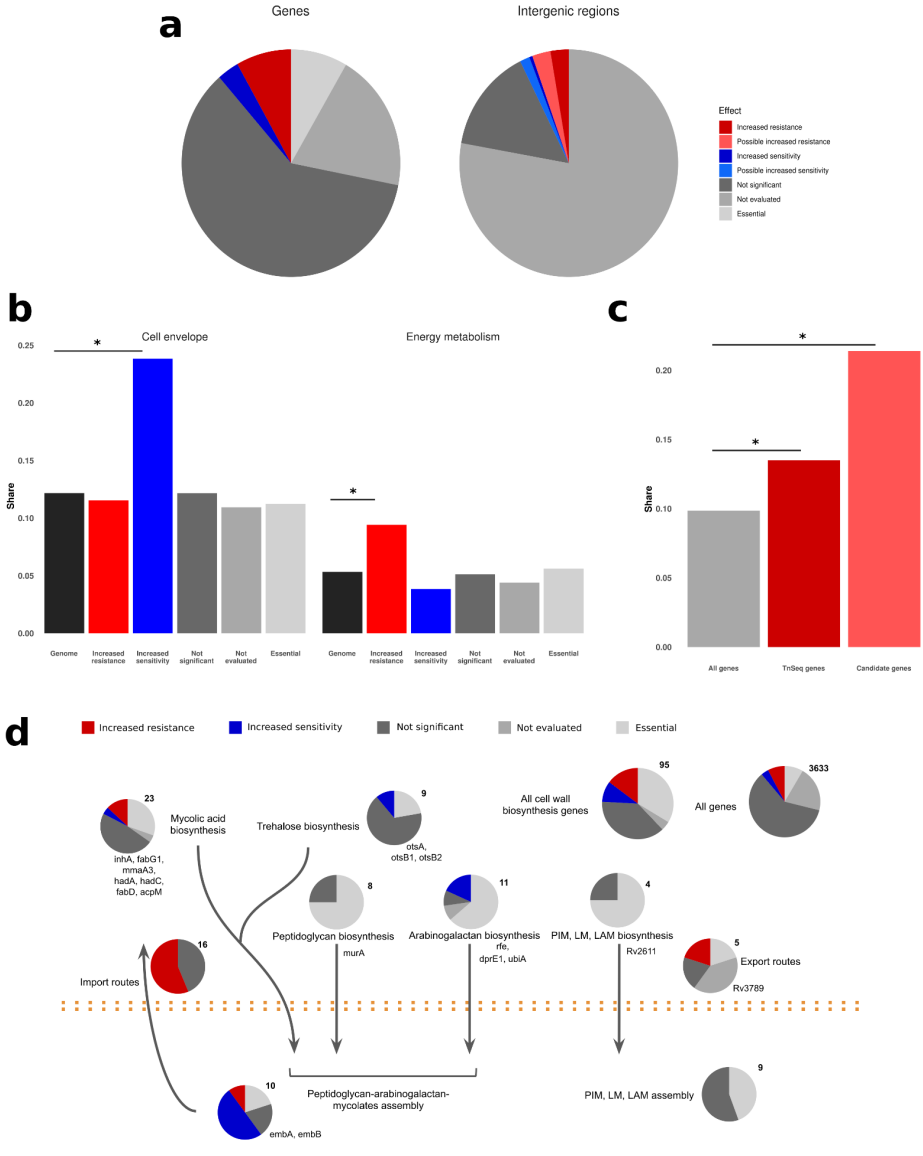

a
Genes
Intergenic regions
Effect
Increased resistance
Possible increased resistance
Increased sensitivity
Possible increased sensitivity
Not significant
Not evaluated
Essential
b
Cell envelope
Energy metabolism
Share
Genome
Increased resistance
Increased sensitivity
Not significant
Not evaluated
Essential
c
Share
All genes
TnSeq genes
Candidate genes
d
Increased resistance
Increased sensitivity
Not significant
Not evaluated
Essential
Mycolic acid biosynthesis
inhA, fabG1, mmaA3, hadA, hadC, fabD, acpM
Trehalose biosynthesis
otsA, otsB1, otsB2
All cell wall biosynthesis genes
All genes
Peptidoglycan biosynthesis
murA
Arabinogalactan biosynthesis
rfe, dprE1, ubiA
PIM, LM, LAM biosynthesis
Rv2611
Import routes
Export routes
Rv3789
Peptidoglycan-arabinogalactan-mycolates assembly
embA, embB
PIM, LM, LAM assembly

Figure 4. Genes associated with isoniazid resistance follow definite functional patterns. a) Classification of all genes and intergenic regions in the M. tuberculosis genome according to the effect of insertions on resistance. Essential genes were obtained from DeJesus *et al [67].* **b**) Genes associated with increased sensitivity were enriched in cell envelope genes (n= 130), while those associated with increased resistance were enriched in energy metabolism genes (n = 329). **c**) Oxidoreductases were even more enriched in the final candidates than they were in the genes detected with functional genomics (n = 57 and 459). **d**) Cell wall biosynthesis genes were enriched in features associated to resistance both functionally and phylogenetically to resistance.

Genes associated with altered resistance follow definite functional patterns

We noticed that resistance-altering features tended to group together on the genome. One explanation for this observation is that functionally related genes sometimes cluster in operons, so they can be transcribed together. To test this we obtained the H37Rv operon annotations from BioCyc [69] and used a sampling approach, finding that significant genes clustered in operons more than expected by chance in 100,000 random samples (388 transcription clusters versus at least 395 in the simulations, p-value < 10^{-5}). Particularly, two large operons were nearly entirely

comprised of resistance-altering features (Fisher's exact test, p-value < 0.01): *nuo* and *mce1*. This proves that these features are not randomly distributed around the genome but show at least some functional relatedness to one another.

We further hypothesized that resistance depends on specific cellular processes. To test this at the most general level, we compared the relative shares of both versions of the TubercuList functional categories [70,71] in the resistance-altering features with their global shares (Figure 4B). We found that although all categories were represented, resistance-increasing features were enriched in energy metabolism (Fisher's exact test, p-value = 0.0029), intermediary metabolism and respiration (Fisher's exact test, p-value = 0.012) and virulence genes (Fisher's exact test, p-value = 0.017) while sensitivity-increasing features were enriched in cell envelope (Fisher's exact test, p-value = 0.0002), and cell wall and cell processes genes (Fisher's exact test, p-value = 0.00001).

To further understand the genetic architecture of isoniazid resistance, we conducted a pathway enrichment analysis using data from both KEGG [72] and BioCyc using resampling. Results were consistent using both databases, and revealed that pathways associated with the mycolic acids and cell wall biosynthesis were significantly enriched among sensitivity-increasing features (1000 samples, p-value < 0.05). This result is not surprising given that isoniazid interferes with the biosynthesis of mycolic acids, one of the main components of the bacterial envelope,

although the main mycolate biosynthesis pathway itself was not significantly enriched (p-value = 0.09). Finally, we collated a dataset with all cell wall biosynthesis genes from published sources [73] and confirmed that sensitivity-increasing features were enriched in those genes (Fisher's exact test, p-value = 0.0001).

Resistance-increasing features were enriched as well (Fisher's exact test, p-value = 0.0069), but that depended mainly on the *mce1* operon as shown in Figure 4D. In contrast, sensitivity-increasing genes can be found all over the cell wall biosynthesis pathway, which demonstrates the central role of the cell envelope in intrinsic resistance and in isoniazid resistance in particular.

Among resistance-increasing genes, oxidative metabolism was enriched as well (p-value < 0.001) with the electron transport chain as the most enriched pathway overall (p-value < 0.001). Additionally, nicotinate and nicotinamide metabolism was enriched for both resistance-increasing and sensitivity-increasing genes (p-value < 0.02), suggesting that NADH metabolism might be of importance. To ensure that the enrichment in oxidative metabolism pathways did not depend entirely on the *nuo* operon, we mined the genomic annotation for H37Rv from the NCBI using the terms: *oxidoreductase, oxidase, reductase, redox, peroxidase, dehydrogenase, NAD, NADH, NADP, NADPH*, and we flagged any feature that contained any of those terms in their name or function description as redox-associated. We found that resistance-increasing genes were significantly enriched in oxidative metabolism genes (Fisher's exact test, p-value = 0.03; Figure 4C). All

these observations point to redox metabolism having a role in isoniazid resistance.

A phylogenetic association test identifies candidate regions associated with clinical resistance

So far, we have successfully linked resistance-altering features to isoniazid resistance at an in vitro and functional level, but we still do not know what their importance in a clinical setting is. We used a phylogenetic test to identify regions associated with resistance in clinical strains. We first set out to obtain a phylogeny that encompassed tuberculosis strain variability using 4,762 globally-distributed, published *M. tuberculosis* complex genomes with around 240,000 polymorphic sites (Supplementary Figure 3). The dataset included 32% of strains resistant to at least one drug. We reconstructed the evolutionary history for each variable site inferring how many substitution events had occurred and where in the phylogeny they had taken place. Finally, we sought to determine which regions in the whole genome are more strongly associated with resistance by calculating the PhyC parameter [74] which acts as an association test and measures the degree of mutation accumulation for a particular gene in predetermined branches of the phylogeny. We first determined in which specific branches an antibiotic resistance mutation had occurred using a comprehensive list of resistance mutations based on PhyResSE [46] and

ReSeqTB. We then tested which mutations tend to appear in resistant versus susceptible subtrees by random sampling.

We identified 511 regions significantly associated with antibiotic resistance in the phylogeny (p-value < 0.05). Most of the top scoring regions were already known resistance genes for first- and second-line antibiotics, which shows that there are still many unidentified resistance mutations in those genes. Some other top scoring regions are known to be associated with compensatory mutations which were also expected to appear after resistance mutations to compensate for their cost. Thus, phylogenetic association does a good job in identifying genes known to be relevant to antibiotic resistance.

Novel isoniazid resistance determinants were identified by combining functional genomics and phylogenetic association

We combined our functional data on isoniazid resistance with phylogenetic convergence results to look for isoniazid resistance candidate genes. Our reasoning was that if resistance-altering regions from our TnSeq experiment accumulated changes specifically in association with resistance mutations then they would probably be involved in the evolution of isoniazid resistance. We found 57 resistance-altering features that had more mutations occuring in resistant subtrees than expected (Table 1). Four of them were well known isoniazid resistance determinants

or associated regions (*katG*, *ahpC* and its promoter region and the promoter region of *fabG1*) which still showed association even though diagnostic mutations had already been removed, thus confirming that the catalog of mutations conferring isoniazid resistance in those features is far from complete. This finding is in agreement with the frequent identification of unidentified, but rare, mutations in *katG* associated with isoniazid resistance in different settings [36,75].

Table 1: Candidate genes. For intergenic regions, the neighboring gene most probable to be regulated by the region is given. IR: increased resistance; IR*: probable increased resistance; IS: increased sensitivity; IS* probable increased sensitivity.

Rv number	Call	Name or associated gene	Function
Rv0001	IR	dnaA	chromosomal replication initiator protein, regulates chromosomal replication
Rv0010c	IS	Rv0010c	conserved membrane protein
IG_Rv0020c_Rv0021c	IR	fhaA	conserved hypothetical protein, thought to be involved in signal transduction
Rv0134	IR	ephF	epoxide hydrolase, thought to be involved in detoxification reactions following oxidative damage to lipids
IG_Rv0237_Rv0238	IS*	Rv0238	transcriptional regulator, tetR-family

Rv0392c	IR	ndhA	membrane NADH dehydrogenase, transfer of electrons from NADH to the respiratory chain
Rv0450c	IR	mmpL4	transmembrane transport protein, thought to be involved in fatty acid transport
Rv0740	IR	Rv0740	conserved hypothetical protein
IG_Rv0767c_Rv0768	IR	aldA	aldehyde dehydrogenase NAD-dependent
Rv0994	IR	moeA1	molybdopterin biosynthesis protein
Rv1022	IS	lpqU	lipoprotein
Rv1053c	IR	Rv1053c	hypothetical protein
Rv1086	IS	Rv1086	short-chain Z-isoprenyl diphosphate synthase, catalyzes the first committed step in the synthesis of decaprenyl diphosphate, a molecule which has a central role in the biosynthesis of most features of the mycobacterial cell wall
Rv1194c	IR	Rv1194c	conserved hypothetical protein
IG_Rv1364c_Rv1365c	IR*	Rv1364c	conserved hypothetical protein
IG_Rv1482c_Rv1483	IS*	fabG1	3-oxoacyl-[acyl-carrier protein] reductase, involved in the fatty acid biosynthesis pathway (first reduction step, mycolic acid biosynthesis). Secondary isoniazid resistance gene
Rv1504c	IR	Rv1504c	conserved hypothetical protein

Rv1512	IR	epiA	nucleotide-sugar epimerase
Rv1692	IR	Rv1692	phosphatase
Rv1767	IR	Rv1767	conserved hypothetical protein
IG_Rv1773c_Rv1774	IR*	Rv1773c	transcriptional regulator
Rv1780	IR	Rv1780	conserved hypothetical protein
Rv1830	IR	Rv1830	conserved hypothetical protein
Rv1836c	IS	Rv1836c	conserved hypothetical protein
IG_Rv1843c_Rv1844c	IR*	guaB1	inosine-5-monophosphate dehydrogenase
IG_Rv1900c_Rv1901	IS*	cinA	competence damage-inducible protein A
Rv1905c	IR	aao	D-amino acid oxidase
Rv1908c	IR	katG	catalase-peroxidase-peroxynitritase T, main isoniazid resistance gene
Rv1928c	IR	Rv1928c	short-chain type dehydrogenase/reductase
Rv2021c	IR	Rv2021c	transcriptional regulator
IG_Rv2208_Rv2209	IR	Rv2209	conserved membrane protein
Rv2214c	IR	ephD	short-chain type dehydrogenase, thought to be involved in detoxification reactions following oxidative damage to lipids.
Rv2333c	IR	stp	conserved membrane transport protein, involved in transport of drug across the membrane (export)
Rv2386c	IR	mbtI	isochorismate synthase, involved in mycobactin siderophore construction

IG_Rv2427A_Rv2428	IS*	ahpC	alkyl hydroperoxide reductase C protein, involved in oxidative stress response and secondary isoniazid resistance gene
Rv2428	IS	ahpC	alkyl hydroperoxide reductase C protein, involved in oxidative stress response and secondary isoniazid resistance gene
IG_Rv2560_Rv2561	IR	Rv2561	conserved hypothetical protein
IG_Rv2709_Rv2710	IR*	sigB	RNA polymerase sigma factor
Rv2710	IR	sigB	RNA polymerase sigma factor
Rv2886c	IR	Rv2886c	resolvase
Rv2994	IS	Rv2994	conserved membrane protein, could be involved in efflux system
Rv3154	IR	nuoJ	NADH dehydrogenase I chain J
IG_Rv3210c_Rv3211	IR	rhlE	ATP-dependent RNA helicase, has a helix-destabilizing activity
IG_Rv3213c_Rv3214	IS*	gpm2	phosphoglycerate mutase
Rv3229c	IS	desA3	linoleoyl-CoA desaturase, thought to be involved in lipid metabolism
IG_Rv3260c_Rv3261	IR	fbiA	F420 biosynthesis protein
Rv3268	IS	Rv3268	conserved hypothetical protein
Rv3272	IR	Rv3272	conserved hypothetical protein
Rv3278c	IS	Rv3278c	conserved membrane protein
Rv3490	IR	otsA	alpha, alpha-trehalose-phosphate synthase, involved in

			osmoregulatory trehalose biosynthesis
Rv3501c	IR	yrbE4A	hypothetical membrane protein
Rv3600c	IS	Rv3600c	conserved hypothetical protein
Rv3777	IR	Rv3777	oxidoreductase
Rv3788	IR	Rv3788	hypothetical protein
Rv3789	IR	Rv3789	conserved membrane protein
Rv3843c	IR	Rv3843c	conserved membrane protein
Rv3908	IR	mutT4	conserved hypothetical protein, possible mutator protein

These candidate resistance features are functionally diverse, showing the different ways in which *M. tuberculosis* can adapt to antibiotics (Table 1). Looking the genome annotation, the probable mechanisms operating here include increased efflux/decreased influx of the antibiotic (*mmpL4*, *stp*), altered transcriptional regulation (*sigB*), mycolic acids biosynthesis (*fabG1*), changes in NADH balance (*ndhA*, *nuoJ* among others) and increased mutagenesis due to changes in DNA repair (*mutT4*), among others. Some of the candidate features have no known function, which means that our strategy allows for discovery of new resistance determinants even if they are poorly characterized.

Only one of the candidate regions (*fabG1*) was involved in cell wall biosynthesis pathways but a further 16 out of the 95 genes in cell wall biosynthesis pathways

showed phylogenetic association with resistance, which is a significantly enriched fraction (Fisher's exact test, p-value < 0.016; Figure 4C). We do not have functional data for some of these regions as they are essential and cannot tolerate insertion, but mutations in these genes probably also affect isoniazid resistance as they are in the same pathway as the antibiotic target itself and our transposon sequencing data show that cell wall biosynthesis pathways are enriched in genes functionally associated with isoniazid resistance. These results highlight the importance of cell wall biosynthesis in isoniazid action and resistance, demonstrating that functional genomics is a powerful tool for discovering important pathways or even determining the mode of action.

We found that 8 out of 42 candidate genes were associated with redox metabolism. This result was mainly due to resistance-increasing genes, which accounted for 7 of the 8 redox genes and represented a significantly enriched fraction (Fisher's exact test, p-value = 0.024; Figure 4C). In contrast, genes phylogenetically associated with resistance as a whole were not enriched in redox genes (44 out of 377 genes; Fisher's exact test, p-value = 0.09). Additionally, 3 of the 15 candidate intergenic regions are next to the start of a redox gene. These results confirm that redox metabolism plays a clinically relevant role in the evolution of isoniazid resistance.

Finally, we demonstrated that the resistance phenotype inferred from the transposon sequencing assay was associated with the expected change in sensitivity by

determining the MICs for a representative sample of the candidate genes using the resazurin microdilution assay (REMA). For this experiment, we used BCG Danish insertion mutants from the BCCM/ITM Mycobacteria Collection. Our results showed that mutants with insertions in resistance-increasing features had higher MICs than mutants with insertions in either sensitivity-increasing (Wilcoxon test, p-value = 0.0346) or non-significant (Wilcoxon test, p-value = 0.0202) features (Figure 5). The result remained significant even when IC50s were used (Wilcoxon test, p-value = 0.0036 and p-value = 0.0202 respectively).

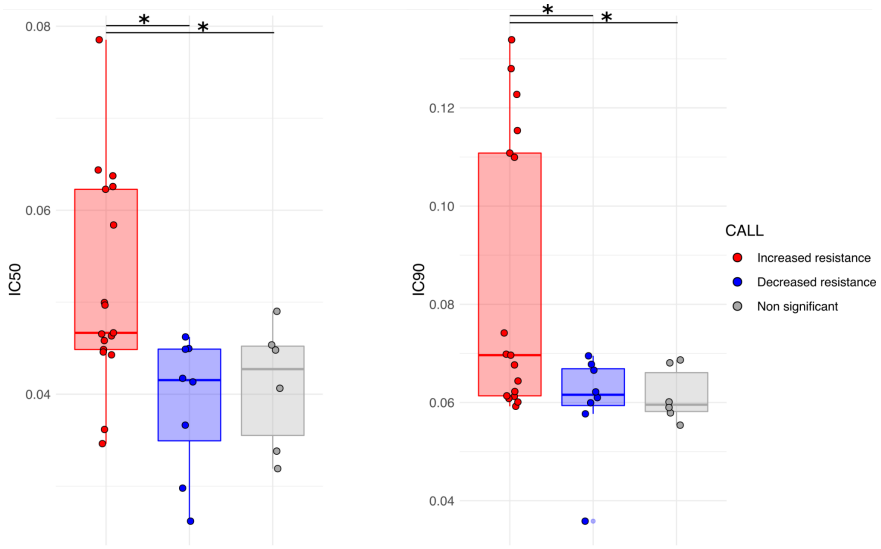


Figure 5. Boxplots of IC50 (left) and IC90 (right) values for a collection of candidate gene insertion mutants on a BCG background.

Novel isoniazid resistance determinants explain resistance in phenotypically resistant strains with no known mutation

Finally, we confirmed that mutations in candidate genes are relevant to clinical resistance. We reasoned that if our list of candidate genes plays a role in clinical resistance we should detect an increment in the sensitivity values to predict isoniazid resistance not explained by available databases. We looked at a selected dataset of strains obtained from the CRyPTIC consortium [58], enriched in isoniazid-resistant strains with no known resistance mutation (362 strains with known mutations, 82 with no known mutation). We found that the sensitivity of candidate genes was significantly greater than a random set of genes both for strains with no known mutations (sensitivity = 0.59, p-value = 0.019; Figure 6A) and for strains harboring well-known resistance mutations (sensitivity = 0.46, p-value = 0.027; Figure 6B). The result suggests that our list of genes is indeed involved in isoniazid resistance in one way or another. In both cases, any non-synonymous mutations in known resistance genes were also included as candidates, but they only contributed a small amount to the total sensitivity. By including rare mutations in our candidate genes list, we could increase global sensitivity from 93.1% to 94.6%, or from 97.1% to 98.9% if we omit genotypes with no clear prediction, further confirming that candidate genes in our analysis are relevant to isoniazid clinical resistance and could help explain uncommon resistant phenotypes.

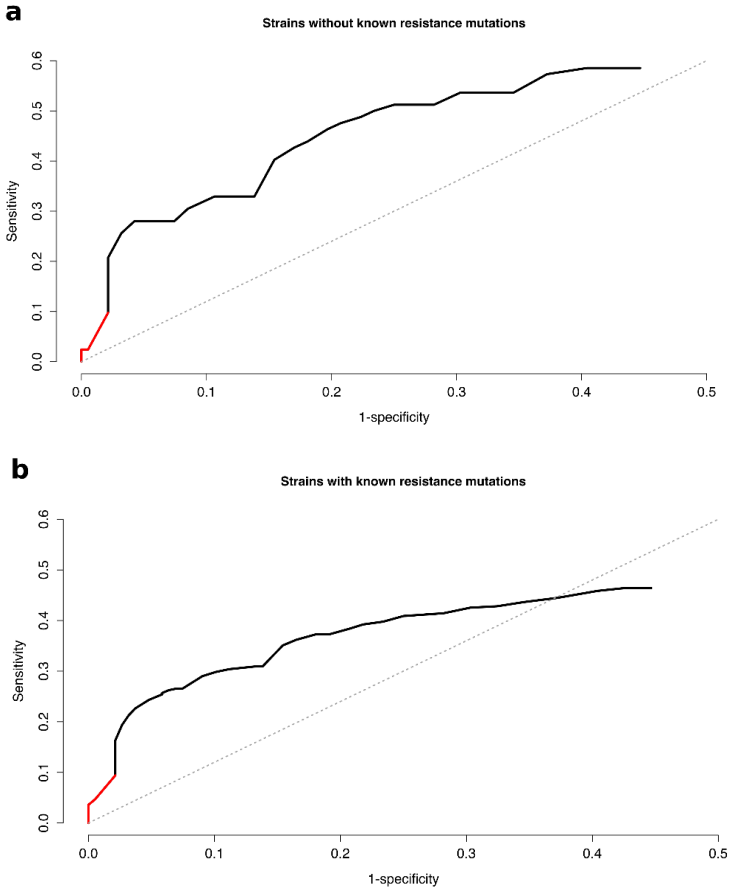


Figure 6. Novel isoniazid resistance determinants explain resistance in clinical strains. a) ROC curve for the dataset of strains without known resistance mutations. **b**) ROC curve for the dataset of strains with known resistance mutations. The red portion of the curve corresponds to candidate mutations in known resistance genes. The black portion corresponds to candidate mutations in novel drug resistance associated genes found in this study.

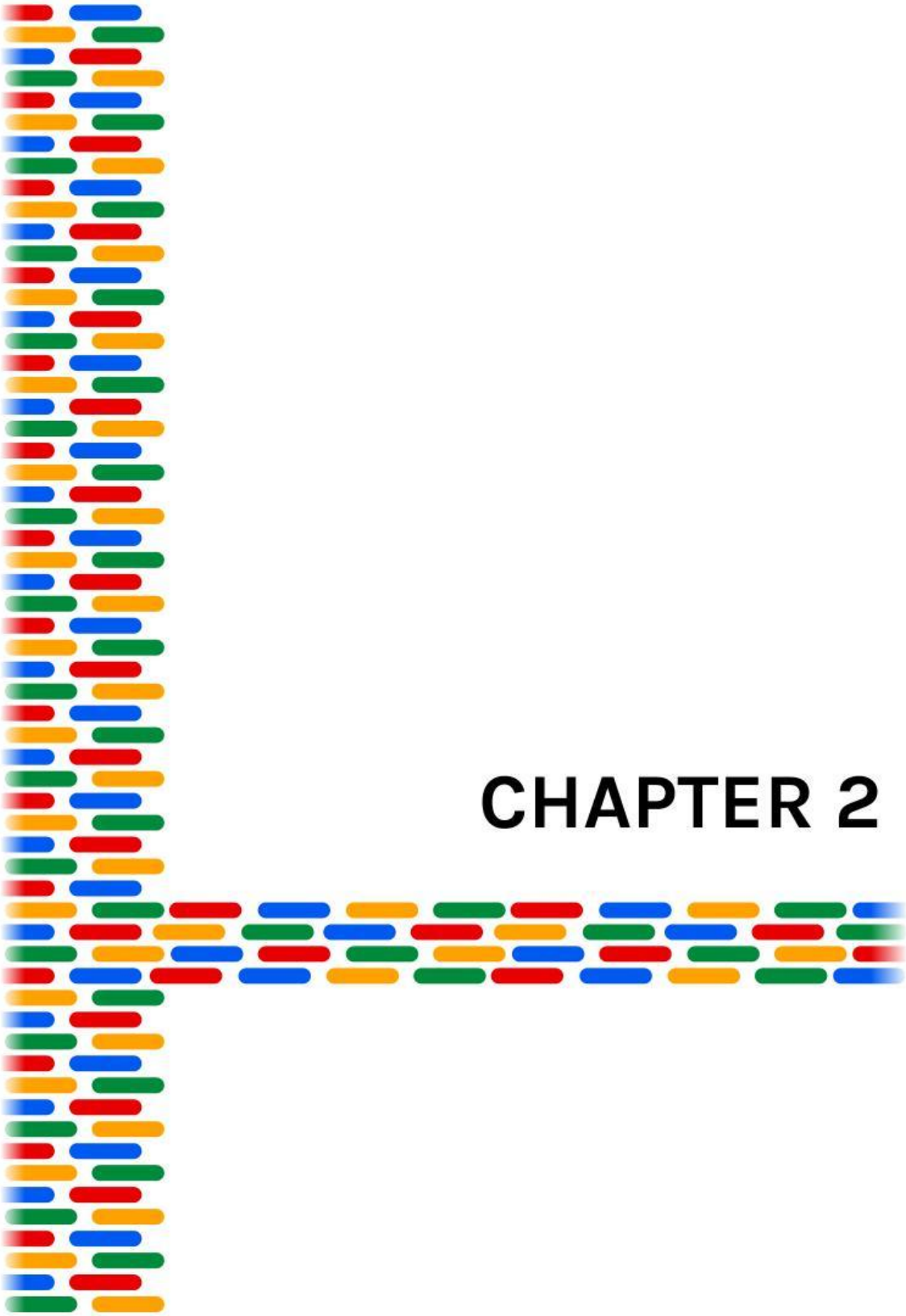
The resistance mechanisms involved seem to be multiple and diverse. For instance, two of the genes with the most mutations in resistant strains are *dnaA* and *mutT4*, which are involved in DNA replication and may affect the acquisition of resistance by increasing mutation rate or indirect mechanisms such as decreasing expression levels of katG [76]. Other interesting genes in the candidate set are those encoding epoxide hydrolases (*ephD* and *ephF*), which are involved in detoxification following oxidative damage to lipids. Also, we find many putative candidate mutations in genes involved in transport such as *mmpL4* and *stp* which can sometimes confer resistance to some antibiotics [77]. Finally, we find several genes with no known function that are accumulating many non-synonymous mutations in resistant strains. These results confirm that there are still many unknown factors affecting the evolution of antibiotic resistance in tuberculosis, even for a very well characterized drug like isoniazid, and highlight the need for systematic studies to uncover them.

DISCUSSION

In this work we show how functional genomics experiments have proven to be a very potent technique for finding resistance determinants in *M. tuberculosis*, as they allowed us to correctly identify five known isoniazid resistance regions as well as discover many other candidate resistance genes. Interestingly, out of these five previously known regions only *katG* insertion mutants showed increased isoniazid resistance, while mutants for the other four regions had higher sensitivity instead. Thus, sensitivity-increasing genes can also act as resistance determinants depending on how mutations affect the protein or its expression levels. One example would be the Rv2170 gene, which in our study increases sensitivity to isoniazid and has been shown to confer resistance when its expression is increased because it encodes an acetyl-transferase that inactivates isoniazid [78]. Although we could not find insertion mutants for target isoniazid gene *inhA* due to its essentiality, non-essential genes in the mycolic acids biosynthesis pathway were functionally more associated with resistance than the rest of the genome, showing that the target pathway can be identified even when the target gene itself is essential. Functional genomics also pointed to redox metabolism as a resistance mechanism for isoniazid, confirming the usefulness of this technique in highlighting resistance determinants not directly tied to the drug's mode of action and in helping complete the resistance mutations catalog. Finally, combining functional genomics with phylogenetic data allowed us to pinpoint which regions and pathways were most important

for resistance in clinical settings and to obtain a series of candidate novel resistance determinants. This last step is very important, it has been shown before that even the most common clinical mutations to isoniazid are difficult to recover in vitro [79]. Likewise, we find regions that have been shown to affect resistance in vitro but are irrelevant in vivo. For instance, the aforementioned Rv2170 has no phylogenetic association with resistance according to our results.

This chapter illustrates how functional genomics is a powerful tool for detecting hundreds of genomic determinants that can modify antibiotic resistance, but we need a clinical readout to determine which of the genes have real relevance to the evolution of antibiotic resistance and the emergence of clinical resistance. Here, we have shown how a systematic approach combining insertion mutants on a genomic scale with phylogenetic association of clinical mutations can reliably detect important resistance features, uncover new resistance candidates and highlight relevant pathways and potential cross-resistances, in this case those most related to the mode of action for the first-line antibiotic isoniazid. We also provide an example of how a thorough understanding of the genetic architecture of isoniazid resistance can help us to prevent its emergence. The approach we describe can be used as a blueprint for studying the genetics of resistance to other antibiotics or describing lineage-specific differences, particularly to provide much-needed knowledge regarding resistance to new antibiotics.



CHAPTER 2

CHAPTER 2

Genomic analyses of *Mycobacterium tuberculosis* from human lung resections reveal a high frequency of polyclonal infections

Publication reference

Moreno-Molina, M., Shubladze, N., Khurtsilava, I. et al. Genomic analyses of Mycobacterium tuberculosis from human lung resections reveal a high frequency of polyclonal infections. Nat Commun 12, 2716 (2021). https://doi.org/10.1038/s41467-021-22705-z

Miguel Moreno-Molina[1], Natalia Shubladze[2], Iza Khurtsilava[2], Zaza Avaliani[2], Nino Bablishvili[2], Manuela Torres-Puente[1], Luis Villamayor[3], Andrei Gabrielian[4], Alex Rosenthal[4], Cristina Vilaplana[5,6,7], Sebastien Gagneux[8,9], Russell R Kempker[10], Sergo Vashakidze[2], Iñaki Comas[1,11]

1 Instituto de Biomedicina de Valencia IBV-CSIC, Valencia, Spain.
2 National Center for Tuberculosis and Lung Diseases of Georgia, Tbilisi, Georgia.
3 FISABIO Public Health, Valencia, Spain.
4 National Institute of Allergy and Infectious Diseases, National Institutes of Health, U.S. Dept. of Health and Human Services, Maryland, USA.
5 Fundació Institut Germans Trias i Pujol (IGTP), Barcelona, Spain.
6 Universitat Autònoma de Barcelona (UAB), Barcelona, Spain.
7 CIBER of Respiratory Diseases, Spain.
8 Swiss Tropical and Public Health Institute, Basel, Switzerland.
9 University of Basel, Basel, Switzerland.
10 Emory University School of Medicine, Department of Medicine, Division of Infectious Diseases, Atlanta, USA.
11 CIBER in Epidemiology and Public Health, Spain.

Chapter summary

Polyclonal infections occur when at least two unrelated strains of the same pathogen are detected in an individual. This has been linked to worse clinical outcomes in tuberculosis, as undetected strains with different antibiotic resistance profiles can lead to treatment failure. Here, we examine the amount of polyclonal infections in sputum and surgical resections from patients with tuberculosis in the country of Georgia. For this purpose, we sequence and analyze the genomes of *Mycobacterium tuberculosis* isolated from the samples, acquired through an observational clinical study (NCT02715271). Access to the lung enhanced the detection of multiple strains (40% of surgery cases) as opposed to just using a sputum sample (0-5% in the general population). We show that polyclonal infections often involve genetically distant strains and can be associated with reversion of the patient's drug susceptibility profile over time. In addition, we find different patterns of genetic diversity within lesions and across patients, including mutational signatures known to be associated with oxidative damage; this suggests that reactive oxygen species may be acting as a selective pressure in the granuloma environment. Our results support the idea that the magnitude of polyclonal infections in high-burden tuberculosis settings is underestimated when only testing sputum samples.

INTRODUCTION

How *Mycobacterium tuberculosis* evolves during the infection and treatment of a patient and transmits is key to understand phenomena like the spread of drug resistance. So far, the diversity of the pathogen has been mostly studied from sputum samples except for one study in the context of HIV/TB co-infection [80]. However, individual sputum samples may underestimate the true bacterial diversity within the lung as they are likely limited to reveal the coexistence of multiple *M. tuberculosis* strains in the same patient. Generally, infections involving two or more unrelated genotypes can be referred to as polyclonal (Figure 7). Polyclonal infections of *Mycobacterium tuberculosis* complicate the diagnosis and treatment of tuberculosis, particularly when the infecting strains differ in their antibiotic susceptibility; which can lead to the total replacement of the susceptibility profile during treatment or to heteroresistance [81].

Polyclonal infections are also relevant to evaluate if and how an initial TB infection protects from a second infection. Recent experiments in macaques suggest that an initial infection is highly protective against reinfection and disease caused by the same *M. tuberculosis* strain [82]. However, whether the initial infection protects against heterologous challenge with a different strain is unknown. Until the advent of molecular epidemiology, the dominant view was that primary TB episodes were due to endogenous reactivation after years of latent infection. However, progression to active disease occurs mainly in the first two

years after infection [83], suggesting that many of the episodes of TB following prolonged exposures are not due to endogenous reactivation but to reinfection. Molecular epidemiological studies also show that in high-burden countries the rate of reinfection is higher than previously recognized [84]. This is probably also true for superinfections with strains resistant to the treatment being used against the first infection.

The difficulty in identifying polyclonal infections is particularly pronounced in high-burden MDR-TB countries. Often, similar genotypes are responsible for a large proportion of recent TB transmission, making it more difficult to distinguish between these closely related strains when they coexist in the same patient [85]. In addition, in the context of drug resistance, patterns of within-host diversity in sputum cultures may be biased towards drug-resistant genotypes with high fitness. As a result, sputum-based cultures may not reflect the true extent of the pathogen diversity inside the patient lung. For obvious reasons, studying *M. tuberculosis* directly from the lung is not usually possible. There are however a few studies, some of them based on post-mortem biopsies, which suggest that the diversity of *M. tuberculosis* within the host is higher than what can be detected in sputa [80,86]. In addition, bacterial diversity within the lungs of TB patients may affect clinical outcomes as TB drugs are known to differ in their capacity to penetrate into the different types of lung lesions [87] or into the variable immune microenvironments of individual granulomas [88–90]. Therefore, the role of within-host diversity in general and of polyclonal infections is pivotal to

TB control as it has implications at many levels [91]. First, effective treatment can be compromised [92]. Second, public health interventions may need adjustments in settings with high rates of exogenous reinfection [92]. And third, there is a need to understand the role of *M. tuberculosis* strain variation in the context of new vaccines and the factors behind the limited success of the existing BCG vaccine [93,94].

The country of Georgia has a yearly incidence of 80 TB cases per 100,000 population. Out of those, 14 are MDR-TB cases (17.5%). Importantly, 12% among all new cases are MDR-TB indicating that transmission plays an important role in the epidemic of MDR-TB in Georgia [53,95]. Like in many countries of the former Soviet Union, adjunctive surgical resections are sometimes performed in patients not responding to treatment [96,97]. Utilizing these resected lung tissue samples, we studied the diversity of *M. tuberculosis* within different parts of cavitary lesions, and compared it to the *M. tuberculosis* diversity seen in sputum samples of the same patients. We achieved this using culture-derived bulk sequencing of the bacteria and detecting minority variants down to 3%.

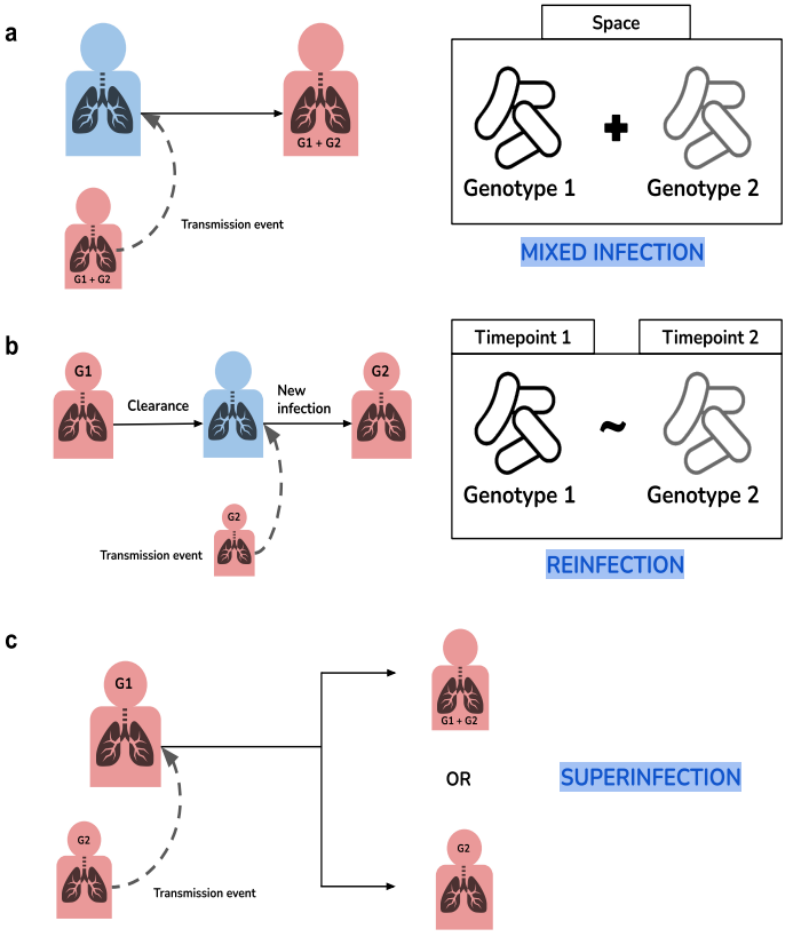


Figure 7. Theoretical scenarios for *Mycobacterium tuberculosis* polyclonal infections. Blue represents healthy patients and red represents infected patients. A) A transmission event of two strains from an infected individual to another results in two different genotypes being present in the same space or sample. B) An infected patient on treatment clears infection and gets infected again resulting in two different genotypes present over time. C) An already infected patient gets superinfected with a different genotype. The second genotype will either coexist with the first one or replace it.

In this work, we show a high frequency of polyclonal infections and important differences between patients in terms of genetic diversity in granulomatous lesions. Some of these differences are driven by an increased mutational supply mediated by host-derived Reactive Oxygen Species (ROS). In most patients infected with multiple strains, these strains differ in their drug resistance profiles. Furthermore, the high genetic distance observed between two strains infecting a given patient suggests an important role of strain genetic diversity in establishing a polyclonal infection. Our results represent a challenge for treatment and control of TB in the setting and highlight a possible limitation of new vaccines against TB.

METHODS

Sampling

We included 275 TB patients in this study (257 with diagnostic sputum and 18 with paired sputum and lung tissue samples) with a total of 370 *M. tuberculosis* clinical isolates from between January 9th 2013 and March 20th 2018 (see Figure 8 for details on the surgical cohort). Sputum samples were collected according to standard clinical practice. The surgical samples used in this study were acquired within the framework of observational clinical study NCT02715271, which is no longer recruiting, and results are reporting on primary outcome #3. Consent to publish clinical information potentially identifying individuals was obtained. The main indication for lung surgery was the persistence of abnormal lung lesions (predominantly cavities) identified on chest X-ray (13/18), both in DS (3/13) and MDR/XDR-TB (10/13), despite good treatment adherence. Additional indications for surgery included treatment failure (2.9%), complications of TB related to pulmonary hemorrhage, spontaneous pneumothorax or empyema [98]. Immediately after lung resection, samples were removed from the following zone of obtention: cavity center (C), cavity internal wall (I), cavity external wall (E), visually healthy tissue around the cavity (H) and nodule (N), placed in sterile tubes and sent to the National Center for Tuberculosis and Lung Disease microbiological laboratory for processing. In total we collected 219 single sputa, 38 pairs S1-S2 (serial sputa), 9 pairs C-S (caseum - sputum) and 9 complete surgical sets (C, I, E, H, N, S).

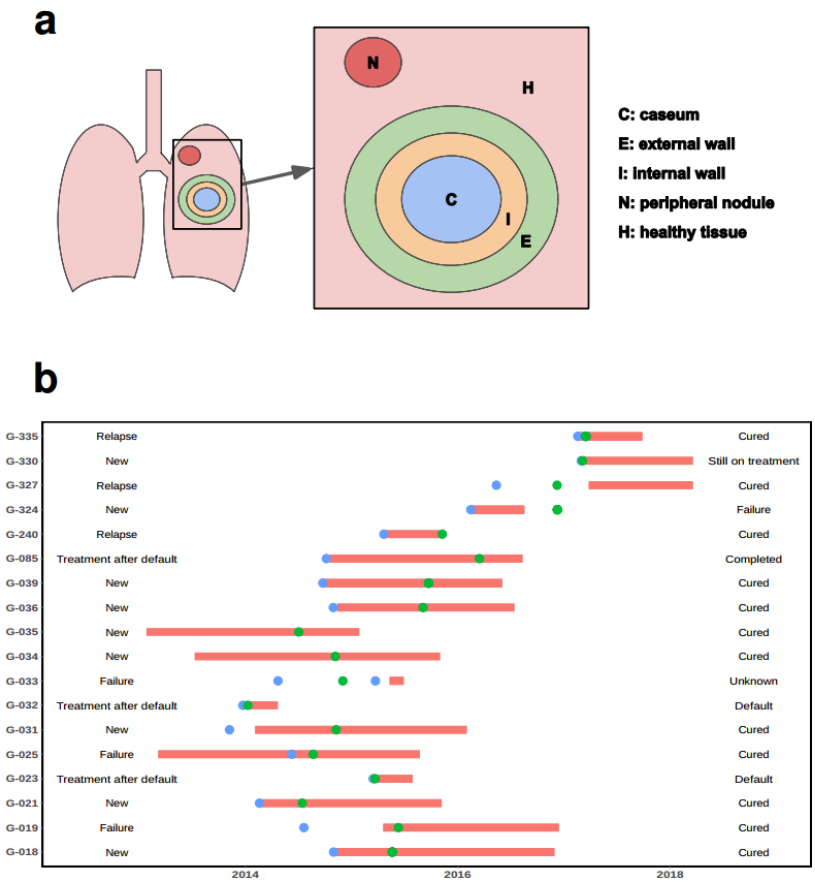


Figure 8. Details on the surgical cohort. A) Sampling sites for the surgical cohort. **B)** Timeline of key events for the surgical cohort. Treatment periods are depicted in red, sputum samples dates in blue points and surgical samples dates in green points. Starting case definition and treatment outcome are also represented.

Culture and drug susceptibility testing

Resected tissue samples were homogenized using Minilys homogenizer, placed in both LJ and BACTEC 960 MGIT™ culture and processed as per manufacturer's instructions. Sputum samples were also cultured in both media. All clinical isolates were confirmed for *M. tuberculosis* complex (MTBC) using the standard microbiological method [99,100]. All isolates underwent drug susceptibility testing (DST) for first and second line drugs in both solid (LJ) with increasing drug concentrations and liquid media BBL® MGIT™, as recommended by the manufacturer.

DNA extraction and sequencing

For both sputum and resected tissue cultures, after reaching sufficient bacterial growth in LJ medium, plates were thoroughly scraped to maximize diversity recovery for bulk sequencing. DNA extraction was performed by standardized protocol in the BSL-2+ laboratory from the AFB positive culture. In short, bacteria were heated at 80°C for 20 min, centrifuged and resuspended in TE buffer [101]. Lysozyme, SDS and proteinase K were added in sequential incubations. DNA was precipitated using chloroform-isoamyl alcohol and isopropanol, and resuspended in TE buffer. When there was no sufficient growth in LJ, the CTAB/chloroform method was used for DNA extraction from 1mL of MGIT culture. The quality and quantity of DNA were

analyzed with the help of spectrophotometers (Thermo Scientific NanoDrop 2000 and Qubit 3.0). Extracted DNAs were sent for whole genome sequencing to Broad Institute, C-PATH, TGen North or Tuberculosis Genomics Unit (IBV-CSIC). The average sequencing depth for this study was 141X. The average sequencing depth for surgical samples was 267X.

Bioinformatic analysis

Read preprocessing was done using *fastp [102]* to scan reads and trim low-quality ends with a mean window quality < 20. We then used Kraken [103] to taxonomically classify reads by means of a custom database, only keeping MTBC sequences to avoid false variants arising due to contaminant DNA. Filtered reads were mapped with BWA [104] to a predicted MTBC ancestor reference sequence [105] using default parameters, and processed using samtools [106] and Picard [107]. After that, we scanned for optical and PCR duplicates to remove them, as this helps to reduce the number of artifactual variants in low frequency ranges. Variant calling for sputum samples was carried out using the software and parameters from the calling module of the pipeline validated for sputum sample cultures at the IBV-CSIC (available at https://gitlab.com/tbgenomicsunit).

For a robust variant calling in surgical samples, we used three different variant callers (VarScan2 [108], GATK's HaplotypeCaller [109], LoFreq [110]) and integrated SNPs reported by at least two of them to get a high-confidence list

of low-frequency variants. VarScan2 was run with parameters '*pileup2snp sample.pileup --min-coverage 20 --min-reads2 4 --min-avg-qual 20 --min-var-freq 0.01 --min-freq-for-hom 0.9 --strand-filter 1*', GATK was run with parameters '*-T HaplotypeCaller -R ref.fasta -I sample.bam -o sample.vcf --min-base-quality-score 20 -ploidy 1*' and LoFreq was run with parameters '*call-parallel --pp-threads 12 -f ref.fasta -o sample.vcf sample.bam*' and '*filter -i sample.vcf -v 20 -A 0.01 -Q 20 -o filtered.vcf*'.

From the initial list of variants, we applied a mapping filter that discarded variants which arose in repetitive genomic regions like the PE/PPE families or phages [66]. To establish a threshold that discarded additional false variants, we performed a synthetic read simulation. Using the ART software package [111], we got the quality distribution and error profiles of our samples (using *art_profiler_illumina* with default parameters), and simulated 100 sequencing runs with the data (*art_illumina -p -1 profile.txt -2 profile.txt -na -iref.fasta -l 150 -f 1000 -m 280 -s 137 -o out.fas*). By analyzing simulations with the same pipeline, we defined a ~3% minimum frequency threshold to validate a variant in surgery samples. In addition, we extended our mapping filter to new regions that showed high-frequency SNPs in the simulations and were due to systematic mapping errors to the predicted ancestor, especially in Lineage 2 strains.

Phylogenetic and population genetics analyses

A maximum likelihood phylogeny of the 370 samples in the dataset was constructed using IQ-TREE 2 [112] from an alignment with gaps and no resistance SNP positions (run parameters: '*-m GTR -bb 1000*', meaning the use of a general time reversible model with unequal rates and unequal base frequencies, plus 1000 ultrafast bootstrap replicates). Additionally, a NJ phylogeny was constructed from the same alignment using MEGA [113] and phylogenetic Hamming distances between pairs of samples were extracted by parsing the branch lengths from the tree using Python's *ete3* toolkit [114].

We parsed SNP files by means of a custom Python script to obtain summary tables for each patient and a general one collecting information about the different numbers of fixed (fSNPs) and variable SNPs (vSNPs). All figures illustrating diversity across sites by patient were produced using R [115] and the ggplot2 package [116]. For comparisons within a patient's samples, we analyzed every SNP in their samples and calculated the differences. This was performed by means of an in-house Python script that works with the SNP files for every patient, analyzing them sequentially and establishing the population dynamic for every variant based on their presence/absence and frequency, obtaining stable, sporadic, ascending or descending dynamics. The data was plotted using R and the *ggplot2* package as well. For multi-sample patients, PCAs were produced with a matrix of SNP frequencies across all

samples and the R's *affycoretools* package using the function '*plotPCA*' for the first two principal components. We also calculated the percentage of explained variation of these two principal components using the '*prcomp'* function.

Prediction of WGS-DST profiles and comparison with culture DST

Using the SNPs obtained from the analysis of genomic data, we performed an antibiotic resistance prediction for every sample. For this, reliable catalogs of resistance-associated variants were used, namely PhyResSE [46] and ReSeqTB databases [117]. We also considered as likely resistance variants any small INDEL present in genes commonly associated with antibiotic resistance. Once the predictions were obtained, we systematically compared them to the available phenotypic DST results to calculate matches and mismatches in the surgery patients dataset. We then computed sensitivity and specificity based on these coincidences and discrepancies.

Identification of polyclonal infections

Identification of polyclonal infections depends on the distribution of the different genotypes among patients samples (Figure 7). When two genotypes are in two samples of the same patient, phylogenetic and genetic distances can be used to differentiate them. This happens mostly in

superinfection cases where one strain prevails over the other. In contrast, when the different genotypes are in one sample, as it happens when both genotypes coexist, then deconvolution methods have to be applied to separate them before calculating their genetic distance. Consequently, we apply different methods to identify multiple strains depending on the number of samples available from a patient.

Identification of polyclonal infections in single samples

Phylogenetic identification. We reasoned that a single sample which shows evidence of two different genotypes in the same isolate should have two characteristics in a phylogeny. First, the sample should show no terminal branch length. Terminal branch lengths represent private fixed SNPs only present in the sample. When there is polyclonal infection no private SNPs are seen as any SNP not shared by the two genotypes will be at intermediate frequencies. The second rule is that they should not be part of a transmission cluster. It is known that transmission clusters are enriched in strains that are zero SNPs apart from other strains in the same cluster. Thus we required the candidate polyclonal infection to be at least 20 SNP apart from another sample in the dataset. We developed a Python script able to analyze a phylogeny, extract branch lengths using Python's ete3 toolkit, calculate phylogenetic distances between all samples, and identify terminal branches with 0 SNPs and more than 20 SNPs apart from any other isolate

(Supplementary Figure 4). For this we generated a phylogeny using a neighbor-joining approach and Hamming distance, which represents branch length proportional to the absolute number of observed differences. Given the low genetic diversity of MTBC and that most positions are biallelic, Hamming distances reconstruct reasonably well the overall phylogeny with respect to maximum likelihood and it is easier to parse. To show the accuracy and the limit of detection of our approach we also generated in-silico mixes of strains with increasing genetic distance between the pairs selected (5, 10, 15, 20, 25 SNP) and added them to our phylogeny to test the script.

Lineage markers identification. To test coexistence of two strains in one sample, we checked the appearance of any phylogenetic markers using a database from the literature [118,119] by means of a custom Python script. For samples that showed evidence of more than one marker from different MTBC lineages at significant frequency (>5%), a polyclonal infection was called by this method and the estimated proportion of the involved genotypes was defined by the approximate frequencies of their markers.

Deconvolution of individual genotypes. For those cases in which we observed a polyclonal infection in just one sample by either the phylogenetic or lineage markers method, we established the proportions of the two strains in the sample and deconvoluted both the individual genotypes if their difference was big enough (eg. 80/20 proportions, obtained by phylogenetic markers). By clustering the frequencies of all variants matching the phylogenetic

markers frequencies and assigning them to its corresponding genotype, we could isolate both and calculate their genetic distance. For this we developed an in-house Python script that uses a phylogenetic marker database from the literature [118,119] and takes the SNP file from the sputum sample to perform the detection and separation of both strains, computing their genetic distance solely based on the number of differences.

Identification of polyclonal infections in multiple samples

Phylogeny manual inspection. To identify polyclonal infections happening in two different samples from the same patient, we manually analyzed the phylogeny looking for isolates of the same patient that sit in different parts of the phylogeny. In addition, we recorded if samples were placed close enough to other patients' samples in the phylogeny to suggest a recent reinfection event involving those two patients.

Analysis of frequency spectra differences. To make the detection process systematic for those cases where we had at least two samples (eg. sputum1-sputum2 or sputum-caseum), we developed Python and R scripts to perform pairwise comparisons using SNP frequency differences obtained from the sample's genomic data, generating differences profiles that were plotted using the ggplot2 package. We generated simple XY plots in which every point is a single variant determined by the frequencies

in each of the samples involved in the comparison, and calculated R^2 as a measure of correlation (Supplementary Figure 5). Polyclonal infections were almost always associated with low values of this index, except in cases where they were very subtle. We then proceeded to expand on the analysis generating a density graph that plots the distribution of SNP differences between the profiles. That is, in which range of frequencies is the comparison more enriched. This procedure allowed us to better define clonal infections when the enriched range was at the very low frequencies, and polyclonal infections when it was in the intermediate variable frequencies or at high fixed frequencies (Supplementary Figure 6).

RESULTS

A high frequency of polyclonal infections in MDR-TB patients from Georgia

A total of 370 *M. tuberculosis* cultured isolates from 275 patients were included in this study (Figure 9A). Lung surgical samples from 18 patients were also available. For 9 patients, two samples were analyzed from sputum and caseum, respectively, and for 9 others, we analyzed multiple samples from cavitary granulomatous lesions, samples from remote tubercular foci, and visually healthy lung tissue surrounding the cavity. In addition, for 38 patients, we analyzed two consecutive cultures from sputum samples. In terms of drug resistance profiles, patient samples were 16% pan-susceptible (44), 2% mono-resistant (3 mono-INH, 2 mono-SM, 1 mono-EMB), 2% poly-resistant (5), 41% MDR (112), 24% pre-XDR (66) and 15% XDR (42).

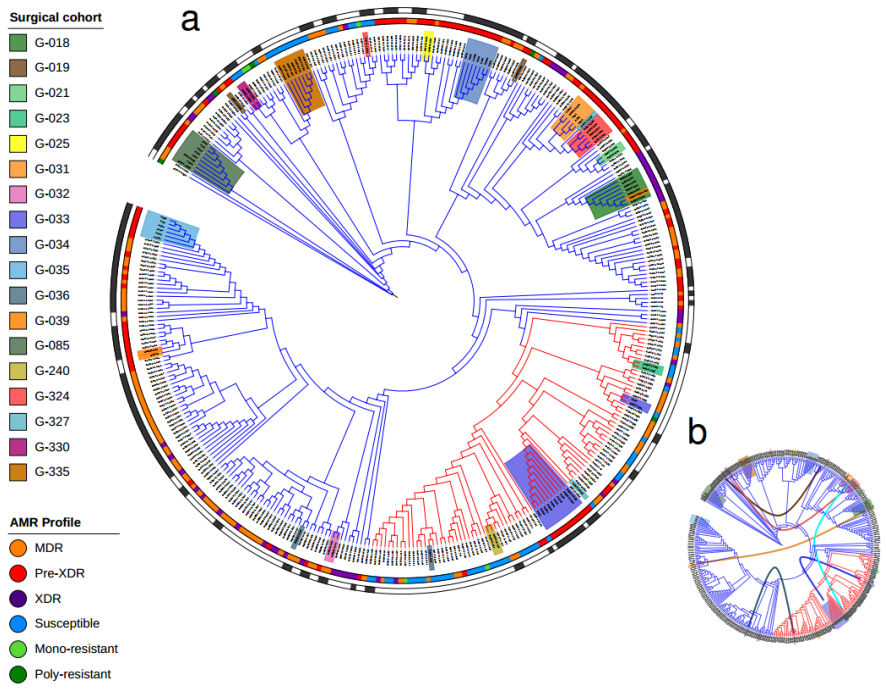


Figure 9. Phylogenetic diversity in Georgia and identification of polyclonal infections. A) Phylogeny of all Georgian *M. tuberculosis* isolates included in this study (L2 in blue, L4 in red). Surgery patients are highlighted in different colors. Drug resistance profiles are illustrated in the inner band (see legend) and transmission clusters represented on the outer band were estimated using a 10 SNP phylogenetic distance threshold. Branch lengths and bootstrap values not represented. B) Curves connecting patient samples located in distant places of the phylogeny, suggesting polyclonal infections.

In regards to the sputum and lung tissue *M. tuberculosis* clinical isolates from patients undergoing adjunctive surgical resection, a genome-based phylogeny of all bulk-sequenced cultures revealed that for seven out of eighteen cases, samples from the same patient did not cluster together (Figure 9B). This strongly suggested the presence of polyclonal infections. Patients harboring different strains in different samples become evident when considering pairwise genetic distances between these strains. It has been suggested that >10 SNPs between isolates of the same patient are unlikely to emerge by clonal diversification from a single *M. tuberculosis* genotype [120]. For all seven patients in which the isolates did not cluster together, the distance between the genotypes was much higher than 10 SNPs (range 105 to 1,132 SNPs, Table 2). By contrast, in the eleven patients whose isolates clustered together, the pairwise SNP distance was between 0 and 2 SNPs (Wilcoxon Rank Sum Test; W 0, p-value 0.001). For example, the caseum isolate from patient G019 clustered 120 SNPs apart from the sputum isolate of the same patient. Extreme examples were patient G036 or G327, in which the sputum and the caseum isolates belonged to different lineages (L2 and L4, more than 1,100 SNPs apart). For three patients, the isolates were not only located in different parts of the phylogeny, but were also assigned to different transmission clusters (Supplementary Figure 7). For patient G324, whose surgical isolates clustered with the caseum isolate from patient G327, the sputum isolate clustered with patient G005 and G312 sputum isolates. Patients G019 and G036 also had paired isolates clustering independently with other isolates in the phylogeny instead of with each other. In

total, we found 6 out of 7 polyclonal infection surgical patients involved in recent transmission clusters (Supplementary Figure 7). These values suggest that superinfections are common among the MDR-TB population of this setting.

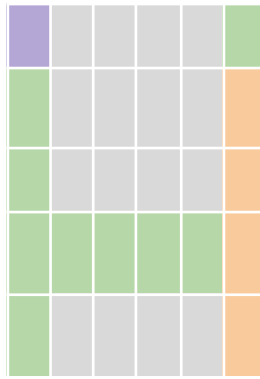
For three surgery patients, the situation was even more complicated. For those patients, we found instances of polyclonal infections of two unrelated genotypes not only between isolates of the same individual, but also within the same lesion. For instance, patient G039 harbored two genotypes in different parts of the same lesion (C, I, E, H), while the nodule (N) and sputum (S) samples harbored a single genotype. Deconvolution of the two genotypes present in the caseum isolate (see *Methods*) revealed that the genotype absent from the nodule and sputum belonged to a different sublineage and was 105 SNPs apart. A similar phenomenon was observed in patient G240, where two genotypes coexisted in the granuloma center, but only one was present in the sputum isolate and clustered with another patient's sputum isolate. Finally, patient G033 also showed evidence of two genotypes belonging to different L4 sublineages, coexisting in the nodule and healthy tissue samples.

Taken together, our results show that seven out of the eighteen surgery patients (39%) showed evidence of infection by two phylogenetically unrelated genotypes, either in the same sample or in separate samples (Table 2). To compare the high percentage of polyclonal infections in our surgical dataset with the percentage expected when using

only a single sputum diagnostic sample, we scanned for possible polyclonal infections in 218 sputum culture positive patients from the country of Georgia. As only one sample per case was available we could only scan for cases of mixed infection. We used two different methods to identify co-existence of two different genotypes in single sputum samples, including a phylogenetic-based approach specifically designed for this analysis (see *Methods*). In total, we identified 11/218 (5%) likely cases of co-existence in these patients based on a single sputum sample. Supporting this result, when we only took into consideration the sputum isolates from surgery patients no mixed infections were detected. This is in contrast with findings in surgical samples where we identify three cases in which two genotypes co-existed (17% in surgical samples vs 5% in sputum samples from the general population; chi-square 4.39, p-value = 0.036). To increase the power to detect additional polyclonal infections in the general population, we analyzed a set of 38 patients with two consecutive sputum samples (Supplementary Table 2) collected from days to months apart (range 1-595 days). In total, we identified seven patients out of the 38 (18.4%) with evidence of polyclonal infection at some point during their TB disease. This was lower than the frequency of polyclonal infections in the surgical samples (39%) but higher than in the single diagnostic sputum samples (5%). Despite limitations, this suggests that the real extent of polyclonal infections cannot be accurately estimated from a single sputum sample.

Table 2. Summary of *M. tuberculosis* genotypes found in the surgical cohort. Lineage (Lin.) and drug resistance profile (DRP) of the different genotypes (G1, G2), with genetic distances (Dist.). Genotype locations indicated with colors (green, G1; orange, G2; purple, mix; grey, no sample available). C - caseum, I - inner wall, E - external wall, H - healthy tissue, N - nodule, S - sputum.

Patient	Infection	G1 Lin	G1 DRP	G2 Lin	G2 DRP	Dist	C	I	E	H	N	S
G018	Clonal	L2.2.10	XDR	-	-	0						
G021	Clonal	L2.2.10	Pre-XDR	-	-	0						
G023	Clonal	L4.2.1	Sus.	-	-	1						
G025	Clonal	L2.2.10	Pre-XDR	-	-	0						
G031	Clonal	L2.2.10	Pre-XDR	-	-	2						
G032	Clonal	L2.2.9	MDR	-	-	0						
G034	Clonal	L2.2.10	Pre-XDR	-	-	0						
G035	Clonal	L.2.2.9	Pre-XDR	-	-	0						
G085	Clonal	L2.2.10	Pre-XDR	-	-	2						
G330	Clonal	L2.2.10	Sus.	-	-	0						
G335	Clonal	L2.2.10	Sus.	-	-	0						
G033	Polyclonal	L4.3.3	Pre-XDR	L4.2.1	Sus.	611						
G039	Polyclonal	L2.2.9	Pre-XDR	L2.2.10	Pre-XDR	105						

G240	Polyclonal	L4.8	Sus.	L4.8	Sus.	162	
G019	Polyclonal	L2.2.10	MDR	L2.2.10	Pre-XDR	120	
G036	Polyclonal	L2.2.9	MDR	L4.8	Sus.	1132	
G324	Polyclonal	L2.2.10	Sus.	L2.2.10	Pre-XDR	128	
G327	Polyclonal	L4.3.3	Sus.	L2.2.10	Pre-XDR	1116	

TB patients undergoing lung surgery show complex infection scenarios

For nine surgery patients, several bacterial cultures from different parts of the cavitary lesion could be analyzed in detail, including caseum (C), inner wall (I), external wall (E), remote nodule (N) and surrounding healthy tissue (H), in addition to the diagnostic sputum culture (S). Analysis of the *M. tuberculosis* genomic diversity within and around lesions showed very different patterns across patients (Supplementary Figures 8 and 9). Further details on the surgical cohort are provided in Supplementary Note 1.

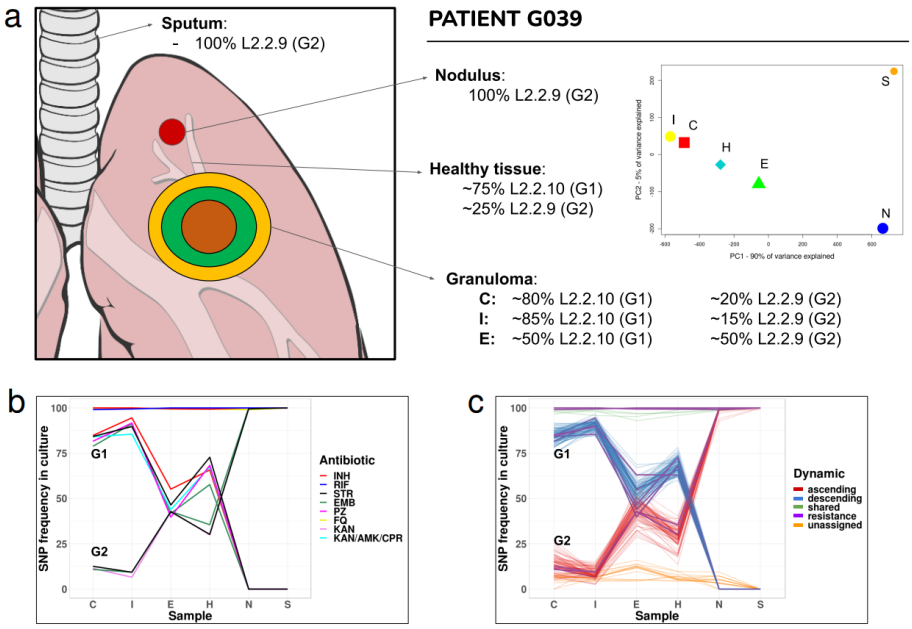


Figure 10. Dynamics of polyclonal infection with two different strains across the granuloma. A) Descriptive summary of patient G039 genotypes frequencies and distribution across the lesion. The PCA graph illustrates their separation: while the X-axis explains ~90% of the variance, leaving N and S apart from the rest, the Y-axis only explains ~5%, mainly defined by low-frequency variants not shared by N and S (due to read depth differences). B) Antibiotic resistance-associated mutations across surgical and sputum samples for the two co-existing genotypes. C) Deconvolution of the two genotypes by frequency patterns clustering. Resistance-associated variants from B are represented in purple. Unassigned subpopulations (in orange) are low-frequency variants that we cannot assign to any of the two genotypes, and shared fixed variants (in green) are common to both genotypes. Abbreviations: C - caseum, I - inner wall, E - external wall, H - healthy tissue, N - nodule, S - sputum.

We chose patient G039 to illustrate how complex infection and transmission patterns can be in a high-burden MDR-TB setting. This patient showed a polyclonal infection of two genotypes in the caseum, the inner and external wall as well as in the healthy tissue, but not in the remote nodule or sputum (Figure 10A). A detailed analysis of the SNP frequencies suggested that there were two genotypes coexisting in the caseum at a ratio of 80:20 (genotype1:genotype2). Tracking of variable positions across samples of the patient revealed that this ratio varied. The proportion of genotype 1 decreased as we moved out of the caseum until reaching a ratio of nearly 50:50 in the external walls and healthy tissue. The situation was reversed in the remote nodule where genotype 2 was fixed at 100% frequency. The sputum of the patient also contained genotype 2 at a 100% frequency. Deconvoluting these two genotypes was challenging as both belonged to the same lineage 2 sublineage. We could extract the genetic base of genotype 2 from the nodule as it represented 100% of the culture but genotype 1 always existed as a mixture with genotype 2 in the other samples. To deconvolute genotype 1, we assigned as a fixed SNP any position that was in the caseum sample at above 75% frequency, thus reflecting the expected frequency of SNPs associated with genotype 1. Comparison with genotype 2 corroborated that both genotypes belonged to the same lineage but different sublineages (2.2.9 and 2.2.10), being only 105 SNPs apart. The frequencies of genotype 1 and genotype 2 correlated with resistance mutations to several antibiotics at the corresponding frequencies (Figure 10B, 10C). Several of the resistance mutations were the same in both genotypes,

including those for isoniazid, rifampicin, ofloxacin and kanamycin. The combined frequency of different resistance mutations to ethambutol and streptomycin across the two genotypes resulted in 100% of the culture being resistant to those drugs. For other drugs like pyrazinamide, resistance mutations were only found in genotype 1.

Patient follow-up treatment can be compromised by polyclonal infection

To analyze the role of polyclonal infections in drug resistance, we used the available DST results for the surgical cohort's samples. First, we explored whether there was variability within a patient in phenotypic DST results. In 13 out of the 18 surgical patients, the DST results did not change across sites even when the sputum sample was taken into account. In five patients, the DST profiles differed across samples. Sometimes the difference involved only one drug like OFX in G019 but on other occasions like patient G327, the DST profile was fully reversed when comparing the caseum and the sputum sample (from pre-XDR to pansusceptible in this case). In other instances, we had polyclonal infections like patient G033, in which the mix between the two genotypes in H and N samples (susceptible + pre-XDR) resulted in an overall profile of pre-XDR masking the second genotype. Although the number of patients was low, three out of five patients with discrepancies in DST versus genomic prediction had a polyclonal infection, suggesting that these infections can mislead individual

treatment of MDR-TB patients. For a comparison of phenotypic and genotypic results as well as identification of novel drug resistance markers see Figure 11 and Supplementary Note 2.

Another complication might result from the replacement over time of one strain by another, which may compromise the treatment if no additional DST is done during follow-up. In support of this notion, in all 4 patients harboring different strains in temporally separated samples (Table 2), the DST profile changed for several drugs. This included cases that were susceptible in the first isolate but MDR/XDR in the second isolate or vice versa, like patients G324 or G327. This data shows that treatment based on the results from one diagnostic sample can be misleading and suggests the need for sequential testing of isolates under programmatic conditions, especially in settings with a high MDR-TB burden such as the country of Georgia.

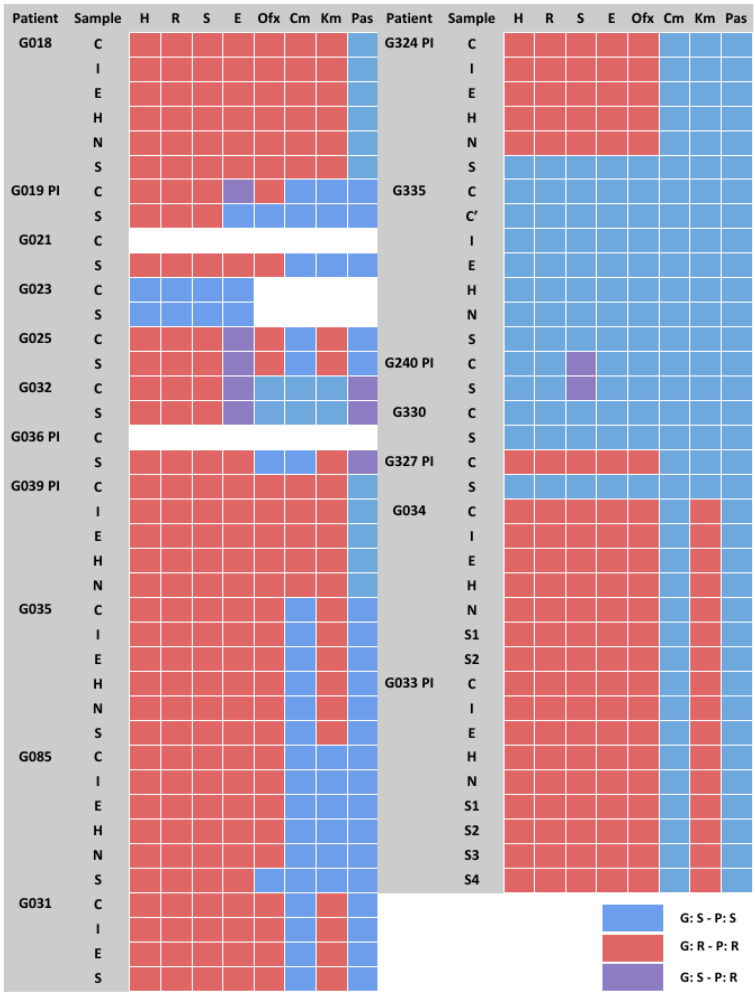


Figure 11. Comparison matrix of resistance prediction based on genotype and drug susceptibility testing (DST). Legend shows three different colors to explain matches and mismatches between them. Blue means susceptibility match, red means resistance match and purple means disagreement in which genomic data failed to predict phenotypic resistance.

Diversity within and around lesions reflects host microenvironment pressures

For those patients with both surgical and sputum samples and with no polyclonal infection detected, we analyzed the M. tuberculosis genetic diversity across their samples. The overall diversity of *M. tuberculosis* within patients was well represented in sputum samples when compared to the caseum sample (correlations >90%). We then calculated the number of bacterial SNPs exclusive to each sample. These are SNPs that have accumulated since the divergence of the sample from the closest isolate in the phylogeny. In general, patients with a high number of exclusive SNPs in the lesions also had a high number of SNPs in the sputum (R^2=0.9623 between cavity centers and corresponding sputa using exclusive SNPs and excluding G021 which showed a high diversity only in sputum). Thus, the general assumption that the *M. tuberculosis* diversity in sputum samples is only a subset of the within-patient diversity is not always true. Indeed, we saw patients in which the sputum harbored exclusive SNPs not found in the surgical samples and vice versa. TB patients widely differed in the amount of within-host MTB genomic diversity, with approximately half of the patients harboring almost no MTB diversity (e.g. G031, G034, G035 with no low-frequency SNPs; Figure 12A and Supplementary Figure 9), while others showed a large diversity across sputum and surgical samples (e.g. G023, G025 with >600 low-frequency SNPs; Figure 12A and Supplementary Figure 9). However, the comparison between sputum and surgery samples must be taken with caution as the sputum and surgical samples were

obtained at different time points and may also reflect the sampling of different lesions.

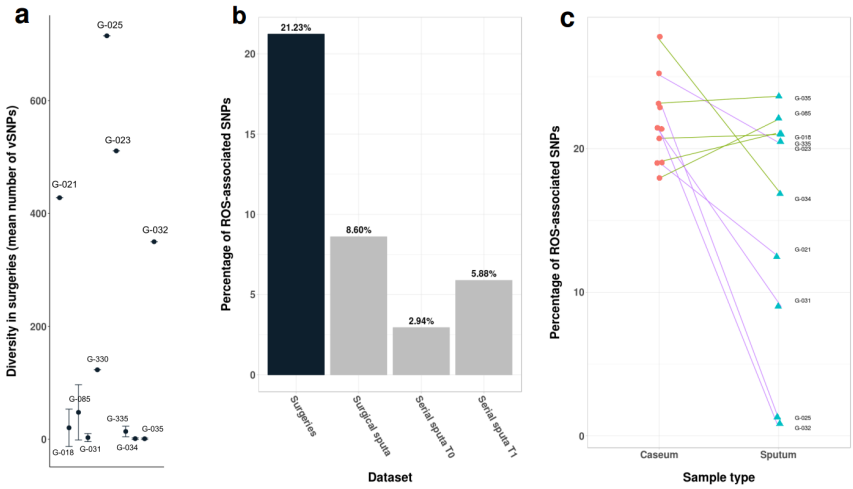


Figure 12. Impact of the granuloma microenvironment on within-host diversity. A) Average diversity of each patient's surgical samples (measured by number of vSNPs and excluding polyclonal infections, n=number of independent surgical samples available for each patient, range 1-3). Data are presented as mean values +/- SD. B) Pooled comparisons of ROS signature in the different datasets. Categories include surgery specimens, sputum samples from surgery patients, and serial sputum samples from the same patient. C) Individual values of caseum and sputum from surgery patients. Purple lines connect those patients whose differences are statistically significant by two-tailed X^2 test.

Several selective forces, including the host response and antibiotic treatment, may affect the diversity within granulomas from the time of diagnosis to the time of

surgery. It has been suggested that the host immune system may exert a mutational pressure on the infecting bacteria through the production of reactive oxygen species (ROS). This has been demonstrated in single colony analyses of cultured sputum samples showing that an elevated mutational supply can be identified in immunocompetent individuals but not in HIV-positive individuals [121]. A hallmark of an elevated mutational rate due to ROS is a mutational signature associated with oxidative damage (increased changes C>T and G>A) [122]. It could be expected that such a ROS mutational signature was amplified in surgical samples compared to sputum. As our study was not based on single-colony sequencing but on bulk sequencing, we reasoned that identifying variants present in a few colonies in a culture should roughly correlate with variants at very low frequency in bulk culture sequencing. Thus for this analysis, we focused on variants with frequency lower than 5% for samples with enough depth of coverage, and excluding patients showing polyclonal infection as frequencies in those cases are not straightforward to interpret (total eligible patients n = 10). Our analysis revealed that, as hypothesized, surgical samples showed a stronger ROS mutational signature compared to the sputum samples from the same patients (21% versus 8.6%, pooled analysis, Figure 12B; chi-square 14.5, p-value <0.0001). As the sputum and surgical samples were obtained at different time points, we also analyzed patients with serial sputum samples to determine the effect of time on the diversity observed in sputum. While there was an effect of time (S1 3% vs. S2 5.9%; chi-square 22.3 p-value <0.0001), the magnitude of the ROS signature in

serial samples was similar to that seen in sputum samples from surgical patients (Figure 12B) but significantly lower than in the surgical samples (chi-square p-value <0.0001 for all comparisons against surgical samples). Some patients had stronger ROS signatures than others (Figure 12C; 5 out of 10 significant at the individual level, G021, G023, G025, G031, G032; chi-square p-value < 0.05). This is in accordance with results by Liu and colleagues [121], where 4 out of 18 cases showed a significant increase in ROS-associated mutations. Thus, our results suggest that the granuloma microenvironment can increase the mutational supply for transitions during the within-host evolution of *M. tuberculosis*.

DISCUSSION

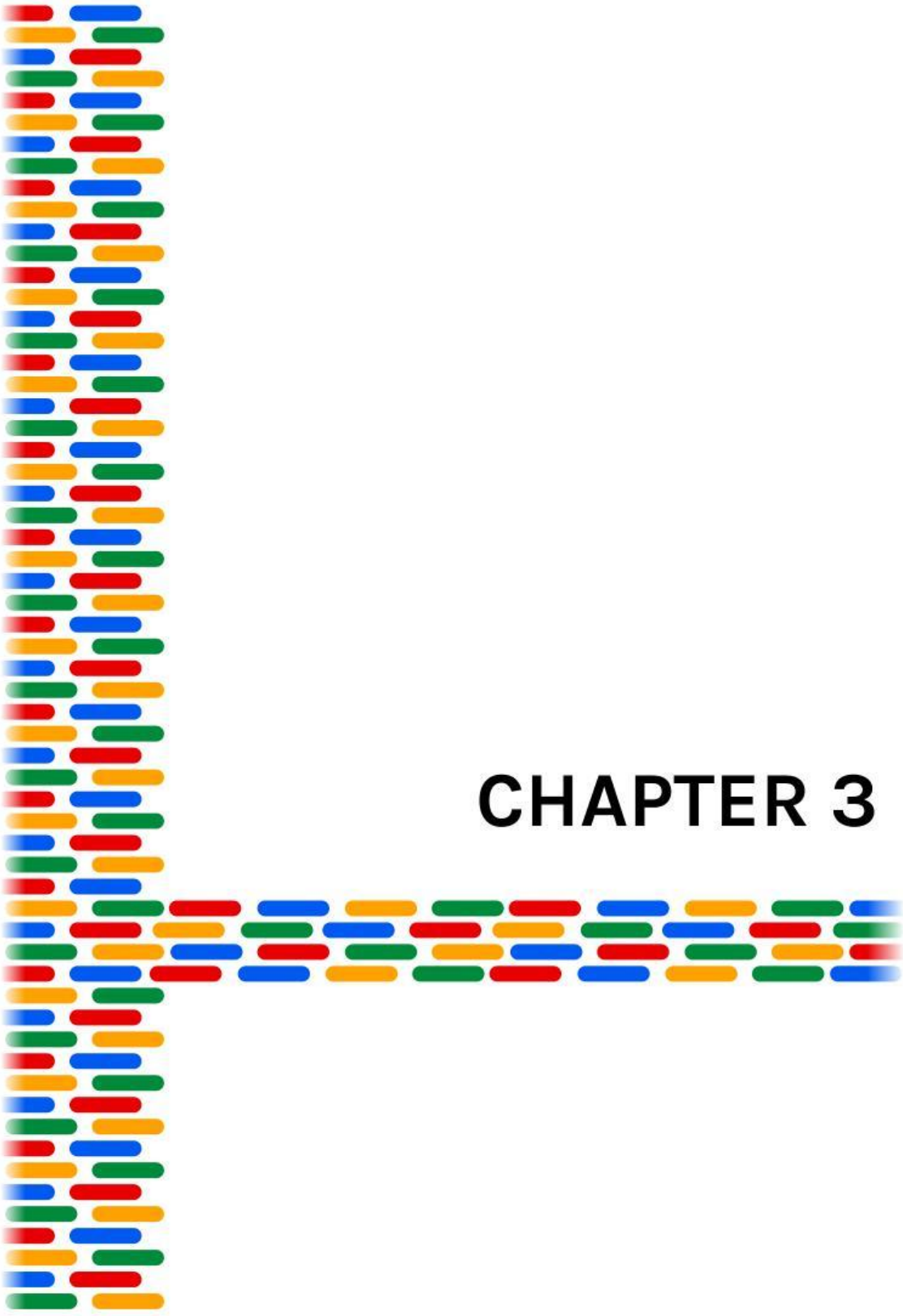
Studies on within-host diversity of *Mycobacterium tuberculosis* are based on cultured samples and very few on non-sputum samples, limiting our capacity to understand diversity patterns at the site of infection. In this study, we analyzed an 18-patient cohort, most of them MDR to XDR-TB cases, with available lung surgical samples from Tbilisi (Georgia). We performed bulk sequencing of cultured bacteria from the surgical resections and sputum isolates of every patient, finding an elevated percentage of polyclonal infections with an impact on drug resistance in the surgical cohort (39%), which cannot be attained by single sputum samples. In addition, we demonstrate that patterns of genetic diversity differ within lesions and across patients suggesting a role of host immune microenvironments.

Our results are necessarily limited by the characteristics of the epidemiological setting and the patients undergoing surgery. Most of the patients who underwent surgery had already been diagnosed at least as MDR-TB and did not respond to treatment. This population is characterized by substantial exposure to multiple antibiotics, a higher frequency of prior TB disease, the frequent presence of cavitary disease (associated with poor antibiotic penetration), and sometimes prolonged hospitalization that can increase the risk of superinfection. In addition, diagnostic sputum and surgical samples have a several months time difference, which limited some of our analyses, but we tried to assess the effect of sampling time by comparing our surgical dataset with the serial sputa dataset

in which the time difference between samples is similar overall. Also, even though we had access to surgery samples, analyses were done on cultured-samples and not directly on the surgery or sputum sample which can impose different biases [123]. Finally, a possible limitation of this study would be cross-contamination as an explanation to the high frequency of polyclonal infections, although we think it is highly unlikely for a number of reasons. First, sample collection dates and processing dates from all patients are not close in time thus they haven't shared the same space. Second, sample homogenization is carried out in closed special, disposable tubes so that samples cannot be mixed. Third, genotypes match their DST phenotypes in nearly all cases, arguing against a general contamination problem. Lastly, not all polyclonal infections are in a transmission cluster and thus they don't match any other strain processed in the laboratory.

In this chapter, we have shown that surgical lung resections from TB patients reveal a more complete picture of the within-patient diversity of *M. tuberculosis* compared to sputum samples. The high frequency of polyclonal infections found may be related to the nature of our patient population, yet this allowed us to study complex infection scenarios with the potential to confound diagnosis and/or DST results, as in many cases the two genotypes involved in the infection had different drug resistance profiles. The surgery patients are often in a transmission cluster suggesting that either there are uncontrolled hotspots of transmission shared by MDR-TB patients, there is increased host susceptibility to reinfection by different genotypes, or both. The fact that

polyclonal infections usually involve strains of different lineages or sub-lineages suggests that vaccine preclinical models must take into account the genetic diversity of the bacteria to assess protection. Finally, as observed in culture-based sputum studies [80,124,125], there are profound differences in diversity between patients and here we show that those differences can also be seen across regions of the cavitary lesions and between patients. In some patients, differences are driven by a mutational signature associated with ROS and by treatment and suggest a link between immune and drug resistance selective pressures. Overall, our results exemplify a challenge for tuberculosis treatment and control, highlighting our knowledge gap on the natural history of the disease in certain settings and the need for better patient follow-up in high burden MDR-TB areas to halt the spread of resistant strains.



CHAPTER 3

CHAPTER 3

HIV coinfection impacts *Mycobacterium tuberculosis* diversity and drug resistance acquisition during early treatment

Unpublished as of July 2023

Miguel Moreno-Molina[1], Carla Mariner-Llicer[1], Belén Saavedra[2,3], Edson Mambuque[2], Neide Gomes[2], Manoli Torres-Puente[1], Luis Villamayor[4], Pablo Cano-Jimenez[1], Mariana G. Lopez[1], Alberto García-Basteiro[2,3], Iñaki Comas[1,5]

1 Instituto de Biomedicina de Valencia (IBV-CSIC), Valencia, Spain
2 Centro de Investigação em Saúde de Manhiça (CISM), Maputo, Mozambique
3 ISGlobal, Barcelona Centre for International Health Research, Hospital Clínic - Universitat de Barcelona, Barcelona, Spain
4 FISABIO Public Health, Valencia, Spain
5 CIBER in Epidemiology and Public Health, Spain

Chapter summary

Tuberculosis (TB) is the primary cause of death among HIV-infected patients, especially in low-income countries. The clinical manifestation of TB in HIV-infected individuals is different from that observed in non-HIV

patients, and in addition to poor treatment adherence in these settings, several studies have also shown a decreased antibiotic bioavailability in HIV+ patients, contributing to the acquisition of drug resistance due to suboptimal levels of drug circulation. However, evidence that associates HIV with increased odds of developing antibiotic resistance and the mechanisms by which this occurs are still unclear. In this study, we analyzed a cohort of 23 HIV+ and 29 HIV- TB patients from Mozambique by using deep-sequencing to study the dynamics of early drug resistance acquisition, in addition to performing 480 MIC determinations. By investigating the 162 TB serial samples from the day of diagnosis up to the first month of treatment, we detected a significant accumulation of low-frequency mutations in genes related to drug resistance for HIV+ individuals at baseline and found evidence of a decreased capability to eliminate the constantly appearing bacterial diversity over the first month of treatment. This suggests lower immune and treatment selective pressures acting over the bacterial populations of the coinfected group, favoring the appearance of low-level heteroresistance as an early indicator of worse treatment outcomes. We were also able to link low-frequency variants in our samples to subtle MIC shifts, which appear associated with poorer outcomes. Our findings give insight into the host-pathogen interacting selective factors that contribute to higher rates of treatment failure in TB-HIV coinfected individuals, and can help to design better treatment regimes for these patients.

INTRODUCTION

Human immunodeficiency virus (HIV) is the most common comorbidity associated with tuberculosis, and it has been reported that HIV-infected individuals are up to 30 times more likely to develop TB [51]. The two diseases potentiate each other, speeding immune decay progress and promoting disseminated or extrapulmonary TB [126]. Also, the lower bacterial loads of HIV-coinfected individuals make TB detection through sputum smear microscopy less effective, and TB drugs combined with antiretroviral therapy (ART) require careful monitorization to avoid side effects or adverse reactions [127]. All of these factors, together with an observed higher likelihood of antibiotic resistance alongside HIV, makes coinfected patients have worse clinical outcomes. However, whether HIV is a risk factor for the arisal of drug resistance in TB has been an ongoing debate for a long time, since findings on the association between HIV and DR-TB across many studies have been inconsistent.

There is a lack of knowledge about how TB bacteria interact with HIV+ hosts, specifically how the population dynamics behave in the context of an immune-deficient environment. In HIV+ patients, less immune cells in which TB bacteria can replicate means a harder competition between subpopulations, and this is an important point that can differentiate coinfection from regular TB from an evolutionary point of view. A hypothesis about drug resistance arisal in this context starts with the appearance of heteroresistance, a phenomenon in which a subpopulation microevolves under the selective pressure of antibiotics

[128]. This would be a temporary state that ends in fixation of the subpopulation that achieves a mutation with the lowest fitness cost and the highest level of resistance. An early detection of heteroresistance can both provide insight into how drug resistance develops within-patient and also serve as a clinical tool to make treatment adjustments if necessary. Even if the variants involved confere low-level resistance through non-canonical mechanisms, they can precede or facilitate the acquisition of a clinically-relevant mutation that affects the ultimate drug target.

Mozambique is a country with a high incidence of both TB and HIV, particularly the district of Manhiça, making it a convenient location to recruit TB-HIV coinfected patients to study the role of HIV in TB microevolution [129]. Therefore, here we sought to assess how antibiotic selective forces shape and affect tuberculosis population dynamics and diversity during the first month of treatment, in the context of evaluating if HIV coinfection has any impact on the early evolution of drug resistance. We did this using deep whole-genome sequencing in serial TB samples from two different cohorts from Manhiça: patients with TB and patients with TB-HIV coinfection. We assess the genetic diversity of the bacterial population, as it constitutes the total variability contained in the sample, representing the fundamental unit of change that determines the microevolutionary potential of subpopulations. We demonstrate that HIV+ patients have a decreased capacity to eliminate TB diversity after 4 weeks of treatment. We then focus on low-frequency variants arising in resistance-related genes and perform MIC determination experiments to weigh

the contribution of these mutations to MIC shifts, and associate those to the outcome of the patient at the end of the first month of treatment to point out that the most likely mechanism for treatment failure in HIV+ patients has to do with the early appearance of low-level heteroresistance that persists after the first month of treatment.

METHODS

Cohort recruitment and sampling

Patients were recruited in the district of Manhiça (Mozambique) from March 2019 to March 2020, and serial sputum samples were collected during the first month of treatment (days 1, 7, 14 and 28). A total of 162 samples were included in this study. These correspond to 52 different patients that met all inclusion criteria to be analyzed: a valid baseline sample, complete drug susceptibility at baseline and no evidence of polyclonal infection. From the final 52 patients, 29 were negative for HIV and 23 were positive. The HIV+ patients had a CD4 count lower than 500 cells per microlitre, thus considered immunocompromised. All sputum samples collected were sent to the Centro de Investigação em Saúde de Manhiça (Mozambique) and processed in their BSL3 facility, where an initial culture step was performed before shipping the decontaminated cultured isolates to the Biomedicine Institute of Valencia (Spain) for further processing.

Culture and MIC determinations

All samples received from Mozambique were cultured and amplified prior to DNA extraction in a BSL3 lab. Cultures were performed in standard Middlebrook 7H10 solid medium (BD) according to the manufacturer's

instructions. For all baseline and last available samples from each patient, MICs to six different drugs were also determined (isoniazid, rifampicin, ethambutol, streptomycin, levofloxacin and amikacin). We used the EUCAST protocol [130] for this with a few modifications but the same plate scheme. Briefly, each isolate was tested in a plate for the 6 antibiotics at the same time with microdilutions ranging from 8 different concentrations in Middlebrook 7H9-10% OADC liquid medium. The inoculum was of $5{\cdot}10^4$ bacteria and it was incubated at 37ºC for 7 days before the first reading. Growth is visually assessed using an inverted mirror at day 7 and also at day 14. The MIC is the lowest concentration that inhibits visual growth at day 14 and is expressed in mg/L. All plates included an internal growth control, consisting of the same strain at two different concentrations (same as test and another at 100X dilution - GC100, GC1) and the plate is considered valid for reading when growth is assessed in GC1.

DNA extraction and sequencing

Cultures were thoroughly scraped to recover the maximum amount of bacterial diversity and the recovered material was inactivated at 90ºC for 30 minutes. DNA extraction was performed on the NucliSENS easyMAG platform using the manufacturer's protocol. Libraries were prepared using Illumina's Nextera XT kit with the standard recommended protocol, and sequencing was performed in a NovaSeq instrument using an S2 flow cell to achieve a high read depth for all samples, resulting in a mean of ~600X.

Bioinformatic analysis

The analysis of deep-sequencing data was performed following the same pipeline as Chapter 2. Briefly, read preprocessing was done using *fastp* to scan reads and trim low-quality ends with a mean window quality < 20. We used Kraken to taxonomically classify reads and keep MTBC sequences, and these filtered reads were mapped with BWA to the predicted MTBC ancestor reference. For a robust variant calling, we again used three different variant callers (VarScan2, GATK's HaplotypeCaller and LoFreq) and integrated SNPs reported by at least two of them to get a high-confidence list of low-frequency variants. To check the exact parameters used with each caller, see Chapter 2, Methods.

Following variant calling, we again applied a mappability filter that discards variants which arise in low-confidence genomic regions. We also performed synthetic read simulations using the ART software package, simulating 100 sequencing runs with the data. By analyzing simulations, we defined a dynamic minimum frequency threshold to validate a variant in each sample depending on its read depth, up to 1.5% frequency for samples at or above 1000X. We parsed variant files using custom Python scripts to obtain summary tables for each patient collecting information about the different numbers of fixed (fSNPs) and variable SNPs (vSNPs). All figures illustrating diversity across

time points by patient were produced using R and the *ggplot2* package.

A maximum likelihood phylogeny of the 162 samples in the dataset was constructed using IQ-TREE 2 from an alignment with gaps and no resistance SNP positions (run parameters: '*-m GTR -bb 1000*', meaning the use of a general time reversible model with unequal rates and unequal base frequencies, plus 1000 ultrafast bootstrap replicates).

We used the amount of diversity detected in the baseline sample vs. follow-up, regardless of the frequency, for a COX regression model (proportional risk model) that allows assessing whether the HIV cohort has a dynamic significantly different from that of its control counterpart. The calculations and graph were made using the *survfit* and *coxph* of the *survival* and *survminer* packages in R.

Prediction of drug resistance profiles and comparison with MIC determinations

With the variants obtained from the analysis of genomic data, we performed an antibiotic resistance prediction for every sample. Drug resistance prediction was carried out using the 2021 WHO catalogue of mutations [49]. We also considered as likely resistance variants any small INDEL present in genes commonly associated with antibiotic resistance. We systematically compared predictions with the MIC results to determine if samples were beyond the clinical

breakpoint for all six antibiotics that were tested and assigned candidate variants to try to explain higher MIC values where it was possible. Additionally, we compared each baseline MIC value to its follow-up to calculate the fold change and define MIC shifts when those were higher than 2. We then aggregated the values of first-line antibiotics to associate their average with the patient's outcome defined as in the Cox regression analysis.

RESULTS

The study initially recruited 86 patients from the district of Manhiça (Mozambique), although 23 were discarded due to not having a valid culture from the day of TB diagnosis or the culture being contaminated. Serial sputum samples were collected at the day of TB diagnosis and along the first month of treatment (days 1, 7, 14 and 28), for a maximum total of four samples per patient. After sequencing, another 11 patients were also excluded from the dataset due to polyclonal infections or other anomalies. The final dataset includes 52 patients having a valid sample set for a total of 162 samples, with 29 patients being HIV- and 23 HIV+. All included patients had a drug susceptible sample at baseline and their treatment outcomes were reported as cured. Figure 13 shows a phylogeny of this dataset and its lineage distribution: 12 patients were L1 (7 HIV- and 9 HIV+), 9 L2 (6 HIV- and 3 HIV+), 2 L3 (both HIV+) and 25 L4 (17 HIV- and 8 HIV+), with no noticeable lineage bias towards any patient group.

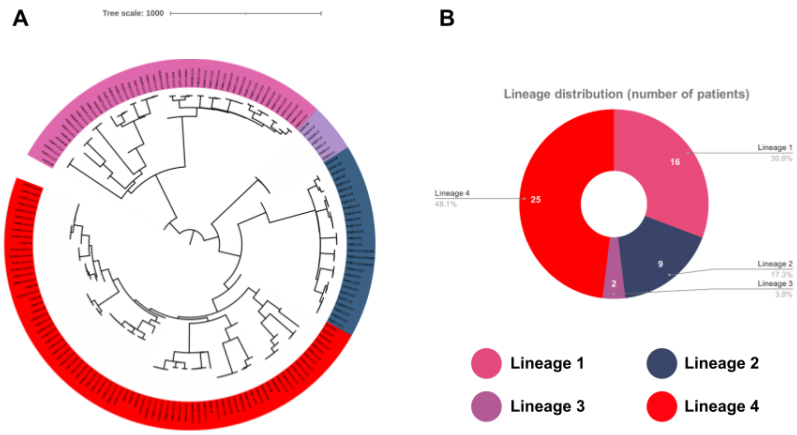


Figure 13. Phylogeny and lineage distribution of the dataset. a) Maximum-likelihood phylogeny of the 162 samples included in the dataset. b) Lineage distribution of the final included patients. Colors represent MTB lineages (see legend).

HIV affects TB within-patient diversity and the ability to reduce it

We initially assessed the overall diversity of each sample as denoted by the number of variable SNPs (vSNPs, frequency ~3% to ~97%) in the population. At very low frequencies, which we were able to achieve thanks to our high read depth across the dataset, this is a reflection of the abundance of different subpopulations that are evolving within-patient during the course of treatment (Figure 14A). The overall diversity followed a decreasing dynamic across the first month in both groups, although quantitatively the

HIV+ group had a lower starting diversity than the HIV- group (median 567 vSNPs for HIV- vs 358 vSNPs for HIV+, T-test p-value 0.046). This difference is mainly driven by a subset of HIV- patients that had a very high total diversity not due to polyclonal infection. However, when looking at the percentage of subpopulations -as vSNPs- that are eliminated during the first month, there is a difference in the rate at which each group loses this diversity. We performed a Cox regression analysis defining a 90% initial diversity reduction -or confirmed negative culture- as the outcome, and the model yielded a +38% hazard ratio for the HIV+ group (Figure 14B). With this analysis we tried to determine if diversity in patients with HIV is purified differentially as a side effect of the patient's condition (longer treatments and relapses) or if it is promoted by the comorbidity in itself. We reasoned that despite the number of positive-culture samples at the end of the first month being low, this result suggests that, over this early stage of treatment, the HIV+ group has a reduced capacity to eliminate diversity, which could translate into a worse prognosis.

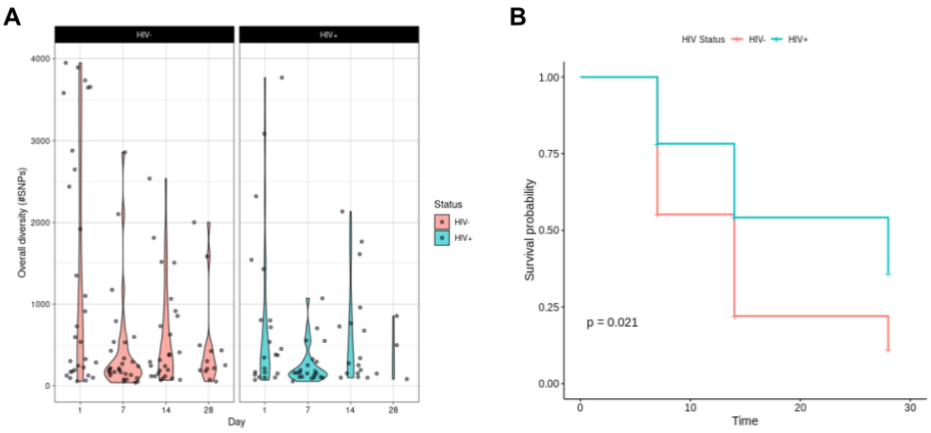


Figure 14. Overall diversity of the dataset and Cox regression analysis. a) Genetic diversity of all samples at each time point, denoted as the total number of vSNPs. b) Survival curve resulting from the Cox regression analysis. Colors represent HIV- and HIV+ groups (see legends).

Heteroresistance is generated at low frequencies under treatment selective pressure

After assessing the overall diversity, we then focused on diversity in resistance-related genes, a set of 58 genes derived from the 2021 WHO drug resistance catalogue. We looked at non-synonymous SNPs in genes such as *katG* or *rpoB* along the serial samples and also observed a decreasing dynamic of diversity during treatment despite

that new low-frequency variants keep appearing (Figure 15A). We then tested if the accumulation of SNPs in this set of genes was significant using permutation tests (n = 1000, see Figure 15B for individual p-values). Although the total diversity at day 1 was higher in HIV- patients, it was in line with the rest of the genome, while HIV+ patients showed a statistically significant accumulation of variants from the start. Finally, both groups showed this enrichment starting from day 7, once the treatment selective pressure starts acting upon resistance-associated genes differentially from the rest of the genome.

Finally, we calculated the dN/dS ratio of genes that confer resistance to first-line antibiotics and examined their trajectories during treatment for both groups (Figure 15C). This ratio gives us an idea of adaptation or functional constraint in protein-coding genes by quantifying the relative excess or deficit of amino acid-replacing versus silent mutations. Both groups of patients started with neutral values but showed a rapid increase in genes related to isoniazid (*katG*, *ahpC*, *ndh*) and ethambutol resistance (*embB*) at day 7, illustrating how synonymous variants are rapidly lost under the evolutionary pressure of antibiotic treatment due to a selective sweep.

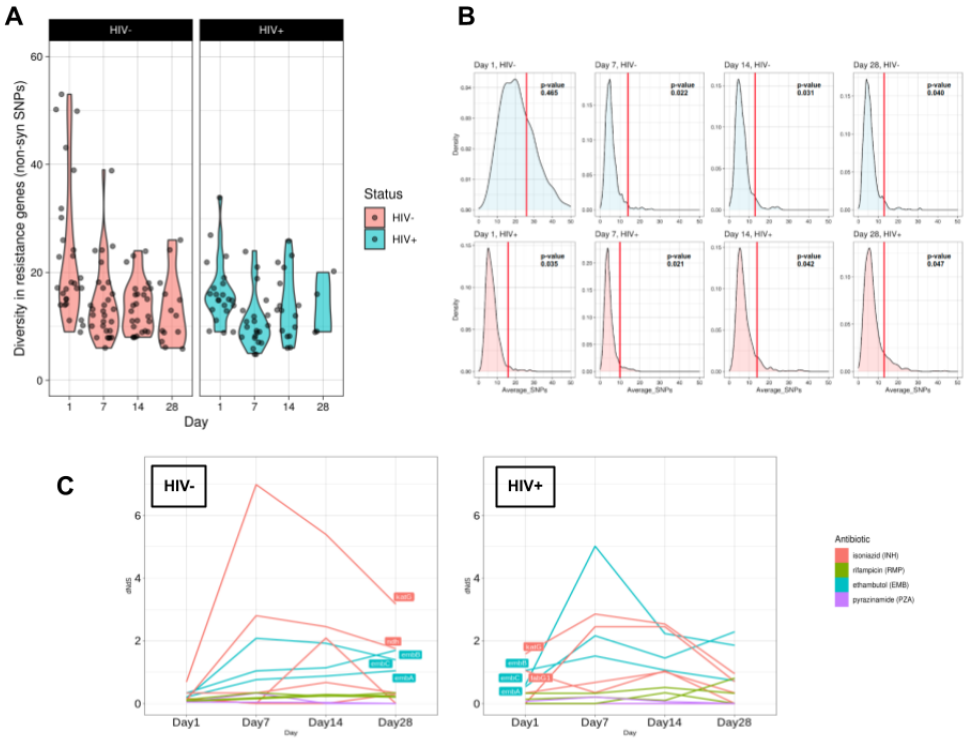


Figure 15. Diversity and evolutionary trajectories of resistance-related genes. a) Genetic diversity of resistance-related genes from the 2021 WHO catalog, denoted as the number of vSNPs. b) Distributions and real values (red lines as the resistance-associated set) of the number of vSNPs in random sets of genes. c) dN/dS trajectories across time for genes associated with resistance to first-line antibiotic treatment.

MIC shifts as early markers of treatment failure

We determined the MIC of six antitubercular drugs -isoniazid, rifampicin, ethambutol, streptomycin, levofloxacin and amikacin- for the baseline sample and a follow-up sample of every patient, depending on availability. We successfully determined the MICs for the pair of samples in 40 out of the 52 patients, for a total of 480 MIC determinations. The follow-up sample varies among patients as they progressively become culture-negative along the first month of treatment, or depending on the success of the re-culture step before the MIC determination. The goal was to detect slight MIC shifts over time that could be associated with low-frequency variants that we identified in many resistance-associated genes. Figure 16 illustrates the MIC distributions for the antibiotics that we tested along with the clinical breakpoints. We observed a slight but general shift to higher concentrations in most antibiotics for the follow-up samples compared to the baseline ones, which supports low-level heteroresistance widely appearing under early treatment selective pressures.

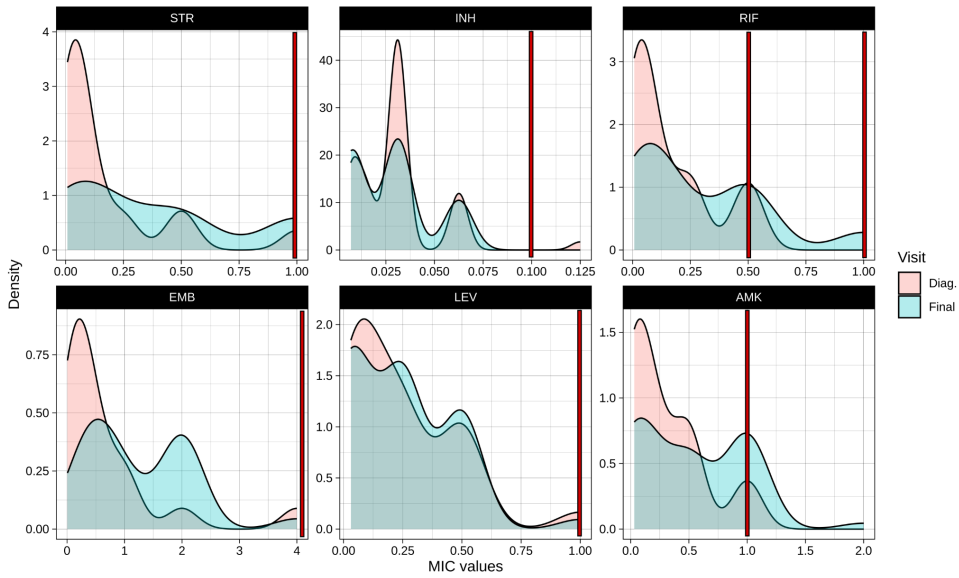


Figure 16. Density distributions of MIC values for all samples and antibiotics tested. Baseline samples MIC distribution in red, and follow-up in blue. Red lines represent each drug's clinical breakpoint (rifampicin has two due to a recent update in the concentration, from 0.5 to 1 mg/L).

In the 52 patients of the dataset, no associated variant from the WHO catalogue was identified at high frequency in any of the samples. We systematically compared the available MIC values of every antibiotic from baseline to follow-up, calculating their individual fold changes. Then, for values equal or greater than 2 that define positive MIC shifts, we searched for variants that could explain them. Table 3 shows the MIC values determined for all drugs, along with their fold change over time and variants

that have no reported confidence score or its score is "Uncertain" in the WHO catalogue that are present in samples with MIC shifts for their respective antibiotics.

Table 3. Results of all MIC determinations and candidate variants. Each patient is included with the MIC concentration for each drug at baseline and follow-up, along with the fold change resulting from the comparison. Variants are provided as candidates to explain positive shifts. In the follow-up column, yellow means the value is at the clinical breakpoint, and red means that it is above it.

Patient	Antibiotic	MIC Baseline	MIC Follow-up	Fold Change	Candidate variants	Other variants
P1	STR	0.5	0.25	0.5		
	INH	0.03125	0.03125	1		
	RIF	0.25	0.125	0.5		
	EMB	1	1	1		
	LEV	0.25	0.125	0.5		
	AMK	0.5	0.5	1		
P2	STR	0.5	0.03125	0.0625		
	INH	0.03125	0.0078125	0.25		
	RIF	0.25	0.0625	0.25		
	EMB	1	0.5	0.5		
	LEV	0.25	0.03125	0.125		
	AMK	0.5	0.0625	0.125		
P3	STR	0.5	-	-		
	INH	0.03125	-	-		
	RIF	0.25	-	-		
	EMB	1	-	-		
	LEV	0.25	-	-		
	AMK	0.5	-	-		

P4	**STR**	1	0.25	0.25		
	INH	0.03125	0.0078125	0.25		
	RIF	0.125	0.03125	0.25		
	EMB	2	0.5	0.25		
	LEV	0.25	0.03125	0.125		
	AMK	1	0.0625	0.0625		
P5	**STR**	0.5	0.03125	0.0625		
	INH	0.03125	0.007	0.25		
	RIF	0.25	0.03125	0.125		
	EMB	1	0.125	0.125		
	LEV	0.25	0.0625	0.25		
	AMK	0.5	0.125	0.25		
P6	**STR**	0.03125	0.125	4		
	INH	0.016	0.03125	2		
	RIF	0.016	0.5	32		
	EMB	0.125	0.5	4	embB_I16F (3.94%)	
	LEV	0.0625	0.5	8		
	AMK	0.03125	0.5	16		
P7	**STR**	0.03125	-	-		
	INH	0.03125	-	-		
	RIF	0.25	-	-		
	EMB	0.25	-	-		
	LEV	0.5	-	-		
	AMK	0.125	-	-		
P8	**STR**	0.008	0.0625	8		
	INH	0.007	0.007	1		
	RIF	0.016	0.125	8		
	EMB	0.125	0.5	4	embR_F376L (8.09%)	
	LEV	0.0625	0.5	8		
	AMK	0.03125	0.5	16		

P9	**STR**	0.03125	0.5	16	rpsL_IG_781479GT (4.94%)	**Low-level resistance:** dnaA_V98A (7.26%)
	INH	0.0078125	0.03125	4	katG_L634I (2.84%)	
	RIF	0.0625	0.25	4		
	EMB	0.25	1	4		
	LEV	0.03125	0.25	8		
	AMK	0.0625	1	16	murA_IG_1471660GA (2.58%)	
P10	**STR**	0.5	0.03125	0.0625		
	INH	0.03125	0.0078125	0.25		
	RIF	0.25	0.03125	0.125		
	EMB	1	0.5	0.5		
	LEV	0.25	0.03125	0.125		
	AMK	0.5	0.0625	0.125		
P11	**STR**	0.0625	0.016	0.25		
	INH	0.0078125	0.03125	4		
	RIF	0.03125	0.008	0.25		
	EMB	0.5	0.25	0.5		
	LEV	0.03125	0.03125	1		
	AMK	0.0625	0.0625	1		
P12	**STR**	0.016	0.016	1		
	INH	0.016	0.03125	2		
	RIF	0.03125	0.03125	1		
	EMB	0.008	0.0625	8	embB_I16F (2.74%)	
	LEV	0.0625	0.0625	1		
	AMK	0.03125	0.03125	1		
P13	**STR**	0.0625	1	16		
	INH	0.008	0.03125	4	fabG1_IG_1673380CG (4.12%)	
	RIF	0.125	0.5	4	rpoB_L449Q (4.15%)	
	EMB	0.25	2	8	embA_V961D (2.49%)	
	LEV	0.25	0.5	2		
	AMK	0.25	1	4	murA_IG_1472226AG (2.76%)	

P14	**STR**	0.5	0.0625	0.125		
	INH	0.03125	0.008	0.25		
	RIF	0.25	0.0625	0.25		
	EMB	1	0.25	0.25		
	LEV	0.25	0.0625	0.25		
	AMK	0.5	0.0625	0.125		
P15	**STR**	0.5	-	-		
	INH	0.03125	-	-		
	RIF	0.25	-	-		
	EMB	1	-	-		
	LEV	0.25	-	-		
	AMK	0.5	-	-		
P16	**STR**	0.5	0.25	0.5		
	INH	0.03125	0.03125	1		
	RIF	0.0625	0.0625	1		
	EMB	1	1	1		
	LEV	0.25	0.25	1		
	AMK	0.5	0.5	1		
P17	**STR**	0.5	1	2	Rv0681_V58L (2.03%)	**Multi-drug tolerance:** prpr_A273T (2.68%) **Other:** pckA_V556A (3.48%)
	INH	0.03125	0.03125	1		
	RIF	0.25	0.5	2		
	EMB	1	2	2	embB_P12Q (2.75%)	
	LEV	0.25	0.5	2		
	AMK	0.5	1	2		
P18	**STR**	0.5	-	-		
	INH	0.03125	-	-		
	RIF	0.25	-	-		
	EMB	1	-	-		
	LEV	0.25	-	-		
	AMK	0.5	-	-		

P19	**STR**	0.5	0.5	1		
	INH	0.03125	0.03125	1		
	RIF	0.25	0.5	2		
	EMB	1	1	1		
	LEV	0.25	0.5	2		
	AMK	0.5	1	2		
P20	**STR**	0.125	0.0625	0.5		
	INH	0.0078125	0.03125	4		
	RIF	0.03125	0.016	0.5		
	EMB	0.5	0.125	0.25		
	LEV	0.03125	0.125	4		
	AMK	0.0625	0.125	2		
P21	**STR**	0.03125	0.0625	2	Rv0681_V58L (2.21%)	**Other:** phoR_P70L (99.55%) Rv2781c_V124A (99.87%)
	INH	0.0078125	0.03125	4	fabG1_IG_1673380CG (2.17%)	
	RIF	0.0625	0.25	4		
	EMB	0.5	0.5	1		
	LEV	0.03125	0.25	8		
	AMK	0.0625	0.25	4	murA_IG_1474425AG (2.51%)	
P22	**STR**	0.5	1	2		**Other:** idsA2_Q152S (2.24%) idsA2_R153P (2.21%) idsA2_V155G (2.26%)
	INH	0.03125	0.0625	2		
	RIF	0.25	1	4		
	EMB	1	2	2	embR_F376L (12.2%)	
	LEV	0.25	0.5	2		
	AMK	0.5	1	2		
P23	**STR**	1	1	1		
	INH	0.0625	0.03125	0.5		
	RIF	0.5	0.5	1		
	EMB	4	4	1		
	LEV	0.5	1	2		
	AMK	1	2	2		

P24	**STR**	1	-	-		
	INH	0.03125	-	-		
	RIF	0.5	-	-		
	EMB	2	-	-		
	LEV	0.5	-	-		
	AMK	1	-	-		
P25	**STR**	1	0.0625	0.0625		
	INH	0.0625	0.0078125	0.125		
	RIF	0.5	0.0625	0.125		
	EMB	2	0.5	0.25		
	LEV	0.5	0.03125	0.0625		
	AMK	1	0.0625	0.0625		
P26	**STR**	0.0625	-	-		
	INH	0.03125	-	-		
	RIF	0.25	-	-		
	EMB	0.25	-	-		
	LEV	0.25	-	-		
	AMK	0.25	-	-		
P27	**STR**	0.25	0.25	1		
	INH	0.03125	0.03125	1		
	RIF	0.5	0.5	1		
	EMB	2	2	1		
	LEV	0.25	0.25	1		
	AMK	1	1	1		
P28	**STR**	0.5	0.25	0.5		
	INH	0.03125	0.125	4		
	RIF	0.25	0.0625	0.25		
	EMB	1	0.5	0.5		
	LEV	0.25	0.5	2		
	AMK	0.5	4	8		

P29	**STR**	0.016	-	-		
	INH	0.016	-	-		
	RIF	0.03125	-	-		
	EMB	0.25	-	-		
	LEV	0.0625	-	-		
	AMK	0.03125	-	-		
P30	**STR**	0.5	1	2		
	INH	0.03125	0.03125	1		
	RIF	0.25	0.03125	0.125		
	EMB	1	1	1		
	LEV	0.25	0.25	1		
	AMK	0.5	0.5	1		
P31	**STR**	0.5	0.25	0.5		**Other:** pckA_E145V (2.16%) pckA_V556A (2.68%)
	INH	0.03125	0.03125	1		
	RIF	0.25	1	4	rpoB_V218L (2.55%)	
	EMB	1	1	1		
	LEV	0.25	0.25	1		
	AMK	0.5	0.5	1		
P32	**STR**	0.016	-	-		
	INH	0.016	-	-		
	RIF	0.03125	-	-		
	EMB	0.125	-	-		
	LEV	0.0625	-	-		
	AMK	0.03125	-	-		
P33	**STR**	1	-	-		
	INH	1	-	-		
	RIF	1	-	-		
	EMB	2	-	-		
	LEV	8	-	-		
	AMK	4	-	-		

P34	**STR**	0.015625	-	-		
	INH	0.0078125	-	-		
	RIF	0.0625	-	-		
	EMB	0.5	-	-		
	LEV	0.03125	-	-		
	AMK	0.0625	-	-		
P35	**STR**	0.03125	0.5	16		**Multi-drug tolerance:** prpr_D401V (2.58%) rnj_I156L (2.87%) **Other:** idsA2_T326A(2.39 %) pyrG_A462D (2.23%) pyrG_G444D (2.57%)
	INH	0.03125	0.125	4		
	RIF	0.0625	0.25	4		
	EMB	0.125	2	16	embB_I16F (5.52%)	
	LEV	0.0625	1	16		
	AMK	0.03125	1	32		
P36	**STR**	0.5	0.5	1		
	INH	0.03125	0.0625	2		
	RIF	0.25	0.5	2		
	EMB	1	2	2	embR_F376L (11.65%)	
	LEV	0.25	0.25	1		
	AMK	0.5	1	2		
P37	**STR**	0.125	0.125	1		
	INH	0.0625	0.0625	1		
	RIF	1	0.5	0.5		
	EMB	2	2	1		
	LEV	0.25	0.25	1		
	AMK	0.5	0.5	1		
P38	**STR**	0.5	0.25	0.5		
	INH	0.03125	0.03125	1		
	RIF	0.25	0.5	2		
	EMB	1	2	2		
	LEV	0.25	0.25	1		
	AMK	0.5	0.5	1		

P39	**STR**	0.015625	0.03125	2		
	INH	0.0078125	0.03125	4		
	RIF	0.0625	0.03125	0.5		
	EMB	0.25	0.125	0.5		
	LEV	0.03125	0.0625	2		
	AMK	0.0625	0.03125	0.5		
P40	**STR**	0.5	0.03125	0.0625		
	INH	0.0625	0.0078125	0.125		
	RIF	0.5	0.0625	0.125		
	EMB	1	0.5	0.5		
	LEV	0.5	0.03125	0.0625		
	AMK	1	0.0625	0.0625		
P41	**STR**	0.5	0.0625	0.125		
	INH	0.03125	0.03125	1		
	RIF	0.25	0.016	0.0625		
	EMB	2	0.007	0.0035		
	LEV	0.25	0.25	1		
	AMK	1	0.125	0.125		
P42	**STR**	0.5	0.03125	0.0625		
	INH	0.03125	0.0078125	0.25		
	RIF	0.25	0.0625	0.25		
	EMB	1	0.5	0.5		
	LEV	0.25	0.03125	0.125		
	AMK	0.5	0.0625	0.125		
P43	**STR**	0.5	0.03125	0.0625		
	INH	0.03125	0.0078125	0.25		
	RIF	0.25	0.0625	0.25		
	EMB	1	0.5	0.5		
	LEV	0.25	0.0625	0.25		
	AMK	0.5	0.0625	0.125		

P44	**STR**	1	1	1		
	INH	0.0625	0.0625	1		
	RIF	0.5	0.5	1		
	EMB	4	4	1		
	LEV	0.5	0.5	1		
	AMK	1	1	1		
P45	**STR**	0.5	0.03125	0.0625		
	INH	0.03125	0.03125	1		
	RIF	0.25	0.008	0.032		
	EMB	1	1	1		
	LEV	0.25	0.25	1		
	AMK	0.5	0.125	0.25		
P46	**STR**	0.5	0.5	1		
	INH	0.03125	0.03125	1		
	RIF	0.25	0.25	1		
	EMB	2	2	1		
	LEV	0.5	0.5	1		
	AMK	0.5	0.5	1		
P47	**STR**	0.5	1	2		**Antibiotic resilience:** resR_V167D (2.88%) **Multi-drug tolerance:** prpr_D401V (19.78%) **Other:** pckA_V556A (3.95%)
	INH	0.03125	0.5	16		
	RIF	0.25	1	4	rpoB_Q975H (46.37%)	
	EMB	1	4	4	embR_F376L (9.72%)	
	LEV	0.25	4	16		
	AMK	0.5	4	8	murA_L67F (3.61%)	
P48	**STR**	0.03125	0.03125	1		**Low-level resistance:** Rv0010c_A26S (5.62%) **Other:** idsA2_R309G (6.56%) Rv2779c_Q165R (4.42%)
	INH	0.03125	0.03125	1		
	RIF	0.03125	0.5	16		
	EMB	0.0625	0.25	4		
	LEV	0.0625	0.25	4		
	AMK	0.03125	0.25	8		

P49	**STR**	0.5	-	-		
	INH	0.03125	-	-		
	RIF	0.25	-	-		
	EMB	1	-	-		
	LEV	0.25	-	-		
	AMK	0.5	-	-		
P50	**STR**	0.5	0.125	0.25		**Multi-drug tolerance:** glpK_A460V (34.68%) glpK_D84E (29.51%) **Other:** pckA_D368A (32.62%)
	INH	0.03125	0.5	16		
	RIF	0.25	1	4		
	EMB	1	2	2	embR_F376L (8.45%)	
	LEV	0.25	1	4		
	AMK	0.5	1	2	Rv0528_G478C (4.76%)	
P51	**STR**	0.5	0.25	0.5		
	INH	0.03125	0.125	4		
	RIF	0.25	0.25	1		
	EMB	1	1	1		
	LEV	0.25	1	4		
	AMK	0.5	0.5	1		
P52	**STR**	0.03125	-	-		
	INH	0.03125	-	-		
	RIF	0.03125	-	-		
	EMB	0.125	-	-		
	LEV	0.0625	-	-		
	AMK	0.03125	-	-		

We then tried to associate these shifts to the patient's outcome at the end of the first month of treatment, defined as in our previous Cox regression. A reduction of 90% of the initial observed diversity or a confirmed negative culture at day 28 is considered a 'positive outcome', and conversely, a positive culture with a smaller reduction is considered a 'negative outcome'. This is a proxy measure of treatment efficacy during the first month, which is thought to correlate with the final clinical outcome of the patient. For calculation purposes, an overall MIC shift for the sample was determined when the average fold change for isoniazid, rifampicin and ethambutol was greater than 2, as the patients were all receiving standard first-line treatment. Using these criteria, 10 out of 12 patients that experienced an MIC shift during early treatment failed to reduce their overall initial diversity and remained culture-positive after a month of treatment (Table 4). This is an early indicator of likely treatment prolongation and perhaps eventual treatment failure according to recent bibliography. As per HIV status we see no bias in the 'MIC shift' category, yet we can observe that for those patients that did not experience MIC shifts, all HIV+ were negative outcomes, further corroborating the result of our previous Cox regression analysis. Comparing by outcome, even though the total numbers are not statistically significant (chi-square p-value = 0.1115), just taking into account the HIV- cases for the calculation (chi-square p-value = 0.0118) suggests there is indeed an association of the MIC shift with the inability to effectively reduce overall genetic diversity after a month of antibiotic treatment. Finally, in quantitative terms we also tested if the MIC shifts of both cohorts were different,

obtaining a negative result (Wilcoxon test p-value = 0.8683). The fact that the overall MIC shift distribution of both groups is similar suggests that the variants that are being selected under treatment pressure are not functionally different, regardless of the patient's immune status.

Table 4. MIC shifts correlate with the patient's outcome. At the end of the first month of treatment we can appreciate differences between groups in terms of diversity reduction.

	Negative outcome			Positive outcome		
	HIV-	**HIV+**	**Total**	**HIV-**	**HIV+**	**Total**
No MIC shift	5	11	**16**	12	0	**12**
MIC shift	6	4	**10**	1	1	**2**

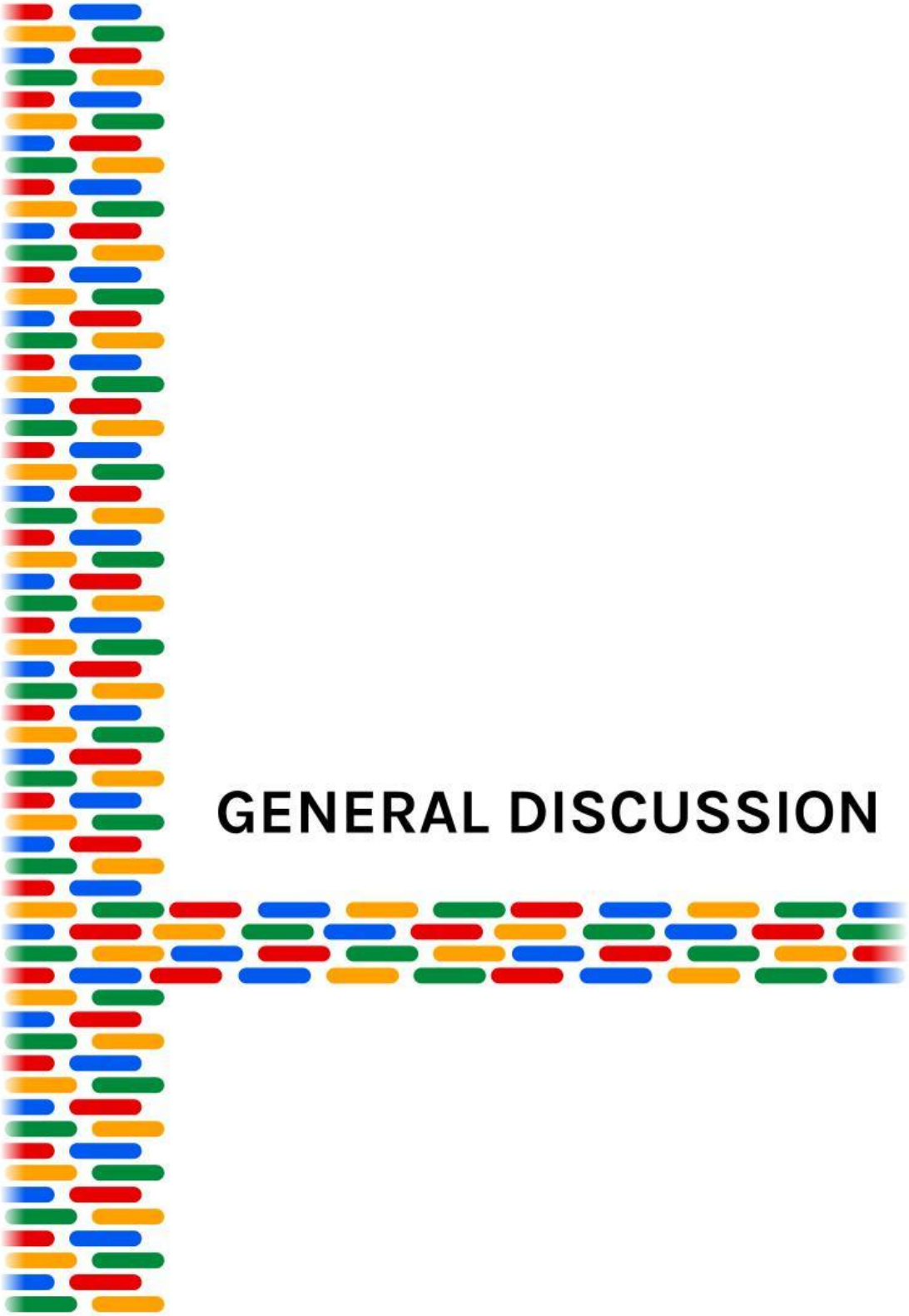
DISCUSSION

The impact of HIV in drug resistance development in TB has been a subject of debate for years, with many inconsistent reports across studies and settings [52]. Here, we evaluated its role in the early evolution of TB bacterial populations by means of deep whole-genome sequencing. To determine if HIV is associated with an early acquisition of resistance to antibiotics, we gathered two cohorts of patients with HIV and their respective control cases from Manhiça (Mozambique). The samples were collected at day 1, 7, 14 and 28 after the start of treatment or until the sputum culture became negative. We characterized the overall genetic diversity in each sample as the number of variable SNPs -or subpopulations- as a measure of the evolutionary potential of the whole bacterial population. We were able to reach very low frequencies of up to ~2% thanks to our superior read depth across the dataset. The initial analysis showed a higher overall diversity in HIV- patients, although there was a decreasing dynamic along the first month of treatment for both of the cohorts (Figure 14A). However, when inputting the data into a Cox regression model, we could appreciate that the HIV+ group had a differentially lower ability to reduce this initial diversity (Figure 14B). This can be used as an early surrogate of eventual treatment failure as it translates into a higher number of positive cultures after one month and, likely, the need for longer treatments. Therefore, with increased treatment lengths and as many studies have pointed out, there is a higher chance for drug resistance to develop, which is the main reason for treatment failure in most cases.

Then, we focused our analysis just on resistance-related genes from the 2021 WHO catalogue [49]. We found a similar pattern than with overall genetic diversity in both groups, with decreasing dynamics along the month of the study (Figure 15A). There are many low-frequency variants detected in these regions, suggesting that the generation of low-level heteroresistance is a common evolutionary pathway to adaptation at a population level. We then tested if there was a statistically significant accumulation of variants in these genes vs. the rest of the genome at every timepoint. This would confirm that these regions are indeed functionally relevant under selective pressure to adapt to the antibiotic treatment. The result of the analysis was that for the HIV- group, the variant enrichment becomes statistically significant starting at day 7 of treatment -which is expected- but the HIV+ group starts at day 1 (Figure 15B). The implications of this result are not fully clear, but it could be due to a transmission bias among HIV+ patients as they tend to have a worse prognosis. Additionally, we calculated the dN/dS trajectories of the resistance-related genes illustrating how, as soon as treatment starts, the evolutionary pressure it creates ends up purifying most of the synonymous variants in resistance-related genes for both groups (Figure 15C).

Finally, we sought to provide experimental evidence for the population dynamics of the low-frequency variants we detected in resistance-related genes. To do so, we determined the MIC values of six different antibiotics for the baseline and a follow-up sample, and tried to correlate the

appearance of MIC shifts to these variants (Table 3). As recent studies have shown, the emergence of low-level heteroresistance is one of the mechanisms responsible for treatment failure and it can be revealed by detecting subtle MIC shifts below the clinical breakpoint [131]. This fact has gained attention as it is completely missed by standard diagnostic tools, and our results show that 12 out of 40 patients experienced an MIC shift to at least one of the first-line antibiotics used in their treatment (Table 4). While HIV+ patients were not a substantial part of this group, 15/16 had a negative outcome whether they experienced MIC shifts or not, reinforcing the notion that they are more likely to end up having worse prognosis and treatment failure. These findings are probably part of the reason why more HIV+ patients end up developing drug resistance. Despite this cohort being very well controlled in terms of treatment adherence and achieving a successful outcome, we reason that many settings with both a high TB and HIV burden have increased odds of resistance acquisition due to the combination of both socioeconomic circumstances and the biological impact of the coinfection [127]. Our results highlight the importance of adapting treatment guidelines for HIV+ patients developing TB in certain settings, providing additional control measures and better follow-up to prevent the emergence of drug resistance.



GENERAL DISCUSSION

GENERAL DISCUSSION

Drug resistance is one of the major health challenges associated with the current tuberculosis epidemic. In some regions up to 50% of tuberculosis cases are drug-resistant and therefore impose a large economic burden on the country and public health services. Thus DR-TB threatens to hamper the progress in TB control that has been made in the last 30 years. One of the biggest problems is that we are largely unaware of the biological factors behind the epidemic and therefore cannot make accurate estimates of how to act in the future. In this dissertation we aimed to study DR-TB from two different but related angles. From the bacterial point of view, we have used large-scale genomic and functional analyses to understand the genetic architecture of isoniazid resistance, expanding our knowledge of the drug's mode of action and making predictions of which regions of the genome could be associated with resistance at the clinical level. From the host's point of view, we have studied the pathogen diversity and its interaction with the host at the site of infection using lung surgical samples, and evaluated the impact of an important comorbidity such as HIV on the early appearance of resistance. In summary, we have sought a better understanding of the bacterial and host factors that impact DR-TB and will allow the development of new diagnostic tools and epidemiological models to control DR-TB in the future.

In Chapter 1, we have developed an evolutionary functional genomics framework coupled with phylogenetic

association to characterize the genomic architecture of isoniazid resistance. We have seen many genes involved in cell wall biosynthesis, including the mycolic acids biosynthesis pathway, that have been highlighted by both the functional and the phylogenetic approaches, even though many of them are not used to predict resistance (Figure 4D). This fits in with the idea that resistance depends on more than the well-characterized diagnostic mutations and that bacteria can also acquire low-level resistance mutations or compensatory mutations. Given that isoniazid specifically targets this pathway, it is plausible that insertions in several genes in the pathway affect isoniazid sensitivity. However, as the cell wall is the first barrier of defense of the bacteria, it is difficult to determine whether these genes are important for isoniazid resistance exclusively or also for resistance to other antibiotics as well. For instance, our data shows that inserting genes in the *mce1* operon increases isoniazid resistance. This operon is proposed to be involved in the transport of cell wall components and disrupting the operon is associated with the accumulation of free mycolic acids [132–134]. This points to a role in cell wall remodeling and recycling and suggests that the operon can be involved in resistance to other antibiotics as well. Other genes such as those in the mycolic acid modification pathways are better candidates to affect isoniazid resistance exclusively. In any case, we expect many of the mutations in these genes to have low diagnostic value even if they actually contribute to resistance, which underlines the importance of systematic studies to understand antibiotic resistance.

Our analysis also highlighted genes involved in redox metabolism, which has previously been associated with isoniazid resistance [135]. This is in accordance with what is already known about the action of the antibiotic, as NADH is required for the formation of the isoniazid-NAD adduct. Additionally, NADH and NADPH are necessary for the activity of two genes in the FAS system, *inhA* and *fabG1*, which we know to be isoniazid resistance genes. However, as most of the oxidative metabolism genes we found to be associated with resistance were not directly involved in any of these pathways we concluded that this association depends on NADH homeostasis. Previous findings have suggested that NADH dehydrogenase gene *ndh* harbors putative resistance mutations [68] and our results showed that many similar genes could have such mutations as well. For instance, we found that genes encoding NADH dehydrogenases present in the electron transport chain, such as several genes in the *nuo* operon and *ndhA*, increased isoniazid resistance when mutated. It has already been shown that inhibiting the action of these genes increases intracellular levels of NADH [136] and that mutations in the NADH dehydrogenase Ndh (isoform to NdhA) lead to higher NADH levels and confer isoniazid resistance, maybe by preventing inhibition of the InhA enzyme [135]. Indeed, we found that mutations in *nuoJ*, *ndhA* and *ndh* have phylogenetic association with resistance, further confirming that NADH homeostasis plays a clinically relevant role in the evolution of isoniazid resistance. NADH can also affect resistance indirectly, as several stress responses specifically use the NADH:NAD+ ratio as a trigger [137]. In many cases, these responses

provide protection against antibiotic stress as well and are triggered as a result of exposure to the antibiotic. We identified several NADH sensors such as *regX3*, *dosT* and *pknG* which were functionally associated with isoniazid resistance. In addition to redox sensors, we also found other genes related to detoxification that were associated with resistance. For instance, insertions in genes involved in mycothiol biosynthesis produced a more resistant phenotype.

Once we understand which pathways are involved in isoniazid resistance, we can start exploring them as targets for future regimens. Many sensitivity-increasing features in the cell wall biosynthesis pathway show increased expression after exposure to isoniazid [138]. The insertion probably compromises the strength and integrity of the cell wall and makes those genes ideal targets for adjuvants that help potentiate the action of the antibiotics or even for new or repurposed therapies. For instance, peptidoglycan biosynthesis gene *ponA1* appears to be the target of the repurposed ceftazidime-avibactam combination [139]. We can also take advantage of the role of redox homeostasis in the action of isoniazid to design treatments and strategies to enhance the performance of the antibiotic. In their work, Flentie *et al.* reported a new compound, C10, that appears to revert the resistant phenotype of *katG* mutants and thus prevents the selection of isoniazid-resistant variants. The compound had been specifically selected to block tolerance to oxidative stress and it was shown to both increase sensitivity to isoniazid and promote the expression of energy metabolism related genes. Here, we have provided

experimental evidence of which specific genes can produce a particular resistance phenotype and which of those are relevant in clinical environments. We can now use this experimental framework and extrapolate it to other antibiotics, including recently-licensed drugs such as bedaquiline and delamanid to design better treatments with them, especially now that the WHO has recommended its use in first- and second-line therapy.

A very powerful feature of functional genomics is that it can be applied comparatively across strains and antibiotics. For instance, functional genomics showed that small genetic differences between characterized strains could be linked to differences in relative importance or essentiality of particular genes and in the way a strain acquires antibiotic resistance [140]. When we use functional genomics with different antibiotics we can find common patterns of resistance across drugs, such as the existence of an intrinsic resistome [64,141] and instances of cross-resistance. In our data we found that the F420 biosynthesis genes *fbiA-C*, which are associated with delamanid resistance [142], also confer isoniazid resistance when inactivated by insertion. These cross-resistance patterns are important as they inform us of the likelihood that the bacteria can develop resistance to two antibiotics that are administered in combination or sequentially, which will ultimately impact treatment success.

By combining a functional and phylogenetic approach we have shown that our candidate resistance determinants could increase sensitivity up to 2%. While the

impact at the population level seems minimal, the impact for the patient is important as isoniazid resistance is a major determinant for adverse clinical outcomes [143]. Furthermore, genome-wide association studies were unable to reveal any of these targets when thousands of genotypes-phenotypes [58] or detailed MIC measurements [144] were implemented. In our approach we are able to identify novel regions that are involved in isoniazid resistance as they increase sensitivity in sets of phenotypically resistant strains with no known causing mutation. Nevertheless, the use of those regions to diagnose isoniazid resistance leads to a decline in specificity suggesting that not all mutations in the target genes are involved in resistance. Furthermore, mutations in some of the regions are predictive of resistance in strains with very well characterized isoniazid resistance mutations suggesting that in some cases they can act either as compensatory mutations or early low-level, facilitating resistance mutations that preceded the well characterized isoniazid resistance mutations.

While it is certainly useful to identify new resistance determinants, we have proved how vital it is to use actual clinical data to discover relevant pathways and mechanisms, and this has always been limited by the geographical bias in the amount of sequencing data available. Despite having been reduced in the last few years, there is an important imbalance by which strains from low-burden countries are vastly overrepresented in comparison with those from high-burden TB settings. However, it is in those settings where we see the highest number of rare drug resistance

variants arising as a consequence of suboptimal treatment adherence, difficulty in its access or other similar factors. These variants often have higher fitness costs but are allowed to thrive in the conditions we just described, and they have increased chances of providing low-level resistance via non-canonical mechanisms. In addition, these types of settings also have a higher probability of multiple transmission events and superinfections are more common, resulting in complex cases. Consequently, if we want to have a better representation of the array of possibilities in which TB can acquire drug resistance, we need to recover as much diversity as we can from high-burden settings and, ideally, do it by sampling the actual site of the infection.

In Chapter 2, we explored the bacterial diversity of a surgical cohort from Georgia, a high burden DR-TB setting, thanks to access to different resected parts of the granuloma geography. We showed distinct patterns of diversity across individuals, and detected that seven out of the eighteen patients (39%) had evidence of infection by two phylogenetically unrelated genotypes, either in the same sample or in separate samples. This represents a high percentage when compared to the 5% found in single sputum patients of this dataset and percentages described by others [81]. Three surgical patients showed true mixed infections (multiple genotypes co-existing) and four are likely the result of superinfections due to multiple transmission events. The high rate of polyclonal infections in these patients suggests deficiencies in infection control in the setting [145] and agrees with epidemiological and model data showing that repeated exposure to infection as seen in

high-burden settings increases the risk of reinfections, disease progression and outbreaks [145,146]. In addition, polyclonal infections in the surgical cohort usually involved strains with different drug resistance patterns (5 out of 7 cases), a fact that can hamper successful treatment if not assessed correctly [124]. There are several studies showing that infection with multiple strains resulted in a poor treatment outcome, since the presence of any undetected resistance during standard treatment can propel the acquisition of further drug resistance [92]. Importantly, in our results polyclonal infections change, and many times fully revert, the drug resistance profile of the patients. Thus, follow-up DST samples should be implemented in these settings for better patient management.

In Figure 7, we show the theoretical scenarios to explain the natural history of infection for these patients. The fact that in several cases the second isolate is in a transmission cluster suggests that superinfection is one of the most common mechanisms. However, we cannot be sure if two strains were already infecting at baseline while only one is detected in sputum as we don't have access to lung samples before treatment. Similarly, we cannot test the hypothesis suggested from mice experiments that a secondary infection drives progression of an asymptomatic primary infection via immune response or by expression of resuscitation-promoting factors [92,147]. However our results contribute to the evidence from mice data [147] (but not from macaques [82]) that reinfection with a second strain can be common after a primary infection under certain

circumstances and calls to re-evaluate the natural history of TB in settings with rampant transmission.

The extent of polyclonal infections also has consequences to understand the protective potential of vaccines. Recent *in vivo* reinfection experiments suggested that a first episode of TB was more protective against a second episode than previously thought [85]. However, our results suggest that in a high-burden setting this might not be the case and future studies designed to corroborate this observation will be needed. Differences may be related to how well the experimental set-up recapitulates natural infection in such settings. In the experiments on macaques the same strain was used for re-challenge [82]. By contrast, in our dataset we frequently observe different strains, usually from different (sub-)lineages, infecting the same patient. In fact we did not see polyclonal infections with strains less than 100 SNP apart despite our phylogenetic approach being designed to identify cases up to 20 SNP apart. This raises the possibility that previous infection might indeed protect against superinfection with the same or a very similar strain, as observed in the macaque model. The differences with animal experiments may also be related to the nature of the re-challenge, as in TB transmission hotspots, the exposure to second infections is recurrent [148]. However, co-morbidities and social determinants may also play an important part in susceptibility of the host to second infections. Lastly, many clinical trials and vaccine studies use sputum as their basis for distinguishing between relapse and reinfection and our results suggest that it may be a poor correlate to ascertain lung polyclonal infection.

Beyond multiple genotypes, analysis of the genomic diversity within and around lesions showed very different patterns across patients. Some showed almost no diversity while others were highly diverse. In some cases (notably G-018) the distribution of drug resistance variants suggests gradients of antibiotic concentration as have been described in recent literature [28,87,149]. The fact that this pattern is not observed in all patients, may be explained by different pharmacodynamics, but also by the time difference until resection between patients as the granuloma walls get hardened over time. This would limit the penetration of drugs inside, which could affect selective pressure and thus the diversity we are able to recover from genomic data [87]. But drug pressure is not the only selective force that the pathogen encounters during infection. It is known that different immune-microenvironments exist within granulomas [150]. Chief among macrophage defense mechanisms is the production of reactive oxygen and nitrogen species [151]. Looking at sputum cultures, it has been proposed that an unintended effect of ROS is to increase the mutation rate of the bacteria. Our data corroborates those results showing that the mutational signature of ROS -increased number of transitions- was overall enriched in surgery samples compared to sputum samples. A recent study suggests that transition bias is linked to the acquisition of drug resistance variants [152]. It is thus tempting to link ROS-increased mutational supply to an accelerated acquisition of drug resistance mutations. Combined analyses of the bacterial population in the lung during infection in humans and relevant animal models [153],

drug penetration and pharmacodynamics [28] and lesion imaging [149] will help to better understand the bacterial population diversity driven by host pressures, how it's linked to the emergence and selection of drug resistant subpopulations and, ultimately, to relapses and treatment failure in clinical settings.

So far we have seen how the identification of the genomic drivers of resistance in tuberculosis and the correct application of the tools at our disposal in complicated settings are essential to build better epidemiological predictive models and to develop population-specific treatments. Until now, the socioeconomic determinants of antibiotic resistance have received most of the attention as major players in the MDR-TB epidemic. These factors, as we have previously discussed, include nonadherence to treatment and/or drug shortages leading to the acquisition of drug resistance due to suboptimal levels of drug circulation in the patient. However, host-pathogen interaction factors are being increasingly recognized as key players in this issue. Host factors implicated in a higher rate of resistance acquisition include specific genetic susceptibilities and comorbidities such as coinfection with HIV, which has been one of the main drivers of the TB epidemic since the resurgence of the disease in the 1980s [51].

Some studies in certain settings have suggested an association between HIV and DR-TB [52]. However, we do not know if the association is directly promoted by the comorbidity or if it is a secondary effect of the impact of the

disease on the patient's evolution. In favor of the first hypothesis, it has been shown that, on many occasions, drug levels are suboptimal in HIV+ patients. In some cases due to interactions with antiretroviral treatment, and in others due to the particular microenvironment in the lung created by the condition (small to no granuloma formation). Whatever the cause may be, if drug levels are below the clinical recommended concentration, the acquisition of drug resistance mutations is evolutionary promoted. In favor of the second hypothesis -drug resistance being a side effect of the HIV condition- it is known that longer treatments and relapses are linked to greater drug resistance rates as a result of the increased difficulty in patient follow-up, especially in low-income settings where there is still significant social stigma. It is important to differentiate between the two hypotheses because the second one implies that drug resistance associated with HIV should be controlled by tuberculosis control programs, while the first hypothesis indicates that action on the global HIV epidemic will have a net benefit in the future DR-TB burden.

In Chapter 3, we had the opportunity to test the former two hypotheses by trying to estimate the emergence rate of extremely low drug-resistance variants during treatment and whether it is significantly different in patients with and without HIV. For this, we recruited two cohorts of patients in Manhiça (Mozambique) that were diagnosed with regular TB infection or TB-HIV coinfection, and collected four clinical samples during the first month of treatment to delve into the evolution of the population’s genetic diversity through deep whole-genome sequencing. The first analysis

showed how there is a quantitative difference in the amount of overall diversity between the two groups (Figure 14A). While both dynamics are descending along the study period, the diversity we recover from the baseline samples is significantly higher in the HIV- group. This is explained by the fact that bacillary loads in HIV-coinfected individuals are usually lower due to their immune status having difficulties in forming granulomas in which TB can replicate faster [154]. So less replication means a smaller bacterial population and thus less chances for variants to randomly appear under no selective pressure, correlating with our observation of less overall genetic diversity.

However, while a lower diversity in HIV+ patients could be an expected result, we chose to evaluate the rate at which it disappears during the first month of treatment and see if there was a difference between the two cohorts. We defined an outcome of >90% initial diversity reduction -or confirmed negative culture, theoretically representing a 100% reduction- for each patient to input the data into a Cox regression model and build a survival curve from the bacterial population's point of view (Figure 14B). The result was a 38% hazard ratio for HIV+ patients, suggesting that their immune status may contribute towards a reduced capacity of variant purification. This result has several clinical implications. First, it supports the notion that HIV+ individuals have an observed worse prognosis, given that despite starting from a smaller pool of diversity, they can have problems to effectively reduce it even with the help of antibiotics. Second, the fact that they have increased chances to remain culture-positive after one month of

following the drug regime means that they could end up having longer treatments, which in turn boosts the probability to acquire drug resistance mutations and eventually fail treatment. And third, this negative clinical outcome may be predicted by an early assessment of the patient's initial diversity and its evolution as a tool to forecast potential treatment failures.

Moving forward from the overall genetic diversity of the samples, we shifted focus to analyze what was happening in resistance-associated regions. The pattern of population dynamics in these genes was similar to that of the rest of the genome (Figure 15A), as we could detect a good amount of variants -most of them at low-frequency- that were present from the beginning and during the course of treatment. We then investigated if there was an enrichment of mutations in this group of genes compared to the rest of the genome. Deploying a permutation test (Figure 15B) we identified a statistically-significant accumulation of variants starting at day 7 for HIV- patients, while it happened from the baseline sample for coinfected individuals. We also characterized the dN/dS dynamics of first-line resistance-related genes to confirm their evolutionary trajectories during treatment (Figure 15C), finding similar patterns between both cohorts. The ratios start at neutral values, rapidly increase once the drug regime is initiated and then start to go down again. We can appreciate a general selective sweep of synonymous variants under treatment pressure during these early stages that is similar regardless of immune status. It is important to remark that resistance to isoniazid has been reported as the first one to appear in

most cases, as we see here, and there is also an arising signal for ethambutol resistance, another antibiotic that targets the cell wall biosynthesis like isoniazid. These results as a whole bolster the conception of low-level heteroresistance as an evolutionary mechanism of adaptation under antibiotic selective pressure, where there is a constant turnover of alleles until a subpopulation can progress into clinical-level canonical resistance and subsequent fixation [41]. This may happen directly by acquiring a high-level resistance variant with little to no fitness cost in a drug target or activator enzyme, or progressing through consecutive but more costly mutations that increase tolerance to the drug/s until the first scenario is reached. In this context, a lower fitness cost of resistance in the absence of adequate immune pressure for HIV+ patients might explain the fact that we observe an enrichment of variants at baseline for this group. It may also have to do with different transmission dynamics or a broader infection bottleneck, but in any case, further dedicated studies with a bigger sample size would be needed to explore this finding.

We next sought to provide experimental evidence for the variants that were apparently causing low-level heteroresistance. To this end, we determined the MIC values of six different drugs (isoniazid, rifampicin, ethambutol, streptomycin, levofloxacin and amikacin) for the baseline sample of each patient and a follow-up sample. Comparing both time points allowed us to detect MIC shifts that developed during the course of treatment. We then could propose candidate variants appearing in those samples to explain the shifts. In general, the MIC values are slightly

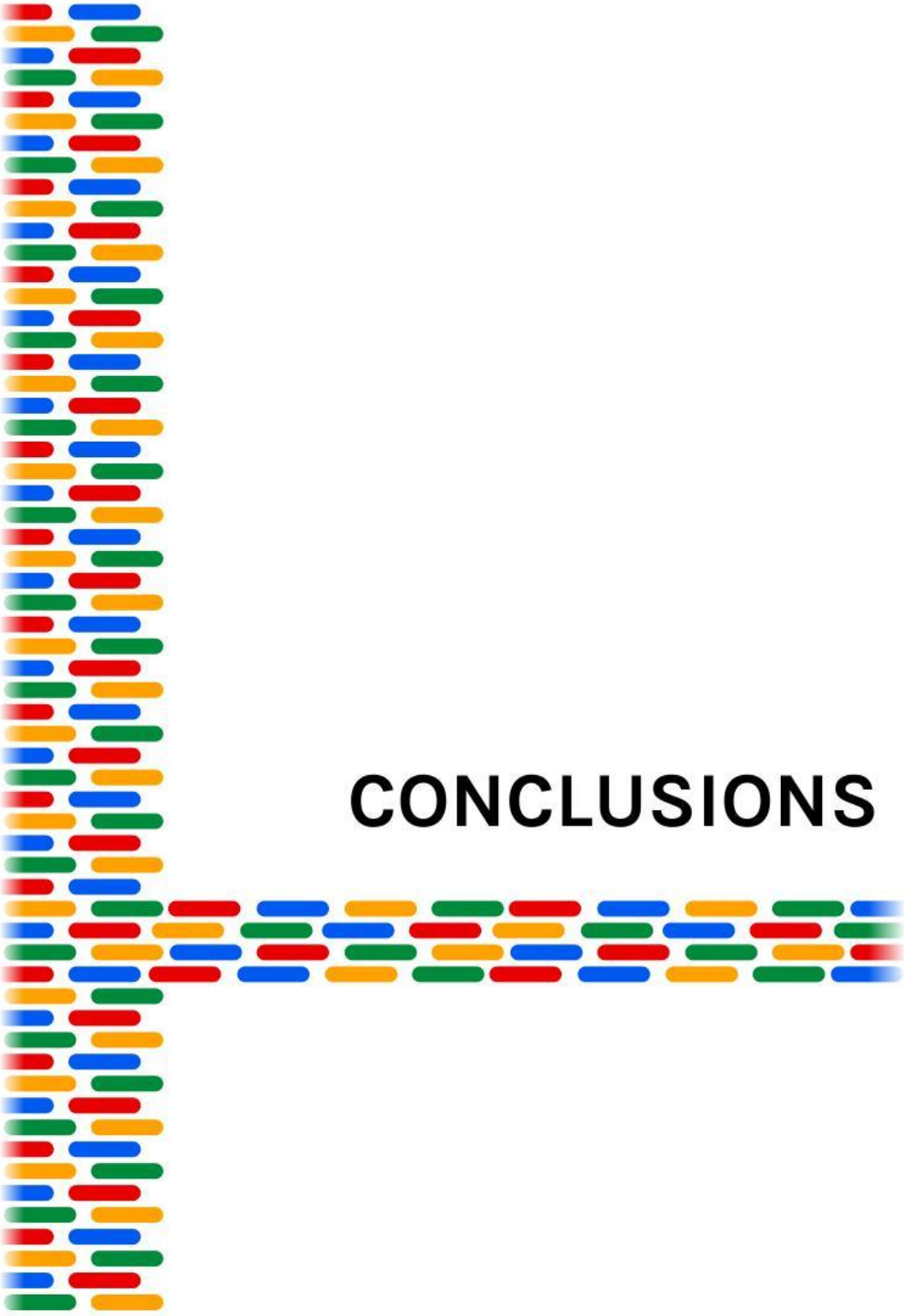
shifted to higher concentration as treatment progresses (Figure 16). Individually, examining Table 3, this is especially true for some patients like P47 or P50, that even surpassed the isoniazid clinical breakpoint to be considered resistant. In most cases, MICs either remain unchanged or are increased as a whole. However, a few cases experimented positive shifts for one or more antibiotics while exhibiting negative shifts for other/s (e.g. P11, P20, P28, P30, P50) hinting at putative underlying mechanisms of collateral sensitivity, a concept that is mostly unexplored in TB [155].

Detecting these subtle MIC shifts below the drugs clinical breakpoint has been proven to serve as a predictor for worse clinical outcomes [131]. However, as all patients in this study were eventually cured, we examined the reduction in overall genetic diversity as a surrogate for treatment efficacy. Table 4 summarizes the result of classifying the 40 eligible patients according to the presence of a MIC shift to first-line antibiotics and their outcome after a month of treatment. As previously explained, these are defined as 'positive outcome' if the patient managed to reduce their initial bacterial diversity at least 90% -or was culture-negative at day 28- and 'negative outcome' if they could not reduce it. 10 out of 12 patients that experienced a MIC shift during early treatment had a 'negative outcome', failing to reduce their overall initial diversity. Even in this early stage of treatment, MIC shifts present themselves as a promising indicator of the eventual clinical outcome of the patient. These results illustrate the notion that there is a percentage of patients that take longer than normal to resolve the infection and they appear to be associated with

higher genetic diversity of the bacterial population and infrequent heteroresistance mutations that eventually drive them towards treatment failure. This is an especially relevant factor to take into account when testing the newer drug regimes that are being sponsored by the WHO, which seek not only to replace old drugs, but also to reduce overall treatment time [11,12]. Rare mutations are probably responsible for an initial worse response to any drug regime, even though patients are eventually cured given enough time.

In this dissertation, we have focused on the urgent problem of drug-resistant tuberculosis and how it remains a deadly disease in defiance of the strict measures taken towards its treatment. The ongoing spread of DR-TB in many regions of the globe is one of the most imperative and crucial challenges to achieve global TB control. Weaker healthcare systems in low-income settings, the amplification of drug resistance through inadequate or suboptimal treatment, and continued transmission are among the main causes for this spread. Bacterial genomics has the potential to revolutionize the diagnosis of DR-TB in the near future, and it has already been implemented in some public health systems like the UK or the Netherlands replacing standard culture. Whether this reality can reach high-burden settings depends on many factors both political and economic. Wet lab and bioinformatic expertise need to be strongly promoted by governments to nourish a solid foundation in which to implement the new diagnostic techniques that whole-genome sequencing enables. We have shown how an early detection of drug resistance, either canonical or

non-canonical, is pivotal to steer patient treatment in an adequate direction, especially in MDR and XDR-TB cases that pose a formidable challenge to clinicians. Timely diagnosis reduces treatment time, ensures better clinical outcomes and prevents further transmission and superinfections. Likewise, we are living a paradigm change by which host factors are increasingly studied and recognised as key players in the success of DR-TB treatment. Discerning the role of comorbidities traditionally associated with tuberculosis will also bring us closer to offering authentic tailored drug regimes to all patients. The techniques and knowledge presented in this dissertation will hopefully contribute towards these goals and serve as a stepping stone towards improving the global control of drug-resistant tuberculosis in the near future.



CONCLUSIONS

CONCLUSIONS

1. Functional genomics allows an accurate detection of antibiotic resistance determinants, target genes and pathways, yet it is always necessary to compare with clinical data to understand which regions are relevant *in vivo*.

2. The remarkably complex mycobacterial cell wall is essential for *Mycobacterium tuberculosis* success, and the biosynthesis of mycolic acids in particular is a key pathway in regards to isoniazid sensitivity that contains a multitude of potential therapeutic targets that are currently uninvestigated.

3. Redox metabolism is another less explored, non-canonical and low-level resistance mechanism with potential for new therapies. Subtle changes in the NADH/NAD+ bacterial balance can provide the necessary evolutionary space for clinical-level isoniazid resistance to be developed.

4. The evolutionary functional genomics framework presented in this dissertation can be applied to new antibiotics early into their deployment to anticipate their possible mechanisms of resistance in a prospective manner.

5. By carefully sampling TB lung lesions, we are able to trace the changing dynamics of different bacterial subpopulations across the geography of the granuloma using genomics, providing a much more complete picture of bacterial diversity than sputum samples.

6. These surgical samples have revealed an elevated number of polyclonal infections in Georgia, a high-burden MDR-TB country, suggesting that we are underestimating this phenomenon in other similar settings.

7. Polyclonal infections can negatively impact treatment outcomes if left undetected, as they often consist of genetically distant strains that can have different drug resistance profiles. Follow-up sputum analysis would be recommended in high-burden MDR-TB settings with significant transmission rates.

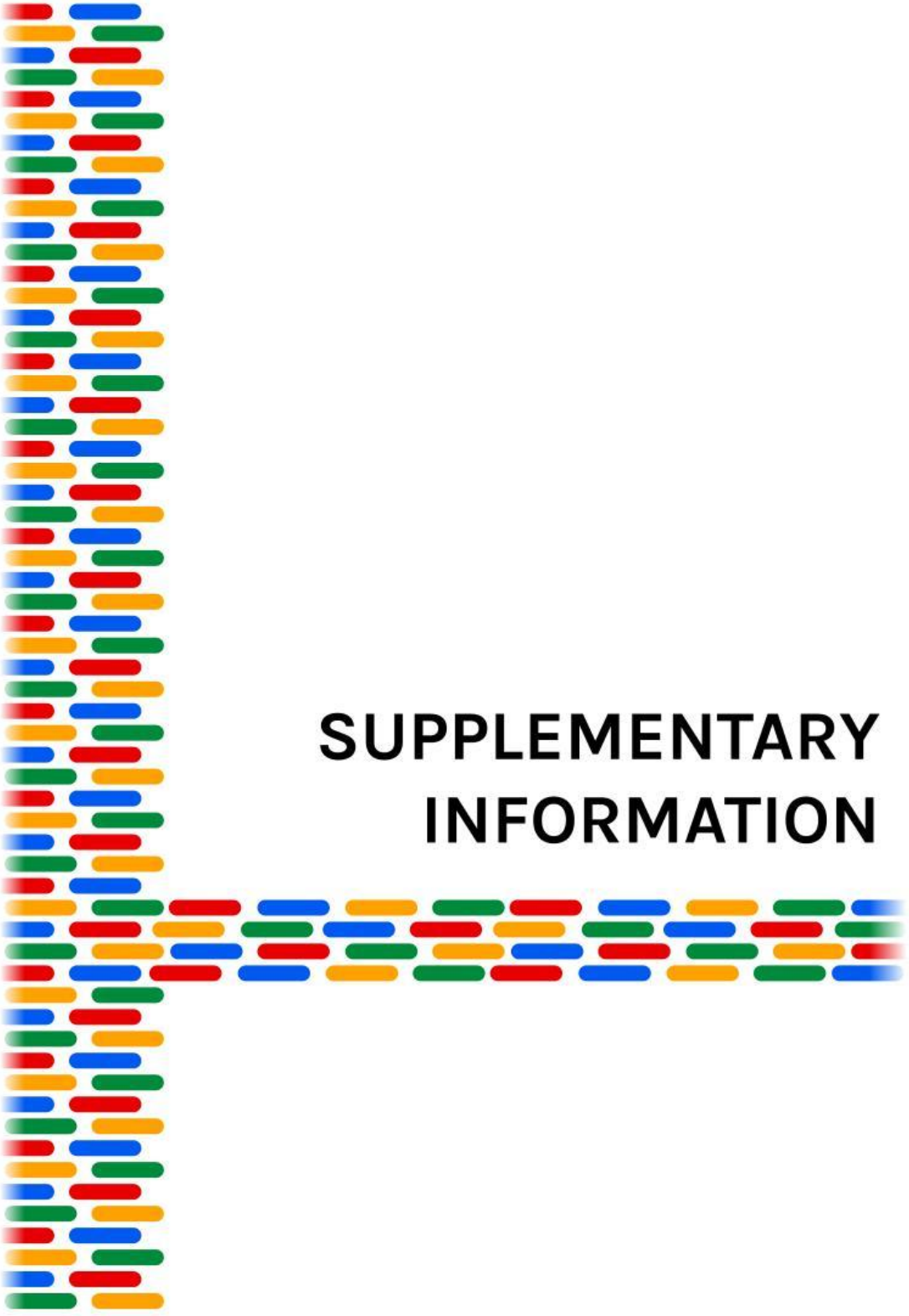
8. The immune pressure of the granuloma microenvironment can create a distinct mutational signature during the within-host evolution of *Mycobacterium tuberculosis* as a result of an accumulation of reactive oxygen species.

9. By investigating serial samples from TB-HIV coinfected patients during the first month of treatment, we detect a higher overall genetic diversity in HIV- patients, which decreases along treatment differentially for HIV+ patients. Immune deficiency

appears to reduce the capacity for diversity reduction and can be translated into worse clinical outcomes.

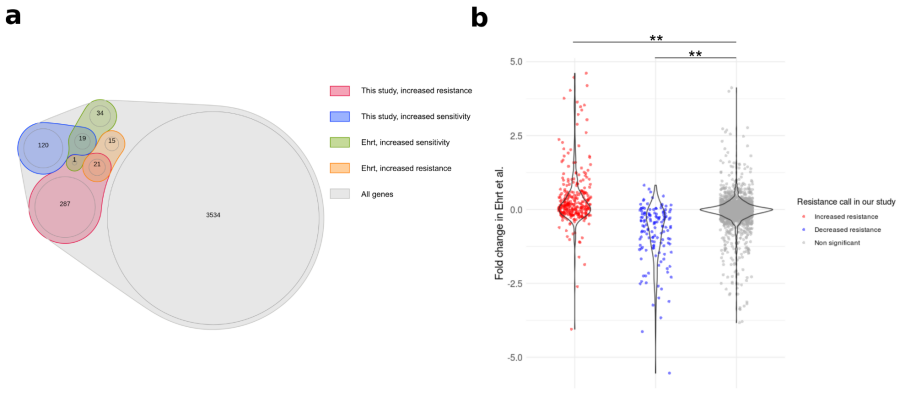
10. Using deep whole-genome sequencing, we can trace the early appearance of heteroresistance as a result of treatment selective pressure. There is a significant accumulation of non-synonymous low-frequency mutations in genes related to drug resistance.

11. Shifts in minimum inhibitory concentration during early treatment can serve as a clinical tool to predict a reduced capacity to eliminate bacterial diversity. This, in turn, could be an indicator of worse prognosis and eventual treatment failure.

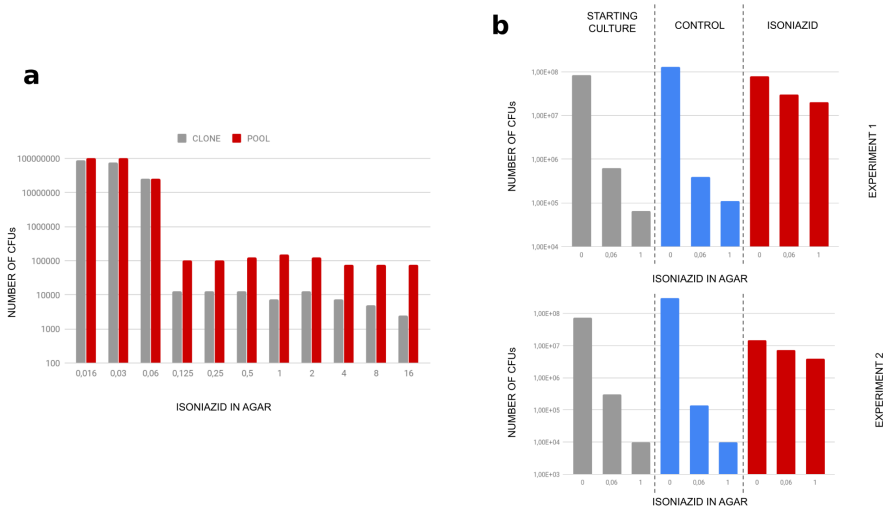


SUPPLEMENTARY INFORMATION

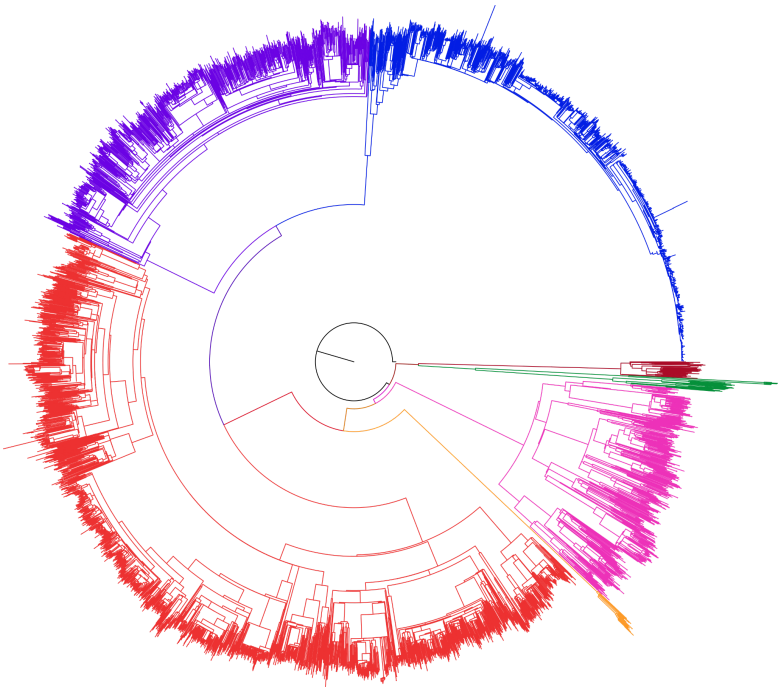
SUPPLEMENTARY INFORMATION



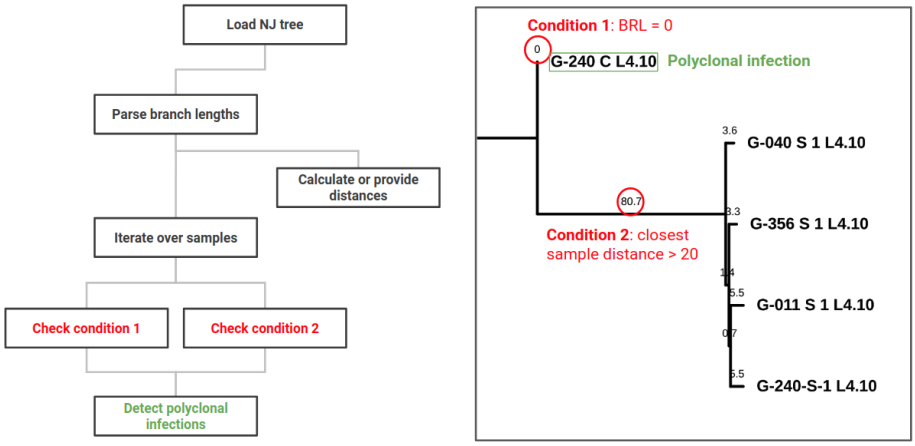
Supplementary Figure 1. a) Coincidence in resistance-altering features detected by functional genomics between our study and that of Ehrt et al. b) Our resistance-altering features showed similar phenotypes in Xu, W. et al. Chemical Genetic Interaction Profiling Reveals Determinants of Intrinsic Antibiotic Resistance in Mycobacterium tuberculosis. Antimicrob. Agents Chemother. 61, (2017).



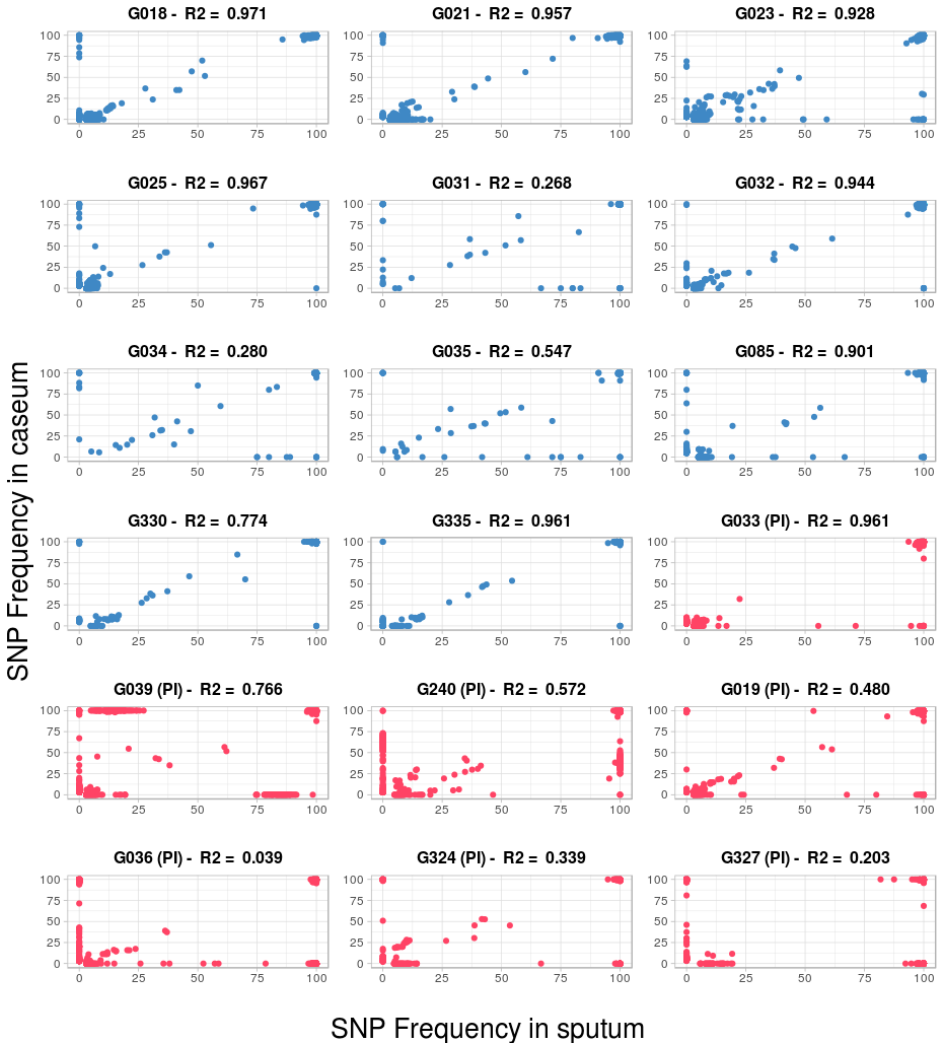
Supplementary Figure 2. a) The pool contains 10 times more isoniazid resistant mutants than its parental clone. b) Throughout the experiment, the frequency of isoniazid resistant mutants increased specifically in populations treated with the antibiotic.



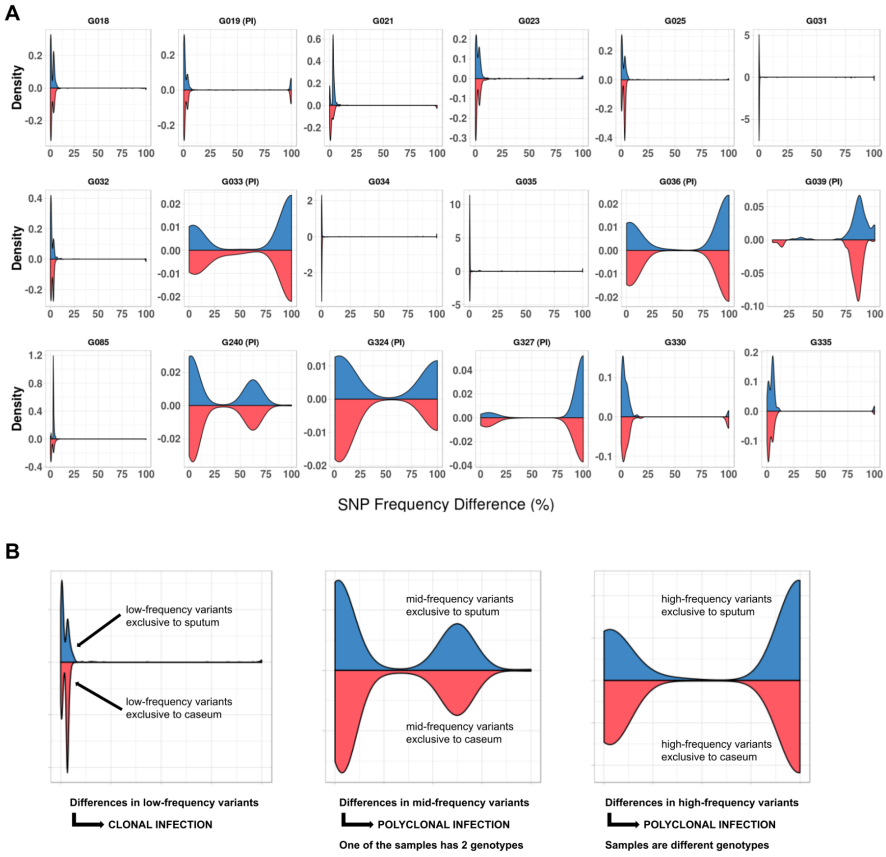
Supplementary Figure 3. Global phylogeny generated with a collated dataset from various sources.



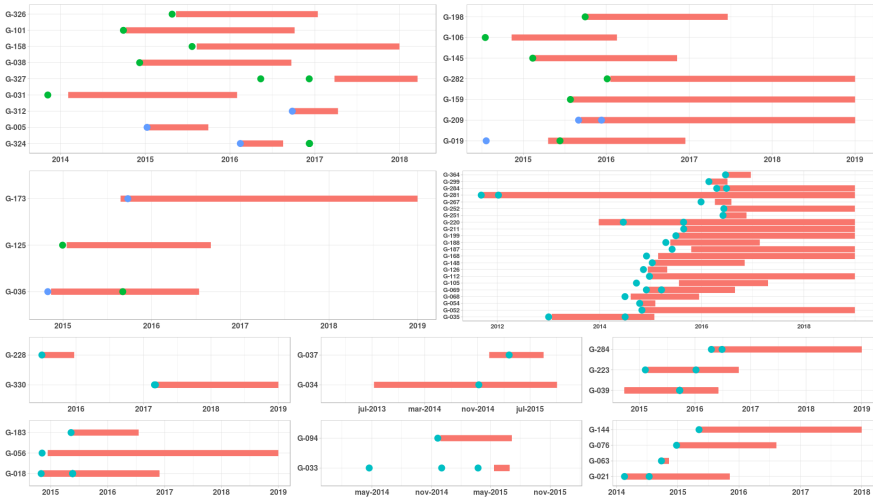
Supplementary Figure 4. Flowchart of phylogenetic identification of polyclonal infections. A Python script loads a NJ tree and scans for isolates that show terminal branch length equal to zero and distance to other isolates higher than 20 fixed SNPs. An example for one of the surgical patients is provided: the caseum sample from patient G240 is a mix of their sputum genotype and G040.



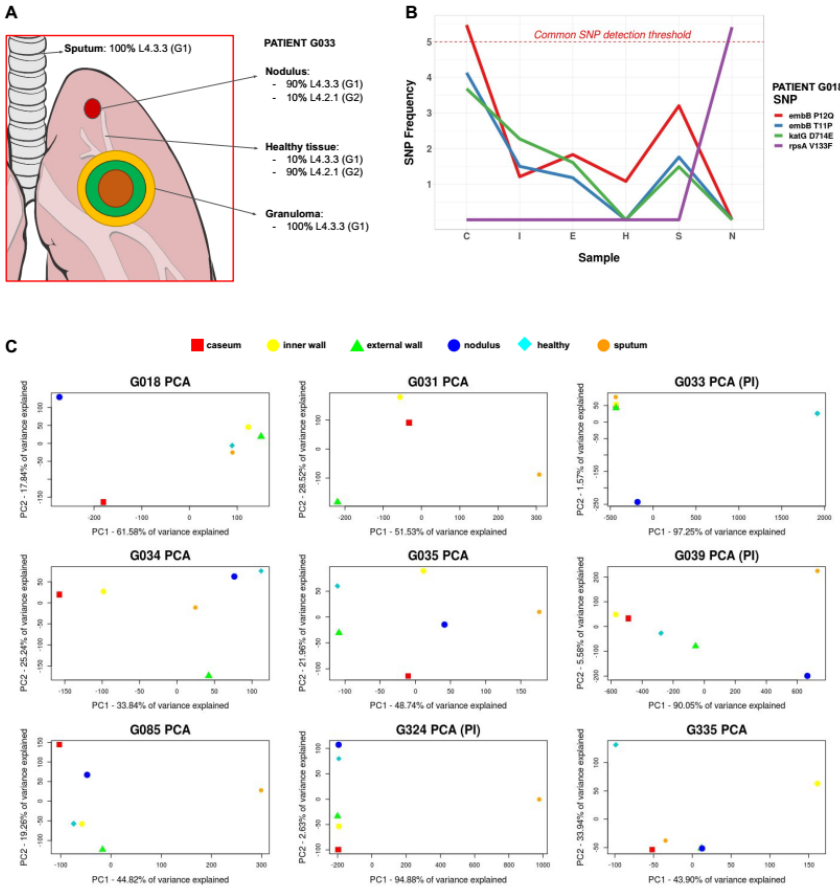
Supplementary Figure 5. Correlation between sputum and caseum SNP frequencies for all surgery patients. Blue graphs correspond to clonal infections and red graphs to polyclonal infections. R^2 usually decreases when multiple genotypes are involved in an infection. Abbreviations: PI, polyclonal infection.



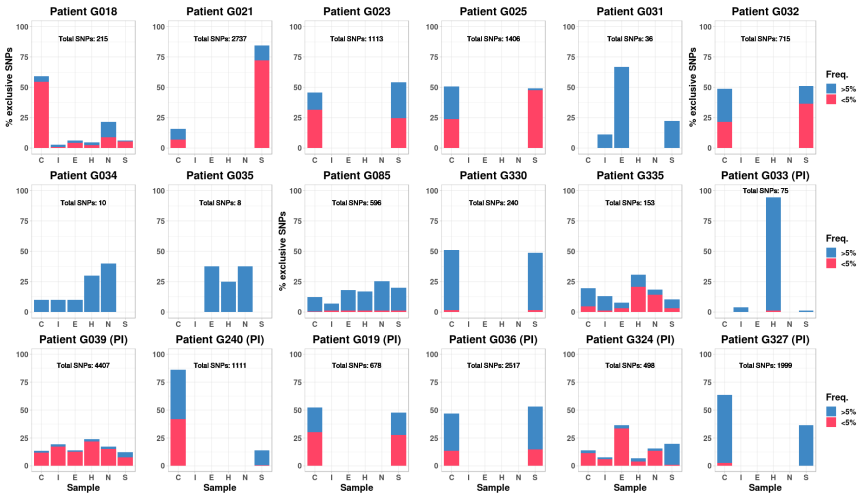
Supplementary Figure 6. Density profiles of SNP frequency differences between sample pairs (sputum in blue, caseum in red) in the surgery cohort. A) Different patterns reflect clonal infections (all differences in very low frequency range) or polyclonal infections (differences found in middle and/or high frequencies). B) Details on how to interpret the density graphs. Abbreviations: PI, polyclonal infection.



Supplementary Figure 7. Timeline of surgical patients involved in transmission clusters. The surgical patient is always depicted at the bottom. If the paired samples belong to different clusters, they are represented in two colors and patients of those two clusters are in the same graph identified with colors (blue = sputum cluster, green = surgical cluster).



Supplementary Figure 8. Additional details on the surgical cohort. A) Details on G033 genotype distribution. B) Example of a set of SNPs in resistance-related genes that are originally above 3% frequency in the caseum sample and how we can trace their frequencies across the rest of the lesion. As we validate those variants as not spurious, it is highly likely that their detection in samples below 3% is accurate. C) PCA analysis for all nine multi-sample surgical patients. Abbreviations: G1, genotype 1; G2, genotype 2; PI, polyclonal infection.



Supplementary Figure 9. Exclusive SNPs among isolates of surgery patients. Bars represent the percentage of the total private SNPs, also noted in each individual graph. Each sample type is divided into blue (SNP freq. >5%) and red bars (SNP freq. 3-5%). Abbreviations: PI, polyclonal infection.

Supplementary Table 1. Oligos used during the experiment for library preparation.

Oligo ID	Sequence	Modification	Comment
Adap1_Lig	TACCACGACCA AGATGTGTATAAGAGACAGTGNNANNANNN	3'AmC7	Ligation to gDNA fragments
Adap2_NX_Lig	TGGTCGTGGTAT	-	Ligation to gDNA fragments
Adap2_NX_PCR	GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAG	-	Tn-Junction pcr, complementary to Adap2
T7_NX_PCR	TCGTCGGCAGCGTCAGATGTGTATAAGAGA CAGCGGGGACTTATCAGCCAACC	-	Tn-Junction pcr, complementary to T7 ending sequence

Supplementary Table 2. Details of serial sputa pairs dataset. * G297 was a special case in which both sputum samples had three different genotypes at once (L4.3.3, L4.8 and L2 for S1; L4.3.3, L2.2.9 and L2.2.10), but only one of them (L4.3.3) was a match.

Patient	G1	G2	G1 Profile	G2 Profile	Distance	Days Diff.	Infection
G-065	L2.2.9	-	Pre-XDR	-	0	267	Clonal
G-069	L2.2.9	-	MDR	-	0	108	Clonal
G-078	L2.2.10	-	MDR	-	0	294	Clonal
G-128	L2.2.9	L1.1.1.1	MDR	MDR	1740	318	Polyclonal
G-169	L4.3.3	L2.2.9	Susceptible	XDR	1185	352	Polyclonal
G-205	L2.2.9	-	MDR	-	0	2	Clonal
G-208	L2.2.10	-	Pre-XDR	-	1	244	Clonal
G-209	L4.8	L2.2.10	Susceptible	Pre-XDR	1168	101	Polyclonal
G-213	L2.2.10	-	Pre-XDR	-	0	595	Clonal
G-214	L2.2.10	-	Pre-XDR	-	0	57	Clonal
G-215	L2.2.9	-	Pre-XDR	-	0	190	Clonal
G-216	L2.2.10	-	Poly-res	-	2	91	Clonal
G-217	L2.2.10	-	Pre-XDR	-	0	247	Clonal
G-219	L2.2.9	-	XDR	-	1	28	Clonal
G-220	L2.2.9	L2.2.9	Pre-XDR	-	10	431	Clonal
G-221	L4.3.3	-	MDR	-	0	62	Clonal
G-223	L4	L2.2.9	MDR	XDR	1128	333	Polyclonal

G-224	L2.2.10	L2.2.10	Pre-XDR	Pre-XDR	1105	38	Polyclonal
G-225	L2.2.9	-	MDR	-	0	41	Clonal
G-227	L2.2.10	-	Mono-INH	-	1	92	Clonal
G-229	L4.6.2	-	Susceptible	-	0	31	Clonal
G-230	L4.8	-	Susceptible	-	0	13	Clonal
G-246	L4.2.1	L2.2.10	Pre-XDR	XDR	811	24	Polyclonal
G-250	L2.2.10	-	MDR	-	0	41	Clonal
G-255	L2.2.9	-	MDR	-	0	75	Clonal
G-258	L4.8	-	Susceptible	-	2	125	Clonal
G-262	L2.2.10	-	Pre-XDR	-	0	95	Clonal
G-270	L2.2.10	-	Pre-XDR	-	0	97	Clonal
G-275	L2.2.9	-	MDR	-	0	45	Clonal
G-276	L2.2.10	-	MDR	-	0	39	Clonal
G-278	L2.2.10	-	XDR	-	3	122	Clonal
G-280	L2.2.10	-	Pre-XDR	-	0	3	Clonal
G-281	L2.2.9	-	MDR	-	1	123	Clonal
G-284	L2.2.9	-	Pre-XDR	-	0	70	Clonal
G-285	L4.2.1	-	MDR	-	0	266	Clonal
G-287	L2.2.10	-	MDR	-	0	68	Clonal
G-297	*	*	*	*	*	117	Polyclonal
G-315	L2.2.10	-	Susceptible	-	0	1	Clonal

Supplementary Note 1: Additional details on surgical patients

When analyzing the surgical patient dataset, we first hypothesized that if there are gradients of drug penetration across the lesion as reported in the literature, this may translate into differences of MTB genomic diversity across these positions in and around TB lesion, and that these differences will be observed both in drug resistance associated genes as well as across the whole MTB genome. As shown in Supplementary Figures 8 and 9, there are important differences in terms of diversity among patient samples. In patient G018, most of the MTB diversity accumulated in the granuloma center. When focusing on exclusive SNPs, most of these were located in the center of the lesion (127/215 SNPs or 59%). The caseum was mainly dominated by variants at 3-5% frequency, suggesting that these were relatively recent. By contrast, the rest of the samples, including the sputum, showed a lower number of low frequency variants, probably reflecting population bottlenecks. In fact, the principal component analysis for this patient (Supplementary Figure 8C) shows that the diagnostic sputum sample was closer to the MTB diversity seen in the sample from the nearby healthy tissue. Additionally, the remote nodule from this patient did not share many of the mutations found in the other samples, which however, may be partially explained by a lower sequencing depth. For this patient, we were also interested in knowing how well the MTB diversity in the sputum sample was represented by the variants in the surgical samples. For

this analysis, we pooled all low frequency variants identified in the surgical samples and compared them to the sputum sample. Only 13 out of the 461 (2.8%) low-frequency variants seen in the sputum were not present in any of the surgical samples. Twelve of these thirteen variants were at 3-5% frequency, aside from one at 74%. In addition, we found some mutations arising in the center of the lesion in resistance-associated genes and not described before in drug resistance catalogs. These cavity center mutations were at very low frequencies in the sputum culture sample of the patient, well below common detection thresholds (Supplementary Figure 8B).

Some other interesting cases are patients G335, G324 or G033. In patient G335, we detected just one genotype and we could not find any drug resistance-associated mutation. This is consistent with the phenotypic DST results of this patient. The PCA showed that the sputum sample was closer to the one from the cavity center, but the differences across sites were minimal and most of the variants were shared between samples (Supplementary Figure 8C). Patient G324 was another case of polyclonal infection, with a sputum sample harboring a different strain from the one present in the other surgical samples. The PCA reflected this, showing surgery samples separating along PC2 and only the sputum isolate far away from them in PC1. We also did not find any trace of the sputum strain in the caseum when looking at low frequencies. Patient G033 was infected with two different L4 sublineages (G1=L4.3.3, G2=L4.2.1) in the nodule (90% G1, 10% G2) and healthy tissue (10% G1, 90% G2). The whole granuloma (C, I, E) was 100% G1

(Supplementary Figure 8A). This was also reflected in the PCA , where both of those samples did not cluster with the other samples, nor with each other. Remarkably, G1 was XDR while G2 was fully drug susceptible.

Supplementary Note 2: Genomic predictions vs DST accuracy evaluation

Predictions of the phenotype for first line drugs using sequencing data and available catalogs (see Chapter 2, *Methods*) was overall accurate in this dataset (sensitivity 0.938, specificity 0.994, see Figure 10). Most drug resistance-associated mutations for first line drugs were already fixed in the population (frequency > 90%) as expected for patients that have undergone surgery after failing their first line treatment. In addition, eight variants present in catalogs were below standard levels of detection (frequency < 10%), however they were always concomitant to other known variants for the same drug so they do not explain potential disagreements between genotype and phenotype. Genotype failed to predict some of the phenotypic DST results. These include one kanamycin, 5 ethambutol, 3 para-aminosalicylic acid, 2 streptomycin and 1 capreomycin discrepancies. We suggest that two new mutations (2747471AG, *folC* I43T; 3073679CA, intergenic *dfrA-thyA*) might explain PAS resistance not detected by available catalogs, as both of this genomic features have been described to be involved in the drug metabolism. Further studies would be needed to assess the role of these mutations. We were not able to pinpoint any mutations to explain the rest of the discrepancies.

REFERENCES

REFERENCES

1. Baquero F. Threats of antibiotic resistance: an obliged reappraisal. Int Microbiol. 2021 Nov;24(4):499–506.

2. Tacconelli E, Sifakis F, Harbarth S, Schrijver R, van Mourik M, Voss A, et al. Surveillance for control of antimicrobial resistance. Lancet Infect Dis. 2018 Mar;18(3):e99–106.

3. Davies J, Davies D. Origins and evolution of antibiotic resistance. Microbiol Mol Biol Rev. 2010 Sep;74(3):417–33.

4. Chahine EB, Dougherty JA, Thornby KA, Guirguis EH. Antibiotic Approvals in the Last Decade: Are We Keeping Up With Resistance? Ann Pharmacother. 2022 Apr;56(4):441–62.

5. Laxminarayan R, Duse A, Wattal C, Zaidi AKM, Wertheim HFL, Sumpradit N, et al. Antibiotic resistance-the need for global solutions. Lancet Infect Dis. 2013 Dec;13(12):1057–98.

6. Uddin TM, Chakraborty AJ, Khusro A, Zidan BRM, Mitra S, Emran TB, et al. Antibiotic resistance in microbes: History, mechanisms, therapeutic strategies and future prospects. J Infect Public Health. 2021 Dec;14(12):1750–66.

7. Pai M, Behr MA, Dowdy D, Dheda K, Divangahi M, Boehme CC, et al. Tuberculosis. Nat Rev Dis Primers. 2016 Oct 27;2:16076.

8. Ngabonziza JCS, Loiseau C, Marceau M, Jouet A, Menardo F, Tzfadia O, et al. A sister lineage of the Mycobacterium tuberculosis complex discovered in the African Great Lakes region. Nat Commun. 2020 Jun 9;11(1):2917.

9. Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, et al. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. Nat Genet. 2013 Oct;45(10):1176–82.

10. Freschi L, Vargas R Jr, Husain A, Kamal SMM, Skrahina A, Tahseen S, et al. Population structure, biogeography and transmissibility of Mycobacterium tuberculosis. Nat Commun. 2021 Oct 20;12(1):6099.

11. World Health Organization. WHO consolidated guidelines on tuberculosis. Module 4: treatment - drug-resistant tuberculosis treatment. World Health Organization; 2020. 120 p.

12. Vanino E, Granozzi B, Akkerman OW, Munoz-Torrico M, Palmieri F, Seaworth B, et al. Update of drug-resistant tuberculosis treatment guidelines: A turning point. Int J Infect Dis [Internet]. 2023 Mar 12; Available from: http://dx.doi.org/10.1016/j.ijid.2023.03.013

13. Van Rie A, Walker T, de Jong B, Rupasinghe P, Rivière E, Dartois V, et al. Balancing access to BPaLM regimens and risk of resistance. Lancet Infect Dis. 2022 Oct;22(10):1411–2.

14. Global tuberculosis report 2021. World Health Organization; 2021. 57 p.

15. Jarlier V, Nikaido H. Mycobacterial cell wall: structure

and role in natural resistance to antibiotics. FEMS Microbiol Lett. 1994 Oct 15;123(1-2):11–8.

16. Abraham EP, Chain E, Fletcher CM, Gardner AD, Heatley NG, Jennings MA, et al. Further observations on penicillin. Lancet. 1941 Aug;238(6155):177–89.

17. Zaunbrecher MA, Sikes RD Jr, Metchock B, Shinnick TM, Posey JE. Overexpression of the chromosomally encoded aminoglycoside acetyltransferase eis confers kanamycin resistance in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A. 2009 Nov 24;106(47):20004–9.

18. Adams KN, Takaki K, Connolly LE, Wiedenhoft H, Winglee K, Humbert O, et al. Drug tolerance in replicating mycobacteria mediated by a macrophage-induced efflux mechanism. Cell. 2011 Apr 1;145(1):39–53.

19. Pule CM, Sampson SL, Warren RM, Black PA, van Helden PD, Victor TC, et al. Efflux pump inhibitors: targeting mycobacterial efflux systems to enhance TB therapy. J Antimicrob Chemother. 2016 Jan;71(1):17–26.

20. Jones RM, Adams KN, Eldesouky HE, Sherman DR. The evolving biology of drug resistance. Front Cell Infect Microbiol. 2022 Oct 5;12:1027394.

21. Campbell EA, Korzheva N, Mustaev A, Murakami K, Nair S, Goldfarb A, et al. Structural mechanism for rifampicin inhibition of bacterial rna polymerase. Cell. 2001 Mar 23;104(6):901–12.

22. Pym AS, Saint-Joanis B, Cole ST. Effect of katG mutations on the virulence of Mycobacterium

tuberculosis and the implication for transmission in humans. Infect Immun. 2002 Sep;70(9):4955–60.

23. Larsen MH, Vilchèze C, Kremer L, Besra GS, Parsons L, Salfinger M, et al. Overexpression of inhA, but not kasA, confers resistance to isoniazid and ethionamide in Mycobacterium smegmatis, M. bovis BCG and M. tuberculosis. Mol Microbiol. 2002 Oct;46(2):453–66.

24. Shea J, Halse TA, Kohlerschmidt D, Lapierre P, Modestil HA, Kearns CH, et al. Low-Level Rifampin Resistance and Mutations in Mycobacterium tuberculosis: an Analysis of Whole-Genome Sequencing and Drug Susceptibility Test Data in New York. J Clin Microbiol [Internet]. 2021 Mar 19;59(4). Available from: http://dx.doi.org/10.1128/JCM.01885-20

25. Lempens P, Meehan CJ, Vandelannoote K, Fissette K, de Rijk P, Van Deun A, et al. Isoniazid resistance levels of Mycobacterium tuberculosis can largely be predicted by high-confidence resistance-conferring mutations. Sci Rep. 2018 Feb 19;8(1):3246.

26. Calver AD, Falmer AA, Murray M, Strauss OJ, Streicher EM, Hanekom M, et al. Emergence of increased resistance and extensively drug-resistant tuberculosis despite treatment adherence, South Africa. Emerg Infect Dis. 2010 Feb;16(2):264–71.

27. Ramachandran G, Swaminathan S. Role of pharmacogenomics in the treatment of tuberculosis: a review. Pharmgenomics Pers Med. 2012 Sep 13;5:89–98.

28. Prideaux B, Via LE, Zimmerman MD, Eum S, Sarathy J, O'Brien P, et al. The association between sterilizing activity and drug distribution into tuberculosis lesions.

Nat Med. 2015 Oct;21(10):1223–7.

29. Nelson KN, Talarico S, Poonja S, McDaniel CJ, Cilnis M, Chang AH, et al. Mutation of and Implications for Using Whole-Genome Sequencing for Investigating Recent Tuberculosis Transmission. Front Public Health. 2021;9:790544.

30. Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, et al. Whole-genome sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. Nat Genet. 2011 Dec 18;44(1):106–10.

31. Borrell S, Teo Y, Giardina F, Streicher EM, Klopper M, Feldmann J, et al. Epistasis between antibiotic resistance mutations drives the evolution of extensively drug-resistant tuberculosis. Evol Med Public Health. 2013 Jan;2013(1):65–74.

32. Nimmo C, Millard J, Faulkner V, Monteserin J, Pugh H, Johnson EO. Evolution of drug resistance in the genomic era. Front Cell Infect Microbiol. 2022 Oct 7;12:954074.

33. Scarparo C, Ricordi P, Ruggiero G, Piccoli P. Evaluation of the fully automated BACTEC MGIT 960 system for testing susceptibility of Mycobacterium tuberculosis to pyrazinamide, streptomycin, isoniazid, rifampin, and ethambutol and comparison with the radiometric BACTEC 460TB method. J Clin Microbiol. 2004 Mar;42(3):1109–14.

34. Dicks KV, Stout JE. Molecular Diagnostics for Mycobacterium tuberculosis Infection. Annu Rev Med. 2019 Jan 27;70:77–90.

35. Osei Sekyere J, Maphalala N, Malinga LA, Mbelle NM, Maningi NE. A Comparative Evaluation of the New Genexpert MTB/RIF Ultra and other Rapid Diagnostic Assays for Detecting Tuberculosis in Pulmonary and Extra Pulmonary Specimens. Sci Rep. 2019 Nov 12;9(1):16587.

36. Cancino-Muñoz I, Moreno-Molina M, Furió V, Goig GA, Torres-Puente M, Chiner-Oms Á, et al. Cryptic Resistance Mutations Associated With Misdiagnoses of Multidrug-Resistant Tuberculosis. J Infect Dis. 2019 Jun 19;220(2):316–20.

37. Gardy JL. Towards genomic prediction of drug resistance in tuberculosis. Lancet Infect Dis. 2015 Oct;15(10):1124–5.

38. Marin M, Vargas R, Harris M, Jeffrey B, Epperson LE, Durbin D, et al. Benchmarking the empirical accuracy of short-read sequencing across the M. tuberculosis genome. Bioinformatics. 2022 Mar 28;38(7):1781–7.

39. Mariner-Llicer C, Goig GA, Zaragoza-Infante L, Torres-Puente M, Villamayor L, Navarro D, et al. Accuracy of an amplicon-sequencing nanopore approach to identify variants in tuberculosis drug-resistance-associated genes. Microb Genom [Internet]. 2021 Dec;7(12). Available from: http://dx.doi.org/10.1099/mgen.0.000740

40. Feuerriegel S, Kohl TA, Utpatel C, Andres S, Maurer FP, Heyckendorf J, et al. Rapid genomic first- and second-line drug resistance prediction from clinical specimens using Deeplex-MycTB. Eur Respir J [Internet]. 2021 Jan;57(1). Available from: http://dx.doi.org/10.1183/13993003.01796-2020

41. Ye M, Yuan W, Molaeipour L, Azizian K, Ahmadi A, Kouhsari E. Antibiotic heteroresistance in Mycobacterium tuberculosis isolates: a systematic review and meta-analysis. Ann Clin Microbiol Antimicrob. 2021 Oct 13;20(1):73.

42. Chen Y, Ji L, Liu Q, Li J, Hong C, Jiang Q, et al. Lesion Heterogeneity and Long-Term Heteroresistance in Multidrug-Resistant Tuberculosis. J Infect Dis. 2021 Sep 1;224(5):889–93.

43. Gygli SM, Borrell S, Trauner A, Gagneux S. Antimicrobial resistance in Mycobacterium tuberculosis: mechanistic and evolutionary perspectives. FEMS Microbiol Rev. 2017 May 1;41(3):354–73.

44. Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, Huang B, et al. Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis. Nat Commun. 2015 Dec 21;6:10063.

45. Walker TM, Kohl TA, Omar SV, Hedge J, Del Ojo Elias C, Bradley P, et al. Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study. Lancet Infect Dis. 2015 Oct;15(10):1193–202.

46. Feuerriegel S, Schleusener V, Beckert P, Kohl TA, Miotto P, Cirillo DM, et al. PhyResSE: a Web Tool Delineating Mycobacterium tuberculosis Antibiotic Resistance and Lineage from Whole-Genome Sequencing Data. J Clin Microbiol. 2015 Jun;53(6):1908–14.

47. Verboven L, Phelan J, Heupink TH, Van Rie A. TBProfiler for automated calling of the association with

drug resistance of variants in Mycobacterium tuberculosis. PLoS One. 2022 Dec 30;17(12):e0279644.

48. Gröschel MI, Owens M, Freschi L, Vargas R Jr, Marin MG, Phelan J, et al. GenTB: A user-friendly genome-based predictor for tuberculosis resistance powered by machine learning. Genome Med. 2021 Aug 30;13(1):138.

49. Walker TM, Miotto P, Köser CU, Fowler PW, Knaggs J, Iqbal Z, et al. The 2021 WHO catalogue of complex mutations associated with drug resistance: A genotypic analysis. Lancet Microbe. 2022 Apr;3(4):e265–73.

50. Cáceres G, Calderon R, Ugarte-Gil C. Tuberculosis and comorbidities: treatment challenges in patients with comorbid diabetes mellitus and depression. Ther Adv Infect Dis. 2022 May 20;9:20499361221095831.

51. Bruchfeld J, Correia-Neves M, Källenius G. Tuberculosis and HIV Coinfection. Cold Spring Harb Perspect Med. 2015 Feb 26;5(7):a017871.

52. Sultana ZZ, Hoque FU, Beyene J, Akhlak-Ul-Islam M, Khan MHR, Ahmed S, et al. HIV infection and multidrug resistant tuberculosis: a systematic review and meta-analysis. BMC Infect Dis. 2021 Jan 11;21(1):51.

53. World Health Organization. Global Tuberculosis Report 2019. 2019. 294 p.

54. Bloemberg GV, Keller PM, Stucki D, Trauner A, Borrell S, Latshang T, et al. Acquired Resistance to Bedaquiline and Delamanid in Therapy for Tuberculosis. N Engl J Med. 2015 Nov 12;373(20):1986–8.

55. de Vos M, Ley SD, Wiggins KB, Derendinger B,

Dippenaar A, Grobbelaar M, et al. Bedaquiline Microheteroresistance after Cessation of Tuberculosis Treatment. N Engl J Med. 2019 May 30;380(22):2178–80.

56. Miotto P, Tessema B, Tagliani E, Chindelevitch L, Starks AM, Emerson C, et al. A standardised method for interpreting the association between mutations and phenotypic drug resistance in. Eur Respir J [Internet]. 2017 Dec;50(6). Available from: http://dx.doi.org/10.1183/13993003.01354-2017

57. Moreno-Molina M, Comas I, Furió V. The Future of TB Resistance Diagnosis: The Essentials on Whole Genome Sequencing and Rapid Testing Methods. Arch Bronconeumol. 2019 Aug;55(8):421–6.

58. CRyPTIC Consortium and the 100,000 Genomes Project, Allix-Béguec C, Arandjelovic I, Bi L, Beckert P, Bonnet M, et al. Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. N Engl J Med. 2018 Oct 11;379(15):1403–15.

59. Flentie K, Harrison GA, Tükenmez H, Livny J, Good JAD, Sarkar S, et al. Chemical disarming of isoniazid resistance in. Proc Natl Acad Sci U S A. 2019 May 21;116(21):10510–7.

60. Long JE, DeJesus M, Ward D, Baker RE, Ioerger T, Sassetti CM. Identifying essential genes in Mycobacterium tuberculosis by global phenotypic profiling. Methods Mol Biol. 2015;1279:79–95.

61. van Opijnen T, Camilli A. Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. Nat Rev Microbiol. 2013 Jul;11(7):435–42.

62. Oppong YEA, Phelan J, Perdigão J, Machado D, Miranda A, Portugal I, et al. Genome-wide analysis of Mycobacterium tuberculosis polymorphisms reveals lineage-specific associations with drug resistance. BMC Genomics. 2019 Mar 29;20(1):252.

63. Hunt M, Bradley P, Lapierre SG, Heys S, Thomsit M, Hall MB, et al. Antibiotic resistance prediction for from genome sequence data with Mykrobe. Wellcome Open Res. 2019 Dec 2;4:191.

64. Xu W, DeJesus MA, Rücker N, Engelhart CA, Wright MG, Healy C, et al. Chemical Genetic Interaction Profiling Reveals Determinants of Intrinsic Antibiotic Resistance in Mycobacterium tuberculosis. Antimicrob Agents Chemother [Internet]. 2017 Dec;61(12). Available from: http://dx.doi.org/10.1128/AAC.01334-17

65. Manson AL, Cohen KA, Abeel T, Desjardins CA, Armstrong DT, Barry CE 3rd, et al. Genomic analysis of globally diverse Mycobacterium tuberculosis strains provides insights into the emergence and spread of multidrug resistance. Nat Genet. 2017 Mar;49(3):395–402.

66. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature. 1998 Jun 11;393(6685):537–44.

67. DeJesus MA, Gerrick ER, Xu W, Park SW, Long JE, Boutte CC, et al. Comprehensive Essentiality Analysis of the Mycobacterium tuberculosis Genome via Saturating Transposon Mutagenesis. MBio [Internet]. 2017 Jan 17;8(1). Available from: http://dx.doi.org/10.1128/mBio.02133-16

68. Coll F, Phelan J, Hill-Cawthorne GA, Nair MB, Mallard K, Ali S, et al. Genome-wide analysis of multi- and extensively drug-resistant Mycobacterium tuberculosis. Nat Genet. 2018 Feb;50(2):307–16.

69. Karp PD, Billington R, Caspi R, Fulcher CA, Latendresse M, Kothari A, et al. The BioCyc collection of microbial genomes and metabolic pathways. Brief Bioinform. 2019 Jul 19;20(4):1085–93.

70. Hasnain SE, Ehtesham NZ, Grover S. Mycobacterium Tuberculosis: Molecular Infection Biology, Pathogenesis, Diagnostics and New Interventions. Springer Nature; 2019. 514 p.

71. Lew JM, Kapopoulou A, Jones LM, Cole ST. TubercuList--10 years after. Tuberculosis . 2011 Jan;91(1):1–7.

72. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000 Jan 1;28(1):27–30.

73. Jankute M, Cox JAG, Harrison J, Besra GS. Assembly of the Mycobacterial Cell Wall. Annu Rev Microbiol. 2015;69:405–23.

74. Farhat MR, Shapiro BJ, Kieser KJ, Sultana R, Jacobson KR, Victor TC, et al. Genomic analysis identifies targets of convergent positive selection in drug-resistant Mycobacterium tuberculosis. Nat Genet. 2013 Oct;45(10):1183–9.

75. Manson AL, Abeel T, Galagan JE, Sundaramurthi JC, Salazar A, Gehrmann T, et al. Mycobacterium tuberculosis Whole Genome Sequences From Southern India Suggest Novel Resistance Mechanisms and the

Need for Region-Specific Diagnostics. Clin Infect Dis. 2017 Jun 1;64(11):1494–501.

76. Hicks ND, Giffen SR, Culviner PH, Chao MC, Dulberger CL, Liu Q, et al. Mutations in dnaA and a cryptic interaction site increase drug resistance in Mycobacterium tuberculosis. PLoS Pathog. 2020 Nov 30;16(11):e1009063.

77. Ramón-García S, Martín C, De Rossi E, Aínsa JA. Contribution of the Rv2333c efflux pump (the Stp protein) from Mycobacterium tuberculosis to intrinsic antibiotic resistance in Mycobacterium bovis BCG. J Antimicrob Chemother. 2007 Mar;59(3):544–7.

78. Arun KB, Madhavan A, Abraham B, Balaji M, Sivakumar KC, Nisha P, et al. Acetylation of isoniazid - a novel mechanism of isoniazid resistance in*Mycobacterium tuberculosis* [Internet]. bioRxiv. bioRxiv; 2020. Available from: http://biorxiv.org/lookup/doi/10.1101/2020.02.10.941252

79. Bergval IL, Schuitema ARJ, Klatser PR, Anthony RM. Resistant mutants of Mycobacterium tuberculosis selected in vitro do not reflect the in vivo mechanism of isoniazid resistance. J Antimicrob Chemother. 2009 Sep;64(3):515–23.

80. Lieberman TD, Wilson D, Misra R, Xiong LL, Moodley P, Cohen T, et al. Genomic diversity in autopsy samples reveals within-host dissemination of HIV-associated Mycobacterium tuberculosis. Nat Med. 2016 Dec;22(12):1470–4.

81. Tarashi S, Fateh A, Mirsaeidi M, Siadat SD, Vaziri F. Mixed infections in tuberculosis: The missing part in a

puzzle. Tuberculosis . 2017 Dec;107:168–74.

82. Cadena AM, Hopkins FF, Maiello P, Carey AF, Wong EA, Martin CJ, et al. Concurrent infection with Mycobacterium tuberculosis confers robust protection against secondary infection in macaques. PLoS Pathog. 2018 Oct;14(10):e1007305.

83. Behr MA, Edelstein PH, Ramakrishnan L. Revisiting the timetable of tuberculosis. BMJ. 2018 Aug 23;362:k2738.

84. Verver S, Warren RM, Beyers N, Richardson M, van der Spuy GD, Borgdorff MW, et al. Rate of reinfection tuberculosis after successful treatment is higher than rate of new tuberculosis. Am J Respir Crit Care Med. 2005 Jun 15;171(12):1430–5.

85. Merker M, Barbier M, Cox H, Rasigade JP, Feuerriegel S, Kohl TA, et al. Compensatory evolution drives multidrug-resistant tuberculosis in Central Asia. Elife [Internet]. 2018 Oct 30;7. Available from: http://dx.doi.org/10.7554/eLife.38200

86. Plazzotta G, Cohen T, Colijn C. Magnitude and sources of bias in the detection of mixed strain M. tuberculosis infection. J Theor Biol. 2015 Mar 7;368:67–73.

87. Dheda K, Lenders L, Magombedze G, Srivastava S, Raj P, Arning E, et al. Drug-Penetration Gradients Associated with Acquired Drug Resistance in Patients with Tuberculosis. Am J Respir Crit Care Med. 2018 Nov 1;198(9):1208–19.

88. Dheda K, Lenders L, Srivastava S, Magombedze G, Wainwright H, Raj P, et al. Spatial Network Mapping of Pulmonary Multidrug-Resistant Tuberculosis Cavities

Using RNA Sequencing. Am J Respir Crit Care Med. 2019 Aug 1;200(3):370–80.

89. Cadena AM, Fortune SM, Flynn JL. Heterogeneity in tuberculosis. Nat Rev Immunol. 2017 Nov;17(11):691–702.

90. Lin PL, Ford CB, Coleman MT, Myers AJ, Gawande R, Ioerger T, et al. Sterilization of granulomas is common in active and latent tuberculosis despite within-host variability in bacterial killing. Nat Med. 2014 Jan;20(1):75–9.

91. Ley SD, de Vos M, Van Rie A, Warren RM. Deciphering Within-Host Microevolution of through Whole-Genome Sequencing: the Phenotypic Impact and Way Forward. Microbiol Mol Biol Rev [Internet]. 2019 May 15;83(2). Available from: http://dx.doi.org/10.1128/MMBR.00062-18

92. Cohen T, van Helden PD, Wilson D, Colijn C, McLaughlin MM, Abubakar I, et al. Mixed-strain mycobacterium tuberculosis infections and the implications for tuberculosis treatment and control. Clin Microbiol Rev. 2012 Oct;25(4):708–19.

93. Andersen P, Scriba TJ. Moving tuberculosis vaccines from theory to practice. Nat Rev Immunol. 2019 Sep;19(9):550–62.

94. Sable SB, Posey JE, Scriba TJ. Tuberculosis Vaccine Development: Progress in Clinical Evaluation. Clin Microbiol Rev [Internet]. 2019 Dec 18;33(1). Available from: http://dx.doi.org/10.1128/CMR.00100-19

95. Maghradze N, Jugheli L, Borrell S, Tukvadze N, Aspindzelashvili R, Avaliani Z, et al. Classifying

recurrent Mycobacterium tuberculosis cases in Georgia using MIRU-VNTR typing. PLoS One. 2019 Oct 18;14(10):e0223610.

96. Kempker RR, Vashakidze S, Solomonia N, Dzidzikashvili N, Blumberg HM. Surgical treatment of drug-resistant tuberculosis. Lancet Infect Dis. 2012 Feb;12(2):157–66.

97. Vashakidze S, Gogishvili S, Nikolaishvili K, Dzidzikashvili N, Tukvadze N, Blumberg HM, et al. Favorable outcomes for multidrug and extensively drug resistant tuberculosis patients undergoing surgery. Ann Thorac Surg. 2013 Jun;95(6):1892–8.

98. Vashakidze S, Despuig A, Gogishvili S, Nikolaishvili K, Shubladze N, Avaliani Z, et al. Retrospective study of clinical and lesion characteristics of patients undergoing surgical treatment for Pulmonary Tuberculosis in Georgia. Int J Infect Dis. 2017 Mar;56:200–7.

99. Siddiqi S, Ahmed A, Asif S, Behera D, Javaid M, Jani J, et al. Direct drug susceptibility testing of Mycobacterium tuberculosis for rapid detection of multidrug resistance using the Bactec MGIT 960 system: a multicenter study. J Clin Microbiol. 2012 Feb;50(2):435–40.

100. Orikiriza P, Nyehangane D, Atwine D, Kisakye JJ, Kassaza K, Amumpaire JM, et al. Evaluation of the SD Bioline TB Ag MPT64 test for identification of complex from liquid cultures in Southwestern Uganda. S Afr J Lab Clin Med. 2017 Mar 31;6(2):383.

101. van Soolingen D, Hermans PW, de Haas PE, Soll DR, van Embden JD. Occurrence and stability of insertion

sequences in Mycobacterium tuberculosis complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. J Clin Microbiol. 1991 Nov;29(11):2578–86.

102. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018 Sep 1;34(17):i884–90.

103. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014 Mar 3;15(3):R46.

104. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754–60.

105. Comas I, Chakravartti J, Small PM, Galagan J, Niemann S, Kremer K, et al. Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved. Nat Genet. 2010 Jun;42(6):498–503.

106. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078–9.

107. Website.

108. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012 Mar;22(3):568–76.

109. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data

to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43(1110):11.10.1–11.10.33.

110. Wilm A, Aw PPK, Bertrand D, Yeo GHT, Ong SH, Wong CH, et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res. 2012 Dec;40(22):11189–201.

111. Huang W, Li L, Myers JR, Marth GT. ART: a next-generation sequencing read simulator. Bioinformatics. 2012 Feb 15;28(4):593–4.

112. Minh BQ, Schmidt H, Chernomor O, Schrempf D, Woodhams M, von Haeseler A, et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era [Internet]. bioRxiv. bioRxiv; 2019. Available from: http://biorxiv.org/lookup/doi/10.1101/849372

113. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018 Jun 1;35(6):1547–9.

114. Huerta-Cepas J, Serra F, Bork P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Mol Biol Evol. 2016 Jun;33(6):1635–8.

115. Tierney L. The R statistical computing environment. In: Lecture Notes in Statistics. New York, NY: Springer New York; 2012. p. 435–47. (Lecture notes in statistics).

116. Wilkinson L. Ggplot2: Elegant graphics for data

analysis by WICKHAM, H. Biometrics. 2011 Jun;67(2):678–9.

117. Cirillo DM, Miotto P, Tagliani E, ReSeqTB Consortium. Reaching consensus on drug resistance conferring mutations. Int J Mycobacteriol. 2016 Dec;5 Suppl 1:S33.

118. Coll F, McNerney R, Guerra-Assunção JA, Glynn JR, Perdigão J, Viveiros M, et al. A robust SNP barcode for typing Mycobacterium tuberculosis complex strains. Nat Commun. 2014 Sep 1;5:4812.

119. Stucki D, Brites D, Jeljeli L, Coscolla M, Liu Q, Trauner A, et al. Mycobacterium tuberculosis lineage 4 comprises globally distributed and geographically restricted sublineages. Nat Genet. 2016 Dec;48(12):1535–43.

120. Guerra-Assunção JA, Crampin AC, Houben RMGJ, Mzembe T, Mallard K, Coll F, et al. Large-scale whole genome sequencing of M. tuberculosis provides insights into transmission in a high prevalence area. Elife [Internet]. 2015 Mar 3;4. Available from: http://dx.doi.org/10.7554/eLife.05166

121. Liu Q, Wei J, Li Y, Wang M, Su J, Lu Y, et al. clinical isolates carry mutational signatures of host immune environments. Sci Adv. 2020 May;6(22):eaba4901.

122. Kreutzer DA, Essigmann JM. Oxidized, deaminated cytosines are a source of C --> T transitions in vivo. Proc Natl Acad Sci U S A. 1998 Mar 31;95(7):3578–82.

123. Shockey AC, Dabney J, Pepperell CS. Effects of Host, Sample, and Culture on Genomic Diversity of Pathogenic Mycobacteria. Front Genet. 2019 Jun

4;10:477.

124. van Rie A, Victor TC, Richardson M, Johnson R, van der Spuy GD, Murray EJ, et al. Reinfection and mixed infection cause changing Mycobacterium tuberculosis drug-resistance patterns. Am J Respir Crit Care Med. 2005 Sep 1;172(5):636–42.

125. Trauner A, Liu Q, Via LE, Liu X, Ruan X, Liang L, et al. The within-host population dynamics of Mycobacterium tuberculosis vary with treatment efficacy. Genome Biol. 2017 Apr 19;18(1):71.

126. Sorra C, McEwan E, Miller C, Farmer M, Nellepalli P. TB and HIV: An Old Disease Takes On a New Partner. Catholic Relief Services; 2009.

127. Vasiliu A, Abelman R, Kherabi Y, Iswari Saktiawati AM, Kay A. Landscape of TB Infection and Prevention among People Living with HIV. Pathogens [Internet]. 2022 Dec 16;11(12). Available from: http://dx.doi.org/10.3390/pathogens11121552

128. Zürcher K, Cox SR, Ballif M, Enane LA, Marcy O, Yotebieng M, et al. Integrating services for HIV and multidrug-resistant tuberculosis: A global cross-sectional survey among ART clinics in low- and middle-income countries. PLOS Glob Public Health [Internet]. 2022 Mar 1;2(3). Available from: http://dx.doi.org/10.1371/journal.pgph.0000180

129. García-Basteiro AL, López-Varela E, Respeito D, González R, Naniche D, Manhiça I, et al. High tuberculosis burden among people living with HIV in southern Mozambique. Eur Respir J. 2015 Feb;45(2):547–9.

130. Schön T, Werngren J, Machado D, Borroni E, Wijkander M, Lina G, et al. Antimicrobial susceptibility testing of Mycobacterium tuberculosis complex isolates - the EUCAST broth microdilution reference method for MIC determination. Clin Microbiol Infect. 2020 Nov;26(11):1488–92.

131. Colangeli R, Jedrey H, Kim S, Connell R, Ma S, Chippada Venkata UD, et al. Bacterial Factors That Predict Relapse after Tuberculosis Therapy. N Engl J Med. 2018 Aug 30;379(9):823–33.

132. A. Cantrell S. Regulated alteration of mycolic acid structure in the cell wall of Mycobacterium tuberculosis. Mycobact Dis [Internet]. 2013;02(03). Available from: https://www.omicsonline.org/regulated-alteration-of-mycolic-acid-structure-in-the-cell-wall-of-mycobacterium-tuberculosis-2161-1068.1000108.php?aid=6308

133. Forrellad MA, McNeil M, Santangelo M de la P, Blanco FC, García E, Klepp LI, et al. Role of the Mce1 transporter in the lipid homeostasis of Mycobacterium tuberculosis. Tuberculosis . 2014 Mar;94(2):170–7.

134. Shimono N, Morici L, Casali N, Cantrell S, Sidders B, Ehrt S, et al. Hypervirulent mutant of Mycobacterium tuberculosis resulting from disruption of the mce1 operon. Proc Natl Acad Sci U S A. 2003 Dec 23;100(26):15918–23.

135. Vilchèze C, Weisbrod TR, Chen B, Kremer L, Hazbón MH, Wang F, et al. Altered NADH/NAD+ ratio mediates coresistance to isoniazid and ethionamide in mycobacteria. Antimicrob Agents Chemother. 2005 Feb;49(2):708–20.

136. Bhat SA, Iqbal IK, Kumar A. Imaging the NADH:NAD

Homeostasis for Understanding the Metabolic Response of Mycobacterium to Physiologically Relevant Stresses. Front Cell Infect Microbiol. 2016 Nov 8;6:145.

137. Bhat SA, Singh N, Trivedi A, Kansal P, Gupta P, Kumar A. The mechanism of redox sensing in Mycobacterium tuberculosis. Free Radic Biol Med. 2012 Oct 15;53(8):1625–41.

138. Briffotaux J, Liu S, Gicquel B. Genome-Wide Transcriptional Responses of to Antibiotics. Front Microbiol. 2019 Feb 20;10:249.

139. Deshpande D, Srivastava S, Chapagain M, Magombedze G, Martin KR, Cirrincione KN, et al. Ceftazidime-avibactam has potent sterilizing activity against highly drug-resistant tuberculosis. Sci Adv. 2017 Aug;3(8):e1701102.

140. Carey AF, Rock JM, Krieger IV, Chase MR, Fernandez-Suarez M, Gagneux S, et al. TnSeq of Mycobacterium tuberculosis clinical isolates reveals strain-specific antibiotic liabilities. PLoS Pathog. 2018 Mar;14(3):e1006939.

141. Fajardo A, Martínez-Martín N, Mercadillo M, Galán JC, Ghysels B, Matthijs S, et al. The neglected intrinsic resistome of bacterial pathogens. PLoS One. 2008 Feb 20;3(2):e1619.

142. Fujiwara M, Kawasaki M, Hariguchi N, Liu Y, Matsumoto M. Mechanisms of resistance to delamanid, a drug for Mycobacterium tuberculosis. Tuberculosis . 2018 Jan;108:186–94.

143. Gegia M, Winters N, Benedetti A, van Soolingen D,

Menzies D. Treatment of isoniazid-resistant tuberculosis with first-line drugs: a systematic review and meta-analysis. Lancet Infect Dis. 2017 Feb;17(2):223–34.

144. Farhat MR, Freschi L, Calderon R, Ioerger T, Snyder M, Meehan CJ, et al. GWAS for quantitative resistance phenotypes in Mycobacterium tuberculosis reveals resistance genes and regulatory regions. Nat Commun. 2019 May 13;10(1):2128.

145. Ackley SF, Lee RS, Worden L, Zwick E, Porco TC, Behr MA, et al. Multiple exposures, reinfection and risk of progression to active tuberculosis. R Soc Open Sci. 2019 Mar;6(3):180999.

146. Lee RS, Proulx JF, Menzies D, Behr MA. Progression to tuberculosis disease increases with multiple exposures. Eur Respir J. 2016 Dec;48(6):1682–9.

147. Mollenkopf HJ, Kursar M, Kaufmann SHE. Immune response to postprimary tuberculosis in mice: Mycobacterium tuberculosis and Miycobacterium bovis bacille Calmette-Guérin induce equal protection. J Infect Dis. 2004 Aug 1;190(3):588–97.

148. Yates TA, Khan PY, Knight GM, Taylor JG, McHugh TD, Lipman M, et al. The transmission of Mycobacterium tuberculosis in high burden settings. Lancet Infect Dis. 2016 Feb;16(2):227–38.

149. Ordonez AA, Wang H, Magombedze G, Ruiz-Bedoya CA, Srivastava S, Chen A, et al. Dynamic imaging in patients with tuberculosis reveals heterogeneous drug exposures in pulmonary lesions. Nat Med. 2020 Apr;26(4):529–34.

150. Marakalala MJ, Raju RM, Sharma K, Zhang YJ, Eugenin EA, Prideaux B, et al. Inflammatory signaling in human tuberculosis granulomas is spatially organized. Nat Med. 2016 May;22(5):531–8.

151. Roca FJ, Ramakrishnan L. TNF dually mediates resistance and susceptibility to mycobacteria via mitochondrial reactive oxygen species. Cell. 2013 Apr 25;153(3):521–34.

152. Payne JL, Menardo F, Trauner A, Borrell S, Gygli SM, Loiseau C, et al. Transition bias influences the evolution of antibiotic resistance in Mycobacterium tuberculosis. PLoS Biol. 2019 May;17(5):e3000265.

153. Martin CJ, Cadena AM, Leung VW, Lin PL, Maiello P, Hicks N, et al. Digitally Barcoding Reveals Infection Dynamics in the Macaque Model of Tuberculosis. MBio [Internet]. 2017 May 9;8(3). Available from: http://dx.doi.org/10.1128/mBio.00312-17

154. Koch AS, Brites D, Stucki D, Evans JC, Seldon R, Heekes A, et al. The Influence of HIV on the Evolution of Mycobacterium tuberculosis. Mol Biol Evol. 2017 Jul 1;34(7):1654–68.

155. Waller NJE, Cheung CY, Cook G, McNeil MB. The evolution of antibiotic resistance is associated with collateral drug phenotypes in Mycobacterium tuberculosis [Internet]. bioRxiv. 2022. Available from: http://biorxiv.org/lookup/doi/10.1101/2022.10.31.514625