Chapter 5

Clade-Specific Distribution of Antibiotic Resistance Mutations in the Population of *Mycobacterium tuberculosis* - Prospects for Drug Resistance Reversion

Karen van Niekerk, Rian Pierneef, Oleg N. Reva, Ilya S. Korostetskiy, Aleksandr I. Ilin and Gulshara K. Akhmetova

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.75181

Abstract

Tuberculosis, caused by *Mycobacterium tuberculosis* (Mtb), is a leading cause of death in humans worldwide. The emergence of antibiotic-resistant strains of Mtb is a threat to tuberculosis control. A general belief is that drug resistance is acquired by Mtb during antibiotic treatment by accumulation of spontaneous mutations. Also, it is known that the drug resistance mutations (DRM) have an associated fitness cost, reducing the transmissibility and virulence of resistant strains. In this work we show that many canonical DRM are clade specific; i.e. they occur only in specific genetic lineages of Mtb and depend on a specific genetic context necessary for the reduction of the fitness cost and sustainability of the drug resistance phenotype. Dependence of the drug resistance on occurrence of genetic variants of multiple genes and specific activities of the encoded proteins allows combating the drug resistance by impairing the global genetic context. A new drug, FS-1, reverses antibiotic resistance by compromising this genetic context and aggravating the fitness cost of DRM.

Keywords: antibiotic resistance, drug resistance mutation, genomic polymorphism, drug resistance reversion

1. Introduction

Tuberculosis (TB), the infectious disease caused by *Mycobacterium tuberculosis* (Mtb), is the leading cause of death from a single infectious disease in humans worldwide. Roughly

one-third of the world's population has latent TB, providing a source of infection. Efforts to curb TB have resulted in 2% annual decline in the global incidence of TB, except in sub-Saharan Africa [1]. However, the emergence of antibiotic-resistant strains of Mtb is a threat to TB control. If antibiotic-resistant TB is not rapidly and appropriately diagnosed, it may lead to an increase in mortality rates and the spread of resistant strains in the population.

The first line antibiotics rifampicin (RIF) and isoniazid (INH), were developed against Mtb in the 1950s and 1960s, and are still the most effective treatments for TB. An estimated 20% of all Mtb isolates are resistant to at least one of the major antibiotics [2]. Multidrug-resistant tuberculosis (MDR-TB) is defined as TB that does not respond to at least rifampicin (RIF) and isoniazid (INH), while extensively drug-resistant TB (XDR-TB) is defined as TB resistant to INH and RIF in addition to resistance to any of the fluoroquinolones (FLQ) and to at least one of the three second-line injectable drugs: amikacin (AMK), capreomycin (CAP) or kanamycin (KAN). Antibiotic resistance arises when bacteria acquire mutations in drug target genes in an infected patient receiving antibiotics, usually as a result of mismanagement of treatment. Primary resistance arises when resistant strains are transmitted from one patient to another.

Efforts to control drug-resistant TB have relied on two beliefs: that most drug resistance is acquired *de novo* during Mtb treatment regiments, i.e., secondary resistance, and that drug-resistance mutations would have an associated fitness cost reducing the transmissibility and virulence of resistant strains [2]. Therefore, TB control has focused on increasing the effectiveness of the first line treatment and of drug-susceptibility tests only in patients who have received anti-TB medication previously. An added challenge is that diagnosing MDR- and XDR-TB requires drug-susceptibility testing with six different drugs, which can take several weeks to months [3].

Results from improved molecular diagnostic methods have challenged these two beliefs. First, an increase in the prevalence of MDR- and XDR-TB appeared to be driving the spread of TB in some areas. For example, primary transmission of MDR- and XDR-TB is the main driving force of drug-resistant TB spread in sub-Saharan Africa [2]. Second, drug-resistance mutations have variable effects on fitness and transmissibility. Mutations associated with resistance to INH, RIF, and streptomycin (SM) have even been associated with low or no fitness costs [4]. Secondary mutations that compensate for drug resistance mutations appear rapidly after the emergence of drug resistance, in the same gene or in genes involved in linked metabolic pathways, and act to restore virulence and may even increase transmissibility [2].

The WHO recommends the Xpert MTB/RIF assay for the diagnosis of rifampicin resistance, and molecular line probe assays for the detection of resistance to first and second line drugs. Many countries with a high TB burden now implement the Xpert MTB/RIF assay, which can be used as a marker for MDR-TB, as INH resistance generally precedes RIF resistance [5]. Microbiological culture is still the reference standard for diagnosis of TB and of drug-resistance. TB remains very difficult to manage in resource-poor areas. Whole-genome sequencing (WGS) and detection of variants holds great promise for characterizing all of the resistance markers (as opposed to a limited range of mutations) as well as genotyping the strain of Mtb, but relies on a more complete understanding of the relationship between genotype, specific drug resistance mutations, activity states of multiple genes and encoded proteins, and the

drug-resistance phenotype [6]. A new promising drug, FS-1, consisting of a nano-molecular complex of iodine atoms ligated to a dextrin-polypeptide network, was reported to cause antibiotic resistance reversion in MDR-TB by compromising the genetic context of the drug resistance phenotype and by aggravating the fitness cost of the drug resistance mutations [7].

2. Genetic mechanisms of drug resistance in Mtb

The major antibiotics for the treatment of TB have four different mechanisms of action: (i) inhibition of RNA synthesis; (ii) inhibition of protein synthesis; (iii) inhibition of cell wall biosynthesis; and (iv) by interfering with the synthesis of cell membranes [8].

Since the early 1990s, numerous studies have described the genetic mechanisms of drug resistance in Mtb, and there is a quantity of data on the polymorphisms found in isolates resistant to specific antibiotics. Mtb is highly clonal, and as such there is little or no horizontal gene transfer, implying that antibiotic resistance is due to point mutations or deletions. Drug-resistance mutations occur in genes coding for the antibiotic target itself (e.g., *gyrA*, *gyrB*, *rrs*), in genes that code for enzymes needed for activating the antibiotic (e.g., *katG*, *inhA*, *rpoB*, *pncA*, *embB*), or in promoter regions of these genes [2, 9]. To date, there are 1031 mutations in the Mtb genome believed to be associated with resistance to nine major groups of antibiotics, with different combinations of mutations causing MDR-TB [10]. Many of the mutations identified are thought to play roles other than causing resistance directly, e.g., compensatory or adaptive roles, to increase fitness, which is being reduced by the drug-resistance mutations [8].

Researchers have not fully elucidated the mechanisms by which drug resistance emerges and is preserved in Mtb populations [11]. Early mathematical models of MDR-TB suggested that DR mutations would impose fitness costs that would tend to select against the mutation in the population and thus limit the spread of TB [12]. However, current research has shown that DR mutations have a variable effect on fitness and transmissibility. INH, RIF and SM resistance have even been associated with low or no fitness costs [2, 4].

Table 1 summarizes the literature data [7, 13–15] on the roles of the major antibiotics used to treat TB and known genes involved in drug-resistance, as well as the mechanisms thought to be responsible for drug-resistance. Drug resistant phenotype in Mtb is associated exclusively with mutations at specific positions in bacterial genomes. No events of a horizontal acquisition of drug resistance genetic determinants were reported for Mtb. Mutations in protein coding genes either alter drug target molecules or reduce activity of enzymes converting pro-drug molecules into active antibiotics, e.g., *katG* gene, which encodes a catalase converting isoniazid to an active isonicotinoyl-NAD adduct [16]. Mutations affecting activities of bacterial enzymes usually reduce viability of bacteria. This phenomenon is known as the fitness cost of drug resistance. Overcoming of the fitness cost requires from bacteria an acquisition of secondary mutations to compensate the side effects of DR mutations. We hypothesized in this paper that the necessity for bacteria to compensate the side effects of DR mutations potentially opens new ways to identify molecular targets for new drugs to induce the reversion of antibiotic resistance in bacterial populations.

	Antibiotic name	Mechanism of action	Some polymorphisms in Mtb causing resistance	Mechanism of drug resistance
First line drugs	Rifampicin, RIF	Inhibits bacterial RNA polymerase by binding it. When RIF binds to the RpoB target, hydroxyl radicals are formed and this has a cytotoxic effect.	Most mutations occur in cluster I of *rpoB*(β subunit of RNA pol), in the 81 bp rifampicin resistance determining region (RRDR)	Drug target is altered. In resistant bacteria, hydroxyl radicals are not formed when RIF binds to RpoB, so cells do not die.
	Ethambutol, EMB	Affects several cellular pathways, mostly arabinogalactan biosynthesis through inhibition of cell wall arabinan polymerization; RNA metabolism, transfer of mycolic acid into cell wall, phospholipid synthesis, spermidine synthesis	Point mutations in the *embCAB* operon or the *emb* genes, affecting expression of the *embA*, *embB*, and *embC* genes	Alteration of the drug target
	Isoniazid, INH	INH is a pro-drug, activated by the catalase-peroxidase enzyme KatG and then binds to InhA. Disrupts multiple pathways, mainly interferes with synthesis of mycolic acid.	Mutations to *katG* gene (50–80%): Mostly S315 T. Mutations to *inhA*, or the promoter region Mutations in *ndh* gene (NADH dehydogenase), *kasA* and *ahpC* genes Mutations in *kasA* gene	*katG*: mutations decrease catalase and peroxidase activity, so reduce activation of INH *inhA* promoter: overexpression of the enzyme
	Pyrazinamide, PZA	Activated by enzyme pyrazinamidase (PZase). Mechanism poorly understood. Disruption of the proton motive force required for essential membrane transport functions by POA at acidic pH.	Mutations in the *pncA* gene encoding PZase, most are in 561-bp region of the open reading frame or in an 82-bp region of its promoter.	Pro-drug cannot be converted to its active form
	Aminoglycosides: streptomycin, SM	Binds to the small 16S rRNA of the 30S subunit of bacterial ribosome, interfering with the binding of tRNA to the 30S subunit	Mutation of the ribosome target binding sites: 50% in the *rpsL* gene, which encodes the ribosomal protein S12, usually K43R 20% mutations to the *rrs* gene. Also mutations in *gidB*, which encodes 16S rRNA methyltransferase	Alteration of the drug target

	Antibiotic name	Mechanism of action	Some polymorphisms in Mtb causing resistance	Mechanism of drug resistance
Second line drugs	Aminoglycosides: kanamycin KAN, amikacin AMK	Binds to the small 16S rRNA of the 30S subunit of bacterial ribosome, interfering with the binding of tRNA to the 30S subunit	Mutation of the ribosome target binding sites genes *rrs*, but not cross-resistant with streptomycin	Alteration of drug target
	Capreomycin, CAP	Inhibits protein synthesis through modification of ribosomal structures at the 16S rRNA	Mutations in the *rrs* gene encoding 16S rRNA mutations in the gene *tlyA* encoding a 2′-O-methyltransferase of 16S rRNA and 23S rRNA	Alteration of drug target
	Ethionamide, ETH	ETH requires activation by monooxygenase EthA, inhibits mycolic acid synthesis by binding the ACP reductase InhA	70% due to mutations in *ethA* or *inhA*	Similar to INH: *inhA* promoter: overexpression of the enzyme
	Fluoroquinolones (FLQ), e.g., ofloxacin (OFX), moxifloxacin (MOX)	Trapping gyrase on DNA as ternary complexes, thereby blocking the movement of replication forks and transcription complexes	Usually multiple mutations in conserved quinolone resistance-determining region (QRDR) of *gyrA* and *gyrB*, most often at positions Ala-90 and Asp-94 in *gyrA* [Mutations at position 80 of *gyrA* cause hypersusceptibility to fluoroquinolones]	FLQ traps the DNA-gyrase complex in which the DNA is broken. Resistant GyrA prevents chromosome breakage.
	Para-aminosalicylic acid, PAS	PAS is a prod-drug and thymidylate synthase A is required for conversion to active form PAS inhibits folic acid biosynthesis and uptake of iron	Mutations in the *thyA* gene encoding the enzyme thymidylate synthesis of the folate biosynthesis pathway, mostly Thr202Ala Also: mutations in *folC*, *ribD*, *dfrA*	Pro-drug cannot be converted to active drug
	Cycloserine, CS	Interrupts peptidoglycan synthesis (for cell wall) by inhibiting the enzymes d-alanine racemase (AlrA) and d-alanine:d-alanine ligase (Ddl)	To be determined *alr*, *ddl*, *cycA*	Unknown

Table 1. The main anti-TB drugs, mechanisms of actions and resistance-conferring polymorphisms.

3. Drug-resistance against the background of Mtb genetic clades and current diagnostic approaches

The disease TB first appeared roughly 70,000 years ago [17]. Studies show that Mtb arose as an obligate human pathogen and that different strains co-evolved with humans, migrated out of Africa, and that the populations expanded with their human hosts [18]. The migrations of modern humans out of Africa and the increased population density during the Neolithic period could be at the origin of its expansion. This theory is consistent with the bacterium's phylogeny and phylogeography [19].

Genetic analyses of global strains have revealed that distinct lineages of Mtb have emerged in different regions of the world. The considerable genetic diversity between these lineages is linked to ancient human migrations out of Africa and to more recent movements and population growth [20]. Hershberg *et al*. demonstrate that there is a greatly reduced selection pressure in Mtb, owing to factors including high clonality of Mtb and serial transmission bottlenecks, both of which reduce the effective population size, increasing the effects of genetic drift [20]. Mutations can reach high functional diversity without being eliminated, which has implications for the emergence of MDR-TB.

During diagnostic procedures, it is helpful to find the lineage of the infecting Mtb strain(s), because some lineages might have acquired specific virulence and/or resistance features before expanding [21]. Clades differ by growth rate and in patterns of host-pathogen interaction in terms of cytokine induction and rate of uptake by macrophages [22]. Lineage 2 (Beijing clade) also is associated with hyper-virulence and with an extended drug resistance pattern [23].

Here we discuss research papers investigating the feasibility of replacing phenotypic drug testing of Mtb with molecular diagnostic techniques. All of them rely on understanding the genetic mechanisms underlying the development and persistence of drug-resistance in Mtb strains, including the context of lineages with varying evolutionary histories.

Köser *et al*. were among the first to publish a method for rapid WGS analysis of an Mtb clinical specimen to reduce the time of XDR-TB diagnosis. They used SNPs to identify lineages, combined with a catalog of well-described DR polymorphisms, demonstrating that WGS is superior to current genotypic tests, but not yet as reliable as phenotypic testing [24].

Rodwell *et al*. of the Global Consortium for Drug-Resistant TB Diagnostics (GCDD) investigated whether a certain collection of mutations can be used as markers of drug resistance in a molecular diagnostic test. They studied a collection of MDR and XDR-TB strains from different regions. Their approach was to select eight genes (*katG*, *inhA*, *rpoB*, *gyrA*, *gyrB*, *rrs*, *eis*, and *thyA*) in which mutations are known to be strongly associated with resistance to the antibiotics INH, RIF, FLQ, AMK, KAN, and CAP. In each specimen, the eight genes were amplified and sequenced, and variants were detected against the H37Rv reference strain. The specificity and sensitivity of the identified variants for drug resistance were determined. They concluded that about 30 mutations in six genes predicted XDR-TB phenotypes with 90–98% sensitivity and almost 100% specificity [3]. However, using these 30 mutations diagnostically would rely on purifying mycobacterial DNA from clinical samples and amplifying the genes of interest before identifying the presence of the mutations. Such a test would rely on broad sequencing coverage and accurate base calling

for the mutations of interest. The study used samples from four geographic regions, but the results do not specify the lineages of the resistant strains. This is problematic, because some DR mutations are lineage specific. More details on the mechanisms of drug resistance summarized from literature sources are in **Table 1**. It should also be noted that the mutations in the identified target genes do not explain all cases of drug resistance.

Genome-wide association studies (GWAS) exploit the rapid turnover and high throughput of NGS, identifying variants in natural populations linked to phenotypic traits by statistical association. Bacterial GWAS have not been frequently used because their population structures reduce the power of association or produce false positives [25]. The clonal nature of bacterial reproduction—especially prevalent in Mtb—means that spurious variations can be strongly associated with particular phenotypes [26]. However, Earle *et al*. have successfully used a linear mixed model approach to perform GWAS on four species of bacteria, including Mtb, to show associations between genetic variation and antibiotic resistant phenotypes. The success of this approach depended on "controlling population structure and boosting power by recovering signals of lineage-level associations" [27]. This method allows the researcher to eliminate signals due only to population, while preserving strain-specific signals that contribute to the DR phenotype.

Coll *et al*. identified a proposed minimum set of SNPs that can be used to differentiate all seven Mtb lineages and 55 sublineages [28]. They identified 88 SNPs in DR candidate regions (two promoters, 21 genes). However, this list of SNPs is aimed at identifying lineages, and is not necessarily informative about drug-resistance in the strains.

Feuerriegel *et al*. showed that many polymorphisms in Mtb previously known to be associated with DR are useful for distinguishing clades, which indicates a lineage specificity of drug resistance [29]. The same team designed the first available web-based drug resistance analysis tool, the Phylo-Resistance Search Engine (PhyResSE) [30]. The tool was evaluated by testing 92 Mtb strains from Sierra Leone with known drug resistance phenotypes, either mono-resistant (RIF, INH or SM) or poly-resistant (RIF, INH, SM, ethambutol (EMB) or pyrazinamide (PZA)). The major advantage of this tool is that it forms a complete analysis pipeline, taking FASTQ files as input: including quality control, mapping and base recalibration prior to genotyping. Thus the end-user need not do complex bioinformatic analysis. This requires considerable computational power. PhyResSE uses a variant catalog based on validated resistance-SNPs from literature as well as from their own experimental data for phylogenetic and drug resistance diagnosis, including lineage-specific resistance mutations. The paper does not give detailed descriptions of the methods used for inclusion or exclusion of specific mutations, or how the sensitivity and specificity were calibrated for mutations, as some mutations may confer only low-level resistance. The program returns a plain-language output which cites the experimental support for the result and also states whether or not there is a high degree of confidence in a particular polymorphism conferring drug resistance. The strains used to evaluate the tool in the paper do not include MDR or XDR strains. Excessive contamination and/or poor sequencing coverage would provide a barrier to correct diagnosis.

One of the few studies using gene pairs associated with drug resistance was by Cui *et al*. [31]. The rationale for the study was that evolution of transmissible drug-resistant Mtb is caused by multiple mutations, many of which interact with each other. This study used nearly 300 Mtb genome sequences from public datasets and their phenotypic drug-sensitivity testing results.

The variants were identified using a standard variant-calling approach. The resulting variants were first filtered using PLINK to remove phylogenetically related variants. The remaining mutations were analyzed using the program GBOOST, which performs a Chi-square test to confirm associations between two variants and phenotype. The resulting gene pairs were screened for the presence of drug target genes and further filtered by non-synonymous mutations. The resulting gene pairs were: one for INH, one for RIF, four for EMB and five for ethionamide (ETH). The authors reported that most of the identified gene pairs containing drug targets consisted of the unique mycobacterial Pro-Pro-Glu (PPE) family proteins, and from this they infer that PPE family proteins play an important role in Mtb drug resistance [30]. The identified mutations were not validated in this study, but the study does show the potential for using pairs of mutations in the diagnosis of drug resistance rather than single mutations. It should be noted that the PPE family proteins make up 10% of the Mtb genome, and they are highly polymorphic, so associations with these genes might occur as a result of genetic drift rather than selection pressure [32]. The value of removing population-specific mutations is unclear, as some lineages of Mtb are strongly associated with drug resistance.

Mortimer *et al*. proposed a method of distinguishing DR loci under positive selection [33]. The rationale behind this is that methods for identifying advantageous mutations usually depend on recombination to differentiate target loci from neutral variants, which is not feasible in the case of Mtb. They analyzed over 1000 Mtb genomes from Russia and South Africa, mostly Lineages 2 (Beijing) and 4, and examined the frequency of different mutations in the populations. They found that resistant sub-populations are less diverse than susceptible subpopulations, which is consistent with the ongoing transmission of resistant Mtb. They classified the DR mutations as either "tight targets" or "sloppy targets" based on their diversity. The authors also noted that lipid metabolism genes are enriched in the list of DR loci under positive selection. This approach has potential for understanding the genetics of resistance in clonal bacteria.

A variety of bioinformatic approaches have been useful for resolving the evolution of the various lineages of *Mtb* over time, for tracing the emergence of pathogenic and more virulent strains, and for identifying variants in Mtb genomes responsible for the development of antibiotic resistance [28, 34–36]. In tandem with methods for rapid identification of drug resistance, researchers are also investigating methods of exploiting our understanding of the evolution of drug resistance. Treatment of TB with antibiotics has had an overall effect of selecting for drug resistance, rather than having the hoped-for effect of selecting DR variants with reduced fitness. Baym *et al*. have reviewed possible mechanisms of selection for drug resistance inversion [37]. These rely on the concept of using combinations of antibiotics and other compounds to inhibit bacterial growth and at the same time reversing the selection for resistance, similar to the combination of penicillin and clavulanic acid to block bacterial β-lactamase, while minimizing or reversing selection for resistance. This avenue shows promise, particularly in combination with quick genotyping of clinical samples.

4. Non-random associations between polymorphic sites in genomes of *M. tuberculosis*

Data for this research was sourced from the GMTV database [17], which consists of SNPs and indels for a large number of Mtb strains for which whole genome sequencing was

performed. Also, this database integrates clinical, epidemiological and microbiological data for all the recorded Mtb isolates. Analysis of this study compared distribution patterns of 58,025 amino acid substitutions in 1089 Mtb strains from the GMTV database. The polymorphisms were determined relative to the H37Rv reference strain [38]. Frequencies of all polymorphisms were calculated for the entire set of 1089 Mtb genomes and for Mtb lineages as they were identified in the GMTV database. Analysis of the data showed that many DR polymorphisms were strongly associated with specific Mtb lineages. A mosaic plot of the data is shown in **Figure 1**. Genomes of the Beijing, Haarlem and Lineage 4.3 clades contained numerous DR mutations, while only a few of them were observed in the Lineage 4.1, Ural and X-type. Bacteria of the latter clades appeared to be mostly drug-susceptible. Statistically significant prevalence of DR mutations in bacteria of specific Mtb clades was confirmed by Fisher's exact test with Bonferroni adjustment. Of these, 25 DR-polymorphism/lineage pairs showed an odds ratio above 1.

Co-occurrence of alleles of different polymorphic sites was identified by calculating the linkage disequilibrium (LD) and χ^2-statistics. In total, 288,840 pairs of polymorphisms showing statistically reliable associations (χ^2 above 6.63 corresponds to a p-value ≤ 0.01) were identified between 823 polymorphic sites including 10 DR mutations [10].Functional associations between DR mutations (denoted as mutations from an initial *A* allele to allele *a* conferring DR) and other genetic polymorphisms (denoted as *B* for the most frequent allele and *b* for all other alternative variants in Mtb population) were identified by Levin's attributable risk statistic [39]. Confidence range values of attributable risks were calculated by Eq. (1).

$$\left[1 - EXP\left(ln\left(1 - R_a\right) - 1.96 \times StdErr\right)\right] \; to \; \left[1 - EXP\left(ln\left(1 - R_a\right) + 1.96 \times StdErr\right)\right] \tag{1}$$

In the case of estimation of the risk of DR mutation from *A* to *a* in a subpopulation of organisms having the allele *b* at the secondary polymorphic site, the parameter R_a was calculated by Eq. (2) and the Fleiss' standard error parameter *StdErr* was calculated by Eq. (3).

$$R_a = \frac{P_{AB}P_{ab} - P_{aB}P_{Ab}}{\left(P_{AB} + P_{aB}\right)\left(P_{aB} + P_{ab}\right)} \tag{2}$$

$$StdErr = \sqrt{\frac{P_{Ab} + R_a\left(P_{AB} + P_{ab}\right)}{N \times P_{aB}}} \tag{3}$$

Risks of secondary mutations *B* to *b* in a DR subpopulation with the genotype *a* were calculated by Eq. (1), but in this case the parameters R_a and *StdErr* were calculated by Eqs. (4) and (5), respectively:

$$R_a = \frac{P_{AB}P_{ab} - P_{aB}P_{Ab}}{\left(P_{AB} + P_{Ab}\right)\left(P_{Ab} + P_{ab}\right)} \tag{4}$$

$$StdErr = \sqrt{\frac{P_{aB} + R_a\left(P_{AB} + P_{ab}\right)}{N \times P_{Ab}}} \tag{5}$$

In Eqs. (2)–(5), values P_{AB}, P_{Ab}, P_{aB} and P_{ab} are the frequencies of corresponding allele combinations; and N is the total number of the analyzed Mtb strains = 1089.

The reasoning behind the further analysis is displayed in **Figure 2**, where two contingency tables of co-distribution of an arginine to leucine replacement at position 463 in the protein KatG rendering INH resistance [40] and two other secondary mutations are shown. Both pairs of mutations are characterized by strong linkage disequilibrium above 0.9. First, the co-distribution of the DR mutation KatG R463L and a polymorphism D69Y in a drug efflux protein Stp (Rv2333c) is considered (**Figure 2-1**). The replacement of the aspartate residue by tyrosine at position 69 of the protein Stp is rather common in the Mtb population and it has not been associated with any DR phenotype. However, this study showed that 91–99% of the DR mutation KatG R463L depends on the presence of the Stp D69Y substitution. In contrast, the likelihood of a D → Y replacement in the protein Stp does not depend significantly on the state of the KatG R463L polymorphism. The estimated attributable risk is in the range of 21–27%. The confidence ranges of attributable risks in **Figure 2** are denoted as $A \to a_{|b}$ and $B \to b_{|a}$, respectively.

Let us consider another co-distribution of the same DR-related polymorphism KatG R463L and a leucine to serine substitution at position 896 in PPE35 protein (Rv1918c), which is shown in **Figure 2-2**. These two mutations are strongly associated with each other, but this

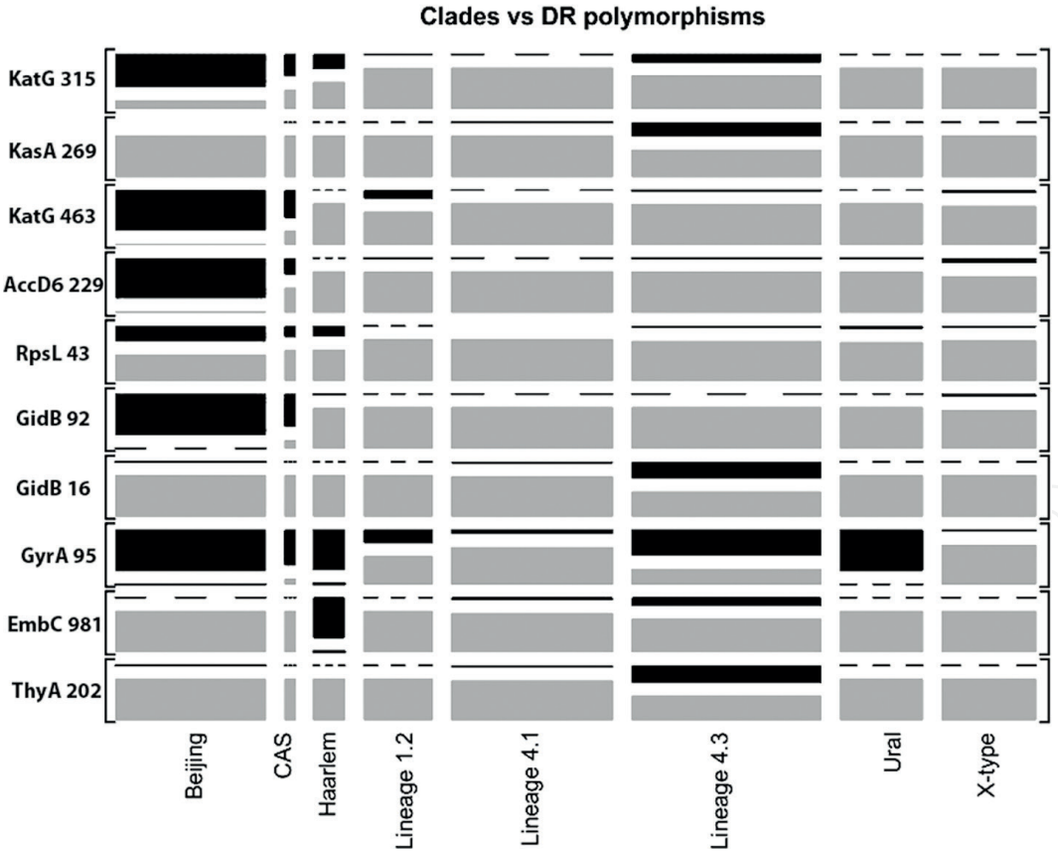

Figure 1. Mosaic plot representing the contingency table for the presence (black) or absence (gray) of each DR polymorphism for each of the 10 loci and the clade for the specimens in the GMTV dataset. The size of the rectangle represents the number of sequences in the category. A dotted line indicates that there were no specimens in that category. Clades with 10 or fewer representatives were omitted.

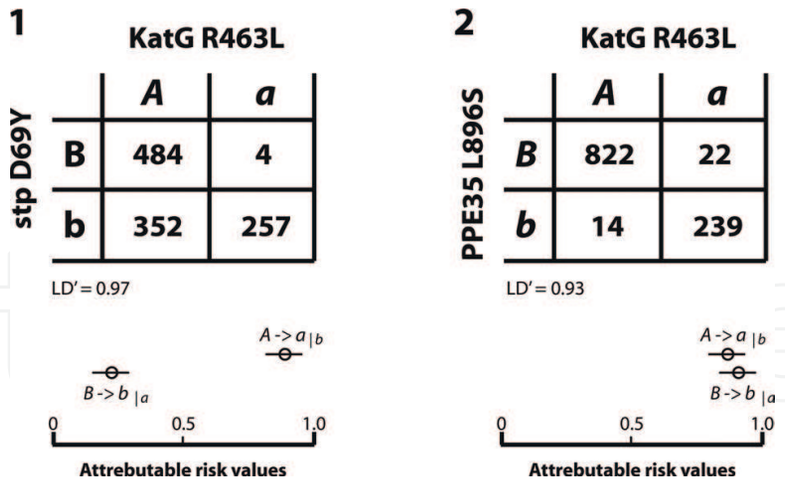

Figure 2. Contingency tables of co-distribution of a DR mutation KatG R463L rendering resistance to INH and two secondary polymorphisms in the (1) efflux drug protein Stp and (2) PPE35 protein. Attributable risks of mutation acquisition were calculated and denoted as $A \rightarrow a_{|b}$ and $B \rightarrow b_{|a}$ for the likelihood of DR mutation acquisition when the secondary site is mutated and the likelihood of secondary mutation in a DR sub-population, respectively.

Secondary mutations	Drug resistance mutations									Annotation
	GyrA S95 T (FLQ)	KatG S315 T,N (INH)	KatG R463L (INH)	AccD6 D229G (INH)	ThyA T202A (PAS)	EmbC V981 L (EMB)	RpsL K43R (SM)	GidB E92D (SM)	GidB L16R (SM)	
Rv0193c K417*,E	86.4 to 94.6	87.5 to 97.7	77.4 to 93.1	71.9 to 90.9	80.7 to 99.6		78.1 to 98.5	79.6 to 99.5	81.1 to 96.0	Hypothetical protein
Rv1186c P207A,T	84.0 to 93.9	82.8 to 96.4	79.8 to 95.7	76.5 to 95.0	64.6 to 94.7		68.7 to 96.6	71.8 to 98.1	74.4 to 93.9	Hypothetical protein
Rv1321 S144R	81.8 to 91.1	84.7 to 96.0	80.6 to 94.5	72.5 to 90.6	67.1 to 92.9		76.7 to 97.4	75.6 to 97.3	76.0 to 92.7	Hypothetical protein
Rv2017 A262E	76.5 to 86.5	83.9 to 95.1	73.5 to 89.4	70.6 to 88.6	72.9 to 95.0		75.7 to 96.4	77.8 to 97.5	76.6 to 92.4	Transcriptional regulator
GalU Q235R	76.6 to 86.6	81.8 to 93.8	74.9 to 90.3	69.4 to 87.8	70.2 to 93.6		72.9 to 95.0	78.0 to 97.6	75.3 to 91.6	UTP-glucose-1-phosphate uridylyltransferase
Rv3204 T34A	74.7 to 84.8	82.5 to 94.1	72.2 to 88.3	69.1 to 87.3	68.5 to 92.4		79.6 to 97.8	75.7 to 96.4	74.8 to 91.0	DNA-methyltransferase
CorA K139*,E	76.2 to 86.1	85.4 to 95.9	70.7 to 87.3	67.3 to 86.1	68.0 to 92.2		73.5 to 95.1	75.4 to 96.3	75.9 to 91.8	Magnesium and cobalt transporter
VapC47 S46 L	74.1 to 84.4	84.5 to 95.3	71.0 to 87.5	71.8 to 89.1	71.2 to 93.8		73.8 to 95.2	75.7 to 96.4	69.3 to 87.4	VapC47 toxin

Secondary mutations	Drug resistance mutations									Annotation
	GyrA S95 T (FLQ)	**KatG S315 T,N (INH)**	**KatG R463L (INH)**	**AccD6 D229G (INH)**	**ThyA T202A (PAS)**	**EmbC V981 L (EMB)**	**RpsL K43R (SM)**	**GidB E92D (SM)**	**GidB L16R (SM)**	
EccC3 P214R	77.6 to 87.4	79.7 to 92.5	71.2 to 87.8	65.1 to 84.8	70.2 to 93.6		72.9 to 95.0	74.8 to 96.2	66.8 to 85.9	Type VII secretion protein
Rv2542 T211A	71.8 to 82.4	81.3 to 93.4	66.4 to 84.1	64.9 to 84.4	71.1 to 93.8		73.8 to 95.2	75.6 to 96.4	70.6 to 88.3	Hypothetical protein
PstA1 M5T	87.2 to 95.3	74.6 to 90.6	82.9 to 96.4	84.1 to 97.6	79.8 to 99.5			78.8 to 99.5	82.3 to 96.8	Phosphate-transport integral membrane ABC transporter
TsnR L232P	88.6 to 96.2	72.4 to 89.0	79.9 to 94.7	76.6 to 93.8	80.1 to 99.5				80.6 to 95.9	23S rRNA methyl-transferase
AroG D265E		81.7 to 92.8	81.5 to 93.4	77.3 to 91.6	73.3 to 93.5		68.5 to 90.8	79.5 to 96.9	83.4 to 95.3	Phenylalanine-repressible DAHP synthetase
ProX L85P		61.4 to 75.4	81.2 to 92.1	80.0 to 92.1		76.2 to 94.2			84.9 to 95.1	Osmoprotectant
UspA V127 L		59.2 to 74.3	82.1 to 93.2	78.2 to 91.3		83.3 to 98.1			82.5 to 94.0	Sugar ABC transporter
Stp D69Y		53.8 to 71.1	90.9 to 98.6	90.7 to 99.0		83.1 to 98.9			88.2 to 97.8	Drug efflux protein
AceAa G179D		53.2 to 70.8	90.8 to 98.6	87.9 to 97.8		82.9 to 98.8			89.4 to 98.4	Isocitratelyase
GalTb T174A		54.0 to 71.1	84.6 to 95.3	87.1 to 97.3		85.3 to 99.7			89.7 to 98.5	Galactose-1-phosphate uridylyl-transferase
Rv0324 T168A		54.8 to 72.3	87.2 to 96.9	90.5 to 98.9		82.6 to 98.8			89.2 to 98.4	Transcriptional regulator
EspK C729S		54.0 to 70.5	83.6 to 94.4	85.5 to 96.2		79.5 to 96.9			83.4 to 94.9	ESX-1 secretion-Associated protein

Polymorphic sites are denoted by names of genes and pairs of amino acid substitutions from the most common allele to one or several alternative allelic states.

Deletions are marked by asterisks (*).Values X_{min} *to* X_{max} in cells represent confidence ranges estimated for p-value ≤0.05 (Eq. (1)).

Table 2. Attributable risk of acquisition of DR mutations in sub-populations of Mtb with secondary mutations.

dependence is highly symmetric: in more than 90% of cases both mutations co-occur in the same genomes. It may indicate a genetic drift event when the DR phenotype is characteristic for a sub-lineage of isolates sharing common ancestry and the neutral mutation in the hyper-mutable PPE35 protein is a genetic marker of the sublineage.

For further analysis, only those secondary polymorphisms which influenced the DR mutations significantly, but were independent, were selected; i.e. cases were selected when confidence

ranges of attributable risks $A \rightarrow a_{|b}$ and $B \rightarrow b_{|a}$do not overlap and $A \rightarrow a_{|b} > B \rightarrow b_{|a}$ as in **Figure 2-1**. In total, 554 secondary polymorphisms were found, which increase likelihood of acquisition of 9 out of 10 studied DR mutations. The mutation KasA G269S, rendering resistance to INH [41], was strongly associated only with the GidB L16R mutation rendering SM resistance [42], which indicates that the former polymorphism is most likely a secondary mutation in multidrug resistant Mtb.

A selection of secondary mutations predetermining acquisition of nine of the most widely distributed DR mutations rendering resistance to FLQ, INH, EMB, SM and para-aminosalisylic acid (PAS) in multidrug resistant Mtb are shown in **Table 2**. Values X_{min} *to* X_{max} in **Table 2** represent confidence ranges estimated for *p*-value ≤ 0.05 (Eq. (1)). It was found that the acquisition of DR mutations require allelic alterations in many other proteins including several transmembrane transporter and efflux proteins, osmoprotectant, transcriptional regulators and some other proteins. Strong cross-associations between DR polymorphisms characteristic for different lineages (**Figure 1**) favors the hypothesis of strong functional associations between these mutations compared to neutral genetic drift. The identified proteins predefining the acquisition of the DR phenotype may be molecular targets for development of new drugs for antibiotic resistance reversion.

5. The concept of the drug resistance reversion and implementation thereof

The concept of drug resistance reversion was applied in recent studies [7, 41]. Drug resistance mutations are often incompatible with one another, as shown by negative linkage disequilibrium values. This suggests that the cumulative fitness cost of mutations is often too high for the resulting strain to be viable. FS-1 is a new drug which seems to exploit this tendency. Active units of FS-1 are aggregated micelles containing complexes of tri-iodide molecules coordinated by metal ions and integrated into a dextrin-polypeptide moiety. The basic formula of the micelle is:

$$\left[\left\{\left(L_n\left(MeJ_3\right)^+\right)_y\left[Me(L_m)J\right]^+_x\right\}\left(Cl^-\right)_{y+x+k}\right]$$

where *L*—dextrin-polypeptide ligand; *Me*—Li/Mg ions; *n*, *m*, *x*, *y* and *k*—variable integers ≥1; molecular mass of the micelles is in the range of 30–300 kD. This molecular complex was designed to prolong the residence time of moderately oxidative iodine molecules in an organism and facilitate their transportation to inner tissues.

Studies of XDR-TB infection in animal models showed the reversion of Mtb pathogens to a more drug sensitive phenotype after treatment with FS-1 despite the remaining DR related mutations in their genomes [7]. Drug resistance reversion was also confirmed on an *in vitro* model with a XDR-TB clinical isolate SCAID 187.0 when cultivated for 60 days in six passages on a medium with a sub-lethal dose of ¼ MIC of FS-1. Reduction of the antibiotic resistance of XDR-TB isolates obtained during the clinical trial of FS-1 was consistent with the results of the above-mentioned laboratory experiments. It was concluded that the DR phenotype requires multiple genes to be in specific activity states controlled either by transcription regulation or resulting from specific mutations. A combination of genetic variants creates a genomic context of drug resistance.

Clinical trials of FS-1 has been undertaken in Kazakhstan and registered in the Clinical Trial database (www.ClinicalTrials.gov) under an accession number NCT02607449. It was found that FS-1 had a high absorption rate after per-oral administration, which was not affected by food intake. Peak plasma concentration of FS-1 was observed within 1–2 h after administration. Gastric juice activated the infusion of FS-1 in stomach. Pharmacokinetic study of FS-1 showed a long residence time of the drug in the blood stream and an elevated accumulation in the liver. The drug was excreted from the test organism mainly by the kidneys.

The preclinical trial of FS-1 included pharmacological studies (primary and secondary pharmacokinetics); general toxicity determination; tests for mutagenesis, inhibition of reproductive performance, immune toxicity, mucous membrane irritation and several other general physicochemical studies of the compound. FS-1 caused no irritation of the stomach mucosa when applied in concentrations of up to 5.0 mg/kg. No ulcerogenic, allergenic, immune toxic, mutagenic or carcinogenic side effects were observed after repeated administration of FS-1. Also, no cytotoxicity or embryonic toxicity was observed. Toxicological studies attributed FS-1 to low toxicity compounds with a reduced accumulation in an organism (drug accumulation coefficient was 1.85). The maximum endurable dose of FS-1 identified in rats was 496 mg/kg, and in mice, 993 mg/kg. The average lethal dose (LD50) in rats was found to be 992 mg/kg for both male and female individuals. Therapeutic doses of FS-1 in clinical trials on humans for the treatment of patients with lung XDR-TB infection ranged from 1.0 to 5.0 mg/kg. During the clinical trials, FS-1 was administrated for up to 6 months in combination with the antibiotics commonly prescribed for XDR-TB treatment. Currently, in the third stage of the clinical trials, FS-1 is administrated at a concentration of 2.5 mg/ml for 6 months. Clinical studies complied with the regulations and recommendations of the Ministry of Health of Kazakhstan and were approved by the respective committees of the Ministry.

The first phase of clinical trials was undertaken in 2009–2010 at the Central Clinical Hospital of the Executive Officer of the President of Kazakhstan. During this phase, the drug tolerance and safety of a unitary and repeated per-oral intake of the drug by healthy volunteers were determined. Hematological parameters, including measuring the concentrations of important microelements, i.e., potassium, sodium, magnesium and calcium; functions of liver and kidney, electro-physiological parameters of myocardium, metabolism of proteins, hydrocarbons and lipids, were monitored. Biochemical parameters of the blood plasma of volunteers remained normal during the study. It was found that the administration of FS-1 activated cellular immunity and synthesis of γ-interferon.

The second phase of clinical trials was conducted in 2010–2012 at the Municipal anti-tuberculosis clinic in Almaty, at the National Centre of Tuberculosis in Almaty and at the Regional anti-tuberculosis clinic of the Karaganda region in Kazakhstan. In total, 220 volunteer patients with active XDR-TB lung tuberculosis were involved in this phase of trials. The volunteers ranged from 18 to 65 years old, with a body mass within 10% of the average body weight of male and female adults, with no contraindications to the common MDR-TB antibiotic therapy. Informed consent principles, which imply voluntariness of participation and understanding of the matter of the trial, were complied with. Contraindications to participation in the trial were: pregnancy; oncological diseases; HIV; 3-fold higher than normal ALT/AST or increased

creatinine in blood; dermatomycosis; mental disorders; hypothyroidism; any allergies, especially an allergy to iodine-containing preparations; and any other cardiovascular, kidney or liver decompensated concomitant diseases.

The therapeutic efficacy of the drug was evaluated by bacteriological examination of sputum samples of patients on Lowenstein-Jensen medium for the presence of Mtb isolates. Other tests performed during the trial were: microscopic examination of sputum smears; controlling the positive dynamics of recovery by regular X-ray examinations and by general clinical tests; positive body weight dynamics; and the efficacy of prevention of disease relapses. The efficacy and safety of the regimen of per-oral administration of FS-1 in concentrations of 2.0–5.0 mg/kg during the 6 months in combination with commonly prescribed antibiotics against XDR-TB were confirmed in the second phase of the trial. No serious side effects of the treatment were recorded. In particular, thyroid gland function was monitored for adverse effects. No statistically reliable alterations in the concentration of thyroid hormones in blood were observed, which indicated no deleterious effect of this iodine-containing drug on thyroid gland functions. The time of complete recovery from XDR-TB was reduced, with no disease relapses during the 12 months surveillance, resulting in a significant reduction of the average cost of XDR-TB treatment (**Table 3**).

Mtb isolates were collected on a regular basis during the second phase of the FS-1 clinical trials. It was found that the percentage of drug resistant isolates decreased continuously during the treatment course with FS-1 despite the administration of the antibiotics. It was hypothesized that the therapeutic activity of FS-1 may be associated with the reversion of antibiotic resistance [37]. This hypothesis was then confirmed in an *in vivo* experiment on guinea pigs, which has been recently published [7].

The third phase of clinical trials began in 2014 and is still in progress. The drug FS-1 has been approved as an antibacterial medicine for per-oral administration in a complex of commonly prescribed anti-tuberculosis drugs for the treatment of XDR-TB in Kazakhstan (approval certificate РК-ЛС-5№021305 from 08-04-2015).

XDR TB treatment expenses	**Conventional antibiotic therapy**	**Combined therapy by antibiotics with FS-1**
Time of 100% sanation from *M. tuberculosis* isolation from sputum	12–24 months	3–6 months
Percentage of relapses in 12 month surveillance period	46.1%	Not observed
Daily therapy cost in clinics of Kazakhstan	$ 11.7	$ 12.5
Full cost of the treatment course including the treatment of disease relapses	$ 4274 or up to $ 8548 in the case of TB relapses	$ 2256 (no TB relapses were recorded)

Table 3. Summarized efficacy of application of FS-1 in the second phase of the clinical trial in terms of reduction of the treatment course duration and cost.

6. Conclusion

The idea that the DR phenotype is determined by multiple genes was supported in a review by Trauner *et al*. and Müller *et al*. [9, 13]. They argued that the genetic background of Mtb is important in determining the phenotypic effect of DR mutations. Epistasis—the genetic interactions that determine a phenotype—and bacterial fitness are the two factors that determine the evolution of drug resistance. It is thus important to study genes that are directly involved in drug metabolism as well as genes that could play a compensatory role, such as those involved in aspects of cell physiology, e.g., permeability of the cell. The complexity of the genetic and epistatic determination of the DR phenotype allows the development of new drugs to induce reversion of drug resistance in populations of pathogens. The phenomenon of DR reversion was defined in a review by Baym *et al*. as an active drug-induced counter-selection of resistant variants from populations of pathogens [37]. Several theoretical assumptions were discussed in this paper to explain the resistance reversion despite the presence of selective antibiotics in a medium. Practical application of the antibiotic resistance reversion approach to combat multidrug resistant tuberculosis was exemplified in this work by an overview of the clinical trial of the new drug FS-1.

Acknowledgements

This research was funded by grant 105996 provided by the National Research Foundation of South Africa and by grant 0115PK00389 of the program "Study on the reversion of antibiotic resistance in pathogenic microorganisms" provided by the Ministry for Investments and Development of Kazakhstan.

Conflict of interest

No conflict of interest was reported by the authors.

Author details

Karen van Niekerk[1], Rian Pierneef[1], Oleg N. Reva[1]*, Ilya S. Korostetskiy[2], Aleksandr I. Ilin[2]* and Gulshara K. Akhmetova[2]

*Address all correspondence to: oleg.reva@up.ac.za and ilin_ai@mail.ru

1 Centre for Bioinformatics and Computational Biology, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, South Africa

2 Scientific Center for Anti-Infectious Drugs, Almaty, Kazakhstan

References

[1] WHO. Global Tuberculosis Report. 2017. DOI: WHO/HTM/TB/2017.23

[2] Dheda K, Gumbo T, Maartens G, Dooley KE, McNerney R, Murray M, Furin J, Nardell EA, London L, Lessem E, Theron G, van Helden P, Niemann S, Merker M, Dowdy D, Van Rie A, Siu GKH, Pasipanodya JG, Rodrigues C, Clark TG, Sirgel FA, Esmail A, Lin H-H, Atre SR, Schaaf HS, Chang KC, Lange C, Nahid P, Udwadia ZF, Horsburgh CR Jr, Churchyard GJ, Menzies D, Hesseling AC, Nuermberger E, McIlleron H, Fennelly KP, Goemaere E, Jaramillo E, Low M, Jara CM, Padayatchi N, Warren RM. The epidemiology, pathogenesis, transmission, diagnosis, and management of multidrug-resistant, extensively drug-resistant, and incurable tuberculosis. The Lancet Respiratory Medicine. 2017;**5**:291-360. DOI: 10.1016/S2213-2600(17)30079-6

[3] Rodwell TC, Valafar F, Douglas J, Qian L, Garfein RS, Chawla A, Torres J, Zadorozhny V, Kim MS, Hoshide M, Catanzaro D, Jackson L, Lin G, Desmond E, Rodrigues C, Eisenach K, Victor TC, Ismail N, Crudu V, Gler MT, Catanzaro A. Predicting extensively drug-resistant *Mycobacterium tuberculosis* phenotypes with genetic mutations. Journal of Clinical Microbiology. 2014;**52**:781-789. DOI: 10.1128/JCM.02701-13

[4] Naidoo CC, Pillay M. Increased *in vitro* fitness of multi- and extensively drug-resistant F15/LAM4/KZN strains of *Mycobacterium tuberculosis*. Clinical Microbiology and Infection. 2014;**20**:O361-O369. DOI: 10.1111/1469-0691.12415

[5] Manson AL, Cohen KA, Abeel T, Desjardins CA, Armstrong DT, Barry CE 3rd, Brand J, TBResist Global Genome Consortium, Chapman SB, Cho SN, Gabrielian A, Gomez J, Jodals AM, Joloba M, Jureen P, Lee JS, Malinga L, Maiga M, Nordenberg D, Noroc E, Romancenco E, Salazar A, Ssengooba W, Velayati AA, Winglee K, Zalutskaya A, Via LE, Cassell GH, Dorman SE, Ellner J, Farnia P, Galagan JE, Rosenthal A, Crudu V, Homorodean D, Hsueh PR, Narayanan S, Pym AS, Skrahina A, Swaminathan S, Van der Walt M, Alland D, Bishai WR, Cohen T, Hoffner S, Birren BW, Earl AM. Genomic analysis of globally diverse *Mycobacterium tuberculosis* strains provides insights into emergence and spread of multidrug resistance. Nature Genetics. 2017;**49**:395-402. DOI: 10.1038/ng.3767

[6] Cabibbe AM, Cirillo DM. Best approaches to drug-resistance surveillance at the country level. International Journal of Mycobacteriology. 2016;**5**:S40-S41. DOI: 10.1016/j.ijmyco.2016.09.010

[7] Ilin AI, Kulmanov ME, Korotetskiy IS, Islamov RA, Akhmetova GK, Lankina MV, Reva ON. Genomic insight into mechanisms of reversion of antibiotic resistance in multidrug resistant *Mycobacterium tuberculosis* induced by a nanomolecular iodine-containing complex FS-1. Frontiers in Cellular and Infection Microbiology. 2017;**7**:151. DOI: 10.3389/fcimb.2017.00151

[8] Zhang Y, Yew W-W. Mechanisms of drug resistance in *Mycobacterium tuberculosis*: Update 2015. The International Journal of Tuberculosis and Lung Disease. 2015;**19**:1276-1289. DOI: 10.5588/ijtld.15.0389

[9] Trauner A, Borrell S, Reither K, Gagneux S. Evolution of drug resistance in tuberculosis: Recent progress and implications for diagnosis and therapy. Drugs. 2014;**74**:1063-1072. DOI: 10.1007/s40265-014-0248-y

[10] Sandgren A, Strong M, Muthukrishnan P, Weiner BK, Church GM, Murray MB. Tuberculosis drug resistance mutation database. PLoS Medicine. 2009;**6**:e2. DOI: 10.1371/journal.pmed.1000002

[11] Fonseca JD, Knight GM, McHugh TD. The complex evolution of antibiotic resistance in *Mycobacterium tuberculosis*. International Journal of Infectious Diseases. 2015;**32**:94-100. DOI: 10.1016/j.ijid.2015.01.014

[12] Cohen T, Murray M. Modeling epidemics of multidrug-resistant *M. tuberculosis* of heterogeneous fitness. Nature Medicine. 2004;**10**:1117-1121. DOI: 10.1038/nm1110

[13] Müller B, Borrell S, Rose G, Gagneux S. The heterogeneous evolution of multidrug-resistant *Mycobacterium tuberculosis*. Trends in Genetics. 2013;**29**:160-169. DOI: 10.1016/j.tig.2012.11.005

[14] Kolyva A, Old KP. New TB drugs: Mechanisms of action and resistance. In: Understanding Tuberculosis—New Approaches to Fighting against Drug Resistance. Rijeka, Croatia: InTech; 2012. pp. 210-232. DOI: 10.5772/30992

[15] Aldred KJ, Blower TR, Kerns RJ, Berger JM, Osheroff N. Fluoroquinolone interactions with *Mycobacterium tuberculosis* gyrase: Enhancing drug activity against wild-type and resistant gyrase. Proceedings of the National Academy of Sciences. 2016;**113**:E839-E846. DOI: 10.1073/pnas.1525055113

[16] Zhao X, Yu H, Yu S, Wang F, Sacchettini JC, Magliozzo RS. Hydrogen peroxide-mediated isoniazid activation catalyzed by *Mycobacterium tuberculosis* catalase-peroxidase (KatG) and its S315T mutant. Biochemistry. 2006;**45**:4131-4140. DOI: 10.1021/bi051967o

[17] Chernyaeva EN, Shulgina MV, Rotkevich MS, Dobrynin PV, Simonov SA, Shitikov EA, Ischenko DS, Karpova IY, Kostryukova ES, Ilina EN, Govorun VM, Zhuravlev VY, Manicheva OA, Yablonsky PK, Isaeva YD, Nosova EY, Mokrousov IV, Vyazovaya AA, Narvskaya OV, Lapidus AL, O'Brien SJ. Genome-wide *Mycobacterium tuberculosis* variation (GMTV) database: A new tool for integrating sequence variations and epidemiology. BMC Genomics. 2014;**15**:308. DOI: 10.1186/1471-2164-15-308

[18] Gagneux S. Host-pathogen coevolution in human tuberculosis. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences. 2012;**367**:850-859. DOI: 10.1098/rstb.2011.0316

[19] Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, Parkhill J, Malla B, Berg S, Thwaites G, Yeboah-Manu D, Bothamley G, Mei J, Wei L, Bentley S, Harris SR, Niemann S, Diel R, Aseffa A, Gao Q, Young D, Gagneux S. Out-of-Africa migration and Neolithic co-expansion of *Mycobacterium tuberculosis* with modern humans. Nature Genetics. 2014;**45**:1176-1182. DOI: 10.1038/ng.2744

[20] Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, Roach JC, Kremer K, Petrov DA, Feldman MW, Gagneux S. High functional diversity in *Mycobacterium*

tuberculosis driven by genetic drift and human demography. PLoS Biology. 2008;**6**:e311. DOI: 10.1371/journal.pbio.0060311

[21] Banuls A-L, Sanou A, Van Anh NT, Godreuil S. *Mycobacterium tuberculosis*: Ecology and evolution of a human bacterium. Journal of Medical Microbiology. 2015;**64**:1261-1269. DOI: 10.1099/jmm.0.000171

[22] Reiling N, Homolka S, Walter K, Brandenburg J, Niwinski L, Ernst M, Herzmann C, Lange C, Diel R, Ehlers S, Niemann S. Clade-specific virulence patterns of *Mycobacterium tuberculosis* complex strains in human primary macrophages and aerogenically infected mice. MBio. 2013;**4**:pii: e00250-pii: e00213. DOI: 10.1128/mBio.00250-13

[23] Hanekom M, Gey van Pittius NC, McEvoy C, Victor TC, Van Helden PD, Warren RM. *Mycobacterium tuberculosis* Beijing genotype: A template for success. Tuberculosis. 2011;**91**:510-523. DOI: 10.1016/j.tube.2011.07.005

[24] Köser CU, Bryant JM, Becq J, Török EM, Ellington MJ, Marti-Renom M, Carmichael AJ, Parkhill J, Smith GP, Peacock J. Whole-genome sequencing for rapid susceptibility testing of *M. tuberculosis*. The New England Journal of Medicine. 2013;**369**:290-292. DOI: 10.1056/NEJMc1215305

[25] Falush D, Bowden R. Genome-wide association mapping in bacteria? Trends in Microbiology. 2006;**14**:353-355. DOI: 10.1016/j.tim.2006.06.003

[26] Falush D. Bacterial genomics: Microbial GWAS coming of age. Nature Microbiology. 2016;**1**:16059. DOI: 10.1038/nmicrobiol.2016.59

[27] Earle SG, Wu CH, Charlesworth J, Stoesser N, Gordon NC, Walker TM, Spencer CCA, Iqbal Z, Clifton DA, Hopkins KL, Woodford N, Smith EG, Ismail N, Llewelyn MJ, Peto TE, Crook DW, McVean G, Walker AS, Wilson DJ. Identifying lineage effects when controlling for population structure improves power in bacterial association studies. Nature Microbiology. 2016;**1**:16041. DOI: 10.1038/nmicrobiol.2016.41

[28] Coll F, McNerney R, Guerra-Assunção JA, Glynn JR, Perdigão J, Viveiros M, Portugal I, Pain A, Martin N, Clark TG. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. Nature Communications. 2014;**5**:4812. DOI: 10.1038/ncomms5812

[29] Feuerriegel S, Koser CU, Niemann S. Phylogenetic polymorphisms in antibiotic resistance genes of the *Mycobacterium tuberculosis* complex. The Journal of Antimicrobial Chemotherapy. 2014;**69**:1205-1210. DOI: 10.1093/jac/dkt535

[30] Feuerriegel S, Schleusener V, Beckert P, Kohl TA, Miotto P, Cirillo DM, Cabibbe AM, Niemann S, Fellenberg K. PhyResSE: A web tool delineating *Mycobacterium tuberculosis* antibiotic resistance and lineage from whole-genome sequencing data. Journal of Clinical Microbiology. 2015;**53**:1908-1914. DOI: 10.1128/JCM.00025-15

[31] Cui ZJ, Yang QY, Zhang HY, Zhu Q, Zhang QY. Bioinformatics identification of drug resistance-associated gene pairs in *Mycobacterium tuberculosis*. International Journal of Molecular Sciences. 2016;**17**:1417-1425. DOI: 10.3390/ijms17091417

[32] Mukhopadhyay S, Balaji KN. The PE and PPE proteins of *Mycobacterium tuberculosis*. Tuberculosis. 2017;**91**:441-447. DOI: 10.1016/j.tube.2011.04.004

[33] Mortimer TD, Weber AM, Pepperell CS. Signatures of selection at drug resistance loci in *Mycobacterium tuberculosis*. mSystems 2018;**3**:e00108-17. DOI: 10.1128/mSystems.00108-17

[34] Coll F, Prestona M, Guerra-Assunção JA, Hill-Cawthorn G, Harris D, Perdigão J, Viveiros M, Portugal I, Drobniewski F, Gagneux S, Glynn JR, Pain A, Parkhill J, McNerney R, Martin N, Clark TG. PolyTB: A genomic variation map for *Mycobacterium tuberculosis*. Tuberculosis. 2014;**94**:346-354. DOI: 10.1016/j.tube.2014.02.005

[35] Faksri K, Tan JH, Chaiprasert A, Teo Y-Y, RT-H O. Bioinformatics tools and databases for whole genome sequence analysis of *Mycobacterium tuberculosis*. Infection, Genetics and Evolution. 2016;**45**:359-368. DOI: 10.1016/j.meegid.2016.09.013

[36] Aminian M, Couvin D, Shabbeer A, Hadley K, Vandenberg S, Rastogi N, Bennett KP. Predicting *Mycobacterium tuberculosis* complex clades using knowledge-based Bayesian networks. BioMed Research International. 2014;**2014**:1-11. DOI: 10.1155/2014/398484

[37] Baym M, Stone LK, Kishony R. Multidrug evolutionary strategies to reverse antibiotic resistance. Science. 2016;**351**:aad3292

[38] Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature. 1998;**393**:537-544. DOI: 10.1038/31159

[39] Fleiss JL, Levin B, Paik MC. Statistical methods for rates and proportions. In: Shewart WA, Wilks SS, editors. Wiley Series in Probability and Statistics. 3rd ed. Wiley Online Library. Hoboken, New Jersey: John Wiley and Sons, Inc; 2004

[40] Heym B, Alzari PM, Honoré N, Cole ST. Missense mutations in the catalase-peroxidase gene, *katG*, are associated with isoniazid resistance in *Mycobacterium tuberculosis*. Molecular Microbiology. 1995;**15**:235-245. DOI: 10.1038/31159

[41] Mdluli K, Slayden RA, Zhu Y, Ramaswamy S, Pan X, Mead D, Crane DD, Musser JM, Barry CE 3rd. Inhibition of a *Mycobacterium tuberculosis* beta-ketoacyl ACP synthase by isoniazid. Science. 1998;**280**:1607-1610. DOI: 10.1126/science.280.5369.1607

[42] Okamoto S, Tamaru A, Nakajima C, Nishimura K, Tanaka Y, Tokuyama S, Suzuki Y, Ochi K. Loss of a conserved 7-methylguanosine modification in 16S rRNA confers low-level streptomycin resistance in bacteria. Molecular Microbiology. 2007;**63**:1096-1106. DOI: 10.1111/j.1365-2958.2006.05585.x