



OPEN ACCESS

EDITED BY

Gavin Huttley,
Australian National University, Australia

REVIEWED BY

Colwyn A Headley,
Stanford University, United States
Leonidas Salichos,
New York Institute of Technology,
United States

*CORRESPONDENCE

Danila Zimenkov,
✉ z@biochip.ru

RECEIVED 14 October 2024

ACCEPTED 04 June 2025

PUBLISHED 02 July 2025

CITATION

Zimenkov D and Ushtanit A (2025)
Dinucleotide codon substitutions as a
signature of diversifying selection in
Mycobacterium tuberculosis.
Front. Mol. Biosci. 12:1511372.
doi: 10.3389/fmolb.2025.1511372

COPYRIGHT

© 2025 Zimenkov and Ushtanit. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with
these terms.

Dinucleotide codon substitutions as a signature of diversifying selection in *Mycobacterium tuberculosis*

Danila Zimenkov* and Anastasia Ushtanit

Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia

The evolution of the human pathogen *Mycobacterium tuberculosis* is shaped by various but interconnected processes of drug treatment pressure and host adaptation. We hypothesize that rarely accounted dinucleotide substitutions within a single codon, which allow for a broader range of amino acid substitutions than single nucleotide changes, are a significant aspect of diversifying selection. From the analysis of 43 studies, comprising 11,730 clinical isolates with resistance to rifampicin, 11 different dinucleotide substitutions were identified in 54 codons of resistance-determining regions of the *rpoB* gene. The prevalence of such substitutions is approaching 4%. Although rifampicin was introduced in treatment regimens in the 1970s, dinucleotide substitutions were also found in resistance determinants for newer drugs, linezolid and bedaquiline, *rplC*, and *atpE*, despite the significantly smaller number of resistant clinical isolates reported. Conducting a genome-wide analysis of dinucleotide mutations in the dataset of 9,941 genomes studied by the CRYpTIC Consortium, in addition to resistance determinants, we discovered three genes with a significantly elevated number of dinucleotide substitutions, which are presumably related to virulence and host adaptation. Two substitutions, *cyp138* P114F and *rv0988* L191A are supposed to occur early in the evolutionary history of lineage 2 and are now under strong selection for reverse substitutions. Two amino acid substitutions in the third gene, *rv2024c* N508T and C514L, could also be obtained by single nucleotide changes and therefore are supposedly being selected based on frequency of codon usage. The signature of dinucleotide mutations introduces a novel approach to understanding the evolution of pathogen and identifying potential targets for antivirulence drugs. They underscore the complexity of the evolutionary dynamics within this pathogen, driven by diverse selection pressures, shedding light on the ongoing battle between *M. tuberculosis* and its human host.

KEYWORDS

tuberculosis, drug resistance, evolution, dinucleotide substitutions, virulence, rifampicin, bedaquiline, linezolid

1 Introduction

Tuberculosis is caused by an obligate airborne pathogen, *Mycobacterium tuberculosis*, particularly by its several human-adapted lineages (Sola et al., 2024). They represent an intriguing model for the study of evolution due to the lack of horizontal transfer of genetic information, associated with transmission from

host to host and absence of an environmental niche (Gagneux, 2018). The main source of diversity in this pathogen arises from accumulation of mutations, recombination between repetitive DNA, and loss of genetic information through deletions (Brosch et al., 2002).

Specific tuberculosis chemotherapy began with the development of tibione (Domagk, 1950) and streptomycin (Schatz et al., 1944) in the middle of the 20th century. The establishment of the complex “first-line therapy” in the 1970s involved the use of four to five effective drugs to prevent the development of resistance (Fox et al., 1999). In this scheme, rifampicin and isoniazid are particularly potent against *M. tuberculosis*, targeting RNA polymerase and mycolic acid biosynthesis, respectively (Van Deun et al., 2018). This treatment regimen is still the cornerstone for susceptible tuberculosis cases (World Health Organization, 2022). However, the global rise of multiple drug resistance and the limited number of antituberculosis drugs (Tahaoğlu et al., 2001) require the identification of molecular mechanisms of resistance and the delineation of clinically significant mutations for the personalization of tuberculosis treatment (World Health Organization, 2021; Köser et al., 2012).

The advent of massive whole genome sequencing since 2013 has provided unprecedented opportunities to study phylogenomics and drug resistance mechanisms (Farhat et al., 2013; Zhang et al., 2013). In addition to genome-wide association studies (GWAS), relying on phenotypic resistance data (Chiner-Oms et al., 2022) accounting for homoplasy indexes allows one to identify mutations that emerge independently in different sublineages or clones under selective pressure using only sequence data (Farhat et al., 2013). Additionally, the dN/dS ratio, which quantifies the number of non-synonymous to synonymous mutations in a specific codon or gene, can be used to identify genes under diversifying or purifying selection (Chiner-Oms et al., 2022; Liu et al., 2022). These methods have confirmed many drug resistance mutations and identified new resistance determinants using sets of hundreds to tens of thousands of genomes (Chiner-Oms et al., 2022; Liu et al., 2022; Hicks et al., 2019). They have also led to the discovery of co-evolved markers that can interact with resistance, increasing minimum inhibitory concentrations (MIC), compensating for the fitness cost of resistance mutations, or improving pathogen tolerance to antibacterial treatments (Mo et al., 2015; Perdigão and Portugal, 2019). Unlike resistance, assessing the virulence and fitness of bacteria is a challenge both *in vitro* and *in vivo* due to the difficulties of phenotype testing (Allué-Guardia et al., 2021).

Mutation frequencies of different types are not uniform, even without amino acid or codon selection pressure (Rogozin et al., 2005). Substituting C for T in genomes is the most common type of mutation, attributed to spontaneous deamination of cytosine to uracil, which can then result in thymine in offspring (Schroeder et al., 2018). The non-uniform frequencies of single nucleotide substitutions shape the spectrum of amino acid substitutions (Cano et al., 2022) and are included in the Kimura two-parameter model and sequence distance measurements (Kimura, 1980). Rarely identified dinucleotide mutations in the same codon (designated further as 2N) are usually considered to evolve sequentially when distinct species are compared (Bazykin et al., 2006; Belinky et al., 2019). The same simplification is used in

methods of evolutionary distance calculations, where the frequency of 2N substitutions is set to zero (Yang, 2014).

However, based on genetic code redundancy and inability to obtain particular amino acid substitutions with single nucleotide substitutions the additional value of 2N substitutions as ‘evolutionary shortcuts’ (Lucaci et al., 2023) helping to ‘cross the adaption valleys’ (Belinky et al., 2019) could be proposed. Protein structure restrictions favour such complex traits over substitutions obtained by single nucleotide changes in specific cases. Particularly, such mutations could occur in codons encoding the structurally important protein residues, forming the small-molecule binding pockets, or involved into interactions with other proteins, either pathogen, or host nature.

As evidenced by studies on eukaryotic DNA polymerases, 2N substitutions may emerge in a single mutation-repair stage (Schridder et al., 2011; Stone et al., 2012), reflecting the adaptation of the cell at the amino acid level (Venkat et al., 2018). Analyzing *M. tuberculosis* genomes, we have noticed that these mutations are not rare indeed and could serve as marker of diversifying selection, previously underestimated. We analyzed the frequencies of such mutations from studies on resistance to rifampicin, then, on few recently emerged cases with resistance to novel drugs bedaquiline and linezolid, and finally, we examined 9,941 genomes from the CRyPTIC project for genome-wide analysis of genes under positive selection reflecting both resistance acquisition and host adaptation process.

2 Hypothesis: dinucleotide substitutions in codon reflect a strong selection process at conservative protein sites

2.1 Dinucleotide substitutions in *rpoB* gene are frequent in *Mycobacterium tuberculosis* clinical isolates

Rifampicin has been a cornerstone drug in first-line tuberculosis treatment since the early 1970s, often used in conjunction with isoniazid, ethambutol, and pyrazinamide (Fox et al., 1999). Resistance mutations to rifampicin occur in the *rpoB* gene encoding the beta subunit of RNA polymerase III (Ovchinnikov et al., 1983; Jin and Gross, 1988; Telenti et al., 1993). These mutations are found in four distinct segments known as the rifampicin resistance determining region, which comprises the exit tunnel of the RNA-polymerase complex (Campbell et al., 2001; Molodtsov et al., 2017) where the drug binding occurs (Lin et al., 2017).

Many studies of clinically resistant isolates have revealed that the mutation most frequently encountered is S450L (codon 531 in *Escherichia coli* numbering) (Goldstein, 2014). This mutation is believed to have the least impact on bacterial fitness, and strains carrying it often exhibit extremely high rifampicin MIC. However, many other mutations are consistently found in clinical isolates (Walker et al., 2022). We decided to analyze the spectra of potential individual amino acid substitutions and reviewed publications starting from a comprehensive TBdreamDB database published in 2009 (Sandgren et al., 2009).

Before the era of whole genome sequencing, various methods were used to detect resistance-associated mutations, such as allele-specific or realtime PCR for already known changes in the genome (Pholwat et al., 2011) or Sanger sequencing for the identification of new ones (Ovchinnikov et al., 1983). We had to exclude data obtained by widely used strip hybridization technology by Hain Lifescience (Hillemann et al., 2007) due to partial information loss on specific mutations, detected by the absence of hybridization signal of the amplicon with wild-type probe (Miotto et al., 2018). Some studies using targeted Sanger sequencing lacked raw data and provided only amino acid substitutions, and thus were also excluded from codon substitution pattern analysis. The general data on the proportion of substitutions is shown in Figure 1. We also tried to exclude isolates with multiple mutations that occur in different places in the RpoB to avoid the effects of epistatic interactions, which would require a larger number of isolates to be properly evaluated. We analyzed mutations in both resistant and susceptible isolates because low-level resistance to rifampicin caused by substitutions in codons such as 435 was probably missed. In 2020, the critical rifampicin concentration was lowered from 1 mg/L to 0.5 mg/L for the Bactec MGIT 960 system, and therefore, isolates previously identified as susceptible could indeed have been resistant (Antimycobacterial Susceptibility Testing Group, 2022). Additionally, we were interested in analyzing all protein sequence variations that do not significantly affect the activity of RNA polymerase or the pathogen's adaptation properties.

In total, our study included 11,730 isolates from 43 studies (Supplementary Table S1). More than half of the isolates had mutations in codon 450, followed by codons 445 and 435. Remarkably, a diverse range of amino acid substitution types was observed in all codons. Even at codon 450, where the S450L mutation dominated and was found in 67% of the cases, eight other amino acid substitutions were identified. Seven of these could not be achieved through single nucleotide substitutions, and the total proportion of 2N mutations for this codon was 2.5% (see Figure 1). The most diverse codon was 445, which codes for histidine in the wild-type isolate. Thirteen different amino acids were tolerated by RNA polymerase in this position, and 2N mutations were found in 7% of isolates with an altered 445 codon.

In general, it can be concluded that 2N substitutions provide amino acid diversity that cannot be achieved through 1N mutations (Belinky et al., 2019). However, it is worth noting that, in some cases, the same amino acid substitution could also result from a single nucleotide change, highlighting the inherent stochastic nature of evolution. The frequencies of alternative 2N are two orders of magnitude lower, as is evident from the analysis H445L and L452P mutations (see Figure 1).

In the specific context of drug-induced selective pressure, these substitutions decrease rifampicin binding to RNA polymerase without adversely affecting polymerase activity (Trinh et al., 2006). The wide range of substitution types observed can be attributed to the mechanism of drug action. Rifampicin binds to the DNA/RNA channel at a distance that exceeds 12 Å from the active site, effectively obstructing RNA elongation beyond 2–3 nucleotides (Campbell et al., 2001). It could be supposed that the variability would be lower for the drugs that bind directly to more conservative active sites of essential bacterial proteins.

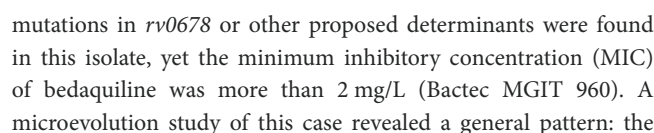
2.2 Dinucleotide substitutions, associated with resistance to bedaquiline and linezolid

Bedaquiline and linezolid were introduced into tuberculosis treatment 40 years after rifampicin, in 2014 and 2010, respectively. Since 2014, they have been used in the BPaL combined therapy scheme for drug-resistant TB around the world (Conradie et al., 2020). The safety and efficacy of this scheme led to the redefinition of tuberculosis with extended drug resistance, now defined as resistance to first-line drugs plus fluoroquinolones and bedaquiline or linezolid (Viney et al., 2021; Veziris et al., 2021). Although acquired drug resistance to these new drugs remains relatively low, there are a growing number of reports of clinical isolates with resistance (Timm et al., 2023).

The primary mutations selected during bedaquiline treatment emerge in the *rv0678* gene, which encodes the repressor of the efflux operon (Andries et al., 2014). Loss-of-function mutations in this repressor have a minimal fitness cost and can be distributed throughout the open reading frame (Hartkoorn et al., 2014). Although most of them are frameshift mutations, amino acid substitutions in critical regions are also observed (Zimenkov et al., 2017). However, the huge number of possibilities to inactivate the gene by frameshifts limits the population frequency of complex and rare substitutions.

In contrast, substitutions in the bedaquiline target, the *c*-subunit of ATP synthase, which comprises the rotor part, are more conservative and rare, limited by the essentiality of the gene and, probably, by the fitness cost (Ghodousi et al., 2022). Mutations in the corresponding *atpE* gene were initially identified *in vitro* and later confirmed in clinical strains of *Mycobacterium intracellulare* (Alexander et al., 2017) and *M. tuberculosis* (Zimenkov et al., 2017). To date, there are ten studies reporting 19 strains with substitutions in codons 25, 28, 61, 63, and 66 of *atpE* in clinical strains of patients who received bedaquiline in treatment schemes and developed decreased susceptibility or resistance (Zimenkov et al., 2017; Martinez et al., 2018; Peretokina et al., 2020; Mokrousov et al., 2021; Chesov et al., 2022; Ghodousi et al., 2022; Le Ray et al., 2022; Shang et al., 2023; Umpeleva et al., 2024; Peng et al., 2024). The WHO 'Catalogue of mutations' reported eleven isolates with non-synonymous mutations in *atpE*, but these sources of data overlap significantly (Walker et al., 2022). AtpE I66M and A63P are the most frequent substitutions found in six (Mokrousov et al., 2021; Chesov et al., 2022; Le Ray et al., 2022; Umpeleva et al., 2024) and four cases (Ghodousi et al., 2022; Chesov et al., 2022; Umpeleva et al., 2024; Peng et al., 2024), respectively. Additional substitutions in these codons, A63V and I66V were also described (Zimenkov et al., 2017; Martinez et al., 2018); however, all reported substitutions resulted from a single nucleotide change.

During the study of bedaquiline resistance in tuberculosis cases, we identified an isolate with two simultaneous mutations in *atpE*. This strain had two single nucleotide mutations at different positions, resulting in G25S and D28G amino acid substitutions (Zimenkov et al., 2017). The simultaneous presence of these two mutations in a single cell was confirmed by whole genome sequencing. We believe such occurrences are less frequent than single-nucleotide mutations, either it was a sequential or simultaneous mutational event.



acquisition of the frameshift mutation t141tc in *rv0678* initially occurred in a mixed state, followed by the extinction of the wild-type strain, and ultimately its elimination with simultaneous development of the AtpE substitution. A similar microevolutionary pathway has been proposed based on clinical studies and *in vitro* mutagenesis (Ismail et al., 2019).

Interestingly, some other mycobacterial species are intrinsically resistant to bedaquiline. It was suggested that the presence of methionine in position 63 of AtpE, compared to alanine in *M. tuberculosis*, is responsible for higher MICs in species such as *Mycobacterium flavescens* and *Mycobacterium xenopi* (Aguilar-Ayala et al., 2017; Petrella et al., 2006). Alanine 63 interacts directly with bedaquiline (Guo et al., 2021) and is substituted with proline or valine in resistant clinical isolates and *in vitro* selected strains (Peretokina et al., 2020; Andries et al., 2005). Both A63P and A63V are the result of 1N substitutions in codon GCA. Other residues, that could be obtained by 1N include E, G, S, and T. The wild-type A, also as P, V and G residues have similar biochemical properties being small, non-polar and hydrophobic. The absence of A63G substitution in clinical strains, resistant to bedaquiline could be explained by the steric properties of glycine, which breaks the alpha helix structure. Dinucleotide substitutions add R, D, Q, I, L, K to the diversity of available amino acids in this position. R, D, Q are hydrophilic, which are unlikely to present of the ATP synthase rotor part exposed to the membrane. Further, K is a positively charged residue, which could have a consequence during the rotation and contact with the stator. Thus, L and I remains the most probable candidates for the 2N substitution of A63 that preserve the function of the enzymatic complex. The substitution of A63 codon for M could be obtained by the change of all three nucleotides in codon (GCA>ATG).

A second example of a rare 2N mutation was observed during a clinical trial in a case of ineffective treatment, resulting in resistance to linezolid. Linezolid inhibits translation by interacting with the ribosome, and resistance mutations are found in distant regions of the 23S rRNA, forming the peptidyl-transferase center. However, most resistant isolates of *M. tuberculosis* carry the substitution C154R in the ribosomal protein RplC, which is typically caused by a single nucleotide change t460c. We identified a resistant isolate with mixed chromatogram data, and Sanger sequencing was unable to distinguish between the simultaneous presence of the nonstandard C154R (CGG codon) and C154S (AGT codon) or two different types of C154R replacements (Zimenkov et al., 2017). Further whole genome sequencing confirmed the latter, where one subpopulation had the usual t460c mutation, while the other had the TGT codon substituted with AGG by the 2N mutation (Figure 2). The simultaneous presence of two different subpopulations was confirmed by the analysis of subsequent isolates; Even after 15 months, a mixed chromatogram was identified in the clinical isolate obtained from the patient.

Both these cases illustrate the rapid emergence of resistance under the pressure of drug treatment. Similarly to resistance to rifampicin, the 2N mutation in *atpE* expands the diversity of the primary protein structure, while in *rplC*, such a mutation does not lead to novel amino acid substitution compared to the commonly found single nucleotide mutation t460c. Although the number of mutated isolates obtained during and after ineffective treatment is not sufficient to provide strong statistical power

for correlation analysis (Nimmo et al., 2022), these cases of 2N mutations undeniably provide evidence of ongoing selection in these loci, and directly point to resistance mechanisms. Consequently, it could be proposed that mutations in these codons are clinically relevant and reflect the development of resistance upon the antibacterial treatment.

2.3 Dinucleotide substitutions for the prediction of diversifying selection sites in the genome

Although the described examples of bacteria adaptation to drug action unequivocally demonstrate the elevated level of 2N mutations in genetic determinants of resistance, it was interesting to perform a genome-wide analysis of dinucleotide mutation frequencies for the identification of all genes under selection pressure. We assumed that additional resistance determinants and host-pathogen interaction could be found by this approach. For the analysis, we used a data set consisting of 9,941 of 12,288 isolates that were sequenced and analyzed by the CRyPTIC Consortium (The CRyPTIC Consortium, 2022b; The CRyPTIC Consortium, 2022a). The remaining 2,347 had an ambiguous description of amino acid substitutions and were omitted.

Two main approaches were used for the analysis of mutation frequencies: allele counting and homoplasy counting (Chen and Shapiro, 2015). Although the former uses frequencies observed in the population and does not require intensive calculations of phylogeny and ancestor reconstruction of mutation events, it has significant disadvantages in its application to *M. tuberculosis*. The clonal nature of the pathogen led to complete linkage disequilibrium and population stratification, which could result in false associations obtained by uncorrected allele counting methods (The CRyPTIC Consortium, 2022b).

The homoplasy count derived from the number of independent events in different branches (Farhat et al., 2013) could be a signal for positive selection (Shapiro et al., 2009). Following this method, we constructed the phylogenetic tree using mutations from all genes, excluding highly repetitive PE/PPE, PE-PGRS genes and insertion elements (Supplementary Figure S1). The reliability of the phylogenetic tree was verified by analyzing the distribution and frequency of selected lineage-specific SNPs (Coll et al., 2014).

The maximal parsimony approach was applied to the set of all variations at a specific codon. All mutations were mapped on the tree and traced by joining neighbors with this mutation and mapping it to the common ancestor. It was supposed that reverse mutations could have occurred if it allowed the joining of branches with mutations split by smaller branches with the wild-type codon. Thus, the final number of all events was minimized for each codon across the whole tree.

As an example, the mutation E112K in the *rv0078A* gene (codon change GAG>AAG, nucleotide substitution c87468t), which could be used as a phylogenetic marker of lineage 2 (Coll et al., 2014), was present in 3,838 of 9,941 isolates. Indeed, they all belong to this lineage ($n = 3,842$ isolates) of the constructed *M. tuberculosis* phylogenetic tree; however, four isolates in L2 did not have this mutation. The phylogeny-adjusted frequency of independent events

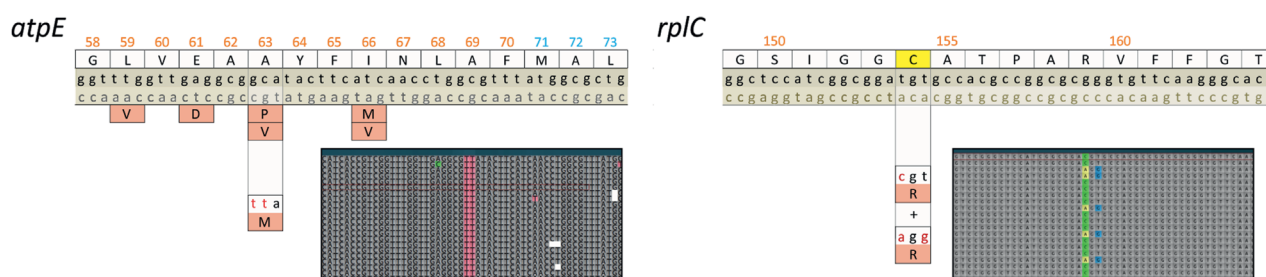


FIGURE 2

Two cases of dinucleotide substitutions in 63 codon of *atpE* and 154 codon of *rplC* genes identified by the whole genome sequences of the clinical *Mycobacterium tuberculosis* isolates. Isolates were obtained from the patients receiving the treatment with bedaquiline and linezolid as core drugs, and these substitutions lead to resistance to bedaquiline and linezolid, respectively. The second case contained two subpopulations of *rplC* C154R substitution—canonical (tgt > cgt) and dinucleotide (tgt > agg) codon substitutions. Screenshots of whole-genome reads (Illumina MiniSeq) aligned to the reference genome of *M. tuberculosis* H37Rv made in UGENE software.

in this case for the entire population was equal to 23. The introduction of the maximal parsimony approach, based on the proposal that in these four isolates the reverse substitution occurred recently, allowed the decrease in homoplasy counts to two for direct and reverse mutations (four genomes with ‘wild-type’ Lys at 112 constitute one branch). Therefore, the direct mutation was traced to the common ancestor of lineage 2.

In such an approach, the consequent 1N mutations leading to final 2N substitution were treated as separate, thus underestimating the final diversity of codon substitutions and decreasing the numbers of 2N events. However, counting the second mutation as 2N will lead to an overestimation due to cases where the first mutation was specific to the branch, for example, the whole lineage 2. Thus, the analysis was limited to the calculation of simultaneous 2N substitutions, providing a more specific but less sensitive result in terms of codon diversity at particular position.

The resulting number of events for 275,531 codons in 3,680 genes analyzed were the following: 142,321 events of synonymous substitutions, 252,702 events of 1N substitutions and 1,645 events of 2N substitutions. The latter occurred in 862 codons in the genome, with a maximum value of 350 for the substitution of *cyp138* P114F. Excluding the genes with single events of 2N substitutions results in 251 codons in 97 genes. The comparison of the number of codons with 2N substitutions and 2N events for each gene is shown in Figure 3. The two most important genes, that cause resistance, *rpoB* and *katG*, have elevated parameters and are clearly distinguishable from the entire set of genes.

Splitting of the counts of nonsynonymous mutations into 1N and 2N mutations allowed the introduction of two metrics, the ratio of dinucleotide to single-nucleotide nonsynonymous substitutions d2N/d1N, and ratio of dinucleotide to synonymous substitutions d2N/dS. In the commonly used GWAS approach, the comparison of dN/dS in resistant and susceptible sets of isolates is performed for identification of resistance determinants or genetic traits that co-evolve with resistance (The CRYPTIC Consortium, 2022b). We used another approach for identification of all genes under selection by the use of codon-based models of evolution and estimation of the difference between measured and expected frequencies of codon substitutions (Delpont et al., 2009; Anisimova and Kosiol, 2009).

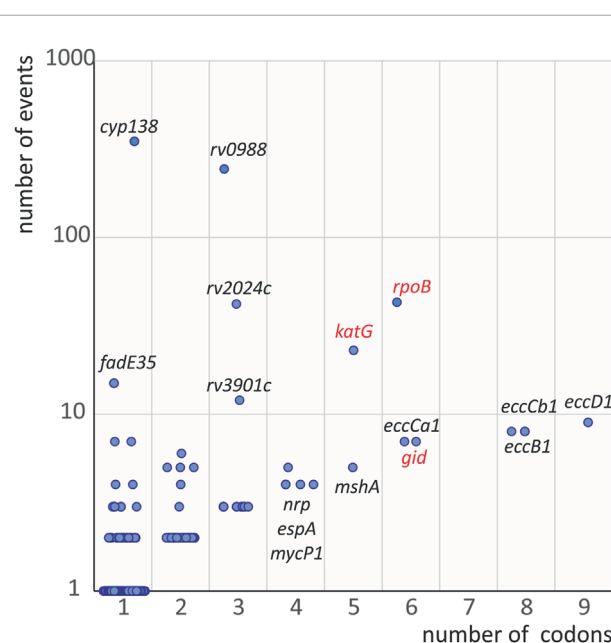


FIGURE 3

The variability of phylogenetically-adjusted frequencies of 2N substitutions and number of affected codons in *Mycobacterium tuberculosis* population (CRYpTIC study). Known drug resistance determinants are marked with red.

The codon substitution matrix for the Monte-Carlo modeling was derived from the whole dataset of phylogeny-adjusted events. Additional correction was introduced by proportional use of all variants of the particular codon in the whole set of isolates as the starting codon in modeling. Otherwise, lineage-specific codons with high prevalence in the genome set would not add to the calculated substitution frequencies. The difference between observed and estimated frequencies allows the identification of any genomic traits that are putatively under selection (Wilson and Consortium, 2020). The statistical approach for the identification of significant traits was analogous to that of the widely used ratio of non-synonymous to synonymous substitutions dN/dS (The CRYPTIC Consortium, 2022b). Therefore, three values for genes were analyzed and

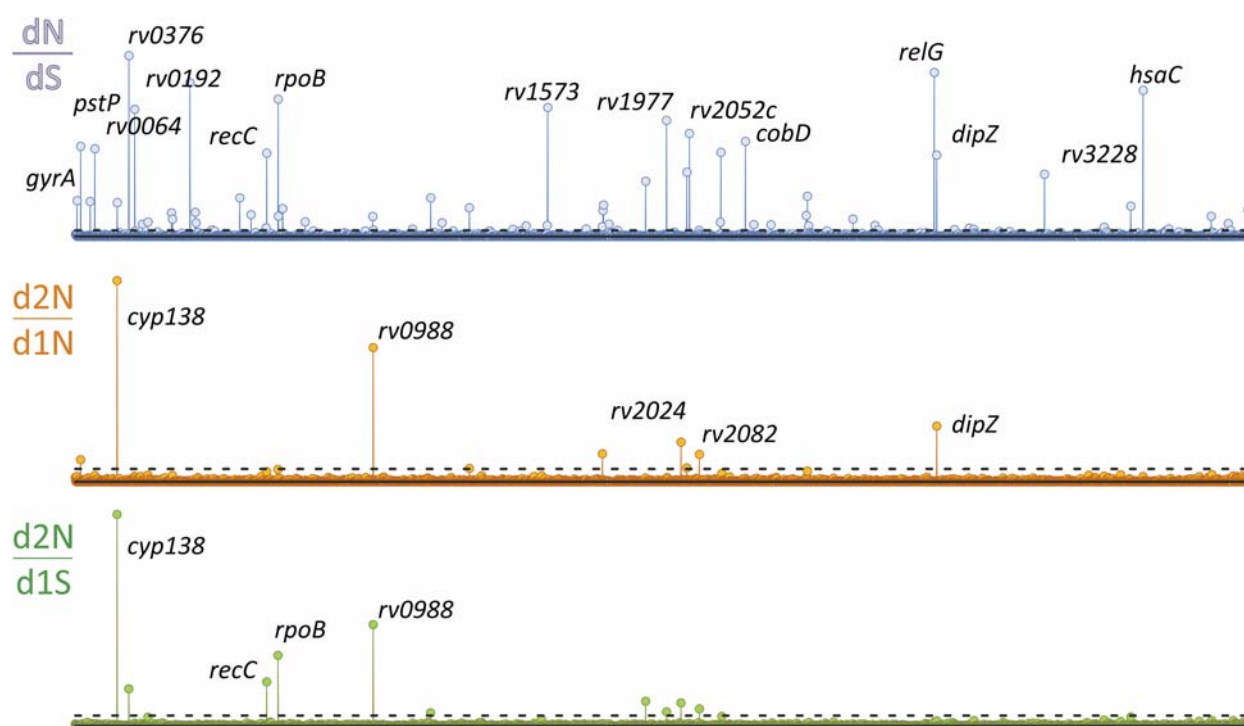


FIGURE 4
Genome-wide study of traits under selection from the analysis of 9,941 *Mycobacterium tuberculosis* genomes of the CRYpTIC study. Manhattan plots for the three ratios were used: total non-synonymous to synonymous (dN/dS), dinucleotide to synonymous (d2N/dS), and dinucleotide to single nucleotide substitutions (d2N/d1N) are plotted along the chromosome. The statistical significance identified by the Bonferroni correction are shown with dotted lines.

compared simultaneously presented as Manhattan plots (Figure 4), two for 2N substitution statistics and conventional dN/dS was used as reference. The significance threshold of $-\log(P)$ value was estimated by Bonferroni correction (Power et al., 2017).

The phylogeny adjusted counting of single-nucleotide mutations of different type and the comparison between genes resulted in distinct statistical signals for 68 genes, including the main resistant determinants *rpoB*, *katG*, *rpsL*, *pncA*, *gid*, *embB*, and *ethA* (Supplementary Table S2). Fourteen genes had a statistically significant change, either increase or decrease, in the dinucleotide events statistics, including the resistance determinants *rpoB*, *katG*, and *pncA*. Two genes, *rv2024c* and *rv2082*, with low *p*-values of d2N/d1N and d2N/dS, were not significant by dN/dS analysis (Table 1).

The statistically significant increase in 2N substitutions was observed for five genes only (Table 1). The *rpoB* gene had the most diverse mutation types and the largest number of isolates due to its role in resistance to rifampicin. Eighty-seven isolates with twelve 2N substitutions in six codons were found. Homoplasmy counts of 2N mutations in the whole gene were 43, while that of 1N mutations was 2,216 (3%). Several types of dinucleotide substitutions were identified in codons 435, 445, and 450, similar to the pattern described in Section 2.1. An additional new 2N mutation, I542K, was also found.

In addition to *rpoB*, the main isoniazid resistance *katG* gene was found to be significant by dN/dS and d2N/dS ratios. The most

prevalent S315T mutation found in resistant isolates is obtained by single nucleotide G>C substitution (codon AGC is changed to ACC). In this set of isolates, three other dinucleotide substitutions were identified that led to the same amino acid change: AGC codon was changed to ACA, ACG, or ACT.

The favored use of particular amino acids in drug-binding pockets with low fitness-cost could explain this observation for resistance genes *rpoB* and *katG*. Other genes associated with resistance to first- and second-line drugs in *M. tuberculosis*, such as *rpsL*, *embB*, *gyrA*, *gid*, and *ethA*, were significant only by dN/dS analysis.

Two more genes, that were not previously linked with resistance, had significant statistical signals for 2N substitutions, *cyp138* and *rv0988*. Both amino acid substitutions, *cyp138* P114F (CCT > TTT) and *rv0988* L191A (CTG > GCG) cannot be obtained by single nucleotide substitution, confirming the importance of particular amino acid residues at these positions.

The predicted cytochrome P450 *cyp138* is involved in pH tolerance mediated by MTS1338 sRNA regulation (Martini et al., 2023). Cytochromes P450s Cyp138 join metabolism, virulence, and susceptibility to antibiotics by altering cell wall composition (Lu et al., 2024). The gene *rv0988* is a subunit of the ABC transporter (Rosas-Magallanes et al., 2006) and negatively regulated in the presence of toxic levels of CuCl₂ (War et al., 2008).

Both *cyp138*(P114F) and *rv0988*(L191A) emerged in many unrelated strains of the *M. tuberculosis* lineage 2 (East-Asian)

TABLE 1 Top significant genes identified by the elevated 2N events in single codon.

Gene	Resistance, or resistance-associated gene ^a	Number of S mutations	Number of 1N mutations	Number of 2N mutations	Statistical significance			2N mutations frequency odds
					dN/dS	d2N/d1N	d2N/dS	
<i>pncA</i>	R	25	886	0	•	•		↓
<i>rpoB</i>	R	108	2,108	43	•		•	↑
<i>cut1</i>	R	9	160	0	•	•		↓
<i>pstP</i>		89	1,705	0	•	•		↓
<i>dipZ</i>		83	2,599	1	•	•		↓
<i>cyp138</i>		64	108	350	•	•	•	↑
<i>katG</i>	R	100	1,396	23	•		•	↑
<i>Rv0988</i>		211	271	244	•	•	•	↑
<i>Rv2082</i>		819	1,223	0		•	•	↓
<i>Rv2024c</i>		102	189	42		•	•	↑
<i>recC</i>		807	183	1	•		•	↓
<i>Rv1190</i>		347	50	0	•		•	↓
<i>mce1F</i>		2,201	205	1	•		•	↓
<i>Rv1977</i>		684	68	0	•		•	↓

^aGene associations taken from WHO, Catalogue of mutations (World Health Organization, 2023) and CRyPTIC, study (The CRyPTIC, consortium, 2022b).

and were absent in all other lineages: lineage 1 (Indo-Oceanic), lineage 3 (East-African-Indian), lineage 4 (Euro-American). Cyp138 (P114F) is present in 3,154 of 3,527 isolates of sublineage L2.1. Such lineage-specific distribution differ from the distributions of resistance-associated variants, which are present in all human-adapted lineages–L1, L2, L3, L4, which is cause by the common therapy used in different regions of the world. Thus, it could be supposed that mutations in *cyp138* and *rv0988* are related to the pathogen adaptation to host pressures from immune system and antituberculosis drug use. Alternatively, these traits could be responsible for the evolutionary success of Lineage 2, which was widely discussed in the phylogeography studies on tuberculosis (Zhu et al., 2023).

Using the maximal parsimony approach, we suppose that P114F occurred only once in the common ancestor of 2.1, and then, in 373 isolates the reverse substitution F114P emerged (Figure 5). The phylogeny adjusted frequency of the reverse mutation is 349, resulting in 350 total events in the codon for this set of isolates. Alternatively, single-stage accumulation of P114F in the 3,154 genomes demands 1,130 mutational events.

The emergence of *cyp138*(P114F) was recently described as having evolved under the pressure of HIV-1 coinfection (Koch et al., 2017). However, taking into account the model of sequential forward and reverse mutations in Lineage 2 strains, the opposite situation

could be supposed: the reversion F114P occurred under the pressure of immunocompetent HIV-naïve hosts.

The same considerations are also relevant for *rv0988*(L191A): forward substitution in the common ancestor of a smaller sublineage within L2.1 and reversion to Leu in many unrelated strains, which is common for all other lineages (Figure 5). The phylogeny-adjusted frequency in the absence of reversion hypothesis was 742, compared to 241 events of reverse substitutions. Interestingly, the distribution of A191L reversion is not unified in L2 sublineages. For example, the Beijing BO/W148 genotype, which is a successive clone spreading in Russia (Mokrousov et al., 2012), has a low frequency of reversions, while some other branches are nearly completely reverted (Figure 5). Such a non-random purge of probably deleterious polymorphisms (Namouchi et al., 2012) demands further studies on correlation with geographical localization of clinical isolates and impact on virulence and pathogenicity.

Contrary to the previous, the substitutions N508T and C514L in *rv2024c* could be achieved also by simple 1N substitutions in corresponding codons, however, the codon usage in these cases is significantly lower compared to obtained by observed 2N substitutions. Thus, the observed AAT>ACG for N508T changes frequency from 5.3 to 35.1, while the codon Thr codon ACT has frequency of 3.7 (Naka et al., 2000). Similarly, for the C154L codon

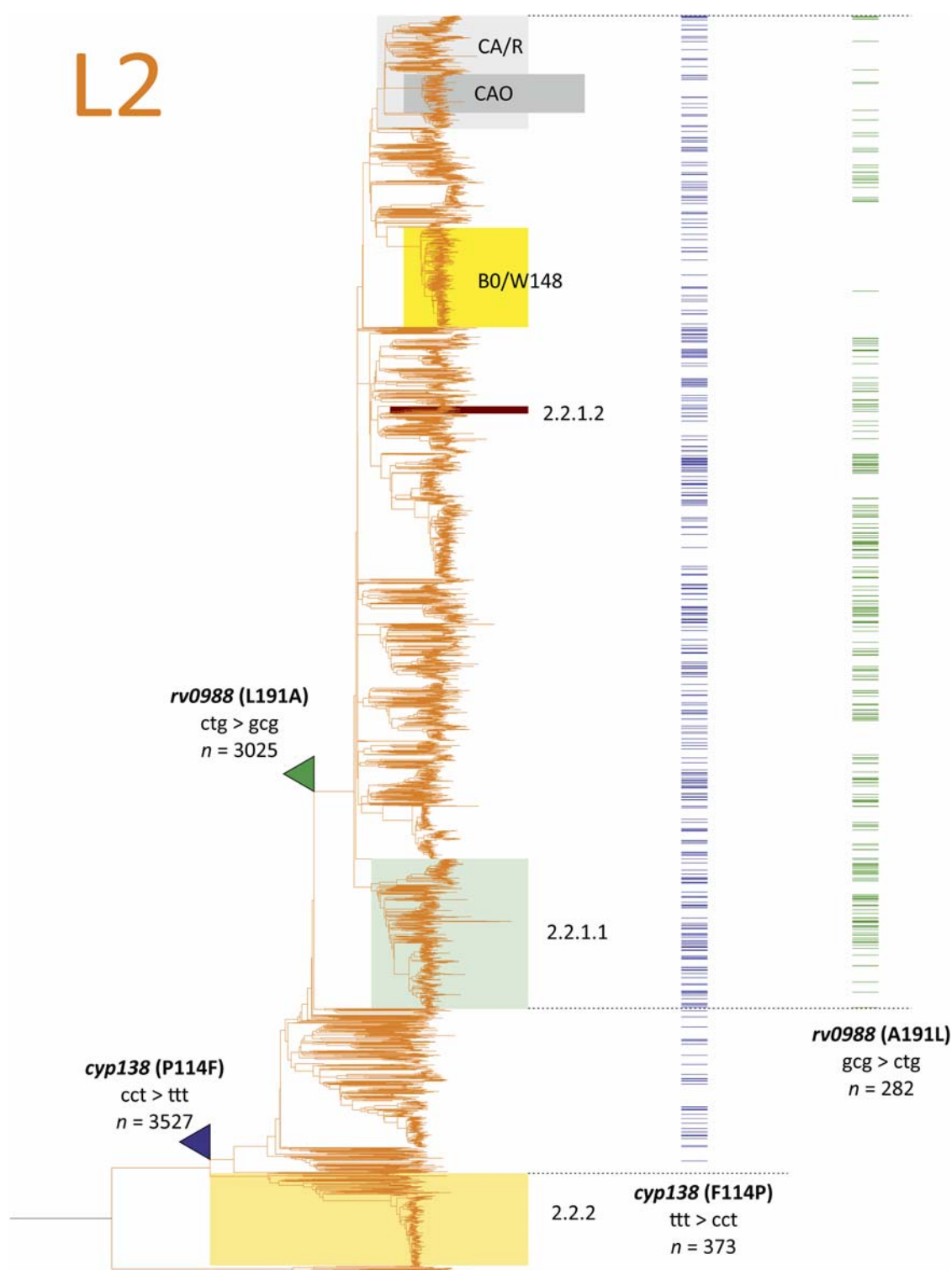


FIGURE 5

The maximal parsimony model of selection of lineage 2 specific 2N substitutions *cyp138*(P114F) and *rv0988*(L191A) from the analysis of 9,941 *Mycobacterium tuberculosis* genomes of the CRYpTIC study. The branches where probable direct substitutions occurred are pointed with color triangles, reversions are shown as distributions in terminal branches (isolate genomes). Number of isolates with mutations in the whole CRYpTIC set are given unadjusted.

usage frequency increased from 8.1 to 50.4, while the alternative Leu CTA codon, which could be obtained by 1N substitution, has frequency of 4.8 (Naka et al., 2000). The exceptional use of 2N substitutions in unrelated strains also points to the preference for particular amino acid at these positions of the protein. Rv2024 has an unknown function but showed diversity in parallel and sequential isolates during the within-host microevolution analysis (Ley et al., 2019) and is believed to have a virulence-associated function, since it was found to be induced in mouse lungs (Dubnau et al., 2005).

Two more classes of genes could be identified among those with a low number of 2N substitutions: under positive selection ($dN \gg dS$) and under purifying selection ($dS \gg dN$). A high number of non-synonymous substitutions were identified for *pncA*, *cut*, *pstP*, *dipZ*, and *Rv2082* in the top-scored list of genes (Table 1). Mutations in *pncA* that lead to the loss of function of the gene are selected during pyrazinamide treatment and lead to resistance. Such mutations are frequent, distributed along the open reading frame, and include frameshifts. Therefore, the low number of 2N substitutions is not an unexpected phenomenon, also as for the *rv0678c* gene alterations leading to the resistance to bedaquiline. In addition to selection for loss of function, genes with high 1N and low 2N counts could encode proteins that are tolerant to substitutions, or the possible changes with low fitness cost are achieved by 1N codon substitutions.

Genes with a high number of synonymous substitutions have low counts of 1N and 2N substitutions, and the statistical significance of the $d2N/dS$ ratio is the direct consequence of the significance of dN/dS ratio.

3 Discussion

Massive whole genome sequencing has allowed unprecedented possibilities for studying evolution and cellular mechanisms, including drug resistance. Various approaches have been developed, such as GWAS (The CRYPTIC Consortium, 2022b), dN/dS (Zhang et al., 2013; Namouchi et al., 2012; Liu et al., 2022) and PhyC (Farhat et al., 2013), which are mostly based on estimating mutation frequencies at specific genomic locations and correlating them with phenotypic characteristics. These methods involve calculating the ratio of non-synonymous substitutions to synonymous substitutions and assessing homoplasy, where the same type of mutation is present in different sublineages of *M. tuberculosis*. Since there is no known horizontal gene transfer and a known environmental niche, mutations in different lineages and sublineages occur independently, reflecting selective pressures.

Bacterial evolution pathways are highly diverse and, in the case of tuberculosis, transmission and survival within a host have shaped the pathogen's adaptation over thousands of years (Brosch et al., 2002). The recent introduction of antibacterial drug treatment, starting with drugs such as streptomycin and tibione, has become a dominant force driving evolution of the pathogen. In this context, diversifying selection has led to changes in amino acids that interact with drugs in their target proteins (Thiltgen et al., 2017). However, it should be noted that in cases of ineffective treatment and the transmission of resistant isolates within a population, we may observe purifying selection, where mutations that result in fitness

costs are eliminated from the population (Comas et al., 2010). On the other hand, fitness-compensating mutations may emerge, representing another form of diversifying selection driven by forces distinct from direct drug action (Gillespie, 2001).

The development of drug resistance is a time-limited event, whereas host pressures drive the ongoing competitive evolution of both the pathogen and the host (Thiltgen et al., 2017). This phenomenon, often referred to as the “Red Queen Effect” (Van Valen, 2014; Solé, 2022), influences various virulence factors in *M. tuberculosis*. These factors enable the pathogen to survive within macrophages by suppressing apoptosis, disseminate throughout the host, utilize diverse nutrient sources, and manipulate host cell processes (Brites and Gagneux, 2012).

Both directions of evolution are highly interconnected, since resistance selection drives the adaptation process through mutagenesis. Recent findings indicate that selection for resistance to rifampicin, for example, leads to an increase in the genome-wide mutation rate, thus accelerating evolution (Cebrián-Sastre et al., 2023).

The diversity of mutations in a population results from the combined effects of chemical mutagenesis, repair mechanisms, and possible selection preventing the essential functions of the cell, ensuring the adaptability and low fitness cost. Other, less measurable or estimated, pressure forces originate from the secondary structure of RNA, its stability, cross-interactions between different RNA molecules, transcription and translation efficiencies, and more (Rahman et al., 2021; Ro et al., 2024).

From a protein perspective, the structural and functionally tolerated variability of amino acids at a specific position cannot be achieved by single nucleotide changes alone. While a single nucleotide change at a particular codon can result in approximately seven amino acid substitutions, more complex dinucleotide and trinucleotide mutations can generate all the range of peptide sequences, albeit with lower probability. As a result, the frequencies of 2N mutations leading to beneficial amino acid substitutions that cannot be obtained by 1N substitution in codon tend to be higher compared to the baseline frequency. Another driving forces for preferential use of 2N substitutions could come from codon usage frequency, secondary mRNA structure, and antisense mRNA interactions. In this case, the same amino acid substitution could be achieved also by single nucleotide substitution.

We observed both types of 2N substitutions in *M. tuberculosis* genomes; however, the total number of mutational events was low. The described approach could provide much informative results if the significantly larger set of genomes will be analyzed. Nowadays, more than 200,000 genomes of *M. tuberculosis* is available in NCBI SRA database ($n = 219,660$, database accessed on 23 November 2024). Since no phenotype data is needed for the described approach, more genetic traits significant for the pathogen adaptation would be identified upon the expanding the analyzed set of genomes. We observed many genes with altered frequency of 2N substitutions, that did not achieved the desired statistical significance threshold. They include already known genes that are associated with resistance, or co-evolved with the emergence of resistance: *gyrA*, *ethA*, *rpoC*. The statistical significance of 2N substitutions in these genes will be improved with the growing number of isolates, as well as for other important unidentified genes and codons. However, the inclusion of mutational data for

hundreds of thousands of genomes compared to ten thousand used in this study will require much more intensive data processing and advanced computing facilities.

While the current knowledge about tuberculosis drug-resistance mechanisms was summarized in two issues of WHO's 'Catalogue of mutations,' virulence-associated traits remains to be obscure in general. An important result of the tuberculosis evolution studies was the description of 'successful' clones or sublineages rapidly spreading in population (Mokrousov et al., 2012). Such clones are associated with resistance, virulence and poor outcomes. The developed molecular methods for the identification of such important genotypes, most probably, have its main clinical significance in preventing of superinfection in hospitals. However, we do not know, which particular mutation, or complex of mutations were responsible for the evolutionary success of these sublineages, while the difference between pathogenic properties of different genotypes has been observed in model experiments (Mokrousov et al., 2023).

Experiments with live *M. tuberculosis* are not easily performed due to the safety considerations, low growth rate, and limited number of genetic tools. Even not every resistance-associated variant was verified by its introduction into wild-type strain and phenotype testing of the recombinant clone. The model organism *Mycobacterium smegmatis* is not suitable for virulence estimation, and the mouse model of tuberculosis infection does not reflect all the aspects of human infection (Kramnik and Beamer, 2016). Thus, statistical approaches of massive genome analysis remains the main source of information of resistance, fitness-compensation and virulence traits of the *M. tuberculosis* genome.

In a massive genome-wide analysis using the dinucleotide statistics we observed the strong signal of selection in two genes not associated with resistance, *cyp138* and *rv0988*. The frequencies of substitutions were elevated in two sublineages of lineage 2. Since we used the hypothesis of possible reversions and maximal parsimony approach minimizing the number of events, we supposed that the direct substitution in the common ancestor occurred first, and then many back substitutions took place reverting to the amino acid that is common to all other lineages in this position. The reversion hypothesis have some limitations. First, only single reversion was allowed, which we suppose to be true due to low mutation rate in *M. tuberculosis*. Second, while the reversion is likely for the case of E112K substitution in the *rv0078A* gene, which were present in 3,838 isolates of 3,842 lineage 2 isolates, the cases of substitutions in *cyp138* and *rv0988* genes have much more comparable counts with or without reversions. However, the number of mutational events without the reversions would have been three times higher, and if we used such higher counts of events, the confirmation of 2N substitutions importance in these cases would have been more statistically significant. By minimizing the events number we thus used more strict approach that could have missed some traits, but the provided result is more specific.

The result obtained in this study provide some clues to some novel mechanisms behind the adaptation of the *M. tuberculosis*. We observed not only the preferential use of complex dinucleotide mutations at particular codons, but also the lineage-specific selection of mutations, which is most probably caused

by some host pressure, either from the host immune response or drug use. The frequency of Rv0988 A191L amino acid substitution is extremely low in 'successful' sublineages Beijing B0/W148 (Mokrousov et al., 2012) and Central Asia/Russian (Mokrousov et al., 2018) compared to other Beijing clones. Future introduction of such traits into diagnostics could allow the personalized therapy accounting the pathogenic properties, and the rapid identification of new epidemiologically significant lineages at their earliest stages of spreading in the population. Moreover, the conserved nature of substitutions caused by dinucleotide mutations give hope for the development of virulence-targeted therapy aimed at these particular residues (Dehbanipour and Ghalavand, 2022).

It should be noted however, that there are few other alternative possible mechanisms explaining the observed mutational events elevation instead of natural selection objected by the interaction with host immune system (Allué-Guardia et al., 2021). While we could reject the 2N substitutions neutrality with high confidence due to the statistics, the nature of substitutions, and the absence of horizontal transfer, such nonrandom distributions could be caused by the coevolution with human population. Besides the adaptation to the particular subpopulations, the human densities and immigration could shape the genetics of the pathogen (Brites and Gagneux, 2012). The lineage specific homoplasy events could also be the result of HIV coinfection or variability of vaccination strategies used in different parts of the world (Comas et al., 2010; Brites and Gagneux, 2012).

Dinucleotide and other more complex mutation occurred in all branches of life, providing the additional diversity for survival and adaptation. To confirm the universality of the finding, and the possibility to apply the approach not only to clonal *M. tuberculosis*, we performed the same analysis for 1,164 genomes of *Neisseria gonorrhoeae* (supplementary data). In this set of genomes, 2N substitutions constituted 1.1% of all nucleotide substitutions. Similarly to the *M. tuberculosis* genomes, the fraction of genes with altered 2N substitutions took only a small part of the whole set of genes with high number of all point mutations, calculated from dN/dS statistics. Most frequently, mutations in genes, encoding monomers of the *Neisseria* type IV pili were identified. Type IV pili of *Neisseria* encoded by a highly variable genes set which diversity is driven by intergenome recombination, and is crucial for virulence. Pili are involved in surface adhesion, bacterial cell aggregation (Kennouche et al., 2019), they are recognized by specific antibodies, and alter host signalling (Criss et al., 2005). Thus, particular changes of specific residues are not unexpected, and are limited by its functional involvement (Kennouche et al., 2019).

In conclusion, the distinct distribution of dinucleotide mutations reveals unique insights into amino acid selection preferences. Genes exhibiting frequent 2N mutations are often pivotal for critical interactions with other molecules or proteins either pathogen, or host nature, playing a fundamental role in the pathogen's survival. The elevated occurrence of 2N mutations signifies diversifying selection, encompassing drug resistance mechanisms and factors associated with host invasion and survival. Exploring these interactions, virulence determinants or their counterparts, presents a promising avenue for innovative

therapeutic strategies (Chen et al., 2010). However, it is crucial to acknowledge that the substantial diversity and rapid evolution within these genes can pose challenges to the long-term effectiveness of such therapies. Therefore, an approach that focuses on more conservative regions of the identified targets might be preferable in drug development.

Data availability statement

The datasets analyzed and generated for this study and the developed scripts can be found in the Zenodo.org repository at <https://doi.org/10.5281/zenodo.15367741>.

Author contributions

DZ: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Visualization, Writing – original draft. AU: Data curation, Formal Analysis, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the Program of fundamental research for state academies for the period 2021–2030 (subprogram # FFEG-2024-0013/ 124032100002-1).

References

- Aguilar-Ayala, D. A., Cnockaert, M., André, E., Andries, K., Gonzalez-Y-Merchand, J. A., Vandamme, P., et al. (2017). *In vitro* activity of bedaquiline against rapidly growing nontuberculous mycobacteria. *J. Med. Microbiol.* 66 (8), 1140–1143. doi:10.1099/jmm.0.000537
- Alexander, D. C., Vasireddy, R., Vasireddy, S., Philley, J. V., Brown-Elliott, B. A., Perry, B. J., et al. (2017). Emergence of mmpT5 variants during bedaquiline treatment of *Mycobacterium intracellulare* lung disease. *J. Clin. Microbiol.* 55 (2), 574–584. doi:10.1128/JCM.02087-16
- Allué-Guardia, A., García, J. I., and Torrelles, J. B. (2021). Evolution of drug-resistant *Mycobacterium tuberculosis* strains and their adaptation to the human lung environment. *Front. Microbiol.* 12, 612675. doi:10.3389/fmicb.2021.612675
- Andries, K., Verhasselt, P., Guillemont, J., Göhlmann, H. W. H., Neefs, J. M., Winkler, H., et al. (2005). A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. *Science* 307 (5707), 223–227. doi:10.1126/science.1106753
- Andries, K., Villellas, C., Coeck, N., Thys, K., Gevers, T., Vranckx, L., et al. (2014). Acquired resistance of *Mycobacterium tuberculosis* to bedaquiline. *PLoS One* 9 (7), e102135. doi:10.1371/journal.pone.0102135
- Anisimova, M., and Kosiol, C. (2009). Investigating protein-coding sequence evolution with probabilistic codon substitution models. *Mol. Biol. Evol.* 26 (2), 255–271. doi:10.1093/molbev/msn232
- Antimycobacterial Susceptibility Testing Group (2022). Updating the approaches to define susceptibility and resistance to anti-tuberculosis agents: implications for diagnosis and treatment. *Eur. Respir. J.* 59 (4), 2200166. doi:10.1183/13993003.00166-2022
- Bazykin, G. A., Dushoff, J., Levin, S. A., and Kondrashov, A. S. (2006). Bursts of nonsynonymous substitutions in HIV-1 evolution reveal instances of positive selection at conservative protein sites. *Proc. Natl. Acad. Sci. U. S. A.* 103 (51), 19396–19401. doi:10.1073/pnas.0609484103
- Belinky, F., Sela, I., Rogozin, I. B., and Koonin, E. V. (2019). Crossing fitness valleys via double substitutions within codons. *BMC Biol.* 17 (1), 105. doi:10.1186/s12915-019-0727-4
- Brites, D., and Gagneux, S. (2012). Old and new selective pressures on *Mycobacterium tuberculosis*. *Infect. Genet. Evol.* 12 (4), 678–685. doi:10.1016/j.meegid.2011.08.010
- Brosch, R., Gordon, S. V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., et al. (2002). A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc. Natl. Acad. Sci. U. S. A.* 99 (6), 3684–3689. doi:10.1073/pnas.052548299
- Campbell, E. A., Korzhova, N., Mustae, A., Murakami, K., Nair, S., Goldfarb, A., et al. (2001). Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell* 104 (6), 901–912. doi:10.1016/s0092-8674(01)00286-0
- Cano, A. V., Rozhoňová, H., Stoltzfus, A., McCandlish, D. M., and Payne, J. L. (2022). Mutation bias shapes the spectrum of adaptive substitutions. *Proc. Natl. Acad. Sci. U. S. A.* 119 (7), e2119720119. doi:10.1073/pnas.2119720119
- Cebrián-Sastre, E., Chiner-Oms, A., Torres-Pérez, R., Comas, I., Oliveros, J. C., Blázquez, J., et al. (2023). Selective pressure by rifampicin modulates mutation rates and evolutionary trajectories of mycobacterial genomes. *Microbiol. Spectr.* 11 (4), e0101723. doi:10.1128/spectrum.01017-23
- Chen, J. M., Pojer, F., Blasco, B., and Cole, S. T. (2010). Towards anti-virulence drugs targeting ESX-1 mediated pathogenesis of *Mycobacterium tuberculosis*. *Drug Discov. Today Dis. Mech.* 7 (1), e25–e31. doi:10.1016/j.ddmec.2010.09.002
- Chen, P. E., and Shapiro, B. J. (2015). The advent of genome-wide association studies for bacteria. *Curr. Opin. Microbiol.* 25, 17–24. doi:10.1016/j.mib.2015.03.002
- Chesov, E., Chesov, D., Maurer, F. P., Andres, S., Utpatel, C., Barilar, I., et al. (2022). Emergence of bedaquiline resistance in a high tuberculosis burden country. *Eur. Respir. J.* 59 (3), 2100621. doi:10.1183/13993003.00621-2021
- Chiner-Oms, Á., López, M. G., Moreno-Molina, M., Furió, V., and Comas, I. (2022). Gene evolutionary trajectories in *Mycobacterium tuberculosis* reveal temporal signs of selection. *Proc. Natl. Acad. Sci. U. S. A.* 119 (17), e2113600119. doi:10.1073/pnas.2113600119
- Coll, F., McNerney, R., Guerra-Assunção, J. A., Glynn, J. R., Perdigão, J., Viveiros, M., et al. (2014). A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat. Commun.* 5, 4812. doi:10.1038/ncomms5812
- Comas, I., Chakravarti, J., Small, P. M., Galagan, J., Niemann, S., Kremer, K., et al. (2010). Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat. Genet.* 42 (6), 498–503. doi:10.1038/ng.590

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2025.1511372/full#supplementary-material>

- Conradie, F., Diacon, A. H., Ngubane, N., Howell, P., Everitt, D., Crook, A. M., et al. (2020). Treatment of highly drug-resistant pulmonary tuberculosis. *N. Engl. J. Med.* 382 (10), 893–902. doi:10.1056/NEJMoa1901814
- Criss, A. K., Kline, K. A., and Seifert, H. S. (2005). The frequency and rate of pilin antigenic variation in *Neisseria gonorrhoeae*. *Mol. Microbiol.* 58 (2), 510–519. doi:10.1111/j.1365-2958.2005.04838.x
- Dehbanipour, R., and Ghalavand, Z. (2022). Anti-virulence therapeutic strategies against bacterial infections: recent advances. *Germes* 12 (2), 262–275. doi:10.18683/germes.2022.1328
- Delpont, W., Scheffler, K., and Seoghe, C. (2009). Models of coding sequence evolution. *Brief. Bioinform* 10 (1), 97–109. doi:10.1093/bib/bbn049
- Domagk, G. (1950). Investigations on the antituberculous activity of the thiosemicarbazones *in vitro* and *in vivo*. *Am. Rev. Tuberc.* 61 (1), 8–19. doi:10.1164/art.1950.61.1.8
- Dubnau, E., Chan, J., Mohan, V. P., and Smith, I. (2005). Responses of mycobacterium tuberculosis to growth in the mouse lung. *Infect. Immun.* 73 (6), 3754–3757. doi:10.1128/IAI.73.6.3754-3757.2005
- Farhat, M. R., Shapiro, B. J., Kieser, K. J., Sultana, R., Jacobson, K. R., Victor, T. C., et al. (2013). Genomic analysis identifies targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis*. *Nat. Genet.* 45 (10), 1183–1189. doi:10.1038/ng.2747
- Fox, W., Ellard, G. A., and Mitchison, D. A. (1999). Studies on the treatment of tuberculosis undertaken by the British Medical Research Council tuberculosis units, 1946–1986, with relevant subsequent publications. *Int. J. Tuberc. Lung Dis.* 3 (10 Suppl. 2), S231–S279.
- Gagneux, S. (2018). Ecology and evolution of *Mycobacterium tuberculosis*. *Nat. Rev. Microbiol.* 16 (4), 202–213. doi:10.1038/nrmicro.2018.8
- Ghodousi, A., Hussain Rizvi, A., Khanzada, F. M., Akhtar, N., Ghafour, A., Trovato, A., et al. (2022). *In vivo* microevolution of *Mycobacterium tuberculosis* and transient emergence of *atpE*_Ala63Pro mutation during treatment in a pre-XDR TB patient. *Eur. Respir. J.* 59 (3), 2102102. doi:10.1183/13993003.02102-2021
- Gillespie, S. H. (2001). Antibiotic resistance in the absence of selective pressure. *Int. J. Antimicrob. Agents* 17 (3), 171–176. doi:10.1016/s0924-8579(00)00340-x
- Goldstein, B. P. (2014). Resistance to rifampicin: a review. *J. Antibiot. (Tokyo)* 67 (9), 625–630. doi:10.1038/ja.2014.107
- Guo, H., Courbon, G. M., Bueler, S. A., Mai, J., Liu, J., and Rubinstein, J. L. (2021). Structure of mycobacterial ATP synthase bound to the tuberculosis drug bedaquiline. *Nature* 589 (7840), 143–147. doi:10.1038/s41586-020-3004-3
- Hartkoorn, R. C., Uplekar, S., and Cole, S. T. (2014). Cross-resistance between clofazimine and bedaquiline through upregulation of *MmpL5* in *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* 58 (5), 2979–2981. doi:10.1128/AAC.00037-14
- Hicks, N. D., Carey, A. F., Yang, J., Zhao, Y., and Fortune, S. M. (2019). Bacterial genome-wide association identifies novel factors that contribute to ethionamide and prothionamide susceptibility in *Mycobacterium tuberculosis*. *mBio* 10 (2), e00616-19. doi:10.1128/mBio.00616-19
- Hillemann, D., Rüscher-Gerdes, S., and Richter, E. (2007). Evaluation of the GenoType MTBDRplus assay for rifampin and isoniazid susceptibility testing of *Mycobacterium tuberculosis* strains and clinical specimens. *J. Clin. Microbiol.* 45 (8), 2635–2640. doi:10.1128/JCM.00521-07
- Ismail, N., Ismail, N. A., Omar, S. V., and Peters, R. P. H. (2019). *In vitro* study of stepwise acquisition of *rv0678* and *atpE* mutations conferring bedaquiline resistance. *Antimicrob. Agents Chemother.* 63 (8), e00292-19. doi:10.1128/AAC.00292-19
- Jin, D. J., and Gross, C. A. (1988). Mapping and sequencing of mutations in the *Escherichia coli* *rpoB* gene that lead to rifampicin resistance. *J. Mol. Biol.* 202 (1), 45–58. doi:10.1016/0022-2836(88)90517-7
- Kennouche, P., Charles-Orszag, A., Nishiguchi, D., Goussard, S., Imhaus, A. F., Dupré, M., et al. (2019). Deep mutational scanning of the *Neisseria meningitidis* major pilin reveals the importance of pilus tip-mediated adhesion. *EMBO J.* 38 (22), e102145. doi:10.15252/embj.2019102145
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16 (2), 111–120. doi:10.1007/BF01731581
- Koch, A. S., Brites, D., Stucki, D., Evans, J. C., Seldon, R., Heekes, A., et al. (2017). The influence of HIV on the evolution of *Mycobacterium tuberculosis*. *Mol. Biol. Evol.* 34 (7), 1654–1668. doi:10.1093/molbev/msx107
- Köser, C. U., Feuerriegel, S., Summers, D. K., Archer, J. A. C., and Niemann, S. (2012). Importance of the genetic diversity within the *Mycobacterium tuberculosis* complex for the development of novel antibiotics and diagnostic tests of drug resistance. *Antimicrob. Agents Chemother.* 56 (12), 6080–6087. doi:10.1128/AAC.01641-12
- Kramnik, I., and Beamer, G. (2016). Mouse models of human TB pathology: roles in the analysis of necrosis and the development of host-directed therapies. *Semin. Immunopathol.* 38 (2), 221–237. doi:10.1007/s00281-015-0538-9
- Le Ray, L. F., Aubry, A., Sougakoff, W., Revest, M., Robert, J., Bonnet, I., et al. (2022). *atpE* mutation in *Mycobacterium tuberculosis* not always predictive of bedaquiline treatment failure. *Emerg. Infect. Dis.* 28 (5), 1062–1064. doi:10.3201/eid2805.212517
- Ley, S. D., de Vos, M., Van Rie, A., and Warren, R. M. (2019). Deciphering within-host microevolution of *Mycobacterium tuberculosis* through whole-genome sequencing: the phenotypic impact and way forward. *Microbiol. Mol. Biol. Rev.* 83 (2), e00062-18. doi:10.1128/MMBR.00062-18
- Lin, W., Mandal, S., Degen, D., Liu, Y., Ebright, Y. W., Li, S., et al. (2017). Structural basis of *Mycobacterium tuberculosis* transcription and transcription inhibition. *Mol. Cell* 66 (2), 169–179.e8. doi:10.1016/j.molcel.2017.03.001
- Liu, Q., Zhu, J., Dulberger, C. L., Stanley, S., Wilson, S., Chung, E. S., et al. (2022). Tuberculosis treatment failure associated with evolution of antibiotic resilience. *Science* 378 (6624), 1111–1118. doi:10.1126/science.abq2787
- Lu, Y., Chen, H., Shao, Z., Sun, L., Li, C., Lu, Y., et al. (2024). Deletion of the *Mycobacterium tuberculosis* *cyp138* gene leads to changes in membrane-related lipid composition and antibiotic susceptibility. *Front. Microbiol.* 15, 1301204. doi:10.3389/fmicb.2024.1301204
- Lucaci, A. G., Zehr, J. D., Enard, D., Thornton, J. W., and Kosakovsky Pond, S. L. (2023). Evolutionary shortcuts via multinucleotide substitutions and their impact on natural selection analyses. *Mol. Biol. Evol.* 40 (7), msad150. doi:10.1093/molbev/msad150
- Martinez, E., Hennessy, D., Jelfs, P., Crichton, T., Chen, S. C. A., and Sintchenko, V. (2018). Mutations associated with *in vitro* resistance to bedaquiline in *Mycobacterium tuberculosis* isolates in Australia. *Tuberc. (Edinb)* 111, 31–34. doi:10.1016/j.tube.2018.04.007
- Martini, B. A., Grigorov, A. S., Skvortsova, Y. V., Bychenko, O. S., Salina, E. G., and Azhikina, T. L. (2023). Small RNA MTS1338 configures a stress resistance signature in *Mycobacterium tuberculosis*. *Int. J. Mol. Sci.* 24 (9), 7928. doi:10.3390/ijms24097928
- Miotto, P., Cabibbe, A. M., Borroni, E., Degano, M., and Cirillo, D. M. (2018). Role of disputed mutations in the *rpoB* gene in interpretation of automated liquid MGIT culture results for rifampin susceptibility testing of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* 56 (5), e01599-17. doi:10.1128/JCM.01599-17
- Mokrousov, I., Akhmedova, G., Molchanov, V., Fundovnaya, E., Kozlova, E., Ostankova, Y., et al. (2021). Frequent acquisition of bedaquiline resistance by epidemic extensively drug-resistant *Mycobacterium tuberculosis* strains in Russia during long-term treatment. *Clin. Microbiol. Infect.* 27 (3), 478–480. doi:10.1016/j.cmi.2020.08.030
- Mokrousov, I., Chernyaeva, E., Vyazovaya, A., Skiba, Y., Solovieva, N., Valcheva, V., et al. (2018). Rapid assay for detection of the epidemiologically important central asian/Russian strain of the *Mycobacterium tuberculosis* Beijing genotype. *J. Clin. Microbiol.* 56 (2), e01551-17. doi:10.1128/JCM.01551-17
- Mokrousov, I., Narvskaya, O., Vyazovaya, A., Otten, T., Jiao, W. W., Gomes, L. L., et al. (2012). Russian “successful” clone B0/W148 of *Mycobacterium tuberculosis* Beijing genotype: a multiplex PCR assay for rapid detection and global screening. *J. Clin. Microbiol.* 50 (11), 3757–3759. doi:10.1128/JCM.02001-12
- Mokrousov, I., Vinogradova, T., Dogonadze, M., Zabolotnykh, N., Vyazovaya, A., Vitovskaya, M., et al. (2023). A multifaceted interplay between virulence, drug resistance, and the phylogeographic landscape of *Mycobacterium tuberculosis*. *Microbiol. Spectr.* 11 (5), e0139223. doi:10.1128/spectrum.01392-23
- Molodtsov, V., Scharf, N. T., Stefan, M. A., Garcia, G. A., and Murakami, K. S. (2017). Structural basis for rifamycin resistance of bacterial RNA polymerase by the three most clinically important *RpoB* mutations found in *Mycobacterium tuberculosis*. *Mol. Microbiol.* 103 (6), 1034–1045. doi:10.1111/nni.13606
- Moura de Sousa, J., Sousa, A., Bourgard, C., and Gordo, I. (2015). Potential for adaptation overrides cost of resistance. *Future Microbiol.* 10 (9), 1415–1431. doi:10.2217/fmb.15.61
- Nakamura, Y., Gojbori, T., and Ikemura, T. (2000). Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* 28 (1), 292. doi:10.1093/nar/28.1.292
- Namouchi, A., Didelot, X., Schöck, U., Gicquel, B., and Rocha, E. P. C. (2012). After the bottleneck: genome-wide diversification of the *Mycobacterium tuberculosis* complex by mutation, recombination, and natural selection. *Genome Res.* 22 (4), 721–734. doi:10.1101/gr.129544.111
- Nimmo, C., Millard, J., Faulkner, V., Monteserin, J., Pugh, H., and Johnson, E. O. (2022). Evolution of *Mycobacterium tuberculosis* drug resistance in the genomic era. *Front. Cell Infect. Microbiol.* 12, 954074. doi:10.3389/fcimb.2022.954074
- Ovchinnikov, Y. A., Monastyrskaya, G. S., Guriev, S. O., Kalinina, N. F., Sverdlov, E. D., Gragerov, A. I., et al. (1983). RNA polymerase rifampicin resistance mutations in *Escherichia coli*: sequence changes and dominance. *Mol. Gen. Genet.* 190 (2), 344–348. doi:10.1007/BF00330662
- Peng, Y., Li, C., Hui, X., Huo, X., Shumuyed, N. A., and Jia, Z. (2024). Phenotypic and genotypic analysis of drug resistance in *M. tuberculosis* isolates in Gansu, China. *PLoS One* 19 (9), e0311042. doi:10.1371/journal.pone.0311042
- Perdigão, J., and Portugal, I. (2019). Genetics and roadblocks of drug resistant tuberculosis. *Infect. Genet. Evol.* 72, 113–130. doi:10.1016/j.meegid.2018.09.023

- Peretokina, I. V., Krylova, L. Y., Antonova, O. V., Kholina, M. S., Kulagina, E. V., Nosova, E. Y., et al. (2020). Reduced susceptibility and resistance to bedaquiline in clinical *M. tuberculosis* isolates. *J. Infect.* 80 (5), 527–535. doi:10.1016/j.jinf.2020.01.007
- Petrella, S., Cambau, E., Chauffour, A., Andries, K., Jarlier, V., and Sougakoff, W. (2006). Genetic basis for natural and acquired resistance to the diarylquinoline R207910 in mycobacteria. *Antimicrob. Agents Chemother.* 50 (8), 2853–2856. doi:10.1128/AAC.00244-06
- Pholwat, S., Heysell, S., Stroup, S., Foongladda, S., and Houpt, E. (2011). Rapid first- and second-line drug susceptibility assay for *Mycobacterium tuberculosis* isolates by use of quantitative PCR. *J. Clin. Microbiol.* 49 (1), 69–75. doi:10.1128/JCM.01500-10
- Power, R. A., Parkhill, J., and de Oliveira, T. (2017). Microbial genome-wide association studies: lessons from human GWAS. *Nat. Rev. Genet.* 18 (1), 41–50. doi:10.1038/nrg.2016.132
- Rahman, S., Kosakovsky, P. S. L., Webb, A., and Hey, J. (2021). Weak selection on synonymous codons substantially inflates dN/dS estimates in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 118 (20), e2023575118. doi:10.1073/pnas.2023575118
- Rodriguez, A., Diehl, J. D., Wright, G. S., Bonar, C. D., Lundgren, T. J., Moss, M. J., et al. (2024). Synonymous codon substitutions modulate transcription and translation of a divergent upstream gene by modulating antisense RNA production. *Proc. Natl. Acad. Sci. U. S. A.* 121 (36), e2405510121. doi:10.1073/pnas.2405510121
- Rogozin, I. B., Malyarchuk, B. A., Pavlov, Y. I., and Milanesi, L. (2005). From context-dependence of mutations to molecular mechanisms of mutagenesis. *Pac Symp. Biocomput.* 409–420. doi:10.1142/978789812702456_0039
- Rosas-Magallanes, V., Deschavanne, P., Quintana-Murci, L., Brosch, R., Gicquel, B., and Neyrolles, O. (2006). Horizontal transfer of a virulence operon to the ancestor of *Mycobacterium tuberculosis*. *Mol. Biol. Evol.* 23 (6), 1129–1135. doi:10.1093/molbev/msj120
- Sandgren, A., Strong, M., Muthukrishnan, P., Weiner, B. K., Church, G. M., and Murray, M. B. (2009). Tuberculosis drug resistance mutation database. *PLoS Med.* 6 (2), e2. doi:10.1371/journal.pmed.1000002
- Schatz, A., Bugle, E., and Waksman, S. A. (1944). Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria. *Proc. Soc. Exp. Biol. Med.* 55, 66–69. doi:10.3181/00379727-55-14461
- Schrider, D. R., Hourmouzdi, J. N., and Hahn, M. W. (2011). Pervasive multinucleotide mutational events in eukaryotes. *Curr. Biol.* 21 (12), 1051–1054. doi:10.1016/j.cub.2011.05.013
- Schroeder, J. W., Yeesin, P., Simmons, L. A., and Wang, J. D. (2018). Sources of spontaneous mutagenesis in bacteria. *Crit. Rev. Biochem. Mol. Biol.* 53 (1), 29–48. doi:10.1080/10409238.2017.1394262
- Shang, Y., Chen, S., Shi, W., Nie, W., Jing, W., Huo, F., et al. (2023). Bedaquiline resistance pattern in clofazimine-resistant clinical isolates of tuberculosis patients. *J. Glob. Antimicrob. Resist.* 33, 294–300. doi:10.1016/j.jgar.2023.04.003
- Shapiro, B. J., David, L. A., Friedman, J., and Alm, E. J. (2009). Looking for Darwin's footprints in the microbial world. *Trends Microbiol.* 17 (5), 196–204. doi:10.1016/j.tim.2009.02.002
- Sola, C., Mokrousov, I., Sahal, M. R., La, K., Senelle, G., Guyeux, C., et al. (2024). Evolution, phylogenetics, and phylogeography of *Mycobacterium tuberculosis* complex. *Genet. Evol. Infect. Dis.*, 683–772. doi:10.1016/b978-0-443-28818-0.00025-2
- Solé, R. (2022). Revisiting leigh van valen's "A new evolutionary law" (1973). *Biol. Theory* 17 (2), 120–125. doi:10.1007/s13752-021-00391-w
- Stone, J. E., Lujan, S. A., Kunkel, T. A., and Kunkel, T. A. (2012). DNA polymerase zeta generates clustered mutations during bypass of endogenous DNA lesions in *Saccharomyces cerevisiae*. *Environ. Mol. Mutagen* 53 (9), 777–786. doi:10.1002/em.21728
- Tahaoglu, K., Törün, T., Sevim, T., Ataç, G., Kir, A., Karasulu, L., et al. (2001). The treatment of multidrug-resistant tuberculosis in Turkey. *N. Engl. J. Med.* 345 (3), 170–174. doi:10.1056/nejm200107193450303
- Telenti, A., Imboden, P., Marchesi, F., Lowrie, D., Cole, S., Colston, M. J., et al. (1993). Detection of rifampicin-resistance mutations in *Mycobacterium tuberculosis*. *Lancet* 341 (8846), 647–650. doi:10.1016/0140-6736(93)90417-f
- The CRyPTIC Consortium (2022a). A data compendium associating the genomes of 12,289 *Mycobacterium tuberculosis* isolates with quantitative resistance phenotypes to 13 antibiotics. *PLoS Biol.* 20 (8), e3001721. doi:10.1371/journal.pbio.3001721
- The CRyPTIC Consortium (2022b). Genome-wide association studies of global *Mycobacterium tuberculosis* resistance to 13 antimicrobials in 10,228 genomes identify new resistance mechanisms. *PLoS Biol.* 20 (8), e3001755. doi:10.1371/journal.pbio.3001755
- Thiltgen, G., Dos Reis, M., and Goldstein, R. A. (2017). Finding direction in the search for selection. *J. Mol. Evol.* 84 (1), 39–50. doi:10.1007/s00239-016-9765-5
- Timm, J., Bateson, A., Solanki, P., Paleckyte, A., Witney, A. A., Rofael, S. A. D., et al. (2023). Baseline and acquired resistance to bedaquiline, linezolid and pretomanid, and impact on treatment outcomes in four tuberculosis clinical trials containing pretomanid. *PLoS Glob. Public Health* 3 (10), e0002283. doi:10.1371/journal.pgph.0002283
- Trinh, V., Langelier, M. F., Archambault, J., and Coulombe, B. (2006). Structural perspective on mutations affecting the function of multisubunit RNA polymerases. *Microbiol. Mol. Biol. Rev.* 70 (1), 12–36. doi:10.1128/MMBR.70.1.12-36.2006
- Umpeleva, T., Chetverikova, E., Belyaev, D., Ereemeeva, N., Boteva, T., Golubeva, L., et al. (2024). Identification of genetic determinants of bedaquiline resistance in *Mycobacterium tuberculosis* in Ural region, Russia. *Microbiol. Spectr.* 12 (3), e0374923. doi:10.1128/spectrum.03749-23
- Ushtanit, A., Mikhailova, Y., Lyubimova, A., Makarova, M., Safonova, S., Filippov, A., et al. (2021). Genetic profile of linezolid-resistant *M. tuberculosis* clinical strains from moscow. *Antibiot. (Basel)* 10 (10), 1243. doi:10.3390/antibiotics10101243
- Van Deun, A., Decroo, T., Piubello, A., de Jong, B. C., Lynen, L., and Rieder, H. L. (2018). Principles for constructing a tuberculosis treatment regimen: the role and definition of core and companion drugs. *Int. J. Tuberc. Lung Dis.* 22 (3), 239–245. doi:10.5588/ijtld.17.0660
- Van Valen, L. (2014). "19 A new evolutionary law (1973)," in *Foundations of macroecology* (University of Chicago Press), 284–314.
- Venkat, A., Hahn, M. W., and Thornton, J. W. (2018). Multinucleotide mutations cause false inferences of lineage-specific positive selection. *Nat. Ecol. Evol.* 2 (8), 1280–1288. doi:10.1038/s41559-018-0584-5
- Veziris, N., Bonnet, I., Morel, F., Guglielmetti, L., Maitre, T., Fournier Le Ray, L., et al. (2021). Impact of the revised definition of extensively drug resistant tuberculosis. *Eur. Respir. J.* 58, 2100641. doi:10.1183/13993003.00641-2021
- Viney, K., Linh, N. N., Gegia, M., Zignol, M., Glaziou, P., Ismail, N., et al. (2021). New definitions of pre-extensively and extensively drug-resistant tuberculosis: update from the World Health Organization. *Eur. Respir. J.* 57 (4), 2100361. doi:10.1183/13993003.00361-2021
- Walker, T. M., Miotto, P., Köser, C. U., Fowler, P. W., Knaggs, J., Iqbal, Z., et al. (2022). The 2021 WHO catalogue of *Mycobacterium tuberculosis* complex mutations associated with drug resistance: a genotypic analysis. *Lancet Microbe* 3 (4), e265–e273. doi:10.1016/S2666-5247(21)00301-3
- Ward, S. K., Hoye, E. A., and Talaat, A. M. (2008). The global responses of *Mycobacterium tuberculosis* to physiological levels of copper. *J. Bacteriol.* 190 (8), 2939–2946. doi:10.1128/JB.01847-07
- Wilson, D. J., and Consortium, CRPTIC (2020). GenomeMap: within-species genome-wide dN/dS estimation from over 10,000 genomes. *Mol. Biol. Evol.* 37 (8), 2450–2460. doi:10.1093/molbev/msaa069
- World Health Organization (2021). *Technical report on critical concentrations for drug susceptibility testing of isoniazid and the rifamycins (rifampicin, rifabutin and rifapentine)*. Geneva: World Health Organization. Available online at: <https://apps.who.int/iris/handle/10665/339275>.
- World Health Organization (2022). *WHO consolidated guidelines on tuberculosis. Module 4: treatment - drug-susceptible tuberculosis treatment*. 1st ed. Geneva: World Health Organization, 1.
- World Health Organization (2023). *Catalogue of mutations in Mycobacterium tuberculosis complex and their association with drug resistance*. Second edition. World Health Organization. Available online at: <https://books.google.ru/books?id=VRz5EAAQBAJ>.
- Yang, Z. (2014). *Molecular evolution: a statistical approach*. 1st ed. Oxford (GB): Oxford University Press.
- Zhang, H., Li, D., Zhao, L., Fleming, J., Lin, N., Wang, T., et al. (2013). Genome sequencing of 161 *Mycobacterium tuberculosis* isolates from China identifies genes and intergenic regions associated with drug resistance. *Nat. Genet.* 45 (10), 1255–1260. doi:10.1038/ng.2735
- Zhu, C., Yang, T., Yin, J., Jiang, H., Takiff, H. E., Gao, Q., et al. (2023). The global success of *Mycobacterium tuberculosis* modern Beijing family is driven by a few recently emerged strains. *Microbiol. Spectr.* 11 (4), e0333922. doi:10.1128/spectrum.03339-22
- Zimenkov, D. V., Nosova, E. Y., Kulagina, E. V., Antonova, O. V., Arslanbaeva, L. R., Isakova, A. I., et al. (2017). Examination of bedaquiline- and linezolid-resistant *Mycobacterium tuberculosis* isolates from the Moscow region. *J. Antimicrob. Chemother.* 72 (7), 1901–1906. doi:10.1093/jac/dkx094