

Original Article

Identification of drug resistance-related virulence gene mutations in 667 clinical *Mycobacterium tuberculosis* isolatesYu Zhang¹, Xinchang Chen¹, Shiyong Wang¹, Ning Jiang², Lingyun Shao¹, Jiazhen Chen¹¹ Department of Infectious Diseases, Shanghai Key Laboratory of Infectious Diseases and Biosafety Emergency Response, National Medical Center for Infectious Diseases, Huashan Hospital, Shanghai Medical College, Fudan University, China² State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai, China**Abstract**

Introduction: Drug-resistant tuberculosis is a severe global public health threat. Virulence factors and antibiotic resistance are generally considered to play a significant role in bacterial pathogenesis. However, the interaction between resistance and virulence in *Mycobacterium tuberculosis* (MTB) remains unclear.

Methodology: Here, we used whole genome sequences from 667 MTB isolates from 14 countries to complete an *in silico* evaluation of the correlations between virulence gene mutations, drug resistance, and lineage classification. The chi-square (χ^2) test was used to determine whether specific virulence gene mutations and drug resistance were related.

Results: Our results showed that Mce1R_G171R and Pks15_V333A, were positively correlated with streptomycin and ethambutol resistance, respectively, and Pks15_T46I was correlated with isoniazid, rifampin, ethambutol, pyrazinamide and streptomycin resistance. We also identified an additional 24 and 40 single nucleotide polymorphisms as well as 6 and 2 insertions or deletions in various virulence genes that are likely to be associated with changes in drug susceptibility in L2 and L4, respectively.

Conclusions: Taken together our data suggest that there may be some degree of co-selection between virulence and resistance factors, which may help MTB more easily adapt to new environments.

Key words: *Mycobacterium tuberculosis*; virulence genes; drug resistance; correlation evaluations.

J Infect Dev Ctries 2024; 18(9):1404-1412. doi:10.3855/jidc.18081

(Received 14 February 2023 – Accepted 14 July 2023)

Copyright © 2024 Zhang *et al.* This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

Tuberculosis, an infectious disease caused by *Mycobacterium tuberculosis* (MTB), is a severe and growing global public health threat. The COVID-19 pandemic has had a continuous damaging impact on access to TB diagnosis and treatment and the burden of TB disease since 2019. According to the Global Tuberculosis Report for 2022, an estimated 10.6 million people fell ill with TB in 2021, an increase of 4.5% from 10.1 million in 2020. Globally, the estimated number of people who developed MDR-TB or RR-TB each year was relatively stable between 2015 and 2020 but grew in 2021. There were an estimated 450,000 incident cases in 2021, up 3.1% from 437,000 in 2020 [1]. MTB is an ancient causative agent and has infected a quarter of the global human population by developing highly sophisticated mechanisms of pathogenesis and immune evasion in its host [2].

Virulence and resistance have evolved over very different timescales. Changes in the virulence mechanisms of this bacteria have developed in response

to the coevolution of this pathogen and its host, with its pathogenic behavior changing to adapt to the evolving immune response of its host environment. However, the evolution and spread of antibiotic resistance is relatively recent having predominantly occurred over the past 50 years following the advent and widespread adoption of various antimicrobial agents. Despite the differences in the evolution of these processes, they share some common characteristics. Genetically encoded antibiotic resistance to some degree can be considered a subtype of virulence factor as it facilitates host pathogenesis, allowing persistent or chronic diseases [3]. Some characteristics including efflux pumps [4], porins [5], and cell wall alterations [6] are also involved in both virulence and resistance [7].

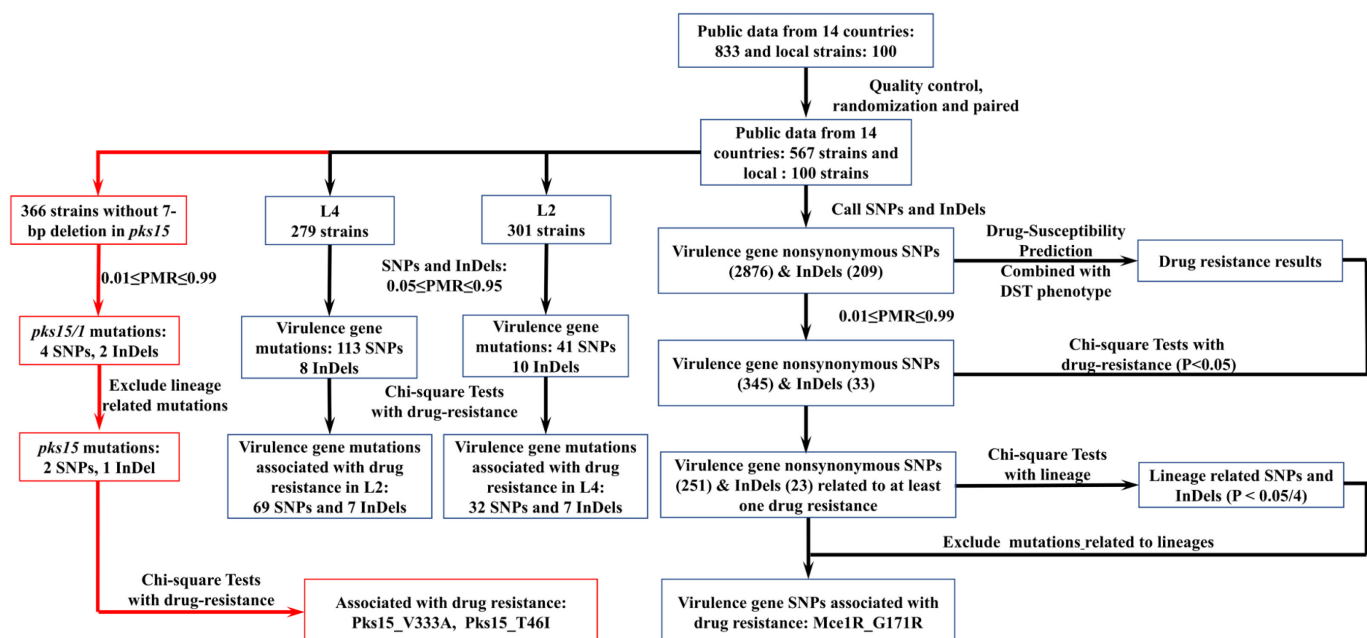
The evolution of these bacteria is the result of adaptation to the widespread, and often unnecessary and nonstandard application, of multiple antibiotics by the host and the variation in immune status caused by multiple chronic diseases, resulting in a series of new lineages of MTB. This is further complicated by the fact

that several of these endemic lineages are usually more drug-resistant and virulent than their common ancestor. As an example, B0/W-148, a rapid transmission subtype found in Russia, is characterized by increased virulence and drug resistance [8]. Many studies have shown that virulence factors are associated with antibiotic resistance in pathogenic bacteria [9,10]. The relationship between virulence genes and drug resistance in MTB remains unclear. It is reasonable to expect that enhanced virulence may be linked to drug resistance. For example, mutations in the mutator genes *mutT2* and *mutT4* in modern Beijing strains increase virulence and enhance drug resistance by improving bacterial adaptation to the host [11]. However, some resistance mutations are likely to compromise virulence, *e.g.*, the deletion of *KatG*, which confers high-level resistance to isoniazid but results in the attenuation of bacterial virulence *in vivo* [12]. Because the targets of anti-tuberculosis drugs often have important physiological functions, mutations in these targets may produce resistance with the cost of reducing fitness [13], which is manifested as slow growth and decreased transmission. Under these conditions, MTB may likely adopt a co-selection approach to allow for improved drug resistance with reduced fitness costs. In this case, we propose that there are some drug resistance-related mutations in the virulence genes, and identifying these mutations will help us better

understand the transmission and development of MDR-TB.

Differences in lipid metabolism and associated genes often contribute to lineage-specific virulence patterns in different MTBC lineages in both *in vitro* and *in vivo* infection models [14]. In particular, the “modern” sub-lineage is globally distributed and is closely associated with MDR-TB, extensively drug-resistant-TB, and hypervirulent TB [15]. Some virulence gene mutations are specific to high-level drug-resistant lineages and may act as potential lineage markers under these conditions. For example, C1881090T, located in virulence gene *pks7*, is considered to be the definitive marker of lineage 2.1 [16]. This study was designed to identify new single nucleotide polymorphisms (SNPs) and insertions and deletions (InDels) in virulence genes that correlate with changes in the drug resistance of their lineage after excluding the specific lineage markers. We accomplished this by analyzing the whole genome sequences of 667 clinical MTB strains from 14 countries worldwide and then evaluating the relationship between specific SNPs in specific virulence genes and drug resistance in each of these isolates. Our data provide important clues for better understanding the evolution of MDR-TB.

Figure 1. Whole workflow of analyzing correlation of virulence gene mutations and drug resistance. The red arrows show the workflow of analyzing correlation of mutations in *pks15* and *pks1* with each drug resistance in the 366 strains without 7-bp deletion (GGGCCGC) in *pks15*.



Methodology

Strain selection

Our study included hundreds of global MTB strains whose genome sequences and phenotypes were available in previous studies from a public database [17-21] or formed part of our laboratory collection. We first randomly included and downloaded 833 whole MTB genome sequences from China, India, Pakistan, South Africa, Congo, Swaziland, Ivory Coast, Belgium, the Netherlands, Italy, Russia, Canada, Peru, and Australia from the Sequence Read Archive (SRA) database and then completed a quality control process on these downloads to ensure that each download included enough reads of comparable quality and depth, at a minimum coverage of 60x. Fastp (v0.20.1, parameters: -q 15 -u 10 -l 90) (Figure 1) was used to exclude low-quality reads. After trimming, samples whose total bases were no fewer than 264,691,920bp (60 times that of MTB H37Rv genome size as reference) were used for further analysis. We then balanced the geographical distribution of our samples to better evaluate the correlations between drug resistance and virulence genes, leaving 30 MDR and 15 non-MDR strains from each country. In the case of countries like Canada, Ivory Coast, and the Netherlands, which did not have enough MDR strains, all the MDR strains meeting the inclusion criteria were enrolled and 15 non-MDR strains were randomly selected, and in the case of Congo-Kinshasa, all 6 MDR and 4 non-MDR strains from this region were included in our evaluations (Table 1).

We also included an additional 100 randomly selected clinical strains from our laboratory collection consisting of 71 MDR strains and 29 non-MDR strains, archived between 2014 and 2019, in this cohort. Thus, our final dataset included 667 clinical MTB strains from

14 countries, including 567 strains from a public database and 100 strains collected by our lab (Table 1).

Virulence gene selection

We evaluated the genetic polymorphisms in 332 virulence genes (Supplementary Table 1), including those encoding toxin-antitoxin (TA) systems, cell-wall proteins, lipid synthesis proteins, and type VII secretion systems and infection survival required genes [22-25], in an effort to establish the correlation between these mutations and drug resistance in our MTB strains.

DNA extraction, sequencing, and SNP calling

MTB colonies were scraped from the Löwenstein–Jensen medium and genomic DNA was extracted using the DNeasy Blood & Tissue Kit (QIAGEN, Valencia, CA, USA), following the manufacturer's instructions. All sequencing libraries were prepared using the Nextera XT Sample Prep Kit (Illumina, San Diego, CA, USA) as described by the manufacturer and sequenced on the Illumina Miseq, Illumina Hiseq, and BGISEQ-500 platforms., with at least a 60-fold coverage. Low-quality reads were then filtered out with fastp (v0.20.1, parameters: -q 15 -u 10 -l 90), and the remaining reads were aligned to MTB H37Rv reference genome (GenBank: NC_000962.3) using Bowtie2 (version 2.3.3.1) set to its default parameters. SNPs and InDels were then detected using SAMtools (version 1.6) with a minimum sequencing depth of 10 reads without strand bias and a frequency of $\geq 85\%$ for SNPs and $\geq 60\%$ for InDels. SNP genotyping was used to determine the lineage of each MTB strain as previously described [16].

Table 1. Country and lineage distribution for all 667 clinical isolates evaluated in this study.

Geography	MDR ^a	Non-MDR ^b	L1	L2	L3	L4	L5	L6
Australia	30	15	2	34	2	7	-	-
Belgium	30	15	1	18	-	26	-	-
Canada	25	15	8	22	2	8	-	-
China	30	15	-	37	1	7	-	-
Congo-Kinshasa	6	4	-	-	1	9	-	-
Italy	30	15	1	18	1	24	-	1
Ivory Coast	18	15	-	-	-	31	1	1
Netherlands	20	15	2	10	6	17	-	-
Pakistan	30	15	2	3	34	6	-	-
Peru	29	15	1	3	-	40	-	-
Russia	30	15	-	32	-	13	-	-
South Africa	30	15	3	7	-	35	-	-
Swaziland	30	15	10	5	2	28	-	-
Thailand	30	15	5	37	-	3	-	-
Our lab (China)	71	29	-	75	-	25	-	-
Total	439	228	35	301	49	279	1	2

a. Multi-drug resistant; b. Non-multi-drug resistant.

Drug-susceptibility testing (DST) and genotypic resistance by prediction

Phenotypic DST of all 100 strains from our lab was performed as described by the Clinical and Laboratory Standards Institute [26,27] which determined that the critical concentration for isoniazid (INH) was 0.2 mg/L; rifampin (RIF), 40.0 mg/L; ethambutol (EMB), 2.0 mg/L; levofloxacin (LFX), 2.0 mg/L; streptomycin (SM), 4.0 mg/L; amikacin (AM), 30.0 mg/L; kanamycin (KM), 30.0 mg/L; and capreomycin (CM), 40.0 mg/L. These results were all determined following 3 weeks of incubation at 37 °C. Apart from pyrazinamide (PZA), the DSTs were performed using the proportion method on Löwenstein–Jensen medium (Baso, Zhuhai, Guangzhou Province, China). The susceptibility of the MTB isolates to PZA was evaluated using an automated Mycobacterial Growth Indicator Tube 960 system (Becton Dickinson Diagnostic Systems, Franklin Lakes, NJ, USA), as described by the manufacturer, at a critical concentration of 100.0 mg/L.

We then evaluated the antimicrobial resistance of the 567 strains from the public database using their phenotypic resistance where available and genetic drug resistance prediction where necessary. Genetic resistance to INH, RIF, PZA, EMB, LFX, AM, KM, CM, and SM was predicted using whole genome sequencing data as previously described [26]. Collectively, previous studies showed that the genotype DST prediction performed well in the mentioned nine drugs with sensitivity higher than 85% and specificity higher than 80% [17,26].

Correlation analysis

In this study, chi-square tests were performed with (minimal case number < 5) or without (minimal case number ≥ 5) Yates's Correction for Continuity. The population mutation rate was determined as the proportion of mutated samples in total samples. SNPs and InDels with too low (< 1%) or too high (> 99%) population mutation rates were excluded from further analysis. We determined the correlation between virulence gene mutations and drug resistance, using the proportions of each SNP or InDel in each virulence gene from each population (drug-resistant/sensitive), where the average mutation rates were between 1% and 99%. These were then compared using a chi-square test, and the *p* value less than 0.05 was considered statistically significant. We then used the chi-square test by SPSS (v26.0) to determine the correlation between specific virulence gene mutations and lineage. In this study, Bonferroni Correction was conducted in Lineage

1 to 4 due to the small population sizes of Lineage 5 and 6. Thus, the *p*-value less than 0.0125 (0.05/4) was considered significant. Since it has been known that a 7-bp deletion (GGGCCGC) in *pks15* disrupted the fused *pks15/1* gene, leading to the abolition of synthesis of phenolic glycolipid (PGL), which is associated with resistance to intracellular killing by macrophage, we excluded the strains with a 7-bp deletion in *pks15* and then reanalyzed the correlation of mutations in *pks15* and *pks1* with each drug resistance in the remaining 366 strains without 7-bp deletion (Figure 1, red arrow part).

Since lineage-specific SNPs may interfere with the analysis of the virulence genes and the sample size of L2 (301 strains) and L4 (279 strains) was relatively large, we performed separate analyses for each of these 2 lineages. We used the chi-square test to evaluate the relationship between drug resistance and virulence gene mutations whose population mutation rates were between 5% and 95%. This whole workflow is described in Figure 1.

Homology models

The homology models for protein mce1R and *pks15*, as well as models for their mutations (Mce1R_G171R, Pks15_T46I, Pks15_V333A), were searched and built using SWISS-MODEL (<https://swissmodel.expasy.org/>) [28], and assessed by MolProbity (ver. 4.4) [29]. Distances between amino acid residues in protein conformations were calculated using UCSF ChimeraX (ver. 1.3, <https://www.cgl.ucsf.edu/chimerax/>) [30].

Results

Dataset characteristics

A total of 667 clinical MTB isolates were suitable for further evaluations with these samples representing strains from 14 countries across five continents, including 567 strains from the public database and 100 strains from our laboratory collection (Figure 2). There were 45 isolates each from Australia, Belgium, China (public data), Italy, Pakistan, Russia, Swaziland, South Africa, and Thailand; 40, 10, 33, 35, and 44 isolates from Canada, Congo-Kinshasa, Ivory Coast, the Netherlands, and Peru, respectively.

Drug resistance in our 667-strain cohort

For nine common antituberculosis drugs, namely, INH, RIF, EMB, PZA, LFX, SM, AM, KM, and CM, we completed drug-resistance testing for 100 local isolates and prediction for strains from a public database for which the drug sensitivity phenotype was not available. Among the 667 strains, 460, 455, 298,

Table 2. Drug susceptibility and resistance to common antibiotics for MTB.

	INH	RIF	EMB	PZA	LFX	SM	AM	KM	CM
R strains	460	455	298	268	125	305	48	60	49
S strains	207	212	369	399	542	362	619	607	618

INH: isoniazid; RIF: rifampin; EMB: ethambutol; PZA: pyrazinamide; LFX: levofloxacin; SM: streptomycin; AM: amikacin; KM: kanamycin; CM: capreomycin; R: resistant; S: susceptible.

268, 125, 305, 48, 60, and 49 strains were resistant (R) to INH, RIF, EMB, PZA, LFX, SM, AM, KM, and CM, respectively (Table 2). Accessions, locations, and drug susceptibility results of phenotype or genotype prediction of all 667 isolates were shown in Supplementary Table 2.

Virulence gene SNPs and InDels correlating with lineage type

We detected 2876 SNPs and 209 InDels in the virulence genes evaluated, and of these, 345 SNPs and 33 InDels had a mutation rate between 1% and 99%. Of the 345 SNPs, 126, 306, 71, and 294 SNPs correlated with lineages L1, L2, L3, and L4, respectively. Among these mutations, only Rv0165c(Mce1R)G171R showed no significant differences in distributions across lineages, whereas the remaining 344 SNPs and 33 InDels were correlated with specific lineages and therefore considered as lineage markers (Supplementary Table 3). For analysis of mutations on *pks15* and *pks1*, 7-bp deletion in *pks15* was found in all L4 strains and some of other lineages. We excluded the

strains with 7-bp deletion in *pks15*, and then analyzed the correlation of mutations in *pks15* and *pks1* with each lineage in the remaining 366 strains without the deletion. Among 4 SNPs and 2 InDels in *pks15*, 2 SNPs and 1 InDel were related to specific lineages (Supplementary Table 3).

Correlation between specific virulence gene mutations and drug resistance

We evaluated the correlation between the 345 nonsynonymous SNPs and 33 InDels with their specific lineages and associated drug resistance phenotype, which revealed that 251 of these SNPs and 23 of the InDels were directly correlated with the strains’ resistance of at least one drug (Supplementary Table 3). All of the InDels were excluded from these evaluations because they demonstrated significant differences in occurrence between the lineages, suggesting that they were lineage-specific mutations. After excluding the lineage-related SNPs, Mce1R_G171R was found to have a significant positive correlation with SM resistance phenotype (Table 3). For analysis of *pks15*

Figure 2. The global distribution of 667 clinical MTB isolates.

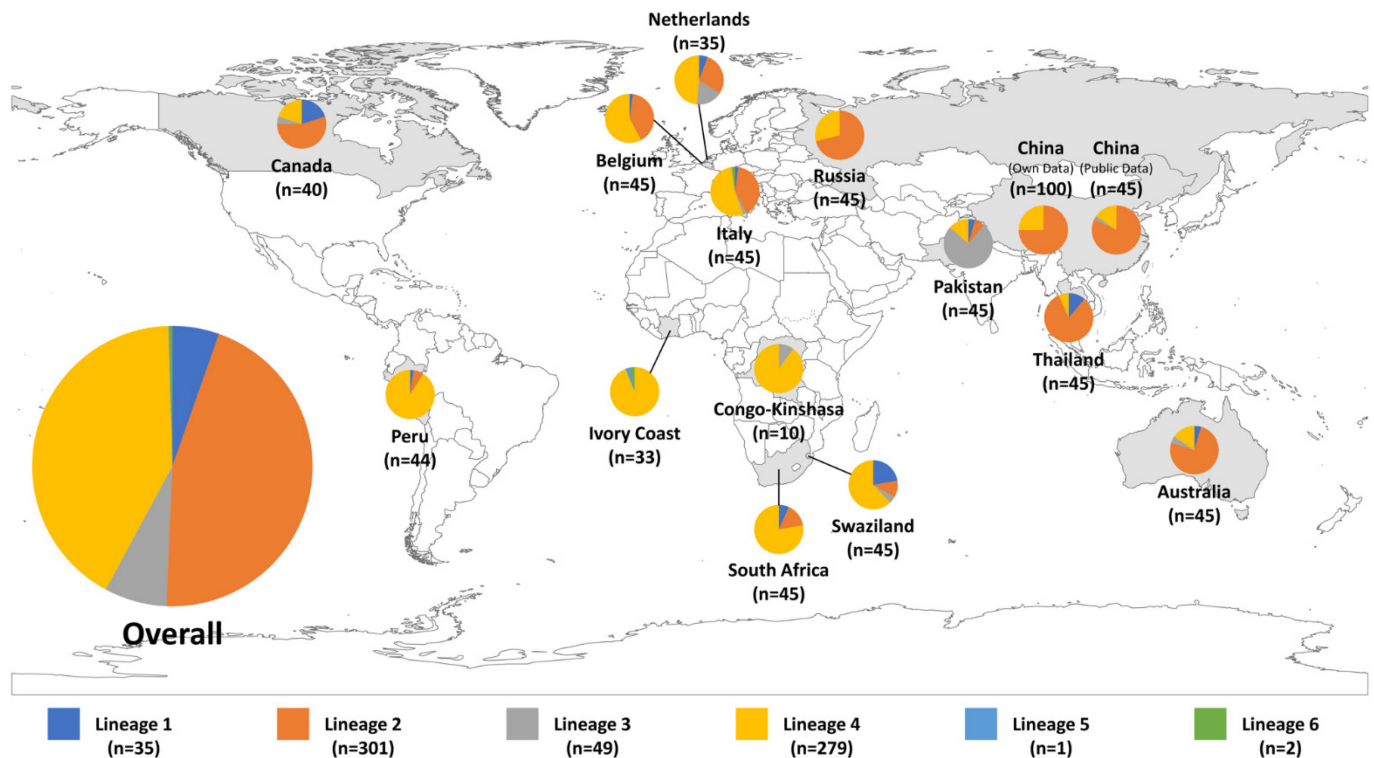


Table 3. Virulence gene mutation rates were significantly different in resistant strains.

Gene	Rv no.	Mutation	Gene function	Mutation rate	Drug	<i>p</i> value	P/N
<i>mce1R</i>	Rv0165c	G171R	Probable transcriptional regulatory protein Mce1R (GntR-family)	1.50%	SM	0.0120	P
<i>pks15^a</i>	Rv2947c	T46I	Putative inactive phenolphthiocerol synthesis polyketide synthase type I Pks15	4.37%	INH	0.0473	P
					RIF	0.0397	P
					PZA	0.0242	N
<i>pks15^a</i>	Rv2947c	V333A	Putative inactive phenolphthiocerol synthesis polyketide synthase type I Pks15	91.80%	SM	0.0035	P
					EMB	0.0331	P

INH: isoniazid; RIF: rifampin; EMB: ethambutol; PZA: pyrazinamide; SM: streptomycin. P/N, mutation is positively or negatively related to corresponding drug resistance. ^a Chi-square test was performed based on the population excluded strains with 7-bp deletion in *pks15*.

and *pks1*, after excluding the strains with 7-bp deletion in *pks15*, the correlation of mutations in *pks15* and *pks1* with each drug resistance was analyzed in the remaining 366 strains. After excluding the lineage-related mutations, the occurrence of Pks15_T46I was found to be associated with resistance of INH, RIF, PZA, and SM, and Pks15_V333A was associated with EMB resistance.

Correlation between virulence gene mutations and drug resistance in L2 and L4 strains

Strains from lineages L2 and L4 accounted for 45.1% and 41.8% of all of the samples, respectively. When we analyzed the correlation between virulence gene mutations and drug resistance in L2 and L4, 32 nonsynonymous SNPs, 7 InDels, and 69 nonsynonymous SNPs, 7 InDels were noted to show significant differences in resistance to at least one drug in L2 and L4, respectively (Supplementary Table 4,5).

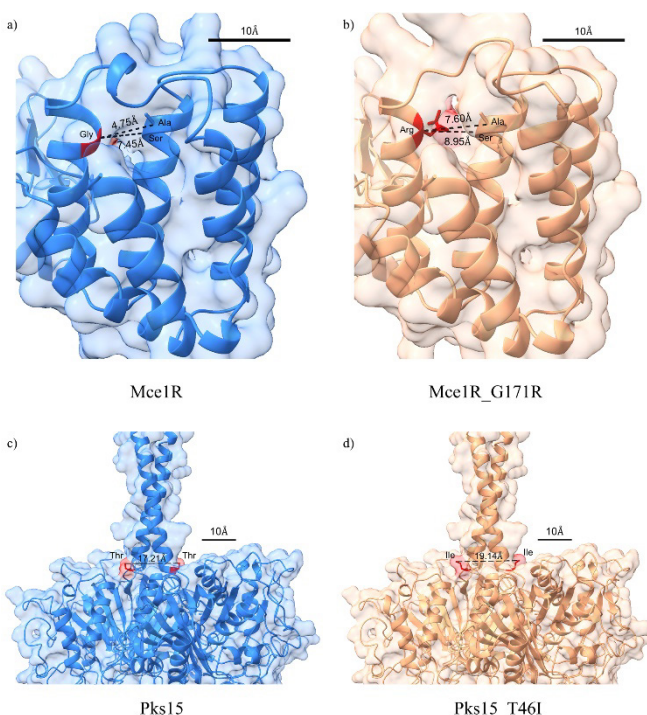
Protein homology models of Mce1R and Pks15

We identified the protein families and determined mutation-affected domains in both Mce1R (UniProtKB entry: Q79G00) and Pks15 (P96284) using EMBL-EBI InterPro (<https://www.ebi.ac.uk/interpro>). The Mce1R_G171R mutation is located within the C-terminal of GntR (InterPro entry: IPR011711), which is known to bind to effector molecules and regulate the transcription of its target genes through the action of its N-terminal DNA-binding domain. Pks15_T46I and Pks15_V333A are located in the N-terminal and C-terminal (IPR014031) of beta-ketoacyl-ACP synthase (EC: 2.3.1.41), which catalyzes the production of long-chain fatty acids.

Homology modelling results revealed very likely conformational changes at Mce1R_G171R and Pks15_T46I (Figure 3). Mce1R_G171R attributes to glycine to arginine substitution that results in the nonpolar side chain of Gly (-H) replaced by a larger and charged aliphatic side chain of Arg (- $(\text{CH}_2)_3$ -

NHC(NH)NH₂). As a part of an α -helix motif, 171Arg is pushed back by 140Ala and 141Ser which are on the opposite side of another α -helix causing an enlarged interval between two motifs and a significant change in its conformation (Figure 3a, b). Pks15_T46I is located at the end of an α -helix motif and next to a random coil. Due to the substitution of a non-polar side chain of Ile (-CH(CH₃)CH₂CH₃) for a smaller polar side chain of Thr (-C(OH)CH₃), the distance between two isoleucine residues of the Pks15_T46I homo-dimers increased leading to conformational changes of the α -helix motif (Figure 3c, d). Models of Pks_V333A substitution

Figure 3. Protein homology models of Mce1R and Pks15. Homology modelling results of a) Mce1R, b) Mce1R_G171R, c) Pks15, d) Pks15_T46I. a) and b) The interval between two α -helices was enlarged in Mce1R_G171R compared to the wild type; c) and d) the distance between two isoleucine residues of Pks15_T46I homo-dimers increased leading to the conformational change.



versus Pks15 did not reveal significant change between their conformations.

Discussion

MTB has not been shown to participate in any form of horizontal gene transfer meaning that it usually acquires MDR via the accumulation of different genetic mutations [31]. In addition, the increasing trend of drug resistance in MTB is a concern since it drastically limits the range of therapeutic alternatives for these infections. Thus, understanding the mechanism and transmission of drug-resistance MTB may be crucial to the development of novel therapeutic strategies. Here, we focused on the first-line anti-tuberculosis drugs (INH, RIF, EMB, PZA, and SM) and important second-line anti-tuberculosis drugs (LFX, AM, KM, and CM), and explored the relationship between virulence gene mutations and drug resistance. Our correlation analysis revealed that two mutations, Mce1R_G171R and Pks15_V333A, were positively correlated with SM and EMB resistance in MTB, respectively.

The *mce1R* gene encodes transcriptional regulatory protein Mce1R, which negatively regulates the *mce1* operon in intracellular MTB [32] that is duplicated four times within the genome (*mce1-4*) [33], because it encodes the ATP-binding cassette (ABC) transporter [34], known to help regulate immunopathological responses in the infected host and is important for lipid metabolism/transport, host invasion, modulation nicety [35]. A previous study revealed that an MTB mutant in which the expression of *mce1R* gene was disrupted *in vivo* induced an accelerated immunopathological response in the infected animal [36], while another study showed that the absence of the Mce1R protein increases MTB virulence in a mouse model [37]. Our finding, that Mce1R_G171R occurs more frequently in SM-resistant strains might correlate with greater virulence which is consistent with the results of a previous study. Homology modeling revealed Mce1R possesses a C-terminal domain comprising a bundle of six α -helices, which was consistent with previous research [35]. Interestingly, G171R of Mce1R is not only located in the α -helix bundle but also very close to a cavity formed by the bundle, which may bind a specific ligand with high affinity [35]. While G171R as well as 140Ala and 141Ser are so close to the residues constituting the cavity. The conformational change caused by G171R will likely lead to the change of the cavity and affect the affinity to the ligand.

The *pks15* gene encodes inactive phenolphthiocerol synthesis polyketide synthase type I and is involved in the biosynthesis of PGL [38]. When *pks15* is fused to

pks1 these isolates produce PGLs, while in strains like H37Rv with split *pks15* and *pks1* loci, PGL is not produced [39]. A 7-bp or 1-bp insertion causes a frameshift in the *pks15* taking H37Rv as a reference, resulting in an intact Pks15/1 with additional codons. Similar results have been shown in other W-Beijing strains, *M. africanum*, *M. bovis*, and in the other *M. tuberculosis* lineages such as EAI, and Delhi. The disruption of the *pks15/1* gene of *M. bovis* BCG abolishes the production of the *M. bovis*-specific PGL [34]. Which, depending on the host genetic background, can be associated with changes in the pro-inflammatory cytokine response in human macrophages [40]. A polyketide synthase-derived PGL produced by a subset of MTB isolates belonging to the W-Beijing family 8 demonstrates “hyperlethality” in murine disease models, while disruption of this PGL synthesis results in the loss of this hypervirulence [41]. Thus, we can reasonably assume that mutations in *pks15* may disturb its expression disrupting PGL synthesis and decreasing virulence.

We also found that 32 nonsynonymous SNPs and 7 InDels showed significant differences in at least one kind of drug resistance in L2 strains (Supplementary Table 4) and that the ins314A and ins726G mutation of *mce2B* were negatively correlated with at least two kinds of drugs resistance, suggesting that these mutations may occur more frequently in sensitive strains. Mce2B (Rv0590) is an *mce* family protein thought to be involved in host cell invasion [42]. Thus, we can infer that the ins314A and ins726G mutation is likely to disturb *mce2B* expression, which is in agreement with the results of a previous study, in which label-free comparative proteomics showed that *mce2B* was specifically upregulated in drug-resistant strains when compared to drug-sensitive strains [43].

Despite the value of our findings, there are a few limitations to this study that are worth mentioning. First, our samples include lineages 1–6, but they are not proportionately represented with L5 and L6 being significantly underrepresented in this dataset. Thus, we failed to analyze the correlation between virulence gene mutations and L5, L6, and other lineages not included in this study. Second, because of the limited accuracy and specificity of molecular resistance predictions, some second-line drugs such as MFX were not included in this study. Third, while a series of mutations were found to be independently correlated with drug resistance in L2 and L4 strains, these mutations might be sub-lineage markers and need further verification. Despite these limitations, this study suggested that there is some relationship between specific virulence gene

mutations and drug resistance in MTB which sheds new light on the importance of identifying the novel resistance mechanisms mediated by virulence genes.

Conclusions

We found some virulence gene mutations of which the occurrence has a certain association with some antibiotic resistance. Our data suggest that there may be some degree of co-selection between virulence and resistance factors, which may help MTB more easily adapt to new environments.

Acknowledgements

This study was supported by the National Science and Technology Major Project of China (2017ZX10302301-001) and the National Natural Science Foundation of China (81761148027).

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

- World Health Organization (2022) Global tuberculosis report 2022. Geneva: World Health Organization. Available: <https://www.who.int/publications/i/item/9789240061729>. Accessed: 14 February 2023.
- Orgeur M, Brosch R (2018) Evolution of virulence in the *Mycobacterium tuberculosis* complex. *Curr Opin Microbiol* 41: 68-75. doi: 10.1016/j.mib.2017.11.021.
- Emaneni M, Jabalameli F, Mirsalehian A, Ghasemi A, Beigverd R (2016) Characterization of virulence factors, antimicrobial resistance pattern and clonal complexes of group B streptococci isolated from neonates. *Microb Pathog* 99: 119-122. doi: 10.1016/j.micpath.2016.08.016.
- Barbosa TM, Levy SB (2000) Differential expression of over 60 chromosomal genes in *Escherichia coli* by constitutive expression of MarA. *J Bacteriol* 182: 3467-3474. doi: 10.1128/JB.182.12.3467-3474.2000.
- Tsai YK, Fung CP, Lin JC, Chen JH, Chang FY, Chen TL, Siu LK (2011) *Klebsiella pneumoniae* outer membrane porins OmpK35 and OmpK36 play roles in both antimicrobial resistance and virulence. *Antimicrob Agents Chemother* 55: 1485-1493. doi: 10.1128/AAC.01275-10.
- Moya B, Juan C, Alberti S, Pérez JL, Oliver A (2008) Benefit of having multiple *ampD* genes for acquiring beta-lactam resistance without losing fitness and virulence in *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* 52: 3694-3700. doi: 10.1128/AAC.00172-08.
- Seral C, Van Bambeke F, Tulkens PM (2003) Quantitative analysis of gentamicin, azithromycin, telithromycin, ciprofloxacin, moxifloxacin, and oritavancin (LY333328) activities against intracellular *Staphylococcus aureus* in mouse J774 macrophages. *Antimicrob Agents Chemother* 47: 2283-2292. doi: 10.1128/AAC.47.7.2283-2292.2003.
- Tsolaki AG, Gagneux S, Pym AS, Goguet de la Salmoniere Y-O, Kreiswirth BN, Van Soolingen D, Small PM (2005) Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis*. *J Clin Microbiol* 43: 3185-3191. doi: 10.1128/JCM.43.7.3185-3191.2005.
- Candan ED, Aksöz N (2015) *Klebsiella pneumoniae*: characteristics of carbapenem resistance and virulence factors. *Acta Biochim Pol* 62: 867-874. doi: 10.18388/abp.2015_1148.
- da Silva GJ, Mendonça N (2012) Association between antimicrobial resistance and virulence in *Escherichia coli*. *Virulence* 3: 18-28. doi: 10.4161/viru.3.1.18382.
- Ebrahimi-Rad M, Bifani P, Martin C, Kremer K, Samper S, Rauzier J, Kreiswirth B, Blazquez J, Jouan M, van Soolingen D, Gicquel B (2003) Mutations in putative mutator genes of *Mycobacterium tuberculosis* strains of the W-Beijing family. *Emerg Infect Dis* 9: 838-845. doi: 10.3201/eid0907.020803.
- Pym AS, Saint-Joanis B, Cole ST (2002) Effect of *katG* mutations on the virulence of *Mycobacterium tuberculosis* and the implication for transmission in humans. *Infect Immun* 70: 4955-4960. doi: 10.1128/IAI.70.9.4955-4960.2002.
- Melynk AH, Wong A, Kassen R (2015) The fitness costs of antibiotic resistance mutations. *Evol Appl* 83: 273-283. doi: 10.1111/eva.12196.
- Moopanar K, Mvubu NE (2020) Lineage-specific differences in lipid metabolism and its impact on clinical strains of *Mycobacterium tuberculosis*. *Microb Pathog* 146: 104250. doi: 10.1016/j.micpath.2020.104250.
- Napier G, Campino S, Merid Y, Abebe M, Woldeamanuel Y, Aseffa A, Hibberd ML, Phelan J, Clark TG (2020) Robust barcoding and identification of *Mycobacterium tuberculosis* lineages for epidemiological and clinical studies. *Genome Med* 12: 114. doi: 10.1186/s13073-020-00817-3.
- Coll F, McNeerney R, Guerra-Assunção JA, Glynn JR, Perdigão J, Viveiros M, Portugal I, Pain A, Martin N, Clark TG (2014) A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat Commun* 5: 4812. doi: 10.1038/ncomms5812.
- The CRyPTIC Consortium and the 100,000 Genomes Project (2018) Prediction of susceptibility to first-line tuberculosis drugs by DNA Sequencing. *N Engl J Med* 379: 1403-1415. doi: 10.1056/NEJMoa1800474.
- Bainomugisa A, Pandey S, Donnan E, Simpson G, Foster J, Lavu E, Hiasihri S, McBryde ES, Moke R, Vincent S, Sintchenko V, Marais BJ, Coin LJM, Coulter C (2019) Cross-border movement of highly drug-resistant *Mycobacterium tuberculosis* from Papua New Guinea to Australia through torres strait protected zone, 2010-2015. *Emerg Infect Dis* 25: 406-415. doi: 10.3201/eid2503.181003.
- Cohen KA, Abeel T, Manson McGuire A, Desjardins CA, Munsamy V, Shea TP, Walker BJ, Bantubani N, Almeida DV, Alvarado L, Chapman SB, Mvelase NR, Duffy EY, Fitzgerald MG, Govender P, Gujja S, Hamilton S, Howarth C, Larimer JD, Maharaj K, Pearson MD, Priest ME, Zeng Q, Padayatchi N, Grosset J, Young SK, Wortman J, Mlisana KP, O'Donnell MR, Birren BW, Bishai WR, Pym AS, Earl AM (2015) Evolution of extensively drug-resistant tuberculosis over four decades: whole genome sequencing and dating analysis of *Mycobacterium tuberculosis* isolates from KwaZulu-Natal. *PLoS Med* 12: e1001880. doi: 10.1371/journal.pmed.1001880.
- Guthrie JL, Delli Pizzi A, Roth D, Kong C, Jorgensen D, Rodrigues M, Tang P, Cook VJ, Johnston J, Gardy JL (2018) Genotyping and whole-genome sequencing to identify tuberculosis transmission to pediatric patients in British Columbia, Canada, 2005-2014. *J Infect Dis* 218: 1155-1163. doi: 10.1093/infdis/jiy278.
- Jabbar A, Phelan JE, de Sessions PF, Khan TA, Rahman H, Khan SN, Cantillon DM, Wildner LM, Ali S, Campino S, Waddell SJ, Clark TG (2019) Whole genome sequencing of

- drug resistant *Mycobacterium tuberculosis* isolates from a high burden tuberculosis region of North West Pakistan. *Sci Rep* 9: 14996. doi: 10.1038/s41598-019-51562-6.
22. Forrellad MA, Klepp LI, Gioffrè A, Sabio y García J, Morbidoni HR, de la Paz Santangelo M, Cataldi AA, Bigi F (2013) Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence* 4: 3-66. doi: 10.4161/viru.22329.
 23. Prozorov AA, Zaichikova MV, Danilenko VN (2013) Systems of genes and proteins affecting Mycobacteria virulence and their homologs participation in conjugation of *Mycobacterium smegmatis*. *Russ J Genet* 49: 110-125. doi: 10.7868/s0016675813010098.
 24. Prozorov AA, Fedorova IA, Bekker OB, Danilenko VN (2014) The virulence factors of *Mycobacterium tuberculosis*: genetic control, new conceptions. *Russ J Genet* 50: 775-797. doi: 10.1134/S1022795414080055.
 25. Sassetti CM, Rubin EJ (2003) Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A* 100: 12989-12994. doi: 10.1073/pnas.2134250100.
 26. Chen X, He G, Wang S, Lin S, Chen J, Zhang W (2019) Evaluation of whole-genome sequence method to diagnose resistance of 13 anti-tuberculosis drugs and characterize resistance genes in clinical multi-drug resistance *Mycobacterium tuberculosis* isolates from China. *Front Microbiol* 10: 1741. doi: 10.3389/fmicb.2019.01741.
 27. Woods GL, Brown-Elliott BA, Conville PS, Desmond EP, Hall GS, Lin G, Pfyffer GE, Ridderhof JC, Siddiqi SH, Wallace RJ, Warren NG, Witebsky FG (2011) Susceptibility testing of mycobacteria, nocardiae, and other aerobic actinomycetes, 2nd edition. Wayne (PA): Clinical and Laboratory Standards Institute. Available: <https://www.ncbi.nlm.nih.gov/books/NBK544374>. Accessed 14 February 2023.
 28. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 46: W296-W303. doi: 10.1093/nar/gky427.
 29. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66: 12-21. doi: 10.1107/S0907444909042073.
 30. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE (2021) UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* 30: 70-82. doi: 10.1002/pro.3943.
 31. Gygli SM, Borrell S, Trauner A, Gagneux S (2017) Antimicrobial resistance in *Mycobacterium tuberculosis*: mechanistic and evolutionary perspectives. *FEMS Microbiol Rev* 41: 354-373. doi: 10.1093/femsre/fux011.
 32. Casali N, White AM, Riley LW (2006) Regulation of the *Mycobacterium tuberculosis mce1* operon. *J Bacteriol* 188: 441-449. doi: 10.1128/JB.188.2.441-449.2006.
 33. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE, Tekaiia F, Badock 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 (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393: 537-544. doi: 10.1038/31159.
 34. Casali N, Riley LW (2007) A phylogenomic analysis of the *Actinomycetales mce* operons. *BMC Genomics* 8: 60. doi: 10.1186/1471-2164-8-60.
 35. Maity D, Katreddy RR, Bandhu A (2021) Molecular cloning, purification and characterization of Mce1R of *Mycobacterium tuberculosis*. *Mol Biotechnol* 63: 200-220. doi: 10.1007/s12033-020-00293-5.
 36. Uchida Y, Casali N, White A, Morici L, Kendall LV, Riley LW (2007) Accelerated immunopathological response of mice infected with *Mycobacterium tuberculosis* disrupted in the mce1 operon negative transcriptional regulator. *Cell Microbiol* 9: 1275-1283. doi: 10.1111/j.1462-5822.2006.00870.x.
 37. Cheigh CI, Senaratne R, Uchida Y, Casali N, Kendall LV, Riley LW (2010) Posttreatment reactivation of tuberculosis in mice caused by *Mycobacterium tuberculosis* disrupted in *mce1R*. *J Infect Dis* 202: 752-759. doi: 10.1086/655224.
 38. Constant P, Perez E, Malaga W, Lanéelle MA, Saurel O, Daffé M, Guilhot C (2002) Role of the *pks15/1* gene in the biosynthesis of phenolglycolipids in the *Mycobacterium tuberculosis* complex. Evidence that all strains synthesize glycosylated p-hydroxybenzoic methyl esters and that strains devoid of phenolglycolipids harbor a frameshift mutation in the *pks15/1* gene. *J Biol Chem* 27: 38148-38158. doi: 10.1074/jbc.M206538200.
 39. Pang JM, Layre E, Sweet L, Sherrid A, Moody DB, Ojha A, Sherman DR (2012) The polyketide PksI contributes to biofilm formation in *Mycobacterium tuberculosis*. *J Bacteriol* 194: 715-721. doi: 10.1128/JB.06304-11.
 40. Sinsimer D, Huet G, Manca C, Tsenova L, Koo MS, Kurepina N, Kana B, Mathema B, Marras SA, Kreiswirth BN, Guilhot C, Kaplan G (2008) The phenolic glycolipid of *Mycobacterium tuberculosis* differentially modulates the early host cytokine response but does not in itself confer hypervirulence. *Infect Immun* 76: 3027-3036. doi: 10.1128/IAI.01663-07.
 41. Reed MB, Domenech P, Manca C, Su H, Barczak AK, Kreiswirth BN, Kaplan G, Barry CE (2004) A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. *Nature* 431: 84-87. doi: 10.1038/nature02837.
 42. Sreenu VB, Kumar P, Nagaraju J, Nagarajaram HA (2006) Microsatellite polymorphism across the *M. tuberculosis* and *M. bovis* genomes: implications on genome evolution and plasticity. *BMC Genomics* 7: 78. doi: 10.1186/1471-2164-7-78.
 43. Ullah N, Hao L, Banga Ndzouboukou JL, Chen S, Wu Y, Li L, Borham Mohamed E, Hu Y, Fan X (2021) Label-free comparative proteomics of differentially expressed *Mycobacterium tuberculosis* protein in rifampicin-related drug-resistant strains. *Pathogens* 10: 607. doi: 10.3390/pathogens10050607.

Corresponding author

Jiazhen Chen
 Department of Infectious Diseases,
 Huashan Hospital, Fudan University,
 Rm 524, Building No.5, No.12 Middle Wulumuqi Rd.,
 Shanghai 200040, China.
 Phone: 86-21-52887961
 Fax: 86-21-62489015
 Email: jiazhen_chen@163.com

Conflict of interests: No conflict of interests is declared.

Annex – Supplementary Items**Supplementary Table 1.** 332 virulence genes included in this study.

Locus	gene name	Locus	gene name	Locus	gene name
Rv0015c	<i>pknA</i>	Rv0587	<i>yrbE2A</i>	Rv1185c	<i>fadD21</i>
Rv0098	<i>Rv0098</i>	Rv0588	<i>yrbE2B</i>	Rv1192	<i>Rv1192</i>
Rv0099	<i>fadD10</i>	Rv0589	<i>mce2A</i>	Rv1204c	<i>Rv1204c</i>
Rv0100	<i>Rv0100</i>	Rv0590	<i>mce2B</i>	Rv1211	<i>Rv1211</i>
Rv0101	<i>nrp</i>	Rv0591	<i>mce2C</i>	Rv1221	<i>sigE</i>
Rv0126	<i>treS</i>	Rv0592	<i>mce2D</i>	Rv1224	<i>Rv1224</i>
Rv0153c	<i>ptbB</i>	Rv0593	<i>lprL</i>	Rv1235	<i>lpqY</i>
Rv0165c	<i>mce1R</i>	Rv0594	<i>mce2F</i>	Rv1236	<i>sugA</i>
Rv0167	<i>yrbE1A</i>	Rv0636	<i>Rv0636</i>	Rv1237	<i>sugB</i>
Rv0168	<i>yrbE1B</i>	Rv0642c	<i>mmaA4</i>	Rv1238	<i>sugC</i>
Rv0169	<i>mce1A</i>	Rv0643c	<i>mmaA3</i>	Rv1244	<i>lpqZ</i>
Rv0170	<i>mce1B</i>	Rv0655	<i>Rv0655</i>	Rv1272c	<i>Rv1272c</i>
Rv0171	<i>mce1C</i>	Rv0670	<i>end</i>	Rv1284	<i>Rv1284</i>
Rv0172	<i>mce1D</i>	Rv0687	<i>Rv0687</i>	Rv1304	<i>atpB</i>
Rv0173	<i>lprK</i>	Rv0735	<i>sigL</i>	Rv1323	<i>fadA4</i>
Rv0174	<i>mce1F</i>	Rv0757	<i>phoP</i>	Rv1332	<i>Rv1332</i>
Rv0175	<i>Rv0175</i>	Rv0758	<i>phoR</i>	Rv1333	<i>Rv1333</i>
Rv0176	<i>Rv0176</i>	Rv0820	<i>phoT</i>	Rv1338	<i>murI</i>
Rv0182c	<i>sigG</i>	Rv0821c	<i>phoY2</i>	Rv1345	<i>mbtM</i>
Rv0198c	<i>zmp1</i>	Rv0899	<i>ompA</i>	Rv1348	<i>irtA</i>
Rv0199	<i>Rv0199</i>	Rv0930	<i>pstA1</i>	Rv1349	<i>irtB</i>
Rv0204c	<i>Rv0204c</i>	Rv0931c	<i>pknD</i>	Rv1371	<i>Rv1371</i>
Rv0211	<i>pckA</i>	Rv0934	<i>pstS1</i>	Rv1405c	<i>Rv1405c</i>
Rv0216	<i>Rv0216</i>	Rv0950c	<i>Rv0950c</i>	Rv1410c	<i>P55</i>
Rv0218	<i>Rv0218</i>	Rv0969	<i>ctpV</i>	Rv1411c	<i>lprG</i>
Rv0249c	<i>Rv0249c</i>	Rv0981	<i>mprA</i>	Rv1422	<i>Rv1422</i>
Rv0326	<i>Rv0326</i>	Rv0982	<i>mprB</i>	Rv1460	<i>Rv1460</i>
Rv0348	<i>Rv0348</i>	Rv0983	<i>pepD</i>	Rv1465	<i>Rv1465</i>
Rv0353	<i>hspR</i>	Rv0990c	<i>Rv0990c</i>	Rv1469	<i>ctpD</i>
Rv0410c	<i>pknG</i>	Rv1013	<i>pks16</i>	Rv1514c	<i>Rv1514c</i>
Rv0414c	<i>thiE</i>	Rv1016c	<i>lpqT</i>	Rv1524	<i>Rv1524</i>
Rv0427c	<i>xthA</i>	Rv1021	<i>Rv1021</i>	Rv1527c	<i>pks5</i>
Rv0432	<i>sodC</i>	Rv1028c	<i>kdpD</i>	Rv1539	<i>lspA</i>
Rv0450c	<i>mmpL4</i>	Rv1092c	<i>coaA</i>	Rv1560	<i>Rv1560</i>
Rv0467	<i>icl1</i>	Rv1099c	<i>Rv1099c</i>	Rv1568	<i>bioA</i>
Rv0470c	<i>pcaA</i>	Rv1109c	<i>Rv1109c</i>	Rv1569	<i>bioF</i>
Rv0475	<i>hbhA</i>	Rv1111c	<i>Rv1111c</i>	Rv1589	<i>bioB</i>
Rv0490	<i>senX3</i>	Rv1128c	<i>Rv1128c</i>	Rv1590	<i>Rv1590</i>
Rv0491	<i>regX3</i>	Rv1144	<i>Rv1144</i>	Rv1640c	<i>lysX</i>
Rv0566c	<i>Rv0566c</i>	Rv1183	<i>mmpL10</i>	Rv1653	<i>argJ</i>
Rv0586	<i>mce2R</i>	Rv1184c	<i>Rv1184c</i>	Rv1660	<i>pks10</i>
Rv1661	<i>pks7</i>	Rv2115c	<i>mpa</i>	Rv2869c	<i>rip</i>
Rv1696	<i>recN</i>	Rv2136c	<i>Rv2136c</i>	Rv2869c	<i>rip</i>
Rv1710	<i>Rv1710</i>	Rv2200c	<i>ctaC</i>	Rv2885c	<i>Rv2885c</i>
Rv1743	<i>pknE</i>	Rv2211c	<i>gcvT</i>	Rv2912c	<i>Rv2912c</i>
Rv1795	<i>eccD5</i>	Rv2224c	<i>caeA</i>	Rv2921c	<i>fisY</i>
Rv1807	<i>Rv1807</i>	Rv2231c	<i>cobC</i>	Rv2930	<i>fadD26</i>
Rv1811	<i>mgcC</i>	Rv2234	<i>ptpA</i>	Rv2936	<i>drdA</i>
Rv1821	<i>secA2</i>	Rv2234	<i>ptpA</i>	Rv2937	<i>drdB</i>
Rv1857	<i>modA</i>	Rv2241	<i>aceE</i>	Rv2938	<i>drdC</i>
Rv1915	<i>aceAa</i>	Rv2246	<i>kasB</i>	Rv2941	<i>fadD28</i>
Rv1916	<i>aceAb</i>	Rv2275	<i>Rv2275</i>	Rv2942	<i>mmpL7</i>
Rv1930c	<i>Rv1930c</i>	Rv2277c	<i>Rv2277c</i>	Rv2945c	<i>lppX</i>
Rv1931c	<i>Rv1931c</i>	Rv2335	<i>cysE</i>	Rv2946c	<i>pks1</i>
Rv1932	<i>tpx</i>	Rv2349c	<i>plcC</i>	Rv2947c	<i>pks15</i>
Rv1936	<i>Rv1936</i>	Rv2350c	<i>plcB</i>	Rv2976c	<i>ung</i>
Rv1937	<i>Rv1937</i>	Rv2351c	<i>plcA</i>	Rv2981c	<i>ddlA</i>
Rv1938	<i>ephB</i>	Rv2359	<i>furB</i>	Rv2998	<i>Rv2998</i>
Rv1939	<i>Rv1939</i>	Rv2374c	<i>hrcA</i>	Rv3042c	<i>serB2</i>
Rv1963c	<i>mce3R</i>	Rv2383c	<i>mbtB</i>	Rv3050c	<i>Rv3050c</i>
Rv1964	<i>yrbE3A</i>	Rv2387	<i>Rv2387</i>	Rv3061c	<i>fadE22</i>
Rv1965	<i>yrbE3B</i>	Rv2388c	<i>hemN</i>	Rv3082c	<i>virS</i>
Rv1966	<i>mce3A</i>	Rv2391	<i>nirA</i>	Rv3083	<i>Rv3083</i>
Rv1967	<i>mce3B</i>	Rv2395	<i>Rv2395</i>	Rv3084	<i>lipR</i>
Rv1968	<i>mce3C</i>	Rv2396	<i>PE_PGERS41</i>	Rv3085	<i>Rv3085</i>
Rv1969	<i>mce3D</i>	Rv2428	<i>ahpC</i>	Rv3086	<i>adhD</i>
Rv1970	<i>lprM</i>	Rv2437	<i>Rv2437</i>	Rv3087	<i>Rv3087</i>
Rv1971	<i>mce3F</i>	Rv2445c	<i>ndkA</i>	Rv3088	<i>tgS4</i>
Rv1974	<i>Rv1974</i>	Rv2472	<i>Rv2472</i>	Rv3089	<i>fadD13</i>
Rv1979c	<i>Rv1979c</i>	Rv2483c	<i>Rv2483c</i>	Rv3103c	<i>Rv3103c</i>
Rv1980c	<i>mpt64</i>	Rv2502c	<i>accD1</i>	Rv3114	<i>Rv3114</i>
Rv1981c	<i>nrdF1</i>	Rv2553c	<i>Rv2553c</i>	Rv3132c	<i>devS</i>
Rv2004c	<i>Rv2004c</i>	Rv2692	<i>trkB</i>	Rv3133c	<i>devR</i>
Rv2027c	<i>dosT</i>	Rv2696c	<i>Rv2696c</i>	Rv3151	<i>nuoG</i>
Rv2031c	<i>hspX</i>	Rv2702	<i>ppgK</i>	Rv3168	<i>Rv3168</i>
Rv2032	<i>acg</i>	Rv2703	<i>sigA</i>	Rv3178	<i>Rv3178</i>

Rv2038c	<i>Rv2038c</i>	Rv2707	<i>Rv2707</i>	Rv3210c	<i>Rv3210c</i>
Rv2040c	<i>Rv2040c</i>	Rv2711	<i>ideR</i>	Rv3223c	<i>sigH</i>
Rv2048c	<i>pks12</i>	Rv2734	<i>Rv2734</i>	Rv3229c	<i>desA3</i>
Rv2051c	<i>ppm1</i>	Rv2745c	<i>clgR</i>	Rv3236c	<i>kefB</i>
Rv2063A	<i>mazF7</i>	Rv2808	<i>Rv2808</i>	Rv3246c	<i>mtrA</i>
Rv2069	<i>sigC</i>	Rv2813	<i>Rv2813</i>	Rv3258c	<i>Rv3258c</i>
Rv2072c	<i>cobL</i>	Rv2845c	<i>proS</i>	Rv3270	<i>ctpC</i>
Rv2097c	<i>pafA</i>	Rv2857c	<i>Rv2857c</i>	Rv3277	<i>Rv3277</i>
Rv3286c	<i>sigF</i>	Rv3542c	<i>Rv3542c</i>	Rv3758c	<i>proV</i>
Rv3310	<i>sapM</i>	Rv3543c	<i>fadE29</i>	Rv3763	<i>lpqH</i>
Rv3335c	<i>Rv3335c</i>	Rv3544c	<i>fadE28</i>	Rv3781	<i>Rv3781</i>
Rv3371	<i>Rv3371</i>	Rv3545c	<i>cyp125</i>	Rv3794	<i>embA</i>
Rv3375	<i>amiD</i>	Rv3551	<i>Rv3551</i>	Rv3804c	<i>fbpA</i>
Rv3400	<i>Rv3400</i>	Rv3556c	<i>fadA6</i>	Rv3805c	<i>Rv3805c</i>
Rv3409c	<i>choD</i>	Rv3560c	<i>fadE30</i>	Rv3810	<i>pirG</i>
Rv3414c	<i>sigD</i>	Rv3563	<i>fadE32</i>	Rv3823c	<i>mmpL8</i>
Rv3416	<i>whiB3</i>	Rv3568c	<i>hsaC</i>	Rv3849	<i>espR</i>
Rv3419c	<i>gcp</i>	Rv3574	<i>Rv3574</i>	Rv3855	<i>Rv3855</i>
Rv3472	<i>Rv3472</i>	Rv3588c	<i>Rv3588c</i>	Rv3864	<i>Rv3864</i>
Rv3484	<i>cpsA</i>	Rv3614c	<i>espD</i>	Rv3866	<i>espG1</i>
Rv3489	<i>Rv3489</i>	Rv3615c	<i>espC</i>	Rv3867	<i>espH</i>
Rv3494c	<i>mce4F</i>	Rv3616c	<i>espA</i>	Rv3868	<i>eccA1</i>
Rv3495c	<i>lprN</i>	Rv3631	<i>Rv3631</i>	Rv3869	<i>eccB1</i>
Rv3496c	<i>mce4D</i>	Rv3649	<i>Rv3649</i>	Rv3870	<i>eccCa1</i>
Rv3497c	<i>mce4C</i>	Rv3651	<i>Rv3651</i>	Rv3871	<i>eccCb1</i>
Rv3498c	<i>mce4B</i>	Rv3663c	<i>dppD</i>	Rv3872	<i>Rv3872</i>
Rv3499c	<i>mce4A</i>	Rv3664c	<i>dppC</i>	Rv3873	<i>Rv3873</i>
Rv3500c	<i>yrbE4B</i>	Rv3665c	<i>dppB</i>	Rv3874	<i>esxB</i>
Rv3501c	<i>yrbE4A</i>	Rv3666c	<i>dppA</i>	Rv3875	<i>esxA</i>
Rv3502c	<i>Rv3502c</i>	Rv3671c	<i>Rv3671c</i>	Rv3876	<i>Rv3876</i>
Rv3519	<i>Rv3519</i>	Rv3682	<i>ponA2</i>	Rv3877	<i>eccD1</i>
Rv3523	<i>Rv3523</i>	Rv3683	<i>Rv3683</i>	Rv3882c	<i>Rv3882c</i>
Rv3534c	<i>Rv3534c</i>	Rv3701c	<i>Rv3701c</i>	Rv3883c	<i>mycP1</i>
Rv3540c	<i>ltp2</i>	Rv3717	<i>Rv3717</i>	Rv3910	<i>Rv3910</i>
Rv3541c	<i>Rv3541c</i>	Rv3723	<i>Rv3723</i>		

Table with 13 columns: ID, Action, Source, Country, Resistance, Lineage, and 7 binary variables (S/R/S/S/R/S/S). Rows include various SAMN and SRR identifiers for laboratory collections and public databases across different countries like China and Congo-Kinshasa.

SRR7517770	SRR7517770	Public database	Peru	MDR	Lineage 4	R	R	S	R	S	S	S	S	S
SRR7517772	SRR7517772	Public database	Peru	MDR	Lineage 4	R	R	S	R	S	S	S	S	S
SRR7517786	SRR7517786	Public database	Peru	MDR	Lineage 4	R	R	S	S	S	R	S	S	S
SRR7517787	SRR7517787	Public database	Peru	MDR	Lineage 4	R	R	S	S	S	R	S	S	S
SRR7517809	SRR7517809	Public database	Peru	MDR	Lineage 4	R	R	S	S	S	S	S	S	S
SRR7517810	SRR7517810	Public database	Peru	MDR	Lineage 4	R	R	S	R	S	S	S	S	S
SRR7517811	SRR7517811	Public database	Peru	non-MDR	Lineage 4	S	S	S	S	S	S	S	S	S
SRR7517813	SRR7517813	Public database	Peru	MDR	Lineage 4	R	R	R	R	S	S	S	S	S
SRR7517814	SRR7517814	Public database	Peru	MDR	Lineage 4	R	R	R	S	S	S	S	S	S
SRR7517815	SRR7517815	Public database	Peru	MDR	Lineage 4	R	R	R	R	S	S	S	S	S
SRR7517816	SRR7517816	Public database	Peru	MDR	Lineage 4	S	R	S	R	S	S	S	S	S
SRR7517831	SRR7517831	Public database	Peru	MDR	Lineage 4	R	R	S	R	S	S	S	S	S
SRR7517834	SRR7517834	Public database	Peru	MDR	Lineage 4	R	R	R	R	S	S	S	S	S
SRR7517835	SRR7517835	Public database	Peru	non-MDR	Lineage 4	S	S	S	S	S	S	S	S	S
SRR7517836	SRR7517836	Public database	Peru	MDR	Lineage 4	S	R	S	R	S	S	S	S	S
ERR067585	ERR067585	Public database	Russia	non-MDR	Lineage 4	S	S	S	S	S	S	S	S	S
ERR067590	ERR067590	Public database	Russia	MDR	Lineage 2	R	R	R	R	S	R	S	S	S
ERR108425	ERR108425	Public database	Russia	MDR	Lineage 4	R	R	S	S	S	R	S	S	S
ERR108441	ERR108441	Public database	Russia	non-MDR	Lineage 2	S	S	S	S	S	S	S	S	S
ERR108447	ERR108447	Public database	Russia	non-MDR	Lineage 2	R	S	S	S	S	R	S	S	S
ERR108458	ERR108458	Public database	Russia	MDR	Lineage 2	R	R	S	S	S	R	S	S	S
ERR108482	ERR108482	Public database	Russia	MDR	Lineage 4	R	R	S	S	S	R	S	S	S
ERR108489	ERR108489	Public database	Russia	MDR	Lineage 2	R	R	S	S	S	R	S	S	S
ERR117454	ERR117454	Public database	Russia	MDR	Lineage 2	R	R	S	R	S	R	R	R	R
ERR117457	ERR117457	Public database	Russia	non-MDR	Lineage 2	S	S	S	S	S	S	S	S	S
ERR133819	ERR133819	Public database	Russia	non-MDR	Lineage 4	S	S	S	S	S	S	S	S	S
ERR133846	ERR133846	Public database	Russia	non-MDR	Lineage 4	S	S	S	S	S	S	S	S	S
ERR133888	ERR133888	Public database	Russia	MDR	Lineage 2	R	R	R	S	S	R	S	S	S
ERR133913	ERR133913	Public database	Russia	non-MDR	Lineage 2	R	S	S	S	S	R	S	S	S
ERR133967	ERR133967	Public database	Russia	non-MDR	Lineage 4	S	S	S	S	S	S	S	S	S
ERR133981	ERR133981	Public database	Russia	MDR	Lineage 4	R	R	R	S	S	S	S	S	S
ERR137192	ERR137192	Public database	Russia	MDR	Lineage 2	R	R	R	R	S	R	S	S	S
ERR137204	ERR137204	Public database	Russia	MDR	Lineage 2	R	R	R	R	S	R	S	S	S
ERR137216	ERR137216	Public database	Russia	non-MDR	Lineage 2	R	S	S	R	S	R	S	S	S
ERR137223	ERR137223	Public database	Russia	MDR	Lineage 2	R	R	R	S	R	R	S	S	S
ERR137274	ERR137274	Public database	Russia	MDR	Lineage 4	R	R	S	S	S	R	S	S	S
ERR144549	ERR144549	Public database	Russia	MDR	Lineage 2	R	R	S	S	S	R	S	S	S
ERR144559	ERR144559	Public database	Russia	non-MDR	Lineage 4	R	S	S	S	S	S	S	S	S
ERR144564	ERR144564	Public database	Russia	non-MDR	Lineage 2	R	S	S	S	S	R	S	S	S
ERR144579	ERR144579	Public database	Russia	MDR	Lineage 2	R	R	R	S	S	R	S	S	S
ERR144620	ERR144620	Public database	Russia	MDR	Lineage 2	R	R	R	R	S	R	S	S	S
ERR144628	ERR144628	Public database	Russia	MDR	Lineage 2	R	R	R	S	S	R	S	S	S
ERR144633	ERR144633	Public database	Russia	MDR	Lineage 2	R	R	R	S	S	R	S	S	S
ERR158572	ERR158572	Public database	Russia	non-MDR	Lineage 2	S	S	R	S	S	S	S	S	S
ERR158579	ERR158579	Public database	Russia	MDR	Lineage 4	R	R	S	R	S	R	R	R	R
ERR229984	ERR229984	Public database	Russia	MDR	Lineage 2	R	R	R	R	S	R	S	S	S
ERR229991	ERR229991	Public database	Russia	non-MDR	Lineage 4	S	S	S	S	S	S	S	S	S
ERR229997	ERR229997	Public database	Russia	MDR	Lineage 2	R	R	S	S	S	S	S	S	S
ERR230003	ERR230003	Public database	Russia	MDR	Lineage 4	R	R	S	S	S	S	S	S	S
ERR234561	ERR234561	Public database	Russia	non-MDR	Lineage 4	S	S	S	R	S	S	S	S	S
ERR234565	ERR234565	Public database	Russia	MDR	Lineage 2	R	R	S	S	S	R	S	S	S
ERR234570	ERR234570	Public database	Russia	MDR	Lineage 2	R	R	S	S	S	R	S	S	S
ERR234574	ERR234574	Public database	Russia	MDR	Lineage 2	R	R	S	S	S	R	S	S	S
ERR234590	ERR234590	Public database	Russia	MDR	Lineage 2	R	R	R	S	R	R	S	S	S
ERR234596	ERR234596	Public database	Russia	MDR	Lineage 2	R	R	R	S	S	R	S	S	S
ERR234598	ERR234598	Public database	Russia	MDR	Lineage 2	R	R	R	S	R	R	S	S	S
ERR234617	ERR234617	Public database	Russia	non-MDR	Lineage 2	S	S	S	S	S	S	S	S	S
ERR234633	ERR234633	Public database	Russia	MDR	Lineage 2	R	R	R	S	R	R	S	S	S
ERR234645	ERR234645	Public database	Russia	MDR	Lineage 2	R	R	R	S	S	R	S	S	S
ERR234646	ERR234646	Public database	Russia	MDR	Lineage 2	R	R	S	S	S	R	S	S	S
ERR2515112	ERR2515112	Public database	South Africa	MDR	Lineage 2	R	R	R	R	S	R	S	S	S
ERR2515182	ERR2515182	Public database	South Africa	MDR	Lineage 4	R	R	R	R	S	S	S	S	S
ERR2515293	ERR2515293	Public database	South Africa	MDR	Lineage 4	R	R	R	R	S	S	S	S	S
ERR2515348	ERR2515348	Public database	South Africa	MDR	Lineage 1	R	R	R	S	S	R	S	S	S
ERR2515498	ERR2515498	Public database	South Africa	non-MDR	Lineage 4	S	R	S	S	S	S	S	S	S
ERR2515517	ERR2515517	Public database	South Africa	non-MDR	Lineage 2	S	S	S	S	S	S	S	S	S
ERR2515542	ERR2515542	Public database	South Africa	MDR	Lineage 4	R	R	R	R	S	S	S	S	S
ERR2515663	ERR2515663	Public database	South Africa	MDR	Lineage 4	R	R	S	S	S	S	S	S	S
ERR2515684	ERR2515684	Public database	South Africa	MDR	Lineage 2	R	R	R	R	R	R	R	R	R
ERR2515735	ERR2515735	Public database	South Africa	MDR	Lineage 2	R	R	R	R	S	R	R	R	R
ERR2515934	ERR2515934	Public database	South Africa	MDR	Lineage 4	R	R	R	S	S	S	S	S	S
ERR2515938	ERR2515938	Public database	South Africa	MDR	Lineage 4	R	R	S	S	S	S	S	S	S
ERR2515950	ERR2515950	Public database	South Africa	MDR	Lineage 4	R	R	R	S	S	S	S	S	S
ERR2515956	ERR2515956	Public database	South Africa	MDR	Lineage 4	R	R	S	S	S	R	S	S	S
ERR2516018	ERR2516018	Public database	South Africa	MDR	Lineage 2	R	R	R	R	S	S	S	S	S
ERR2516031	ERR2516031	Public database	South Africa	MDR	Lineage 4	R	R	R	R	S	S	S	S	S
ERR2516069	ERR2516069	Public database	South Africa	MDR	Lineage 4	R	R	R	R	S	S	S	S	S
SRR1011449	SRR1011449	Public database	South Africa	MDR	Lineage 2	R	R	R	R	S	R	S	R	R
SRR1062863	SRR1062863	Public database	South Africa	non-MDR	Lineage 1	S	S	S	S	S	S	S	S	S
SRR1140949	SRR1140949	Public database	South Africa	MDR	Lineage 4	R	R	R	R	S	S	S	S	S
SRR1140950	SRR1140950	Public database	South Africa	non-MDR	Lineage 4	S	R	S	S	S	S	S	S	S
SRR1140959	SRR1140959	Public database	South Africa	non-MDR	Lineage 1	R	S	R	S	S	S	S	S	S
SRR1140965	SRR1140965	Public database	South Africa	MDR	Lineage 4	R	R	R	R	S	S	S	S	S
SRR1180396	SRR1180396	Public database	South Africa	MDR	Lineage 2	R	R	R	R	S	R	S	S	S
SRR1181168	SRR1181168	Public database	South Africa	non-MDR	Lineage 4	S	S	S	S	S	S	S	S	S
SRR1181211	SRR1181211	Public database	South Africa	non-MDR	Lineage 4	S	S	S	S	S	S	S	S	S
SRR1181301	SRR1181301	Public database	South Africa	non-MDR	Lineage 4	S	S	S	S	S	S	S	S	S
SRR1181313	SRR1181313	Public database	South Africa	non-MDR	Lineage 4	S	S	S	S	S	S	S	S	S
SRR1184002	SRR1184002	Public database	South Africa	MDR	Lineage 4	R	R	R	S	R	R	S	S	S
SRR1184027	SRR1184027	Public database	South Africa	non-MDR	Lineage 4	R	S	R	S	S	S	S	S	S
SRR1184340	SRR1184340	Public database	South Africa	MDR	Lineage 4	R	R	S	R	S	R	R	R	R
SRR1184356	SRR1184356	Public database	South Africa	non-MDR	Lineage 4	S	S	S	S	S	S	S	S	S
SRR833055	SRR833055	Public database	South Africa	MDR	Lineage 4	R	R	R	R	S	S	S	S	S
SRR833080	SRR833080	Public database	South Africa	MDR	Lineage 4	R	R	R	R	S	S	S	S	S
SRR833095	SRR833095	Public database	South Africa	non-MDR	Lineage 4	S	S	S	R	S	S	S	S	S

SRR833148	SRR833148	Public database	South Africa	MDR	Lineage 4	R	R	R	R	S	R	S	S	S
SRR833173	SRR833173	Public database	South Africa	MDR	Lineage 4	R	R	R	R	R	S	R	R	R
SRR924206	SRR924206	Public database	South Africa	MDR	Lineage 4	R	R	R	R	R	S	R	R	R
SRR924217	SRR924217	Public database	South Africa	MDR	Lineage 4	R	R	R	R	R	S	S	R	S
SRR924218	SRR924218	Public database	South Africa	MDR	Lineage 4	R	R	R	R	R	S	S	S	S
SRR924235	SRR924235	Public database	South Africa	MDR	Lineage 4	R	R	R	R	S	S	S	S	S
SRR924697	SRR924697	Public database	South Africa	MDR	Lineage 4	R	R	R	R	R	S	R	R	R
SRR924699	SRR924699	Public database	South Africa	non-MDR	Lineage 4	S	S	S	S	S	S	S	S	S
SRR924700	SRR924700	Public database	South Africa	non-MDR	Lineage 4	S	S	S	S	S	S	S	S	S
SRR924705	SRR924705	Public database	South Africa	non-MDR	Lineage 4	S	S	S	S	S	S	S	S	S
ERR2199787	ERR2199787	Public database	Swaziland	non-MDR	Lineage 1	S	S	S	S	S	S	S	S	S
ERR2199923	ERR2199923	Public database	Swaziland	non-MDR	Lineage 1	S	S	S	S	S	S	S	S	S
ERR2199924	ERR2199924	Public database	Swaziland	non-MDR	Lineage 4	S	S	S	S	S	S	S	S	S
ERR2199942	ERR2199942	Public database	Swaziland	non-MDR	Lineage 1	S	S	S	S	S	S	S	S	S
ERR2199950	ERR2199950	Public database	Swaziland	non-MDR	Lineage 1	S	S	S	S	S	S	S	S	S
ERR2199977	ERR2199977	Public database	Swaziland	non-MDR	Lineage 2	S	S	S	S	S	S	S	S	S
ERR2199981	ERR2199981	Public database	Swaziland	non-MDR	Lineage 4	S	S	S	S	S	S	S	S	S
ERR2199982	ERR2199982	Public database	Swaziland	non-MDR	Lineage 4	S	S	S	S	S	S	S	S	S
ERR2199990	ERR2199990	Public database	Swaziland	non-MDR	Lineage 4	S	S	S	S	S	S	S	S	S
ERR2200002	ERR2200002	Public database	Swaziland	non-MDR	Lineage 3	S	S	S	S	S	S	S	S	S
ERR2200014	ERR2200014	Public database	Swaziland	non-MDR	Lineage 2	S	S	S	S	S	S	S	S	S
ERR2200017	ERR2200017	Public database	Swaziland	non-MDR	Lineage 4	S	S	S	R	S	S	S	S	S
ERR2200034	ERR2200034	Public database	Swaziland	non-MDR	Lineage 4	S	S	S	S	S	S	S	S	S
ERR2200078	ERR2200078	Public database	Swaziland	non-MDR	Lineage 4	S	S	S	S	S	S	S	S	S
ERR2200105	ERR2200105	Public database	Swaziland	non-MDR	Lineage 3	S	S	S	S	S	S	S	S	S
ERR550619	ERR550619	Public database	Swaziland	MDR	Lineage 4	R	R	R	R	R	R	S	S	S
ERR550734	ERR550734	Public database	Swaziland	MDR	Lineage 4	R	R	R	R	S	R	S	S	S
ERR550861	ERR550861	Public database	Swaziland	MDR	Lineage 4	R	R	R	R	S	R	S	S	S
ERR550869	ERR550869	Public database	Swaziland	MDR	Lineage 1	R	R	R	R	S	R	S	S	S
ERR550959	ERR550959	Public database	Swaziland	MDR	Lineage 4	R	R	R	R	S	S	S	S	S
ERR551014	ERR551014	Public database	Swaziland	MDR	Lineage 4	R	R	R	R	S	R	S	S	S
ERR551045	ERR551045	Public database	Swaziland	MDR	Lineage 4	R	R	S	R	S	R	S	S	S
ERR551280	ERR551280	Public database	Swaziland	MDR	Lineage 4	R	R	R	R	S	R	S	S	S
ERR551393	ERR551393	Public database	Swaziland	MDR	Lineage 4	R	R	R	R	S	R	S	S	S
ERR551545	ERR551545	Public database	Swaziland	MDR	Lineage 2	R	R	R	R	S	R	S	S	S
ERR551578	ERR551578	Public database	Swaziland	MDR	Lineage 4	R	R	S	S	S	S	S	S	S
ERR551739	ERR551739	Public database	Swaziland	MDR	Lineage 1	R	R	R	R	S	R	S	S	S
ERR551801	ERR551801	Public database	Swaziland	MDR	Lineage 1	R	R	R	R	S	R	S	S	S
ERR552000	ERR552000	Public database	Swaziland	MDR	Lineage 1	R	R	R	R	S	R	S	S	S
ERR552010	ERR552010	Public database	Swaziland	MDR	Lineage 4	R	R	S	R	S	R	S	S	S
ERR552124	ERR552124	Public database	Swaziland	MDR	Lineage 4	R	R	R	R	S	R	S	S	S
ERR552361	ERR552361	Public database	Swaziland	MDR	Lineage 4	R	R	R	R	S	S	S	S	S
ERR552456	ERR552456	Public database	Swaziland	MDR	Lineage 4	R	R	S	R	R	R	S	S	S
ERR552495	ERR552495	Public database	Swaziland	MDR	Lineage 1	R	R	R	R	S	R	S	S	S
ERR552518	ERR552518	Public database	Swaziland	MDR	Lineage 4	R	R	R	R	S	R	S	S	S
ERR552599	ERR552599	Public database	Swaziland	MDR	Lineage 4	R	R	R	R	S	R	S	S	S
ERR552612	ERR552612	Public database	Swaziland	MDR	Lineage 4	R	R	R	R	S	R	S	S	S
ERR552780	ERR552780	Public database	Swaziland	MDR	Lineage 4	R	R	R	R	S	R	S	S	S
ERR552789	ERR552789	Public database	Swaziland	MDR	Lineage 4	R	R	S	R	S	S	S	S	S
ERR553058	ERR553058	Public database	Swaziland	MDR	Lineage 2	R	R	R	S	S	R	S	S	S
ERR553230	ERR553230	Public database	Swaziland	MDR	Lineage 4	R	R	R	R	S	R	S	S	S
ERR553264	ERR553264	Public database	Swaziland	MDR	Lineage 4	R	R	R	R	S	S	S	S	S
ERR553313	ERR553313	Public database	Swaziland	MDR	Lineage 2	R	R	R	R	S	R	S	S	S
ERR553321	ERR553321	Public database	Swaziland	MDR	Lineage 1	R	R	R	R	S	R	S	S	S
ERR553357	ERR553357	Public database	Swaziland	MDR	Lineage 4	R	R	S	R	S	R	S	S	S
SRR5709740	SRR5709740	Public database	Thailand	non-MDR	Lineage 4	S	S	S	S	S	S	S	S	S
SRR5709741	SRR5709741	Public database	Thailand	non-MDR	Lineage 2	S	S	S	S	S	S	S	S	S
SRR5709742	SRR5709742	Public database	Thailand	non-MDR	Lineage 2	S	S	S	S	S	R	S	S	S
SRR5709746	SRR5709746	Public database	Thailand	non-MDR	Lineage 2	S	S	S	S	S	S	S	S	S
SRR5709772	SRR5709772	Public database	Thailand	MDR	Lineage 2	R	R	S	S	S	S	S	S	S
SRR5709776	SRR5709776	Public database	Thailand	MDR	Lineage 2	R	R	R	S	S	S	S	S	S
SRR5709777	SRR5709777	Public database	Thailand	MDR	Lineage 1	R	R	R	S	S	S	S	S	S
SRR5709783	SRR5709783	Public database	Thailand	non-MDR	Lineage 2	S	S	S	S	S	S	S	S	S
SRR5709786	SRR5709786	Public database	Thailand	non-MDR	Lineage 2	S	S	S	S	S	S	S	S	S
SRR5709787	SRR5709787	Public database	Thailand	non-MDR	Lineage 2	S	S	S	S	S	S	S	S	S
SRR5709799	SRR5709799	Public database	Thailand	MDR	Lineage 2	R	R	R	R	R	R	S	S	S
SRR5709801	SRR5709801	Public database	Thailand	MDR	Lineage 2	R	R	R	R	R	R	S	S	S
SRR5709817	SRR5709817	Public database	Thailand	MDR	Lineage 2	R	R	S	R	S	R	S	S	S
SRR5709823	SRR5709823	Public database	Thailand	MDR	Lineage 2	R	R	R	R	R	S	S	S	S
SRR5709853	SRR5709853	Public database	Thailand	non-MDR	Lineage 2	R	S	S	S	S	R	S	S	S
SRR5709855	SRR5709855	Public database	Thailand	non-MDR	Lineage 2	S	S	S	S	S	S	S	S	S
SRR5709858	SRR5709858	Public database	Thailand	non-MDR	Lineage 1	R	S	S	R	S	R	S	S	S
SRR5709862	SRR5709862	Public database	Thailand	MDR	Lineage 2	R	R	R	R	R	R	S	S	S
SRR5709863	SRR5709863	Public database	Thailand	MDR	Lineage 2	R	R	R	R	S	R	S	R	S
SRR5709865	SRR5709865	Public database	Thailand	MDR	Lineage 2	R	R	R	R	R	R	S	S	S
SRR5709867	SRR5709867	Public database	Thailand	MDR	Lineage 2	R	R	R	R	R	R	S	S	S
SRR5709881	SRR5709881	Public database	Thailand	MDR	Lineage 2	R	R	S	S	S	S	S	S	S
SRR5709898	SRR5709898	Public database	Thailand	MDR	Lineage 2	R	R	S	R	R	R	S	S	S
SRR5709907	SRR5709907	Public database	Thailand	MDR	Lineage 2	R	R	R	R	S	R	S	R	S
SRR5709923	SRR5709923	Public database	Thailand	MDR	Lineage 2	R	R	R	S	S	S	S	S	S
SRR5709924	SRR5709924	Public database	Thailand	MDR	Lineage 1	R	R	S	S	S	R	S	S	S
SRR5709927	SRR5709927	Public database	Thailand	MDR	Lineage 2	R	R	R	S	S	R	S	S	S
SRR5709941	SRR5709941	Public database	Thailand	non-MDR	Lineage 1	S	S	S	S	S	S	S	S	S
SRR5709951	SRR5709951	Public database	Thailand	MDR	Lineage 2	R	R	S	R	S	R	S	S	S
SRR5709953	SRR5709953	Public database	Thailand	MDR	Lineage 2	R	R	S	S	S	S	S	S	S
SRR5709956	SRR5709956	Public database	Thailand	non-MDR	Lineage 2	R	S	S	S	S	S	S	S	S
SRR5709958	SRR5709958	Public database	Thailand	non-MDR	Lineage 2	S	S	S	S	S	S	S	S	S
SRR5709962	SRR5709962	Public database	Thailand	MDR	Lineage 2	R	R	S	S	S	S	S	S	S
SRR5709968	SRR5709968	Public database	Thailand	MDR	Lineage 1	R	R	R	S	S	S	S	S	S
SRR5709970	SRR5709970	Public database	Thailand	MDR	Lineage 2	R	R	R	S	S	S	S	S	S
SRR5709971	SRR5709971	Public database	Thailand	non-MDR	Lineage 4	S	S	S	S	S	S	S	S	S
SRR5709977	SRR5709977	Public database	Thailand	non-MDR	Lineage 2	S	S	S	S	S	S	S	S	S
SRR5709994	SRR5709994	Public database	Thailand	MDR	Lineage 2	R	R	S	S	R	S	S	S	S
SRR5709996	SRR5709996	Public database	Thailand	MDR	Lineage 2	R	R	R	R	S	S	S	S	S
SRR5710009	SRR5710009	Public database	Thailand	MDR	Lineage 2	R	R	R	S	S	R	S	S	S

SRR5710017	SRR5710017	Public database	Thailand	MDR	Lineage 4	R	R	R	R	R	S	S	S	S
SRR5710025	SRR5710025	Public database	Thailand	MDR	Lineage 2	R	R	S	S	S	R	S	S	S
SRR5837707	SRR5837707	Public database	Thailand	MDR	Lineage 2	R	R	R	R	S	R	S	S	S
SRR5837710	SRR5837710	Public database	Thailand	MDR	Lineage 2	R	R	R	R	R	R	R	R	R
SRR5837712	SRR5837712	Public database	Thailand	MDR	Lineage 2	R	R	S	S	S	R	S	S	S

Table with 14 columns representing genetic markers and their associated values across various *Mtb* strains. The table contains a dense grid of numerical data for each strain-variant combination.

P value of InDels

Virulence gene InDels	Population mutation rate	Lineage					Drug resistance							
		L1	L2	L3	L4	INH	RIF	EMB	PZA	LFX	SM	AM	KM	CM
Rv0165c(510InCC)	49.93%	2.82E-09	4.59E-51	7.55E-13	1.2E-104	0.162462	0.420749	0.260523	0.448419	0.906213	8.77E-11	0.756225	0.172105	0.87342
Rv0427c(567InGTA)	1.80%	1.45E-45	0.003844	0.66748	0.007344	0.887765	0.844219	0.015307	0.483804	0.191718	0.003376	0.681935	0.55521	0.670081
Rv0590(314InA)	4.50%	0.365912	2.4E-09	0.220629	4.6E-06	0.495032	0.829965	0.597961	0.983591	0.856466	0.915818	0.230422	0.896795	0.222412
Rv0590(762InG)	49.48%	2.01E-09	2.07E-49	4.64E-13	8.3E-103	0.081289	0.251988	0.182202	0.570251	0.975315	2.08E-11	0.707545	0.150267	0.822178
Rv0592(1388InC)	11.39%	0.055806	7.11E-16	5.32E-40	0.025156	0.153862	0.314406	0.001496	0.008818	0.8131	0.001812	0.824875	0.721609	0.845646
Rv0670(171DelC)	7.05%	0.180553	2.55E-10	0.085782	2.38E-16	0.397691	0.200654	0.361032	0.001799	0.276294	0.001148	0.034182	0.046085	0.001296
Rv0934(64InA)	2.70%	0.631269	0.000236	0.448984	1.5E-06	0.830764	0.88621	0.645192	0.388953	0.078514	0.712197	0.462096	0.349952	0.451352
Rv1028c(200DelCA)	5.25%	4.2E-142	8.29E-08	0.16648	5.78E-07	0.120386	0.028441	0.240167	0.464967	0.007013	0.726234	0.174949	0.107975	0.168027
Rv1192(470DelCG)	2.70%	0.631269	0.000236	0.448984	1.5E-06	0.077866	0.092454	0.616509	0.909823	0.319228	0.284598	0.462096	0.349952	0.451352
Rv1204c(105DelGGT)	5.25%	4.2E-142	8.29E-08	0.16648	5.78E-07	0.120386	0.028441	0.240167	0.464967	0.007013	0.726234	0.174949	0.107975	0.168027
Rv1660(835DelG)	4.95%	2E-133	2.14E-07	0.18622	1.33E-06	0.287025	0.083656	0.126354	0.646481	0.009299	0.97428	0.195208	0.123446	0.18786
Rv1661(170DelG)	1.80%	0.862825	0.003844	0.66748	0.000138	0.151845	0.171461	0.61378	0.915575	0.191718	0.245213	0.065091	0.148091	0.070747
Rv1661(579DelGGGC)	1.50%	1.75E-37	0.009838	0.771807	0.016862	0.191488	0.212591	0.508259	0.736383	0.261901	0.184943	0.786543	0.656359	0.774456
Rv1915(882InT)	64.92%	8.19E-05	1.16E-15	1.89E-07	2.01E-37	0.553413	0.5565	0.929033	0.013177	0.020471	0.00197	0.191486	0.988806	0.134724
Rv2275(647InG)	5.25%	4.2E-142	8.29E-08	0.16648	5.78E-07	0.120386	0.028441	0.240167	0.464967	0.007013	0.726234	0.174949	0.107975	0.168027
Rv2275(791DelCGC)	2.55%	0.663188	0.000375	0.47826	3.2E-06	0.883527	0.831487	0.43051	0.35895	0.2885	0.702811	0.491628	0.376806	0.480673
Rv2351c(1066InCAGG)	6.75%	0.195861	6.76E-10	1.2E-130	8.51E-09	0.747325	0.153726	0.513374	0.109632	0.070841	0.156088	0.659167	0.767497	0.633496
Rv2388c(225InG)	1.20%	3.83E-22	0.12865	0.902152	0.039199	0.449898	0.973962	0.036326	0.35096	0.362403	0.188489	0.917004	0.784814	0.90482
Rv2553c(171DelGGT)	4.05%	0.416923	1.82E-08	0.261997	1.58E-05	0.099469	0.276018	0.414932	0.388598	0.778004	0.000643	0.27262	0.185358	0.263927
Rv2553c(171InGGT)	7.65%	0.004922	1.58E-09	0.068874	1.25E-09	0.066442	0.315159	0.034479	0.180283	0.184044	0.006406	0.5094	0.833694	0.890446
Rv2703(248DelCCG)	1.20%	0.900792	0.005645	0.902152	0.039199	0.127402	0.425728	0.957656	0.835844	0.999455	0.006087	0.917004	0.784814	0.90482
Rv3872(4DelA)	4.80%	0.91214	0.000416	0.208459	8.1E-06	0.449512	0.648682	0.913853	0.428516	0.246339	0.389264	0.206258	0.132026	0.198692
Rv3876(1395DelG)	1.35%	1.85E-33	0.015803	0.833095	0.025658	0.608093	0.644895	0.746427	0.936604	0.307478	0.27647	0.847916	0.716497	0.835759
Rv3876(1453InA)	1.50%	1.75E-37	0.009838	0.771807	0.016862	0.191488	0.212591	0.508259	0.736383	0.261901	0.184943	0.786543	0.656359	0.774456
Rv1028c(2543DelAC)	3.75%	0.4556	6.95E-08	0.294285	3.59E-05	0.001382	0.017113	0.005097	1.38E-05	0.005496	8.91E-08	1.34E-08	1.86E-14	2.18E-08
Rv1128c(918DelC)	44.53%	1.2E-07	4.4E-142	1.62E-10	5.31E-86	4.74E-06	9.04E-05	0.01645	0.008088	3.93E-10	2.09E-24	0.000456	0.000101	0.006111
Rv1371(614DelCC)	28.49%	0.000256	4.15E-71	8.96E-06	2.51E-43	0.0265	0.083638	0.847774	0.32213	2.01E-05	1.64E-10	0.150906	0.241229	0.501936
Rv1971(64DelGTGCT)	6.75%	0.195861	7.73E-14	0.095744	8.51E-09	0.001586	0.005913	0.002123	0.331993	0.158269	6.68E-10	0.09902	0.032987	0.110912
Rv1979c(1418InGCC)	1.80%	0.862825	0.003844	0.66748	0.000138	0.000276	0.003373	0.004393	0.010234	0.191718	0.003542	0.681935	0.55521	0.670081
Rv2027c(775DelG)	36.13%	1.05E-05	1.8E-100	9.51E-08	5.48E-61	0.016256	0.136586	0.30501	0.602658	0.704664	7.81E-16	0.146391	0.008645	0.184498
Rv2383c(2576DelCCA)	1.80%	0.862825	0.003844	0.66748	0.000138	0.000276	0.003373	0.004393	0.010234	0.191718	0.003542	0.681935	0.55521	0.670081
Rv2437(106DelA)	44.53%	1.2E-07	4.4E-142	1.62E-10	5.31E-86	3.75E-07	1E-05	0.010604	0.001777	3.93E-10	7.95E-26	0.000456	0.000101	0.006111
Rv2885c(1306InC)	54.27%	2.23E-06	6.59E-74	2.53E-09	2.3E-124	4.31E-05	0.000119	0.025718	0.13007	5.42E-07	1.72E-20	0.073579	0.021921	0.310208

Those in Lineage column with p values less than 0.05/4 are highlighted in bold and in Drug resistance column less than 0.05 are highlighted in bold and italic.

P values of mutations on pks15 1

pks15/1 SNPs & InDels	Population mutation rate	Lineage					Drug resistance						
		L1	L2	L3	INH	RIF	EMB	PZA	LFX	SM	AM	KM	CM
Rv2947c(V333A)	91.80%	0.469612	0.747531	0.840894	0.697325	0.496907	0.033072	0.95973	0.136255	0.606108	0.89156	0.972636	0.977624
Rv2946c(Q918R)	7.38%	3.88E-70	1.71E-24	0.076137	0.092438	0.014446	0.454643	0.522714	0.002565	0.046744	0.176837	0.103058	0.199382
Rv2947c(G374R)	7.65%	5.57E-73	1.88E-25	0.068812	0.122351	0.021887	0.348546	0.687918	0.002044	0.07144	0.164502	0.094131	0.18621
Rv2947c(T46I)	4.37%	0.449492	0.072218	0.234835	0.047271	0.039662	0.256465	0.024204	0.659954	0.003517	0.400141	0.283871	0.432329
Rv2946c(4126InACGG G)	1.91%	0.918256	0.362426	0.649078	0.453522	0.502437	0.48986	0.701397	0.758743	0.497952	0.861279	0.716503	0.898682
Rv2946c(714InC)	2.46%	1.29E-21	4.48E-08	0.50827	0.280664	0.324286	0.933832	0.805726	0.149389	0.03051	0.713591	0.572627	0.750526

Those in Lineage column with p values less than 0.05/3 are highlighted in bold and in Drug resistance column less than 0.05 are highlighted in bold and italic.