A standardised method for interpreting the association between mutations and phenotypic
drug-resistance in *Mycobacterium tuberculosis*

Authors:

Paolo Miotto, Belay Tessema, Elisa Tagliani, Leonid Chindelevitch, Angela M. Starks, Claudia Emerson, Debra Hanna, Peter S. Kim, Richard Liwski, Matteo Zignol, Christopher Gilpin, Stefan Niemann, Claudia M. Denkinger, Joy Fleming, Rob Warren, Derrick Crook, James Posey, Sebastien Gagneux, Sven Hoffner, Camilla Rodrigues, Iñaki Comas, David M. Engelthaler, Megan Murray, David Alland, Leen Rigouts, Christoph Lange, Keertan Dheda, Rumina Hasan, Uma Devi K. Ranganathan, Ruth McNerney, Matthew Ezewudo, Daniela M. Cirillo, Marco Schito, Claudio U. Köser, Timothy C. Rodwell.
Table S1.1. Loci of interest considered for retrieving mutations associated with drug resistance. References reported in the table have been chosen to justify the selection of the loci for each drug.

| Drug   | Gene name | Gene locus | References |
|--------|-----------|------------|------------|
| Amikacin (AM) | rrs | MTB000019 | Maus 2005 | 1-2; Reeves 2015 |
| Capreomycin (CM) | rrs, tlyA | MTB000019, Ry1694 | Maus 2005 | 1-2; Johansen 2006 |
| Kanamycin (KM) | rrs, eis, whiB7 | MTB000019, Rv2416c, Rv3197A | Maus 2005 | 1-2; Reeves 2015 | 3 |
| Streptomycin (S) | rpsL, tap, rrs, whiB7, gidB | Rv0682, Rv1258c, Rv3197A | Finken 1993 | 8 |
| Levofloxacin (LEF) | gyrA, gyrB | Rv0006, Rv0005 | Takiff 1994 | 11; Nebenzahl-Guimaraes 2014 |
| Ofloxacin (OFX) | gyrA, gyrB | Rv0006, Rv0005 | Takiff 1994 | 11; Nebenzahl-Guimaraes 2014 |
| Moxifloxacin (MFX) | gyrA, gyrB | Rv0006, Rv0005 | Takiff 1994 | 11; Nebenzahl-Guimaraes 2014 |
| Isoniazid (H) | mshA, mabA (fabG1), inhA, katG, furA | Rv0486, Rv1483, Rv1484, Rv1908c, Rv1909c | Vilchèze 2008 | 15; Vilchèze 2011 |
| Ethionamide (ETO) | mshA, mabA (fabG1), inhA, ethA | Rv0846, Rv1483, Rv1484, Rv3199c | Vilchèze 2008 | 15; Vilchèze 2011 |
| Prothionamide (PTO) | mshA, mabA (fabG1), inhA, ethA | Rv0846, Rv1483, Rv1484, Rv3199c | Vilchèze 2008 | 14; Vilchèze 2011 |
| Pyrazinamide (Z) | pncA | Rv2043c | Scorpio 1996 |
| Rifampicin (R) | rpoB | Rv0667 | Telenti 1993 | 25; Williams 1998; Schön 2013 |

References

1. Maus CE, Plikaytis BB, Shinnick TM. Molecular analysis of cross-resistance to capreomycin, kanamycin, amikacin, and viomycin in Mycobacterium tuberculosis. Antimicrob Agents Chemother 2005; 49(8): 3192-7.
2. Maus CE, Plikaytis BB, Shinnick TM. Mutation of tlyA confers capreomycin resistance in Mycobacterium tuberculosis. Antimicrob Agents Chemother 2005; 49(2): 571-7.
3. Reeves AZ, Campbell PJ, Willby MJ, Posey JE. Disparities in capreomycin resistance levels associated with the rrs A1401G mutation in clinical isolates of Mycobacterium tuberculosis. Antimicrob Agents Chemother 2015; 59(1): 444-9.
4. Johansen SK, Maus CE, Plikaytis BB, Douthwaite S Capreomycin binds across the ribosomal subunit interface using tlyA-encoded 2'-O-methylations in 16S and 23S rRNAs. Mol Cell 2006; 23(2): 173-82.
5. Köser CU, Bryant JM, Parkhill J, Peacock SJ. Consequences of whiB7 (Rv3197A) mutations in Beijing genotype isolates of the Mycobacterium tuberculosis complex. Antimicrob Agents Chemother 2013; 57(7): 3461.
6. Reeves AZ, Campbell PJ, Sultana R, et al. Aminoglycoside cross-resistance in Mycobacterium tuberculosis due to mutations in the 5' untranslated region of whiB7. Antimicrob Agents Chemother 2013; 57(4): 1857-65.
7. Zaunbrecher MA, Sikes RD, Metchock B, Shinnick TM, Posey JE. Overexpression of the chromosomally encoded aminoglycoside acetyltransferase eis confers kanamycin resistance in Mycobacterium tuberculosis. Proc Natl Acad Sci USA 2009; 106(47): 20004-9.
8. Finken M, Kirchner P, Meier A, Wrede A, Böttger EC. Molecular basis of streptomycin resistance in Mycobacterium tuberculosis: alterations of the ribosomal protein S12 gene and point mutations within a functional 16S ribosomal RNA pseudoknot. Mol Microbiol 1993; 9(6): 1239-46.
Okamoto S, Tamaru A, Nakajima C, et al. Loss of a conserved 7-methylguanosine modification in 16S rRNA confers low-level streptomycin resistance in bacteria. Mol Microbiol 2007; 63(4): 1096-106.

Wong SY, Lee JS, Kwak HK, Via LE, Boshoff HI, Barry CE. Mutations in gidB confer low-level streptomycin resistance in Mycobacterium tuberculosis. Antimicrob Agents Chemother 2011; 55(6): 2515-22.

Takiff HE, Salazar L, Guerrero C, et al. Cloning and nucleotide sequence of Mycobacterium tuberculosis gyrA and gyrB genes and detection of quinolone resistance mutations. Antimicrob Agents Chemother 1994; 38(4): 773-80.

Nebenzahl-Guimaraes H, Jacobson KR, Farhat MR, Murray MB. Systematic review of allelic exchange experiments aimed at identifying mutations that confer drug resistance in Mycobacterium tuberculosis. J Antimicrob Chemother 2014; 69(2): 331-42.

Kocagöz T, Hackbarth CJ, Unsal I, Rosenberg EY, Nikaido H, Chambers HF. Gyrase mutations in laboratory-selected mycobacterial isolates. J Antimicrob Agents Chemother 1996; 38(8): 1768-74.

Vilchêze C, Av-Gay Y, Attarian R, et al. Mycothiol biosynthesis is essential for ethionamide susceptibility in Mycobacterium tuberculosis. Mol Microbiol 2008; 69(5): 1316-29.

Vilchêze C, Av-Gay Y, Barnes SW, et al. Coresistance to isoniazid and ethionamide maps to mycothiol biosynthetic genes in Mycobacterium bovis. Antimicrob Agents Chemother 2011; 55(9): 4422-3.

Ando H, Miyoshi-Akiyama T, Watanabe S, Kikikae T. A silent mutation in gyrA confers isoniazid resistance on Mycobacterium tuberculosis. Mol Microbiol 2014; 91(3): 538-47.

Banerjee A, Dubnau E, Quemard A, et al. inhA, a gene encoding a target for isoniazid and ethionamide in Mycobacterium tuberculosis. Science 1994; 263(5144): 227-30.

Kapur V, Li LL, Hamrick MR, et al. Rapid Mycobacterium species assignment and unambiguous identification of mutations associated with antimicrobial resistance in Mycobacterium tuberculosis by automated DNA sequencing. Arch Pathol Lab Med 1995; 119(2): 131-8.

Zhang Y, Heym B, Allen B, Young D, Cole S. The catalase-peroxidase gene and isoniazid resistance of Mycobacterium tuberculosis. Nature 1992; 358(6387): 591-3.

Ando H, Kitao T, Miyoshi-Akiyama T, Kato S, Mori T, Kikikae T. Downregulation of katG expression is associated with isoniazid resistance in Mycobacterium tuberculosis. Mol Microbiol 2011; 79(6): 1615-28.

DeBarber AE, Mdluli K, Bosman M, Bekker LG, Barry CE. Ethionamide activation and sensitivity in multidrug-resistant Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 2000; 97(17): 9677-82.

Dover LG, Alahari A, Gratraud P, et al. EthA, a common activator of thiocarbamidine-containing drugs acting on different mycobacterial targets. Antimicrob Agents Chemother 2007; 51(3): 1055-63.

Gopal P, Dick T. The new tuberculosis drug Perchozone® shows cross-resistance with thiacetazone. Int J Antimicrob Agents 2015; 45(4): 430-3.

Scorpio A, Zhang Y. Mutations in pncA, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antitubercular drug pyrazinamidine in tubercle bacillus. Nat Med 1996; 2(6): 662-7.

Telenti A, Imboden P, Marchesi F, et al. Detection of rifampicin-resistance mutations in Mycobacterium tuberculosis. Lancet 1993; 341(8846): 647-50.

Williams DL, Spring L, Collins L, et al. Contribution of rpoB mutations to development of rifamycin cross-resistance in Mycobacterium tuberculosis. Antimicrob Agents Chemother 1998; 42(7): 1853-7.

Schon T, Jureen P, Chryssanthou E, et al. Rifampicin-resistant and rifabutin-susceptible Mycobacterium tuberculosis strains: a breakpoint artefact? J Antimicrob Chemother 2013; 68(9): 2074-7.
Supplementary 4

Assessment Methods and utilization of data

| Main possible sources of bias                                      | Bias mitigation                                                                 |
|-------------------------------------------------------------------|--------------------------------------------------------------------------------|
| **Screening**                                                     | **Screening**                                                                  |
| - Papers not accessible have been excluded                        | - High number of papers considered should compensate for the papers excluded a priori |
| **Sample selection**                                              | **Sample selection**                                                           |
| - Sampling strategy has not been systematically tracked           | - Studies providing DST only for mutated strains have been excluded            |
| - Due to their high number, studies providing sequencing only of drug resistant cases have been included | - Studies providing sequencing data for discrepancies or for a subset of data selected as representative of the entire data set (e.g. % of susceptible cases or random selected subgroup and similar) have been excluded |
| **Data sets**                                                     | **Data sets**                                                                  |
| - Unless clearly specified in the study, reporting of multiple mutations was considered based on the data reported in the tables | - Only data produced for the published study have been considered in order to avoid data duplication (e.g. data reported as produced in references have not been considered) |
| - In almost all the cases, double mutations reported in the studies are not clearly classified as “mixed infection” or “real presence of multiple mutations on the same genome” | - Whenever identifiable, replicated data sets have been excluded |
| - Hetero-resistance defined as the presence of wild-type + mutated alleles has not been considered and only the mutated allele has been reported | |

Country representativeness was considered as “indicative” because (i) origin of samples is not always clearly stated and/or (ii) origin of samples is often referred to by regions (e.g. Europe, North America) rather than by specific countries.

The quality of included studies was appraised with a modified Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool. QUADAS-2 consists of four domains: patient selection, index test, reference standard, and flow and timing. Risk of bias was determined.

**Domain 1: Sample Selection**
Studies included in the systematic review considered culture isolates which probably introduced bias. Most published data is focused on drug resistant and/or MDR – TB patients and only a relatively small number of studies consider appropriate proportions of susceptible cases in order to provide statistical confidence to assessments made correlating genotype to phenotype. Thus, the predominant selection strategy was by convenience or based on a case-control design, representing a bias source. Including surveillance data could provide a mitigation strategy.

Time and geographical regions could potentially represent a source of bias in terms of circulating strains (e.g. samples over-representing specific geographical settings and/or specific time frames). However, timing and geographical regions of isolation has been tracked.

Cases not matching the review question are likely to be successfully excluded: studies providing partial data sets (e.g. sequencing/phenotypic data for discrepancies or for data subsets) have been excluded.

**Domain 2: Index Test**
Different index tests have been used. Despite the main purpose of the review, to provide a correlation between a given mutation observed and it’s observed phenotypic drug susceptibility results, rather than evaluate the performance of a test/technology in detecting the mutations, the possibility to have studies reporting errors during the analysis of NGS results could not be excluded.

Considering the sample selection strategy, most of the cases have been analysed in an un-blinded manner. This could affect the results because the operator could pay more attention in identifying mutations in resistant cases (e.g. looking for smaller peaks in the electropherogram or lower percentage of nucleotides in NGS data) or considering clean wild-type sequences in susceptible cases (e.g. overlooking double signals).
In addition, assessment of reporting of multiple mutations could be biased due to the performance limits of the sequencing chemistry and the technology utilized (e.g., inability to detect and discriminate between mixed populations and real double mutations) as well as the data reporting system (e.g., most of the studies reporting mutations known or assumed to be involved in drug resistance and do not report any additional mutations or mutations known or assumed to be not involved in drug resistance including silent mutations or mutations known to be lineage-specific. Also most of the studies relied on targeted sequencing, thus mutations outside the targeted regions are not detected); therefore the data presented is not truly indicative of the true genomic sequence.

**Domain 3: Reference Standard**

If the phenotypic reference standard used was WHO-endorsed and performed as per WHO recommendations, it was judged as 'low risk'; if the phenotypic reference standard was performed by other methods or procedures, it was judged to be at 'high risk' for inaccuracies.

Due to the well-known limits of the culture-based PZA testing methods, for PZA resistance the enzymatic-based assay (Wayne’s method) was also used as alternative reference test.

**Domain 4: Flow and Timing (time frame)**

The interval between when the sequencing and phenotypic reference standard was done is not considered a source of bias for this study.

Different reference standards were used in the different studies. Stratifying the data by reference standard methods used (liquid/solid culture, absolute/proportion method, critical concentration/minimum inhibitory concentration) should minimize biases introduced by the technique used for phenotypic DST testing. Quality of reference standard results could not be easily and objectively assessed; however, whenever possible and clearly stated in the study, participation of reference laboratories with external quality assurance programs was noted. In addition, we considered only studies providing phenotypic testing and sequencing for all of the tested samples, thus minimizing potential biases introduced by the inclusion of partial data sets.
Supplementary 5
Development of a standardized protocol for the assessment of resistance-associated mutations

Traditionally, likelihood ratios (LRs) are a metric measuring the strength of association between the outcome of a test and a diagnosis, and are often used to guide clinical decision-making. In this manuscript we used them as a measure of the strength of association between the presence of a mutation and the drug resistance phenotype. Under the null hypothesis of no association, the LR is expected to be 1, but deviations from this value can be due to both a true association as well as a variety of factors such as sampling effects. Generally speaking, the higher the likelihood ratio, the more the probability of the outcome of interest (here, drug resistance) increases after a positive test or observation (here, detection of the mutation). However, the baseline probabilities must not be ignored, even a mutation with a high LR is not conclusive proof of drug resistance in settings where the prior (or pre-test) probability of drug resistance is low.

In practice of evidence-based medicine, it is generally assumed that LRs from 2 to 5 yield small increases in the post-test probability, whereas ratios 5 to 10 denote moderate increases, and LRs above 10 indicate large increases. For ratios less than unity, the post-test probability is in fact lower than the pre-test probability, and the smaller the LR, the lower the post-test probability of the outcome of interest.

Data from the literature were used to calculate the frequency of each mutation in resistant (DR) and susceptible (DS) MTBC isolates and then used to derive a LR, which, in this case, represented the probability of observing a given mutation in a phenotypically resistant isolate divided by the probability of observing the mutation in a phenotypically susceptible isolate:

\[
LR^+ = \frac{Pr(M|DR^+)}{Pr(M|DR^-)} \quad \text{or equivalently} \quad LR^+ = \frac{sensitivity}{1 - specificity}
\]

In addition, odds ratios (OR), another metric commonly adopted in genome-wide association studies, were used to evaluate the association of the systematically reviewed genotypic and phenotypic data. The OR is the ratio of the odds of an event (“the presence of a mutation”) occurring in one group (“phenotypically drug resistant isolates”) to the odds of it occurring in another group (“phenotypically drug susceptible isolates”). It is a measure of effect size, describing the strength of association between two binary values, and is defined as:

\[
OR = \frac{Pr(M|DR^+)}{Pr(M|DR^-)} \quad \text{or equivalently} \quad OR = \frac{\text{sensitivity} \times \text{specificity}}{(1 - \text{sensitivity}) \times (1 - \text{specificity})}
\]

In the above formulas, “M+” and “M−” denote the presence and absence of a mutation, respectively, whereas “DR+” and “DR−” denote a resistant phenotype and a susceptible phenotype, respectively.

The interpretation of OR is similar to the interpretation of the LR, and can be summarized as follows:

- OR=1 the presence of a mutation does not affect the odds of phenotypic resistance
- OR>1 the presence of a mutation associated with higher odds of phenotypic resistance
- OR<1 the presence of a mutation associated with lower odds of phenotypic resistance

In our approach, LRs were used for objectively evaluating whether mutations were positively associated with phenotypic resistance. Using this rationale, the thresholds that are most commonly used in evidence-based medicine to grade MTBC mutations (Figure 1) were adopted.

In addition, p-values associated with the respective LR and OR scores were considered; the p-value can be interpreted as the probability of occurrence of a result as the one observed if the null hypothesis of no association is true. For the purposes of this evaluation, a p-value below 0.05 was designated as the threshold for indicating a nominally statistically significant difference between drug resistant and drug susceptible groups with respect to the presence of a mutation, whereas a p-value above 0.05 indicates that there is no evidence of a statistically significant difference between the groups. If the p-value associated with the LR and the OR was not significant (above 0.05), the association between the mutation and the phenotype was considered to be “indeterminate”. The calculated p-values provide the putative significance level of the association between the presence of a mutation and drug resistance. These nominal p-values (also referred as “uncorrected p-values” in the manuscript and related Supplementary material) may, however, be misleading given that multiple mutations are simultaneously evaluated for association with the drug resistance phenotype. Even under the null hypothesis of no association, when the p-values are uniformly distributed in the interval (0, 1), on average one in every 20 association tests will appear to be significant. From a clinical point of view, this means that some of the mutations will be incorrectly pointing towards drug resistance when they should actually not. For this reason we adopted a correction procedure known as the FDR (false discovery rate), which effectively confers statistical significance only to those associations that pass a more stringent threshold, adjusted based on the number of tests performed. The FDR-corrected p-values were used as the basis for the final significance level of association between the presence of a mutation and drug resistance. However, not all tests were counted for the purposes of this correction, as described below in “criteria used to classify genetic variants for the correction of the p-value”.

---

\[Pr(M|DR^+)\cdot Pr(DR^-|M)\]
Nonsense mutations (i.e. premature stop codons) and frameshift mutations in non-essential resistance genes can be predicted with confidence to cause a loss-of-function phenotype⁴⁻⁷. Therefore, in addition to analyzing these changes individually, these variants were pooled and treated as one class of changes. This rule was only applied to isolates in which indels or nonsense mutations occurred on their own (i.e. isolates with indels or nonsense mutations that coincided with further changes were not considered for this calculation). By contrast, in-frame indels in essential genes were only analyzed individually given that it was not possible to predict the effect of such variants a priori. Unless otherwise stated, indels are frameshift (whereas inframe mutations have been explicitly specified).

Essential and non-essential genes were defined according to the information reported in the Tuberculist database⁸, available at http://tuberculist.epfl.ch/ (version 2.6, Release 27. Accessed Nov 18th, 2016). Silent mutations (i.e. nucleotide changes that do not alter the amino acid sequence of the protein) were tracked but considered as wild-type in the analysis. For this reason, variants found to be in association with silent mutations were merged if also found not in combination with silent mutations. (e.g. the entries “S531L” and “S531L+G536G” were combined to form “S531L (merged entry)”).

The synonymous mutations that were reported in the included studies can be found in the following table and were considered as wild-type throughout the analysis.

| Gene | Synonymous mutations |
|------|----------------------|
| eis  | S48S, P90P           |
| gidB | A205A                |
| gyrA | A125A, L197L, L198L, T272T, A343A, A384A, E485E, L549L, I568I, A581A, V678V, G809G |
| gyrB | T625T                |
| inhA | G3G, L44L, L88L, S142S, G205G |
| mabA | L203L, G241G         |
| katG | D33D, E99E, A110A, G205G, E261E, K239K, P241P, A281A, A291A, T308T, T326T, K327K, T344T, L611L |
| mshA | G106G, D208D, A298A, C409C, L244L, L261L |
| pncA | F13F, A20A, A26A, L27L, A36A, A38A, A39A, A46A, G55G, K96K, S65S, S74S, G75G, T76T, G78G, G124G, D129D, C138C, G150G, E173E |
| rpoB | D184D, F514F, L521L, T525T, R529R, L533L, G534G, G536G, S539S, R542R, G544G, R548R, V550V, P552P, E562E, G566G, S576S, P657P, L859L, G957G, A1156A |
| rpsL | R12R, S17S, A23A, L24L, R30R, G32G, C34C, T39T, T41T, P45P, S47S, A48A, V52V, L57L, S59S, V61V, V63V, T64T, A65A, I67I, G71G, S78S, L81L, R83R, G84G, P91P, R94R, K121K |
| tlyA | L11L                |

**Criteria used to classify genetic variants for the correction of the p-value**

The following criteria were used to classify genetic variants for the correction of the p-value:

1. **to be included**: single variant to be considered for the correction
2. **to be excluded - variant already considered in merged entry for silent mutation**: variant observed in combination with silent mutations, thus a merged entry has been already considered for the correction
3. **to be excluded – other**: genetic variants showing uncertainties and/or not defined (rationale: we cannot use an undefined mutation as a standard variant)
4. **to be excluded - multiple mutations**: genetic variants observed in combination with other variants (rationale: these entries are at risk of bias as reported in our QUADAS analysis)
5. **to be excluded - WT or silent**: WT or entries showing only silent mutations (rationale: these entries do not represent variants. Silent mutations are at risk of bias as reported in our QUADAS analysis and have been considered as wild-type throughout the analysis)
6. **to be excluded – duplicated**: entries highlighted with a “(b)” in the “Mutation” column as described in the following
paragraph (rationale: entries at risk of bias).

7. considered as pooled: as previously mentioned, frameshift mutations and premature stop codon were only considered as pooled entries

The nominal p-value was used to identify mutations likely to be associated with drug resistance (putative Individual Confidence Values), whereas the corrected p-value was used for the final classification (Final Individual Confidence Values).

Data analysis

The following types of studies were included if sequencing and phenotypic DST data were available: randomized controlled trials, cohort studies, case-control studies, cross-sectional studies, case reports, animal studies, and in vitro research.

Types of specimens included: clinical MTBC isolates (i.e. in vitro mutants were excluded).

For sequencing data, the following definitions were considered:
- wild-type: wild-type refers to a strain, gene, or genetic characteristic that prevails among individuals in natural conditions, as distinct from an atypical or mutant type;
- mutant: mutant refers to any heritable change in the nucleotide sequence in a locus, gene, or genome;
- genetic patterns: combination of different mutations either within the same gene or in different genes.

Index testing for mutation detection: MTBC genomic mutations included in this review were detected and characterized by comparing MTBC nucleotide sequence data with a wild-type, standard reference sequence. Although a variety of sequencing technologies exist, Sanger sequencing, pyrosequencing, and next generation sequencing technologies were the most commonly utilized methods considered in this systematic review. M. tuberculosis H37Rv (accession number AL123456.2) was used as the reference for the sequence and annotation in this review, with the exception of rifampicin and fluoroquinolones, a different numbering system for relevant codons was used. For rifampicin (R), the E. coli codon numbering system was employed9,10 with the exception of 3 mutations outside the RRDR and reported using the MTB numbering system (H323Y, P206R, and Y314C), whereas for fluoroquinolones (moxifloxacin – MFX, ofloxacin – OFX, levofloxacin – LFX) we used the gyrase B numbering system proposed by Maruri et al11:

| Drug       | Gene | Region   | M. tuberculosis numbering system | Reference numbering system used | Reference   |
|------------|------|----------|----------------------------------|---------------------------------|-------------|
| R          | rpoB | N-term   | Cod. 167-172                     | Cod. 143-148                    | Campbell 200110 |
| R          | rpoB | Cluster I| Cod. 424-456                     | Cod. 505-537                    | Campbell 200110 |
| R          | rpoB | Cluster II| Cod. 481-494                    | Cod. 562-575                    | Campbell 200110 |
| R          | rpoB | Cluster III| Cod. 604-610                   | Cod. 684-610                    | Campbell 200110 |
| MFX/OFX/LFX| gyrB | QRDR     | Cod. 461-499                     | Cod. 461-499                    | Maruri 201211 |

The reference test for the detection of phenotypic resistance was culture-based phenotypic drug susceptibility testing (refer to Supplementary Material – Data Gathering Form for details). Accordingly, we considered 3 different datasets:
- WHO-endorsed liquid phenotypic drug susceptibility testing
- WHO-endorsed solid phenotypic drug susceptibility testing
- combined phenotypic drug susceptibility testing (derived by combining liquid + solid + WHO-endorsed not specified testing method).

Given the aforementioned limitations of the phenotypic drug susceptibility testing method, we used the following rules for the combined dataset for Z:

| Liquid DST | Wayne | Combined |
|------------|-------|----------|
| R          | R     | R        |
| n.a.       | R or S| Wayne result |
| R or S     | n.a.  | Liquid DST result |
| S          | S     | S        |
| S          | R     | R according to Wayne because of the limitations of liquid DST |
| R          | S     | S according to Wayne because liquid DST is more prone to false positives |
The amino acid substitution S100F in the gidB gene was not considered as a SNP because it represents an error in the earlier H37Rv reference sequences. Therefore, all isolates with this mutation were considered to be wild-type. Some studies reported mutations that fell just outside of the regions that were sequenced based on the methods sections of the studies in question (e.g. outside of the aforementioned rpoB clusters). The mutations were highlighted with a ‘(b)’ in the ‘Mutation’ column in Supplementary S7.1. They were analysed separately given that these could have been genuine mutations because sequencing primers are usually not immediately adjacent to the targeted regions (the primer sequences were not provided to confirm this).

The sensitivity and specificity of predicting phenotypic ofloxacin (OFX) and levofloxacin (LFX) resistance by sequencing were found to be independent of the phenotypic method used, whereas there were substantial differences in specificity for moxifloxacin (MFX) depending on whether liquid or solid DST was used as the reference method (data not shown). Results for OFX and LFX from both testing methods were therefore pooled whereas MFX results were analysed separately for each DST method. To maximise the number of isolates studied and thus increase statistical power, results for ethionamide (ETO) and prothionamide (PTO) were pooled.

The frequency of each mutation in DR and DS cases was determined and stratified according to i) the phenotypic testing method used as the reference (liquid or solid conventional DST) and ii) the geographical origin of the isolates in question. The denominator to calculate these frequencies could differ for different positions in the same gene, as illustrated in the following example:

- Study A reports sequencing results for 100 clinical isolates; target sequence: rifampicin resistance-determining region (RRDR: codons 505-537 of rpoB)
- Study B reports sequencing results for the entire rpoB gene for 50 clinical isolates

For the mutations in RRDR, the denominator would be 150 (because both the studies A and B are interroigated this regions), whereas the denominator for codons outside this range (i.e. codon >505 and >537) would be 50 (because only study B covered these positions).

Statistical analyses were done using the R statistical computing language. Sensitivity, specificity, positive, negative predictive values, p-value estimates, and 95% confidence intervals for each mutation and associated phenotype were calculated using the binomial test (null hypothesis: proportion = ½). Confidence intervals and p-values for the odds ratio (OR were calculated using the Fisher exact test (null hypothesis: odds ratio = 1), whereas for the LR, they were calculated using the melded binomial test (null hypothesis: likelihood ratio = 1). Performance characteristics (sensitivity, specificity, and diagnostic accuracy) of different categories of mutations were calculated using the same R script.

The number of isolates with multiple and silent mutations was likely underreported and this estimate should be considered as a risk of potential bias (refer to SUPPL4: QUADAS-2, Domain 2 risk assessment). A comprehensive analysis of multiple mutations could not be provided due the limitation of the data sets analyzed (refer to SUPPL4: QUADAS-2, Domain 2 risk assessment).

**Applying the grading system to special cases**

- For antimycobacterial drugs with multiple resistance genes “individual confidence value” (ICV) “indeterminate” values were over-ruled by the highest confidence score available. In example 1, the ICVs, for which rrs alone was sequenced, overruled the ICVs for both rrs and tlyA, the latter showing an “indeterminate” confidence value.
- For antimycobacterial drugs with multiple resistance genes “individual confidence value” (ICV) were prioritized for which the greatest number of resistance genes were sequenced (unless the ICVs were “indeterminate” as mentioned above). In examples 2-4, the ICVs, for which both katG and inhA were sequenced, overruled the ICVs for inhA alone, despite the fact that the latter had higher confidence values. The resulting “medium confidence values” (MCV) were calculated for each medium (i.e. liquid, solid, and composite). For antimycobacterial drugs with only a single resistance gene, ICV were by definition equal to the MCV.
- Where the ICV or MCV differed between media the highest confidence value was chosen as the “best confidence value” (BCV) (see cases 24 & 5). If data were only available for a single medium, the corresponding ICV or MCV was by definition equal to the BCV.
- Having assessed variants singly or in combination using ICV, MCV, and BCV, we also calculated equivalent “interpretative” values to extrapolate the data of single mutations to isolates with multiple variants (not considered for the correction of the p-value, and thus not provided with an ICV). A variant was assumed to be necessary and sufficient to cause resistance on its own. So for mutations that harboured multiple variants, the highest confidence value of the individual mutations replaced the confidence level of the combined mutations. This was done for individual media (i.e. by comparing ICVs for antimycobacterial drugs with only a single resistance gene, as shown in case 6, and by comparing MCVs for drugs with multiple resistance gens, as shown in case 7) to yield an “interpretative ICV” (iICV) and “interpretative MCV” (iMCV). Moreover, equivalent “interpretative BCVs” were generated using the BCV for an
overall assessment (case 8; please note that in this case the ICV corresponded to the MCV because a single resistance gene was considered). For isolates with single variants, confidence values were by definition identical to interpretative values (case 8). Indels and nonsense mutations for non-essential genes were considered as pooled for the p-value correction and represent a special case. Here the pooled confidence value overruled the value for an indel/nonsense mutation that occurred either alone or in combination (case 9), provided that pooled confidence was higher. The “interpretative BCVs” (iBCV), which excluded the non-causative mutations from Table 3, was used to calculate the sensitivities and specificities of the mutations (e.g. in Figure 2).
### Example cases:

| Case | Drug | Mutation | Medium | ICV  | MCV  |
|------|------|----------|--------|------|------|
| 1    | CM   | rrs c1402t + tlyA n.a. | Composite |      | indeter |
|      | CM   | rrs c1402t + tlyA WT | Composite |      |       |

| Case | Medium | Less informative Mutation | ICV | More informative Mutation | ICV | MCV | BCV |
|------|--------|--------------------------|-----|---------------------------|-----|-----|-----|
| 2    | H - composite | katG n.a. + inhA c-15t + mabA n.a. + furA n.a. + mshA n.a. | | katG WT + inhA c-15t + mabA n.a. + furA n.a. + mshA n.a. | | | |
| 3    | H - liquid | katG n.a. + inhA c-15t + mabA n.a. + furA n.a. + mshA n.a. | | katG WT + inhA c-15t + mabA n.a. + furA n.a. + mshA n.a. | | | |
| 4    | H - solid | katG n.a. + inhA c-15t + mabA n.a. + furA n.a. + mshA n.a. | | katG WT + inhA c-15t + mabA n.a. + furA n.a. + mshA n.a. | | | |

| Case | Drug | Gene | Mutation | Medium | Interpretative ICV |
|------|------|------|----------|--------|-------------------|
| 5    | R    | rpoB | L533P    | Composite |                  |
|      | R    | rpoB | L533P    | Liquid  | indeter          |
|      | R    | rpoB | L533P    | Solid   |                  |

| Case | Drug | Gene | Mutation | DST | ICV  | Interpretative ICV |
|------|------|------|----------|-----|------|-------------------|
| 6    | R    | rpoB | S531L    | Liquid |      |                   |
|      | R    | rpoB | L511P    | Liquid |      |                   |
|      | R    | rpoB | L511P+S531L | Liquid | n.a. |                   |

| Case | Drug | Gene | Mutation | DST  | MCV  | Interpretative MCV |
|------|------|------|----------|------|------|-------------------|
| 7    | H    | katG+inhA | katG S315N + inhA WT | Composite |      |                  |
|      | H    | katG+inhA | katG WT + inhA c-15t | Composite |      |                  |
|      | H    | katG+inhA | katG S315N + inhA c-15t | Composite | n.a. |                  |
| Case | Drug | Gene     | Mutation                      | Medium | ICV  | Interpretative ICV | BCV | Interpretative BCV |
|------|------|----------|-------------------------------|--------|------|--------------------|-----|--------------------|
|      |      |          | M515I                         | Composite | indeter | indeter           |     | indeter            |
|      |      |          | M515I                         | Solid   | indeter | indeter           |     | indeter           |
|      |      |          | M515I                         | Liquid  | indeter | indeter           |     | indeter           |
| 8    | R    | rpoB     | H526N                         | Composite |        |                    |     |                    |
|      |      |          | H526N                         | Solid   | indeter |                    |     | Indeter            |
|      |      |          | H526N                         | Liquid  |        |                    |     |                    |
|      |      |          | M515I+H526N                   | Composite | n.a.   | n.a.               |     |                    |
|      |      |          | M515I+H526N                   | Solid   | n.a.   | n.a.               |     | Indeter            |
|      |      |          | M515I+H526N                   | Liquid  | n.a.   | n.a.               |     |                    |

| Case | Drug | Gene | Mutation                      | Medium | ICV  | Interpretative ICV |
|------|------|------|-------------------------------|--------|------|--------------------|
| 9    | Z    | pncA | pooled frameshift and premature Stop | Composite |        |                    |
|      |      |      | V128ins                       | Composite | n.a. |                    |
|      |      |      | G24del                        | Composite |      |                    |
|      |      |      | A102T                         | Composite | n.a. | Indeter            |
|      |      |      | A102T+G105ins                 | Composite | n.a. |                    |
|      |      |      | T168I                         | Composite | indeter | Indeter       |
|      |      |      | W119Stop                      | Composite | n.a. |                    |
|      |      |      | W119Stop+T168I                | Composite | n.a. |                    |
References

1. Hayden SR, Brown MD. Likelihood ratio: A powerful tool for incorporating the results of a diagnostic test into clinical decisionmaking. *Ann Emerg Med* 1999; 33(5): 575-80.
2. Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. *Ann Intern Med* 1999; 130(12): 1005-13.
3. Brown MD, Reeves MJ. Evidence-based emergency medicine/skills for evidence-based emergency care. Interval likelihood ratios: another advantage for the evidence-based diagnostician. *Ann Emerg Med* 2003; 42(2): 292-7.
4. Scorpio A, Zhang Y. Mutations in *pncA*, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculous drug pyrazinamide in tubercle bacillus. *Nat Med* 1996; 2(6): 662-7.
5. DeBarber AE, Mduli K, Bosman M, Bekker LG, Barry CE. Ethionamide activation and sensitivity in multidrug-resistant Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* 2000; 97(17): 9677-82.
6. Okamoto S, Tamaru A, Nakajima C, et al. Loss of a conserved 7-methylguanosine modification in 16S rRNA confers low-level streptomycin resistance in bacteria. *Mol Microbiol* 2007; 63(4): 1096-106.
7. Maus CE, Plikaytis BB, Shinnick TM. Mutation of tlyA confers capreomycin resistance in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* 2005; 49(2): 571-7.
8. Lew JM, Kapopoulou A, Jones LM, Cole ST. TuberculList--10 years after. *Tuberculosis (Edinb)* 2011; 91(1): 1-7.
9. Telenti A, Imboden P, Marchesi F, et al. Detection of rifampicin-resistance mutations in *Mycobacterium tuberculosis*. *Lancet* 1993; 341(8846): 647-50.
10. Campbell EA, Korzheva N, Mustaev A, et al. Structural mechanism for rifampicin inhibition of bacterial rna polymerase. *Cell* 2001; 104(6): 901-12.
11. Maruri F, Sterling TR, Kaiga AW, et al. A systematic review of gyrase mutations associated with fluoroquinolone-resistant *Mycobacterium tuberculosis* and a proposed gyrase numbering system. *J Antimicrob Chemother* 2012; 67(4): 819-31.
12. Köser CU, Niemann S, Summers DK, Archer JA. Overview of errors in the reference sequence and annotation of *Mycobacterium tuberculosis* H37Rv, and variation amongst its isolates. *Infect Genet Evol* 2012; 12(4): 807-10.
13. Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing 2013.
14. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 1934; 26: 404-13.
15. Fisher RA. Confidence limits for a cross-product ratio. *Australian Journal of Statistics* 1962; 4: 41.
16. Fay MP, Proschan MA, Brittain E. Combining one-sample confidence procedures for inference in the two-sample case. *Biometrics* 2015; 71(1): 146-56.
Of the 95 studies included, the type of laboratory performing the testing was only known for 36%. The breakdown concerning the knowledge of the laboratories in provided in Figure 3.
The number of countries of collection and isolation for isolates is reported in Figure 4. The geographical representativeness reported refer only to information provided in the studies and does not reflect speculation based upon the publication of the reference. However, given that many studies lack information on the origin of samples, it is conceivable that the geographic distribution of collection is broader than reported.

Geographical origin of clinical strains was available for 1851 isolates (22.9% of the total strains considered for analysis). An additional 264 clinical isolates were reported to be collected across different countries and/or regions, however specific details at the country level were not available.
### Table S6.1: Clinical Isolates Stratified by Phenotypic DST Method (Critical Concentration DST) - Rifampicin

| Testing method                                                                 | Combined | Liquid | Solid |
|--------------------------------------------------------------------------------|----------|--------|-------|
| BACTEC MGIT 960 - 1 µg/mL                                                      | 2642      | 2642   | 0     |
| BACTEC 460TB or MGIT 960 - 2 or 1 µg/mL                                        | 541       | 541    | 0     |
| BACTEC 460TB - 2 µg/mL                                                        | 189       | 189    | 0     |
| Agar proportion - 1 µg/mL                                                      | 881       | 0      | 881   |
| LJ proportion - 40 µg/mL                                                       | 7151      | 0      | 7151  |
| Agar proportion - other                                                        | 53        | 0      | 53    |
| LJ absolute - 40 µg/mL                                                         | 172       | 0      | 172   |
| LJ proportion - other                                                          | 98        | 0      | 98    |
| not specified                                                                   | 1536      | 0      | 0     |
| BACTEC MGIT 960 - 1 µg/mL or LJ proportion - 40 µg/mL                         | 161       | 0      | 0     |
| **Tot**                                                                        | **13424** | **3372** | **8355** |
ISONIAZID (H), ETHIONAMIDE (ETO), PROTHIONAMIDE (PTO)

**FIGURE 5: PRISMA DIAGRAM OF THE DATABASE QUERY**

**FIGURE S6.6: OVERVIEW OF THE CRITERIA FOR EXCLUDING PAPERS (N= 523)**

Of the 127 studies included, the quality of the laboratory performing the testing was known for 28% (Figure 7).
The number of countries of collection and isolation for isolates is reported in Figure 8. The geographical representativeness reported refer only to information provided in the studies and does not reflect speculation based upon the publication of the reference. However, given that many studies lack information on the origin of samples, it is conceivable that the geographic distribution of collection is broader than reported.

**FIGURE 8: GLOBAL REPRESENTATIVENESS OF DATA SET**

Geographical origin of clinical isolates was available for 11528 isolates (93% of the total isolates considered for analysis).
| Country | Medium | DST | Method | Drug | Concentration | Resistance | Notes |
|---------|--------|-----|--------|------|---------------|------------|-------|
| WHO     |        |     |        |      |               |            |       |
| MS      |        |     |        |      |               |            |       |
| WHO     |        |     |        |      |               |            |       |
| MS      |        |     |        |      |               |            |       |

**TABLE S6.2: CLINICAL ISOLATES STRATIFIED BY PHENOTYPIC DST METHOD (CRITICAL CONCENTRATION DST)**
Of the 75 studies included, the type of laboratory performing the testing was only known for 33%. The breakdown concerning the knowledge of the laboratories in provided in Figure 11.
The number of countries of collection and isolation for isolates is reported in Figure 12. The geographical representativeness reported refer only to information provided in the studies and does not reflect speculation based upon the publication of the reference. However, given that many studies lack information on the origin of samples, it is conceivable that the geographic distribution of collection is broader than reported.

**Figure S6.12: Global representativeness of data set**

Geographical origin of clinical strains was available for 1171 isolates (13.4% of the total strains considered for analysis).
### Table S6.3: Clinical Isolates Stratified by Phenotypic DST Method (Critical Concentration DST)

#### gyrA

| Medium                | Phenotypic testing method                                         | MFX | OFX | LFX |
|-----------------------|-------------------------------------------------------------------|-----|-----|-----|
| unknown               | BACTEC MGIT 960 or LJ proportion - 2 µg/mL                       | 0   | 106 | 0   |
| unknown               | not specified                                                    | 0   | 78  | 0   |
| unknown               | proportion and absolute - not specified                          | 0   | 36  | 0   |
| unknown               | WHO recommended                                                  | 0   | 247 | 0   |
| liquid                | BACTEC 460 - 2 µg/mL                                             | 0   | 26  | 0   |
| liquid                | BACTEC 460 - not specified                                       | 0   | 53  | 0   |
| liquid                | BACTEC MGIT 960 - 10 µg/mL                                       | 0   | 4   | 0   |
| liquid                | BACTEC MGIT 960 - 2 µg/mL                                        | 0   | 1566| 4   |
| liquid                | BACTEC MGIT 960 - 2 and 10 µg/mL                                 | 0   | 21  | 0   |
| liquid                | BACTEC MGIT 960 - 0.5 µg/mL                                     | 22  | 0   | 0   |
| liquid                | BACTEC MGIT 960 - 0.125 µg/mL                                    | 14  | 0   | 0   |
| liquid                | BACTEC MGIT 960 - 0.5 µg/mL                                     | 14  | 0   | 0   |
| liquid                | BACTEC MGIT 960 - 1.5 µg/mL                                     | 0   | 0   | 12  |
| solid                 | 7H10 submerged-disk proportion method - 2 µg/mL                   | 0   | 102 | 0   |
| solid                 | 7H11 agar proportion - 2 µg/mL                                   | 0   | 306 | 0   |
| solid                 | 7H11 multiple proportion - 2 µg/mL                                | 0   | 10  | 10  |
| solid                 | LJ absolute - 10 µg/mL                                           | 0   | 133 | 0   |
| solid                 | LJ absolute - 2 µg/mL                                            | 0   | 769 | 0   |
| solid                 | LJ proportion - 2 µg/mL                                          | 0   | 2021| 226 |
| solid                 | 7H10 agar proportion - 0.5 µg/mL                                 | 152 | 0   | 0   |
| solid                 | 7H10 agar proportion - 1 µg/mL                                   | 170 | 0   | 42  |
| solid                 | 7H11 agar proportion - 0.5 µg/mL                                 | 40  | 0   | 0   |
| solid                 | 7H11 multiple proportion - 0.5 µg/mL                             | 10  | 0   | 0   |
| solid                 | LJ proportion - 0.75 µg/mL                                       | 68  | 0   | 0   |
| solid                 | 7H10 agar proportion - 2 µg/mL                                   | 3   | 433 | 0   |
| solid                 | LJ absolute - 1 µg/mL                                            | 0   | 0   | 93  |
| solid                 | LJ proportion - 10 µg/mL                                         | 0   | 0   | 62  |
| **Tot**               |                                                                  | 1019| 5911| 449 |

#### gyrB

| Medium                | Phenotypic testing method                                         | MFX | OFX | LFX |
|-----------------------|-------------------------------------------------------------------|-----|-----|-----|
| unknown               | not specified                                                    | 0   | 78  | 0   |
| unknown               | proportion and absolute - not specified                          | 0   | 36  | 0   |
| liquid                | BACTEC MGIT 960 - 10 µg/mL                                       | 0   | 4   | 0   |
| liquid                | BACTEC MGIT 960 - 2 µg/mL                                        | 0   | 861 | 0   |
| liquid                | BACTEC MGIT 960 - 2 and 10 µg/mL                                 | 0   | 21  | 0   |
| liquid                | BACTEC MGIT 960 - 0.25 µg/mL                                     | 120 | 0   | 0   |
| liquid                | BACTEC MGIT 960 - 1 µg/mL                                        | 150 | 0   | 0   |
| liquid                | BACTEC MGIT 960 - 0.125 µg/mL                                    | 14  | 0   | 0   |
| liquid                | BACTEC MGIT 960 - 0.5 µg/mL                                     | 14  | 0   | 0   |
| solid                 | 7H11 agar proportion - 2 µg/mL                                   | 0   | 72  | 0   |
| solid                 | 7H11 multiple proportion - 2 µg/mL                                | 0   | 10  | 10  |
| solid                 | LJ absolute - 10 µg/mL                                           | 0   | 26  | 0   |
| solid                 | LJ absolute - 2 µg/mL                                            | 0   | 111 | 0   |
| solid                 | LJ proportion - 2 µg/mL                                          | 0   | 1442| 11  |
| solid                 | 7H10 agar proportion - 0.5 µg/mL                                 | 149 | 0   | 0   |
| solid                 | 7H10 agar proportion - 1 µg/mL                                   | 170 | 0   | 42  |
| solid                 | 7H11 agar proportion - 0.5 µg/mL                                 | 40  | 0   | 0   |
| solid                 | 7H11 multiple proportion - 0.5 µg/mL                             | 10  | 0   | 0   |
| solid                 | LJ proportion - 0.75 µg/mL                                       | 68  | 0   | 0   |
| solid                 | 7H10 agar proportion - 2 µg/mL                                   | 0   | 417 | 0   |
| solid                 | LJ proportion - 1 µg/mL                                          | 0   | 0   | 93  |
| solid                 | LJ proportion - 10 µg/mL                                         | 0   | 0   | 62  |
| **Tot**               |                                                                  | 735 | 3078| 218 |
Of the 81 studies included, the type of laboratory performing the testing was only known for 33%. The breakdown concerning the knowledge of the laboratories in provided in Figure 15.
The number of countries of collection and isolation for isolates is reported in Figure 16. The geographical representativeness reported refer only to information provided in the studies and does not reflect speculation based upon the publication of the reference. However, given that many studies lack information on the origin of samples, it is conceivable that the geographic distribution of collection is broader than reported.

**Figure S6.16: Global representativeness of data set**

(Phenotypic DST on liquid media) (Wayne pyrazinamidase activity assay)

Geographical origin of clinical strains was available for 2252 isolates for liquid phenotypic DST (46% of the total strains considered for the analysis) and 1447 isolates for Wayne assay (83% of the total strains considered for analysis).

**Table S6.4: Clinical isolates stratified by phenotypic DST method (critical concentration DST) - pyrazinamide**

| Phenotypic testing method         | DR | DS |
|----------------------------------|----|----|
| 7H12B broth - 100 μg/mL          | 6  | 5  |
| 7H9 broth - 100 μg/mL            | 82 | 18 |
| BACTEC 460 - 100 μg/mL           | 534| 583|
| BACTEC MGIT960 - 100 μg/mL       | 1796| 1925|
| **tot**                          | **2418** | **2531** |
AMINOGLYCOSIDES (STREPTOMYCIN (S), CAPREOMYCIN (CM), AMIKACIN (AM), KANAMYCIN (KM))

**Figure S6.17: Results from Database Query**

![Diagram showing the process of study identification and selection]

- **Studies identified through PubMed search** ($n = 1956$)
- **Studies after duplicates removed** ($n = 423$)
- **Studies screened** ($n = 423$)
- **Studies excluded** ($n = 173$)
- **Full text studies excluded**
  - Other focus ($n = 20$)
  - Incomplete for dataset ($n = 76$)
  - Biochemistry/metabolomics/transcriptomics ($n = 8$)
  - Assay development/lab procedure ($n = 5$)
  - Genotyping/epidemiology ($n = 24$)
  - In vitro selected mutants/mutagenesis ($n = 4$)
  - Sample numbers not clearly reported ($n = 5$)
  - Not relevant for seq-DST correlation ($n = 4$)
  ($n = 146$)
- **Studies included** ($n = 104$)

**Figure S6.18: Overview of the Studies Included (N=104)**

![Pie chart showing the distribution of studies by laboratory type]

The number of countries of collection and isolation for isolates is reported in Figure 19. The geographical representativeness reported refer only to information provided in the studies and does not reflect speculation based upon
the publication of the reference. However, given that many studies lack information on the origin of samples, it is conceivable that the geographic distribution of collection is broader than reported.

**FIGURE S6.19: GLOBAL REPRESENTATIVENESS OF DATA SET**

AM

![AM map with color legend](image)

CM

![CM map with color legend](image)

KM

![KM map with color legend](image)
| Method                  | Medium | Drug | rs  | tlyA | ris | rpsL | gidB | tap | whib? |
|------------------------|--------|------|-----|------|-----|------|------|-----|-------|
| unknown                | Unknown| AM   | 28  | 29   | 0   | 0    | 0    | 0   | 0     |
| WHO endorsed           | Unknown| AM   | 44  | 44   | 43  | 0    | 0    | 0   | 0     |
| BACTEC 460             | Liquid | AM   | 1   | 7    | 0   | 0    | 0    | 1   | 7     |
| BACTEC MGIT 960        | Liquid | AM   | 11  | 8    | 0   | 0    | 0    | 11  | 8     |
| BACTEC MGIT 960 - 1 µg/mL | Liquid | AM   | 95  | 239  | 0   | 0    | 0    | 0   | 0     |
| BACTEC MGIT 960 - 1.0 µg/mL | Liquid | AM   | 4   | 34   | 4   | 7    | 4   | 0   | 27    |
| BACTEC MGIT 960 - 1.5 µg/mL | Liquid | AM   | 17  | 74   | 0   | 0    | 17  | 74   | 0     |
| Micobact - 1 µg/mL     | Liquid | AM   | 1   | 12   | 0   | 0    | 1   | 12   | 0     |
| 7H10 agar proportion - 4 µg/mL | Solid | AM   | 132 | 257  | 62  | 13   | 62  | 12   | 0     |
| 7H10 agar proportion - 4 µg/mL and 8 µg/mL | Solid | AM   | 84  | 155  | 0   | 0    | 0   | 0    | 0     |
| 7H10 agar proportion - 5 µg/mL | Solid | AM   | 37  | 0    | 37  | 0    | 0    | 37   | 0     |
| 7H10 agar proportion - 6 µg/mL | Solid | AM   | 28  | 10   | 0   | 0    | 0   | 0    | 0     |
| 7H11 agar proportion - 4 µg/mL | Solid | AM   | 30  | 20   | 0   | 0    | 0   | 0    | 0     |
| 7H11 agar proportion - 6 µg/mL | Solid | AM   | 50  | 0    | 0   | 0    | 0   | 0    | 0     |
| agar proportion - 20 µg/mL | Solid | AM   | 20  | 27   | 0   | 0    | 20  | 27   | 0     |
| LJ method; 40 µg/mL    | Solid  | AM   | 113 | 87   | 103 | 54   | 113 | 87   | 113  |
| LJ proportion - 8µg/mL | Solid  | AM   | 35  | 173  | 25  | 173  | 0   | 0    | 0     |
| LJ proportion - 40µg/mL | Solid  | AM   | 46  | 60   | 0   | 0    | 0   | 0    | 0     |
| LJ proportion method; 40 µg/mL | Solid | AM   | 16  | 3    | 0   | 0    | 0   | 0    | 0     |

| Method                  | Medium | Drug | rs  | tlyA | ris | rpsL | gidB | tap | whib? |
|------------------------|--------|------|-----|------|-----|------|------|-----|-------|
| unknown                | Unknown| CM   | 17  | 40   | 0   | 0    | 0    | 0   | 0     |
| WHO endorsed           | Unknown| CM   | 43  | 44   | 43  | 0    | 0    | 0   | 0     |
| BACTEC 460 - 1.25 µg/mL | Liquid | CM   | 26  | 0    | 26  | 0    | 0    | 0   | 0     |
| BACTEC 460 or BACTEC MGIT 960 - 1.25 µg/mL, 960: 3 µg/mL | Liquid | CM   | 10  | 91   | 0   | 0    | 10  | 91   | 0     |
| BACTEC MGIT 960        | Liquid | CM   | 11  | 8    | 0   | 0    | 0   | 11  | 8     |
| BACTEC MGIT 960 - 1.25 µg/mL | Liquid | CM   | 4   | 25   | 4   | 25   | 4   | 25   | 0     |
| BACTEC MGIT 960 - 2.5 µg/mL | Liquid | CM   | 57  | 94   | 7   | 4    | 7   | 94   | 0     |
| RENAM - 2.5 µg/mL      | Liquid | CM   | 0   | 21   | 0   | 0    | 0   | 21   | 0     |
| 7H10 agar proportion - 4 µg/mL | Solid | CM   | 66  | 86   | 66  | 86   | 66  | 86   | 0     |
| 7H11 agar proportion - 30 µg/mL | Solid | CM   | 82  | 678  | 82  | 678  | 0   | 0    | 0     |
| agar proportion - 40 µg/mL | Solid | CM   | 24  | 23   | 0   | 0    | 24  | 23   | 0     |
| LJ absolute - 40 µg/mL | Solid  | CM   | 13  | 0    | 0   | 0    | 0   | 0    | 0     |
| LJ proportion - 40 µg/mL | Solid  | CM   | 187 | 949  | 129 | 282  | 100 | 57   | 100  |
| LJ proportion method; 40 µg/mL | Solid | CM   | 11  | 41   | 0   | 0    | 11  | 41   | 0     |
| 7H10 agar proportion - 6 µg/mL | Solid | CM   | 22  | 16   | 0   | 0    | 0   | 0    | 0     |
| 7H10 agar proportion - 10 µg/mL | Solid | CM   | 155 | 289  | 119 | 259  | 56  | 18   | 14    |

| Method                  | Medium | Drug | rs  | tlyA | ris | rpsL | gidB | tap | whib? |
|------------------------|--------|------|-----|------|-----|------|------|-----|-------|
| BACTEC 460 - 5 µg/mL   | Liquid | KM   | 27  | 12   | 24  | 2    | 24  | 2    | 10    |
| BACTEC 460 or BACTEC MGIT 960 - 5 µg/mL, 960: 2.5 µg/mL | Liquid | KM   | 30  | 71   | 0   | 0    | 30  | 71   | 0     |
| BACTEC MGIT 960        | Liquid | KM   | 22  | 7    | 22  | 7    | 22  | 7    | 0     |
| BACTEC MGIT 960 - 2.5 µg/mL | Liquid | KM   | 2   | 4    | 2   | 4    | 2   | 4    | 0     |
| BACTEC MGBT 960 - 5 µg/mL | Liquid | KM   | 0   | 27   | 0   | 0    | 0   | 27   | 0     |
| RENAM - 2.5 µg/mL      | Liquid | KM   | 3   | 18   | 0   | 0    | 3   | 18   | 0     |
| 7H10 agar proportion - 4 µg/mL | Solid | KM   | 216 | 250  | 105 | 47   | 152 | 248  | 0     |
| 7H10 agar proportion - 5 µg/mL | Solid | KM   | 63  | 12   | 63  | 12   | 63  | 12   | 0     |
| 7H10 agar proportion - 6 µg/mL | Solid | KM   | 104 | 27   | 66  | 27   | 66  | 27   | 37    |
| 7H10 agar proportion - not specified | Solid | KM   | 14  | 29   | 0   | 0    | 14  | 29   | 0     |
| 7H11 agar proportion - 30 µg/mL | Solid | KM   | 0   | 0    | 0   | 0    | 0   | 0    | 0     |
| 7H11 agar proportion - 6 µg/mL | Solid | KM   | 50  | 0    | 0   | 0    | 0   | 0    | 0     |
| LJ absolute - 30 µg/mL | Solid  | KM   | 13  | 0    | 0   | 0    | 0   | 0    | 0     |
| LJ absolute - 40 µg/mL | Solid  | KM   | 20  | 86   | 20  | 86   | 20  | 86   | 0     |
| LJ method; 20 µg/mL    | Solid  | KM   | 10  | 0    | 0   | 0    | 10  | 0    | 0     |
| LJ proportion - 30 µg/mL | Solid  | KM   | 135 | 266  | 0   | 0    | 117 | 265  | 33   |
| LJ proportion - 40 µg/mL | Solid  | KM   | 110 | 47   | 110 | 47   | 110 | 47   | 47    |
| LJ proportion method; 20 µg/mL | Solid | KM   | 10  | 42   | 0   | 0    | 10  | 42   | 0     |
| Method                                      | Medium | Drug | rrs | tlyA | els | rpsL | gidB | tap | whiB |
|--------------------------------------------|--------|------|-----|------|-----|------|------|-----|------|
| BACTEC 460 or 7H10 agar proportion - BACTEC: 6 µg/mL, 7H10: 2 and 10 µg/mL | Unknown | S    | 65  | 50   | 0   | 0    | 0    | 70  | 59   |
| not specified                              | Unknown | S    | 115 | 100  | 0   | 0    | 0    | 115 | 100  |
| proportion - 2 µg/mL                       | Unknown | S    | 33  | 0    | 0   | 0    | 0    | 31  | 0    |
| proportion method                          | Unknown | S    | 0   | 0    | 0   | 0    | 0    | 7   | 10   |
| unknown                                    | Unknown | S    | 0   | 0    | 0   | 0    | 0    | 104 | 20   |
| BACTEC 460                                 | Liquid | S    | 54  | 16   | 0   | 0    | 0    | 44  | 16   |
| BACTEC 460 or BACTEC MGIT 960 - 460: 2-6 mg/ml; MGIT:1-4 mg/ml | Liquid | S    | 69  | 0    | 0   | 0    | 0    | 69  | 0    |
| BACTEC 12B                                 | Liquid | S    | 0   | 0    | 0   | 0    | 0    | 61  | 3    |
| BACTEC 460 - 10 µg/mL                      | Liquid | S    | 4   | 9    | 0   | 0    | 0    | 4   | 9    |
| BACTEC 460 - 2 µg/mL                       | Liquid | S    | 148 | 24   | 26  | 7    | 26   | 7   | 193  |
| BACTEC 460 - not specified                 | Liquid | S    | 5   | 16   | 0   | 0    | 0    | 5   | 16   |
| BACTEC 460 - 6 µg/mL                       | Liquid | S    | 102 | 46   | 0   | 0    | 0    | 104 | 46   |
| BACTEC MGIT 960                            | Liquid | S    | 13  | 6    | 0   | 0    | 0    | 13  | 6    |
| BACTEC MGIT 960 - 1 µg/mL                  | Liquid | S    | 76  | 23   | 10  | 1    | 30   | 1   | 212  |
| BACTEC MGIT 960 - 1 µg/mL; 4 µg/mL         | Liquid | S    | 31  | 11   | 31  | 11   | 31   | 11  | 0    |
| BACTEC MGIT 960 - not specified            | Liquid | S    | 121 | 108  | 0   | 0    | 0    | 121 | 108  |
| REMA - 1 µg/mL                             | Liquid | S    | 49  | 81   | 0   | 0    | 0    | 49  | 81   |
| 7H10 agar proportion - 2 µg/mL              | Solid  | S    | 69  | 10   | 50  | 6    | 0    | 61  | 12   |
| 7H10 agar proportion - not specified        | Solid  | S    | 73  | 108  | 0   | 0    | 0    | 73  | 108  |
| LJ absolute - 4 µg/mL                      | Solid  | S    | 38  | 1    | 0   | 0    | 0    | 66  | 0    |
| LJ method - 4 µg/mL                        | Solid  | S    | 37  | 19   | 0   | 0    | 37   | 19  | 37   |
| LJ proportion - 4 µg/mL                    | Solid  | S    | 359 | 237  | 0   | 0    | 0    | 465 | 213  |
| LJ proportion - 4 µg/mL and 8 µg/mL         | Solid  | S    | 40  | 57   | 0   | 0    | 0    | 40  | 57   |
| LJ proportion - not specified               | Solid  | S    | 0   | 0    | 0   | 0    | 0    | 7   | 3    |
| LJ proportion -4µg/mL                      | Solid  | S    | 27  | 100  | 0   | 0    | 0    | 27  | 100  |
| LJ proportion method - 4 µg/mL              | Solid  | S    | 38  | 14   | 0   | 0    | 0    | 38  | 14   |
Figure S6.20. Distribution of genetic variants in resistant (panel A, left) and susceptible (panel B, right) clinical isolates across the different drugs considered.
### Supplementary 11

**Table S11.1. Comparison of the list of mutations predictive of resistance described in this study and in Farhat et al.** Genetic variants common to both studies are highlighted in yellow. Association between mutation and drug resistance confirmed by False Discovery Rate corrected p-values and used for comparing performance characteristics are reported in bold. R: rifampicin; AM: amikacin; CM: capreomycin; ETO/PTO: ethionamide and/or prothionamide; H: isoniazid; KM: kanamycin; MXF: moxifloxacin; OFX/LFX: ofloxacin and/or levofloxacin; Z: pyrazinamide; S: streptomycin.

| Drug (phenotypic testing) | Gene | This study (high, moderate, and minimal confidence mutations) | Farhat et al 2016 (Minimum list of predictive variables) |
|---------------------------|------|---------------------------------------------------------------|-------------------------------------------------------|
| **First-line** R          | rpoB | F505V+D516Y, L511P, S512T, Q513H+L533P, Q513-F514ns, Q513K, Q513L, Q513P, F514dup, M515H+D516Y, D516A, D516F, D516G, D516G+L533P, D516ns, D516N, D156V, D162Y, Del N518, S522L, S522Q, H526C, H526D, H526F, H526L, H526N, H526R, H526Y, S531F, S531L, S531Q, S531W, S531Y, L533P, E572F, D562E |
| H                         | inhA-mabA, katG, embB, katA, ahpC | S315L, S315N, S315T, Frameshifts and premature Stop codons (Genomic locus not considered) | g-17l, c-18l, t-8g, S94A, S315T, S315N, S315R |
| **Second-line** MFX (group A) | gyrA | G88C, A90V, S91P, D94A, D94G, D94N, D94Y |
| OFX/LFX (group B)         | gyrA, gyrB | G88A, G88C, D89N, A90V, S91P, D94A, D94G, D94H, D94N, D94Y |
| AM                        | rrs | a1401g, g1484t (Genomic locus not considered) | a1401g, g1484t |
| KM                        | rrs, tlyA | c-14l, g-10a, c-12l, g-37t (Genomic locus not considered) | a1401g, c1402t, g1484t |
| S                         | rpoL, rrs, gidB | K43R, K43T, K88Q, K88R, T401A, a514, a514a, g462t, c513t, a514 |
| **Second-line** ETO/PTO (group C) | inhA-mabA, esha | c-15t, c-15t+H194T, c-15t+G49A |
| **Second-line** Z (group D) | pncA | t-12c, a-11g, t-7c, a3E, L106T, V7G, D88G, D88N, Q101P, Q101R, D12A, D12G, D12N, C14R, G71D, L19P, G24D, Y34D, A46V, K48T, D49G, D49N, H51Q, H51R, P45L, P54S, R54L, H57R, H57Y, F58L, S59P, P62L, P62Q, D63G, S65P, S67P, W66C, W68G, W68R, H71D, H71Q, H71R, H71Y, C72R, C76P, H82R, L83P, L85R, F94L, F94S, R96P, K96R, K96T, G97C, G97D, V103H, S104R, G108R, L116P, L116R, L120P, R123P, V125F, V125G, V128G, G132A, G132D, G132E, G132T, A134V, T135N, T135P, H137P, C138Y, V139A, V139G, V139L, Q141P, T142A, T142K, T142M, indel - R148ns (inframe), L151S, V155G, L159P, T160P, G162D, T168F, A171E, L172P, M175I, M175T, M175V, V180F, V180G, Pooled frameshifts and premature Stop codons |
| H                         | inhA-mabA | c-15t, c-15t+H194T, c-15t+G49A |

Farhat et al 2016 (Minimum list of predictive variables): g-17l, c-18l, t-8g, S94A, S315T, S315N, S315R, M306L, M306V, G406A, G406S, S927A, c-12t, A70T, G269S, V49G, (Drug not considered)

Farhat et al 2016 (Minimum list of predictive variables): g-17l, c-18l, t-8g, S94A, S315T, S315N, S315R, M306L, M306V, G406A, G406S, S927A, c-12t, A70T, G269S, V49G, (Drug not considered)
References

1. Farhat MR, Sultana R, Iartchouk O, et al. Genetic Determinants of Drug Resistance in *Mycobacterium tuberculosis* and Their Diagnostic Value. *Am J Respir Crit Care Med* 2016; **194**(5): 621-30.