

The 2021 WHO catalogue of *Mycobacterium tuberculosis* complex mutations associated with drug resistance: A new global standard for molecular diagnostics

Timothy M Walker<sup>1,2†</sup>, Paolo Miotto<sup>3†</sup>, Claudio U Köser<sup>4†</sup>, Philip W Fowler<sup>1</sup>, Jeff Knaggs<sup>1,5</sup>, Zamin Iqbal<sup>5</sup>, Martin Hunt<sup>1,5</sup>, Leonid Chindelevitch<sup>6</sup>, Maha Farhat<sup>7</sup>, Daniela Maria Cirillo<sup>3</sup>, Iñaki Comas<sup>8,9</sup>, James Posey<sup>10</sup>, Shaheed Vally Omar<sup>11</sup>, Timothy EA Peto<sup>1,12</sup>, Anita Suresh<sup>13</sup>, Swapna Uplekar<sup>13</sup>, Sacha Laurent<sup>13</sup>, Rebecca Colman<sup>13</sup>, Carl-Michael Nathanson<sup>14</sup>, Matteo Zignol<sup>14</sup>, Ann Sarah Walker<sup>1,12</sup>, The CRyPTIC Consortium, The Seq&Treat Consortium, Derrick W Crook<sup>1,12‡</sup>, Nazir Ismail<sup>14‡</sup> and Timothy C Rodwell<sup>15‡</sup>

†‡ Contributed equally

<sup>1</sup> Nuffield Department of Medicine, University of Oxford, Oxford, UK\*

<sup>2</sup> Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam\*

<sup>3</sup> San Raffaele Scientific Institute, Milano, Italy\*

<sup>4</sup> Department of Genetics, University of Cambridge, Cambridge, UK

<sup>5</sup> European Bioinformatics Institute, Hinxton, UK\*

<sup>6</sup> Imperial College London, UK

<sup>7</sup> Harvard Medical School, Boston, USA

<sup>8</sup> Biomedicine Institute of Valencia IBV-CSIC, Spain

<sup>9</sup> CIBER Epidemiology and Public Health, Madrid, Spain

<sup>10</sup> Centres for Disease Control and Prevention, Atlanta, USA\*

<sup>11</sup> National Institute for Communicable Diseases, Johannesburg, South Africa\*

<sup>12</sup> NIHR Oxford Biomedical Research Centre, Oxford, UK\*

<sup>13</sup> The Foundation for Innovative New Diagnostics, Geneva, Switzerland\*\*

<sup>14</sup> Global Tuberculosis Programme, World Health Organization, Geneva, Switzerland\*,\*\*

<sup>15</sup> Division of Pulmonary, Critical Care and Sleep Medicine, University of California, San Diego, USA\*,\*\*

\* CRyPTIC (<http://www.crypticproject.org>)

\*\* Seq&Treat (<https://www.finddx.org/at-risk-populations/seq-treat/>)

Other members of the CRyPTIC consortium and of the Seq&Treat consortium are listed at the end.

Correspondence to: T.M.W. ([timothy.walker@ndm.ox.ac.uk](mailto:timothy.walker@ndm.ox.ac.uk)) or T.C.R. ([trodwell@health.ucsd.edu](mailto:trodwell@health.ucsd.edu))

## **Abstract** (250 words)

### Background

Molecular diagnostics are considered the most promising route to achieving rapid, universal drug susceptibility testing for *Mycobacterium tuberculosis*. There is currently no WHO endorsed catalogue of mutations to serve as a global standard for interpreting molecular information for drug resistance prediction.

### Methods

A candidate gene approach was used to identify mutations as associated with resistance, or consistent with susceptibility, for 13 WHO endorsed anti-tuberculosis drugs. 38,215 *M. tuberculosis* isolates with paired whole-genome sequencing and phenotypic drug susceptibility data were amassed from 41 countries. Mutations were graded according to their odds ratio and associated false discovery rate corrected p-value, and their positive predictive value for phenotypic resistance. A series of expert rules were applied for final confidence grading of each mutation.

### Findings

15,667 associations were computed for 13,211 unique mutations linked to one or more drugs. 1,149/15,667 (7·3%) mutations were classified as associated with phenotypic resistance and 107/15,667 (0·7%) were deemed consistent with susceptibility. For rifampicin, isoniazid, ethambutol, fluoroquinolones, and streptomycin, the mutations' pooled sensitivity was >80%. Specificity was over 95% for all drugs except ethionamide (91·4%), moxifloxacin (91·6%) and ethambutol (93·3%). Only two resistance mutations were classified for bedaquiline, delamanid, clofazimine, and linezolid as prevalence of phenotypic resistance was low for these drugs.

### Interpretation

This first WHO endorsed catalogue of molecular targets for *M. tuberculosis* drug susceptibility testing provides a global standard for resistance interpretation. It will be expanded as data emerge on phenotypically resistant isolates, including for new and repurposed drugs.

### Funding

UNITAID, Wellcome, MRC, BMGF.

---

## **Introduction**

Due to disruptions related to the COVID-19 pandemic, it is estimated that in 2020 1.4 million fewer people received treatment for tuberculosis compared to 2019. The estimated additional 500,000 deaths over the previous year's total of 1.4 million will set the world back to levels of mortality levels not seen since 2010.<sup>1,2</sup> The diagnosis and appropriate treatment of patients with rifampicin resistance was already challenging pre-pandemic, with fewer than half of an estimated 500,000 patients benefitting.<sup>2</sup> The availability of several new treatment options and strategies for the first time in decades is an advance, but getting the right drugs to the right patients in time to positively impact outcomes is essential, and requires access to rapid and accurate diagnostics that meet the emerging needs.<sup>3,4</sup>

The World Health Organization (WHO) set an important but challenging target on universal drug susceptibility testing (DST), which includes testing for the new and repurposed drugs in the WHO revised definition of extensively drug-resistant tuberculosis.<sup>5</sup> Phenotypic DST for *Mycobacterium tuberculosis*, although still the reference standard for most drugs, can take over a month to complete and requires expensive, complex laboratory capacity. For many countries these challenges remain prohibitive. Genotypic approaches to susceptibility testing can be rapid, accurate, automated and cost-effective.<sup>6</sup> However, no WHO-endorsed catalogue of mutations for the interpretation of genotypic DST results has hitherto been available.

The combination of rapid and accurate diagnostic tools and supportive WHO policies can have significant global impact.<sup>2</sup> The WHO previously published a guide to *M. tuberculosis* next-generation sequencing data interpretation, adopting the findings from a previous systematic review of the literature.<sup>7,8</sup> However, that review did not cover new or repurposed drugs and relied on Sanger sequencing results, which meant that the genomic regions interrogated were inconsistent, increasing the chance of false associations. Gaps in knowledge thus remain. We describe the methods and results of a systematic analysis of a large, globally diverse set of isolates using whole-genome sequencing (WGS) data and accompanying DST, intended to generate more knowledge and create the first WHO endorsed catalogue of mutations for 13 anti-tuberculosis drugs.<sup>9</sup>

## **Methods**

### **Data sources**

Existing *M. tuberculosis* WGS data were collected worldwide from academic groups and consortia, reference laboratories, public health organizations and the published literature, along with associated phenotypic DST data. Data were accepted whether locally

representative or enriched for resistance. Genomically clustered samples were not excluded, provided these had been assayed independently.

### Phenotypic data

Both categorical (resistant/susceptible) and quantitative (minimum inhibitory concentrations (MICs)) phenotypic data were collected. Different DST methods and resistance-defining critical concentrations were accepted. To ensure comparability and optimise quality, phenotypes were categorised as follows: methods and critical concentrations currently endorsed by WHO (category 1); critical concentrations previously endorsed by WHO for those methods (category 2); methods or critical concentrations not currently endorsed by WHO (category 3). Results in downstream analyses were weighted accordingly (Supplementary Tables, S1).

Category 1 methods included Löwenstein-Jensen, Middlebrook 7H10, Middlebrook 7H11, and MGIT, using critical concentrations from current WHO DST guidelines.<sup>10-14</sup> Phenotypes derived from these media were classified as category 2 if critical concentrations used were outdated, or reported to have relied on WHO guidance without providing the concentration tested.<sup>15-17</sup> If critical concentrations were unknown, phenotypes were classified as category 3, along with MIC data obtained from Thermo Fisher Scientific broth microdilution plates developed for, and validated by, CRyPTIC, which were converted into categorical results using plate-specific epidemiological cut-offs.<sup>18</sup> Phenotypic results that did not fit categories 1-3 were excluded.

Where data from more than one phenotypic method were available for an isolate, phenotypes from category 1 were selected over category 2, and these in turn over category 3. Within each category, solid media were ranked above liquid media, which in turn had its own hierarchy, ranking MGIT over microscopic observation drug susceptibility (MODS) over CRyPTIC plates on the basis of historical WHO endorsements.

### Genotypic data

Only WGS data derived from Illumina sequencers were considered. Raw WGS data were processed by Clockwork,<sup>(19)</sup> a variant calling pipeline developed by CRyPTIC.<sup>20-21</sup> As a sanity check and means of measuring quality of the variant calls thereby obtained, 17 well characterised *M. tuberculosis* isolates with high quality polished single-contig hybrid PacBio and Illumina assemblies were used as controls.<sup>22</sup> These were added to the cohort at the start, and then their variant calls were compared with the truth assemblies following the methodology previously described.<sup>21</sup> Across these 17 samples, mean precision and recall was 99.8% and 94.7% respectively (after filters, and excluding the untrustworthy masked part of the genome).

### Identification of variants

For each drug, a set of candidate genes and corresponding promoter sequences with a high prior probability of being associated with resistance were identified by an expert panel based on the published literature (Table S2). The number of sequence positions not passing a pipeline filter was quantified for each gene for every isolate as a measure of likely local sequencing noise. Assuming a Poisson distribution, where the probability of seeing a given number of calls failing a filter in a given gene was  $<1\%$  (i.e. a comparatively large number), the phenotype associated with that gene was excluded from further consideration. Isolates with *katG* S315T or *rpoB* S450L variants and susceptible phenotypes to isoniazid (n=128) or rifampicin (n=118), respectively, were also excluded on the assumption that these discrepancies were likely due to sample mislabelling.<sup>23</sup>

An algorithm was then used to categorise the candidate sequences using methods similar to those previously described.<sup>24</sup> The approach reflects the definite defectives method from the field of group testing,<sup>25,26</sup> identifying bacterial isolates containing just a single genetic mutation among candidate genes, and associating this with the phenotype.

To maximise the number of isolates in which a mutation can be isolated as the only mutation among candidate genes, a series of pre-processing steps identified mutations consistent with phenotypic susceptibility, and masked these prior to analysis. This was based on the upper bound of the 95% confidence interval on the positive predictive value (PPV) for phenotypic resistance being  $<10\%$  for any given mutation. Synonymous mutations, loci at otherwise invariant sites for which no base could be called, and variants previously reported as phenotypically neutral were also masked.<sup>27</sup> Genes were also divided into two tiers for hierarchical analysis, with tier 1 sequences representing those considered by an expert panel to have a higher prior probability of association with resistance, analysed first (Table 3).<sup>9</sup> For isolates with no tier 1 genomic explanation for resistance, tier 2 sequences were analysed.

The algorithmic approach characterised mutations in two passes, as follows: Resistant ('R') if a mutation was identified as the only mutation (a 'solo') across candidate genes in at least one drug resistant isolate; Susceptible ('S') if the variant was only ever seen in susceptible isolates, or only ever in susceptible isolates when solo; or unknown ('U') when the variant was never seen solo and not exclusively found in susceptible isolates. Variants characterised as 'S' were then masked and the algorithm run a second time (pass 2), identifying additional mutations now exposed as solo and characterising these. Two-by-two tables were generated from the number of susceptible and resistant phenotypes with and without each characterised variant as solo, and odds ratios (OR) and corresponding p-values generated using Fisher's exact test. Benjamini-Hochberg corrections were used to assess

statistical significance using a false discovery rate (FDR) of 5%. PPV and binomial exact 95% confidence intervals were computed from the contingency tables.

The algorithm was run once for category 1 phenotypes, again for category 1 and 2 phenotypes combined, and a third time for all phenotypes together. To avoid perfect prediction for the *katG* S315T or *rpoB* S450L variants (for which any susceptible isolates were excluded assuming sample mislabelling, as described above), phenotypes from excluded isolates were added back in only to compute the OR and PPV for these variants for each iteration.

### Confidence grading

The ORs, PPVs and associated FDR corrected p-values and confidence intervals formed the basis for the confidence grading approach in which variants were assigned to one of five groups: 1) associated with resistance; 2) associated with resistance – interim; 3) uncertain significance; 4) not associated with resistance – interim; 5) not associated with resistance (i.e. ‘consistent with susceptibility’).<sup>28</sup> Mutations identified as solo with a category 1 or 2 phenotype (i.e. a WHO endorsed method) on at least five occasions; that had a 95% CI lower bound of  $\geq 0.25$  for the positive predictive value; and with an OR $>1$  and a significant FDR-corrected p-value, were classified as ‘associated with resistance’. A mutation was graded as ‘associated with resistance - interim’ if fewer than 5 of the solos were associated with a category 1 or 2 phenotype, or if the mutation was only identified as solo on the second pass of the algorithm.

Mutations were graded as ‘not associated with resistance’ if solos met the pre-processing criteria described above (95% confidence interval upper bound on the positive predictive value for phenotypic resistance  $<10\%$ ). All other mutations were graded as having ‘uncertain significance’.

Unlike many drugs which have hotspots in which resistance mutations cluster in a relevant gene, resistance to pyrazinamide can be conferred by a large number of individually infrequent mutations dispersed across *pncA*.<sup>28</sup> Confidence grading criteria established for other drugs excluded most of these mutations and were therefore relaxed. Mutations present as solo in *pncA* in at least two resistant isolates and with  $\geq 50\%$  PPV were classified as ‘associated with resistance – interim’, whereas those with a PPV  $<40\%$  (and upper bound 95% CI  $<75\%$ ) were classified as ‘not associated with resistance – interim’.

Finally, a set of expert rules were applied to mutations of ‘uncertain significance’ whereby any nonsynonymous mutation in the rifampicin resistance determining region of *rpoB*, and any premature stop codon or insertion/deletion in *ethA*, *gid*, *katG* or *pncA* was interpreted as ‘associated with resistance – interim’.<sup>8,28</sup> Mutations not already graded, but for which there was previous guidance from WHO, were graded according to that external

evidence, with an ‘interim’ caveat unless these were so-called borderline mutations in the rifampicin resistance determining region, or unless there was more recent evidence in the literature to suggest previous WHO guidance should be revised (see detailed methods in WHO document).<sup>9</sup>

### Role of the funding source

The funders of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

### Findings

We analysed 41,137 *M. tuberculosis* isolates with phenotypic and WGS data from 41 countries across six continents. Thirty countries contributed data on more than 100 isolates and ten countries contributed more than 1,000 isolates. 38,215 isolates passed quality control steps and were included in the final analysis.

Not all isolates had phenotypic DST results for all drugs. Data on first-line drugs, isoniazid, rifampicin, ethambutol and pyrazinamide were most common. Phenotypic data on new and repurposed drugs, bedaquiline, delamanid, clofazimine and linezolid, were least common and almost exclusively derived from CRyPTIC plates (Table 1).<sup>(18)</sup> The prevalence of resistance to first-line drugs ranged from 14% (pyrazinamide) to 35% (isoniazid). The prevalence of resistance to new (bedaquiline, delamanid) and repurposed (linezolid, clofazimine) drugs was lower than for other drugs ( $\leq 1.2\%$ ).

15,667 associations were computed for 13,211 unique mutations relevant to one or more of 13 drugs. 1,149/15,667 (7.3%) mutations were graded as group 1 or 2, and 107/15,667 (0.7%) were graded as group 4 or 5. The majority were graded as ‘uncertain significance’ (group 3) (Figure 1; Table S3). All group 1 and 2 mutations were derived from tier 1 sequences (Table S3), and for most drugs from only one or two genes (Table S4). Except for *inhA* promoter mutations, all upstream group 1 and 2 mutations were within 12bp of a start codon. Although there were many group 3 mutations, individually these were seen far less frequently than the comparatively small number of group 1 or 2, or group 4 or 5 mutations (Figure 2). For pyrazinamide, only 7% of isolates contained a mutation of uncertain significance, climbing to 42% for bedaquiline (Table S5).

As no independent dataset was available to test the catalogue, sensitivity and specificity were assessed by predicting phenotypic resistance for the same data from which the catalogue was derived. For rifampicin, isoniazid, ethambutol, fluoroquinolones, and streptomycin, the mutations’ pooled sensitivity was  $>80\%$ . Specificity was over 95% for all drugs except ethionamide (91.4%), moxifloxacin (91.6%) and ethambutol (93.3%) (Figure 3).

In most cases the contribution of expert rules to sensitivity was small. For isoniazid, 10,978/12,199 (90%) resistant isolates contained one of five data-derived resistance mutations, with an additional 148 (1.2%) isolates subject to an expert rule. For rifampicin 9,047/9,868 (91.7%) resistant isolates contained one of the 23 data-derived resistance mutations, with 207 (2.1%) additional isolates subject to expert rules (Figure 3; Table S3). For drugs where resistance can be caused by loss of function mutations in non-essential genes, the expert rules played a greater role. For pyrazinamide, 284/2,329 (12.2%) resistant isolates were subject to an expert rule, and for ethionamide 530/2,965 (17.9%) (Table S6).

Only one mutation was classified as a group 1 resistance associated mutation for linezolid and one group 2 mutation for delamanid. No resistance mutations were identified from the data for bedaquiline or clofazimine. Figure 4 shows how the correspondingly low sensitivity relates to the number of isolates analysed for each drug and to the prevalence of phenotypic resistance among those isolates. Not only were fewer isolates analysed for the novel use drugs, the prevalence of resistance was markedly lower, between 0.9%-1.2%, and  $\leq 2\%$  even among isolates resistant to rifampicin and isoniazid. For legacy-use drugs it was between 7.6%-35.4%.

All genomic data are available here (*in the process of being uploaded of ENA*) and associated phenotypic data in Table S7.

## **Discussion**

This catalogue represents the first WHO endorsed list of genomic mutations associated with drug resistance, or consistent with susceptibility, in *M. tuberculosis*.<sup>9</sup> It is derived from the largest, globally sourced dataset of *M. tuberculosis* genome sequences and associated phenotypes published to date. This catalogue provides a common starting point and serves as a public resource for a wide array of users from TB reference laboratories, to molecular diagnostics developers, to surveillance programmes.

The approach adopted to classify mutations combined three previously published schemes that were refined during several rounds of consultation with an international panel of clinicians and researchers.<sup>8,24,28</sup> The analysis was designed to be deliberately stringent to minimise the chances of having to reverse the grading of any variants in the future. There will therefore be mutations described in the literature, including in known drug resistance genes for new and repurposed drugs, that are not in the catalogue (e.g. *rrl* for linezolid).<sup>3,29,30</sup>

This work was not designed as a systematic review of the literature, although decisions by WHO, based on the literature, were incorporated unless more recent literature contradicted these.<sup>9</sup> Following these precedents, a series of expert rules intended to cover all possible mutations that share functional features, such as premature stop codons, insertions or deletions resulting in the loss of function of non-essential genes such as *katG* or *pncA* were



used.<sup>8</sup> Although the rules are evidence-based, it is possible that exceptions exist. For instance, not all nonsense mutations may result in a loss of function. Expert rules could also have been applied to other non-essential genes relevant to other drugs, but we chose a stringent approach. Rules may need updating as new data accumulate.

While large numbers of isolates can help overcome the variability introduced by random error in measurements, big data remain susceptible to systematic error such as from changes in recommended critical concentrations over time, as happened for the fluoroquinolones in 2018.<sup>12</sup> The phenotypic data here were largely derived from WHO endorsed methods, albeit from a variety of media with varying critical concentrations. The broth microdilution plates by the CRyPTIC consortium are not WHO-endorsed, but contributed phenotypes for over 20% of these isolates, including the vast majority of data for the new and repurposed drugs.<sup>18</sup> The hierarchical prioritization of phenotypes was an attempt to manage the diversity in phenotypic methods.

Two tiers of candidate genes were analysed, with mainly canonical targets in tier 1 and more recently identified genes in tier 2.<sup>31,32</sup> No mutations outside of tier 1 genes and their promoters were classified as associated with resistance in this analysis. It could be that tier 2 variants were either too rare, or that the mutations in these genes result in smaller increases in the MIC, producing a more variable binary phenotype that failed to meet the established thresholds.

The specificity of graded mutations was low for some drugs (Figure 3), which is likely due to a combination of four factors.<sup>33</sup> First, inappropriately high critical concentrations between 2014-18 played a major role for moxifloxacin, despite our hierarchical approach prioritising phenotypic DST results.<sup>10</sup> Second, some resistance mutations only confer modest MIC increases, which means that their MIC distributions overlap with that of genuinely susceptible isolates. For these mechanisms, phenotypic DST, even using the correct critical concentration, is not a reliable confirmatory method, as recognised by WHO's endorsement of a composite reference standard for rifampicin.<sup>11</sup> Third, epistasis could have played a role for amikacin and potentially bedaquiline and clofazimine.<sup>34</sup> Finally, the expert rules may have overcalled resistance in some cases, as outlined above. It should also be noted that both sensitivity and specificity might be overestimated since these were assessed on the same data from which mutations were graded.

Despite these limitations, progress is to be expected as more resistant isolates are collected. This is especially important for new and repurposed drugs. Diminishing returns are however to be expected from analysing many more isolates resistant to legacy-use drugs, especially where the very major error rate (the gap between observed sensitivity and 100%) starts to overlap with the expected rate of phenotypic or sample labelling error.<sup>23</sup>

There are national TB programmes that already use WGS in place of phenotypic testing to direct the use of first-line drugs,<sup>23</sup> but further work is required to expand the catalogue for all drugs, and meet the desired target product profiles for molecular DST.<sup>7,33</sup> This should focus on collections with more phenotypic resistance to new and repurposed drugs, and could involve *in vivo* and *in vitro* selection experiments. Future analyses should also focus on the association between mutations, and combinations of mutations, with MICs.<sup>32,33</sup> Such data would help facilitate the tailoring of individual drug doses based on molecular diagnostics. The WHO plans to regularly update the catalogue and will endeavour to incorporate these advances and address existing gaps to strengthen the public health response. The effort will depend on future contributions from researchers, funding agencies and data custodians to this global effort.

## **Research in Context**

### **Evidence before this study**

Much work has been undertaken to describe molecular causes of resistance to anti-tuberculosis drugs over the past three decades. Available tools have changed over the years and whole-genome sequencing technology has emerged and taken over from PCR based techniques as the dominant approach to large scale genomic analyses. There have been various independent efforts to consolidate research findings into lists of mutations associated with drug resistance. The TBDreamDB was an early attempt and the systematic review of the literature by the ReSeqTB consortium was a more recent effort. The findings of the latter work were reflected in a WHO guide to the interpretation of *M. tuberculosis* WGS data. However, there was no globally endorsed standard of *M. tuberculosis* mutations and interpretation for national TB programmes or industry to refer to in the design of their services or products.

### **Added value of this study**

The diverse nature of the data accumulated over the past decades using different technologies and different platforms has made comparability challenging. This study leveraged established knowledge to define a set of candidate genes for each drug and sought to assemble as large a dataset as possible, from as many countries as possible, to perform a *de novo* analysis of all the available WGS data and associated phenotypes. The result is the first WHO endorsed catalogue of mutations associated with drug resistance and consistent with drug susceptibility that constitutes an international reference point for national TB programmes and assay manufacturers.

### **Implication of all the available evidence**

The WHO catalogue is a deliberately conservative effort and is not exhaustive. The weight of evidence supporting each mutation in the catalogue should however provide confidence to the TB community. Great accuracy has been achieved when predicting drug resistance in the past, but not every mutation in previous catalogues has necessarily been as well supported as is the case here. Having a WHO endorsed mutation list should support the field of molecular drug susceptibility testing in *M. tuberculosis* by providing all countries a template for the interpretation of their data. WHO plans to update the catalogue on a regular basis, and as more data are generated as a consequence, there are expected to be more data to feed into future analyses, hopefully initiating a virtuous circle.

## **Figure Legends**

**Figure 1:** The number of Tier 1 mutations in each group for each drug (no Tier 2 mutations were graded as group 1 or 2 so these are not shown for any groups). To reflect the minimum number of isolates a mutation must have been seen in for it to be graded as group 1 or 2, the mutations counted here are those that were seen in 5 or more isolates. The exceptions are mutations relevant to pyrazinamide (counted if seen in 2 or more isolates) and mutations subject to an expert rule (counted if seen in any number of isolates). The actual number of Tier 1 group 3 mutations ('Uncertain significance'), regardless of the frequency with which these were seen, is written within each grey bar. Group 4 mutations graded as such by an expert rule are not shown separately here; see supplementary table S3 for details on those. \*All expert rule mutations were group 2, with the exception of rpoB borderline mutations.

**Figure 2:** Number of isolates per mutation, by group and drug. Box and whisker plots exclude outside values. Mutations graded by expert rules are not shown as the number of isolates such mutations are seen in is not relevant to their classification. (Group 1 = Associated with Resistance; Group 2 = Associated with Resistance – interim; Group 3 = Uncertain significance; Group 4 = Not associated with Resistance – interim; Group 5 = Not Associated with Resistance)

**Figure 3:** Sensitivity and specificity are shown for all drugs. Sensitivity is represented by bars going upwards from zero, and specificity by bars going downwards from 100. As in figure 1, red and orange represent group 1 (Associated with Resistance) and group 2 (Associated with Resistance – Interim) mutations respectively. Mutations subject to expert rules are separated from their group and shown independently in light blue. The colour scheme shows the incremental sensitivity gained, and corresponding specificity lost, by expanding the catalogue to include first group 1 and then group 2 mutations, in each case without the use of expert rules, and then adding in the expert rules. With the exception of borderline rpoB mutations, all mutation subject to an expert rule were graded as group 2. Confidence intervals are for the total effect of all mutations shown.

**Figure 4:** Total number of isolates for each drug plotted against sensitivity. Each drug is represented by a coloured circle that is weighted by the prevalence of phenotypic resistance to that drug in this dataset. The centre of each circle shows the intersection between values on the X and Y axes. Group 1 = Associated with Resistance; Group 2 = Associated with Resistance – interim

**Table 1.** Summary of phenotypic DST data reported by drug; number and percentage of the isolates reported with resistant (R) phenotypes. (Table copied from <https://www.who.int/publications/i/item/9789240028173>)

Drug	Level of Support for CC	Total # Isolates	Total # Resistant	Percentage R (95% CI)
Rifampicin	WHO CURRENT	4,107	1,387	33.8% (32.3-35.2%)
	WHO CURRENT+PAST	27,063	6,736	24.9% (24.4-25.4%)
	ALL	34,375	9,868	28.7% (28.2-29.2%)
Isoniazid	WHO CURRENT	14,252	3,657	25.7% (24.9-26.4%)
	WHO CURRENT+PAST	26,727	8,440	31.6% (31.0-32.1%)
	ALL	34,437	12,199	35.4% (34.9-35.9%)
Ethambutol	WHO CURRENT	11,028	1,307	11.9% (11.3-12.5%)
	WHO CURRENT+PAST	23,706	3,615	15.2% (14.8-15.7%)
	ALL	30,708	4,900	16.0% (15.5-16.4%)
Pyrazinamide	WHO CURRENT	8,416	851	10.1% (9.5-10.8%)
	WHO CURRENT+PAST	15,903	2,329	14.6% (14.1-15.2%)
	ALL*	15,902	2,329	14.6% (14.1-15.2%)
Levofloxacin	WHO CURRENT	2,407	194	8.1% (7.0-9.2%)
	WHO CURRENT+PAST	10,305	2,019	19.6% (18.8-20.4%)
	ALL	18,277	3,108	17.0% (16.5-17.6%)
Moxifloxacin	WHO CURRENT	164	12	7.3% (3.8-12.4%)
	WHO CURRENT+PAST	6,904	1,094	15.8% (15.0-16.7%)
	ALL	13,351	1,869	14.0% (13.4-14.6%)
Bedaquiline	WHO CURRENT	0	0	NA
	WHO CURRENT+PAST	88	3	3.4% (0.7-9.6%)
	ALL	8,321	73	0.9% (0.7-1.1%)
Linezolid	WHO CURRENT	72	0	0.0% (0.0-5.0%)
	WHO CURRENT+PAST	1,131	9	0.8% (0.4-1.5%)
	ALL	11,018	123	1.1% (0.9-1.3%)
Clofazimine	WHO CURRENT	0	0	NA
	WHO CURRENT+PAST	3,635	23	0.6% (0.4-0.9%)
	ALL	10,179	125	1.2% (1.0-1.5%)
Delamanid	WHO CURRENT	0	0	NA
	WHO CURRENT+PAST	89	2	2.2% (0.3-7.9%)
	ALL	7,778	82	1.1% (0.8-1.3%)
Amikacin	WHO CURRENT	1,015	60	5.9% (4.5-7.5%)
	WHO CURRENT+PAST	8,040	664	8.3% (7.7-8.9%)
	ALL	16,978	1,288	7.6% (7.2-8.0%)
Streptomycin	WHO CURRENT	1,577	144	9.1% (7.8-10.7%)

	WHO CURRENT+PAST	9,043	2,562	28.3% (27.4-29.3%)
	ALL	13,984	4,635	33.1% (32.4-33.9%)
Ethionamide	WHO CURRENT	45	17	37.8% (23.8-53.5%)
	WHO CURRENT+PAST	2,184	884	40.5% (38.4-42.6%)
	ALL	13,918	2,965	21.3% (20.6-22.0%)

\* The 'ALL' dataset has one phenotype fewer than the WHO current and past dataset. This is because a whole strain was removed at the stage when category 3 phenotypes were added. One strain had a category 3 isoniazid phenotype, with a susceptible isoniazid phenotype and a katG S315T mutation, but a phenotypes from a higher category for pyrazinamide. The pyrazinamide phenotype was therefore included as part of the WHO current and past dataset. When the isoniazid phenotype was added, and the whole strain was removed, including the pyrazinamide phenotype.

## Acknowledgements – funders

This work was supported by Wellcome Trust/Newton Fund-MRC Collaborative Award (200205/Z/15/Z); and Bill & Melinda Gates Foundation Trust (OPP1133541); and supported by a grant from Unitaid. This work was funded by National Institutes of Health/ National Institute of Allergy and Infectious Disease (NIH/NIAID R01 AI55765 to MRF). J.A. has received funding from NIH (R01 AI130058). The NIH/NIAID had no role in study design, data collection, and interpretation, or the decision to submit the work for publication. Oxford CRyPTIC consortium members are funded/supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC), the views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health, and the National Institute for Health Research (NIHR) Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance, a partnership between Public Health England and the University of Oxford, the views expressed are those of the authors and not necessarily those of the NIHR, Public Health England or the Department of Health and Social Care. J.M. is supported by the Wellcome Trust (203919/Z/16/Z). Z.Y. is supported by the National Science and Technology Major Project, China Grant No. 2018ZX10103001. K.M.M. is supported by EMBL's EIPOD3 programme funded by the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska Curie Actions. T.C.R. is funded in part by funding from Unitaid Grant No. 2019-32-FIND MDR. R.S.O. is supported by FAPESP Grant No. 17/16082-7. L.F. received financial support from FAPESP Grant No. 2012/51756-5. B.Z. is supported by the National Natural Science Foundation of China (81991534) and the Beijing Municipal Science & Technology Commission (Z201100005520041). N.T.T.T. is supported by the Wellcome Trust International Intermediate Fellowship (206724/Z/17/Z). G.T. is funded by the Wellcome Trust. R.W. is supported by the South African Medical Research Council. J.C. is supported by the Rhodes Trust and Stanford Medical Scientist Training Program (T32 GM007365). T.C. has received grant funding and salary support from US NIH, CDC, USAID and Bill and Melinda Gates Foundation. I.C. laboratory is supported by ERC CoG 101001038 and Ministerio de Ciencia project PID2019-104477RB-I00 L.G. was supported by the Wellcome Trust (201470/Z/16/Z), the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number 1R01AI146338, the GOSH Charity (VC0921) and the GOSH/ICH Biomedical Research Centre ([www.nihr.ac.uk](http://www.nihr.ac.uk)). A.L. is supported by the National Institute for Health Research (NIHR) Health Protection Research Unit in Respiratory Infections at Imperial College London. J.R. is funded by MinCiencias, Colombia (No. 221389666216 CT-783-2018). S.G.L. is supported by the Fonds de Recherche en Santé du Québec. C.N. is funded by Wellcome Trust Grant No. 203583/Z/16/Z. A.V.R. is supported by Research Foundation Flanders (FWO) under Grant No. G0F8316N (FWO Odysseus). G.M. was supported by the Wellcome Trust (098316, 214321/Z/18/Z, and 203135/Z/16/Z), and the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation (NRF) of South Africa (Grant No. 64787). The funders had no role in the study design, data collection, data analysis, data interpretation, or writing of this report. The opinions, findings and conclusions expressed in this manuscript reflect those of the authors alone. A.B. is funded by the NDM Prize Studentship from the Oxford Medical Research Council Doctoral Training Partnership and the Nuffield Department of Clinical Medicine. L.C. acknowledges funding from the MRC Centre for Global Infectious Disease Analysis (reference MR/R015600/1), jointly funded by the UK Medical Research Council (MRC) and the UK Foreign, Commonwealth & Development Office (FCDO), under the MRC/FCDO Concordat agreement and is also part of the EDCTP2 programme supported by the European Union. D.J.W. is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (Grant No. 101237/Z/13/B) and by the Robertson Foundation. A.S.W. is an NIHR Senior Investigator. T.M.W. is a Wellcome Trust Clinical Career Development Fellow (214560/Z/18/Z). C.U.K. is a research associate at Wolfson College and visiting scientist at the Department of Genetics, University of Cambridge. C.U.K. received an observership by the European Society of Clinical Microbiology and Infectious Diseases. I.C. is supported by ERC CoG 101001038 and

Ministerio de Ciencia project PID2019-104477RB-I00. A.S.L. is supported by the Rhodes Trust. M.G.W is supported by the National Institute for Health Research (NIHR) Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at Imperial College London in partnership with Public Health England (PHE) in collaboration with, Imperial Healthcare Partners, University of Cambridge and University of Warwick. R.J.W. receives funding from the Francis Crick Institute which is supported by Wellcome Trust, (FC0010218), UKRI (FC0010218), and CRUK (FC0010218). The computational aspects of this research were supported by the Wellcome Trust Core Award Grant Number 203141/Z/16/Z and the NIHR Oxford BRC. Parts of the work were funded by the German Center of Infection Research (DZIF). The Scottish Mycobacteria Reference Laboratory is funded through National Services Scotland. The Wadsworth Center contributions were supported in part by Cooperative Agreement No. U60OE000103 funded by the Centers for Disease Control and Prevention through the Association of Public Health Laboratories and NIH/NIAID grant AI-117312. Additional support for sequencing and analysis was contributed by the Wadsworth Center Applied Genomic Technologies Core Facility and the Wadsworth Center Bioinformatics Core. SYNLAB Holding Germany GmbH for its direct and indirect support of research activities in the Institute of Microbiology and Laboratory Medicine Gauting. N.R. thanks the Programme National de Lutte contre la Tuberculose de Madagascar.

### **Conflicts of interest**

C.U.K. is a consultant Becton Dickinson, the Foundation for Innovative New Diagnostics and the TB Alliance. C.U.K. is collaborating with PZA Innovation and Thermo Fisher Scientific. C.U.K. worked as a consultant for QuantuMDx, the Stop TB Partnership, the World Health Organization (WHO) Global TB Programme and the WHO Regional Office for Europe. C.U.K. gave a paid educational talk for Oxford Immunotec. Hain Lifescience covered C.U.K.'s and accommodation to present at a meeting. C.U.K. is an unpaid advisor to BioVersys and GenoScreen. E.R. is employed by Public Health England and holds an honorary contract with Imperial College London. I.F.L. is Director of the Scottish Mycobacteria Reference Laboratory. S.N. receives funding from German Center for Infection Research, Excellenz Cluster Precision Medicine in Chronic Inflammation, Leibniz Science Campus Evolutionary Medicine of the LUNG (EvoLUNG)tion EXC 2167. P.S. is a consultant at Genoscreen. T.R. is funded by NIH and DoD and receives salary support from the non-profit organization FIND. T.R. is a co-founder, board member and shareholder of Verus Diagnostics Inc, a company that was founded with the intent of developing diagnostic assays. Verus Diagnostics was not involved in any way with data collection, analysis or publication of the results. T.R. has not received any financial support from Verus Diagnostics. UCSD Conflict of Interest office has reviewed and approved T.R.'s role in Verus Diagnostics Inc. T.R. is a co-inventor of a provisional patent for a TB diagnostic assay (provisional patent #: 63/048.989). T.R. is a co-inventor on a patent associated with the processing of TB sequencing data (European Patent Application No. 14840432.0 & USSN 14/912,918). T.R. has agreed to "donate all present and future interest in and rights to royalties from this patent" to UCSD to ensure that he does not receive any financial benefits from this patent. S.S. is working and holding ESOPs at HaystackAnalytics Pvt. Ltd. (Product: Using whole genome sequencing for drug susceptibility testing for Mycobacterium tuberculosis). G.F.G. is listed as an inventor on patent applications for RBD-dimer-based CoV vaccines. The patents for RBD-dimers as protein subunit vaccines for SARS-CoV-2 have been licensed to Anhui Zhifei Longcom Biopharmaceutical Co. Ltd, China.

### **Wellcome Trust Open Access**

*This research was funded in part, by the Wellcome Trust/Newton Fund-MRC Collaborative Award [200205/Z/15/Z]. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.*



This research was funded, in part, by the Wellcome Trust [214321/Z/18/Z, and 203135/Z/16/Z]. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

## **Ethics Statement**

Approval for CRyPTIC study was obtained by Taiwan Centers for Disease Control IRB No. 106209, University of KwaZulu Natal Biomedical Research Ethics Committee (UKZN BREC) (reference BE022/13) and University of Liverpool Central University Research Ethics Committees (reference 2286), Institutional Research Ethics Committee (IREC) of The Foundation for Medical Research, Mumbai (Ref nos. FMR/IEC/TB/01a/2015 and FMR/IEC/TB/01b/2015), Institutional Review Board of P.D. Hinduja Hospital and Medical Research Centre, Mumbai (Ref no. 915-15-CR [MRC]), scientific committee of the Adolfo Lutz Institute (CTC-IAL 47-J / 2017) and in the Ethics Committee (CAAE: 81452517.1.0000.0059) and Ethics Committee review by Universidad Peruana Cayetano Heredia (Lima, Peru) and LSHTM (London, UK).

## **Additional authors not named on the title page:**

### **Contributing through the CRyPTIC consortium (in alphabetical order)**

Ivan Barilar,<sup>29</sup> Simone Battaglia<sup>1</sup>, Emanuele Borroni<sup>1</sup>, Angela Pires Brandao<sup>2,3</sup>, Alice Brankin<sup>4</sup>, Andrea Maurizio Cabibbe<sup>1</sup>, Joshua Carter<sup>5</sup>, Pauline Claxton<sup>6</sup>, David A Clifton<sup>4</sup>, Ted Cohen<sup>7</sup>, Jorge Coronei<sup>8</sup>, Viola Dreyer<sup>29</sup>, Sarah G Earle<sup>4</sup>, Vincent Escuyer<sup>9</sup>, Lucilaine Ferrazoli<sup>3</sup>, George Fu Gao<sup>10</sup>, Jennifer Gardy<sup>11</sup>, Saheer Gharbia<sup>12</sup>, Kelen Teixeira Ghisi<sup>3</sup>, Arash Ghodousi<sup>1,13</sup>, Ana Luíza Gibertoni Cruz<sup>4</sup>, Clara Grazian<sup>14</sup>, Jennifer L Guthrie<sup>15,16</sup>, Wencong He<sup>10</sup>, Harald Hoffmann<sup>17,18</sup>, Sarah J Hoosdally<sup>4</sup>, Lisa Jarrett<sup>21</sup>, Lavana Joseph<sup>20</sup>, Ruwen Jou<sup>22</sup>, Priti Kamli<sup>23</sup>, Rukhsar Khot<sup>23</sup>, Anastasia Koch<sup>24</sup>, Thomas Andreas Kohl<sup>29</sup>, Donna Kohlerschmidt<sup>9</sup>, Samaneh Kouchaki<sup>4,25</sup>, Alexander S Lachapelle<sup>4</sup>, Ajit Lalvani<sup>26</sup>, Louis Grandjean,<sup>33</sup> Simon Grandjean Lapierre<sup>27</sup>, Ian F Laurenson<sup>6</sup>, Brice Letcher<sup>19</sup>, Wan-Hsuan Lin<sup>22</sup>, Chunfa Liu<sup>10</sup>, Dongxin Liu<sup>10</sup>, Kerri M Malone<sup>19</sup>, Ayan Mandal<sup>28</sup>, Daniela Matias,<sup>21</sup> Graeme Meintjes<sup>24</sup>, Flávia de Freitas Mendes<sup>3</sup>, Matthias Merker<sup>29</sup>, Marina Mihalic,<sup>18</sup> James Millard<sup>30</sup>, Nerges Mistry<sup>28</sup>, David Moore<sup>31,8</sup>, Kimberlee A Musser<sup>9</sup>, Dumisani Ngcamu<sup>20</sup>, Hoang Ngoc Nhung<sup>32</sup>, Stefan Niemann<sup>29, 48</sup>, Kayzad Soli Nilgiriwala<sup>28</sup>, Camus Nimmo<sup>30,33</sup>, Nana Okozi<sup>20</sup>, Rosangela Siqueira Oliveira<sup>3</sup>, Nicholas Paton<sup>34</sup>, Juliana Maira Watanabe Pinhata<sup>3</sup>, Sara Plesnik<sup>18</sup>, Zully M Puyen<sup>35</sup>, Marie Sylvianne Rabodoarivelo<sup>36</sup>, Niaina Rakotosamimanana<sup>36</sup>, Paola MV Rancoita<sup>13</sup>, Priti Rathod<sup>21</sup>, Esther Robinson,<sup>21</sup> Gillian Rodger<sup>4</sup>, Camilla Rodrigues<sup>23</sup>, Aysha Roohi<sup>4</sup>, David Santos-Lazaro<sup>35</sup>, Sanchi Shah<sup>28</sup>, , Grace Smith<sup>21,12</sup>, Walter Solano<sup>8</sup>, Andrea Spitaleri<sup>1,13</sup>, Philip Supply<sup>39</sup>, Utkarsha Surve<sup>23</sup>, Sabira Tahseen<sup>40</sup>, Nguyen Thuy Thuong Thuong<sup>32</sup>, Guy Thwaites<sup>32,4</sup>, Katharina Todt<sup>18</sup>, Alberto Trovato<sup>1</sup>, Christian Utpatel<sup>29</sup>, Annelies Van Rie<sup>41</sup>, Srinivasan Vijay<sup>42</sup>, Robin Warren<sup>43</sup>, Jim Werngren<sup>44</sup>, Robert J Wilkinson<sup>45,46,26</sup>, Daniel J Wilson<sup>4</sup>, Penelope Wintringer<sup>19</sup>, Yu-Xin Xiao<sup>22</sup>, Yang Yang<sup>4</sup>, Zhao Yanlin<sup>10</sup>, Shen-Yuan Yao<sup>20</sup>, Baoli Zhu<sup>47</sup>

(CRyPTIC authors named on the title page are identified there by an asterisk against their affiliation)

### **Contributing through the Seq&Treat consortium (in alphabetical order)**

Heidi Albert<sup>49</sup>, Emmanuel Andre<sup>50</sup>, Jason Andrews<sup>51</sup>, Irena Arandjelovic<sup>52</sup>, Arnold Bainomugisa<sup>53</sup>, Marie Ballif<sup>54</sup>, Anna Barbova<sup>55</sup>, Draurio Barreira Cravo Neto<sup>56</sup>, Alain Baulard<sup>39</sup>, Erik Böttger<sup>57</sup>, Maxine Caws<sup>58</sup>, Angkana Chaiprasert<sup>59</sup>, Darren Chetty<sup>30</sup>, Francesc Coll<sup>31</sup>, Victoria Cook<sup>60</sup>, Teresa Cortes<sup>31</sup>, Chris Coulter<sup>53</sup>, Helen Cox<sup>24</sup>, Alan Cristoffels<sup>61</sup>, Roland Diel<sup>62</sup>, Carla Duncan<sup>16</sup>, Sarah Dunstan<sup>63</sup>, Matthias Egger<sup>54</sup>, Kiatichai Faksri<sup>97</sup>, Margaret Fitzgibbon<sup>64</sup>, Victoria Furió<sup>65</sup>, Sebastien Gagneux<sup>66,96</sup>, Patricia Hall,<sup>95</sup> Tanya Halse<sup>9</sup>, Zahra Hasan<sup>67</sup>, Kathryn Holt<sup>68</sup>, Michael Inouye<sup>69</sup>, Frances Jamieson<sup>16</sup>, Babak Javid<sup>70</sup>, James Johnston<sup>60</sup>, Moses Joloba<sup>71</sup>, Danielle Jorgensen<sup>60</sup>, Mostofa Kamal<sup>72</sup>, George Kasule<sup>71</sup>, Peter M Keller<sup>54</sup>, Ellis Kelly<sup>73</sup>, Donna Kohlerschmidt<sup>9</sup>, Clare Kong<sup>60</sup>, Julianne Kus<sup>16</sup>, Pascal Lapierre<sup>9</sup>, Gian Maria Rossolini<sup>74</sup>, Marguerite Massinga Loembé<sup>75</sup>, Vanessa Mathys<sup>76</sup>, Fabrizio Menardo<sup>66,96</sup>, Alberto Mendoza<sup>77</sup>, John Metcalfe<sup>70</sup>, Satoshi Mitarai<sup>78</sup>, Max O'Donnell<sup>79</sup>, Rick Ong Twee Hee<sup>34</sup>, Sushil Pandey<sup>53</sup>, Amy Piatek<sup>80</sup>, Therdsak Prammananan<sup>81</sup>, Youwen Qin<sup>63,69</sup>, Leen Rigouts<sup>82</sup>, Jaime Robledo<sup>83</sup>, Mabel Rodrigues<sup>60</sup>, Tom Rogers<sup>64,84</sup>, Emma Roycroft<sup>64</sup>, Vonthanak Saphonn<sup>93</sup>, Marco Schito<sup>85</sup>, Joseph Shea<sup>9</sup>, Vitali Sintchenko<sup>14</sup>, Alena Skrahina<sup>86</sup>, Prapaporn Srilohasin<sup>59</sup>, Adrie Steyn<sup>30</sup>, Prapat Suriyaphol<sup>59</sup>, Patrick Tang<sup>87</sup>, Jill Taylor<sup>9</sup>, Dang Thi Minh Ha<sup>88</sup>, Dick van Soolingen<sup>89</sup>, Wayne van Gemert<sup>90</sup>, Phan Vuong Khac Thai<sup>88</sup>, Michael G Whitfield<sup>26,41</sup>, Mark Wilcox<sup>91</sup>, Timothy William<sup>94</sup>, Mei-Hua Wu<sup>22</sup>, Teo Yik Ying<sup>34</sup>, Danila Zimenkov<sup>92</sup>

(Seq&Treat authors named on the title page are identified there by a double asterisk against their affiliation)

1 IRCCS San Raffaele Scientific Institute, Milan, Italy

2 Oswaldo Cruz Foundation, Rio de Janeiro, Brazil

3 Instituto Adolfo Lutz, São Paulo, Brazil

4 University of Oxford, Oxford, UK

5 Stanford University School of Medicine, Stanford, USA  
6 Scottish Mycobacteria Reference Laboratory, Edinburgh, UK  
7 Yale School of Public Health, Yale, USA  
8 Universidad Peruana Cayetano Heredia, Lima, Perú  
9 Wadsworth Center, New York State Department of Health, Albany, USA  
10 Chinese Center for Disease Control and Prevention, Beijing, China  
11 Bill & Melinda Gates Foundation, Seattle, USA  
12 UK Health Security Agency, London, UK  
13 Vita-Salute San Raffaele University, Milan, Italy  
14 University of New South Wales, Sydney, Australia  
15 The University of British Columbia, Vancouver, Canada  
16 Public Health Ontario, Toronto, Canada  
17 SYNLAB Gauting, Munich, Germany  
18 Institute of Microbiology and Laboratory Medicine, IMLred, WHO-SRL Gauting, Germany  
19 EMBL-EBI, Hinxton, UK  
20 National Institute for Communicable Diseases, Johannesburg, South Africa  
21 Public Health England, Birmingham, UK  
22 Taiwan Centers for Disease Control, Taipei, Taiwan  
23 Hinduja Hospital, Mumbai, India  
24 University of Cape Town, Cape Town, South Africa  
25 University of Surrey, Guildford, UK  
26 Imperial College, London, UK  
27 Université de Montréal, Canada  
28 The Foundation for Medical Research, Mumbai, India  
29 Research Center Borstel, Borstel, Germany  
30 Africa Health Research Institute, Durban, South Africa  
31 London School of Hygiene and Tropical Medicine, London, UK  
32 Oxford University Clinical Research Unit, Ho Chi Minh City, Viet Nam  
33 University College London, London, UK  
34 National University of Singapore, Singapore  
35 Instituto Nacional de Salud, Lima, Perú  
36 Institut Pasteur de Madagascar, Antananarivo, Madagascar  
39 Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019 - UMR 9017 -  
CILL - Center for Infection and Immunity of Lille, F-59000 Lille, France  
40 National TB Reference Laboratory, National TB Control Program, Islamabad, Pakistan  
41 University of Antwerp, Antwerp, Belgium  
42 University of Edinburgh, Edinburgh, UK  
43 SAMRC Centre for Tuberculosis Research, Stellenbosch University, Cape Town, South Africa  
44 Public Health Agency of Sweden, Solna, Sweden  
45 Wellcome Centre for Infectious Diseases Research in Africa, Cape Town, South Africa  
46 Francis Crick Institute, London, UK  
47 Institute of Microbiology, Chinese Academy of Sciences, Beijing, China  
48 German Center for Infection Research (DZIF), Hamburg-Lübeck-Borstel-Riems, Germany  
49 FIND, Cape Town, South Africa  
50 University Hospitals Leuven, Belgium  
51 Stanford University, California, USA  
52 University of Belgrade, Belgrade, Serbia  
53 Queensland Department of Health, Brisbane, Australia  
54 University of Bern, Switzerland  
55 Ministry of Health of Ukraine, Kyiv, Ukraine  
56 Unitaïd, Geneva, Switzerland  
57 University of Zürich, Switzerland  
58 Liverpool School of Tropical Medicine, UK  
59 Mahidol University, Thailand  
60 British Columbia Centre for Disease Control, Canada  
61 University of the Western Cape, Bellville, South Africa  
62 University Medical Hospital Schleswig-Holstein, Germany  
63 University of Melbourne, Australia  
64 Irish Mycobacteria Reference Laboratory, Dublin, Ireland  
65 Universitat de València, Spain  
66 Swiss Tropical and Public Health Institute Basel, Switzerland  
67 The Aga Khan University, Karachi, Pakistan  
68 Monash University, Melbourne, Australia  
69 Baker Institute, Melbourne, Australia  
70 University of California, San Francisco, USA  
71 National TB and Leprosy Programme, Kampala, Uganda  
72 National Institute of Diseases of the Chest and Hospital, Dhaka, Bangladesh

73 University of Cambridge, Cambridge, UK  
 74 Careggi University Hospital, Florence, Italy  
 75 Africa CDC, Libreville, Gabon  
 76 Belgian Reference Laboratory for *M. tuberculosis*, Sciensano, Belgium  
 77 Ministry of Health, Lima, Peru  
 78 Research Institute of Tuberculosis, Tokyo, Japan  
 79 CAPRISA MRC HIV-TB Pathogenesis and Treatment Research Unit, Durban, South Africa  
 80 US Agency for International Development (USAID), Washington, USA  
 81 The National Science and Technology Development Agency, Thailand  
 82 Institute of Tropical Medicine, Antwerp, Belgium  
 83 Corporación Para Investigaciones Biológicas, Universidad Pontificia Bolivariana, Medellín, Colombia  
 84 Trinity College Dublin, Ireland  
 85 C-PATH, Tucson, Arizona, USA  
 86 Republican Scientific and Practical Centre for Pulmonology and TB, Minsk, Belarus  
 87 Sidra Medican and Research Center, Doha, Qatar  
 88 Pham Ngoc Thach Hospital, Ho Chi Minh City, Vietnam  
 89 National Institute for Public Health and the Environment, Bilthoven, The Netherlands  
 90 Stop TB Partnership, Geneva Switzerland  
 91 Leeds Teaching Hospital NHS Trust, Leeds, UK  
 92 Engelhardt Institute of Molecular Biology, Moscow, Russia  
 93 University of Health Sciences, Cambodia  
 94 Clinical Research Centre, Queen Elizabeth Hospital, Sabah, Malaysia  
 95 Centres for Disease Control and Prevention, Atlanta, USA  
 96 University of Basel, Basel, Switzerland  
 97 Khon Kaen University, Khon Kaen, Thailand

## Author contributions

All named authors were involved in conceptualization. Walker (T), Miotto, Köser, Fowler, Knaggs, Iqbal, Hunt, Malone, Wintringer, Brankin, Rodwell and Crook curated the data, and performed the formal analysis together with Chindelevitch, Farhat, Comas, Walker (S), Ismail and Rodwell. Funding for the overarching initiative was acquired by all authors affiliated to CRYPTIC, FIND, and University of California, San Diego. Funding sources for individual data contributions are listed in the funding section. All authors named on the title page contributed towards methodology, along with Albert, André, Bainomugisa, Ballif, Barilar, Neto, Baulard, Böttger, Coll, Cortes, Coulter, Cristoffels, Dawson, Egger, Furió, Gagneux, van Gemert, Hall, Hasan, Hoffman, Javid, Joloba, Kasule, Keller, Kelly, Letcher, Loembé, Mendoza, Merker, Metcalf, Mitarai, Niemann, Pandey, Piatek, Rancoita, Rigouts, Robledo, Rodrigues, Whitfield, Zhao, Zimenkov. Project administration was undertaken by Walker (T), Hoosdally, Suresh, Uplekar, Nathanson, Crook, Ismail, Rodwell, and Roohi. Resources and software were provided by the CRYPTIC consortium, European Bioinformatics Institute, FIND, WHO. Data were produced by members of the CRYPTIC consortium and the Seq&Treat consortium. Iqbal, Cirillo, Zignol, Crook, Ismail and Rodwell supervised the project, or components of the project. Walker (T) wrote the original draft. All authors contributed to the review and editing of the manuscript.

## Acknowledgements

We would like to thank Solange Alves Vinhas, Marcel Behr, Joconiah Chirenda, Tom Connor, Marcelo Cordeiro dos Santos, Lauren Cowan, Julio Croda, Chhavarath Dary, James Dawson, Bouke de Jong, Han de Neeling, Keertan Dheda, Tatiana Dutra, Konstantin Edokimov, David Englethaler, Neel Gandhi, Amy Gaskin, Solomon Ghebremichael, Li-Yang Hsu, Agathe Jouet, Faisal Masood Khanzada, HeeJin Kim, Afrânio Kritski, Connie Lam, Maria Lucia Rossetti, Maria Luiza Bazzo, Flora Martinez, Elena Martinez, Neil Martinson, Barun Mathema, Vidya Mave, Conor Meehan, Isabela Neves de Almeida, Anna P. Ralph, Alamdar Hussain Rizvi, Moises Palaci, Eddy Qing-Hao Miow, Elis Regina Dalla Costa, Alex Rosenthal, Richard Salvato, Giri Shan Rajahram, Soyoun Shin, Elisangela Silva, Mel Spigelman, Silvana Spindola

de Miranda, Angela Starks, Grant Theron, Juan Carlos Toro, Oren Tzfadia, Mirela Verza, and Yoshiyuki Yamada for efforts to contribute data and other support.

## References:

1. World Health Organization. (2021). Impact of the COVID-19 pandemic on TB detection and mortality in 2020. World Health Organization. [https://cdn.who.int/media/docs/default-source/hq-tuberculosis/impact-of-the-covid-19-pandemic-on-tb-detection-and-mortality-in-2020.pdf?sfvrsn=3fdd251c\\_16&download=true](https://cdn.who.int/media/docs/default-source/hq-tuberculosis/impact-of-the-covid-19-pandemic-on-tb-detection-and-mortality-in-2020.pdf?sfvrsn=3fdd251c_16&download=true) [accessed 12/09/2021]
2. World Health Organization. (2020). Global Tuberculosis Report, 2020. World Health Organization. <https://www.who.int/publications/i/item/9789240013131> [accessed 12/09/2021]
3. Kadura S, King N, Nakhoul M, Zhu H, Theron G, Köser CU, et al. Systematic review of mutations associated with resistance to the new and repurposed Mycobacterium tuberculosis drugs bedaquiline, clofazimine, linezolid, delamanid and pretomanid. *J Antimicrob Chemother.* 2020;**75**(8):2031–43.
4. Mohamed S, Köser CU, Salfinger M, Sougakoff W, Heysell SK. Targeted next-generation sequencing: a Swiss army knife for mycobacterial diagnostics? *Eur Respir J.* 2021 Mar;**57**(3).
5. World Health Organization. (2020). Meeting report of the WHO expert consultation on the definition of extensively drug-resistant tuberculosis. World Health Organization. <https://www.who.int/publications/i/item/meeting-report-of-the-who-expert-consultation-on-the-definition-of-extensively-drug-resistant-tuberculosis> [accessed 12/09/2021]
6. Mugwagwa T, Abubakar I, White PJ. Using molecular testing and whole-genome sequencing for tuberculosis diagnosis in a low-burden setting: a cost-effectiveness analysis using transmission-dynamic modelling. *Thorax.* 2021 Mar;**76**(3):281–91.
7. World Health Organization. (2018). The use of next-generation sequencing technologies for the detection of mutations associated with drug resistance in Mycobacterium tuberculosis complex: technical guide. World Health Organization. <https://apps.who.int/iris/handle/10665/274443> [accessed 12/09/2021]
8. Miotto P, Tagliani E, Chindelevitch L, Starks AM, Emerson C, Hanna D, et al. A standardised method for interpreting the association between mutations and phenotypic drug resistance in Mycobacterium tuberculosis. *Eur Respir J.* 2017 Dec **18**;1–13.
9. World Health Organization. (2021). Catalogue of mutations in Mycobacterium tuberculosis complex and their association with drug resistance. World Health Organization. <https://www.who.int/publications/i/item/9789240028173> [accessed 12/09/2021]
10. World Health Organization. (2018). Technical manual for drug susceptibility testing of medicines used in the treatment of tuberculosis. World Health Organization. <http://apps.who.int/iris/bitstream/handle/10665/275469/9789241514842-eng.pdf> [accessed 12/09/2021]

11. World Health Organization. (2021). Technical Report on critical concentrations for drug susceptibility testing of isoniazid and the rifamycins (rifampicin, rifabutin and rifapentine). World Health Organization. [https://www.who.int/publications/i/item/technical-report-on-critical-concentrations-for-drug-susceptibility-testing-of-isoniazid-and-the-rifamycins-\(rifampicin-rifabutin-and-rifapentine\)](https://www.who.int/publications/i/item/technical-report-on-critical-concentrations-for-drug-susceptibility-testing-of-isoniazid-and-the-rifamycins-(rifampicin-rifabutin-and-rifapentine)) [accessed 12/09/2021]
12. World Health Organization. (2018). Technical Report on critical concentrations for drug susceptibility testing of medicines used in the treatment of drug-resistant tuberculosis. World Health Organization. <https://apps.who.int/iris/handle/10665/260470> [accessed 12/09/2021]
13. World Health Organization. (2011). Noncommercial culture and drug-susceptibility testing methods for screening patients at risk for multidrug-resistant tuberculosis: policy statement. World Health Organization. <https://apps.who.int/iris/handle/10665/44601> [accessed 12/09/2021]
14. Martin L, Coronel J, Faulx D, Valdez M, Metzler M, Crudder C, et al. A Field Evaluation of the Hardy TB MODS Kit™ for the Rapid Phenotypic Diagnosis of Tuberculosis and Multi-Drug Resistant Tuberculosis. Delogu G, editor. *PLoS ONE*. 2014;**9**(9):e107258.
15. World Health Organization. Communicable Diseases Cluster. (2001). Guidelines for drug susceptibility testing for second-line anti-tuberculosis drugs for dots-plus. World Health Organization. <https://apps.who.int/iris/handle/10665/83807> [accessed 12/09/2021]
16. World Health Organization & World Health Organization. (2008). Policy guidance on drug-susceptibility testing (DST) of second-line antituberculosis drugs. World Health Organization. <https://apps.who.int/iris/handle/10665/70500> [accessed 12/09/2021]
17. World Health Organization. (2014). Companion handbook to the WHO guidelines for the programmatic management of drug-resistant tuberculosis. World Health Organization. [https://apps.who.int/iris/bitstream/handle/10665/130918/9789241548809\\_eng.pdf](https://apps.who.int/iris/bitstream/handle/10665/130918/9789241548809_eng.pdf) [accessed 12/09/2021]
18. The CRyPTIC Consortium, Fowler PW. Epidemiological cutoff values for a 96-well broth microdilution plate for high-throughput research antibiotic susceptibility testing of *M. tuberculosis*. *medRxiv* 2021; <http://medrxiv.org/lookup/doi/10.1101/2021.02.24.21252386> [accessed 12/09/2021]
19. Clockwork. Available from: <https://github.com/iqbal-lab-org/clockwork>
20. The CRyPTIC consortium. A data compendium of *M. tuberculosis* antibiotic resistance. [in submission to *BioRxiv*, 2021 – doi pending]
21. Hunt, M.H., Letcher, B., Malone, KM, Nguyen, G., Hall, M.B., Colquhoun, R.M., Lima, L., Schatz, M., Ramakrishnan, S., CRyPTIC consortium, Iqbal, Z. Minos: graph adjudication and joint genotyping of cohorts of bacterial genomes. [in submission to *BioRxiv*, 2021 – doi pending]

22. Letcher B, Hunt M, Iqbal Z. Gramtools enables multiscale variation analysis with genome graphs. *Genome Biology*. 2021;**22**:259
23. The CRyPTIC Consortium and the 100,000 Genomes Project. Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. *N Engl J Med* 2018; **379**:1403-1415.
24. Walker TM, Kohl TA, Omar SV, Hedge J, Del Ojo Elias C, Bradley P, et al. Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study. *Lancet Infect Dis*. 2015;**15**(10):1193–1202.
25. Dorfman R. The Detection of Defective Members of Large Populations. *Ann Math Stat*. 1943;**14**(4):436–40.
26. Zabeti H, Dexter N, Safari AH, Sedaghat N, Libbrecht M, Chindelevitch L. INGOT-DR: an interpretable classifier for predicting drug resistance in M. tuberculosis. *Algorithms Mol Biol*. 2021;**16**(1):17.
27. Merker M, Kohl TA, Barilar I, Andres S, Fowler PW, Chryssanthou E, et al. Phylogenetically informative mutations in genes implicated in antibiotic resistance in Mycobacterium tuberculosis complex. *Genome Med*. 2020;**12**(1):27.
28. Köser CU, Cirillo DM, Miotto P. How To Optimally Combine Genotypic and Phenotypic Drug Susceptibility Testing Methods for Pyrazinamide. *Antimicrob Agents Chemother*. 2020 ;**64**(9):e01003-20.
29. Somoskovi A, Bruderer V, Hömke R, Bloemberg GV, Böttger EC. A mutation associated with clofazimine and bedaquiline cross-resistance in MDR-TB following bedaquiline treatment. *Eur Respir J*. 2015;**45**(2):554–7.
30. Bloemberg GV, Keller PM, Stucki D, Trauner A, Borrell S, Latshang T, et al. Acquired Resistance to Bedaquiline and Delamanid in Therapy for Tuberculosis. *N Engl J Med*. 2015;**373**(20):1986–8.
31. Coll F, Phelan J, Hill-Cawthorne GA, Nair MB, Mallard K, Ali S, et al. Genome-wide analysis of multi- and extensively drug-resistant Mycobacterium tuberculosis. *Nat Genet*. 2018;**50**(2):307–16.
32. Farhat MR, Freschi L, Calderon R, Ioerger T, Snyder M, Meehan CJ, et al. GWAS for quantitative resistance phenotypes in Mycobacterium tuberculosis reveals resistance genes and regulatory regions. *Nat Commun*. 2019;**10**(1):2128.
33. World Health Organization. (2021). Target product profile for next-generation drug-susceptibility testing at peripheral centres. World Health Organization. <https://www.who.int/publications/i/item/9789240032361> [accessed 12/09/2021]
34. Vargas R, Freschi L, Spitaleri A, Tahseen S, Barilar I, Niemann S, et al. The role of epistasis in amikacin, kanamycin, bedaquiline, and clofazimine resistance in Mycobacterium tuberculosis complex. *Antimicrob Agents Chemother*. 2021;AAC0116421.



35. Gygli SM, Keller PM, Ballif M, Blöchliger N, Hömke R, Reinhard M, et al. Whole-Genome Sequencing for Drug Resistance Profile Prediction in Mycobacterium tuberculosis. *Antimicrob Agents Chemother*. 2019;**63**(4).

Figure1

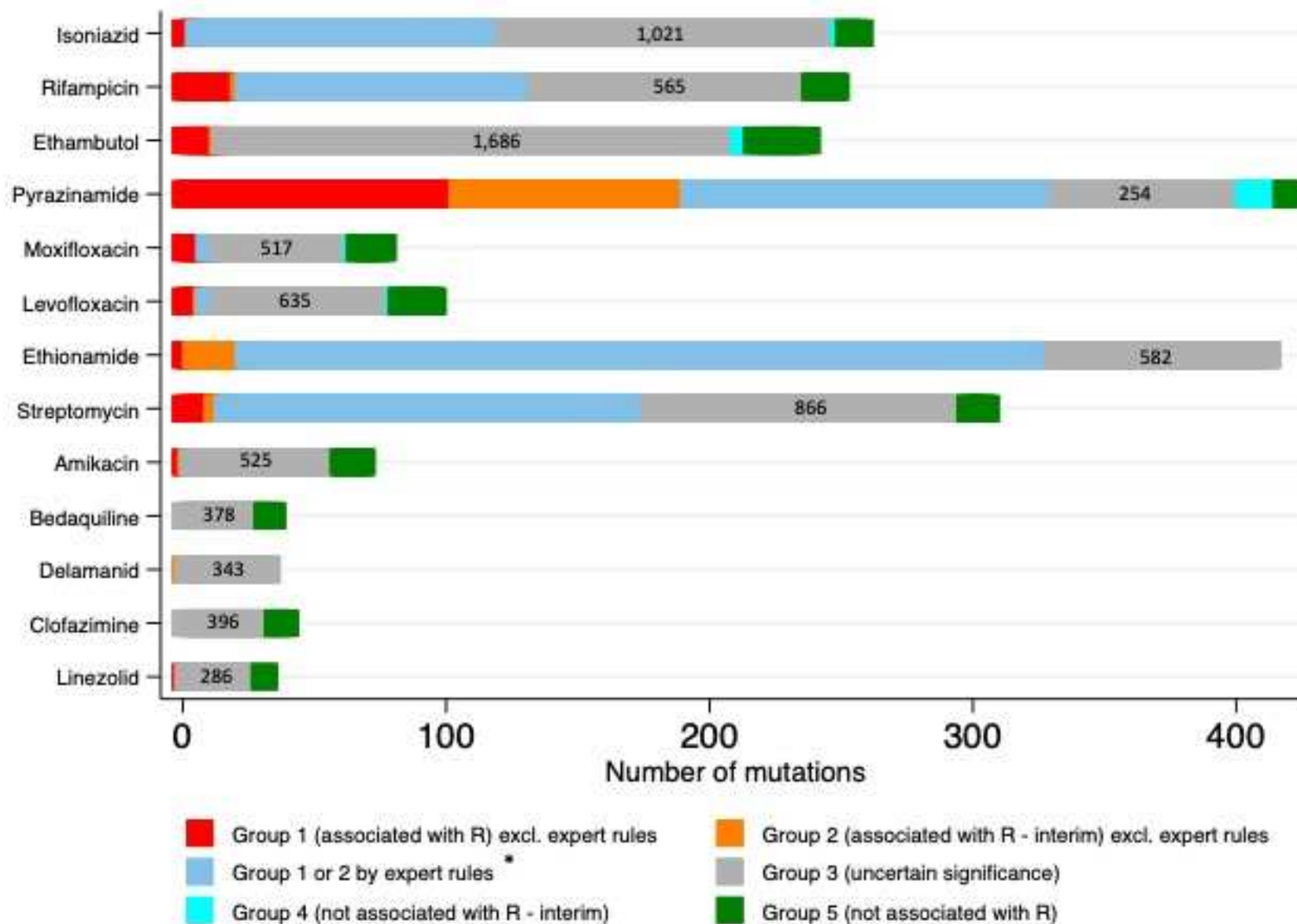
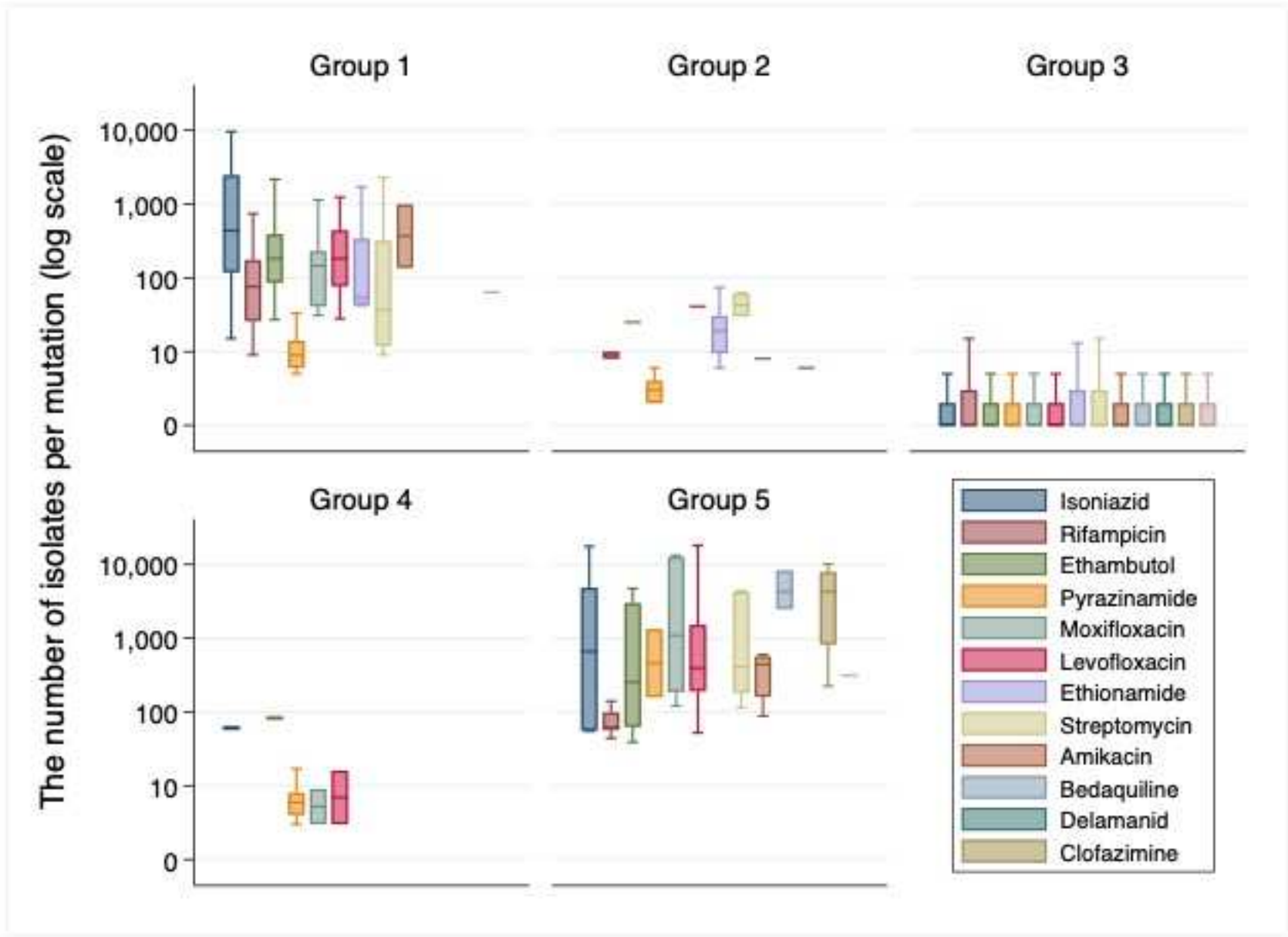
[Click here to access/download;Figure;Figure1.jpg](#)

Figure2



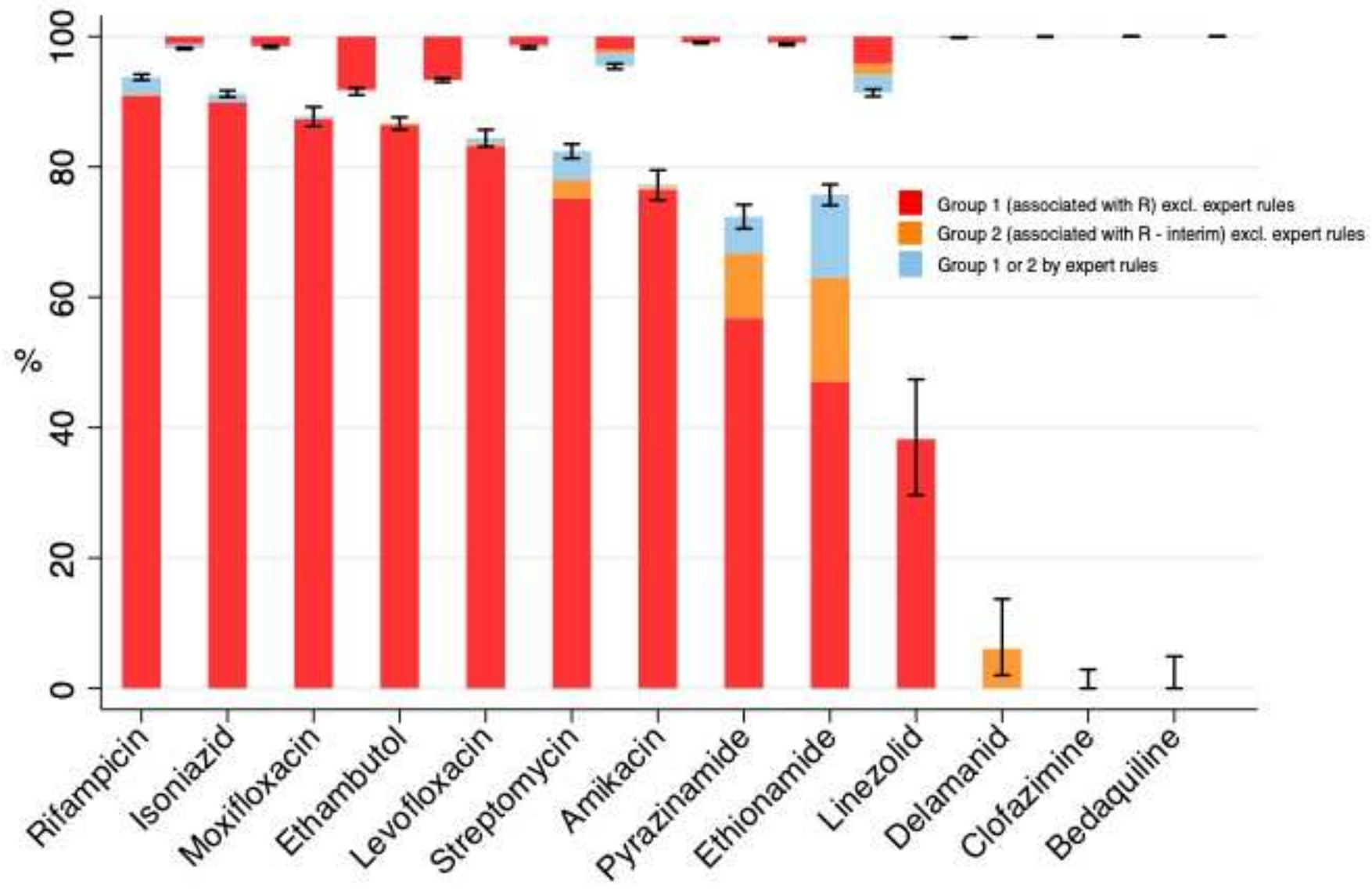


Figure4

[Click here to access/download;Figure;Figure4.jpg](#)

