AMERICAN SOCIETY FOR MICROBIOLOGY | Antimicrobial Agents and Chemotherapy®

MECHANISMS OF RESISTANCE

Distribution of Common and Rare Genetic Markers of Second-Line-Injectable-Drug Resistance in *Mycobacterium tuberculosis* Revealed by a Genome-Wide Association Study

Derek Conkle-Gutierrez,[a] Calvin Kim,[a] Sarah M. Ramirez-Busby,[a] Samuel J. Modlin,[a] Mikael Mansjö,[b] Jim Werngren,[b] Leen Rigouts,[c,d] Sven E. Hoffner,[a,e] Faramarz Valafar[a]

[a]Laboratory for Pathogenesis of Clinical Drug Resistance and Persistence (LPCDRP), Biomedical Informatics Research Center, Division of Epidemiology, School of Public Health, San Diego State University, San Diego, California, USA
[b]Department of Microbiology, Public Health Agency of Sweden, Solna, Sweden
[c]Department of Biomedical Sciences, Antwerp University (UA), Antwerp, Belgium
[d]Mycobacteriology Unit, Department of Biomedical Sciences, Institute of Tropical Medicine (ITM), Antwerp, Belgium
[e]Department of Global Public Health, Karolinska Institute, Stockholm, Sweden

ABSTRACT Point mutations in the *rrs* gene and the *eis* promoter are known to confer resistance to the second-line injectable drugs (SLIDs) amikacin (AMK), capreomycin (CAP), and kanamycin (KAN). While mutations in these canonical genes confer the majority of SLID resistance, alternative mechanisms of resistance are not uncommon and threaten effective treatment decisions when using conventional molecular diagnostics. In total, 1,184 clinical *Mycobacterium tuberculosis* isolates from 7 countries were studied for genomic markers associated with phenotypic resistance. The markers *rrs*:A1401G and *rrs*:G1484T were associated with resistance to all three SLIDs, and three known markers in the *eis* promoter (*eis*:G-10A, *eis*:C-12T, and *eis*:C-14T) were similarly associated with kanamycin resistance (KAN-R). Among 325, 324, and 270 AMK-R, CAP-R, and KAN-R isolates, 274 (84.3%), 250 (77.2%), and 249 (92.3%) harbored canonical mutations, respectively. Thirteen isolates harbored more than one canonical mutation. Canonical mutations did not account for 103 of the phenotypically resistant isolates. A genome-wide association study identified three genes and promoters with mutations that, on aggregate, were associated with unexplained resistance to at least one SLID. Our analysis associated *whiB7* 5′-untranslated-region mutations with KAN resistance, supporting clinical relevance for this previously demonstrated mechanism of KAN resistance. We also provide evidence for the novel association of CAP resistance with the promoter of the *Rv2680-Rv2681* operon, which encodes an exoribonuclease that may influence the binding of CAP to the ribosome. Aggregating mutations by gene can provide additional insight and therefore is recommended for identifying rare mechanisms of resistance when individual mutations carry insufficient statistical power.

KEYWORDS drug resistance, *Mycobacterium tuberculosis*, rare mechanisms, amikacin, capreomycin, injectables, kanamycin, second-line antibiotics

Tuberculosis (TB) remains a constant global public health threat due to rising cases of drug resistance among various strains of *Mycobacterium tuberculosis*. Half a million estimated TB cases were rifampicin resistant in 2020, including 3 to 4% of new TB cases and 18 to 21% of previously treated cases (1). This trend is exacerbated in countries of the former Soviet Union, where over half of the isolates from previously treated TB patients were rifampicin resistant (1). In 2018, 78% of rifampicin-resistant cases were also resistant to isoniazid, making them multidrug-resistant tuberculosis (MDR-

Address correspondence to Faramarz Valafar, faramarz@sdsu.edu.

The authors declare no conflict of interest.

Received 23 November 2021
Returned for modification 20 December 2021
Accepted 24 March 2022
Published 9 May 2022

TABLE 1 Known SLID-R markers in *M. tuberculosis*[a]

Antibiotic	Known markers
AMK	*rrs*:A1401G, *rrs*:G1484T, *eis*:C-14T
CAP	*rrs*:A1401G, *rrs*:G1484T
KAN	*rrs*:A1401G, *rrs*:G1484T, *eis*:G-10A, *eis*:C-12T, *eis*:C-14T

[a]This set of mutations was derived from the WHO-endorsed catalogue of resistance-associated mutations (18) and used to determine the expected resistance of clinical isolates to AMK, CAP, and KAN.

TB) cases (2). In 2018, an estimated 6.2% of MDR-TB cases were extensively drug resistant (XDR), then defined as MDR-TB strains that were additionally resistant to at least a fluoroquinolone and a second-line injectable drug (SLID) (2).

Successful tuberculosis treatment relies on early identification and effective regimens, which can be ensured by rapid and accurate drug susceptibility testing (DST) to identify potential MDR-TB cases. MDR-TB should be treated with combinations of drugs shown *in vitro* to be effective (3). Phenotypic DST takes weeks, during which time patients may face ineffective treatment regimens with often debilitating side effects. Molecular diagnostics, in contrast, can be performed rapidly. As these rely on genetic markers of resistance, to improve their accuracy, we must understand the mechanisms behind resistance and comprehensively identify all their markers.

The SLIDs amikacin (AMK) and kanamycin (KAN) kill *M. tuberculosis* cells by binding to the 16S ribosomal subunit and disrupting translation (4). The bactericidal mechanism of the SLID capreomycin (CAP) is less understood. It is hypothesized to bind to the 70S ribosomal subunit and limit mRNA-tRNA translocation (5). In 2018, the WHO recommended against the use of CAP and KAN due to their side effects such as ototoxicity and nephrotoxicity (6) and their significant association with treatment failure (7). Even though AMK is now being phased out in favor of bedaquiline (8), it is still widely used. Although CAP and KAN are currently not recommended in treatment regimens, understanding the mechanisms of resistance to these drugs can inform and corroborate our understanding of AMK resistance (AMK-R), as the three drugs have similar mechanisms of action.

Genetic markers can rapidly identify SLID resistance (SLID-R) in clinical isolates (9). Across multiple studies, the most frequently observed SLID-R marker within the *M. tuberculosis* complex is *rrs*:A1401G (10–13). This single nucleotide polymorphism (SNP) in the 16S rRNA gene *rrs* typically causes cross-resistance to all three SLIDs (11, 13). Other mutations in *rrs* are also associated with SLID-R but are far rarer (14).

While *rrs*:A1401G typically causes resistance to all three SLIDs, not all isolates resistant to one or more SLIDs have this marker (11). Many KAN-resistant (KAN-R) isolates, for example, harbor variants in the promoter of *eis* (11, 15, 16). The *eis* gene encodes an *N*-acetyltransferase, which can inactivate KAN when overexpressed (15). Spectrophotometric assays have further shown that *eis* acetylates KAN (17). The *eis* promoter mutation *eis*:C-14T, a C to T transition located 14 nucleotides upstream of *eis*, has also been shown to cause resistance to amikacin (18). The amikacin MIC, when increased by *eis*:C-14T, is just below the critical concentration, and resistance is thus inconsistently detected by culture-based DST (18). These *eis* promoter markers are common in *M. tuberculosis* strains from the former Soviet Union, where the use of KAN has been high (16). Together with *rrs*:A1401G and *rrs*:G1484T, these known markers (Table 1) explain most SLID-R *M. tuberculosis* isolates (18).

Previous studies have found SLID-R isolates with no known markers (11, 13). Mutations in several genes have been suggested to cause resistance in these isolates, including *tlyA* (19), *whiB7* (20), *vapC21* (21), and *bfrB*, also known as ferritin (22, 23). Knockout variants in *tlyA* induce CAP resistance (CAP-R) *in vitro* (5). Loss-of-function mutations in *tlyA* (5, 19, 24) are proposed to prevent CAP from binding to the 16S subunit (19) by preventing the methylation of 16S (25).

To find alternative mechanisms of SLID-R, we performed a genome-wide

TABLE 2 Phenotypic drug susceptibility testing results[a]

Drug	No. of resistant isolates/no. of isolates tested	% resistant isolates
AMK	325/1,163	27.9
CAP	324/1,159	27.9
KAN	270/496	54.4

[a]The number of total resistant *M. tuberculosis* isolates out of all isolates with available phenotypic drug susceptibility testing (DST) results for each drug and the percentage of isolates with DST data for a given drug that show resistance are shown.

association study (GWAS) on 1,184 clinical *M. tuberculosis* isolates, including 111 SLID-R isolates with no known SLID-R markers. Our methods corroborated the association of *whiB7* with kanamycin resistance and identified several putative amikacin resistance markers in *ppe51*, a transport mediator previously implicated in resistance to the drug candidate 3,3-bis-di(methylsulfonyl)propionamide (26).

RESULTS

This study surveyed 1,184 clinical isolates for injectable-drug resistance markers. AMK and CAP phenotypic DST data were available for 1,163 and 1,159 isolates, respectively (Table 2). Only 496 isolates had phenotypic DST data available for KAN (Table 2). The isolates were categorized based on phenotypic DST and the presence of known resistance markers. No isolate tested for all three SLIDs was monoresistant to AMK (see Fig. S1 at https://doi.org/10.5281/zenodo.6149149).

Mutations in *rrs* and the *eis* promoter are associated with SLID resistance and specific lineages. Known SLID-R markers in the *rrs* gene and the *eis* promoter were associated strongly with resistance (Table 3). They predicted SLID-R with a sensitivity similar to that in a 2018 GWAS (10), although the sensitivity was lower than those in previous studies (5, 24). The lower sensitivity is likely due to the deliberate selection of clinical isolates with discordant genotypes and phenotypes. Variant *rrs*:A1401G was the most frequent marker (n = 260) (Table 4), while *rrs*:G1484T was the least frequent (n = 4) (Table 4). However, no isolate with *rrs*:G1484T was susceptible to any SLID that it was tested for (Table 4). The known marker *rrs*:A1401G was more frequent (181/700) within lineage 2 (East Asian) than within any other lineage (Fig. 1 and Table 5). Despite *rrs*:A1401G typically conferring full cross-resistance, *rrs*:A1401G was carried by 22 CAP-susceptible (CAP-S) isolates (Table 4), all of which were AMK-R.

The *eis* promoter mutation C-14T was carried by 11 AMK-R and 18 AMK-S isolates (Table 4). Ten of the AMK-R isolates with *eis*:C-14T carried no other SLID-R marker in *rrs* (Table 4). Three unexplained AMK-R isolates had evidence of heteroresistance, with *rrs*:A1401G being supported by 10 to 20% of the mapped reads (see Table S1 at https://doi.org/10.5281/zenodo.6149149). No unexplained AMK-R isolates had reads supporting *rrs*:G1484T.

Mutations in the *eis* promoter were especially common in isolates from Moldova (Table 6). Among the 270 KAN-R isolates, isolates with known KAN-R markers in the *eis* promoter were 27.8 times more likely to be from Moldova than isolates without known markers in the *eis* promoter (95% confidence interval [CI], 12.8 to 64.8; $P < 2.2e{-}16$ by Fisher's exact test)

TABLE 3 Sensitivity and specificity of SLID-R prediction with known markers[a]

Drug	No. of isolates				Sensitivity (%) (95% CI)	Specificity (%) (95% CI)
	Explained resistant	Explained susceptible	Unexplained resistant	Unexplained susceptible		
AMK	276	829	51	27	84.3 (82.1–86.3)	96.8 (95.6–97.7)
CAP	250	813	74	22	77.2 (74.6–79.5)	97.4 (96.2–98.2)
KAN	249	223	21	3	92.3 (89.6–94.4)	98.7 (97.1–99.4)

[a]A 95% confidence interval for each estimate was calculated using the score method with continuity correction (64). Isolates with genotypic-phenotypic concordance were classified as "explained." Otherwise, they are "unexplained."

TABLE 4 Number of *M. tuberculosis* isolates with each known marker[a]

Variant	No. of isolates					
	AMK-R	AMK-S	CAP-R	CAP-S	KAN-R	KAN-S
rrs:A1401G	260	9	246	22	188	1
rrs:G1484T	4	0	4	0	3	0
eis:G-10A	5	49	3	51	10	0
eis:C-12T	10	95	19	84	42	2
eis:C-14T	11	18	3	26	10	0

[a]Isolates are classified based on their resistance or susceptibility to each SLID. R, resistant; S, susceptible. Note that some isolates carried multiple markers.

(Table 6). The 89 isolates from Moldova belonged exclusively to lineage 4 (Euro-American, $n = 51$; and East Asian, $n = 38$). Similarly, among all 1,184 isolates, the only isolates with variant *eis*:C-12T were East Asian (54/700) or Euro-American (53/350) (Fig. 1). Ten isolates with known markers in the *eis* promoter also carried the marker *rrs*:A1401G (see Table S3 at the URL mentioned above). Of the three KAN-S isolates with known KAN-S markers, one had both *eis*:C-12T and a nonsynonymous *eis* mutation, *eis*:V163I. Promoter mutations in *eis* do not confer resistance if *eis* has a loss-of-function mutation (27); however, *eis*:V163I is not a frameshift or early stop codon and is thus less likely to be a loss-of-function mutation.

High cooccurrence of *rrs*:A1401G with the streptomycin resistance marker *rrs*:A514C. Variants *rrs*:C492T, *rrs*:C517T, and *rrs*:A514C were the most frequent noncanonical *rrs* mutations among the clinical isolates (Fig. 2). The three mutations were not associated with resistance to any SLID. Both *rrs*:A514C and *rrs*:C517T are known streptomycin resistance markers (28). Variant *rrs*:A514C was frequently carried by isolates that also carried the SLID-R marker *rrs*:A1401G. Among all clinical isolates, *rrs*:A514C was 6.0 times more likely to be carried in isolates with *rrs*:A1401G than in isolates without *rrs*:A1401G ($P = 2.424e{-}11$ by Fisher's exact test).

Insufficient statistical power to identify individual alternative resistance markers. As there were only 21 unexplained KAN-R isolates in this study (Table 3), there was

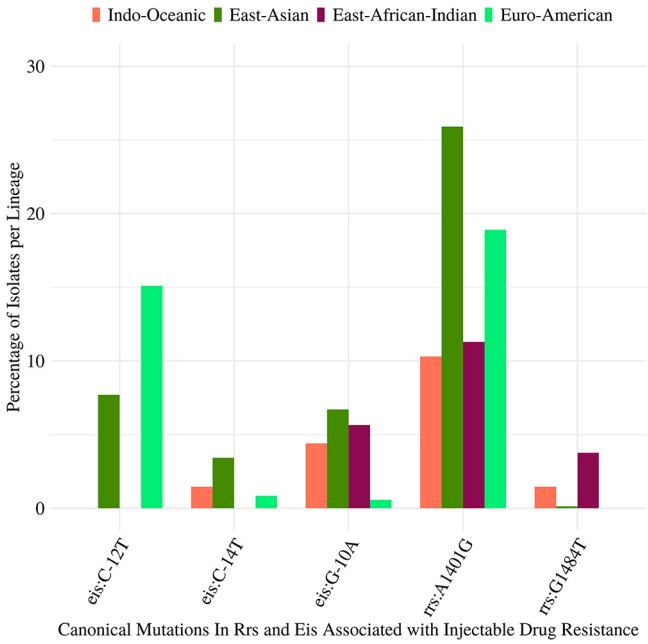

FIG 1 Frequency of known resistance markers within each lineage among 1,184 clinical *M. tuberculosis* isolates. The *y* axis represents the percentage of clinical *M. tuberculosis* isolates from each lineage that possess the mutations represented on the *x* axis. The lineages of 12 isolates could not be determined, including 1 isolate with *eis*:C-14T and 1 isolate with *rrs*:A1401G. One other isolate belonged to the *M. africanum* West African 1 lineage.

TABLE 5 Frequency of known resistance markers within each lineage[a]

Variant	Indo-Oceanic	East Asian	East African-Indian	Euro-American
rrs:A1401G	7	181	6	66
rrs:G1484T	1	1	2	0
eis:C-12T	0	54	0	53
eis:C-14T	0	25	0	3
eis:G-10A	3	47	3	2
Total	68	700	53	350

[a]The number of clinical isolates in each lineage carrying each known SLID-R marker, regardless of phenotypic DST, is shown. The lineages of 12 isolates could not be determined, including 1 isolate with *eis*:C-14T and 1 isolate with *rrs*:A1401G. One other isolate belonged to the *Mycobacterium africanum* West African 1 lineage.

insufficient statistical power to identify potential rare KAN-R mechanisms. In this isolate set, aside from known markers, no significant association was found between KAN-R and any other mutations in the *eis* promoter. While previous studies found the mutation *eis*:G-37T in isolates with at least low-level resistance to KAN (15, 16), only one of our isolates carried it (Fig. 2). This isolate was phenotypically KAN-S, although it is possible that the MIC would reveal low-level resistance to KAN. Similarly, the mutation *rrs*:C1402T was carried by only three isolates, including 1 AMK-R, 1 AMK-S, 1 AMK-untested, 2 CAP-R, and 1 CAP-S isolates (Fig. 2).

Mutations in the RNA methylase gene *tlyA* confer CAP-R *in vitro* (5) and have previously been observed in clinical isolates (29). However, in this isolate set, no significant association was found between CAP-R and any mutation in *tlyA*. Only six nonsynonymous *tlyA* mutations were carried by any CAP-R isolates that did not also carry the known marker *rrs*:A1401G (see Table S4 at https://doi.org/10.5281/zenodo.6149149). None of these six mutations were carried by more than two CAP-R isolates (see Table S4 at the URL mentioned above).

Excluding isolates with known SLID-R markers, no individual mutation in the genome predicted resistance to any SLID with >21.8% sensitivity in this isolate set. No mutation with at least a 50% positive predictive value had more than 5.03% sensitivity, and no mutation exclusive to resistant isolates had more than 1.34% sensitivity. No AMK-R-exclusive mutation was carried by more than two unexplained AMK-R isolates. Similarly, no CAP-R-exclusive mutation was carried by more than two unexplained CAP-R isolates, and no KAN-R-exclusive mutation was carried by more than four unexplained KAN-R isolates. Due to their rarity, there was insufficient statistical power to determine whether any of these mutations were alternative resistance markers.

Our set of clinical isolates included 333 isolates sequenced on single-molecule real-time (SMRT) sequencers, which can resolve many known blind spots in the *M. tuberculosis* genome, such as in the proline-glutamic acid (PE) and proline-proline-glutamic acid (PPE)

TABLE 6 Geographic specificity of KAN resistance markers in the *eis* promoter[a]

	No. of isolates		
Specificity	Moldovan	Not Moldovan	Total
Kanamycin resistant			
Carries known marker in the *eis* promoter	48 (2)	13 (1)	61 (3)
No known marker in the *eis* promoter	24 (21)	185 (170)	209 (191)
Total	72	198	270
Kanamycin susceptible			
Carries known marker in the *eis* promoter	2	0	2
No known marker in the *eis* promoter	13	211	224
Total	15	211	226

[a]This contingency table reports the association between the possession of a known KAN-R marker in the *eis* promoter and collection from Moldova, stratified by KAN DST. Two Moldovan isolates had low sequencing coverage in the *eis* promoter and were thus excluded from this contingency table and all KAN analyses. The numbers of isolates with either of the SLID-R markers *rrs*:A1401G and *rrs*:G1484T are included in parentheses.

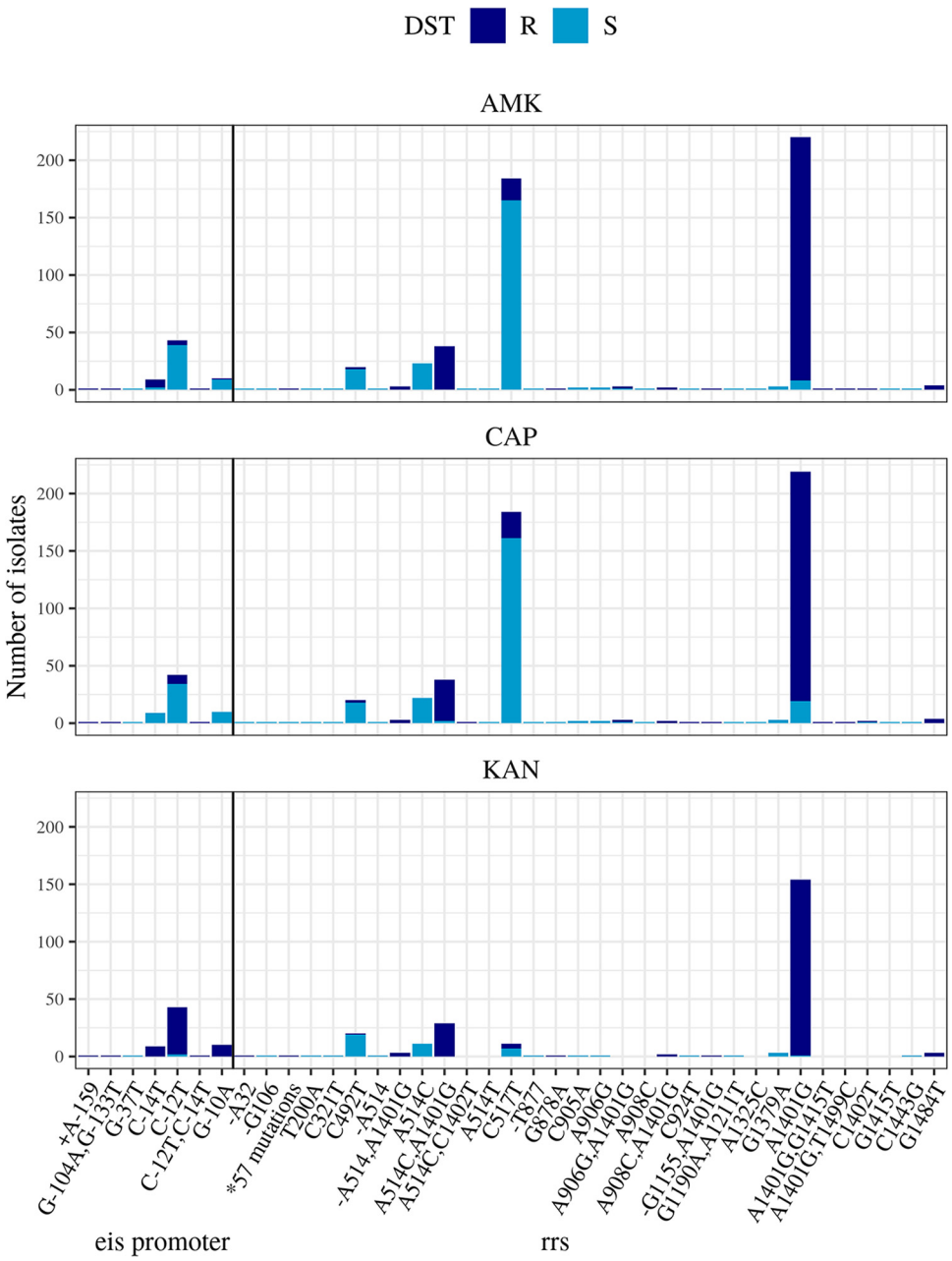

FIG 2 Mutations in the *eis* promoter and *rrs*. Shown are numbers of clinical *M. tuberculosis* isolates with mutations in the *rrs* gene or the *eis* promoter, with resistance or susceptibility to AMK, CAP, and KAN. Each column reports the numbers of resistant (R) and susceptible (S) isolates carrying each mutation. For each drug, only isolates with phenotypic DST results for that drug were counted. A vertical line separates the mutations found in *rrs* from those in the *eis* promoter. Note that column "*57 mutations" represents a set of 57 *rrs* variants called in a single isolate, which were combined for brevity.

gene families (30). However, no mutations in PE or PPE genes were associated with resistance, even for a subset of only the SMRT-sequenced isolates.

Several genes have previously been suggested to affect SLID-R. A support vector machine approach identified variants in three genes as potential determinants of an XDR phenotype: *vapC21*, *Rv3471c*, and *Rv3848* (21). Comparative proteomics suggested that 12 genes may be involved in resistance to AMK or KAN: *atpA*, *tig*, *lpdC*, *tuf*, *moxR1*, *Rv2005c*, *35kd_ag*, *prcA*, *Rv0148*, *bfrA*, *bfrB*, and *hspX* (22). The efflux pump *Rv1258c* (31), the transcriptional regulator *whiB7* (32), and the virulence gene *whiB6* (33) have also

TABLE 7 Top genes from the genome-wide association study, aggregated by gene[a]

Gene	SLID	No. of unexplained resistant isolates with resistance-exclusive mutations	Total no. of unexplained resistant isolates with mutations in the gene	Proportion	No. of unique resistance-exclusive mutations in unexplained resistant isolates	No. of isolates with resistance-exclusive mutations carried by unexplained resistant isolates
plcB	AMK	3	5	0.60	5	3
Rv0584	AMK	3	5	0.60	3	7
recB	AMK	3	6	0.50	3	3
smc	AMK	3	7	0.43	3	3
Rv2680_prom	CAP	7	7	1	15	8
cysG_prom	CAP	5	7	0.71	26	17
thrB_prom	CAP	5	8	0.62	30	29
plcB	CAP	4	7	0.57	10	4
cut1	CAP	4	7	0.57	4	12
Rv1726	CAP	4	7	0.57	5	4
whiB7_prom	KAN	9	9	1	9	16
cyp142	KAN	5	5	1	5	6
accE5	KAN	5	5	1	13	21
prfB	KAN	5	5	1	4	5
folC	KAN	5	5	1	4	37

[a]The number of unexplained resistant isolates with resistance-exclusive mutations for each SLID and each gene is the number of isolates that have unexplained resistance to that SLID and carry a mutation in that gene that is not carried by any isolate susceptible to that SLID. The total number of unexplained resistant isolates with mutations in the gene for each gene and each SLID is the number of isolates that have unexplained resistance to that SLID and carry any mutation in that gene. Genes below the mean for this count were removed (5 for AMK, 7 for CAP, and 5 for KAN). "Proportion" is the first count divided by the second. Genes associated with first-line-drug resistance were excluded.

been associated with SLID resistance. However, no individual variant within these 18 genes was present in more than one AMK-R or CAP-R isolate after removing variants also present in susceptible isolates. Variants present in one or more susceptible isolates also were not significantly associated with resistance in this isolate set.

***whiB7* variants in aggregate are associated with kanamycin resistance.** Different variants in the same gene can cause the same change in phenotype (34). If multiple rare variants in the same gene cause resistance, they may be missed by a genome-wide association study on individual variants. To account for this, we measured the association between SLID-R and the variants in each *M. tuberculosis* gene in aggregate (Table 7).

For KAN, the *whiB7* 5′ untranslated region (UTR) had the strongest signal, with nine unexplained KAN-R isolates carrying resistance-exclusive mutations (Table 7). Each of the nine isolates carried a different *whiB7* 5′-UTR mutation. WhiB7 regulates *eis* (35), which is in turn associated with KAN-R. Two of these *whiB7* 5′-UTR mutations, *whiB7*:T-116C and *whiB7*:C-56T, were previously reported in KAN-R spontaneous mutants (32). Two of the nine isolates carried additional mutations in the coding DNA sequence (CDS) of *whiB7*, the missense mutation *whiB7*:E69A and an in-frame deletion of codons 84 through 87. While these are not nonsense mutations, if they still cause a loss of function, then these two KAN-R isolates remain unexplained despite their *whiB7* 5′-UTR mutations (27). None of the nine isolates carried mutations in the CDS of *eis*.

Meanwhile, seven unexplained CAP-R isolates carried CAP-R-exclusive mutations in the upstream region of *Rv2680* (Table 7). These seven mutations were all within 82 bp of both the *Rv2680* start codon and the *Rv2680* transcription start site reported previously by Cortes et al. (36). While the function of *Rv2680* is unknown, it is likely in an operon with *Rv2681* (37), a homologue of the exoribonuclease RNase D (38). *Rv2681* is on the same strand as *Rv2680*, and the start codon of *Rv2681* is 2 bp after the stop codon of *Rv2680*.

DISCUSSION

While known markers in the *eis* promoter and *rrs* were associated strongly with SLID-R, 111 SLID-R isolates lacked known markers in this study of 1,184 clinical *M. tuberculosis* isolates. Some of the discordant isolates may be due to errors in phenotypic DST. Phenotypic/genotypic DST results sometimes disagree (39) and have shown

an error rate as high as 2.2% for AMK (40). These discordant isolates also suggest the existence of rare, alternative mechanisms of resistance. However, identifying rare mechanisms is difficult as no single variant is carried by enough isolates for a strong association with resistance. To overcome this difficulty, we searched for rare mechanisms by aggregating the mutations around each gene.

The known marker *rrs*:A1401G was ubiquitous and strongly associated with cross-resistance to all three SLIDs (see Table S2 at https://doi.org/10.5281/zenodo.6149149). However, while the variant was carried by 246 CAP-R isolates, it was also carried by 22 CAP-S isolates (Table 4). The discordant isolates may still have low-level resistance to CAP, as a previous study reported a wide range of CAP MIC values (8 to 40 μg/mL) among clinical isolates carrying *rrs*:A1401G (41). However, the cause of this variable MIC is still unknown. The same study reported that mutagenesis of reference strains consistently resulted in a 40-μg/mL CAP MIC, suggesting that the inconsistency was due to the genetic background of the clinical isolates rather than the mechanism of *rrs*:A1401G itself (41). However, we observed no mutations common to the genetic background of our 22 discordant isolates.

The known marker *rrs*:G1484T was rare, carried by only four clinical isolates (Table 4), all cross-resistant. The low prevalence of *rrs*:G1484T has been reported previously despite a high MIC (14, 42). In a mutagenesis study on *Mycobacterium smegmatis*, *rrs*:G1484T mutants grew slower than *rrs*:A1401G mutants, suggesting that this rarity is due to a fitness cost of *rrs*:G1484T (43). However, a later study found that *rrs*:G1484T conferred no growth disadvantage to the *M. tuberculosis* reference strain H37Rv, although it did confer a disadvantage to a strain of the Beijing F2 sublineage (41). The full extent of the fitness cost of *rrs*:G1484T across clinical isolates is unknown.

KAN-R was most often explained by *rrs*:A1401G (Table 4), except among isolates from Moldova, which were enriched for known markers in the *eis* promoter (Table 6). Moldova is representative of countries in the region of the former Soviet Union, where this geographic trend has been reported previously (16). The prevalence of *eis* promoter mutations in this region is thought to be the result of the extensive use of KAN (16).

While known markers in the *eis* promoter and *rrs* were associated strongly with resistance, no other mutations within these genes were significantly associated with SLID-R in this isolate set (Fig. 2). The mutation *rrs*:C1402T was infrequent and carried by both resistant and susceptible isolates (Fig. 2). Mutagenesis has previously shown that *rrs*:C1402T reduces susceptibility to a level near the critical concentration (41). Both AMK-R and AMK-S isolates carried *rrs*:C1402T in a recent WHO study (18), and the marker is considered resistant by the GenoType MTBDRsl platform (Hain Lifescience).

The known *eis* promoter and *rrs* mutations (Table 1) were not carried by 21 of the 270 KAN-R isolates (Table 3), leaving the genetic basis of their resistance unexplained. Of these, seven isolates carried *whiB7* 5′-UTR mutations (Table 7), although these mutations were unique in each isolate. Thus, while no single *whiB7* 5′-UTR mutation was associated strongly with KAN-R, their aggregate signal suggests that *whiB7* 5′-UTR mutations are an alternative mechanism of KAN-R. This finding in clinical isolates is supported by previous mutagenesis experiments. Increased expression of *whiB7* causes low-level streptomycin resistance and KAN-R in H37Rv mutants (32), while deletion of *whiB7* in *Mycobacterium abscessus* lowers the MICs of erythromycin, tetracycline, streptomycin, and AMK (44). The *whiB7* gene encodes a transcriptional activator that regulates ribosome protection and efflux pump genes (44). Moreover, *whiB7* regulates *eis*, providing a plausible mechanism for KAN-R (32).

The promoter of the *Rv2680-Rv2681* operon was associated with unexplained CAP-R, with a signal comparable to that of the *whiB7* 5′ UTR and KAN-R (Table 7). While the function of *Rv2680* is unknown, Rv2681 is a homologue of the exoribonuclease RNase D (38). The sequence homology between *Rv2681* and RNase D (38) is further supported by the similarity of its predicted protein structure (45). Mutations in the promoter of the *Rv2680-Rv2681* operon may affect the transcription of *Rv2681*, in turn altering the structure of rRNA processed by the exoribonuclease encoded by *Rv2681*. CAP disrupts

protein synthesis by binding to the ribosome (25). By altering rRNA, these mutations may prevent CAP from binding to the ribosome. This potential mechanism would explain seven CAP-R isolates that lacked known CAP-R markers but carried mutations in the promoter of the *Rv2680-Rv2681* operon that were absent from all CAP-S isolates.

The known SLID-R markers are accurate predictors of resistance. However, they still do not explain all SLID-R cases. Rare, alternative mechanisms, such as *whiB7* 5′-UTR mutations, are likely responsible for these unexplained SLID-R isolates. For molecular diagnostics to fully replace phenotypic diagnostics, these rare mechanisms must also be understood. Finding these rare mechanisms will require sequences from larger sets of unexplained resistant isolates and more sensitive methods of association, such as the machine learning approaches employed previously (21) or aggregating mutations by gene as done here. This method independently corroborated the association between *whiB7* 5′-UTR mutations and KAN-R (44) and identified a new association between CAP-R and mutations in the promoter of the *Rv2680-Rv2681* operon, which encodes an exoribonuclease (38) and may impact the binding between CAP and the ribosome.

MATERIALS AND METHODS

Sample collection. As part of a previous study, 323 clinical *M. tuberculosis* isolates were collected for long-read PacBio sequencing (24). There were 89 isolates that originated from Hinduja National Hospital (PDHNH) in Mumbai, India; 89 that came from the Phthisiopneumology Institute (PPI) in Chisinau, Moldova; 48 that were from the Tropical Disease Foundation (TDF) in Manila, Philippines; and 97 that were from the National Health Laboratory Service of South Africa (NHLS) in Johannesburg, South Africa. All raw sequences were uploaded to the NCBI Sequence Read Archive (SRA) database under the BioProject accession number PRJNA353873. An additional 10 *M. tuberculosis* clinical isolates were collected from the Supranational Reference Laboratories in Stockholm, Sweden, and Antwerp, Belgium. These isolates were originally genotyped with a Hain Lifescience GenoType MTBDRsl line probe assay (9) and were chosen for sequencing due to discordance between their genotype and phenotypic DST results for any SLID. Another 851 whole-genome sequences were downloaded from the NCBI SRA database using the SRA Toolkit's fastqdump (46). These 851 downloaded raw reads were previously sequenced on Illumina short-read platforms (47–50).

Phenotypic drug susceptibility testing. DST for the PacBio-sequenced isolates was performed on the Bactec mycobacterial growth indicator tube (MGIT) 960 platform (BD Diagnostic Systems, Franklin Lakes, NJ, USA) using the 2008 WHO-recommended critical concentrations of 1.0 mg/L (AMK) and 2.5 mg/L (CAP/KAN) as described in previous studies (24, 51, 52). DST for Illumina-sequenced isolates was also performed on the MGIT 960 platform using contemporary WHO-recommended critical concentrations, as described previously (47–50). As of 2018, the recommended critical concentrations for AMK, CAP, and KAN remain at 1.0 mg/L, 2.5 mg/L, and 2.5 mg/L, respectively (53). Bacterial isolates were excluded from analysis if DST data were not available for at least one SLID.

DNA extraction and sequencing. The DNAs of all 333 isolates collected for long-read PacBio sequencing, including those from the WHO Supranational Reference Laboratories in Stockholm and Antwerp and the NCBI SRA database, were extracted as described in a previous study (54). The SMRT sequencing protocol was described previously (55, 56). Sixty-four isolates were later resequenced due to low coverage. DNA extraction for the 851 downloaded public genomes was previously described (47–50). The downloaded genomes were sequenced on the Illumina Genome Analyzer, MiSeq, or HiSeq platform.

Genome assembly, alignment, and variant calling. Genome assembly, alignment, and variant calling methods are described in the supplemental material at https://doi.org/10.5281/zenodo.6149149. Briefly, PBHoover (57) aligned 64 SMRT-sequenced isolates to H37Rv and called variants. Later, 269 SMRT-sequenced isolates were *de novo* assembled with canu (58) or HGAP2 (Pacific Biosciences), and their assembled genomes were then aligned to reference strain H37Rv using dnadiff (v1.3) (59) for variant calling, with the output converted to VCF format by a custom script, mummer-snps2vcf (https://gitlab.com/LPCDRP/mummer-extras/-/blob/master/src/mummer-snps2vcf). Reads from Illumina-sequenced isolates were aligned to H37Rv using bowtie2 (v2.2.4) (60), and variants were then called with VarScan2 (v2.3) (61).

Lineage identification. Mycobacterial interspersed repetitive-unit–variable-number tandem-repeat (MIRU-VNTR) analysis and spoligotyping were previously performed on the initial 323 SMRT-sequenced isolates collected (51). The 10 isolates sent from Stockholm and Antwerp and the 851 downloaded Illumina-sequenced isolates underwent MIRU-VNTR analysis and spoligotyping with MiruHero, a custom Python script (https://gitlab.com/LPCDRP/miru-hero). MiruHero used the rule-based criteria from TB-Insight (62) to classify lineages.

Identifying known resistance-conferring mutations. After variant calling, known SLID-R markers were searched for in the VCF file of each clinical isolate. The *eis* promoter mutations C-14T, C-12T, and G-10A are known KAN-R markers, and the *rrs* mutations G1484T and A1401G are known markers of resistance to all three SLIDs (24). The genomic positions and orientation of *rrs* and the *eis* promoter are noted

in Table S5 at https://doi.org/10.5281/zenodo.6149149 (63). Known markers and phenotypic DST were used to estimate the sensitivity and specificity of predicting resistance to each SLID using known markers. A 95% confidence interval was calculated for sensitivity and specificity estimates using the score method with continuity correction (64).

Genome-wide association. For each of the three drugs, separate genome-wide association studies were performed using a custom Python script (https://gitlab.com/LPCDRP/gwa) to identify novel markers for alternative mechanisms of resistance. To remove the overriding signal of known resistance markers, our GWAS excluded isolates that had their resistance explained by known markers (Table 1). This exclusion was necessary to avoid confounding associations with potential alternative mechanisms of resistance. Isolates with additional *eis* promoter resistance markers (Table 1) were excluded from our KAN-R analysis for similar reasons. We calculated the sensitivity and specificity of each variant's prediction of resistance to each SLID (the proportion of unexplained resistant isolates with the variant and the proportion of susceptible isolates without the variant, respectively). For each SLID, we identified variants absent from susceptible isolates and ranked them by the number of unexplained resistant isolates that carried them.

In the gene-based association, for each SLID and each gene, we counted the number of unexplained resistant isolates with at least one variant in that gene that was absent from isolates susceptible to that SLID. Genes below the mean for this count were removed. This count was then divided by the total number of unexplained resistant isolates with any variant in that gene. Genes with known markers of resistance to first-line drugs were excluded, as most SLID-R isolates are also resistant to first-line drugs due to prior treatment with first-line drug regimens.

ACKNOWLEDGMENTS

Phenotypic confirmation of 10 isolates without common molecular mechanisms of resistance was performed by Solomon Ghebremichael in the Department of Microbiology at the Public Health Agency of Sweden in Stockholm.

This project was funded by a grant (R01AI105185) from the National Institute of Allergy and Infectious Diseases.

REFERENCES

1. World Health Organization. 2021. Global tuberculosis report 2021. World Health Organization, Geneva, Switzerland.
2. World Health Organization. 2019. Global tuberculosis report 2019. Document number WHO/CDS/TB/2019.15. World Health Organization, Geneva, Switzerland.
3. World Health Organization. 2014. Companion handbook to the WHO guidelines for the programmatic management of drug-resistant tuberculosis. Document number WHO/HTM/TB/2014.11. World Health Organization, Geneva, Switzerland.
4. Palomino J, Martin A. 2014. Drug resistance mechanisms in Mycobacterium tuberculosis. Antibiotics (Basel) 3:317–340. https://doi.org/10.3390/antibiotics3030317.
5. Engström A, Perskvist N, Werngren J, Hoffner SE, Juréen P. 2011. Comparison of clinical isolates and in vitro selected mutants reveals that tlyA is not a sensitive genetic marker for capreomycin resistance in Mycobacterium tuberculosis. J Antimicrob Chemother 66:1247–1254. https://doi.org/10.1093/jac/dkr109.
6. Seddon JA, Godfrey-Faussett P, Jacobs K, Ebrahim A, Hesseling AC, Schaaf HS. 2012. Hearing loss in patients on treatment for drug-resistant tuberculosis. Eur Respir J 40:1277–1286. https://doi.org/10.1183/09031936.00044812.
7. World Health Organization. 2018. Rapid communication: key changes to treatment of multidrug- and rifampicin-resistant tuberculosis. Document number WHO/CDS/TB/2018.18. World Health Organization, Geneva, Switzerland.
8. World Health Organization. 2019. Rapid communication: key changes to the treatment of drug-resistant tuberculosis. Document number WHO/CDS/TB/2019.26. World Health Organization, Geneva, Switzerland.
9. Lacoma A, García-Sierra N, Prat C, Maldonado J, Ruiz-Manzano J, Haba L, Gavin P, Samper S, Ausina V, Domínguez J. 2012. GenoType MTBDRsl for molecular detection of second-line-drug and ethambutol resistance in Mycobacterium tuberculosis strains and clinical samples. J Clin Microbiol 50:30–36. https://doi.org/10.1128/JCM.05274-11.
10. Coll F, Phelan J, Hill-Cawthorne GA, Nair MB, Mallard K, Ali S, Abdallah AM, Alghamdi S, Alsomali M, Ahmed AO, Portelli S, Oppong Y, Alves A, Bessa TB, Campino S, Caws M, Chatterjee A, Crampin AC, Dheda K, Furnham N, Glynn JR, Grandjean L, Minh Ha D, Hasan R, Hasan Z, Hibberd ML, Joloba M, Jones-López EC, Matsumoto T, Miranda A, Moore DJ, Mocillo N, Panaiotov S, Parkhill J, Penha C, Perdigão J, Portugal I, Rchiad Z, Robledo J, Sheen P, Shesha NT, Sirgel FA, Sola C, Oliveira Sousa E, Streicher EM, Van Helden P, Viveiros M, Warren RM, McNerney R, Pain A, Clark TG. 2018. Genome-wide analysis of multi- and extensively drug-resistant Mycobacterium tuberculosis. Nat Genet 50:307–316. https://doi.org/10.1038/s41588-017-0029-0.
11. Kambli P, Ajbani K, Nikam C, Sadani M, Shetty A, Udwadia Z, Georghiou SB, Rodwell TC, Catanzaro A, Rodrigues C. 2016. Correlating rrs and eis promoter mutations in clinical isolates of Mycobacterium tuberculosis with phenotypic susceptibility levels to the second-line injectables. Int J Mycobacteriol 5:1–6. https://doi.org/10.1016/j.ijmyco.2015.09.001.
12. Brossier F, Pham A, Bernard C, Aubry A, Jarlier V, Veziris N, Sougakoff W, CNR-MyRMA. 2017. Molecular investigation of resistance to second-line injectable drugs in multidrug-resistant clinical isolates of Mycobacterium tuberculosis in France. Antimicrob Agents Chemother 61:e01299-16. https://doi.org/10.1128/AAC.01299-16.
13. Jugheli L, Bzekalava N, de Rijk P, Fissette K, Portaels F, Rigouts L. 2009. High level of cross-resistance between kanamycin, amikacin, and capreomycin among Mycobacterium tuberculosis isolates from Georgia and a close relation with mutations in the rrs gene. Antimicrob Agents Chemother 53:5064–5068. https://doi.org/10.1128/AAC.00851-09.
14. Perdigão J, Portugal I. 2019. Genetics and roadblocks of drug resistant tuberculosis. Infect Genet Evol 72:113–130. https://doi.org/10.1016/j.meegid.2018.09.023.
15. Zaunbrecher MA, Sikes RD, Jr, Metchock B, Shinnick TM, Posey JE. 2009. Overexpression of the chromosomally encoded aminoglycoside acetyltransferase eis confers kanamycin resistance in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 106:20004–20009. https://doi.org/10.1073/pnas.0907925106.
16. Gikalo MB, Nosova EY, Krylova LY, Moroz AM. 2012. The role of eis mutations in the development of kanamycin resistance in Mycobacterium tuberculosis isolates from the Moscow region. J Antimicrob Chemother 67:2107–2109. https://doi.org/10.1093/jac/dks178.
17. Pricer RE, Houghton JL, Green KD, Mayhoub AS, Garneau-Tsodikova S. 2012. Biochemical and structural analysis of aminoglycoside acetyltransferase Eis from Anabaena variabilis. Mol Biosyst 8:3305–3313. https://doi.org/10.1039/c2mb25341k.
18. World Health Organization. 2021. Catalogue of mutations in Mycobacterium tuberculosis complex and their association with drug resistance. World Health Organization, Geneva, Switzerland. https://www.who.int/publications/i/item/9789240028173.

19. Maus CE, Plikaytis BB, Thomas M, Shinnick TM. 2005. Mutation of tlyA confers capreomycin resistance in Mycobacterium tuberculosis. Antimicrob Agents Chemother 49:571–577. https://doi.org/10.1128/AAC.49.2.571-577.2005.
20. Köser CU, Bryant JM, Parkhill J, Peacock SJ. 2013. Consequences of whiB7 (Rv3197A) mutations in Beijing genotype isolates of the Mycobacterium tuberculosis complex. Antimicrob Agents Chemother 57:3461. https://doi.org/10.1128/AAC.00626-13.
21. Kavvas ES, Catoiu E, Mih N, Yurkovich JT, Seif Y, Dillon N, Heckmann D, Anand A, Yang L, Nizet V, Monk JM, Palsson BO. 2018. Machine learning and structural analysis of Mycobacterium tuberculosis pan-genome identifies genetic signatures of antibiotic resistance. Nat Commun 9:4306. https://doi.org/10.1038/s41467-018-06634-y.
22. Sharma D, Kumar B, Lata M, Joshi B, Venkatesan K, Shukla S, Bisht D. 2015. Comparative proteomic analysis of aminoglycosides resistant and susceptible Mycobacterium tuberculosis clinical isolates for exploring potential drug targets. PLoS One 10:e0139414. https://doi.org/10.1371/journal.pone.0139414.
23. Sharma D, Lata M, Faheem M, Khan AU, Joshi B, Venkatesan K, Shukla S, Bisht D. 2016. M. tuberculosis ferritin (Rv3841): potential involvement in amikacin (AK) & kanamycin (KM) resistance. Biochem Biophys Res Commun 478:908–912. https://doi.org/10.1016/j.bbrc.2016.08.049.
24. Rodwell TC, Valafar F, Douglas J, Qian L, Garfein RS, Chawla A, Torres J, Zadorozhny V, Kim MS, Hoshide M, Catanzaro D, Jackson L, Lin G, Desmond E, Rodrigues C, Eisenach K, Victor TC, Ismail N, Crudu V, Gler MT, Catanzaro A. 2014. Predicting extensively drug-resistant Mycobacterium tuberculosis phenotypes with genetic mutations. J Clin Microbiol 52:781–789. https://doi.org/10.1128/JCM.02701-13.
25. Johansen SK, Maus CE, Plikaytis BB, Douthwaite S. 2006. Capreomycin binds across the ribosomal subunit interface using tlyA-encoded 2′-O-methylations in 16S and 23S rRNAs. Mol Cell 23:173–182. https://doi.org/10.1016/j.molcel.2006.05.044.
26. Wang Q, Boshoff HIM, Harrison JR, Ray PC, Green SR, Wyatt PG, Barry CE. 2020. PE/PPE proteins mediate nutrient transport across the outer membrane of Mycobacterium tuberculosis. Science 367:1147–1151. https://doi.org/10.1126/science.aav5912.
27. Vargas R, Jr, Freschi L, Spitaleri A, Tahseen S, Barilar I, Niemann S, Miotto P, Cirillo DM, Köser CU, Farhat MR. 2021. Role of epistasis in amikacin, kanamycin, bedaquiline, and clofazimine resistance in Mycobacterium tuberculosis complex. Antimicrob Agents Chemother 65:e01164-21. https://doi.org/10.1128/AAC.01164-21.
28. Wang Y, Li Q, Gao H, Zhang Z, Liu Y, Lu J, Dai E. 2019. The roles of rpsL, rrs, and gidB mutations in predicting streptomycin-resistant drugs used on clinical Mycobacterium tuberculosis isolates from Hebei Province, China. Int J Clin Exp Pathol 12:2713–2721.
29. Li Q, Gao H, Zhang Z, Tian Y, Liu T, Wang Y, Lu J, Liu Y, Dai E. 2019. Mutation and transmission profiles of second-line drug resistance in clinical isolates of drug-resistant Mycobacterium tuberculosis from Hebei Province, China. Front Microbiol 10:1838. https://doi.org/10.3389/fmicb.2019.01838.
30. Tyler AD, Christianson S, Knox NC, Mabon P, Wolfe J, Van Domselaar G, Graham MR, Sharma MK. 2016. Comparison of sample preparation methods used for the next-generation sequencing of Mycobacterium tuberculosis. PLoS One 11:e0148676. https://doi.org/10.1371/journal.pone.0148676.
31. Cloete R, Kapp E, Joubert J, Christoffels A, Malan SF. 2018. Molecular modelling and simulation studies of the Mycobacterium tuberculosis multidrug efflux pump protein Rv1258c. PLoS One 13:e0207605. https://doi.org/10.1371/journal.pone.0207605.
32. Reeves AZ, Campbell PJ, Sultana R, Malik S, Murray M, Plikaytis BB, Shinnick TM, Posey JE. 2013. Aminoglycoside cross-resistance in Mycobacterium tuberculosis due to mutations in the 5′ untranslated region of whiB7. Antimicrob Agents Chemother 57:1857–1865. https://doi.org/10.1128/AAC.02191-12.
33. Farhat MR, Freschi L, Calderon R, Ioerger T, Snyder M, Meehan CJ, de Jong B, Rigouts L, Sloutsky A, Kaur D, Sunyaev S, van Soolingen D, Shendure J, Sacchettini J, Murray M. 2019. GWAS for quantitative resistance phenotypes in Mycobacterium tuberculosis reveals resistance genes and regulatory regions. Nat Commun 10:2128. https://doi.org/10.1038/s41467-019-10110-6.
34. Oppong YEA, Phelan J, Perdigão J, Machado D, Miranda A, Portugal I, Viveiros M, Clark TG, Hibberd ML. 2019. Genome-wide analysis of Mycobacterium tuberculosis polymorphisms reveals lineage-specific associations with drug resistance. BMC Genomics 20:252. https://doi.org/10.1186/s12864-019-5615-3.
35. Burian J, Yim G, Hsing M, Axerio-Cilies P, Cherkasov A, Spiegelman GB, Thompson CJ. 2013. The mycobacterial antibiotic resistance determinant WhiB7 acts as a transcriptional activator by binding the primary sigma factor SigA (RpoV). Nucleic Acids Res 41:10062–10076. https://doi.org/10.1093/nar/gkt751.
36. Cortes T, Schubert OT, Rose G, Arnvig KB, Comas I, Aebersold R, Young DB. 2013. Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in Mycobacterium tuberculosis. Cell Rep 5:1121–1131. https://doi.org/10.1016/j.celrep.2013.10.031.
37. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ, von Mering C. 2011. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res 39:D561–D568. https://doi.org/10.1093/nar/gkq973.
38. Zuo Y, Wang Y, Malhotra A. 2005. Crystal structure of Escherichia coli RNase D, an exoribonuclease involved in structured RNA processing. Structure 13:973–984. https://doi.org/10.1016/j.str.2005.04.015.
39. Hillemann D, Hoffner S, Cirillo D, Drobniewski F, Richter E, Rüsch-Gerdes S, Baltic-Nordic TB-Laboratory Network, TB PAN-NET, ECDC ERLN-TB Networks. 2013. First evaluation after implementation of a quality control system for the second line drug susceptibility testing of Mycobacterium tuberculosis joint efforts in low and high incidence countries. PLoS One 8:e76765. https://doi.org/10.1371/journal.pone.0076765.
40. Nikolayevskyy V, Hillemann D, Richter E, Ahmed N, van der Werf MJ, Kodmon C, Drobniewski F, Ruesch-Gerdes S, ERLTB-Net Network. 2016. External quality assessment for tuberculosis diagnosis and drug resistance in the European Union: a five year multicentre implementation study. PLoS One 11:e0152926. https://doi.org/10.1371/journal.pone.0152926.
41. Reeves AZ, Campbell PJ, Willby MJ, Posey JE. 2015. Disparities in capreomycin resistance levels associated with the rrs A1401G mutation in clinical isolates of Mycobacterium tuberculosis. Antimicrob Agents Chemother 59:444–449. https://doi.org/10.1128/AAC.04438-14.
42. Georghiou SB, Magana M, Garfein RS, Catanzaro DG, Catanzaro A, Rodwell TC. 2012. Evaluation of genetic mutations associated with Mycobacterium tuberculosis resistance to amikacin, kanamycin and capreomycin: a systematic review. PLoS One 7:e33275. https://doi.org/10.1371/journal.pone.0033275.
43. Shcherbakov D, Akbergenov R, Matt T, Sander P, Andersson DI, Böttger EC. 2010. Directed mutagenesis of Mycobacterium smegmatis 16S rRNA to reconstruct the in vivo evolution of aminoglycoside resistance in Mycobacterium tuberculosis. Mol Microbiol 77:830–840. https://doi.org/10.1111/j.1365-2958.2010.07218.x.
44. Hurst-Hess K, Rudra P, Ghosh P. 2017. Mycobacterium abscessus WhiB7 regulates a species-specific repertoire of genes to confer extreme antibiotic resistance. Antimicrob Agents Chemother 61:e01347-17. https://doi.org/10.1128/AAC.01347-17.
45. Modlin SJ, Elghraoui A, Gunasekaran D, Zlotnicki AM, Dillon NA, Dhillon N, Kuo N, Robinhold C, Chan CK, Baughn AD, Valafar F. 2021. Structure-aware Mycobacterium tuberculosis functional annotation uncloaks resistance, metabolic, and virulence genes. mSystems 6:e00673-21. https://doi.org/10.1128/mSystems.00673-21.
46. Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database Collaboration. 2011. The Sequence Read Archive. Nucleic Acids Res 39:D19–D21. https://doi.org/10.1093/nar/gkq1019.
47. Walker TM, Kohl TA, Omar SV, Hedge J, Del Ojo Elias C, Bradley P, Iqbal Z, Feuerriegel S, Niehaus KE, Wilson DJ, Clifton DA, Kapatai G, Ip CLC, Bowden R, Drobniewski FA, Allix-Béguec C, Gaudin C, Parkhill J, Diel R, Supply P, Crook DW, Smith EG, Walker AS, Ismail N, Niemann S, Peto TEA, Modernizing Medical Microbiology (MMM) Informatics Group. 2015. Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study. Lancet Infect Dis 15:1193–1202. https://doi.org/10.1016/S1473-3099(15)00062-6.
48. Casali N, Nikolayevskyy V, Balabanova Y, Ignatyeva O, Kontsevaya I, Harris SR, Bentley SD, Parkhill J, Nejentsev S, Hoffner SE, Horstmann RD, Brown T, Drobniewski F. 2012. Microevolution of extensively drug-resistant tuberculosis in Russia. Genome Res 22:735–745. https://doi.org/10.1101/gr.128678.111.
49. Casali N, Nikolayevskyy V, Balabanova Y, Harris SR, Ignatyeva O, Kontsevaya I, Corander J, Bryant J, Parkhill J, Nejentsev S, Horstmann RD, Brown T, Drobniewski F. 2014. Evolution and transmission of drug-resistant tuberculosis in a Russian population. Nat Genet 46:279–286. https://doi.org/10.1038/ng.2878.
50. Chernyaeva EN, Shulgina MV, Rotkevich MS, Dobrynin PV, Simonov SA, Shitikov EA, Ischenko DS, Karpova IY, Kostryukova ES, Ilina EN, Govorun

VM, Zhuravlev VY, Manicheva OA, Yablonsky PK, Isaeva YD, Nosova EY, Mokrousov IV, Vyazovaya AA, Narvskaya OV, Lapidus AL, O'Brien SJ. 2014. Genome-wide Mycobacterium tuberculosis variation (GMTV) database: a new tool for integrating sequence variations and epidemiology. BMC Genomics 15:308. https://doi.org/10.1186/1471-2164-15-308.
51. Garfein RS, Catanzaro DG, Rodwell TC, Avalos E, Jackson RL, Kaping J, Evasco H, Rodrigues C, Crudu V, Lin S-YG, Groessl E, Groessel E, Hillery N, Trollip A, Ganiats T, Victor TC, Eisenach K, Valafar F, Channick J, Qian L, Catanzaro A. 2015. Phenotypic and genotypic diversity in a multinational sample of drug-resistant Mycobacterium tuberculosis isolates. Int J Tuberc Lung Dis 19:420–427. https://doi.org/10.5588/ijtld.14.0488.
52. World Health Organization. 2008. Policy guidance on drug-susceptibility testing (DST) of second-line antituberculosis drugs. World Health Organization, Geneva, Switzerland.
53. World Health Organization. 2018. Technical report on critical concentrations for TB drug susceptibility testing of medicines used in the treatment of drug-resistant TB. World Health Organization, Geneva, Switzerland.
54. van Helden PD, Victor TC, Warren RM, van Helden EG. 2001. Isolation of DNA from Mycobacterium tuberculosis. Methods Mol Med 54:19–29. https://doi.org/10.1385/1-59259-147-7:019.
55. Torres JN, Paul LV, Rodwell TC, Victor TC, Amallraja AM, Elghraoui A, Goodmanson AP, Ramirez-Busby SM, Chawla A, Zadorozhny V, Streicher EM, Sirgel FA, Catanzaro D, Rodrigues C, Gler MT, Crudu V, Catanzaro A, Valafar F. 2015. Novel katG mutations causing isoniazid resistance in clinical M. tuberculosis isolates. Emerg Microbes Infect 4:e42. https://doi.org/10.1038/emi.2015.42.
56. Elghraoui A, Modlin SJ, Valafar F. 2017. SMRT genome assembly corrects reference errors, resolving the genetic basis of virulence in Mycobacterium tuberculosis. BMC Genomics 18:302. https://doi.org/10.1186/s12864-017-3687-5.
57. Ramirez-Busby SM, Elghraoui A, Kim YB, Valafar F. 2018. PBHoover and CigarRoller: a method for confident haploid variant calling on legacy Pacific Biosciences data and its application to heterogeneous population analysis. bioRxiv. https://doi.org/10.1101/360370.
58. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.
59. Delcher AL, Kasif S, Fleischmann RD, Peterson J, White O, Salzberg SL. 1999. Alignment of whole genomes. Nucleic Acids Res 27:2369–2376. https://doi.org/10.1093/nar/27.11.2369.
60. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
61. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22:568–576. https://doi.org/10.1101/gr.129684.111.
62. Shabbeer A, Cowan LS, Ozcaglar C, Rastogi N, Vandenberg SL, Yener B, Bennett KP. 2012. TB-Lineage: an online tool for classification and analysis of strains of Mycobacterium tuberculosis complex. Infect Genet Evol 12:789–797. https://doi.org/10.1016/j.meegid.2012.02.010.
63. Kapopoulou A, Lew JM, Cole ST. 2011. The MycoBrowser portal: a comprehensive and manually annotated resource for mycobacterial genomes. Tuberculosis (Edinb) 91:8–13. https://doi.org/10.1016/j.tube.2010.09.006.
64. Newcombe RG. 1998. Two-sided confidence intervals for the single proportion: comparison of seven methods. Stat Med 17:857–872. https://doi.org/10.1002/(SICI)1097-0258(19980430)17:8%3C857::AID-SIM777%3E3.0.CO;2-E.