




Rapid Identification of Drug Resistance and Phylogeny in *M. tuberculosis*, Directly from Sputum Samples

Martín Barbosa-Amezcuca,^a Betzaida Cuevas-Córdoba,^{a,b} Cristóbal Fresno,^{c,d} Joshua I. Haase-Hernández,^c Karol Carrillo-Sánchez,^e Minerva Mata-Rocha,^f Marcela Muñoz-Torrero,^g Claudia Bäcker,^h Vanessa González-Covarrubias,^a Carmen Alaez-Verson,^e  Xavier Soberón^{a,i}

^aLaboratorio de Farmacogenómica, Instituto Nacional de Medicina Genómica (INMEGEN), Mexico City, México

^bInstituto de Investigaciones Biológicas, Universidad Veracruzana, Xalapa, Veracruz, México

^cDepartamento de Desarrollo Tecnológico, Instituto Nacional de Medicina Genómica (INMEGEN), Mexico City, México

^dCentro de Investigación en Ciencias de la Salud (CICSA), Universidad Anáhuac México, Naucalpan de Juárez, México

^eLaboratorio de Diagnóstico Genómico, Instituto Nacional de Medicina Genómica (INMEGEN), Mexico City, México

^fUnidad de Investigación Médica en Epidemiología Clínica, UMAE Hospital de Pediatría, Centro Médico Nacional "Siglo XXI", Instituto Mexicano del Seguro Social (IMSS), Mexico City, México

^gClínica de Tuberculosis y Enfermedades Pleurales, Instituto Nacional de Enfermedades Respiratorias (INER), Mexico City, México

^hLaboratorio de Micobacterias, Instituto de Diagnóstico y Referencia Epidemiológicos (InDRE), Secretaría de Salud, Mexico City, México

ⁱInstituto de Biotecnología, Universidad Nacional Autónoma de México (UNAM), Cuernavaca, Morelos, México

ABSTRACT Tuberculosis (TB) remains one of the most important infectious diseases globally. Establishing a resistance profile from the initial TB diagnosis is a priority. Rapid molecular tests evaluate only the most common genetic variants responsible for resistance to certain drugs, and Whole Genome Sequencing (WGS) needs culture prior to next-generation sequencing (NGS), limiting their clinical value. Targeted sequencing (TS) from clinical samples avoids these drawbacks, providing a signature of genetic markers that can be associated with drug resistance and phylogeny. In this study, a proof-of-concept protocol was developed for detecting genomic variants associated with drug resistance and for the phylogenetic classification of *Mycobacterium Tuberculosis* (Mtb) in sputum samples. Initially, a set of Mtb reference strains from the WHO were sequenced (WGS and TS). The results from the protocol agreed >95% with WHO reported data and phenotypic drug susceptibility testing (pDST). Lineage genetics results were 100% concordant with those derived from WGS. After that, the TS protocol was applied to sputum samples from TB patients to detect resistance to first- and second-line drugs and derive phylogeny. The accuracy was >90% for all evaluated drugs, except Eto/Pto (77.8%), and 100% were phylogenetically classified. The results indicate that the described protocol, which affords the complete drug resistance profile and phylogeny of Mtb from sputum, could be useful in the clinical area, advancing toward more personalized and more effective treatments in the near future.

IMPORTANCE The COVID-19 pandemic negatively affected the progress in accessing essential Tuberculosis (TB) services and reducing the burden of TB disease, resulting in a decreased detection of new cases and increased deaths. Generating molecular diagnostic tests with faster results without losing reliability is considered a priority. Specifically, developing an antimicrobial resistance profile from the initial stages of TB diagnosis is essential to ensure appropriate treatment. Currently available rapid molecular tests evaluate only the most common genetic variants responsible for resistance to certain drugs, limiting their clinical value. In this work, targeted sequencing on sputum samples from TB patients was used to identify *Mycobacterium tuberculosis* mutations in genes associated with drug resistance and to derive a phylogeny of the infecting strain. This protocol constitutes a proof-of-concept toward the goal of helping clinicians select a timely and appropriate treatment by providing them with actionable information beyond current molecular approaches.

Editor S. Wesley Long, Houston Methodist Hospital

Copyright © 2022 Barbosa-Amezcuca et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Xavier Soberón, xavier.soberon@ibt.unam.mx.

The authors declare no conflict of interest.

Received 13 April 2022

Accepted 5 August 2022

Published 14 September 2022

KEYWORDS drug resistance, NGS, targeted sequencing, tuberculosis

Tuberculosis (TB) is the 13th global cause of death and, before COVID-19, the first one caused by a single infectious agent. In recent years, the number of new cases has remained stable, with a slight increase in deaths from TB in 2020, despite implementing policies to control it (1).

Although antituberculosis treatments have been available for decades, control of TB has not yet been achieved due, in part, to drug resistance. Drug-resistant tuberculosis (DR-TB) will cause a quarter of the deaths associated with this condition by 2050 (2). In 2019, the WHO (3) reported that 3.4% of new TB cases and 18% of previously treated ones were resistant to rifampicin (RIF) and isoniazid (INH) (MDR), or at least resistant to RIF (RR-TB). Additionally, 6.2% of MDR cases were XDR (MDR with additional resistance to at least one fluoroquinolone and one injectable). From 2018 to 2020, only 32% of people diagnosed with MDR/RR-TB received adequate treatment (1).

Establishing a resistance profile through drug susceptibility testing (DST) from the initial TB diagnosis is essential to ensure appropriate treatment, thus reducing the selection and transmission of resistant *Mtb*, therapeutic failure, and death. Currently, phenotypic DST (pDST) based on bacterial culture is being replaced by genetic tests, such as the Xpert MTB/RIF in its different versions and Line Probe Assays (LiPA), changing the turnaround of drug susceptibility results from several weeks to days or hours (4). However, these molecular methods evaluate only the most common genetic variants responsible for resistance to certain drugs, limiting their clinical value.

The development of rapid and reliable methods for the early determination of DR-TB remains a necessity. Next-generation sequencing (NGS) has the potential to speed up and simultaneously detect all known resistance markers, generating a complete resistance profile, which will contribute to the design of effective treatments from the beginning (5).

Whole Genome Sequencing (WGS) of clinical samples is an expensive and complicated process due to the low bacillary load, which results in high human to bacterial DNA content ratio with consequently low depth of coverage and poor identification of resistance (6). Culture in specialized media for *Mtb* before NGS ensures sufficient genetic material, but at the expense of limiting the speed at which clinically relevant data are obtained, plus the requirement of biosafety level 3 facilities.

Culture time and expenses can be reduced by sequencing only regions associated with drug resistance (7). Targeted sequencing (TS) panels have shown effectiveness in obtaining resistance profiles directly from clinical samples, avoiding the need for an initial culture, providing a signature of genetic markers associated with drug resistance, and promoting a personalized treatment (8, 9). TS also allows increasing the sequencing depth, facilitating the identification of subpopulations of susceptible and resistant bacteria (heteroresistance), which could impact the early diagnosis of DR-TB (8, 10).

Although the genomes of different members of the *Mycobacterium tuberculosis* Complex (MTBC) are more than 99.9% identical to each other (11, 12), MTBC strains exhibit important differences in intrinsic resistance to antibiotics, development of drug resistance, pathogenicity, transmissibility, virulence, and the immune response they elicit. These differences that affect the clinical presentation of TB are associated with different phylogenetic lineages (13). Several molecular methods have been used to identify the MTBC, assigning them to lineages (spoligo-typing, IS6110-RFLP, MIRU-VNTR, and sequence-based detection of single-nucleotide polymorphisms [SNPs] [14]). SNPs are stable, phylogenetically informative markers that proved capable of differentiating with great discriminatory power the main MTBC lineages and can be easily obtained with an NGS approach (15, 16). Their use for lineage identification is straightforward and can contribute to better clinical management and TB treatment.

Genomic technologies that speed up the diagnosis and simultaneously identify drug-resistant *Mtb* and lineages will help reduce the spread of resistant strains and achieve better disease control by allowing specific and timely treatment.

This work aimed to develop a proof-of-concept protocol for detecting genomic variants associated with drug resistance and phylogenetic classification of *Mtb* in sputum samples.

TABLE 1 Indicators of the diagnostic test obtained by WGS^a compared to pDST^b of Mtb reference strains

Drug	Resistance identified by:		Sensitivity (95% CI ^g)	Specificity (95% CI)	PPV ^c (95% CI)	NPV ^d (95% CI)	Accuracy (95% CI)	Cohen's Kappa (95% CI)
	Phenotype ^e	Genotype ^f						
RIF	13/20	13/20	100 (96.2–100)	100 (92.7–100)	100 (96.2–100)	100 (92.7–100)	100 (97.5–100)	1.0 (1.0–1.0)
INH	14/20	13/20	92.9 (75.8–100)	100 (91.7–100)	100 (96.16–100)	85.7 (52.7–100)	95.0 (83.0–100)	0.89 (0.7–1.0)
PZA	9/20	9/20	100 (94.4–100)	100 (95.5–100)	100 (94.4–100)	100 (95.5–100)	100 (97.5–100)	1.0 (1.0–1.0)
EMB	11/20	11/20	100 (95.5–100)	100 (94.4–100)	100 (95.5–100)	100 (94.4–100)	100 (97.5–100)	1.0 (1.0–1.0)
SM	9/20	9/20	100 (94.4–100)	100 (95.5–100)	100 (94.4–100)	100 (95.5–100)	100 (97.5–100)	1.0 (1.0–1.0)
FQs ^h	7/20	7/20	100 (92.9–100)	100 (96.2–100)	100 (92.9–100)	100 (96.2–100)	100 (97.6–100)	1.0 (1.0–1.0)
AM ⁱ	7/20	7/20	100 (92.9–100)	100 (96.2–100)	100 (92.9–100)	100 (96.2–100)	100 (97.6–100)	1.0 (1.0–1.0)
ETH/PTH	3/20	3/20	100 (83.3–100)	100 (97.1–100)	100 (83.3–100)	100 (97.1–100)	100 (97.5–100)	1.0 (1.0–1.0)

^aWGS, whole genome sequencing.^bpDST, phenotypic drug susceptibility testing.^cPPV, positive predictive value.^dNPV, negative predictive value.^eResistant samples identified by pDST/Samples with pDST results.^fResistant samples identified by sequencing/Samples with sequencing results.^gCI, confidence interval.^hFQs, fluoroquinolones (includes MFX, OFX, and LFX).ⁱAM, aminoglycosides (includes kanamycin [KAN], amikacin [AMK] and capreomycin [CAP]).

RESULTS

Reference strains. As a first step, the protocol was tested by identifying genetic variants in a set of well-characterized strains. The drug resistance and phylogeny databases were employed to assign them drug resistance profiles and lineage classification. The results were compared with the phenotypic results for drug resistance and also with WGS-derived spoligotyping *in silico* for lineage.

(i) Sequencing process. All reference strains ($n = 30$) were processed, and the DNA recovered was 34.4 ng/ μ L, on average, which was sufficient for the preparation of TS libraries. Libraries were generated with TruSeq Custom Amplicon and sequenced, obtaining a mean depth of coverage of 2040x and >99% of curated reads mapped to the reference genome.

For the WGS libraries, 3 WHO reference strains and one *H37Rv* showed low DNA concentration and precluded the generation of WGS. The DNA recovered from the 26 strains after library generation with the Nextera XT system was, on average, 9.0 ng/ μ L. Sequencing afforded, after curating reads, a mean depth of coverage of 63x with >99% of reads mapping to the reference genome.

(ii) Prediction of drug resistance and phylogeny. The 28 WHO reference strains comprised only 18 genetically different samples: variants obtained by TS revealed that 10 samples were represented twice. That these were indeed repeated samples (a normal practice with validating panels of the WHO) was confirmed with data derived from WGS. These results were not known in advance by the experimenters because a simple blinded approach was used intentionally to assure the reproducibility of the proposed method. Only the results from 20 reference strains, 18 unique from WHO set, plus two *H37Rv*, are reported in the TS results. Regarding the strains used for WGS, only 17 were available due to the previous exclusion of 3 samples during library generation (16 different WHO strains and one *H37Rv*) and were included for subsequent analysis.

Variants from TS were identified, and the genotypic resistance was determined. In 20 independent samples, statistical indicators of diagnostic tests were calculated for the drugs evaluated. All indicators were 100% for RIF, pyrazinamide (PZA), streptomycin (SM), fluoroquinolones (FQs), and aminoglycosides (AM), showing lower concordance for INH (92.9%) (Table 1). Cohen's kappa coefficient in all drugs showed very good concordance (>0.89)

TABLE 2 Lineage prediction of the Mtb reference strains by TS or WGS^a, validated with SpoTyping

Frequency	Lineage identified by:		
	Targeted-SNPs ^b	WGS-SNPs	SpoTyping
1	Wild-type ^c	Wild-type	<i>H37Rv</i>
2	Haarlem	Haarlem	T1, H1
2	East African Indian	East African Indian	CAS1-Delhi
2	Indo Oceanic	Indo Oceanic	EAI1-SOM, EAI6-BGD1
4	LAM	LAM	LAM1, LAM6, T5-RUS1
5	East Asian	East Asian	Beijing
1	East Asian	East Asian	ND ^d
2	Wild-type	Not sequenced by WGS	
1	East Asian	Not sequenced by WGS	

^aWGS, whole genome sequencing.^bSNPs, single nucleotide polymorphisms.^cWild-type, strain without phylogenetic variants (identical to *H37Rv*).^dND, not determined.

between the results predicted from the genotypic tests and those obtained by pDST. Except for a false negative sample to INH, all identified resistance variants were congruent with the WHO report for these reference strains.

Reads obtained from WGS of 17 Mtb strains were analyzed with the SpoTyping program to get *in silico* spoligotyping results. These results completely agree with the proposed phylogenetic variants set, although *in silico* spoligotyping results, which use the whole genome, had better resolution (Table 2).

The *H37Rv* strain is considered the wild-type strain by our protocol and, by definition, contained no phylogenetic variants. An East Asian strain was identified with our phylogenetic variants set but not by SpoTyping (ND). Three strains (2 non-duplicated WHO reference strains and one *H37Rv*) with low DNA concentration were not sequenced by WGS (Table 2).

Clinical samples. After the initial validation with the reference strains, the protocol and the databases were applied to clinical, sputum samples, in order to provide a proof-of-concept for its prospective use in the clinical area.

(i) Sequencing process. The study included 46 clinical samples. Of them, 23 had pDST for first- and second-line drugs by Bactec MGIT 960; 21 had results for first-line drugs only, and 2 were negative controls by smear microscopy and negative-culture (Table S4).

On average, the DNA recovered was 40.8 ng/ μ L. The TS libraries of 46 sputum samples were obtained with TruSeq Custom Amplicon, and the DNA recovered after libraries generation was an average of 30.5 ng/ μ L. After sequencing, quality-filtering and mapping to the Mtb reference genome, an average coverage depth of 862x was achieved.

70% of the sputum samples (31 of 44) were helpful for the subsequent analyses, such as variant calling and annotation. In these samples, about 55% of the reads mapped to the Mtb reference genome. However, from the remaining 15 sputum samples (including 2 negative controls), the reads obtained showed poor mapping to *H37Rv* (<10% of reads), resulting in a low depth of coverage (<22x) and a high percentage of lost variants (>20%) during GATK analysis, probably due to low Mtb load. These samples were discarded from further analysis.

(ii) Prediction of drug resistance. The genotypic resistance profile for each drug was obtained and compared to the gold standard, pDST results. Table 3 shows a high sensitivity to first-line drugs (>80%), except for SM (50%). Specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy were high (>70%) for all the drugs analyzed.

All clinical samples in which Xpert MTB/RIF identified resistance ($n = 14$) or sensitivity ($n = 4$) were concordant with the TS system and the bioinformatic analysis developed in this study (sensitivity and specificity of 100%).

Cohen's kappa coefficient showed very good concordance (0.80-1.0) between phenotypic and genotypic results for RIF, INH, and ofloxacin (OFX)/levofloxacin (LFX). Good agreement (0.60-0.79) was identified for PZA, ethambutol (EMB), SM, and moxifloxacin (MFX); and

TABLE 3 Indicators of the diagnostic test obtained by TS compared to pDST^a of clinical samples

Drug	Resistance identified by:		Sensitivity (95% CI) ^f	Specificity (95% CI)	PPV ^b (95% CI)	NPV ^c (95% CI)	Accuracy (95% CI)	Cohen's Kappa (95% CI)
	Phenotype ^d	Genotype ^e						
RIF	20/31	21/31	100 (97.5–100)	90.9 (69.4–100)	95.2 (83.8–100)	100 (95.0–100)	96.8 (88.9–100)	0.93 (0.8–1.0)
INH	12/31	10/31	83.3 (58.0–100)	100 (97.4–100)	100 (95.0–100)	90.5 (75.5–100)	93.6 (83.3–100)	0.86 (0.7–1.0)
PZA	5/31	7/31	100 (90.0–100)	92.3 (80.1–100)	71.4 (30.8–100)	100 (97.9–100)	93.6 (83.3–100)	0.79 (0.5–1.0)
EMB	5/31	8/31	100 (90.0–100)	88.5 (74.3–100)	62.5 (22.7–100)	100 (97.8–100)	90.3 (78.3–100)	0.71 (0.4–1.0)
SM	6/31	3/31	50.0 (1.7–98.3)	100 (98.0–100)	100 (83.3–100)	89.3 (76.0–100)	90.3 (78.3–100)	0.62 (0.2–1.0)
MFX	4/11	3/31	75.0 (20.0–100)	100 (92.9–100)	100 (83.3–100)	87.5 (58.3–100)	90.9 (69.4–100)	0.79 (0.2–1.0)
OFX/LFX	4/18	3/31	75.0 (20.0–100)	100 (96.4–100)	100 (83.3–100)	93.3 (77.4–100)	94.4 (81.0–100)	0.82 (0.5–1.0)
KAN	1/18	0/31	ND ^g	ND	ND	ND	ND	ND
AMK	1/18	0/31	ND	ND	ND	ND	ND	ND
CAP	1/11	0/31	ND	ND	ND	ND	ND	ND
ETH/PTH	7/18	3/31	42.8 (0.0–86.7)	100 (95.5–100)	100 (83.3–100)	73.3 (47.6–99.0)	77.8 (55.8–99.8)	0.48 (0.1–0.9)

^apDST, phenotypic drug susceptibility testing.^bPPV, positive predictive value.^cNPV, negative predictive value.^dResistant samples identified by pDST/Samples with pDST results.^eResistant samples identified by sequencing/Samples with sequencing results.^fCI, confidence interval.^gND, not determined.

moderate agreement (0.40–0.59) for ethionamide (ETH)/prothionamide (PTH). Concordance for AM was not evaluated due to the absence of genotypically resistant sputum samples.

Table 4 shows the distribution of variants predicted to cause drug resistance observed in the clinical samples analyzed. Variants *rpoB* S450L (S531L), *katG* S315T, *pncA* H57D, *embB* M306I, and *rpsL* K88R were the most frequently found for first-line drugs (RIF, INH, PZA, EMB, and SM, respectively). Two samples with variant *pncA* H57D (a marker for *M. bovis*) identified in this study were reported as *M. bovis* by microbiology tests.

Five clinical samples were assumed to be false-susceptible, as follows. Four samples were reported as sensitive to RIF by pDST; however, they harbored the disputed variants H445N ($n = 2$) and L452P ($n = 2$) in the *rpoB* gene. Also, one sample was reported as phenotypically sensitive to INH by pDST but it harbored the *katG* S315T mutation. These variants are widely considered true drug resistance markers (17).

Three of the four RIF false-susceptible samples were drug-resistant by Xpert MTB/RIF. The remaining sample only had pDST results available.

(iii) Phylogeny prediction. A maximum-likelihood phylogenetic tree was generated to classify the sputum samples according to their genotype. As expected, the different lineages were grouped (Fig. 1), showing high congruence with our protocol for the phylogenetic variants set.

The presence of 2 lineages of Mtb was observed in sputum samples. The Euro-American lineage was the highest with 20 samples (8 LAM, 6 Xtype, 5 Haarlem, and 1 Stype), followed by the Indo Oceanic with 3. Five samples grouped within the Euro-American lineage by the tree were considered *H37Rv* by our method due to the absence of phylogenetic variants. Variants consistent with *M. bovis* were found in 2 samples. In 1 sample, variants of *M. bovis* and the Stype lineage were observed.

Costs. Among the factors that need to be considered when developing new diagnostic tests are costs and feasibility. The cost for the test was 287 USD per sample sequenced on a MiSeq (11 samples per run) or 197 USD per sample sequenced on a NextSeq (84 samples per run). The price includes reagents for DNA extraction and quantification (Qubit high sensitivity dsDNA assay), library generation (TruSeq), quality control (Agilent 4200 TapeStation

TABLE 4 Distribution of variants associated with DR-TB observed in clinical samples

Gen	Variants (codon in <i>E. coli</i> ^a)	Frequency	Antibiotic
<i>rpoB</i>	S450L (S531L) ^b	6	Rifampicin
	Q432P (Q513P)	3	
	L452P (L533P)	3	
	H445N (H526N)	2	
	H445L (H526L)	2	
	H445D (H526D)	2	
	D435Y (D516Y)	2	
	D435V (D516V) ^b	1	
<i>inhA</i>	V170F (V146F)	1	Isoniazid/Ethionamide
	c-15t ^c	3	
	I21T	1	
<i>katG</i>	g-17t ^c	1	Isoniazid
	S315T	5	
<i>pncA</i>	W191R	1	Pyrazinamide
	H57D	2	
	H71R	2	
	V163G	1	
	Y41H	1	
<i>embB</i>	D136N	1	Ethambutol
	M306I	6	
	G406D	1	
<i>rpsL</i>	M306V	1	Streptomycin
	K88R	3	
<i>gyrA</i>	A90V	2	Fluoroquinolones (MFX, OFX, LFX)
	D94Y ^{d,e}	2	
	S91P ^{d,e}	2	
	D94G ^d	1	
	G88C	1	

^aFor *rpoB*, codon numbering in parentheses is based on *E. coli*.

^bVariants combined in a single sample resistant to RIF.

^cVariants combined in a single sample resistant to INH.

^dVariants combined in a single sample resistant to FQs.

^eVariants combined in a single sample resistant to FQs.

System), and sequencing (MiSeq reagent kit V2 300 cycles or NextSeq Mid Output V2 300 cycle kit).

DISCUSSION

The incomplete resistance profile of current routine molecular methods such as Xpert MTB/RIF and the Line Probe Assay (LiPA) (18) limits their value for acquiring relevant clinical information. Additionally, different members of the MTBC have been shown to have variations in the clinical phenotype of TB (19) due in part to differences in virulence (20) and the development of drug resistance (21).

Ideally, WGS of *Mtb* from clinical samples would allow deriving drug resistance and phylogenetic information, but it remains challenging due to the low bacillary load, which complicates the analysis and raises the cost, complicating its use in the clinical context (6, 18). This study proposes a rapid TS protocol, applied to sputum samples, capable of detecting resistance to all first- and some second-line drugs of *Mtb* and the phylogeny. The results presented show the feasibility of advancing its application as a tool of translational medicine, allowing more effective treatments for TB.

The experimental and bioinformatics pipelines initially tested with *Mtb* reference strains showed high accuracy (>95%) between the genotype predicted from this protocol and the pDST. Similarly, the reference strain variants reported by the WHO were entirely congruent with those obtained here. Even in clinical samples, the accuracy was >90% for all evaluated drugs, except for ETH/PTH (77.8%). The performance of our genotypic test was similar to other reports (22–25).

Only a few samples disagree between phenotypic and genotypic results. Two clinical samples RIF-susceptible harbored the L452P variant in the *rpoB* gene. This mutation confers

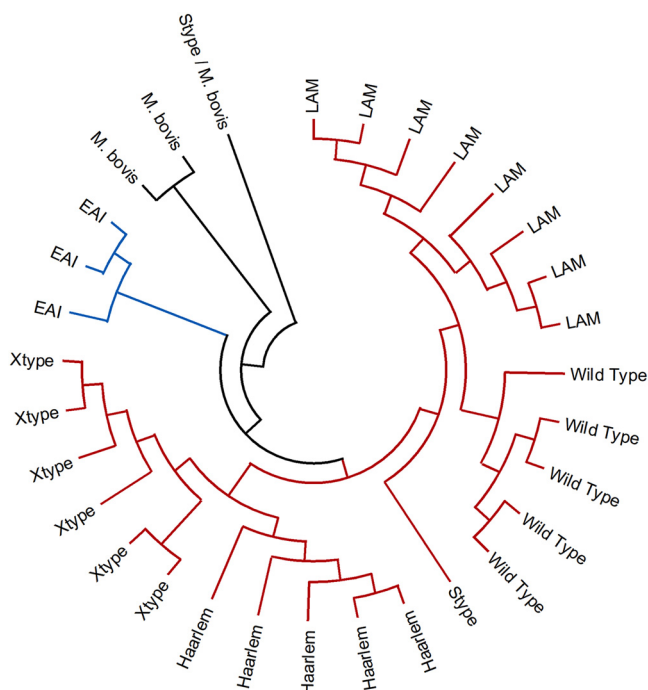


FIG 1 Maximum-likelihood phylogenetic tree of Mtb, obtained by targeted sequencing of clinical samples. Wild-type: Strain without phylogenetic variants (identical to *H37Rv*); Lineage 4 is colored in red; Lineage 3 is colored in blue; *M. bovis* samples are colored in black; One sample is probably a mixed infection (Stype/*M. bovis*).

a low-level of RIF-resistance (MIC 0.25–1.0 $\mu\text{g/mL}$) (26, 27) but remains susceptible by pDST. Reducing the critical concentration from 1.0 to 0.5 $\mu\text{g/mL}$ may improve the capacity of phenotypic tests to detect these clinically relevant strains (26).

On the other hand, previous studies have shown that the *rpoB* V170F mutation, located outside the resistance-determining region (RRDR), confers high-level resistance to RIF (MIC >16 $\mu\text{g/mL}$) (26–28). However, a clinical sample with this mutation was reported as susceptible by pDST. Additional evidence is required to confirm whether this variant is associated with resistance to RIF.

Based on a meta-analysis, Seifert et al. (29) concluded that the mutations in *katG* 315 and *inhA* -15 could explain 83% of resistance to INH. These mutations also were the most common variants in our study with similar sensitivity results (83%).

The evaluation of resistance to PZA, an important drug with sterilizing activity, is complex due to the frequent false positives results obtained by pDST (30, 31). In this study, the sensitivity and specificity to PZA resistance observed in clinical samples (100 and 92%, respectively) were similar to those obtained with WGS from cultures (22, 25).

FQs are part of Group A drugs recommended to treat multidrug-resistant tuberculosis due to their bactericidal effect and good tolerance. Therefore, the WHO (32) recommends, to increase success rates, ruling out FQs resistance before initiating the shorter MDR regimes. Mutations in the quinolone resistance-determining region of *gyrA* and *gyrB* are the main resistance mechanism to FQs (MFX, OFX/LFX). However, these mutations do not fully reflect the complexity of fluoroquinolone resistance (33, 34). Mutations in other genes, including efflux pumps, could explain the observation in this study of a phenotypically resistant sample without mutation in DNA gyrase.

Some resistant samples to SM present variants in the *rrs* and *rpsL* genes. Cuevas-Córdoba et al. (35) found mutations in these genes only in 48% of resistant samples to SM, which suggested the presence of additional resistance mechanisms (e.g., *gidB*). Similarly, in this study, mutations in *rrs* or *rpsL* genes were identified in only half of the phenotypically SM-resistant samples.

As reported by other authors with clinical samples and similar sequencing methods

(9, 36), the sensitivity was low for ETH/(PTH) (43%). Although only *inhA* gene mutations are currently considered resistance markers for these drugs, the TS panel was designed to capture additional genes related to ETH/PTH resistance (e.g., *ethA*). As with SM, the addition of under-characterized genes or variants to increase the sensitivity to ETH/PTH could lead to a reduction in specificity, potentially excluding useful drugs from the treatment of the disease (17, 37, 38). Therefore, before including poorly characterized mutations in the analysis, additional studies about their impact on drug resistance are required.

The *eis*, *rrs*, and *tlyA* genes were included in the TS panel to determine resistance to AM. However, sensitivity to AM was not evaluated because the sole phenotypically resistant sample was classified as genotypically sensitive (false negative). Analysis of additional resistance mechanisms is necessary to increase the accuracy, such as the Rv1258c efflux pump (included in this panel), which has AM as substrates (39).

Three discordant samples (phenotypically resistant but genotypically sensitive) for INH and SM showed moderate impact variants in *katG* (G428R) and *gidB* (L16R) genes. Intergenic mutation *inhA* g-17t was observed in one ETH false negative sample. Although these variants could explain the resistance observed in pDST, they are not part of the resistance variants database because they have not been fully characterized. Today, these variants would be considered of "uncertain significance" regarding their role in resistance determination.

Twenty-four variants identified in clinical samples were classified as high-impact (frameshifts, premature stop codons) and 167 as moderate impact variants (missense) by the SnpEff software. These variants were not part of the resistance variants database but could easily be incorporated into future reports when they are better described and validated. Fifteen of these mutations have previously been reported with the potential of generating drug resistance.

In this study, 32% of the clinical samples were MDR, a percentage considerably higher than the 7.94% reported by other authors in México (40). This may be due to the samples' origin in this study, which came from INER, a high specialty hospital, where severe or complicated (e.g., resistant) TB cases are referred.

The sensitivity and specificity for the MDR in clinical samples were 83.3% and 100%, respectively. Within the limits of this study, the sensitivity was only slightly lower than that reported by Madrazo-Moya et al. (23) with WGS from bacterial culture (89%) but with higher specificity (100% versus 97%).

In this study, only 48% (10/21) of the genotypically RIF resistant samples were also INH resistant. Bisimwa et al. (41) reported that only 61.6% of the samples RIF resistant by Xpert MTB/RIF were also INH resistant by MTBDRplus. According to our results, using resistance to RIF as an MDR marker could deprive patients of receiving INH in their treatment, a crucial antituberculosis drug.

Regarding the final terminal efficiency of the method, 31 sputum samples from 44 smear-positive (70%) were successfully processed by TS. Even 2 of the 3 smear-negative and culture-positive samples generated enough data to establish resistance prediction and phylogeny profile. These results are similar to those reported by Doyle et al. (7), who obtained resistance prediction in 74% of the sputum samples analyzed. Feuerriegel et al. (9) and Kayomo et al. (42), using a similar methodology (DeeplexMyc-TB assay), reported a resistance prediction for 78% of the clinical samples.

The presence of heteroresistant variants in *Mtb* isolates has been reported worldwide for several clinically important drugs (43). However, current molecular methods have limited utility in their detection; Xpert and Xpert Ultra can detect heteroresistance greater than 20% and GenoType greater than 5% (44). Nevertheless, poor detection of heteroresistance can lead to incorrect treatment, amplification of resistance (45), and treatment failure (10).

Detection of minority subpopulations of *Mtb* (1%) requires a depth of coverage of at least 2000x (10). When sequencing genetically diverse samples, such as sputum, contaminating reads are frequently present (belonging to other microorganisms genetically similar to *Mtb*), even using TS (46). Based on this consideration, deep sequencing (4000x) was used to ensure that enough reads of *Mtb* were obtained.

Due to the stringency used to minimize false positive results, approximately 50% of the sequences obtained were quality filtered. After that, contaminating sequences from

humans (~3%), viral (0.1%), non-MTB bacteria (~35%), and other microorganisms (~7%) present in the clinical samples were removed; about 55% of high-quality reads mapped to the Mtb reference genome. However, the sequences obtained were sufficient for an adequate phylogenetic and drug resistance determination and to identify heteroresistance in clinical samples.

TS from clinical samples allows heteroresistance identification, improves the phenotype-genotype relation (47), and avoids culture-induced loss of resistant subpopulations (48). In this work, 62% of the variants associated with resistance were fixed (present in $\geq 95\%$ of the reads [48]), and 38% showed macroheteroresistance (present in 5 to 94% [48]). This method could identify heteroresistance with similar sensitivity compared to current methods.

Interestingly, heteroresistance was observed in FQs-associated variants, even in patients who had not received second-line treatment. Most of these variants are in the fixation process (macroheteroresistance), and samples with 2 or more resistance variants to these drugs were even observed. These data agree with those reported by Zhang et al. (49), who mentioned that this is due to the monotherapeutic use of FQs to treat other infections in patients with undiagnosed TB.

Although WGS is currently the best method used in TB epidemiology (14), it has several drawbacks, as previously discussed. Thus, different studies have focused on SNPs as markers to classify Mtb isolates in their main lineages with a high resolution (15, 16, 50). The TS protocol reported here for sputum samples was able to get similar results to WGS from the bacterial culture at a lower cost.

In this study, a set of 68 SNPs allowed the identification of all MTBC species and Mtb lineages, resulting in a precise method compared with the referenced data. As expected, the Euro-American lineage had the highest presence in the clinical samples, followed by the Indo Oceanic lineage and *M. bovis*. One sample is probably a mixed infection (Stype/*M. bovis*). These results coincide with Amlerova et al. (14), who mentions that the LAM sublineage is prevalent in Latin America, with a significant incidence of Xtype and Haarlem in North America and, to a lesser extent, the Indo Oceanic. The prevalence of the Stype sublineage is very low in the world.

Any technological advance must also be evaluated regarding the feasibility of being introduced in the clinical field, considering cost factors and technical practicality for its implementation. In Mexico, the cost of pDST and an XpertMTB/RIF, if not subsidized by international agencies, reaches 250 USD per sample (23). Using our methodology, the cost per sample is 287 USD and could be reduced to 197 USD by sequencing on a NextSeq Mid Output. Considering the constant reduction in sequencing costs and savings derived from processing larger batches, TS of clinical samples could become a viable alternative to obtain effective and timely clinical information, especially in third-level reference hospitals.

Unlike conventional molecular methods, like Xpert MTB/RIF and LiPA, TS and WGS require relatively high computational and laboratory infrastructure (51). This inconvenience could be solved by considering an operation in 2 phases; (i) sample processing and DNA extraction are carried out in first and second-level hospitals, and (ii) the generation of libraries, sequencing, bioinformatic analysis, and reporting of results are carried out in central reference laboratories.

Limitations. This study focused on the available information on resistant variants validated microbiologically and genetically. For a better evaluation of the diagnostic method, it will be necessary for future studies to evaluate the performance of this protocol by increasing the number of samples (sensitive and resistant) and including the ability to evaluate discordant results by repeated culture and other molecular tests. Although the panel includes some genes related to resistance to other drugs of clinical importance (linezolid, bedaquiline, and clofazimine), they were not included in the analysis due to the absence of samples with pDST and poor characterization of the variants associated with their resistance. Future studies should add new genes and variants as they are characterized, mainly in drugs with low detection of genotypic resistance. The inclusion of genes such as Mpt64 (31) or hsp65 (52) in the targeted sequencing panel could simplify the identification of Mtb from other nontuberculous mycobacteria (NTM).

Although the spoligotyping results obtained by WGS provided excellent results, having phenotypic lineage assignment for the reference strains and clinical samples would provide double confirmation.

Finally, this proof-of-concept was carried out with a low number of samples; however, it was enough to move toward the next phase involving studies with a greater number of samples with more variability of resistance and phylogeny to validate the TS protocol.

Conclusions. Direct sputum sequencing is one attractive option, slightly slower than current molecular methods but with a much better time response than cultures. TS can generate a complete profile of drug resistance, phylogeny, and heteroresistance. Access to this type of molecular information would aid clinicians in selecting the adequate treatment for each person from the beginning of the disease, reducing the transmission of drug resistance between individuals at a manageable cost.

An added advantage of this approach is that, unlike other published methods (e.g., Deeplex Myc-TB), the oligonucleotides in the panel used for TS can be easily synthesized by various suppliers, which provides further opportunities for lowering costs. Furthermore, due to the rapid pace at which new knowledge is generated, with more genes and mutations validated to be associated with drug resistance, these can be readily added to the sequencing panel and the bioinformatics pipeline, increasing the capacity for resistance detection.

MATERIALS AND METHODS

Samples. (i) Reference strains. As a first step, well-characterized *Mtb* reference strains were used to test the protocol and identify drug resistance and phylogeny. Cultures from two *H37Rv* strains and 28 *Mtb* reference strains were donated by the Instituto de Diagnóstico y Referencia Epidemiológicos (InDRE) from the panel sent annually to Supranational Laboratories by the WHO. The WHO strains are provided with a report with genotypic resistance information; additionally, the InDRE performed pDST for first-line and second-line drugs.

(ii) Clinical samples (sputum). In a second stage, the protocol was applied to clinical samples from 46 adult patients collected from February 2014 to February 2016 at the Instituto Nacional de Enfermedades Respiratorias (INER). Forty four of those samples were determined to be TB caused by members of the MTBC. The 2 remaining samples were negative for smear microscopy and culture. Routine tests such as smear microscopy, culture, Xpert MTB/RIF, and pDST were all performed in the Microbiology Laboratory of the INER. An aliquot of the sputum samples ($\geq 500 \mu\text{L}$) was inactivated at 80°C for 20 min, transported to Instituto Nacional de Medicina Genómica (INMEGEN), and stored at -80°C until use for the genetic analysis.

DNA extraction. DNA extraction from cultures of reference strains and sputum samples was performed following the protocol by Warren et al. (53). The purity of the DNA was evaluated using a NanoDrop One (Thermo Fisher Scientific), the integrity was verified in a 1% agarose gel, and the concentration in a Qubit 3.0 fluorometer (Thermo Fisher Scientific).

Targeted panel design. The Illumina Custom Product Design Services was used to design a targeted sequencing panel with 73 genetic regions of *Mtb* (93 kbp), containing structural genes and regulatory regions, selected from the literature based on their potential association with DR-TB and phylogeny (Table S1). The coordinates of the regions of interest according to the *Mtb H37Rv* genome (NC_000962.3) were sent to Illumina to design and synthesize primers capable of amplifying the full-length of selected genes in a single multiplex PCR assay.

Libraries generation. (i) TS. TruSeq Custom Amplicon Low Input Kit (Illumina) and the targeted panel described above were used for library generation from reference strains and clinical samples, following the manufacturer's recommendations.

(ii) WGS. This was performed on reference strains only. WGS libraries were prepared using the Nextera XT system (Illumina), with 1.0 ng of input DNA, according to the manufacturer's recommendations. To obtain the *in silico* spoligotypes, the sequences obtained were used as input for the SpOTyping software (54); these results were used as the gold standard for lineage classification.

Sequencing. Sequencing was performed in a NextSeq 500/550, using a Mid Output Kit (2×150 bp) (Illumina). A depth of 4000x was estimated for targeted panel libraries (clinical samples and reference strains) to detect minor populations of *Mtb* in co-infections and possible heteroresistance (10). The depth of WGS libraries (reference strains only) was at least 200x.

The quality of all libraries was evaluated on an Agilent 4200 TapeStation System, and their concentration determined with the Qubit high sensitivity dsDNA assay (Qubit 3.0, Thermo Fisher) before sequencing.

Bioinformatic analysis. The bioinformatic analysis of the clinical samples and the strains was carried out following the methodology of Cuevas-Córdoba et al. (46). Briefly, reads were quality-filtered, and human and viral DNA were eliminated. Reads were mapped to the *Mtb* reference genome NC_000962.3. Specific conditions were applied to identify and discard reads from other microorganisms genetically similar to *Mtb* (Fig. S1). Reads regarded as true *Mtb* were used for the variant calling with the GATK software, and the variants were annotated with the SnpEff software.

Resistance variants and phylogeny databases. A resistance variants database of 585 mutations (Table S2) associated with phenotypic resistance was created using the variants reported by Miotto et al. (17), WHO (5) and the CRyPTIC Consortium, and the 100,000 Genomes Project (22).

Additionally, a phylogenetic variants database with 68 SNPs, based on the SNPs reported by Coll et al. (15),

Lipworth et al. (50), and Feuerriegel et al. (16), was generated. Within the variants present in the genes of the TS panel, those shared by more than 1 lineage were eliminated, until a minimum of 1 and a maximum of 3 SNPs were kept capable of classifying each of the different species and subspecies of the MTBC (Table S3).

Genotypic-phenotypic correlation of the resistance and phylogeny variants in Mtb. (i) Reference strains. Resistance variants of the reference strains were identified by TS. Genotypic resistance results were assigned to each sample using the resistance variants database and compared with those reported by the WHO and the pDST results (that were considered the gold standard).

WGS of reference strains was also used for *in silico* spoligotyping. Raw reads were used for spoligotyping with the SpoTyping program (54). Results were compared with the lineages predicted by TS, using the phylogenetic variants database generated in this study.

(ii) Clinical samples. Resistance variants of the clinical samples were identified by TS. Genotypic resistance was determined for each sample using the resistance variants database. The results were compared with the pDST, which was taken as the gold standard.

The phylogenetic variants were derived from the TS and the lineages were predicted using the phylogenetic variants database.

Statistical analysis. The genotypic resistance for each drug, for all samples, was compared with their corresponding pDST. Based on these results, statistical indicators for diagnostic tests such as sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy for each drug were calculated using the Epidat 3.2 software. The genotype-phenotype agreement was evaluated with Cohen's kappa test using SPSS statistical program.

Clinical samples that were phenotypically susceptible, but harbored mutations associated with drug resistance according to Miotto et al. (17), were assumed to be false-susceptible and considered resistant, disregarding the phenotypic result (5, 17).

Phylogenetic tree. The method proposed by Walker et al. (55) was utilized to classify the samples and identify them according to their genotype. Briefly, all the SNPs identified in each clinical sample by TS were concatenated to generate a signature sequence for each sample. Using the Mega X program, a multiple sequence alignment was performed with the signature sequences. A maximum-likelihood phylogenetic tree was inferred by applying the general reversible time model with 1000 bootstraps.

Ethics. All participants read and signed an informed consent to participate in this study. All patients' personal information was removed, and the samples were tagged with a code for identification. Both INER and INMEGEN bioethics committees approved the study.

Data availability. All relevant data that support the findings described in this study are fully available in GenBank, BioProject accession number [PRJNA826553](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA826553).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.4 MB.

ACKNOWLEDGMENTS

We thank Eugenia Silva-Herzog for her critical reading of the manuscript.

We thank the members of the Laboratorio de Diagnóstico Genómico from INMEGEN for technical support in the generation of libraries and the Unidad de Secuenciación from INMEGEN for sequencing technical support.

X.S. was funded by CONACyT (grant numbers GACD 264693-C1 and PDCPN_2015_1012) and by INMEGEN (internal funds).

REFERENCES

- World Health Organization. 2021. Global tuberculosis report. World Health Organization, Geneva, Switzerland. <https://www.who.int/publications/i/item/9789240037021>.
- O'Neill J. 2019. Tackling drug-resistant infections globally: final report and recommendations: the review on antimicrobial resistance; 2016.
- World Health Organization. 2019. Global tuberculosis report. World Health Organization, Geneva, Switzerland. <https://www.who.int/publications/i/item/9789241565714>.
- Horsburgh CR, Barry CE, Lange C. 2015. Treatment of tuberculosis. *N Engl J Med* 373:2149–2160. <https://doi.org/10.1056/NEJMra1413919>.
- 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, Geneva, Switzerland. <https://apps.who.int/iris/handle/10665/274443>.
- Brown AC, Bryant JM, Einer-Jensen K, Holdstock J, Houniet DT, Chan JZM, Depledge DP, Nikolayevskiy V, Broda A, Stone MJ, Christiansen MT, Williams R, McAndrew MB, Tutill H, Brown J, Melzer M, Rosmarin C, McHugh TD, Shorten RJ, Drobniewski F, Speight G, Breuer J. 2015. Rapid whole-genome sequencing of *Mycobacterium tuberculosis* isolates directly from clinical samples. *J Clin Microbiol* 53:2230–2237. <https://doi.org/10.1128/JCM.00486-15>.
- Doyle RM, Burgess C, Williams R, Gorton R, Booth H, Brown J, Bryant JM, Chan J, Creer D, Holdstock J, Kunst H, Lozewicz S, Platt G, Romero EY, Speight G, Tiberi S, Abubakar I, Lipman M, McHugh TD, Breuer J. 2018. Direct whole-genome sequencing of sputum accurately identifies drug-resistant mycobacterium tuberculosis faster than MGIT culture sequencing. *J Clin Microbiol* 56:e00666-18. <https://doi.org/10.1128/JCM.00666-18>.
- Colman RE, Anderson J, Lemmer D, Lehmkuhl E, Georghiou SB, Heaton H, Wiggins K, Gillece JD, Schupp JM, Catanzaro DG, Crudu V, Cohen T, Rodwell TC, Engelthaler DM. 2016. Rapid drug susceptibility testing of drug-resistant mycobacterium tuberculosis isolates directly from clinical samples by use of amplicon sequencing: A proof-of-concept study. *J Clin Microbiol* 54:2058–2067. <https://doi.org/10.1128/JCM.00535-16>.
- Feuerriegel S, Kohl TA, Utpatel C, Andres S, Maurer FP, Heyckendorf J, Jouet A, Badalato N, Foray L, Kamara RF, Conteh OS, Supply P, Niemann S. 2021. Rapid genomic first-and second-line drug resistance prediction from clinical *Mycobacterium tuberculosis* specimens using Deeplex-MycTB. *Eur Respir J* 57: 2001796. <https://doi.org/10.1183/13993003.01796-2020>.

10. Colman RE, Schupp JM, Hicks ND, Smith DE, Buchhagen JL, Valafar F, Crudu V, Romancenco E, Noroc E, Jackson L, Catanzaro DG, Rodwell TC, Catanzaro A, Keim P, Engelthaler DM. 2015. Detection of low-level mixed-population drug resistance in *Mycobacterium tuberculosis* using high fidelity amplicon sequencing. *PLoS One* 10:e0126626. <https://doi.org/10.1371/journal.pone.0126626>.
11. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, Musser JM. 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci U S A* 94:9869–9874. <https://doi.org/10.1073/pnas.94.18.9869>.
12. Brosch R, Gordon S v, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, Garnier T, Gutierrez C, Hewinson G, Kremer K, Parsons LM, Pym AS, Samper S, van Soolingen D, Cole ST. 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci U S A* 99:3684–3689. <https://doi.org/10.1073/pnas.052548299>.
13. Niemann S, Supply P. 2014. Diversity and evolution of *Mycobacterium tuberculosis*: moving to whole-genome-based approaches. *Cold Spring Harb Perspect Med* 4:a021188. <https://doi.org/10.1101/cshperspect.a021188>.
14. Amlerova J, Hrabak J. 2018. Genotyping of *Mycobacterium tuberculosis* using whole genome sequencing. *Folia Microbiol* 63:537–545. <https://doi.org/10.1007/s12223-018-0599-y>.
15. Coll F, Mc Nerney R, Guerra-Assunção JA, Glynn JR, Perdigão J, Viveiros M, Portugal I, Pain A, Martin N, Clark TG. 2014. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat Commun* 5:1–5. <https://doi.org/10.1038/ncomms5812>.
16. Feuerriegel S, Köser CU, Niemann S. 2014. Phylogenetic polymorphisms in antibiotic resistance genes of the *Mycobacterium tuberculosis* complex. *J Antimicrob Chemother* 69:1205–1210. <https://doi.org/10.1093/jac/dkt535>.
17. Miotto P, Tessema B, Tagliani E, Chindelevitch L, Starks AM, Emerson C, Hanna D, Kim PS, Liwski R, Zignol M, Gilpin C, Niemann S, Denkinge CM, Fleming J, Warren RM, Crook D, Posey J, Gagneux S, Hoffner S, Rodrigues C, Comas I, Engelthaler DM, Murray M, Alland D, Rigouts L, Lange C, Dheda K, Hasan R, Ranganathan UDK, Mc Nerney R, Ezewudo M, Cirillo DM, Schito M, Köser CU, Rodwell TC. 2017. A standardised method for interpreting the association between mutations and phenotypic drug resistance in *Mycobacterium tuberculosis*. *Eur Respir J* 50:1701354. <https://doi.org/10.1183/13993003.01354-2017>.
18. Votintseva AA, Bradley P, Pankhurst L, del Ojo Elias C, Loose M, Nilgiriwala K, Chatterjee A, Smith EG, Sanderson N, Walker TM, Morgan MR, Wyllie DH, Walker AS, Peto TEA, Crook DW, Iqbal Z. 2017. Same-day diagnostic and surveillance data for tuberculosis via whole-genome sequencing of direct respiratory samples. *J Clin Microbiol* 55:1285–1298. <https://doi.org/10.1128/JCM.02483-16>.
19. Caws M, Thwaites G, Dunstan S, Hawn TR, Lan NTN, Thuong NTT, Stepniewska K, Huyen MNT, Nguyen DB, Tran HL, Gagneux S, van Soolingen D, Kremer K, van der Sande M, Small P, Anh PTH, Nguyen TC, Hoang TQ, Duyen NTH, Dau QT, Hieu NT, Torok E, Tran TH, Nguyen HD, Nhu NTQ, Phan MD, Chau NVV, Farrar J. 2008. The influence of host and bacterial genotype on the development of disseminated disease with *Mycobacterium tuberculosis*. *PLoS Pathog* 4:e1000034. <https://doi.org/10.1371/journal.ppat.1000034>.
20. Reiling N, Homolka S, Walter K, Brandenburg J, Niwinski L, Ernst M, Herzmann C, Lange C, Diel R, Ehlers S, Niemann S. 2013. Clade-specific virulence patterns of *Mycobacterium tuberculosis* complex strains in human primary macrophages and aerogenically infected mice. *mBio* 4:e00250-13. <https://doi.org/10.1128/mBio.00250-13>.
21. Ford CB, Shah RR, Maeda MK, Gagneux S, Murray MB, Cohen T, Johnston JC, Gardy J, Lipsitch M, Fortune SM. 2013. *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nat Genet* 45:784–790. <https://doi.org/10.1038/ng.2656>.
22. CRYPTIC Consortium and the 100,000 Genomes Project. 2018. Prediction of susceptibility to first-line tuberculosis drugs by DNA sequencing. *N Engl J Med* 379:1403–1415. <https://doi.org/10.1056/NEJMoa1800474>.
23. Madrazo-Moya CF, Cancino-Muñoz I, Cuevas-Córdoba B, González-Covarrubias V, Barbosa-Amezcuca M, Soberón X, Muñoz-Salazar R, Martínez-Guarneros A, Bäcker C, Zarrabal-Meza J, Sampieri-Ramírez C, Enciso-Moreno A, Lauzardo M, Comas I, Zenteno-Cuevas R. 2019. Whole genomic sequencing as a tool for diagnosis of drug and multidrug-resistance tuberculosis in an endemic region in Mexico. *PLoS One* 14:e0213046. <https://doi.org/10.1371/journal.pone.0213046>.
24. Walker TM, Kohl TA, Omar S v, 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, Davies J, Crichton C, Acharya M, Madrid-Marquez L, Eyre D, Wyllie D, Golubchik T, Munang M. 2015. Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study. *Lancet Infectious Diseases* 15:1193–1202. [https://doi.org/10.1016/S1473-3099\(15\)00062-6](https://doi.org/10.1016/S1473-3099(15)00062-6).
25. Shea J, Halse TA, Lapiere P, Shudt M, Kohlerschmidt D, van Roey P, Limberger R, Taylor J, Escuyer V, Musser KA. 2017. Comprehensive whole-genome sequencing and reporting of drug resistance profiles on clinical cases of *Mycobacterium tuberculosis* in New York State. *J Clin Microbiol* 55:1871–1882. <https://doi.org/10.1128/JCM.00298-17>.
26. Shea J, Halse TA, Kohlerschmidt D, Lapiere P, Modestil AHA, Kearns CH, Dworkin FF, Rakeman JL, Escuyer V, Musser KA. 2021. Low-level rifampin resistance and rpoB mutations in *Mycobacterium tuberculosis*: an analysis of whole-genome sequencing and drug susceptibility test data in New York. *J Clin Microbiol* 59:e01885-20. <https://doi.org/10.1128/JCM.01885-20>.
27. Liu D, Huang F, Zhang G, He W, Ou X, He P, Zhao B, Zhu B, Liu F, Li Z, Liu C, Xia H, Wang S, Zhou Y, Walker TM, Liu L, Crook DW, Zhao Y. 2022. Whole-genome sequencing for surveillance of tuberculosis drug resistance and determination of resistance level in China. *Clin Microbiol Infect* 28:731.e9–731.e15. <https://doi.org/10.1016/j.cmi.2021.09.014>.
28. Ma P, Luo T, Ge L, Chen Z, Wang X, Zhao R, Liao W, Bao L. 2021. Compensatory effects of *M. tuberculosis* rpoB mutations outside the rifampicin resistance-determining region. *Emerging Microbes and Infections* 10:743–752. <https://doi.org/10.1080/22221751.2021.1908096>.
29. Seifert M, Catanzaro D, Catanzaro A, Rodwell TC. 2015. Genetic mutations associated with isoniazid resistance in *Mycobacterium tuberculosis*: A systematic review. *PLoS One* 10:e0119628. <https://doi.org/10.1371/journal.pone.0119628>.
30. Zhang Y, Mitchison D. 2003. The curious characteristics of pyrazinamide: a review. *Int J Tuberculosis and Lung Dis* 7:6–21.
31. Soundararajan L, Kambli P, Priyadarshini S, Let B, Murugan S, Iravatham C, Tornheim JA, Rodrigues C, Gupta R, Ramprasad VL. 2020. Whole genome enrichment approach for rapid detection of *Mycobacterium tuberculosis* and drug resistance-associated mutations from direct sputum sequencing. *Tuberculosis (Edinb)* 121:101915. <https://doi.org/10.1016/j.tube.2020.101915>.
32. World Health Organization. 2019. WHO consolidated guidelines on drug-resistant tuberculosis treatment. World Health Organization, Geneva, Switzerland. <https://apps.who.int/iris/handle/10665/311389>.
33. Maruri F, Sterling TR, Kaiga AW, Blackman A, van der heijden YF, Mayer C, Cambau E, Aubry A. 2012. A systematic review of gyrase mutations associated with fluoroquinolone-resistant *Mycobacterium tuberculosis* and a proposed gyrase numbering system. *J Antimicrob Chemother* 67:819–831. <https://doi.org/10.1093/jac/dkr566>.
34. Chaiyachai P, Chaiprasert A, Nonghanphithak D, Smithikarn S, Kamolwat P, Punggrassami P, Reechaipichitkul W, Ong RTH, Teo YY, Faksri K. 2021. Whole-genome analysis of drug-resistant *Mycobacterium tuberculosis* reveals novel mutations associated with fluoroquinolone resistance. *Int J Antimicrobial Agents* 58:106385. <https://doi.org/10.1016/j.ijantimicag.2021.106385>.
35. Cuevas-Córdoba B, Cuellar-Sánchez A, Passisi-Crivelli A, Santana-álvarez CA, Hernández-Illezcas J, Zenteno-Cuevas R. 2013. Rrs and rpsL mutations in streptomycin-resistant isolates of *Mycobacterium tuberculosis* from Mexico. *J Microbiology, Immunology and Infection* 46:30–34. <https://doi.org/10.1016/j.jmii.2012.08.020>.
36. Jouet A, Gaudin C, Badalato N, Allix-Béguec C, Duthoy S, Ferré A, Diels M, Laurent Y, Contreras S, Feuerriegel S, Niemann S, André E, Kaswa MK, Tagliani E, Cabibbe A, Mathys V, Cirillo D, de Jong BC, Rigouts L, Supply P. 2021. Deep amplicon sequencing for culture-free prediction of susceptibility or resistance to 13 anti-tuberculous drugs. *Eur Respir J* 57:2002338. <https://doi.org/10.1183/13993003.02338-2020>.
37. Chen X, He G, Wang S, Lin S, Chen J, Zhang W. 2019. Evaluation of whole-genome sequence method to diagnose resistance of 13 anti-tuberculosis drugs and characterize resistance genes in clinical multi-drug resistance *Mycobacterium tuberculosis* isolates from China. *Front Microbiol* 10. <https://doi.org/10.3389/fmicb.2019.01741>.
38. Coll F, Mc Nerney R, Preston MD, Guerra-Assunção JA, Warry A, Hill-Cawthorne G, Mallard K, Nair M, Miranda A, Alves A, Perdigão J, Viveiros M, Portugal I, Hasan Z, Hasan R, Glynn JR, Martin N, Pain A, Clark TG. 2015. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome Med* 7. <https://doi.org/10.1186/s13073-015-0164-0>.
39. Balganes M, Dinesh N, Sharma S, Kuruppath S, Nair A v, Sharma U. 2012. Efflux pumps of *Mycobacterium tuberculosis* play a significant role in anti-tuberculosis activity of potential drug candidates. *Antimicrob Agents Chemother* 56:2643–2651. <https://doi.org/10.1128/AAC.06003-11>.
40. Lopez-Avalos G, Gonzalez-Palomar G, Lopez-Rodriguez M, Vazquez-Chacon CA, Mora-Aguilera G, Gonzalez-Barrios JA, Villanueva-Arias JC, Sandoval-Diaz

- M, Miranda-Hernández U, Alvarez-Maya I. 2017. Genetic diversity of *Mycobacterium tuberculosis* and transmission associated with first-line drug resistance: a first analysis in Jalisco, Mexico. *J Glob Antimicrob Resist* 11:90–97. <https://doi.org/10.1016/j.jgar.2017.07.004>.
41. Bisimwa BC, Nachega JB, Warren RM, Theron G, Metcalfe JZ, Shah M, Diacon AH, Sam-Agudu NA, Yotebieng M, Bulabula ANH, Katoto PDMC, Chirambiza JP, Nyota R, Birembano FM, Musafiri EM, Byadunia S, Bahizire E, Kaswa MK, Callens S, Kashongwe ZM. 2021. Xpert *Mycobacterium tuberculosis*/rifampicin-detected rifampicin resistance is a suboptimal surrogate for multidrug-resistant tuberculosis in Eastern Democratic Republic of the Congo: diagnostic and clinical implications. *Clinical Infectious Diseases* 73:e362–e370. <https://doi.org/10.1093/cid/ciaa873>.
 42. Kayomo MK, Mbula VN, Aloni M, André E, Rigouts L, Boutachkourt F, de Jong BC, Nkiere NM, Dean AS. 2020. Targeted next-generation sequencing of sputum for diagnosis of drug-resistant TB: results of a national survey in Democratic Republic of the Congo. *Sci Rep* 10:1–8. <https://doi.org/10.1038/s41598-020-67479-4>.
 43. Ye M, Yuan W, Molaeipour L, Azizian K, Ahmadi A, Kouhsari E. 2021. Antibiotic heteroresistance in *Mycobacterium tuberculosis* isolates: a systematic review and meta-analysis. *Ann Clin Microbiol Antimicrob* 20:1–9. <https://doi.org/10.1186/s12941-021-00478-z>.
 44. Ng KCS, Supply P, Cobelens FGJ, Gaudin C, Gonzalez-Martin J, de Jong BC, Rigouts L. 2019. How Well Do Routine Molecular Diagnostics Detect Rifampin Heteroresistance in *Mycobacterium tuberculosis*? *J Clin Microbiol* 57:e00717–19. <https://doi.org/10.1128/JCM.00717-19>.
 45. Black PA, de Vos M, Louw GE, van der Merwe RG, Dippenaar A, Streicher EM, Abdallah AM, Sampson SL, Victor TC, Dolby T, Simpson JA, van Helden PD, Warren RM, Pain A. 2015. Whole genome sequencing reveals genomic heterogeneity and antibiotic purification in *Mycobacterium tuberculosis* isolates. *BMC Genomics* 16:1–14. <https://doi.org/10.1186/s12864-015-2067-2>.
 46. Cuevas-Córdoba B, Fresno C, Haase-Hernández JI, Barbosa-Amezcuca M, Mata-Rocha M, Muñoz-Torrico M, Salazar-Lezama MA, Martínez-Orozco JA, Narváez-Díaz LA, Salas-Hernández J, González-Covarrubias V, Soberón X. 2021. A bioinformatics pipeline for *Mycobacterium tuberculosis* sequencing that cleans contaminant reads from sputum samples. *PLoS One* 16:e0258774. <https://doi.org/10.1371/journal.pone.0258774>.
 47. Nimmo C, Shaw LP, Doyle R, Williams R, Brien K, Burgess C, Breuer J, Balloux F, Pym AS. 2019. Whole genome sequencing *Mycobacterium tuberculosis* directly from sputum identifies more genetic diversity than sequencing from culture. *BMC Genomics* 20:1–9. <https://doi.org/10.1186/s12864-019-5782-2>.
 48. Metcalfe JZ, Streicher E, Theron G, Colman RE, Penaloza R, Allender C, Lemmer D, Warren RM, Engelthaler DM. 2017. *Mycobacterium tuberculosis* subculture results in loss of potentially clinically relevant heteroresistance. *Antimicrob Agents Chemother* 61:e00888–17. <https://doi.org/10.1128/AAC.00888-17>.
 49. Zhang D, Gomez JE, Chien JY, Haseley N, Desjardins CA, Earl AM, Hsueh PR, Hung DT. 2016. Genomic analysis of the evolution of fluoroquinolone resistance in *Mycobacterium tuberculosis* prior to tuberculosis diagnosis. *Antimicrob Agents Chemother* 60:6600–6608. <https://doi.org/10.1128/AAC.00664-16>.
 50. Lipworth S, Sajou R, de Neeling A, Bradley P, van der Hoek W, Maphalala G, Bonnet M, Sanchez-Padilla E, Diel R, Niemann S, Iqbal Z, Smith G, Peto T, Crook D, Walker T, van Soolingen D. 2019. SNP-IT tool for identifying subspecies and associated lineages of *Mycobacterium tuberculosis* complex. *Emerg Infect Dis* 25:482–488. <https://doi.org/10.3201/eid2503.180894>.
 51. Dheda K, Gumbo T, Maartens G, Dooley KE, McNerney R, Murray M, Furin J, Nardell EA, London L, Lessem E, Theron G, van Helden P, Niemann S, Merker M, Dowdy D, van Rie A, Siu GKH, Pasipanodya JG, Rodrigues C, Clark TG, Sirgel FA, Esmail A, Lin HH, Atre SR, Schaaf HS, Chang KC, Lange C, Nahid P, Udhwadia ZF, Horsburgh CR, Churchyard GJ, Menzies D, Hesselting AC, Nuernberger E, McIlleron H, Fennelly KP, Goemaere E, Jaramillo E, Low M, Jara CM, Padayatchi N, Warren RM. 2017. The epidemiology, pathogenesis, transmission, diagnosis, and management of multidrug-resistant, extensively drug-resistant, and incurable tuberculosis. *Lancet Respiratory Medicine* 5:291–360. [https://doi.org/10.1016/S2213-2600\(17\)30079-6](https://doi.org/10.1016/S2213-2600(17)30079-6).
 52. Dai J, Chen Y, Lauzardo M. 2011. Web-accessible database of hsp65 sequences from *Mycobacterium* reference strains. *J Clin Microbiol* 49:2296–2303. <https://doi.org/10.1128/JCM.02602-10>.
 53. Warren R, de Kock M, Engelke E, Myburgh R, van Pittius NG, Victor T, van Helden P. 2006. Safe *Mycobacterium tuberculosis* DNA extraction method that does not compromise integrity. *J Clin Microbiol* 44:254–256. <https://doi.org/10.1128/JCM.44.1.254-256.2006>.
 54. Xia E, Teo YY, Ong RTH. 2016. SpoTyping: fast and accurate in silico *Mycobacterium* spoligotyping from sequence reads. *Genome Med* 8:1–9. <https://doi.org/10.1186/s13073-016-0270-7>.
 55. Walker TM, Ip CLC, Harrell RH, Evans JT, Kapatai G, Dedicato MJ, Eyre DW, Wilson DJ, Hawkey PM, Crook DW, Parkhill J, Harris D, Walker AS, Bowden R, Monk P, Smith EG, Peto TEA. 2013. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infectious Diseases* 13:137–146. [https://doi.org/10.1016/S1473-3099\(12\)70277-3](https://doi.org/10.1016/S1473-3099(12)70277-3).