AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Clinical Microbiology®

MYCOBACTERIOLOGY AND AEROBIC ACTINOMYCETES

Automatic Identification of Individual *rpoB* Gene Mutations Responsible for Rifampin Resistance in *Mycobacterium tuberculosis* by Use of Melting Temperature Signatures Generated by the Xpert MTB/RIF Ultra Assay

Yuan Cao,[a] Heta Parmar,[a] Ann Marie Simmons,[b] Devika Kale,[b] Kristy Tong,[b] Deanna Lieu,[b] David Persing,[b] Robert Kwiatkowski,[b] David Alland,[a] Soumitesh Chakravorty[a,b]

[a]Department of Medicine, Rutgers New Jersey Medical School, Newark, New Jersey, USA
[b]Cepheid Inc., Sunnyvale, California, USA

ABSTRACT Molecular surveillance of rifampin-resistant *Mycobacterium tuberculosis* can help to monitor the transmission of the disease. The Xpert MTB/RIF Ultra assay detects mutations in the rifampin resistance-determining region (RRDR) of the *rpoB* gene by the use of melting temperature (T_m) information from 4 *rpoB* probes which can fall in one of the 9 different assay-specified T_m windows. The large amount of T_m data generated by the assay offers the possibility of an RRDR genotyping approach more accessible than whole-genome sequencing. In this study, we developed an automated algorithm to specifically identify a wide range of mutations in the *rpoB* RRDR by utilizing the pattern of the T_m of the 4 probes within the 9 windows generated by the Ultra assay. The algorithm builds a RRDR mutation-specific "T_m signature" reference library from a set of known mutations and then identifies the RRDR genotype of an unknown sample by measuring the T_m distances between the test sample and the reference T_m values. Validated using a set of clinical isolates, the algorithm correctly identified RRDR genotypes of 93% samples with a wide range of *rpoB* single and double mutations. Our analytical approach showed a great potential for fast RRDR mutation identification and may also be used as a stand-alone method for ruling out relapse or transmission between patients. The algorithm can be further modified and optimized for higher accuracy as more Ultra data become available.

KEYWORDS Xpert MTB/RIF Ultra, *rpoB*, tuberculosis

Molecular surveillance of rifampin-resistant (RR) tuberculosis (TB) can aid in the detection of disease transmission and help to monitor the effectiveness of public health interventions (1). Although whole-genome sequencing (WGS) has been increasingly used as a surveillance tool (2), the cost and complexity of this approach limit its applicability, especially in the low- and middle-income countries where TB is most prevalent. Fortunately, approximately 95% of rifampin-resistant tuberculosis (RR-TB) cases are caused by a limited number of mutations in a short (81-bp) "core region" of the *Mycobacterium tuberculosis rpoB* gene (3, 4), also known as the rifampin resistance-determining region (RRDR). This suggests that RR-TB monitoring may be usefully performed using a more focused genotyping approach that limits analysis to the RRDR. As an alternative to WGS, this method would be especially useful if RRDR genotyping could be automatically incorporated into standard diagnostic procedures.

The increasing use of the Xpert MTB/RIF Ultra assay (CE *in vitro* diagnostic [IVD] medical device not available in the United States) to detect *M. tuberculosis* and RR may provide an automated means to perform *M. tuberculosis* RRDR genotyping. The Ultra

Citation Cao Y, Parmar H, Simmons AM, Kale D, Tong K, Lieu D, Persing D, Kwiatkowski R, Alland D, Chakravorty S. 2020. Automatic identification of individual *rpoB* gene mutations responsible for rifampin resistance in *Mycobacterium tuberculosis* by use of melting temperature signatures generated by the Xpert MTB/RIF Ultra assay. J Clin Microbiol 58:e00907-19. https://doi.org/10.1128/JCM.00907-19.

Editor Geoffrey A. Land, Carter BloodCare & Baylor University Medical Center

Address correspondence to Soumitesh Chakravorty, soumitesh.chakravorty@cepheid.com.

For a commentary on this article, see 10.1128/JCM.01504-19.

Received 14 June 2019
Returned for modification 18 July 2019
Accepted 12 September 2019

Accepted manuscript posted online 18 September 2019
Published 23 December 2019

assay uses 4 overlapping sloppy molecular beacons (SMBs) to probe the *M. tuberculosis rpoB* RRDR for mutations associated with RR-TB. Rifampin resistance due to a mutation in the RRDR is detected by a shift in the melting temperature (T_m) of at least one of these probes away from the T_m expected to occur in the presence of wild-type (WT), rifampin-susceptible (RS) RRDR sequences (5). Several small-scale studies of clinical *M. tuberculosis* isolates have suggested that the T_m of individual Ultra assay *rpoB* SMBs can indeed be used to identify several specific *rpoB* mutations and to "bin" other *rpoB* mutations into T_m-defined categories (5, 6). However, the complex effects of 4 different SMBs hybridizing to overlapping regions of the *rpoB* RRDR amplicon suggested that a single RRDR mutation might destabilize multiple Ultra SMBs. This raised the possibility that individual mutations could produce complex patterns of T_m shifts among the Ultra SMBs, which would enable more-specific differentiation among RRDR mutations and provide a more useful surveillance tool. Further, since different RRDR mutations may be associated with different rifampin MICs (7), a more reliable tool for identifying specific mutations might aid in customizing rifamycin-containing therapy.

We have developed an analytic method that uses the T_m values generated by the Ultra assay to specifically identify most *rpoB* RRDR mutations. This method combines the Ultra assay T_m values into a "T_m signature" so that the T_m signature of an unknown sample can be compared to the T_m signatures in a reference library of known *rpoB* RRDR mutants and, thus, the mutation present in the unknown sample be identified. The rare mutations that cannot be specifically identified can be grouped into one of a small number of "mutation groups" that are usually present in the same *rpoB* RRDR codon. Here, we describe this analytic method and test the overall approach of using Ultra T_m values to identify different *RRDR* mutations in a wide panel of RR *M. tuberculosis* isolates.

MATERIALS AND METHODS

Defining the T_m distance (TD) value. The 4 SMB probes in the Ultra assay can produce many different patterns of T_m values. At its simplest, an RS *M. tuberculosis* strain produces one T_m value for each probe. For each Ultra test performed on RS *M. tuberculosis*, the T_m associated with each probe clusters around the mean T_m value produced by that probe in multiple reference tests with *rpoB*-wild-type *M. tuberculosis*; thus, any probe that produces a T_m which falls into a larger, assay-defined, probe-specific "wild-type T_m window" is considered to have hybridized to a wild-type target sequence. A sample is defined as RS *M. tuberculosis* when all the 4 *rpoB* probes produce T_m values that lie within the 4 wild-type T_m windows. The Ultra assay also defines 5 "mutant T_m" windows. The assay identifies a sample as containing RR *M. tuberculosis* when at least one of the 4 assay probes produces a T_m value that falls within at least one of the mutant T_m windows. Samples that contain mixtures of RS and RR *M. tuberculosis* or that contain *rpoB* core sequences with multiple mutations produce more-complex patterns where one or more probes can produce T_m values in both wild-type and mutant windows at the same time. Mixtures and double mutants can also produce multiple different T_m values within the same T_m window.

We had previously defined the relatedness of a reference to test T_m signatures as being represented a "distance index" or "D" value (8), which we have renamed here the TD value (T_m distance value) to avoid confusion with Cohen's *d*, which measures the effect size of the difference between two means. The TD value for a pair of samples was defined as the Euclidean distance between the samples in Euclidean space. Because the Ultra assay classified T_m values into 9 different T_m windows, we could localize each T_m signature produced by the assay in a 9-dimensional Euclidean space. T_m windows that did not contain a T_m value (for example, a wild-type T_m window for a probe that produced a T_m value in its mutant T_m window due to the presence of a mutation) were classified as having a T_m value of zero. A library of reference T_m values was created as described in Results (Table 1). To calculate the TD value for an unknown test sample, the distance between the 9 data points in the reference T_m signature and the unknown test T_m signature was derived using the following formula:

$$\text{TD value} = \sqrt{\sum_{i=1}^{9} \Delta Tm_i^2}$$

In this context, the 9 ΔT_m values correspond to the differences of *rpo1*, *rpo2*, *rpo3*, *rpo4*, *rpo1* Mut, *rpo2* Mut, *rpo3* Mut, *rpo4* MutA or *rpo4* MutB T_m values of a reference (known) T_m signature with the corresponding T_m values of an unknown test T_m signature. A series of TD values are generated by calculating the TD value of the sample against each reference in the library, and the smallest TD value, i.e., the TD-1 value corresponding to the difference between a specific reference T_m signature and a test T_m signature, identifies the most likely mutation present in the test sample. The next closest match is identified by the TD-2 value. The calculation of TD values and matching TD-1 and TD-2 values was done using an R script which is available upon request.

TABLE 1 Reference T_m values with standard deviations and T_m windows as defined by the Ultra assay[a]

rpoB genotype	Mean or SD	Value (°C) for T_m windows as defined by the Ultra assay for:								
		rpo1	*rpo1* Mut	*rpo2*	*rpo2* Mut	*rpo3*	*rpo3* Mut	*rpo4*	*rpo4* MutA	*rpo4* MutB
Genotypes with one set of reference T_m values										
WT	Mean	69.57	0	73.29	0	75.96	0	67.27	0	0
	SD	0.16	0	0.16	0	0.22	0	0.15	0	0
511CCG	Mean	0	63.52	73.22	0	75.93	0	67.28	0	0
	SD	0	0.05	0.05	0	0.05	0	0.05	0	0
513CAG[b]	Mean	68.17	0	75.1	0	75.93	0	67.47	0	0
	SD	0.06	0	0.17	0	0.21	0	0.15	0	0
516GTC	Mean	70.06	0	0	69.72	75.81	0	67.42	0	0
	SD	0.12	0	0	0.1	0.11	0	0.18	0	0
516TAC	Mean	69.71	0	0	68.91	75.7	0	67.39	0	0
	SD	0.12	0	0	0.17	0.14	0	0.13	0	0
522CAG	Mean	70.12	0	0	68.83	0	73.95	67.12	0	0
	SD	0.1	0	0	0.1	0	0.06	0.05	0	0
522TTG	Mean	69.6	0	0	70.2	0	73.66	67.26	0	0
	SD	0.07	0	0	0.07	0	0.09	0.05	0	0
526CCC	Mean	69.47	0	73.45	0	0	71.7	66.08	0	0
	SD	0.05	0	0.06	0	0	0	0.1	0	0
526TAA	Mean	69.27	0	73.17	0	0	71.7	67.57	0	0
	SD	0.06	0	0.12	0	0	0.1	0.06	0	0
526TAC	Mean	69.44	0	73.49	0	0	72.83	67.29	0	0
	SD	0.1	0	0.13	0	0	0.13	0.16	0	0
526TCC	Mean	69.37	0	73.37	0	0	70.63	67.3	0	0
	SD	0.06	0	0.15	0	0	0.12	0.1	0	0
526TGC	Mean	69.4	0	73.43	0	0	71.7	68.43	0	0
	SD	0.1	0	0.12	0	0	0.1	0.06	0	0
531CAG	Mean	69.27	0	73.33	0	0	71.23	67.1	0	0
	SD	0.06	0	0.06	0	0	0.06	0	0	0
531TGG	Mean	69.62	0	73.78	0	0	73.54	0	71.06	0
	SD	0.08	0	0.08	0	0	0.09	0	0.05	0
531TTG	Mean	69.5	0	74.27	0	0	73.44	0	73.89	0
	SD	0.12	0	0.16	0	0	0.19	0	0.13	0
533CCG	Mean	69.64	0	73.3	0	75.98	0	0	0	61.68
	SD	0.13	0	0.12	0	0.1	0	0	0	0.13
529AAA[c]	Mean	69.53	0	72.98	0	0	0	0	0	64.08
	SD	0.05	0	0.08	0	0	0	0	0	0.08
518AAC del	Mean	69.53	0	0	65.73	75.5	0	66.93	0	0
	SD	0.21	0	0	0.38	0.26	0	0.06	0	0
510CTG_516TAC	Mean	0	65.73	0	69.07	75.77	0	67.47	0	0
	SD	0	0.06	0	0.12	0.12	0	0.06	0	0
511CCG_516GGC	Mean	0	65.9	71.87	0	75.83	0	67.47	0	0
	SD	0	0.1	0.15	0	0.12	0	0.06	0	0
512CGC_526AAC	Mean	0	63.7	73.67	0	0	72.53	67.53	0	0
	SD	0	0.1	0.12	0	0	0.15	0.06	0	0
512GGC_531TGG	Mean	0	66.75	74.05	0	0	73.65	0	71.1	0
	SD	0	0.06	0.06	0	0	0.13	0	0.08	0
513CTA_523GAG	Mean	0	0	71.55	0	0	72.28	67.08	0	0
	SD	0	0	0.06	0	0	0.05	0.13	0	0
513CTA_526TAA	Mean	0	0	73.23	0	0	71.8	67.7	0	0
	SD	0	0	0.15	0	0	0.1	0.1	0	0
515ATT_516TAC	Mean	69.87	0	0	68.74	75.81	0	67.2	0	0
	SD	0.17	0	0	0.19	0.12	0	0.1	0	0
515ATT_526AAC	Mean	69.6	0	0	70.03	0	72.5	67.17	0	0
	SD	0.1	0	0	0.06	0	0	0.12	0	0
516GGC_533CCG	Mean	70.85	0	71.77	0	75.78	0	0	0	61.62
	SD	0.1	0	0.05	0	0.08	0	0	0	0.13
516TAC_531TTG	Mean	69.8	0	0	68.5	0	73.23	0	73.8	0
	SD	0.1	0	0	0.1	0	0.06	0	0.1	0
522TTG_531GCG	Mean	69.57	0	0	70.3	0	72.87	67.7	0	0
	SD	0.31	0	0	0.1	0	0.12	0.1	0	0
530ATG_531TTC	Mean	69.5	0	73.63	0	0	69.53	0	0	64
	SD	0.17	0	0.21	0	0	0.12	0	0	0.1
512ACC_515ATT_526AAC	Mean	0	61.62	0	69.9	0	72.53	67.48	0	0
	SD	0	0.04	0	0.06	0	0.08	0.1	0	0

(Continued on next page)

TABLE 1 (Continued)

rpoB genotype	Mean or SD	Value (°C) for T_m windows as defined by the Ultra assay for: *rpo1*	*rpo1* mut	*rpo2*	*rpo2* mut	*rpo3*	*rpo3* mut	*rpo4*	*rpo4* MutA	*rpo4* MutB
Genotypes with two sets of reference T_m values										
526AGC-1	Mean	69.4	0	73.5	0	0	71.5	68.9	0	0
	SD	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*
526AGC-2[c]	Mean	69.53	0	73.61	0	0	71.59	0	0	0
	SD	0.11	0	0.18	0	0	0.13	0	0	0
526CGC-1	Mean	69.51	0	73.46	0	0	74.25	0	69.36	0
	SD	0.03	0	0.05	0	0	0.05	0	0.07	0
526CGC-2[c]	Mean	69.55	0	73.49	0	0	0	0	69.39	0
	SD	0.08	0	0.08	0	0	0	0	0.09	0
509AGG_526CGC-1	Mean	0	66.1	73.47	0	0	74.27	0	69.43	0
	SD	0	0.15	0.05	0	0	0.05	0	0.1	0
509AGG_526CGC-2[c]	Mean	0	66.09	73.55	0	0	0	0	69.3	0
	SD	0	0.14	0.14	0	0	0	0	0.21	0
Genotype with three sets of reference T_m values										
518GAC_533CCG-1[c]	Mean	69.42	0	0	0	75.77	0	0	0	61.45
	SD	0.13	0	0	0	0.12	0	0	0	0.14
518GAC_533CCG-2	Mean	69.6	0	71.5	0	75.9	0	0	0	61.7
	SD	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*
518GAC_533CCG-3	Mean	69.2	0	0	71.1	75.6	0	0	0	61.3
	SD	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*
Grouped genotypes										
526AAC or 526CTC	Mean	69.38	0	73.29	0	0	72.41	67.47	0	0
	SD	0.13	0	0.1	0	0	0.06	0.09	0	0
526GAC or 531TTC	Mean	69.42	0	73.56	0	0	72.19	68.15	0	0
	SD	0.17	0	0.23	0	0	0.21	0.15	0	0
516GAC del or 516GTC_518GAC	Mean	69.65	0	0	66.75	75.75	0	67.14	0	0
	SD	0.09	0	0	0.13	0.09	0	0.11	0	0

[a]$n = 9$ T_m windows. SD information was unavailable for 3 references because the cases are too rare to provide meaningful SD data.
[b]Silent mutation.
[c]References with only 3 positive T_m values.

Samples that have more than 4 positive T_m values are identified as "mixtures." If the T_m signature includes T_m values in all 4 wild-type windows plus a T_m value in one mutant T_m window, the mutant in the mixture would also be identified. Any other sample with more than 4 positive T_m values would be identified simply as a heteroresistant mixture because the specific mutations in the mixture could not be identified in these cases. This process was done using the R script.

Testing of DNA, BCG, and clinical isolates with the Ultra assay. The mutant DNA samples used to generate the reference library of T_m signatures and to perform the initial validation study were obtained from the TDR-TB Strain Bank, International Tuberculosis Research Center (Masan, South Korea), and Foundation of Innovative New Diagnostics (FIND). A *Mycobacterium bovis* BCG strain (ATCC catalog no. 35734) was used to generate the representative reference T_m signature for the WT T_m pattern. The *M. tuberculosis* cultures used in the blind validation study were obtained from FIND. For the DNA samples, the Ultra cartridge was preloaded with the DNA of interest and was then placed into a GeneXpert instrument (Cepheid, USA). A modified assay protocol was used that allows direct testing of DNA instead of sputum or cells (5). BCG cells were processed in a biosafety cabinet and tested using Ultra cartridges following the standard procedure. Each genotype was tested at least 3 times, including multiple tests performed on the same sputum sample and/or on multiple sputa. For algorithm evaluation, clinical isolates provided by FIND were processed in a biosafety level 3 (BSL3) facility and tested on Ultra cartridges using version 2 of the Ultra protocol following the standard procedure specified by Cepheid.

***T_m* reference samples.** Clinical DNA samples ($n = 81$) containing 40 different *rpoB* core region mutants as well as BCG were used to build the reference. The samples that had been tested using version 1 of the Ultra protocol were reanalyzed using the version 2 protocol. T_m values were recorded, and windows with no T_m value were given a value of zero so that every sample had 9 T_m values, and means of each of the 9 T_m values were calculated for each reference in the R script.

Human subject approval. This study did not use any human subjects.

RESULTS

Deriving reference *T_m* signatures specific to rifampin resistance mutations. Each Ultra assay generates a series of T_m values. Each of the generated T_m values that falls inside a T_m window can be localized into one of 9 T_m windows and used

FIG 1 Mean Ultra T_m value ± 1 SD determined for 41 *rpoB* genotypes. Mutations from 511CCG through 512ACC_515ATT_526AAC showed unique mutant T_m window patterns. Mutations from 516TAC through 513CTA_523GAG had undistinguishable mutant T_m window patterns, but they showed different WT T_m window patterns. Colored labels represent 3 mutant groups, each containing 2 mutations that have similar T_m patterns. The single asterisk (*) indicates that 513CAG is a silent mutation (Q513Q) and all 4 T_m values fall within WT windows. The double asterisks (**) indicate that the *rpo2* T_m value of 518GAC_533CCG could fall within the WT window or the Mut window or neither.

for subsequent analysis; those that fall outside a T_m window are excluded. The T_m values that are localized into T_m windows can be combined into a T_m signature. We investigated whether T_m signatures specific to individual RRDR mutations could be identified. In order to use T_m signatures generated by the Ultra assay for this purpose, we first needed to generate a reference library of T_m signatures that were specific for each mutation or each small "mutation group" of related mutations. We tested 82 samples comprising 41 different *rpoB* genotypes (1 WT genotype and 40 mutant genotypes) and recorded the mean T_m value present in each window for each mutation (Table 1). Most samples generated 4 positive T_m values; however, some mutations were found to generate only 3 T_m values because the 4th T_m value fell in a range that did not correspond to the defined T_m windows of the assay. A T_m value of zero was assigned to any window that did not contain a T_m value generated by the assay. An initial analysis performed using mean values for all test replicates showed that combining the T_m values in the 5 mutant and 4 wild-type windows enabled us to distinguish almost all mutations (Fig. 1, codons in black)

except for 3 "mutation groups" consisting of two mutations each (Fig. 1, codons in color).

A repeat analysis of individual (not mean) test results showed that slight variations in T_m values between some replicates (Fig. 1) would result in some errors in RRDR mutation identification unless a larger library of all possible T_m signature patterns corresponding to individual mutations was developed. For example, certain mutations such as 526CGC generated 3 T_m values in some replicates but 4 T_m values in others because the fourth T_m was very close to the edge of a defined T_m window and could be detected only if it fell inside the T_m window and was not detected if it fell outside the window. In such cases, we created 2 sets of T_m reference patterns: one that contained 4 positive T_m values and an alternate set that contained only 3 T_m values. In rare cases, we found samples that, upon repeat testing, produced T_m values that were very close to those within two or more T_m windows (for example, if 2 adjacent windows were in close proximity with a gap of $<$0.2 degree, a T_m at the very edge of one window would be able to cross to the other window due to minor T_m variations within standard deviation [SD] limits, during repeat runs). In these cases, 3 sets of alternative reference T_m signatures were made: one with the probe T_m value in one window, one with the probe T_m in an adjacent window, and one with the probe T_m in a temperature range that was between these two windows and which was not detected by the assay (T_m of zero for that window). However, this was observed only in case of the double mutant 518GAC_533CCG, for which we designated three T_m signature variants in the reference library (Table 1). In total, we determined that 43 different T_m signatures would best identify the 41 different *rpoB* RRDR genotypes in our test set. Of these, 31 genotypes could be identified by a single T_m signature each, one double mutant was identified by any of 3 possible signatures, and 3 mutations were identified by either of 2 possible signatures. The remaining 6 mutations produced T_m signatures that were not significantly different (P value $>$ 0.19) but were found to match in pairs. This enabled us to divide these 6 mutations into 3 groups of two different mutations in each group, with each group being identified by its group-specific T_m signature (Table 1).

Internal validation of T_m signatures. We next assessed how well the T_m signature reference library could be used to identify *rpoB* RRDR mutations using our TD value calculation method to determine the relatedness of a test sample to all the T_m signatures in the reference library. In this initial validation study, as a challenge set, we retested the samples used to generate the reference T_m signature library. This was done due to the limited availability of additional RR mutant samples. The output generated by these multiple comparisons consisted of the closest (TD-1) and the next-closest (TD-2) matches of the test sample to the mutations in the T_m signature library along with the TD values for each match. The genotype of a sample was identified using TD-1 and TD-2 values following a method illustrated in Fig. 2 and following processing through the use of the R script described in Materials and Methods. In this method, a perfect match results in a TD value of zero, and the cutoff value for TD-1 and TD-2 for calling a match is set to 0.8 because the maximum TD-1 value observed was 0.69 ($\log_{10}$ 0.69 $=$ $-$0.16 in Fig. 3). We first used a reference library without grouping the 6 mutations described earlier to test whether grouping was actually necessary. The majority of the 41 genotypes tested were correctly identified by examination of their TD-1 values alone (Fig. 3), which was expected since these tests were performed on the same samples as were used to generate the T_m signature references. Only 6 genotypes (526AAC, 526CTC, 526GAC, 531TTC, 516GAC deletion, and double mutant 516GTC 518GAC) in the challenge set were not matched to the correct references by TD-1 values alone in some replicates (colored labels in Fig. 3). These were the same 6 mutations belonging to the 3 sets that were grouped together in the final T_m reference library (Grouped genotypes in Table 1). The samples were tested against the final T_m reference library (which now included these 6 mutations grouped in 3 different sets of 2 mutations each), and all genotypes were correctly identified with TD-1 values

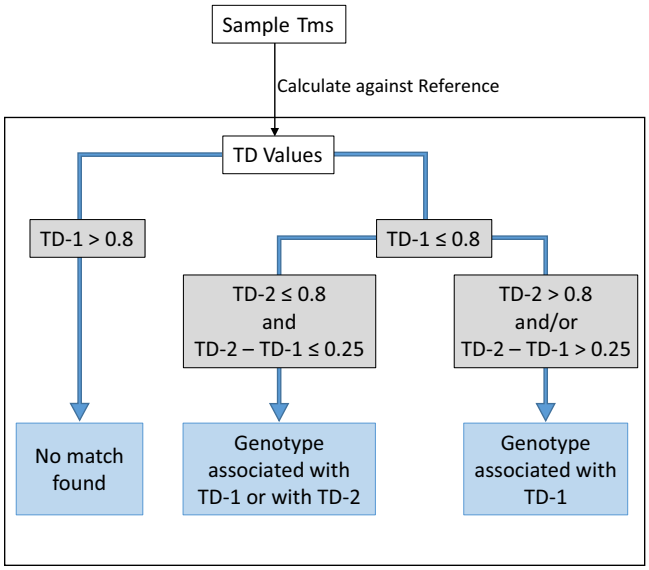

FIG 2 Algorithm for identifying mutants based on TD values.

between 0 and 0.69, of which 21 replicates from 7 genotypes had TD-2 – TD-1 = <0.25 (see Table S1 in the supplemental material).

Confirming the TD algorithm on clinical isolates. To validate the performance of our TD distance calculation approach to RRDR mutation identification, a new and previously untested set of 33 clinical isolates with various RRDR mutations was tested in a blind manner. Each RRDR mutation was also confirmed by Sanger sequencing. All isolates were tested using the Ultra assay. The T_m values generated by each assay were then analyzed by running the R script. Of the 33 isolates tested, 3 (2 514TTT and 1 531ATG) did not generate a match (i.e., all TD-1 values were >0.8) because they were not in the T_m reference library. Of the remaining 30 isolates, 19 were correctly identified by their TD-1 values alone (range, 0.1 to 0.56) and 8 were correctly identified by either their TD-1 value or their TD-2 value (ranges, 0.22 to 0.53 and 0.35 to 0.66, respectively), with one of these two closest matches representing the correct RRDR mutation. Finally, 3 samples were misidentified (Table 2), one of which (Isolate-15 in Table 2) had a mutation (515ATA_526AAC) that did not exist in our T_m reference library. This isolate was, therefore, incorrectly matched to an existing reference (515ATT_526AAC) with a TD-1 value of 0.15. Note that even in the absence of this particular double mutation in our reference library, a similar double mutation in the same two codons was identified as a possible correct match. The other 2 misidentified samples (Isolate-3 and Isolate-21 in Table 2) had mutations that existed in the T_m reference library; however, their correct references generated TD values that were higher than their TD-2 values for the mismatched mutations. However, the codons containing these mutations were identified correctly in each case also, despite the errors in identifying specific mutations within that codon. Thus, excluding samples whose genotypes did not exist in our reference library, the TD algorithm identified 27 of 29 mutations (93%), including 19/29 (65.5%) where a single mutation was correctly identified and 8/29 (27.6%) where a mutation group was correctly identified.

DISCUSSION

We have developed a method of identifying mutations in the *M. tuberculosis rpoB* core region. Our TD value-based method identified a wide range of mutations with high resolution, using an approach that is easy to automate. Our approach looks at the sequence-specific patterns generated by all the possible T_m values from four SMB probes covering the entire RRDR and not individual T_m values or T_m differences between WT and mutant sequences. Thus, it is not limited to identifying mutations

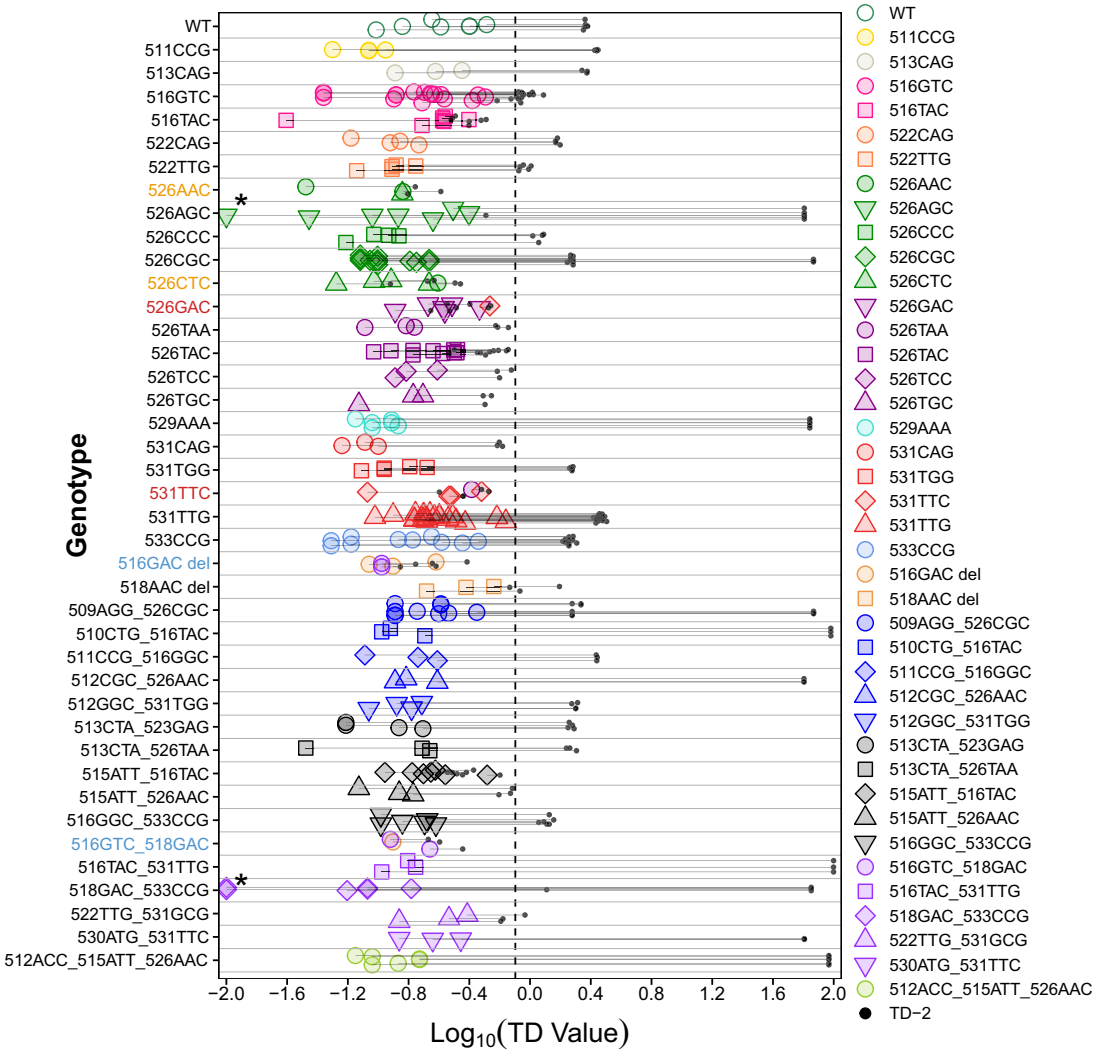

FIG 3 TD-1 and TD-2 values determined for each sample in logarithmic transformation. Each genotype was tested on multiple samples or for multiple times on one sample to generate at least 3 replicates. Horizontal lines indicate the distance between TD-1 (colored data points) and TD-2 (black dots) of individual replicates. Colors and shapes of each data point represent the reference mutant corresponding to TD-1 values. The vertical dashed line indicates the cutoff for calling a mutant. *, 3 samples of 2 genotypes had TD-1 values of 0; therefore, 0.01 was added to their TD-1 values before logarithmic transformation was performed.

associated only with RR, and silent mutations in the target region of the Ultra assay were able to be similarly identified. For example, the Ultra assay was designed to call the silent mutation 513CAG (Q513Q) as "RIF resistance not detected" (5) and the T_m values of all 4 probes tested against this silent mutation fell into WT T_m windows. However, our method distinguished this mutation from others with a true WT sequence due to specific patterns generated by WT and Q513Q silent mutation, which were recognized robustly using our algorithm. Our approach also showed improved mutation detection compared to alternative approaches that are based purely on whether a T_m value is present or absent in a T_m window or on T_m values present only in the mutant windows. For example, *rpoB* mutants 526AGC, 526CCC, 526TAA, and 526TGC generated the same T_m window patterns (*rpo1* WT, *rpo2* WT, *rpo3* Mut, and *rpo4* WT) and had almost identical mutant T_m values in the same mutant T_m window (*rpo3*). In spite of that, our method was able to specifically identify each mutation because our approach uses both the presence of a T_m and the actual T_m values within each window to identify a mutation (Fig. 1). Results showing differing T_m values in the *rpo4* WT window for these mutations even in cases in which they fell within the same window

TABLE 2 Clinical isolates tested using TD value-based algorithm

Sample	*rpoB* genotype	Result	Correct call?	Note
Isolate-1	526AAC	526AAC or 526CTC or 526TAC	Yes	
Isolate-2	516GTC	516GTC	Yes	
Isolate-3	526GAC	526TAA or 526AAC or 526CTC	No	TD-1 = 0.39, TD-2 = 0.56, TD value with correct reference = 0.58
Isolate-4	531TTG	531TTG	Yes	
Isolate-5	526TAC	526AAC or 526CTC or 526TAC	Yes	
Isolate-6	531TTG	531TTG	Yes	
Isolate-7	531TGG	531TGG	Yes	
Isolate-8	516GTC	516GTC or 516TAC	Yes	
Isolate-9	516GTC	516GTC	Yes	
Isolate-10	531TTG	531TTG	Yes	
Isolate-11	511CCG	511CCG (mixture)	Yes	
Isolate-12	526GAC	526GAC or 531TTC	Yes	
Isolate-13	531TTG	531TTG	Yes	
Isolate-14	526TAC	526AAC or 526CTC or 526TAC	Yes	
Isolate-15	515ATA_526AAC	515ATT_526AAC	No	
Isolate-16	516GTC	516GTC or 516TAC	Yes	
Isolate-17	526CTC	526AAC or 526CTC	Yes	
Isolate-18	514TTT	No match		TD-1 = 2.58
Isolate-19	514TTT	No match		TD-1 = 2.64
Isolate-20	516GTC	516GTC	Yes	
Isolate-21	516GTC	516TAC or 515ATT_516TAC	No	TD-1 = 0.53, TD-2 = 0.66, TD value with correct reference = 0.71
Isolate-22	531TTG	531TTG	Yes	
Isolate-23	531TTG	531TTG	Yes	
Isolate-24	526TAC	526AAC or 526CTC or 526TAC	Yes	
Isolate-25	531TGG	531TGG	Yes	
Isolate-26	516GTC	516GTC or 516TAC	Yes	
Isolate-27	531TTG	531TTG	Yes	
Isolate-28	511CCG	511CCG	Yes	
Isolate-29	526AAC	526AAC or 526CTC	Yes	
Isolate-30	531ATG	No match		TD-1 = 3.04
Isolate-31	526TAC	526AAC or 526CTC or 526TAC	Yes	
Isolate-32	531TTG	531TTG	Yes	
Isolate-33	531TTG	531TTG	Yes	

were enough for the algorithm to distinguish these mutation variants in the same codon (Table 1). Most of the *M. tuberculosis* isolates with so-called "disputed mutations," which can be difficult to identify as RR using phenotypic methods due to their borderline RR MICs (511CCG, 516TAC, 533CCG, and 526AGC) (9), were also specifically identified. Two other such mutations, 526AAC and 526CTC (9, 10), were identified as a group due to their identical T_m signatures. Differentiating these disputed mutations from the others may help to explain at least some of the occasional discordances between Ultra and phenotypic results, where samples that were identified as RR by Ultra assay were identified as RS in phenotypic tests (11).

The other advantage of our method is that it can correctly identify the RRDR mutations in samples that generate only 3 positive T_m values if at least one of the 3 T_m values falls into a mutant T_m window. The latest version of the Ultra assay (Xpert MTB/RIF Ultra V2) detects samples with 3 T_m values as RR but does so only when either the *rpo3* or *rpo4* T_m values fall within the "Mut" or "MutA" T_m window, respectively, and when the *rpo1* and *rpo2* T_m values fall into wild-type windows (internal Rutgers/Cepheid study data: this information has not yet been published). The rare RRDR mutations that result in 3 T_m values but which do not meet these criteria, such as 529AAA, would be detected as "RIF Indeterminate" by the Ultra assay, but these mutations would still be identified as RR by our method. In our T_m reference library, there are 4 RRDR genotypes with only 3 positive T_m values, and each has a unique T_m signature (Table 1). Our method can also identify RRDR mutant genotypes in many heteroresistant samples, i.e., test samples that contain DNA from both WT and RRDR mutant *M. tuberculosis* strains. Heteroresistance is identified by the presence of >4 T_m values in the 9 T_m windows. The RRDR mutant genotype can be identified in these

mixed samples if the Ultra assay produces a T_m value in each WT window plus an additional T_m value in a mutant window. All other signatures produced by heteroresistant samples are identified simply as a mixture.

A current limitation of our approach is that it is entirely based on known information. A mutation that has not yet been included in our T_m signature library would either be called "No Match" or, if it has a T_m signature close enough to one of the T_m signatures in the library, would be misidentified as the RDRR mutant with the closest-matching signature. In our study performed with clinical isolates, mutations of 3 isolates could not be identified and a double mutation was misidentified because their genotypes were not present in our library. Assays that produce T_m values that represent larger-than-expected variations from reference T_m values may also result in RRDR misidentifications. Two RRDR mutations in the clinical isolates appear to have been misidentified for this reason (see Fig. S1 in the supplemental material). In these two cases, the TD values of the correct mutation were slightly greater than the TD-2 values generated by our approach (0.55 versus 0.53 and 0.73 versus 0.67, respectively). However, the mutant codon was still correctly identified in these two samples, which underscores the utility of our approach. Furthermore, an expansion of our T_m signature library resulting from the addition of more *M. tuberculosis* samples and including clinically relevant "deletion" mutants is likely to produce more-precise mean values that could improve mutation identification. It is also possible that once enough data have been collected, it would be possible to modify our R identification tool to include the distribution of reference T_m values in addition to the mean values and to calculate the possibilities corresponding to each genotype call, which could further improve the specificity of our mutation identification.

Our study demonstrated the feasibility of using the wide range of T_m profiles generated in the Ultra assay and over 100 independent GeneXpert modules to specifically identify clinically prevalent RR mutations and opened the possibility of using Ultra data as an epidemiological tool. Given the performance of this approach applied to a large number of GeneXpert modules, this method should be widely applicable to others using a shared T_m signature reference library. As more T_m data are collected and added to the T_m signature reference library, the enrichment of the data will further evolve and optimize the algorithm for higher accuracy and improve its applicability. In clinical practice, the use of TD values may also be valuable as a stand-alone method to distinguish two clinical samples that clearly have different genotypes from samples that may have the same genotype, and this information could rule out transmission or relapse.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.01 MB.

SUPPLEMENTAL FILE 2, XLSX file, 0.1 MB.

ACKNOWLEDGMENTS

This work was funded by U.S. National Institutes of Health (NIH) grant AI111397, a grant from the Foundation for Innovative New Diagnostics. Research support was also provided by Cepheid.

A.M.S., D.K., K.T., D.L., D.P., R.K., and S.C. are employed by Cepheid Inc., which sells the Xpert MTB/RIF Ultra product. D.A. and S.C. earn royalties from the sale of Xpert MTB/RIF Ultra.

REFERENCES

1. Conceicao EC, Guimaraes AED, Lopes ML, Furlaneto IP, Rodrigues YC, da Conceicao ML, Barros WA, Cardoso NC, Sharma A, Lima L, Gomes HM, Duarte RS, Frota C, Rutaihwa LK, Gagneux S, Suffys PN, Lima K. 2018. Analysis of potential household transmission events of tuberculosis in the city of Belem, Brazil. Tuberculosis (Edinb) 113:125–129. https://doi.org/10.1016/j.tube.2018.09.011.
2. Kohl TA, Diel R, Harmsen D, Rothganger J, Walter KM, Merker M, Weniger T, Niemann S. 2014. Whole-genome-based Mycobacterium tuberculosis

surveillance: a standardized, portable, and expandable approach. J Clin Microbiol 52:2479–2486. https://doi.org/10.1128/JCM.00567-14.
3. Musser JM. 1995. Antimicrobial agent resistance in mycobacteria: molecular genetic insights. Clin Microbiol Rev 8:496–514. https://doi.org/10.1128/CMR.8.4.496.
4. Nusrath Unissa A, Hanna LE. 2017. Molecular mechanisms of action, resistance, detection to the first-line anti tuberculosis drugs: rifampicin and pyrazinamide in the post whole genome sequencing era. Tuberculosis (Edinb) 105:96–107. https://doi.org/10.1016/j.tube.2017.04.008.
5. Chakravorty S, Simmons AM, Rowneki M, Parmar H, Cao Y, Ryan J, Banada PP, Deshpande S, Shenai S, Gall A, Glass J, Krieswirth B, Schumacher SG, Nabeta P, Tukvadze N, Rodrigues C, Skrahina A, Tagliani E, Cirillo DM, Davidow A, Denkinger CM, Persing D, Kwiatkowski R, Jones M, Alland D. 2017. The New Xpert MTB/RIF Ultra: improving detection of Mycobacterium tuberculosis and resistance to rifampin in an assay suitable for point-of-care testing. mBio 8:e00812-17. https://doi.org/10.1128/mBio.00812-17.
6. Ng KCS, van Deun A, Meehan CJ, Torrea G, Driesen M, Gabriels S, Rigouts L, Andre E, de Jong BC. 2018. Xpert Ultra can unambiguously identify specific rifampin resistance-conferring mutations. J Clin Microbiol 56:e00686-18. https://doi.org/10.1128/JCM.00686-18.
7. Jamieson FB, Guthrie JL, Neemuchwala A, Lastovetska O, Melano RG, Mehaffy C. 2014. Profiling of rpoB mutations and MICs for rifampin and rifabutin in Mycobacterium tuberculosis. J Clin Microbiol 52:2157–2162. https://doi.org/10.1128/JCM.00691-14.
8. Chakravorty S, Aladegbami B, Burday M, Levi M, Marras SAE, Shah D, El-Hajj HH, Kramer FR, Alland D. 2010. Rapid universal identification of bacterial pathogens from clinical cultures by using a novel sloppy molecular beacon melting temperature signature technique. J Clin Microbiol 48:258–267. https://doi.org/10.1128/JCM.01725-09.
9. Van Deun A, Barrera L, Bastian I, Fattorini L, Hoffmann H, Kam KM, Rigouts L, Rusch-Gerdes S, Wright A. 2009. Mycobacterium tuberculosis strains with highly discordant rifampin susceptibility test results. J Clin Microbiol 47:3501–3506. https://doi.org/10.1128/JCM.01209-09.
10. Van Deun A, Aung KJM, Bola V, Lebeke R, Hossain MA, de Rijk WB, Rigouts L, Gumusboga A, Torrea G, de Jong BC. 2013. Rifampin drug resistance tests for tuberculosis: challenging the gold standard. J Clin Microbiol 51:2633–2640. https://doi.org/10.1128/JCM.00553-13.
11. Dorman SE, Schumacher SG, Alland D, Nabeta P, Armstrong DT, King B, Hall SL, Chakravorty S, Cirillo DM, Tukvadze N, Bablishvili N, Stevens W, Scott L, Rodrigues C, Kazi MI, Joloba M, Nakiyingi L, Nicol MP, Ghebrekristos Y, Anyango I, Murithi W, Dietze R, Lyrio Peres R, Skrahina A, Auchynka V, Chopra KK, Hanif M, Liu X, Yuan X, Boehme CC, Ellner JJ, Denkinger CM, Dorman SE, Schumacher SG, Alland D, Nabeta P, Armstrong DT, King B, Hall SL, Chakravorty S, Cirillo DM, Tukvadze N, Bablishvili N, Stevens W, Scott L, Rodrigues C, Kazi MI, Joloba M, Nakiyingi L, Nicol MP, et al. 2018. Xpert MTB/RIF Ultra for detection of Mycobacterium tuberculosis and rifampicin resistance: a prospective multicentre diagnostic accuracy study. Lancet Infect Dis 18:76–84. https://doi.org/10.1016/S1473-3099(17)30691-6.