ORIGINAL ARTICLE

Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing

The CRyPTIC Consortium and the 100,000 Genomes Project

ABSTRACT

BACKGROUND

The World Health Organization recommends drug-susceptibility testing of *Mycobacterium tuberculosis* complex for all patients with tuberculosis to guide treatment decisions and improve outcomes. Whether DNA sequencing can be used to accurately predict profiles of susceptibility to first-line antituberculosis drugs has not been clear.

METHODS

We obtained whole-genome sequences and associated phenotypes of resistance or susceptibility to the first-line antituberculosis drugs isoniazid, rifampin, ethambutol, and pyrazinamide for isolates from 16 countries across six continents. For each isolate, mutations associated with drug resistance and drug susceptibility were identified across nine genes, and individual phenotypes were predicted unless mutations of unknown association were also present. To identify how whole-genome sequencing might direct first-line drug therapy, complete susceptibility profiles were predicted. These profiles were predicted to be susceptible to all four drugs (i.e., pansusceptible) if they were predicted to be susceptible to isoniazid and to the other drugs or if they contained mutations of unknown association in genes that affect susceptibility to the other drugs. We simulated the way in which the negative predictive value changed with the prevalence of drug resistance.

RESULTS

A total of 10,209 isolates were analyzed. The largest proportion of phenotypes was predicted for rifampin (9660 [95.4%] of 10,130) and the smallest was predicted for ethambutol (8794 [89.8%] of 9794). Resistance to isoniazid, rifampin, ethambutol, and pyrazinamide was correctly predicted with 97.1%, 97.5%, 94.6%, and 91.3% sensitivity, respectively, and susceptibility to these drugs was correctly predicted with 99.0%, 98.8%, 93.6%, and 96.8% specificity. Of the 7516 isolates with complete phenotypic drug-susceptibility profiles, 5865 (78.0%) had complete genotypic predictions, among which 5250 profiles (89.5%) were correctly predicted. Among the 4037 phenotypic profiles that were predicted to be pansusceptible, 3952 (97.9%) were correctly predicted.

CONCLUSIONS

Genotypic predictions of the susceptibility of *M. tuberculosis* to first-line drugs were found to be correlated with phenotypic susceptibility to these drugs. (Funded by the Bill and Melinda Gates Foundation and others.)

The members of the writing group (Timothy M. Walker, D.Phil., A. Sarah Walker, Ph.D., and Tim E.A. Peto, D.Phil.) assume responsibility for the overall content and integrity of this article. The authors' full names and academic degrees are listed in the Appendix. The authors' affiliations are listed in the Supplementary Appendix, available at NEJM.org. Address reprint requests to Dr. Timothy Walker at the Department of Microbiology, Level 7, John Radcliffe Hospital, Headley Way, Headington, Oxford, OX3 9DU, United Kingdom, or at timothy.walker@ndm.ox.ac.uk.

This article was published on September 26, 2018, at NEJM.org.

This is the *New England Journal of Medicine* version of record, which includes all *Journal* editing and enhancements. The Author Final Manuscript, which is the author's version after external peer review and before publication in the *Journal*, is available under a CC BY license at PMC6121966.

DOI: 10.1056/NEJMoa1800474

Mycobacterium tuberculosis killed more people than any other pathogen in 2016, a year in which there were estimated to be more than 10 million active cases and 1.7 million patients who died from tuberculosis.[1] In 2014, the World Health Organization (WHO) set a target to "END TB" by 2035, acknowledging that success depends on the development of better preventative, diagnostic, and therapeutic interventions. The global emergence of antimicrobial resistance poses a major challenge. Despite a call for universal access to drug-susceptibility testing to guide individualized therapy, the high costs of the testing and shortages of people with the skills necessary to conduct it mean that it is unavailable in many countries with the greatest need. Consequently, only 22% of an estimated 600,000 patients requiring treatment for multidrug-resistant tuberculosis received diagnoses and were treated in 2016,[1] which facilitated the onward transmission of multidrug-resistant strains.[2]

The Xpert MTB/RIF assay (Cepheid) has partially eased the global diagnostic need. It uses polymerase chain reaction (PCR) technology to identify both *M. tuberculosis* complex and mutations in the *rpoB* gene (predictive of multidrug resistance) directly from clinical samples.[3] However, because the assay targets only a few potential resistance-conferring mutations, antimicrobial susceptibility cannot be reliably inferred from a negative result.[4] To devise individualized therapies, a diagnostic assay is needed to determine which drugs to give, in addition to which drugs to avoid.

Advances in whole-genome sequencing mean that it is now feasible to consider how this technology can aid in the assessment of drug susceptibility. Whole-genome sequencing is faster, more scalable, and likely to become less expensive than phenotypic testing.[5] If all resistance-conferring mutations were known, it should be possible to infer *M. tuberculosis* antimicrobial susceptibility from their absence, because the number of genomic sites that whole-genome sequencing covers is virtually unrestricted,[6] although resistance mechanisms with complex underlying gene interactions may not be detected. Here, we assess how well whole-genome sequencing performs for the detection of susceptibility to first-line antituberculosis drugs, given existing knowledge, as compared with the standards set forth in WHO target product profiles for new molecular assays[7]; we also assess whether whole-genome sequencing can be used to accurately guide antituberculosis therapy.

Methods

Sample Selection

We analyzed a total of 23 collections of *M. tuberculosis* complex isolates from 16 countries, each sequenced as part of population-based or diagnostic studies (Table 1, and Table S1 in the Supplementary Appendix, available with the full text of this article at NEJM.org). Six collections from Germany, Italy, the Netherlands, and the United Kingdom were unenriched for antimicrobial resistance and were sequenced largely prospectively. Seventeen other collections from across six continents were enriched for antimicrobial resistance. Analyses of both the unenriched and the complete collection were planned.

Sequencing

Isolates were sequenced on Illumina platforms, and the reads were processed by the Public Health England bioinformatics pipeline at Genomics England,[8] as described previously.[6] Stampy, version 1.0.17,[9] was used to map reads (with repetitive regions masked) to the *M. tuberculosis* reference genome (GenBank accession number, NC_000962.2), which is susceptible to the four first-line antituberculosis drugs isoniazid, rifampin, ethambutol, and pyrazinamide (i.e., pansusceptible). SAMtools mpileup, version 0.1.18,[10] was used to make variant calls based on a minimum read depth of 5× and at least one read on each strand. Mixed calls were assigned where minority alleles composed more than 10% of the read depth. Insertions and deletions were identified with Cortex, version 1.0.5.21.[11]

Drug-Susceptibility Testing and Prediction

Phenotypic drug-susceptibility testing was performed locally with the use of an MGIT 960 system (Becton Dickinson), by culture on 7H10 or Löwenstein–Jensen agar, or by microscopic-observation drug-susceptibility (MODS) assay, with method-specific critical concentrations for isoniazid (MGIT, 0.1 to 0.2 μg per milliliter; agar, 0.2 μg per milliliter; and MODS, 0.4 μg per milliliter), rifampin (MGIT, 1.0 μg per milliliter; agar, 40 μg per milliliter), ethambutol (MGIT, 5.0 μg

Table 1. Numbers of Isolates According to Country of Sample Origin and Drug-Resistance Profile.

Country	Period Isolated	Enriched for Resistance	Susceptible to All Four Drugs	Susceptible to Three Drugs*	Isoniazid-Resistant, Rifampin-Susceptible	Isoniazid-Susceptible, Rifampin-Resistant	Isoniazid-Resistant, Rifampin-Resistant	Other Pattern	Total
Australia	2006–2016	Yes	0	0	4	0	38	0	42
Belgium	2007–2015	Yes	121	0	2	0	97	14	234
Canada	2003–2014	Yes	11	1,118	164	14	24	12	1,343
China	2009–2012	Yes	0	44	0	0	236	0	280
Germany	1998–2015	No	248	0	9	1	13	2	273
Italy	2008–2016	Yes and no†	82	1	9	0	132	2	226
Netherlands	1993–2016	Yes and no†	420	42	24	1	149	31	667
Pakistan	2014–2015	Yes	47	5	11	6	345	1	415
Peru	1997–2009	Yes	24	12	49	18	199	13	315
Russia	2008–2010	Yes	282	0	116	15	407	22	842
Serbia	2008–2014	Yes	0	0	0	0	105	0	105
South Africa	2012–2014	Yes	593	11	37	69	151	130	991
Spain	2013–2015	Yes	45	3	5	2	8	1	64
eSwatini‡	2009–2010	Yes	2	130	14	4	116	7	273
Thailand	1998–2013	Yes	0	53	7	4	188	0	252
United Kingdom	2009–2017	Yes and no†	3,036	82	167	6	442	154	3,887
Total			4,911	1,501	618	140	2,650	389	10,209

* Isolates in this category were missing results for pyrazinamide.
† More than one collection was derived from Italy, the Netherlands, and the United Kingdom, some of which were enriched and some of which were not enriched for resistance. Details are provided in the Supplementary Appendix.
‡ Until recently, eSwatini was known as Swaziland.

per milliliter; agar, 0.2 μg per milliliter), and pyrazinamide (100 μg per milliliter). Not all laboratories routinely tested all agents (Table S1 in the Supplementary Appendix). Genotypic predictions were based on mutations in or upstream from genes associated with resistance to isoniazid (*ahpC*, *inhA*, *fabG1*, and *katG*), rifampin (*rpoB*), ethambutol (*embA*, *embB*, and *embC*), and pyrazinamide (*pncA*).[6] A knowledge base of mutations that are predictive of resistance or consistent with susceptibility was informed by the molecular targets of WHO-recommended line-probe assays (MTBDR*plus* or MTBDR*sl*, version 1.0 [HAIN Lifesciences]), a systematic literature review,[12] the Centers for Disease Control and Prevention (CDC) panel, and two recent studies that had no isolates in common with the present study (Table S2 in the Supplementary Appendix),[6,13] of which one became available after the present study commenced.[13]

Isolates containing resistance mutations were predicted to be phenotypically resistant, whereas isolates containing only wild-type sequence, phylogenetic mutations,[6] or mutations that were considered to be consistent with susceptibility were predicted to be susceptible. Predictions were withheld for isolates containing mutations that affect target genes but that are of unknown association or in instances in which no nucleotide call could be determined at a resistance-associated site. In these circumstances, the genotype was reported as "unknown" or "failed," respectively. Using phenotypic results as the standard, we calculated the sensitivity, specificity, and negative and positive predictive values for the correct assignment of susceptibility or resistance. For the primary analyses, we excluded phenotypes without a prediction.

Laboratory error was assumed in instances in which three or more phenotypes were discordant with the genotype of an isolate or in which susceptible phenotypes were recorded despite the presence of the high-level resistance mutation *katG* S315T for isoniazid or *rpoB* S450L for rifampin.[14] Such isolates were excluded from further analysis.

The analysis was performed with the use of Stata software, version 13.1 (StataCorp). No institutional-review-board approval was required, because this study used only data from mycobacteria. In Thailand, approval was granted through Mahidol University as part of a larger study.

RESULTS

PREDICTION OF PHENOTYPIC SUSCEPTIBILITY OR RESISTANCE TO INDIVIDUAL DRUGS

A total of 10,290 isolates were available for the study, of which 38 were associated with three or four phenotype–genotype discrepancies. High-level resistance mutations were found in 37 phenotypically susceptible isolates: 25 with the *katG* S315T mutation, which confers resistance to isoniazid, and 12 with the *rpoB* S450L mutation, which confers resistance to rifampin; 6 additional phenotypically susceptible isolates contained both of these mutations. All 81 of these isolates (0.8% of the total sample) were excluded from further analysis because of likely laboratory mislabeling. Of the 10,209 isolates that remained, full first-line phenotypic profiles were available for 7516 (73.6%), and partial profiles were available for the remainder. A total of 4911 (48.1%) isolates were phenotypically susceptible to all drugs (Table 1).

For each isolate, the complete sequence of nine genes and their promoter regions was interrogated to make genotypic predictions of each available phenotypic result. Predictions could be made for 8405 (93.6%) of 8976 phenotypic test results indicating resistance and 26,879 (93.5%) of 28,746 phenotypic test results indicating susceptibility; the remainder were from isolates that had uncharacterized mutations or were missing key nucleotide calls. For isoniazid, rifampin, ethambutol, and pyrazinamide, the sensitivity of genotypic prediction (i.e., the percentage of phenotypic test results indicating resistance that had concordant genotypic predictions) was 97.1%, 97.5%, 94.6%, and 91.3%, respectively, and the specificity (i.e., the percentage of phenotypic test results indicating susceptibility that had concordant genotypic predictions) was 99.0%, 98.8%, 93.6%, and 96.8%. In comparison, the results expected from WHO-recommended molecular assays (Xpert MTB/RIF, MTBDR*plus*, and MTBDR*sl*, version 1.0) on the basis of the mutations they probe having been identified from the genome-sequence data showed a significantly lower sensitivity than whole-genome sequencing for isoniazid, rifampin, and ethambutol ($P<0.001$) but a greater specificity for isoniazid and ethambutol ($P<0.001$) (Table 2).

The negative predictive value of whole-genome sequence analysis (i.e., the percentage of geno-

typic predictions of susceptibility that were correct) was greater than 98.5% for all four drugs. Although it is necessarily dependent on the prevalence of resistance, the negative predictive value for each drug also varied according to the phenotypes of susceptibility or resistance to the other three drugs. For example, at a prevalence of pyrazinamide resistance of 20%, the expected negative predictive value for pyrazinamide was 93.6% and 99.0% for isolates that were susceptible and resistant, respectively, to the other three drugs (Table 3, and Table S3 in the Supplementary Appendix).

Because some collections included clustered isolates, the analysis was repeated after random selection of one representative among genomically indistinguishable isolates and again from isolates that were within five single-nucleotide polymorphisms (SNPs) of another isolate. No significant change in sensitivity or specificity was observed for any drugs ($P>0.1$) (Table S4 in the Supplementary Appendix).

To reflect the emerging practice of routinely sequencing isolates for clinical care, the analysis was repeated for the subset of 4397 isolates from German, Italian, Dutch, and U.K. collections that were not enriched for resistance. Among these isolates, 335 (7.6%) were isoniazid-resistant and 125 (2.8%) were multidrug-resistant. For each drug, the specificity and negative predictive values were higher and the positive predictive values (the percentage of genotypic predictions of resistance that were correct) lower than in the overall results. There was no significant difference in sensitivity (Table 2).

PREDICTION OF COMPLETE PHENOTYPIC SUSCEPTIBILITY PROFILES

In order for DNA sequencing to be useful for the individualization of therapy, a minimum requirement is that all phenotypes of resistance or susceptibility to first-line antimicrobial agents are predicted. Phenotypic profiles were thus predicted for 7516 isolates that had phenotypic data available for all first-line drugs (Tables S1 and S6 in the Supplementary Appendix). "Unknown" or "failed" predictions for at least one drug were reported for 1651 profiles (22.0%). A total of 5865 profiles (78.0%) were predicted completely, of which 5250 (89.5%) were predicted correctly (Table S5 in the Supplementary Appendix). Among the 5865 phenotypic profiles with complete genotypic predictions, 4037 were predicted to be susceptible to all four drugs, of which 3952 (97.9%) were predicted correctly; these 3952 correctly predicted profiles account for 98.6% of the 4007 phenotypically pansusceptible isolates for which complete predictions were made (Table 4).

Because the percentage of incompletely predicted profiles was substantial (22.0%), we assessed whether pansusceptibility could still be accurately predicted for some of these isolates. Because susceptibility to isoniazid predicts susceptibility to other first-line drugs,[15] we maximized the confidence in isoniazid predictions by making predictions only in the absence of "unknown" mutations in isoniazid-related genes. Unknown mutations that were relevant to other drugs were permitted. When this was done, pansusceptibility was correctly predicted for 4481 (97.8%) of 4582 isolates, including 545 (33.0%) of 1651 previously incompletely predicted profiles (Table 4). Among the collections that were unenriched for resistance, 3439 (99.7%) of 3450 profiles were thereby correctly predicted to be susceptible to all four drugs (Table S7 in the Supplementary Appendix).

To simulate how this approach would perform in contexts with differing burdens of antimicrobial resistance, we assessed the decline in negative predictive value associated with an increasing prevalence of resistance to individual drugs and with an increasing prevalence of any resistance within drug profiles. We randomly subsampled 1000 isolates to represent every 1-percentage-point increment in the prevalence of antimicrobial resistance between 10% and 90% and repeated this step 1000 times for each drug and for complete drug profiles. The negative predictive value declined further for ethambutol and pyrazinamide than for complete drug profiles, but it declined least for isoniazid and rifampin. Below a 47.0% prevalence of resistance to any drug, the simulated negative predictive value remained above 95% for 97.5% of drug profiles (Fig. 1).

DISCREPANCY ANALYSES

In Australia, 11 ethambutol-susceptible isolates containing *embB* mutations associated with resistance to ethambutol were rephenotyped. Three

Table 2. Prediction of Phenotypes of Resistance or Susceptibility to Individual Drugs.*

Analysis and Drug	Resistant Phenotype					Susceptible Phenotype					Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Sensitivity, All†	Specificity, All†	NGP	RP
	R	S	U	F	Total	R	S	U	F	Total								
	number of isolates										*percent*							
WGS, all isolates																		
Isoniazid	3067	90	93	44	3294	65	6313	215	117	6710	97.1 (96.5–97.7)	99.0 (98.7–99.2)	97.9 (97.4–98.4)	98.6 (98.3–98.9)	93.1	94.1	4.7	32.9
Rifampin	2743	69	7	84	2903	85	6763	232	147	7227	97.5 (96.9–98.1)	98.8 (98.5–99.0)	97.0 (96.3–97.6)	99.0 (98.7–99.2)	94.5	93.6	4.6	28.7
Ethambutol	1410	81	94	55	1640	468	6835	781	70	8154	94.6 (93.3–95.7)	93.6 (93.0–94.1)	75.1 (73.0–77.0)	98.8 (98.5–99.1)	86.0	83.8	10.2	16.7
Pyrazinamide	863	82	117	77	1139	204	6146	197	108	6655	91.3 (89.3–93.0)	96.8 (96.3–97.2)	80.9 (78.4–83.2)	98.7 (98.4–99.0)	75.8	92.4	6.4	14.6
WRAs, all isolates‡																		
Isoniazid	2886	355	—	53	3294	27	6675	—	8	6710	89.0 (87.9–90.1)§	99.6 (99.4–99.7)§	99.1 (98.7–99.4)§	95.0 (94.4–95.5)§	—	—	0.6	32.9
Rifampin	2669	143	—	91	2903	129	6826	—	272	7227	94.9 (94.0–95.7)§	98.1 (97.8–98.4)¶	95.4 (94.5–96.1)¶	97.9 (97.6–98.3)§	—	—	3.6	28.7
Ethambutol	961	641	—	38	1640	241	7895	—	18	8154	60.0 (57.5–62.4)§	97.0 (96.6–97.4)§	80.0 (77.6–82.2)¶	92.5 (91.9–93.0)§	—	—	0.6	16.7
WGS, unenriched‖																		
Isoniazid	314	8	9	4	335	15	3770	104	90	3979	97.5 (95.2–98.9)	99.6 (99.3–99.8)§	95.4 (92.6–97.4)¶	99.8 (99.6–99.9)§	93.7	94.7	4.8	7.8
Rifampin	126	0	0	9	135	31	3958	103	116	4208	100.0 (97.1–100.0)	99.2 (98.9–99.5)**	80.3 (73.2–86.2)§	100.0 (99.9–100.0)§	93.3	94.1	5.2	3.1
Ethambutol	72	1	0	0	73	47	3711	458	36	4252	98.6 (92.6–100.0)	98.7 (98.3–99.1)§	60.5 (51.1–69.3)§	100.0 (99.8–100.0)§	98.6	87.3	11.4	1.7
Pyrazinamide	109	6	4	6	125	30	4003	14	58	4105	94.8 (89.0–98.1)	99.3 (98.9–99.5)§	78.4 (70.6–84.9)	99.9 (99.7–99.9)§	87.2	97.5	1.9	3.0

WRAs, unenriched††																		
Isoniazid	295	36	—	4	335	10	3965	—	4	3979	89.1 (85.3–92.3)§	99.7 (99.5–99.9)	96.7 (94.1–98.4)	99.1 (98.8–99.4)§	—	—	0.2	—
Rifampin	114	11	—	10	135	22	3957	—	229	4208	91.2 (84.8–95.6)§	99.4 (99.2–99.7)	83.8 (76.5–89.6)	99.7 (99.5–99.9)§	—	—	5.5	—
Ethambutol	57	16	—	0	73	29	4220	—	3	4252	78.1 (66.9–86.9)§	99.3 (99.0–99.5)**	66.3 (55.3–76.1)	99.6 (99.4–99.8)§	—	—	0.1	—

* NGP denotes no genotypic prediction, NPV negative predictive value, PPV positive predictive value, and RP resistance prevalence. Unless otherwise indicated, percentages are based on genotypic predictions of resistant (R) or susceptible (S) only (i.e., excluding isolates with mutations of unknown association [U] and genotypic predictions that failed because of missing data around a genomic resistance locus [F]). F was reported in the presence of minority alleles at relevant sites for the results expected from the World Health Organization (WHO)–recommended molecular assays just as for the whole-genome sequencing (WGS) predictions.

† Percentages were calculated with the total number of isolates (R, S, U, and F) as the denominator.

‡ Data are from predictions of the performance of the WHO-recommended assays (WRAs) (MTB/RIF Xpert and MTBDR*plus* or MTBDR*sl*) for all isolates. Expected predictions of resistance for the Xpert and MTBDR assays were based on the presence of non–wild type sequence within the genomic regions interrogated by these assays. P values are for comparisons with the analysis of all isolates. No results are shown for pyrazinamide, since there is no WRA for the detection of resistance or susceptibility to this drug.

§ P≤0.001.

¶ P≤0.01.

‖ Data are from collections of isolates from Germany, Italy, the Netherlands, and the United Kingdom that were not enriched for resistance. P values are for comparisons with the analysis of all isolates.

** P≤0.05.

†† Data are from predictions of the performance of the WRAs for collections that are not enriched for resistance. Expected predictions of resistance for the Xpert and MTBDR assays were based on the presence of non–wild type sequence within the genomic regions interrogated by these assays. P values are for comparisons with the analysis of collections isolates from Germany, Italy, the Netherlands, and the United Kingdom that were not enriched for resistance. No results are shown for pyrazinamide, since there is no WRA for the detection of resistance or susceptibility to this drug.

The New England Journal of Medicine

Table 3. Individual Drug Predictions against Different Background Phenotypic Profiles.*

Drug and Phenotypic Profile†	Resistant Phenotype					Susceptible Phenotype					Prevalence of Resistance	Sensitivity	Specificity	PPV	NPV	95% CI for Expected NPV at Given Prevalence of Resistance‡	Calculated NPV at 20% Prevalence of Resistance	Calculated NPV at 40% Prevalence of Resistance
	R	S	U	F	Total	R	S	U	F	Total								
	number of isolates										*percent*							
Isoniazid																		
–SSS	391	30	18	12	451	21	4653	133	104	4911	8.4	93	100	95	99.4	99.3–100	98.2	95.4
–RSS	459	21	20	6	506	7	85	5	1	98	83.8	96	92	98	80.2	83.5–100	98.8	96.9
–RRS	424	3	13	4	444	2	2	2	0	6	98.7	99	50	100	40.0	73.7–85.6	99.6	99.1
–SRS	24	4	1	0	29	0	10	1	0	11	72.5	86	100	100	71.4	90.5–95.6	96.6	91.3
–SSR	24	1	2	1	28	0	95	6	3	104	21.2	96	100	100	99.0	98.5–99.7	99.0	97.4
–RRR	662	3	11	4	680	0	0	0	0	0	100.0	100	—	100	0.0	73.7–85.6	—	—
–RSR	217	3	5	5	230	0	3	0	0	3	98.7	99	100	100	50.0	73.7–85.6	99.7	99.1
–SRR	13	0	0	2	15	0	0	0	0	0	100.0	100	—	100	—	73.7–85.6	—	—
Rifampin																		
S–SS	74	16	0	8	98	30	4632	126	123	4911	2.0	82	99	71	99.7	99.3–100	95.7	89.3
S–RS	6	0	0	0	6	1	9	1	0	11	35.3	100	90	86	100.0	97.8–99.5	100.0	100.0
S–SR	1	2	0	0	3	0	100	3	1	104	2.8	33	100	100	98.0	99.3–100	85.7	69.2
S–RR	0	0	0	0	0	0	0	0	0	0	—	—	—	—	—	—	—	—
R–SS	464	20	1	21	506	18	424	3	6	451	52.9	96	96	96	95.5	95.8–98.6	98.9	97.2
R–RS	424	7	2	11	444	4	25	0	0	29	93.9	98	86	99	78.1	76.2–86.6	99.5	98.8
R–SR	218	4	0	8	230	7	20	0	1	28	89.1	98	74	97	83.3	77.9–87.9	99.4	98.4
R–RR	665	2	0	13	680	10	3	0	2	15	97.8	100	23	99	60.0	76.2–86.6	99.7	99.1

Ethambutol																		
SS–S	1	9	1	0	11	4	4399	472	36	4911	0.2	10	100	20	99.8	98.8–99.9	81.6	62.5
RS–S	21	5	3	0	29	31	376	40	4	451	6.0	81	92	40	98.7	98.8–99.9	95.1	87.8
SR–S	4	2	0	0	6	1	93	3	1	98	5.8	67	99	80	97.9	98.8–99.9	92.2	81.7
RR–S	375	20	30	19	444	203	241	48	14	506	46.7	95	54	65	92.3	93.4–96.7	97.7	94.1
SS–R	0	0	0	0	0	1	81	22	0	104	0.0	—	99	0	100.0	98.8–99.9	—	—
RS–R	12	2	1	0	15	7	20	1	0	28	34.9	86	74	63	90.9	95.7–98.1	95.4	88.6
SR–R	0	0	0	0	0	0	3	0	0	3	0.0	—	100	—	100.0	98.8–99.9	—	—
RR–R	625	9	26	20	680	150	50	25	5	230	74.7	99	25	81	84.7	82.0–88.2	98.6	96.4
Pyrazinamide																		
SSS–	74	28	0	2	104	12	4826	13	60	4911	2.1	73	100	86	99.4	98.6–99.6	93.6	84.5
RSS–	13	8	4	3	28	5	431	2	13	451	5.8	62	99	72	98.2	98.6–99.6	91.2	79.6
RRS–	166	25	22	17	230	49	374	68	15	506	31.3	87	88	77	93.7	95.5–97.7	96.4	91.0
SRS–	0	3	0	0	3	0	97	0	1	98	3.0	0	100	—	97.0	98.6–99.6	80.0	60.0
RRR–	532	15	83	50	680	107	216	105	16	444	60.5	97	67	83	93.5	87.3–91.0	99.0	97.3
SRR–	0	0	0	0	0	0	6	0	0	6	0.0	—	100	—	100.0	98.6–99.6	—	—
RSR–	10	2	1	2	15	0	28	0	1	29	34.1	83	100	100	93.3	95.0–97.3	96.0	90.0
SSR–	0	0	0	0	0	0	11	0	0	11	0.0	—	100	—	100.0	98.6–99.6	—	—

* Percentages are based on R and S genotypic predictions only (excluding U and F).

† Phenotypic profiles are listed in the following order: isoniazid, rifampin, ethambutol, pyrazinamide. A dash (–) in one of the four positions indicates the drug-resistance phenotype being assessed.

‡ The expected NPV was calculated as specificity × (1 – prevalence) / [specificity × (1 – prevalence) + (1 – sensitivity) × prevalence]. For prevalences of less than 10% or greater than 90%, simulated values are given for 10% and 90%, respectively, because simulations were not performed below or above these values.

The New England Journal of Medicine

Table 4. Genotypic Drug Profile Predictions of Pansusceptibility.*

Prediction and Genotypic Drug Profile				No. of Isolates Predicted to Have Profile	No. of Phenotypically Pansusceptible Isolates Predicted to Have Profile (% Predicted Correctly)
Isoniazid	Rifampin	Ethambutol	Pyrazinamide		
Predicted to be pansusceptible					
S	S	S	S	4037	3952 (97.9)
Predicted to be pansusceptible when U mutations are inferred to be consistent with susceptibility					
S	S	S	U	11	11 (100)
S	S	U	S	410	399 (97.3)
S	S	U	U	2	2 (100)
S	U	S	S	93	88 (94.6)
S	U	U	S	29	29 (100)
Total				4582	4481 (97.8)
Predicted to have some phenotypic resistance					
R	S	R or S	R or S	397	18 (4.5)
S	At least one R, no U or F			158	36 (22.8)
R	R	R or S	R or S	1273	1 (0.1)
Total				1828	55 (3.0)
No prediction made; drug profile prediction incomplete					
U	S or U	S or U	S or U	150	126 (84.0)
At least one F, no R				280	240 (85.7)
At least one R and one U, no F				499	6 (1.2)
At least one R and one F, no U				159	3 (1.9)
At least one R, one U, and one F				18	0
Total				1106	375 (33.9)

* Among the 5865 profiles with complete predictions, the sensitivity of genetic prediction was 95.4%, specificity 98.6%, PPV 97.0%, and NPV 97.9%, with predictions made for 78.0% of isolates. When predictions were made only in the absence of U mutations in isoniazid-related genes (with U mutations that were relevant to other drugs permitted), the sensitivity was 94.6%, specificity 98.8%, PPV 97.0%, and NPV 97.8%, with predictions made for 85.1% of isolates.

repeat assays failed, but 7 of the remaining 8 yielded now-consistent resistant phenotypes. In Peru, 10 of 16 repeated assays continued to indicate phenotypic susceptibility by MODS, despite the presence of *fabG1* C–15T or G–17T mutations. In isolates from the Netherlands, 6 resistant phenotypes that had been predicted to be susceptible were identified as clerical errors, and 3 susceptible phenotypes that had been predicted to be resistant tested phenotypically resistant by means of alternative phenotypic assays (Table S8 in the Supplementary Appendix). Although additional rephenotyping was not possible, we conducted a "per mutation" analysis to further assess discrepancies.

Of the 322 resistant phenotypes that had been predicted to be susceptible, 290 (90.1%) were in isolates that had no mutations affecting targeted genes, and 32 (9.9%) were in isolates that had 1 or more of 15 mutations that had previously been characterized as being consistent with antimicrobial susceptibility. In support of this finding, across all isolates in which no mutation other than 1 or more of these 15 was found, the presence of the mutations correctly predicted susceptibility to isoniazid in 286 (97.6%) of 293 isolates and susceptibility to ethambutol in 95 (79.8%) of 119 isolates. The 1 mutation that was relevant to pyrazinamide was found in 2 isolates, both of which were phenotypically resistant. None of these

mutations were relevant to rifampin (Table S9 in the Supplementary Appendix).

Among the isolates with the 822 susceptible phenotypes that had been predicted to be resistant, 145 different resistance-conferring mutations were found. Of these, 142 (97.9%) featured as the only resistance-conferring mutation in at least 1 isolate in the data set, which allowed for the assessment of individual predictive performance. The presence of these mutations correctly predicted resistance to isoniazid in 308 (83.0%) of 371 isolates, to rifampin in 548 (87.4%) of 627, to ethambutol in 1280 (73.4%) of 1743, and to pyrazinamide in 459 (69.2%) of 663 (Table S9 in the Supplementary Appendix). Of the 17 mutations leading to predictions of resistance to rifampin in phenotypically susceptible isolates, 14 (82.4%) were in the genetic region targeted by Xpert MTB/RIF and MTBDR*plus*.

Mislabeling of laboratory samples probably also contributed to discrepant results. This possibility was assessed for each collection on the basis of the proportion of isolates that were excluded because of *katG* S315T or *rpoB* S450L mutations being associated with susceptible phenotypes, the discrepancy rate within the collection, and the prevalence of antimicrobial resistance (Table S10 in the Supplementary Appendix). Overall, approximately 43% of discrepancies for isoniazid and 12% of discrepancies for rifampin were thereby judged to be attributable to mislabeling.

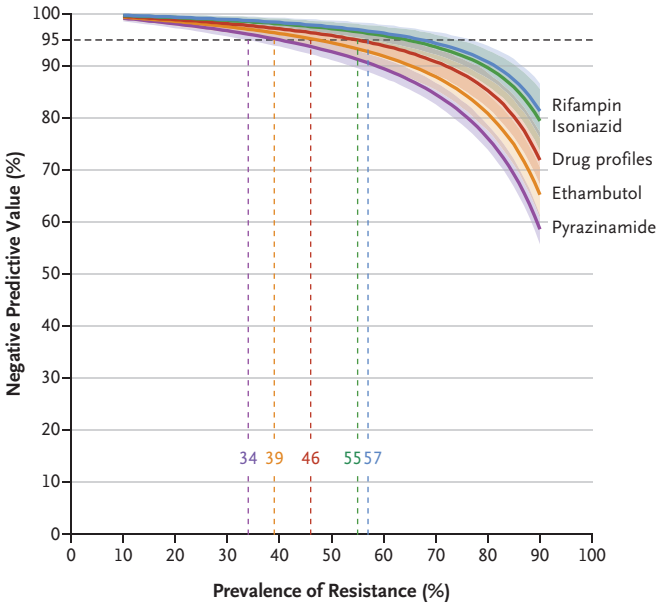

Figure 1. Simulated Negative Predictive Values for Individual Drugs and Complete Drug Profiles.

Negative predictive values are shown for individual drugs and complete drug profiles, according to the simulated prevalence of resistance to each drug, or within each drug profile (any resistance). For each percentage prevalence between 10% and 90%, 1000 isolates were randomly selected, 1000 times. Solid lines indicate the median, and shaded areas indicate the 95% confidence intervals. Vertical dashed lines indicate the prevalence at which the 95% confidence interval intersects a negative predictive value of 95%.

DISCUSSION

This analysis of more than 10,000 *M. tuberculosis* isolates collected from 16 countries across six continents and representing all major lineages (Table S1 in the Supplementary Appendix) suggests that whole-genome sequencing can now characterize profiles of susceptibility to first-line antituberculosis drugs with a degree of accuracy sufficient for clinical use. The importance of this is twofold. First, it shows that the genomic approach could be used to guide the choice of which drugs to prescribe and not just which drugs to avoid, in a way similar to phenotyping. Second, the data can be used to support plans to reduce the workload associated with culture and susceptibility analysis in places where routine whole-genome sequencing is performed.

The WHO target product profiles for new molecular assays for *M. tuberculosis* require more than 90% sensitivity and 95% specificity.[7] Overall, both these targets were met for all drugs with the exception of specificity for ethambutol (93.6%). This is no surprise, since phenotyping is an imperfect standard, particularly for isolates with *embB* mutations.[6,13,16] For the collections that were not enriched for resistance, however, these targets were met for all drugs, and they were also met for the predictions of pansusceptibility in all collections. Only categorical agreement was assessed for complete drug-susceptibility profile predictions, because of the number of permutations. These predictions met the external quality assurance criterion (>80% concordance) for the European tuberculosis reference laboratory network.[17]

There are three reasons for the predictions regarding pansusceptibility being approximately 98% correct. First, the knowledge base included both resistance-associated genomic mutations and mutations that were compatible with phenotypic susceptibility. Second, antituberculosis drug-sus-

ceptibility phenotypes are not independent of one another, which allows for the use of isoniazid susceptibility to predict susceptibility to other drugs. Third, no predictions were attempted for isolates that contained genomic variation of unknown association in genes affecting isoniazid. This maximized confidence in the isoniazid predictions that were made. Consequently, the performance in the prediction of drug profiles was better than that in the per-drug analysis for ethambutol and pyrazinamide, and although there was a slight corresponding decline in performance for isoniazid and rifampin, simulations showed that the prevalence of resistance would have to exceed that seen in most of the worst-affected countries in the world before these predictions no longer satisfied the WHO targets.[1]

Our findings showed substantially better performance of sequencing analysis relative to the the sensitivity that could be expected from WHO-recommended PCR-based assays because whole-genome sequencing is able to identify many more mutations. These additional mutations were, however, simultaneously responsible for the losses in specificity, largely because of the number of mutations for which a minority of isolates did not manifest a resistant phenotype. A typical example is the *rpoB* I491F mutation, which is frequently associated with a result indicating susceptibility to rifampin in liquid culture but has been linked to treatment failure.[4,18,19]

The broader discrepancy analysis highlighted the same phenomenon. Although the predictive performance of individual mutations, whether probed by WHO-recommended assays or not, was good, each mutation has the potential to be associated with an unexpected phenotype in a minority of isolates. This is most likely where a mutation elevates the minimum drug concentration required to inhibit bacterial growth to close to the concentration above which an isolate is considered resistant. Canonical ethambutol mutations are a classic example,[20] but there are many others, including the mutations missed by the MODS assay in Peru.[16,21,22] Such phenomena are thus likely to explain the majority of isolates that were predicted to be resistant yet were phenotypically susceptible. They are also the most likely reason for the prediction of pansusceptible drug profiles being more accurate than the prediction of profiles that are apparently resistant to one or more drugs.

One limitation of our study was our inability to definitively resolve most discrepancies because of the scale and cost of repeat sequencing and phenotyping. This was most worrisome for phenotypically resistant isolates that were predicted to be susceptible. For these discrepancies, possible explanations include further limitations of our study — namely, phenotypic error, resistant minority bacterial populations that went undetected by sequencing, mechanisms of resistance unknown to us, or laboratory labeling error. To maintain or improve accuracy, ongoing surveillance for the phenotypic effect of new mutations will be required. Another limitation is the use of phenotypic susceptibility data as the standard. The lack of clinical outcome data to link the antimicrobial-resistance phenotypes to treatment failure requires us to infer potential clinical benefit.

More work remains to be done before predictions can be extended to second- and third-line drugs and to newer compounds. However, after an external review, Public Health England has already decided to stop phenotyping isolates that are predicted to be susceptible to all first-line drugs (Crook D, National Infection Service: personal communication). Similar decisions have been made in the Netherlands (van Soolingen D, Rijksinstituut voor Volksgezondheid en Milieu: personal communication) and New York (Musser K, Wadsworth Center, New York State Department of Health: personal communication).

These data show how our understanding of the molecular determinants of resistance to first-line antituberculosis drugs allows us to consider using DNA sequencing to guide therapy. Similar performance must now be replicated for the remaining drugs.

The views expressed in this article are those of the authors and not necessarily those of the National Health Service, the National Institute for Health Research, the U.K. Department of Health, or Public Health England. Use of trade names is for identification only and does not constitute endorsement by the U.S. Department of Health and Human Services, the U.S. Public Health Service, or the Centers for Disease Control and Prevention (CDC). The findings and conclusions expressed by authors contributing to this journal do not necessarily reflect the official opinion of the CDC or the authors' affiliated institutions.

Supported by grants from the Bill and Melinda Gates Foundation (OPP1133541, to CRyPTIC, plus separate support to Dr. Rodwell), a Wellcome Trust/Newton Fund–MRC Collaborative Award (200205/Z/15/Z, to CRyPTIC), the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC) and NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at University of Oxford in partnership with Public Health England, the NIHR Biomedical Research Centre at Barts, the NIHR Biomedical Research Centre at Imperial, the NIHR and NHS England (to the 100,000 Genomes Project, which is managed by Genomics

England, a wholly owned company of the U.K. Department of Health), the Wellcome Trust, the Medical Research Council, Public Health England, a grant from the National Science and Technology Key Program of China (2014ZX10003002), a grant from the National Basic Research program of China (2014CB744403), a grant from the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB29020000), a grant from the European Commission Seventh Framework Program (FP7/2007-2013, to Borstel under grant agreement 278864 in the framework of the Patho-NGen-Trace project), the German Center for Infection Research (to Borstel), Leibniz Science Campus Evolutionary Medicine of the Lung (EvoLUNG), the Belgian Ministry of Social Affairs (to the Belgian Reference Center for Tuberculosis and Mycobacteria from Bacterial Diseases Service through a fund within the Health Insurance System), the French governmental program "Investing for the Future" (to Genoscreen), a grant from the European Commission Seventh Framework Program (FP7/2007-2013, to Genoscreen under grant agreement 278864 in the framework of the Patho-NGen-Trace project), grants from the Drug Resistant Tuberculosis Fund (R015833003, to Dr. Chaiprasert), the Faculty of Medicine, Siriraj Hospital, Mahidol University (to Dr. Chaiprasert), a grant from the Ministry of Economy and Competitiveness (MINECO), Spain (SAF2016-77346-R, to Dr. Comas), a grant from the European Research Council (638553-TB-ACCELERATE, to Dr. Comas), a grant from the BC Centre for Disease Control Foundation for Population and Public Health (to Dr. Gardy), a grant from the British Colombia Lung Association (to Dr. Gardy), grants from the Wellcome Trust and the Royal Society (101237/Z/13/Z and 102541/A/13/Z, to Drs. Wilson and Iqbal [Sir Henry Dale Fellows]), a grant from the National University of Singapore Yong Loo Lin School of Medicine Aspiration Fund (NUHSRO/2014/069/AF-New Idea/04, to Drs. Ong and Teo), a European Commission Seventh Framework Program European Genetic Network (EUROGEN) grant (201483, to Dr. Drobniewski), and the National Institute of Allergy and Infectious Diseases, National Institutes of Health (to Dr. Rodwell). Dr. T. Walker is an NIHR Academic Clinical Lecturer, and Drs. Crook, Peto, and Caulfield are NIHR Senior Investigators.

No potential conflict of interest relevant to this article was reported.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank Stéphanie Duthoy, Carina Hahn, Alamdar Hussain, Yannick Laurent, Mathilde Mairey, Vanessa Mohr, and Mahmood Qadir for technical assistance and George F. Gao, Director of the Chinese Center for Disease Control and Prevention, for directing the Chinese grant and sequencing program.

APPENDIX

The authors' full names (listed in alphabetical order) and academic degrees are as follows: Caroline Allix-Béguec, Ph.D., Irena Arandjelovic, Ph.D., Lijun Bi, Ph.D., Patrick Beckert, Ph.D., Maryline Bonnet, Ph.D., Phelim Bradley, D.Phil., Andrea M. Cabibbe, Ph.D., Irving Cancino-Muñoz, Ph.D., Mark J. Caulfield, F.Med.Sci., Angkana Chaiprasert, Ph.D., Daniela M. Cirillo, M.D., David A. Clifton, D.Phil., Iñaki Comas, Ph.D., Derrick W. Crook, F.R.C.Path., Maria R. De Filippo, Ph.D., Han de Neeling, Ph.D., Roland Diel, Ph.D., Francis A. Drobniewski, Ph.D., Kiatichai Faksri, Ph.D., Maha R. Farhat, M.D., Joy Fleming, Ph.D., Philip Fowler, Ph.D., Tom A. Fowler, Ph.D., Qian Gao, Ph.D., Jennifer Gardy, Ph.D., Deborah Gascoyne-Binzi, Ph.D., Ana-Luiza Gibertoni-Cruz, M.R.C.P., Ana Gil-Brusola, Ph.D., Tanya Golubchik, Ph.D., Ximena Gonzalo, M.D., Louis Grandjean, Ph.D., Guangxue He, Ph.D., Jennifer L. Guthrie, M.Sc., Sarah Hoosdally, Ph.D., Martin Hunt, Ph.D., Zamin Iqbal, D.Phil., Nazir Ismail, F.C.Path., James Johnston, M.D., Faisal M. Khanzada, M.Phil., Chiea C. Khor, D.Phil., Thomas A. Kohl, Ph.D., Clare Kong, Ph.D., Sam Lipworth, M.B., B.S., Qingyun Liu, Ph.D., Gugu Maphalala, Ph.D., Elena Martinez, Ph.D., Vanessa Mathys, Ph.D., Matthias Merker, Ph.D., Paolo Miotto, Ph.D., Nerges Mistry, Ph.D., David A.J. Moore, F.R.C.P., Megan Murray, M.D., Stefan Niemann, Ph.D., Rick T.-H. Ong, Ph.D., Tim E.A. Peto, D.Phil., James E. Posey, Ph.D., Therdsak Prammananan, Ph.D., Alexander Pym, Ph.D., Camilla Rodrigues, M.D., Mabel Rodrigues, Ph.D., Timothy Rodwell, Ph.D., Gian M. Rossolini, M.D., Elisabeth Sánchez Padilla, M.D., Marco Schito, Ph.D., Xin Shen, Ph.D., Jay Shendure, Ph.D., Vitali Sintchenko, Ph.D., Alex Sloutsky, Ph.D., E. Grace Smith, F.R.C.Path., Matthew Snyder, Ph.D., Karine Soetaert, Ph.D., Angela M. Starks, Ph.D., Philip Supply, Ph.D., Prapat Suriyapol, Ph.D., Sabira Tahseen, M.B., B.S., Patrick Tang, Ph.D., Yik-Ying Teo, Ph.D., Thuong N.T. Thuong, Ph.D., Guy Thwaites, F.R.C.P., Enrico Tortoli, Ph.D., Shaheed V. Omar, Ph.D., Dick van Soolingen, Ph.D., A. Sarah Walker, Ph.D., Timothy M. Walker, D.Phil., Mark Wilcox, M.D., Daniel J. Wilson, D.Phil., David Wyllie, Ph.D., Yang Yang, Ph.D., Hongtai Zhang, Ph.D., Yanlin Zhao, Ph.D., and Baoli Zhu, Ph.D.

REFERENCES

1. World Health Organization. Global tuberculosis report 2017. Geneva: World Health Organization, 2017 (http://www.who.int/tb/publications/global_report/en/).
2. Shah NS, Auld SC, Brust JCM, et al. Transmission of extensively drug-resistant tuberculosis in South Africa. N Engl J Med 2017;376:243-53.
3. Boehme CC, Nicol MP, Nabeta P, et al. Feasibility, diagnostic accuracy, and effectiveness of decentralised use of the Xpert MTB/RIF test for diagnosis of tuberculosis and multidrug resistance: a multicentre implementation study. Lancet 2011;377:1495-505.
4. Sanchez-Padilla E, Merker M, Beckert P, et al. Detection of drug-resistant tuberculosis by Xpert MTB/RIF in Swaziland. N Engl J Med 2015;372:1181-2.
5. Pankhurst LJ, Del Ojo Elias C, Votintseva AA, et al. Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. Lancet Respir Med 2016;4:49-58.
6. Walker TM, Kohl TA, Omar SV, et al. Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study. Lancet Infect Dis 2015;15:1193-202.
7. High-priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting. Geneva: World Health Organization, 2014 (http://www.who.int/tb/publications/tpp_report/en/).
8. The 100,000 Genomes Project Protocol v3. London: Genomics England, 2017 (https://www.genomicsengland.co.uk/100000-genomes-project-protocol/).
9. Lunter G, Goodson M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res 2011;21:936-9.
10. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078-9.
11. Iqbal Z, Caccamo M, Turner I, Flicek P, McVean G. De novo assembly and genotyping of variants using colored de Bruijn graphs. Nat Genet 2012;44:226-32.
12. Miotto P, Tessema B, Tagliani E, et al. A standardised method for interpreting the association between mutations and phenotypic drug resistance in Mycobacterium tuberculosis. Eur Respir J 2017;50(6):pii:1701354.
13. Yadon AN, Maharaj K, Adamson JH, et al. A comprehensive characterization of

PncA polymorphisms that confer resistance to pyrazinamide. Nat Commun 2017,8:588.
14. Casali N, Nikolayevskyy V, Balabanova Y, et al. Evolution and transmission of drug-resistant tuberculosis in a Russian population. Nat Genet 2014;46:279-86.
15. Manson AL, Cohen KA, Abeel T, et al. Genomic analysis of globally diverse Mycobacterium tuberculosis strains provides insights into the emergence and spread of multidrug resistance. Nat Genet 2017;49:395-402.
16. Schön T, Miotto P, Köser CU, Viveiros M, Böttger E, Cambau E. Mycobacterium tuberculosis drug-resistance testing: challenges, recent developments and perspectives. Clin Microbiol Infect 2017;23:154-60.
17. Nikolayevskyy V, Hillemann D, Richter E, et al. External quality assessment for tuberculosis diagnosis and drug resistance in the European Union: a five year multicentre implementation study. PLoS One 2016;11(4):e0152926.
18. Rigouts L, Gumusboga M, de Rijk WB, et al. Rifampin resistance missed in automated liquid culture system for Mycobacterium tuberculosis isolates with specific rpoB mutations. J Clin Microbiol 2013;51:2641-5.
19. André E, Goeminne L, Colmant A, Beckert P, Niemann S, Delmee M. Novel rapid PCR for the detection of Ile491Phe rpoB mutation of Mycobacterium tuberculosis, a rifampicin-resistance-conferring mutation undetected by commercial assays. Clin Microbiol Infect 2017;23(4):267.e5-267.e7.
20. Sreevatsan S, Stockbauer KE, Pan X, et al. Ethambutol resistance in Mycobacterium tuberculosis: critical role of embB mutations. Antimicrob Agents Chemother 1997;41:1677-81.
21. Miotto P, Cabibbe AM, Feuerriegel S, et al. Mycobacterium tuberculosis pyrazinamide resistance determinants: a multicenter study. MBio 2014;5(5):e01819-e14.
22. Coronel J, Roper M, Mitchell S, et al. MODS accreditation process for regional reference laboratories in Peru: validation by GenoType® MTBDRplus. Int J Tuberc Lung Dis 2010;14:1475-80.