



# Whole-Genome Sequencing of *Mycobacterium tuberculosis* Provides Insight into the Evolution and Genetic Composition of Drug-Resistant Tuberculosis in Belarus

Kurt R. Wollenberg,<sup>a</sup> Christopher A. Desjardins,<sup>b</sup> Aksana Zalutskaya,<sup>c</sup> Vervara Slodovnikova,<sup>c</sup> Andrew J. Oler,<sup>a</sup> Mariam Quiñones,<sup>a</sup> Thomas Abeel,<sup>b</sup> Sinead B. Chapman,<sup>b</sup> Michael Tartakovsky,<sup>a</sup> Andrei Gabrielian,<sup>a</sup> Sven Hoffner,<sup>d</sup> Aliaksandr Skrahin,<sup>e</sup> Bruce W. Birren,<sup>b</sup> Alexander Rosenthal,<sup>a</sup> Alena Skrahina,<sup>c</sup> Ashlee M. Earl<sup>b</sup>

Office of Cyber Infrastructure & Computational Biology, National Institute of Allergy and Infectious Disease, National Institutes of Health, Bethesda, Maryland, USA<sup>a</sup>; The Broad Institute of MIT & Harvard, Cambridge, Massachusetts, USA<sup>b</sup>; Republican Scientific and Practical Centre for Pulmonology and Tuberculosis, Minsk, Belarus<sup>c</sup>; Department of Microbiology, The Public Health Agency of Sweden, Solna, Sweden<sup>d</sup>; Belarusian State Medical University, Minsk, Belarus<sup>e</sup>

**ABSTRACT** The emergence and spread of drug-resistant *Mycobacterium tuberculosis* (DR-TB) are critical global health issues. Eastern Europe has some of the highest incidences of DR-TB, particularly multidrug-resistant (MDR) and extensively drug-resistant (XDR) TB. To better understand the genetic composition and evolution of MDR- and XDR-TB in the region, we sequenced and analyzed the genomes of 138 *M. tuberculosis* isolates from 97 patients sampled between 2010 and 2013 in Minsk, Belarus. MDR and XDR-TB isolates were significantly more likely to belong to the Beijing lineage than to the Euro-American lineage, and known resistance-conferring loci accounted for the majority of phenotypic resistance to first- and second-line drugs in MDR and XDR-TB. Using a phylogenomic approach, we estimated that the majority of MDR-TB was due to the recent transmission of already-resistant *M. tuberculosis* strains rather than repeated *de novo* evolution of resistance within patients, while XDR-TB was acquired through both routes. Longitudinal sampling of *M. tuberculosis* from 34 patients with treatment failure showed that most strains persisted genetically unchanged during treatment or acquired resistance to fluoroquinolones. HIV+ patients were significantly more likely to have multiple infections over time than HIV- patients, highlighting a specific need for careful infection control in these patients. These data provide a better understanding of the genomic composition, transmission, and evolution of MDR- and XDR-TB in Belarus and will enable improved diagnostics, treatment protocols, and prognostic decision-making.

**KEYWORDS** *Mycobacterium tuberculosis*, multidrug resistant, extensively drug resistant, Belarus

Despite recent reductions in tuberculosis (TB) incidence and mortality, the emergence of drug-resistant *Mycobacterium tuberculosis* strains, responsible for 480,000 cases of multidrug-resistant (MDR) TB disease in 2014, is a severe threat to individual patients and to effective TB control (1). MDR-TB is defined as TB resistant to at least the two major first-line drugs, rifampin and isoniazid. Extensively drug-resistant (XDR)-TB is defined as MDR-TB with additional resistances to second-line fluoroquinolones and

**Received** 18 October 2016 **Returned for modification** 4 November 2016 **Accepted** 18 November 2016

**Accepted manuscript posted online** 30 November 2016

**Citation** Wollenberg KR, Desjardins CA, Zalutskaya A, Slodovnikova V, Oler AJ, Quiñones M, Abeel T, Chapman SB, Tartakovsky M, Gabrielian A, Hoffner S, Skrahin A, Birren BW, Rosenthal A, Skrahina A, Earl AM. 2017. Whole-genome sequencing of *Mycobacterium tuberculosis* provides insight into the evolution and genetic composition of drug-resistant tuberculosis in Belarus. *J Clin Microbiol* 55:457–469. <https://doi.org/10.1128/JCM.02116-16>.

**Editor** Karen C. Carroll, The Johns Hopkins University School of Medicine

**Copyright** © 2017 Wollenberg 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 Alena Skrahina, [alena.skrahina@gmail.com](mailto:alena.skrahina@gmail.com), or Ashlee M. Earl, [aearl@broadinstitute.org](mailto:aearl@broadinstitute.org).

K. R. Wollenberg and C. A. Desjardins contributed equally to this work. A. Skrahina and A. M. Earl contributed equally to this work.

injectable drugs (2). XDR-TB has been identified in 105 countries and represents 9.7% of MDR cases globally (2).

Eastern Europe has the highest incidence of MDR-TB worldwide (1), representing up to 20% of new cases and 60% of retreatment cases. This is especially true for countries of the former Soviet Union, such as Belarus. Despite the relatively low rate of reported TB incidence (45/100,000), the high and increasing level of MDR-TB presents a very serious challenge for TB control in Belarus. A drug resistance survey carried out in the city of Minsk in 2009 and 2010 showed the highest-ever-reported level of MDR-TB, with almost half (48%) of all smear-positive pulmonary TB patients infected with a *M. tuberculosis* strain presenting as MDR-TB (3). More recently, a nationwide survey revealed that the MDR-TB problem in Belarus was not localized to Minsk but was widely spread throughout the country, with both new and previously treated patients having astonishingly high rates of MDR-TB (35.3% and 76.5%, respectively) (4). Previous reports implicated specific clones, usually belonging to the Beijing lineage (5), as a primary driver of MDR-TB cases in Eastern Europe, although a survey of DR-TB in Minsk found 31% of MDR cases were caused by a specific clone within the T1 spoligotype (SIT 266) (3). XDR-TB is also a critical problem in Belarus; of all countries with  $\geq 10$  reported cases of XDR-TB in 2014, Belarus had the highest proportion (29%), followed by Latvia and Lithuania (2).

A recent study, which explored the distribution of resistance-related mutations at four specific loci associated with MDR-TB in isolates from seven countries on four continents, showed that the frequency and distribution of mutations detected varied considerably among the studied countries (5). Although Belarus had the highest prevalence of MDR-TB of all the countries, it also had the narrowest spectrum of rifampin and isoniazid resistance-related mutations compared to isolates from countries where MDR-TB was much less frequent, such as Honduras, Iran, Iraq, and Uganda. In these other countries, the distribution of rifampin and isoniazid resistance-related mutations was generally broader and more heterogeneous. The limited diversity of resistance-causing mutations for rifampin and isoniazid in Belarus suggested ongoing transmission of select MDR-TB strains rather than the importation or *de novo* evolution of new MDR variants.

To gain a more complete understanding of the genomic diversity, evolution, and transmission of MDR- and XDR-TB in Belarus, we conducted whole-genome sequencing of single *M. tuberculosis* isolates from 97 patients in Belarus collected between 2010 and 2013. Specifically, we selected patients representing MDR-, XDR-, and susceptible-TB cases in order to examine the genetic basis of drug resistance and the relatedness and transmission of strains among patients being treated in Belarus. We also sequenced an additional 41 isolates from 34 of these same patients who were reported to have failed treatment and who were sampled over time to investigate the evolution of *M. tuberculosis* within patients over the course of treatment.

## RESULTS

**Known resistance-conferring genotypes explained most phenotypic resistance to first-line drugs in MDR and XDR isolates.** To better understand the genetic basis of drug resistance in MDR- and XDR-TB in Belarus, we sequenced 97 isolates that each came from a unique patient. This set included 32 MDR, 40 XDR, and 25 susceptible isolates (see Table 1 for phenotypic drug-resistant profiles). We called variants for each isolate using the H37Rv genome as a reference (Materials and Methods), searched variants against a list of known resistance-conferring mutations (given in Table S1), and compared results to each strain's antimicrobial susceptibility testing (AST) profile for first-line drugs (rifampin, isoniazid, ethambutol, and pyrazinamide). Additionally, we identified all mutations present in the genes searched described above, regardless of whether they were known to confer resistance or not (Data Set S2). Mutations known to confer resistance to rifampin and isoniazid were highly predictive of resistance to these two drugs (Table 2). While 97% of isoniazid resistance could be explained by a

**TABLE 1** Antibiotic resistance phenotype counts for each of the drugs tested

Drug	No. of phenotypically resistant	No. of phenotypically sensitive	No. not tested
First-line drugs			
Isoniazid	73	24	0
Rifampin	71	25	1
Ethambutol	69	28	0
Pyrazinamide	19	7	71
Second-line drugs			
Amikacin	39	58	0
Capreomycin	44	47	6
Kanamycin	51	46	0
Ethionamide	23	74	03
Ofloxacin	50	47	0
Cycloserine	33	57	7
Other drugs			
Streptomycin	75	22	0

single mutation, *katG* S315T, six different *rpoB* mutations, including the common S450L, combined to explain 99% of the rifampin resistance in this data set.

The sensitivity for predicting ethambutol resistance was slightly lower (90%), with the primary resistance-conferring mutations in *embB* being M306I, M306V, and Q497R (Table 2). Of the seven isolates for which known ethambutol resistance-conferring mutations could not explain phenotypic resistance, two isolates carried other polymorphisms in *embB* that were found only within ethambutol-resistant strains (N296H and D354A; Data Set S2). Although these mutations were not included in our list, both mutations were recently implicated in ethambutol resistance in a study from France (6). Mutations known to confer resistance to pyrazinamide had the lowest sensitivity for predicting phenotypic resistance to this drug; only 16% of resistance could be explained when *pncA* loss-of-function mutations were included. However, when all 10 nonsynonymous mutations in *pncA* present in pyrazinamide-resistant isolates were considered, the sensitivity and specificity for detecting resistance rose to 100% (Table 3 and Data Set S2). All of these mutations have been previously implicated in pyrazinamide resistance ([http://www.moleculartb.org/gb/pdf/doc/Moltb-tableau-pncA-17\\_12\\_14.pdf](http://www.moleculartb.org/gb/pdf/doc/Moltb-tableau-pncA-17_12_14.pdf)), with the exception of *pncA* L151W, which was present in only a single isolate.

While most strains had resistance-conferring mutations that matched their first-line AST profiles, isolate XTB13-299 was phenotypically MDR but was genetically susceptible based on our list of known mutations. Examination of sequencing reads for XTB13-299 revealed low-frequency polymorphisms at known resistance-conferring sites within *rpoB* (5 of 237 aligned reads contained the *rpoB* S450W variant), the *inhA* promoter (2 of 266 aligned reads contained the *inhA* c-15t variant), and *embB* (1 of 229 aligned reads contained the *embB* M306I variant). These low-frequency mutations suggest that this patient might have been infected with a mixed population of *M. tuberculosis*, with the minority strain leading to the MDR phenotype.

We also evaluated strains for mutations in *rpoA* and *rpoC*, mutant forms that are known to compensate for reductions in fitness caused by *rpoB* mutations (7, 8). We identified 11 unique single nucleotide polymorphisms (SNPs) in *rpoC* and 2 SNPs in *rpoA* present only in rifampin-resistant isolates (Data Set S2). *rpoC* I491T was the most prevalent SNP cooccurring with the *rpoB* S450L mutation and was previously predicted to compensate for the *rpoB* S450L (*Escherichia coli* S531L) mutation (8). Similarly, most other potential compensatory mutations seen here (Data Set S2) were previously reported (7–11). The exceptions were *rpoC* E518D, E757G, and A1213E and *rpoA* G31S variants, which may represent novel compensatory sites in these genes.

**Known resistance-conferring genotypes explained most phenotypic resistance to second-line fluoroquinolones and injectables in MDR and XDR isolates.** Second-

**TABLE 2** Diagnostic potential of mutations at known resistance-conferring loci<sup>a</sup>

Drug	Gene	Mutation	Frequency	PPV	Sensitivity	Specificity	
Rifampin	Any		70	1	0.99	1	
	<i>rpoB</i>	D435Y <sup>b,c</sup>	2	1	0.03	1	
		H445D/L/Y <sup>b,c</sup>	19	1	0.27	1	
		S450L/W <sup>b,c</sup>	49	1	0.69	1	
Isoniazid	Any		72	0.99	0.97	0.96	
	<i>katG</i>	S315T <sup>c</sup>	71	1	0.97	1	
	<i>inhA</i>	c-15t <sup>c</sup>	26	0.96	0.34	0.96	
		t-8a/c <sup>c</sup>	26	1	0.36	1	
Ofloxacin	Any		43	1	0.91	1	
	<i>gyrA</i>	A90V <sup>c</sup>	14	1	0.28	1	
		S91P <sup>c</sup>	4	1	0.08	1	
		D94A/G/H/N/Y <sup>c</sup>	25	1	0.50	1	
	<i>gyrB</i>	D461N	2	1	0.04	1	
N499T		1	1	0.02	1		
Kanamycin	Any		40	0.90	0.73	0.91	
	<i>rrs</i>	a1400g <sup>c</sup>	31	1	0.61	1	
		c1483t <sup>c</sup>	1	1	0.02	1	
	<i>eis</i>	g-10a	2	0	0	0.96	
c-14t		6	0.67	0.08	0.96		
Amikacin	Any		32	1	0.86	1	
	<i>rrs</i>	a1400g <sup>c</sup>	31	1	0.79	1	
		c1483t <sup>c</sup>	1	1	0.03	1	
Capreomycin	Any		35	0.97	0.83	0.98	
	<i>rrs</i>	a1400g <sup>c</sup>	31	1	0.72	1	
		c1483t <sup>c</sup>	1	1	0.02	1	
<i>tlyA</i>	LOF	3	0.67	0.05	0.98		
Streptomycin	Any		70	1	0.93	1	
	<i>rpSL</i>	K43R	41	1	0.55	1	
		K88R	3	1	0.04	1	
	<i>rrs</i>	a513c	19	1	0.25	1	
		a516t	6	1	0.08	1	
<i>gidB</i>	LOF	2	1	0.03	1		
Pyrazinamide	Any		3	1	0.16	1	
	<i>pncA</i>	LOF	3	1	0.16	1	
Ethambutol	Any		63	0.98	0.90	0.96	
	<i>embB</i>	M306I/V <sup>c</sup>	38	0.97	0.53	0.96	
		G406D	1	1	0.01	1	
		Q497R	24	1	0.34	1	
Ethionamide	Any		65	0.35	1	0.42	
	<i>ethA</i>	LOF	47	0.36	0.74	0.59	
		<i>inhA</i>	c-15t <sup>c</sup>	26	0.50	0.57	0.82
			t-8a/c <sup>c</sup>	26	0.23	0.26	0.73

<sup>a</sup>Isolates were included only if they had both phenotypic and genotypic resistance predictions. For each mutation, we show frequency of the mutation in the data set, positive predictive value (PPV), sensitivity, and specificity for predicting phenotypic resistance. Mutations are listed as either specific changes or as any loss-of-function (LOF) mutation, including nonsense mutations and frameshifts.

<sup>b</sup>Mutation detectable by GeneXpert MTB/RIF.

<sup>c</sup>Mutation detectable by MTBDR<sub>plus</sub> version 2.0 or MTBDR<sub>sl</sub> version 1.0.

line treatment of TB for the patient cohort included an aminoglycoside/cyclic peptide, a fluoroquinolone, ethionamide, cycloserine, and *para*-aminosalicylic acid. Of all second-line drugs, fluoroquinolone (ofloxacin) resistance was most often (91%; Table 2) successfully predicted by known resistance-conferring mutations from our list. Seven different resistance-conferring polymorphisms were identified in *gyrA*, and two were identified in *gyrB*, the most common being *gyrA* A90V, D94A, and D94G. Known resistance mutations had a lower sensitivity for predicting resistance to injectables

**TABLE 3** Nonsynonymous mutations in *pncA* and *alr* improve diagnosis of resistance to pyrazinamide and cycloserine, respectively<sup>a</sup>

Drug	Gene	Mutation	Frequency	PPV	Sensitivity	Specificity
Pyrazinamide	Any		19	1	1	1
	<i>pncA</i>	LOF	3	1	0.16	1
		D8N	1	1	0.05	1
		Q10R	1	1	0.05	1
		D12A	1	1	0.05	1
		V21A	1	1	0.05	1
		D49G/Y	5	1	0.26	1
		H57D	1	1	0.05	1
		W68G/R	5	1	0.26	1
		L151W	1	1	0.05	1
Cycloserine	<i>alr</i>		15	0.73	0.33	0.93
		L89R	15	0.73	0.33	0.93

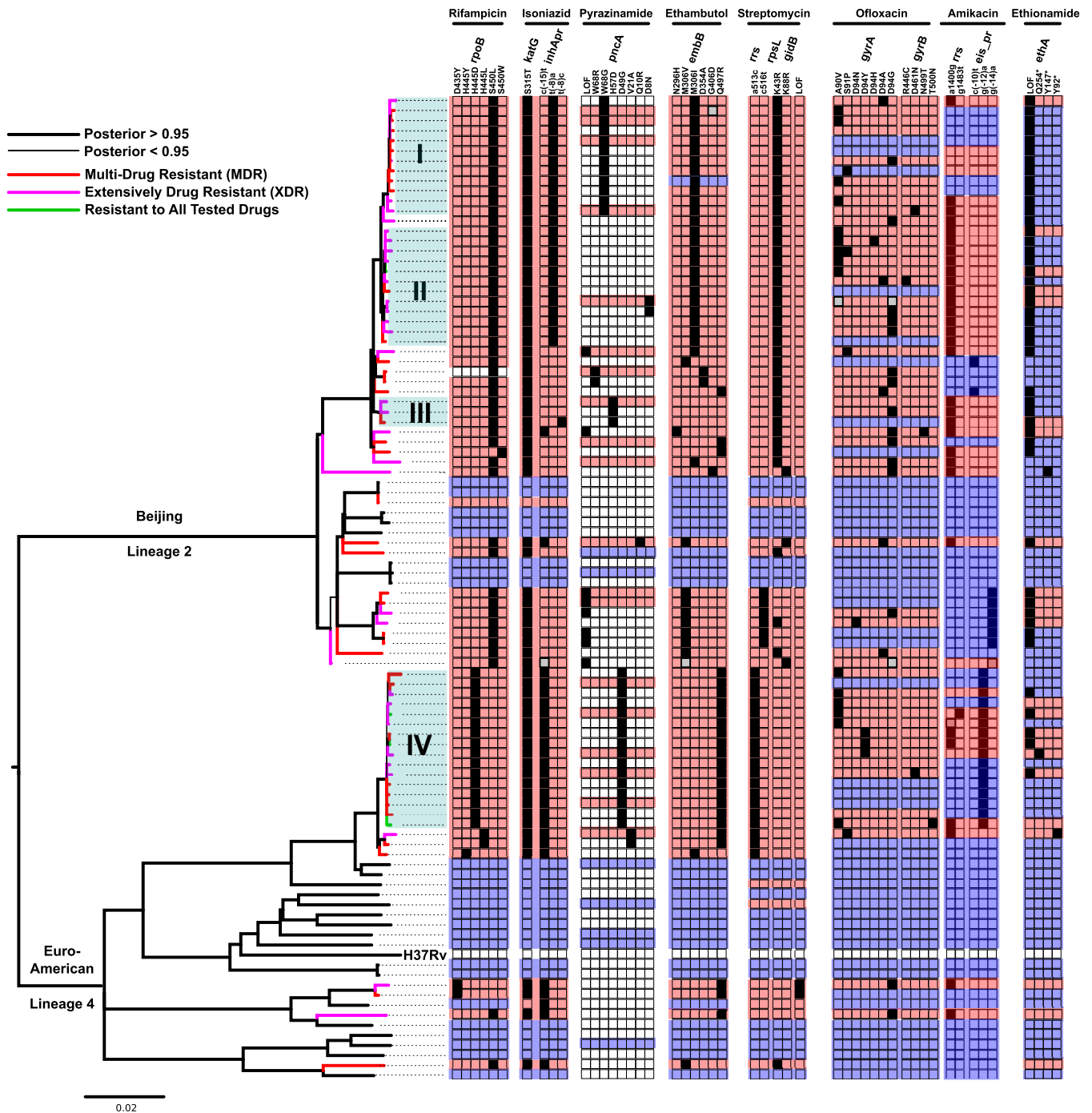
<sup>a</sup>Isolates were included only if they had both phenotypic and genotypic resistance predictions. For each mutation, we show the frequency of the mutation in the data set, positive predictive value (PPV), sensitivity, and specificity for predicting phenotypic resistance. Mutations are either listed as specific changes or as any loss-of-function (LOF) mutation, including nonsense mutations and frameshifts.

(aminoglycosides and cyclic peptides), ranging from 73% for kanamycin to 83% for capreomycin and 86% for amikacin. The most common mutation was *rrs* a1400g, which provides cross-resistance to all three aforementioned aminoglycosides and cyclic peptides. Some strains also had kanamycin-specific (*eis* promoter) or capreomycin-specific (*tlyA* loss of function) mutations.

Ethionamide resistance was poorly predicted in this data set; while the sensitivity was 100%, the specificity was only 42%. Loss-of-function mutations in *ethA* had a particularly low specificity of 59%. This effect was not driven by any single mutation but rather poor specificity across a wide range of loss-of-function mutations. No cycloserine resistance was initially predicted, as the only mutation on our list, loss of function of *ald*, was absent in all samples. However, 15 cycloserine-resistant strains harbored an *alr* L89R, a mutation that was previously identified in cycloserine-resistant clinical isolates from patients treated in South Africa (12). Although not yet experimentally validated, the addition of this mutation to our list increased sensitivity in this data set to 33%, with 100% specificity (Table 3).

**Phylogenetic analysis suggests uncontrolled transmission of MDR-TB in Belarus.** To better understand whether the spread of MDR- and XDR-TB in Belarus is driven by person-to-person transmission or *de novo* evolution of resistance within individual patients, we estimated a Bayesian phylogeny from SNPs predicted in the 97 isolates. The phylogeny revealed that all isolates fell into either the Beijing lineage (lineage 2) or the Euro-American lineage (lineage 4; Fig. 1). MDR isolates were significantly more likely to belong to the Beijing lineage than the Euro-American lineage ( $P < 0.03$ , Fisher's exact test), although XDR isolates did not show such a bias ( $P < 0.68$ ; Fisher's exact test). The phylogenetic analysis also demonstrated that many MDR and XDR isolates clustered into four shallow well-supported (posterior probability  $>0.95$ ) clades containing closely related strains (mean pairwise distance, 11 to 15 SNPs; Fig. 1, clades I, II, III, and IV). This close similarity among drug-resistant *M. tuberculosis* strains was comparable to the level of variation recently observed in a single patient in which drug resistance evolved over the course of treatment (13), suggesting an important role for the transmission of MDR- and XDR-TB among patients included in this study.

To quantify the relative importance of person-to-person transmission and *de novo* evolution of resistance within individual patients for each drug, we used parsimony and the data presented in Fig. 1 to estimate the number of times genotypic resistance evolved in isolates with AST results. For each drug, we divided this estimate by the total number of phenotypically resistant isolates to determine the fraction of resistance acquired through *de novo* evolution (Table 4). Our analysis revealed that most of the resistance to isoniazid and rifampin was acquired by patients through transmission



**FIG 1** Phylogeny and drug resistance diversity among *M. tuberculosis* isolated from patients in Belarus from 2010 to 2013. The phylogenetic tree was created using MrBayes and rooted between the Beijing and Euro-American lineages. Thicker branches indicate a posterior probability support of >0.95, and strains phenotyped as MDR, XDR, or resistant to all tested drugs are indicated by color at the terminal branches of the tree. Four closely related clades of MDR and XDR isolates are labeled (I, II, III, and IV). The results of computational spoligotyping (see Materials and Methods) are also shown along representative branches and nodes. A heatmap is shown that illustrates, for each strain, the presence of known drug resistance-conferring variants detected in isolate genomes. Clear cells in this heatmap indicate that no known drug resistance variant was detected; black cells indicate that a high-confidence drug resistance variant was detected, and gray cells indicate that there was ambiguity in the variant call made at those positions. A red background indicates the isolate was phenotypically resistant, a blue background indicates the isolate was phenotypically sensitive, and a white background indicates the isolate did not undergo AST for that drug.

(90% and 87%, respectively). This was not the case for second-line drugs. While we predicted that the majority (74%) of amikacin-resistant TB was transmitted to these patients, the majority (54%) of ofloxacin resistance was predicted to have arisen *de novo* within this population, suggesting a role for both transmission and intrahost evolution in the rising incidence of XDR-TB in Belarus.

**TABLE 4** Fraction of drug resistance acquired through person-to-person transmission versus *de novo* evolution of resistance within individual patients for each drug<sup>a</sup>

Drug	No. of acquisitions	Estimated % <i>de novo</i>	Estimated % transmission
First-line drugs			
Isoniazid	7	10	90
Rifampin	9	13	87
Ethambutol	15	22	78
Pyrazinamide	9	47	53
Second-line drugs			
Amikacin	10	26	74
Ethionamide	9	39	61
Ofloxacin	27	54	46
Other drugs			
Streptomycin	10	13	87

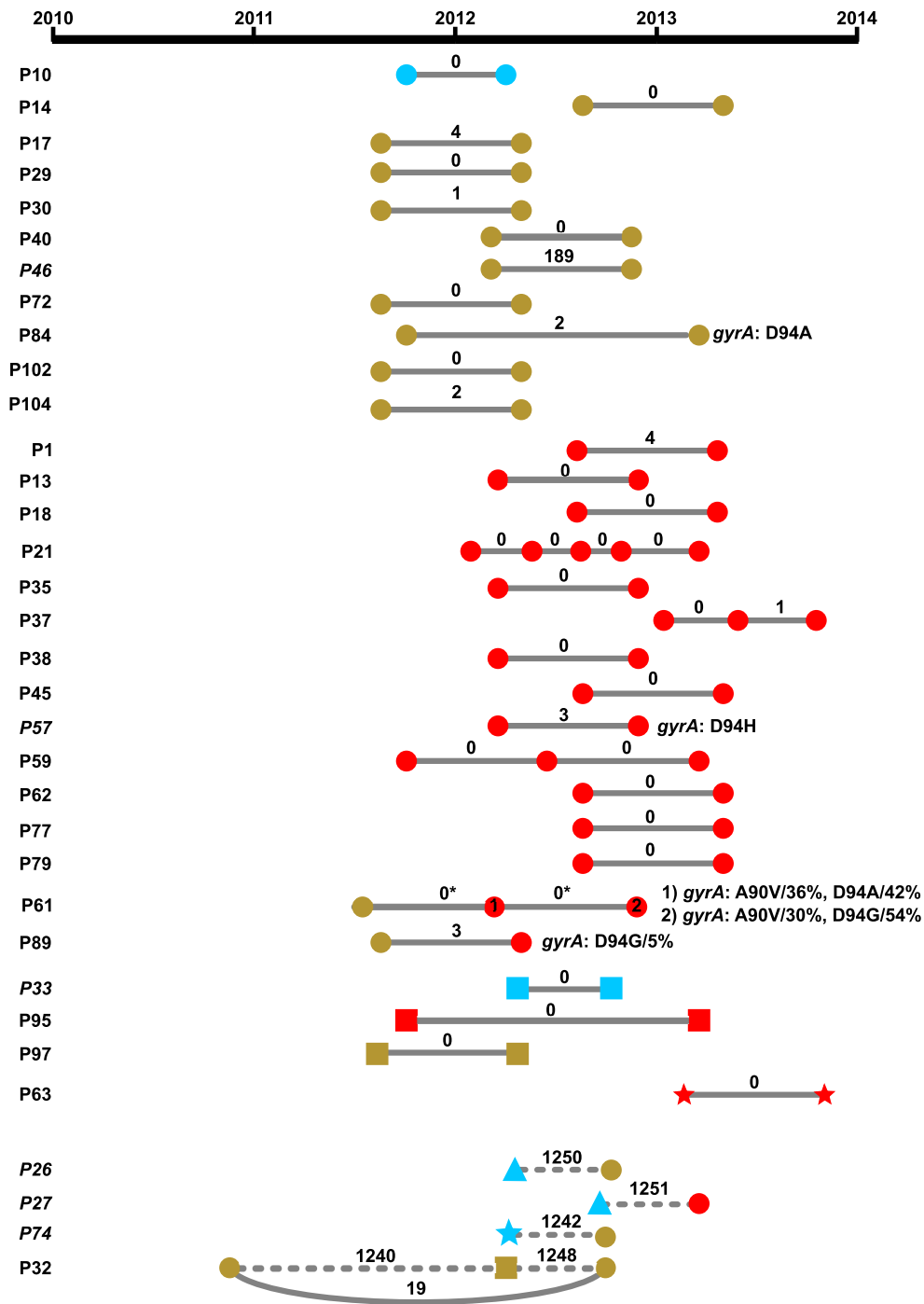
<sup>a</sup>Parsimony was used to estimate the number of times genotypic resistance was gained in isolates with AST results. This number was divided by the total number of phenotypically resistant isolates to estimate the fraction of resistance to each drug acquired through *de novo* evolution, while the remaining fraction was assigned to acquisition from person-to-person transmission.

While streptomycin was not given to patients as either a first- or second-line therapy during the course of treatment for this study, all MDR and XDR isolates harbored resistance to streptomycin through a number of known resistance-conferring mutations in *rpsL*, *rrs*, and *gidB* (Table 2). While we estimated that most streptomycin resistance was acquired through person-to-person transmission (Table 4), parsimony still predicted 10 independent instances of streptomycin resistance, suggesting that strains acquired this resistance through exposure to streptomycin through treatment prior to this study. As streptomycin was widely used for the treatment of TB in Belarus up to 2012, exposure of circulating *M. tuberculosis* strains to this drug could have frequently occurred.

**Longitudinal samples highlight intrahost evolution and multiple infections.** In addition to understanding how MDR- and XDR-TB are evolving and spreading among patients within Belarus, it is also critical to understand the genomic context of a TB infection when treatment fails. We, therefore, sequenced an additional 41 isolates from the original 97 isolates among 34 patients who maintained positive sputum cultures over time, i.e., their infections persisted despite treatment. These included 29 patients with TB that persisted but did not gain additional drug resistances over time and 5 patients with TB that became more resistant over time (Fig. 2).

When we identified SNPs that differed across same-patient isolates, we observed that most patients (29 of 34) maintained strains that either remained genetically unchanged (0 SNPs) or showed a pattern consistent with intrahost evolution (1 to 4 SNPs) (Fig. 2). In three of these cases, a new phenotypic resistance was acquired, and in all of those cases, the resistance gained was to ofloxacin. However, only one case of phenotypic ofloxacin could be explained using our list of known resistance-conferring mutations (patient 84; Fig. 2). However, examination of sequencing reads revealed resistant subpopulations in the other two patients (61 and 89). In the sample from patient 89, the *gyrA* D94G mutation was detected at 5% frequency in the sequencing reads. In the sample from patient 61, known resistant alleles, *gyrA* A90V and D94A, comprised 36% and 42% of the sequencing reads, respectively, aligning across these sites. A third sample taken at least 6 months later from patient 61 showed approximately the same level of heterogeneity at these two positions (Fig. 2) indicating that ofloxacin-resistant and -susceptible subpopulations can coexist in a patient for many months. Daily fluoroquinolone was part of each MDR patient's treatment over the period when these samples were collected (Data Set S1).

Five patients showed a clear pattern of multiple infections rather than evidence for intrahost evolution of the same strain, having >150 SNP differences between isolates



Lineage/Spoligotype - 2/Beijing ● 4/T1 ■ 4/H3 ▲ 4/Unknown ★  
 Drug Resistance - Susceptible □ MDR ■ XDR ■

**FIG 2** Genetic diversity and drug resistance in serial *M. tuberculosis* isolates. Individual samples from the same patient are shown as nodes connected by lines. Spoligotypes are indicated by node shape, and same-patient samples from the same or different spoligotype are connected by solid and dashed lines, respectively. Node colors correspond to phenotypic drug resistance. Samples are ordered by spoligotype and by patient number within each spoligotype, and mixed-spoligotype samples are presented last in the figure. The number of SNPs that differ between samples is indicated on the connecting lines. SNPs involving changes that are predicted to impact drug resistance are listed; patient 57 was initially fluoroquinolone resistant yet acquired a second fluoroquinolone resistance-conferring SNP. Percentages indicate prevalence. Asterisks indicate cases where no SNPs were identified between samples, but evaluation of read data indicated the presence of low-frequency variants conferring drug resistance. HIV+ patients are indicated by italics.



sampled at different time points (patients 26, 27, 32, 46, and 74; Fig. 2). In four patients, longitudinal isolates differed by over 800 SNPs and were part of distinct lineages, while the remaining patient isolates belonged to the Beijing lineage and differed by 189 SNPs. In three of the five cases, the earlier isolates were more drug susceptible than the later isolates, which were genotypically predicted to survive the patient's first-line treatment (treatment details are given in Data Set S1). In one of the remaining two cases with three time points, an MDR strain was present at the first and third time points, while the second sample was unrelated but also MDR, suggesting this patient may have been simultaneously harboring multiple MDR strains. In the final remaining case, the patient harbored two XDR strains at different time points. Strikingly, of the five patients showing a pattern consistent with multiple infections, four patients were HIV+ (of 5 total HIV+ patients in this study), representing a significantly greater likelihood of multiple infections in HIV+ patients than in patients not infected with HIV ( $P < 0.001$ , Fisher's exact test).

## DISCUSSION

To better understand the genetic basis and transmission of MDR- and XDR-TB in Belarus, we sequenced and analyzed 97 isolates distributed across MDR, XDR, and susceptible phenotypes. Known resistance-conferring loci accounted for the majority of phenotypic resistance to first- and second-line drugs for MDR- and XDR-TB isolates in our study. Phylogenetic analysis suggested that most MDR-TB was due to the transmission of already-resistant *M. tuberculosis* strains rather than *de novo* evolution of resistance within patients, while XDR-TB was acquired through both routes. Sequencing of an additional 41 longitudinally sampled isolates from 34 patients with treatment failure showed that most strains persisted genetically unchanged during treatment or acquired resistance to fluoroquinolones in progression toward an XDR status.

The vast majority of phenotypic resistance to first-line drugs (rifampin, isoniazid, and ethambutol) in MDR- and XDR-TB was well predicted by known resistance-conferring mutations. Additionally, while known resistance-conferring mutations performed well in predicting resistance to second-line fluoroquinolones in MDR- and XDR-TB, they overpredicted resistance to ethionamide and underpredicted resistance to aminoglycosides, cycloserine, and pyrazinamide. Pyrazinamide resistance detection could be improved by the incorporation of a number of rare nonsynonymous changes, which has been noted previously (14). It should also be noted that sampling focused on MDR- and XDR-TB, forms of TB that we considered the highest priority for diagnosis, treatment, and infection control, and not the relative prevalence of individual phenotypic drug resistances in Belarus. Therefore, our sensitivity and specificity results should be interpreted only for MDR- and XDR-TB.

Our phylogenetic analysis supported the hypothesis that MDR-TB in Belarus was primarily being acquired through person-to-person transmission. Additionally, the small level of genetic diversity found within a number of MDR clades in this data set was consistent with levels found across time course isolates from a single patient (13). A previous study of DR-TB detected limited variation at drug resistance-conferring loci in samples from Belarus compared to other countries and hypothesized that this was due to a greater proportion of transmitted resistant TB (as opposed to *de novo* evolution of resistance) (5). While we also conclude there is extensive transmission of drug-resistant TB in Belarus, we observed a much greater variety of resistance mutations among sequenced strains than previously reported. This is likely due to increased sample size, which allowed us to capture less-prevalent mutations. However, we acknowledge that our estimation of the relative role of transmission and *de novo* acquisition of resistance is an estimate, and it could be biased if the selected samples were not representative of all circulating strains. In particular, estimates could be biased toward transmission if large hospital outbreaks of MDR-TB were oversampled.

Unlike MDR-TB, the acquisition of XDR-TB appeared to be driven by both patient-to-patient transmission and *de novo* evolution of resistance. *De novo* evolution of resistance was particularly true for fluoroquinolones, as we observed repeated acqui-

sition of resistance in both the cross-sectional and longitudinal data sets. Strains in three patients evolved fluoroquinolone resistance over the course of treatment, two of which were present as a mixed community of susceptible and resistant isolates. Heteroresistance to fluoroquinolones was also seen in one additional patient in the 97-isolate set and has been previously noted in studies of fluoroquinolone-resistant TB (15, 16). A previous study from Russia also noted that fluoroquinolone resistance appeared to be independently acquired more frequently than resistances to other drugs (17). These patterns suggest that resistance to fluoroquinolones may have a greater fitness cost than resistance to other drugs, making it more difficult for fluoroquinolone resistance-conferring mutations to become fixed in intra- and interhost populations. Furthermore, this heteroresistance can lead to discrepancies between results obtained from AST and genome sequencing.

Analysis of time course data of cases of treatment failure revealed that most strains, despite appropriate treatment, remained stable in patients; 85% of the 34 treatment failure cases had phenotypic and genomic susceptibility to at least one drug they were being given. This included two fully drug-susceptible strains, indicating that treatment failure is not limited to drug-resistant cases. The limited *de novo* evolution of resistance in this patient population (aside from fluoroquinolone resistance acquisition) is in contrast to previous studies, where extensive intrahost evolution of drug resistance was well documented (13, 18). While we documented cases of mixed infection of both treatment-susceptible and -resistant strains, the fact that susceptible strains were not eliminated by treatment indicates possible problems with treatment adherence, absorption of the drugs in the gastrointestinal tract, or drug quality, as well as host immunosuppression.

Despite most strains persisting unchanged in patients, three patients had drug-susceptible strains isolated at earlier time points convert to unrelated drug-resistant strains at later time points. Our primary hypothesis is the patient was cured of the susceptible TB but was later reinfected with a more resistant strain. However, mixed infections could have been present and gone undetected due to limited sampling, as hypothesized for the one patient harboring two drug-resistant strains at multiple time points, and we cannot rule out the possibility that contamination or laboratory error resulted in mixed samples. Furthermore, this potential recolonization by drug-resistant strains was particularly apparent in HIV+ patients, in line with a recent study of TB in South America that showed that HIV positivity increased the likelihood of contracting tuberculosis (19). Therefore, it may be particularly important that immunocompromised TB patients limit exposure to drug-resistant cases of TB, particularly within health care settings.

Overall, these data provide insight into the genomic composition and modes of acquisition of MDR- and XDR-TB, critical health care problems in Belarus. Improved infection control will be of critical importance to stop the growing spread of MDR-TB and reinfection of particularly susceptible populations. However, these data also suggest that *de novo* evolution of XDR-TB is still frequently occurring and must be accounted for in strategies to control drug-resistant TB in Belarus.

## MATERIALS AND METHODS

**Sample acquisition and selection.** Initial samples from patients in Minsk, Belarus, from 2010 to 2013, were taken at the time of TB diagnosis and treatment initiation, and the resulting cultured isolates were stored at  $-70^{\circ}\text{C}$  at the Belarusian National Reference Laboratory (NRL) in Minsk. Follow-up samples were taken from patients who maintained a positive sputum culture over time, i.e., failed treatment, every 5 to 6 months postdiagnosis for patients with susceptible TB and 6 months to 1 year postdiagnosis for patients with MDR- or XDR-TB, although in a small number of cases, samples were taken more frequently. Sampling time points, HIV status, and drug therapy frequency and dosage are presented in Data Set S1. An initial set of 97 isolates in which each isolate came from a unique patient was selected to provide a balance among drug resistance profiles; this set included 32 MDR, 40 XDR, and 25 additional isolates, including 22 isolates susceptible to all tested drugs, 2 isolates resistant to only streptomycin, and 1 isolate resistant to isoniazid and streptomycin. For simplicity, we refer to these 25 isolates as "susceptible." From patients who maintained a positive sputum culture over time, i.e., failed treatment, an additional 41 isolates were selected, resulting in a second set of 75 isolates from longitudinally

sampled patients. In total, 138 isolates were sequenced, with 34 isolates overlapping both sets. For all longitudinally sampled patients, the sample with the latest date was used in the 97-isolate set.

**Antimicrobial susceptibility testing.** The *in vitro* antimicrobial susceptibility testing (AST) was carried out at the NRL in Minsk, which was externally quality assured by the WHO Supranational Reference Laboratory at the Public Health Agency of Sweden (PHAS) in Stockholm. AST was done with the absolute concentration method on Lowenstein-Jensen (LJ) medium with the following critical concentrations: 1 mg/liter isoniazid (INH), 40 mg/liter rifampin (RIF), 2 mg/liter ethambutol (EMB), 4 mg/liter streptomycin (STR), 30 mg/liter kanamycin (KAN), 30 mg/liter amikacin (AMK), 40 mg/liter capreomycin (CM), 4 mg/liter ofloxacin (OFX), 40 mg/liter ethionamide (ETH), 30 mg/liter cycloserine (CS), and 1 mg/liter *para*-aminosalicylic acid (PAS). Pyrazinamide (PZA) testing was performed using Bactec MGIT PZA kits (BD) at a concentration of 100  $\mu$ g/ml. AST was conducted on the initial isolate and in most cases was only done once. However, isolates that were found to have a resistant genotype during the study and a corresponding susceptible phenotype were retested for that drug and the data were corrected if the phenotypic retesting differed from the original results. Furthermore, isolates that had multiple resistance phenotypes and corresponding susceptible genotypes were retested for those drugs and the data were corrected if the phenotypic retesting differed from the original results. The AST results for each isolate are given in Data Set S1.

**DNA extraction.** Isolates were cultured on LJ medium at 37°C for 3 to 4 weeks at PHAS. To begin DNA extraction, two inoculation loops of bacteria were resuspended in 450  $\mu$ l of TE buffer (10 mM Tris-HCl, 1 mM EDTA [pH 8.0]) and heated at 80°C for 20 min. The cells were freeze-thawed twice, and 30  $\mu$ l of lysozyme (20 mg/ml) was added and incubated for 2 h at 37°C. Seventy microliters of 10% sodium dodecyl sulfate (SDS) and 5  $\mu$ l of proteinase K (10 mg/ml) were added to the lysate, vortexed, and incubated for 10 min at 65°C, followed by addition of 100  $\mu$ l of 5 mol/liter NaCl and 100  $\mu$ l of 10% *N*-cetyl-*N,N,N*-trimethyl ammonium bromide. The tubes were vortexed until the solution was white and then incubated for 10 min at 65°C. DNA was extracted by two chloroform-isoamyl alcohol (24:1 [vol/vol]) treatments and precipitated by the addition of cold isopropanol, and the pellet was redissolved in TE buffer.

**Genome sequencing, variant analysis, and computational spoligotyping.** DNA samples were submitted from PHAS to the Broad Institute (Cambridge, MA) for whole-genome sequencing. For each sample, the DNA concentration was determined, and two sequencing libraries were created: a fragment library (~180-bp inserts) and a jumping library (~3- to 5-kb insert). Libraries were sequenced on Illumina HiSeq 2000 instruments generating 101-bp paired-end reads at ~140 $\times$  coverage. Reads from both libraries were aligned against the H37Rv reference genome (GenBank accession no. CP003248.2) using BWA (20), and variants were called using Pilon version 1.5 (21). Variant effects were predicted using the vcf annotator (<https://sourceforge.net/projects/vcfannotator/>). Genomic polymorphisms known to confer drug resistance in each strain were identified using a previously published list of polymorphisms based on current diagnostics, as well as variants implicated in resistance with experimental support in the literature (Table S1) (12). In a separate analysis, we identified all variants present in genes from Table S1, regardless of whether they were known to confer resistance or not. Finally, spoligotypes were computationally predicted by statistically testing for the presence of each of the 43 unique spacer sequences used in classical spoligotyping from sequence reads (22). The resulting profiles were matched to those at SITVITWEB ([http://www.pasteur-guadeloupe.fr:8081/SITVIT\\_ONLINE/](http://www.pasteur-guadeloupe.fr:8081/SITVIT_ONLINE/)) to assign a spoligotype.

**Phylogenetic analysis.** An initial maximum clade compatibility Bayesian phylogenetic tree was constructed from the full complement of *M. tuberculosis* genomic SNPs using the program MrBayes version 3.2.4 (23). Support for individual clusters was quantified by clade posterior probabilities. To control for the influence of selection at loci involved in antibiotic resistance, the original set of genomic SNPs for these samples was filtered to remove SNPs from known antibiotic resistance loci. These loci were those identified as having a significant correlation with antibiotic resistance by two or more analytical methods in Farhat et al. (9). The remaining subset of sites from the *M. tuberculosis* genome was the resistance-neutral sequence data used for further phylogenetic analysis. A comparison of the full-genomic SNP Bayesian tree against the resistance-neutral Bayesian tree found no significant differences in the placement of well-supported clusters, so the resistance-neutral tree was used for further analysis. Evolutionary acquisitions of genotypic resistance were counted on the phylogenetic tree using the parsimony principle, and the resulting number was divided by the total number of phenotypically resistant isolates to determine the fraction of resistance acquired through *de novo* evolution.

**Accession number(s).** Short read data were submitted to NCBI under BioProject PRJNA200335.

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/JCM.02116-16>.

**TEXT S1**, PDF file, 0.08 MB.

**DATA SET S1**, XLSX file, 0.09 MB.

**DATA SET S2**, XLSX file, 0.06 MB.

## ACKNOWLEDGMENTS

We express our sincere gratitude to the patients in Belarus who provided clinical samples, without which this study would not have been possible. Additionally, we

would like to acknowledge the Broad Sequencing Platform for assistance with data acquisition.

This project has been funded in part with federal funds from the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health, Department of Health and Human Services, under contract HHSN272200900018C and grant U19AI110818 to the Broad Institute and under BCBB Support Services contract GS35F0373X to MSC, Inc., and under the U.S. Civilian Research and Development Foundation (CRDF) agreement BOB1-31120-MK-13. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

A. Skrahina and A. M. Earl designed the study; Z. Zalutskaya, V. Slodovnikova, S. B. Chapman, and A. Skrahin performed laboratory experiments and were responsible for sample handling; K. R. Wollenberg, C. A. Desjardins, A. J. Oler, M. Quiñones, and T. Abeel performed the data analysis; K. R. Wollenberg, C. A. Desjardins, A. Skrahina, and A. M. Earl wrote the manuscript; M. Tartakovsky, A. Gabrielian, A. Rosenthal, B. W. Birren, A. Skrahina, and A. M. Earl supervised the study.

## REFERENCES

- World Health Organization. 2015. Global tuberculosis report. World Health Organization, Geneva, Switzerland. <https://www.health-e.org.za/wp-content/uploads/2015/10/Global-TB-Report-2015-FINAL-2.pdf>.
- World Health Organization. 2008. Guidelines for the programmatic management of drug-resistant tuberculosis. World Health Organization, Geneva, Switzerland.
- Zalutskaya A, Wijkander M, Jureen P, Skrahina A, Hoffner S. 2013. Multidrug-resistant *Mycobacterium tuberculosis* caused by the Beijing genotype and a specific T1 genotype clone (SIT no. 266) is widely transmitted in Minsk. *Int J Mycobacteriol* 2:194–198. <https://doi.org/10.1016/j.ijmyco.2013.08.001>.
- Skrahina A, Hurevich H, Zalutskaya A, Sahalchik E, Astrauko A, Hoffner S, Rusovich V, Dadu A, de Colombani P, Dara M, van Gemert W, Zignol M. 2013. Multidrug-resistant tuberculosis in Belarus: the size of the problem and associated risk factors. *Bull World Health Organ* 91:36–45. <https://doi.org/10.2471/BLT.12.104588>.
- Rosales-Klintz S, Jureen P, Zalutskaya A, Skrahina A, Xu B, Hu Y, Pineda-Garcia L, Merza MA, Muntean I, Bwanga F, Joloba M, Hoffner SE. 2012. Drug resistance-related mutations in multidrug-resistant *Mycobacterium tuberculosis* isolates from diverse geographical regions. *Int J Mycobacteriol* 1:124–130. <https://doi.org/10.1016/j.ijmyco.2012.08.001>.
- Brossier F, Sougakoff W, Bernard C, Petrou M, Adeyema K, Pham A, Amy de la Breteque D, Vallet M, Jarlier V, Sola C, Veziris N. 2015. Molecular analysis of the *embCAB* locus and *embR* gene involved in ethambutol resistance in clinical isolates of *Mycobacterium tuberculosis* in France. *Antimicrob Agents Chemother* 59:4800–4808. <https://doi.org/10.1128/AAC.00150-15>.
- Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, Galagan J, Niemann S, Gagneux S. 2012. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat Genet* 44:106–110.
- Casali N, Nikolayevskyy V, Balabanova Y, Ignatyeva O, Kontsevaya I, Harris SR, Bentley SD, Parkhill J, Nejentsev S, Hoffner SE, Horstmann RD, Brown T, Drobniewski F. 2012. Microevolution of extensively drug-resistant tuberculosis in Russia. *Genome Res* 22:735–745. <https://doi.org/10.1101/gr.128678.111>.
- Farhat MR, Shapiro BJ, Kieser KJ, Sultana R, Jacobson KR, Victor TC, Warren RM, Streicher EM, Calver A, Sloutsky A, Kaur D, Posey JE, Plikaytis B, Oggioni MR, Gardy JL, Johnston JC, Rodrigues M, Tang PK, Kato-Maeda M, Borowsky ML, Muddukrishna B, Kreiswirth BN, Kurepina N, Galagan J, Gagneux S, Birren B, Rubin EJ, Lander ES, Sabeti PC, Murray M. 2013. Genomic analysis identifies targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis*. *Nat Genet* 45:1183–1189. <https://doi.org/10.1038/ng.2747>.
- Ali A, Hasan Z, Mc Nerney R, Mallard K, Hill-Cawthorne G, Coll F, Nair M, Pain A, Clark TG, Hasan R. 2015. Whole genome sequencing based characterization of extensively drug-resistant *Mycobacterium tuberculosis* isolates from Pakistan. *PLoS One* 10:e0117771. <https://doi.org/10.1371/journal.pone.0117771>.
- Zhang H, Li D, Zhao L, Fleming J, Lin N, Wang T, Liu Z, Li C, Galwey N, Deng J, Zhou Y, Zhu Y, Gao Y, Wang T, Wang S, Huang Y, Wang M, Zhong Q, Zhou L, Chen T, Zhou J, Yang R, Zhu G, Hang H, Zhang J, Li F, Wan K, Wang J, Zhang XE, Bi L. 2013. Genome sequencing of 161 *Mycobacterium tuberculosis* isolates from China identifies genes and intergenic regions associated with drug resistance. *Nat Genet* 45:1255–1260. <https://doi.org/10.1038/ng.2735>.
- Desjardins CA, Cohen KA, Munsamy V, Abeel T, Maharaj K, Walker BJ, Shea TP, Almeida DV, Manson AL, Salazar A, Padayatchi N, O'Donnell MR, Mlisana KP, Wortman J, Birren BW, Grosset J, Earl AM, Pym AS. 2016. Genomic and functional analyses of *Mycobacterium tuberculosis* strains implicate *ald* in D-cycloserine resistance. *Nat Genet* 48:544–551. <https://doi.org/10.1038/ng.3548>.
- Eldholm V, Norheim G, von der Lippe B, Kinander W, Dahle UR, Caugant DA, Mannsaker T, Mengshoel AT, Dyrhol-Riise AM, Balloux F. 2014. Evolution of extensively drug-resistant *Mycobacterium tuberculosis* from a susceptible ancestor in a single patient. *Genome Biol* 15:490. <https://doi.org/10.1186/s13059-014-0490-3>.
- Farhat MR, Sultana R, Iartchouk O, Bozeman S, Galagan J, Sisk P, Stolte C, Nebenzahl-Guimaraes H, Jacobson K, Sloutsky A, Kaur D, Posey J, Kreiswirth BN, Kurepina N, Rigouts L, Streicher EM, Victor TC, Warren RM, van Soolingen D, Murray M. 2016. Genetic determinants of drug resistance in *Mycobacterium tuberculosis* and their diagnostic value. *Am J Respir Crit Care Med* 194:621–630. <https://doi.org/10.1164/rccm.201510-2091OC>.
- Eilertson B, Maruri F, Blackman A, Herrera M, Samuels DC, Sterling TR. 2014. High proportion of heteroresistance in *gyrA* and *gyrB* in fluoroquinolone-resistant *Mycobacterium tuberculosis* clinical isolates. *Antimicrob Agents Chemother* 58:3270–3275. <https://doi.org/10.1128/AAC.02066-13>.
- 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*, in press. <https://doi.org/10.1128/aac.00664-16>.
- Casali N, Nikolayevskyy V, Balabanova Y, Harris SR, Ignatyeva O, Kontsevaya I, Corander J, Bryant J, Parkhill J, Nejentsev S, Horstmann RD, Brown T, Drobniewski F. 2014. Evolution and transmission of drug-resistant tuberculosis in a Russian population. *Nat Genet* 46:279–286. <https://doi.org/10.1038/ng.2878>.
- Köser CU, Bryant JM, Becq J, Torok ME, Ellington MJ, Marti-Renom MA, Carmichael AJ, Parkhill J, Smith GP, Peacock SJ. 2013. Whole-genome sequencing for rapid susceptibility testing of *M. tuberculosis*. *N Engl J Med* 369:290–292. <https://doi.org/10.1056/NEJMc1215305>.
- Eldholm V, Rieux A, Monteserin J, Lopez JM, Palmero D, Lopez B, Ritacco V, Didelot X, Balloux F. 2016. Impact of HIV co-infection on the evolution and transmission of multidrug-resistant tuberculosis. *Elife* 5:p16644. <https://doi.org/10.7554/eLife.16644>.
- Li H, Durbin R. 2010. Fast and accurate long-read alignment with

- Burrows-Wheeler transform. *Bioinformatics* 26:589–595. <https://doi.org/10.1093/bioinformatics/btp698>.
21. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
  22. Cohen KA, Abeel T, Manson McGuire A, Desjardins CA, Munsamy V, Shea TP, Walker BJ, Bantubani N, Almeida DV, Alvarado L, Chapman SB, Mvelase NR, Duffy EY, Fitzgerald MG, Govender P, Gujja S, Hamilton S, Howarth C, Larimer JD, Maharaj K, Pearson MD, Priest ME, Zeng Q, Padayatchi N, Grosset J, Young SK, Wortman J, Mlisana KP, O'Donnell MR, Birren BW, Bishai WR, Pym AS, Earl AM. 2015. Evolution of extensively drug-resistant tuberculosis over four decades: whole-genome sequencing and dating analysis of *Mycobacterium tuberculosis* isolates from KwaZulu-Natal. *PLoS Med* 12:e1001880. <https://doi.org/10.1371/journal.pmed.1001880>.
  23. Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574. <https://doi.org/10.1093/bioinformatics/btg180>.