



OPEN ACCESS

ORIGINAL ARTICLE

Extensive global movement of multidrug-resistant *M. tuberculosis* strains revealed by whole-genome analysis

Keira A Cohen,¹ Abigail L Manson,² Thomas Abeel,^{2,3} Christopher A Desjardins,² Sinead B Chapman,² Sven Hoffner,⁴ Bruce W Birren,² Ashlee M Earl²

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/thoraxjnl-2018-211616>).

¹Division of Pulmonary and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

²Broad Institute of Harvard and M.I.T, Cambridge, Massachusetts, USA

³Delft Bioinformatics Lab, Technische Universiteit Delft Faculteit Technische Natuurwetenschappen, Delft, Netherlands

⁴Department of Public Health Sciences, Karolinska Institute, Stockholm, Sweden

Correspondence to

Dr Ashlee M Earl, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; aearl@broadinstitute.org

Received 12 February 2018

Revised 28 January 2019

Accepted 25 February 2019

Published Online First

2 May 2019



► <http://dx.doi.org/10.1136/thoraxjnl-2019-213098>



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Cohen KA, Manson AL, Abeel T, et al. *Thorax* 2019;**74**:882–889.

ABSTRACT

Background While the international spread of multidrug-resistant (MDR) *Mycobacterium tuberculosis* strains is an acknowledged public health threat, a broad and more comprehensive examination of the global spread of MDR-tuberculosis (TB) using whole-genome sequencing has not yet been performed.

Methods In a global dataset of 53 10 *M. tuberculosis* whole-genome sequences isolated from five continents, we performed a phylogenetic analysis to identify and characterise clades of MDR-TB with respect to geographic dispersion.

Results Extensive international dissemination of MDR-TB was observed, with identification of 32 migrant MDR-TB clades with descendants isolated in 17 unique countries. Relatively recent movement of strains from both Beijing and non-Beijing lineages indicated successful global spread of varied genetic backgrounds. Migrant MDR-TB clade members shared relatively recent common ancestry, with a median estimate of divergence of 13–27 years. Migrant extensively drug-resistant (XDR)-TB clades were not observed, although development of XDR-TB within migratory MDR-TB clades was common.

Conclusions Application of genomic techniques to investigate global MDR migration patterns revealed extensive global spread of MDR clades between countries of varying TB burden. Further expansion of genomic studies to incorporate isolates from diverse global settings into a single analysis, as well as data sharing platforms that facilitate genomic data sharing across country lines, may allow for future epidemiological analyses to monitor for international transmission of MDR-TB. In addition, efforts to perform routine whole-genome sequencing on all newly identified *M. tuberculosis*, like in England, will serve to better our understanding of the transmission dynamics of MDR-TB globally.

INTRODUCTION

Multidrug-resistant (MDR) tuberculosis (MDR-TB) is defined when *Mycobacterium tuberculosis* isolated from a patient exhibit in vitro resistance to both isoniazid and rifampicin, two key first-line antitubercular drugs. With an estimated 480 000 new cases and approximately 250 000 deaths from MDR-TB globally in 2015,¹ MDR-TB is a major threat to global TB control efforts. In 2015, the WHO classified 30 countries as high burden for MDR-TB in an effort to better focus attention on

Key messages

What is the key question?

► Does combining distinct datasets of *Mycobacterium tuberculosis* whole genome sequences from around the globe improve understanding of the international spread of multidrug-resistant tuberculosis?

What is the bottom line?

► In a global genomic analysis, extensive international spread of multidrug-resistant tuberculosis was observed, including relatively recent movement of strains from varied genetic backgrounds between high-burden and low-incidence countries.

Why read on?

► In nearly all examples of multidrug-resistant (MDR) clades that had dispersed internationally, strains from different countries were isolated and sequenced as part of separate investigations, highlighting the power of combining data across diverse studies and the need for international cooperation in order to contextualise the global MDR-tuberculosis epidemic and halt the spread of drug-resistant strains.

the MDR-TB crisis; these countries account for 89% of MDR-TB cases worldwide.²

The global epidemic of MDR-TB is due to both de novo acquisition of resistance during treatment and person-to-person transmission. Globalisation has facilitated movement of people, and consequently movement of drug-resistant strains around the world. Prior investigations have documented transmission of MDR-TB within countries;^{3–7} however, international dissemination of MDR-TB strains has not been well studied. Few studies in low-burden settings have investigated whether or not MDR-TB was ‘imported’, meaning that it was acquired before immigration to a low-burden country.^{8,9} A survey of variable number tandem repeat (patterns) molecular fingerprints by the European Centre for Disease Prevention and Control concluded that international transmission of MDR-resistant and extensively drug-resistant (XDR)-TB plays an important role in TB incidence in Europe,¹⁰ especially import



of Beijing isolates from Eastern European countries.^{10 11} Other studies have linked MDR-TB and XDR-TB isolates from the U.K. to strains circulating in Russia, Eastern Europe or South Africa^{5 7 9 12} as well as linking MDR-TB strains in Thailand and the USA.¹³

Several of these existing studies have looked for international links in a targeted manner.¹¹⁻¹³ To date, more comprehensive examinations of the global spread of MDR-TB using whole-genome sequencing have not been performed. In Manson *et al* (2017),¹⁴ we constructed a global whole genome dataset of 5310 diverse *M. tuberculosis* isolates from patients in 48 countries on five continents. Comparative analysis of these geographically distinct isolates supported the existence and emergence of MDR-TB and XDR-TB throughout the globe and provided a unique opportunity to investigate evolutionary relationships among drug-resistant strains on a global scale. Here, we identify migratory clades of MDR-TB, where MDR strains have disseminated globally and estimate the timing of their geographic divergence.

METHODS

We used the global dataset of 5310 strains from Manson *et al* (2017), together with their database of drug resistance mutations and parsimony analysis to determine phylogenetic nodes at which genotypic drug resistance evolved.¹⁴ As described previously, sequence reads were mapped onto *M. tuberculosis* reference strain H37Rv (GenBank accession number CP003248.2) using BWA V.0.7.10¹⁵ and variant calls were performed using Pilon V.1.11.¹⁶ For phylogenetic reconstruction, all sites with unambiguous single nucleotide polymorphisms (SNPs) in at least one strain were combined into a concatenated alignment containing 231 898 polymorphic sites. Ambiguous positions were treated as missing data. The concatenated alignment was then used to generate a midpoint rooted phylogenetic tree using FastTree.¹⁷ Genotypic MDR was defined by identification of resistance mutations to both isoniazid and rifampicin (online supplementary appendix table 1). Diversity among strains was assessed using pairwise SNP differences (online supplementary appendix figure 1).

Clades of MDR-TB were defined as groups of isolates sharing a single shared ancestral evolution of MDR-level resistance that resulted in at least two descendant strains with genotypic MDR. We excluded 8 of 40 (20%) MDR clades with descendants from multiple countries due to their having low bootstrap support that fell below our threshold for inclusion of 0.8. For each high confidence node (bootstrap value ≥ 0.8) at which genotypic MDR evolved within our global phylogenetic tree, we examined the geographic origins of the descendent strains with respect to country of origin and UN Geoscheme designation.¹⁸ Countries were classified as high burden for MDR-TB if they were designated as such by the WHO in the 2016 Global Tuberculosis Report.¹

As done previously,¹⁹ each strain's 'digital' spoligotype was predicted by determining the presence or absence of 43 unique spacer sequences among the sequence read data.

We defined 'migrant' MDR clades as MDR clades in which descendant strains were identified in two or more countries. Migrant clades were further characterised with respect to number of descendants as well as characteristics of descendant strains including spoligotype, country and UN geographic region of isolation, genotypic XDR status and study source. Within migrant MDR clades, we determined the number of SNPs that separated the two most closely related isolates from different

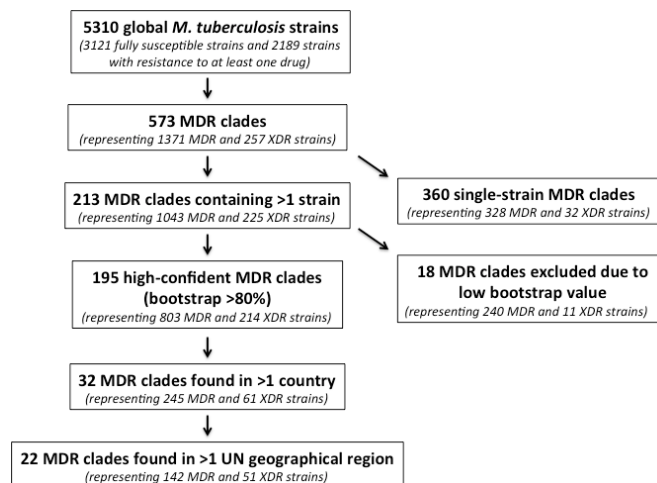


Figure 1 Flow diagram of included study strains, numbers of MDR evolutions, numbers of MDR clades and MDR migrant clades. MDR, multidrug-resistant; XDR, extensively drug-resistant.

countries as a proxy for when the isolates diverged geographically. We estimated divergence times using estimated mutation rates of 0.3 and 0.6 SNPs/genome/year.¹⁹

RESULTS

In an analysis of 5310 clinical *M. tuberculosis* isolates obtained from patients in 48 countries, we identified 573 independent evolutions of MDR-TB.¹⁴ Of these 573 independent evolutions of MDR-TB, 18 were excluded from further analysis due to our inability to confidently evaluate phylogenetic structure within that part of the tree (Methods), and 360 were associated with a single MDR isolate. This left us with 195 MDR clades with high-confidence nodes throughout the phylogeny (represented by a total of 1017 member strains) (figure 1), suggesting that MDR-TB had been transmitted between at least two individuals hundreds of times.

We observed descendants of MDR evolution in 24 countries. Seventeen of these countries, in Africa, Asia and Europe, harboured descendants of MDR clades that were isolated in two or more countries, which we define as globally 'migrant' MDR clades. Sixteen per cent (32 of 195) of MDR clades, containing a total of 306 isolates, were classified as migrant, indicating repeated global spread of MDR-TB (figure 2; table 1A and B); the remaining 163 MDR clades (84%) were confined to a single country (online supplementary appendix table 2). While we did not observe any examples of internationally migrant XDR clades in this dataset (online supplementary appendix figure 2), de novo emergence of XDR was frequently observed within migrant MDR clades. More than 50% (17 of 32) of migrant MDR-TB clades contained at least one XDR strain. In total, 61 XDR strains were identified within migrant MDR clades.

Nearly 70% (22 of 32) of migrant MDR clades showed evidence of movement between UN geographic regions (figure 1, table 1A, and online supplementary appendix figure 2) and over 80% (26 of 32) showed evidence of movement between non-adjacent countries, which accounts for migration between non-contiguous countries within the same UN geographic region. While we did not impute the geographic directionality of strain movement, most (73%; 16 of 22) migrant MDR clades having members moving between UN geographic regions indicated migration between countries designated by the WHO as high burden and non-high burden for MDR-TB (table 1A),

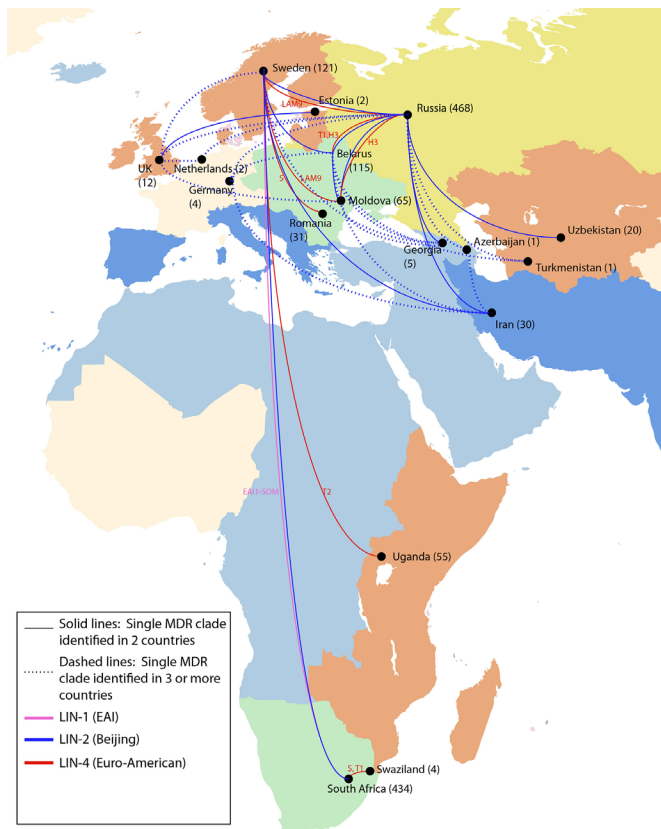


Figure 2 International spread of MDR *Mycobacterium tuberculosis* strains. In this dataset of 5310 strains, there were 32 examples in which a single evolution of MDR resulted in descendent strains isolated in more than one country, implying movement of MDR strains between countries. The numbers in parentheses indicate the total number of MDR strains in our dataset from each country, of which a subset was involved in geographic movement. The lines showing international spread are coloured by lineage. The spoligotype designation for all lineage 2 strains is Beijing; lineage 1 and 4 spoligotypes are indicated. The map is coloured by UN geographic subregion. MDR, multidrug-resistant.

likely indicating import of MDR-TB from high burden MDR-TB settings into low-incidence countries.

In every case of MDR migration, evidence for migration came from drawing together genomic data from separate studies into one¹⁴ (online supplementary appendix figure 3). For example, strains isolated from individuals in Sweden²⁰ belonged to 13 distinct MDR clades that each shared recent MDR ancestry with isolates from patients from 10 other countries, identified across 9 different studies. Thus, by combining data from distinct sequencing efforts in different areas of the globe, we were able to link isolates from patients treated in Sweden to those treated in South Africa, described in Cohen *et al.*¹⁹ in Russia, described in Casali *et al.*⁵ in Belarus, described in Wollenberg *et al.*²¹ in Moldova, Romania and Iran, described in Manson *et al.*¹⁴ and in other countries, described in Clark *et al.*²² and Merker *et al.*¹¹

Spoligotype analysis revealed that lineage 2 (Beijing) strains, which represented 28% of the strains in our global dataset and were previously documented to have extensive global spread,¹¹ represented more than half (20 of 32) of the observed instances of MDR movement between countries. However, there were also numerous examples involving lineage 4 (Euro-American) strains, which represented 53% of our global dataset—including spoligotypes H, LAM9, T1, T2 and S—and one example

involving a lineage 1 (EAI) strain, which represented 9% of our global dataset (see figure 3; table 1A and B). Lineage 4 strains were observed to disseminate among locations in Russia, Eastern Europe, Africa and Sweden, and a lineage 1 clade was observed to have shared descendants in Africa and Sweden.

For migrant MDR-TB clades, we calculated the minimum SNP distances between descendant strains isolated in different countries as a proxy for the timing of geographic divergence. For migrant MDR moving between UN geographic regions, the median SNP difference was 17 (range 4–39 SNPs), similar to that calculated for internationally migrant MDR clades (median difference 16 SNPs; range 1–116 SNPs). Assuming a mutation rate of 0.3–0.6 SNPs per genome per year, a difference of four SNPs, the smallest genetic distance between migrant MDR strains we observed, suggests close common ancestry for MDR isolates from patients in Sweden and Uganda, dating back approximately 3–7 years. The median estimate of divergence for migrant MDR-TB clade members was 13–27 years. Thus, whole-genome sequencing can provide both a geospatial and a temporal analysis of the global dissemination of drug-resistant *M. tuberculosis* strains.

DISCUSSION

In a global dataset of whole-genome sequences from 5310 clinical isolates of *M. tuberculosis*, we investigated the evolutionary relationships among MDR strains. Combining sequencing data from published studies and disparate global regions allowed us to disentangle de novo evolutions of MDR within a country from person-to-person spread of MDR into a country or region, in a way which could not be done within a single, geographically constrained study. By examining the distinct geographic sites of isolation of all descendants from a single MDR evolution, we saw evidence for 32 internationally migrant MDR clades, of which 22 had migrated between UN geographic regions.

The relationship between the global *M. tuberculosis* population structure and geographical distribution of strains has been the subject of prior investigation.^{23–29} Each of the main lineages of *M. tuberculosis* has been observed to associate with a particular region of the globe.²³ Geographic dissemination of *M. tuberculosis* has been attributed to recent political events, human migration patterns and other ecological drivers of evolution.^{28–31} Previous sublineage phylogeographic investigations have focused on the widespread dissemination of Beijing (lineage 2) strains,^{10,11} and the differential geographic restriction of certain lineage 4 sublineages.³² However, to date, few studies have linked phylogeography to global dispersal of MDR-TB or have been unable to find evidence for migration of drug-resistant strains. For example, in a recent global analysis of 1669 lineage 4 *M. tuberculosis* strains, Brynildsrud *et al.*²⁶ did not observe drug-resistant strains crossing international borders.

In addition to human migration patterns, global dissemination of *M. tuberculosis* is affected by the relative fitness of the infecting strains to be transmitted person-to-person and cause active TB disease. Primary transmission of drug-resistant TB, including MDR and XDR has now been well documented in multiple settings.^{19,33–36} However, it is less well established whether drug-resistant *M. tuberculosis* strains are more or less transmissible compared with drug susceptible strains. In a transmission network analysis conducted in Malawi, isoniazid resistance was not associated with increased risk of transmission.³⁷ And there is conflicting evidence from household contact studies that have observed both increased and decreased transmissibility of drug-resistant strains, which potentially underscores

Table 1 Movement of MDR strains around the globe

MDR-defining node	Lineage	Spoligotype	Bootstrap value	Total # MDR strains	# Strains per country (# XDR)	# Strains per UN region	# Strains per dataset	min SNP dist. bt. countries
(A) Movement of MDR strains between UN geographic regions. This table lists MDR clades with >1 UN geographic region represented.								
1	LIN-4	T2	1.00	3	Uganda 2 Sweden 1	E. Africa 2 N. Europe 1	Clark <i>et al</i> , 2013 2 TB-ARC Sweden 1	Uganda-Sweden 4
2	LIN-1	EAI1-SOM	1.00	3	South Africa 2 Sweden 1	S. Africa 2 N. Europe 1	Cohen <i>et al</i> , 2015 2 TB-ARC Sweden 1	South Africa-Sweden 5
3	LIN-2	Beijing	0.99	8	Russia 1 Uzbekistan 7	E. Europe 1 Central Asia 7	Casali <i>et al</i> , 2014 1 Merket <i>et al.</i> , 2015 7	Russia-Uzbekistan 11
4	LIN-2	Beijing	0.88	5	Russia 2 Moldova 1 Georgia 2	E. Europe 3 W. Asia 2	TB-ARC Moldova 1 Merker <i>et al.</i> , 2015 3 Casali <i>et al.</i> , 2014 1	Russia-Moldova 12 Russia-Georgia 14 Moldova-Georgia 20
5	LIN-2	Beijing	0.99	23	Russia 19 Belarus 2 (2) Sweden 1 Georgia 1	E. Europe 21 N. Europe 1 W. Asia 1	Merker <i>et al.</i> , 2015 1 Casali <i>et al.</i> , 2014 19 TB-ARC Belarus 2 TB-ARC Sweden 1	Russia-Sweden 13 Russia-Georgia 17 Russia-Belarus 20 Belarus-Georgia 26 Belarus-Sweden 29 Sweden-Georgia 32
6	LIN-2	Beijing	1.00	3	UK 1 (1) Sweden 1 Moldova 1	N. Europe 2 E. Europe 1	TB-ARC Moldova 1 Casali <i>et al.</i> , 2014 1 TB-ARC Sweden 1	UK-Moldova 14 UK-Sweden 19 Sweden-Moldova 24
7	LIN-4	LAM9	0.92	3	Russia 1 Sweden 2 (2)	E. Europe 1 N. Europe 2	Casali <i>et al</i> , 2014 1 TB-ARC Sweden 2	Russia-Sweden 14
8	LIN-2	Beijing	0.99	2	Russia 1 Iran 1 (1)	E. Europe 1 S. Asia 1	Casali <i>et al</i> , 2014 1 TB-ARC Iran 1	Russia-Iran 14
9	LIN-4	LAM9	0.99	3	Sweden 1 Moldova 2	N. Europe 1 E. Europe 2	TB-ARC Moldova 2 TB-ARC Sweden 1	Sweden-Moldova 15
10	LIN-2	Beijing	0.86	6	UK 1 (1) Russia 4 Netherlands 1	N. Europe 1 E. Europe 4 W. Europe 1	Bryant <i>et al</i> , 2013 1 Casali <i>et al</i> , 2014 5	UK-Russia 16 Russia-Netherlands 16 UK-Netherlands 17
11	LIN-2	Beijing	0.84	7	Russia 2 Iran 4 (4) Azerbaijan 1	E. Europe 2 S. Asia 4 W. Asia 1	Merker <i>et al</i> , 2015 1 Casali <i>et al</i> , 2014 2 TB-ARC Iran 4	Iran-Azerbaijan 17 Russia-Azerbaijan 24 Russia-Iran 30
12	LIN-4	T2	1.00	2	Uganda 1 Sweden 1	N. Europe 1 E. Africa 1	Clark <i>et al</i> , 2013 1 TB-ARC Sweden 1	Uganda-Sweden 17
13	LIN-4	LAM9	1.00	2	Russia 1 Sweden 1	N. Europe 1 E. Europe 1	Casali <i>et al</i> , 2014 1 TB-ARC Sweden 1	Russia-Sweden 17
14	LIN-2	Beijing	0.94	37	Russia 32 (6) Belarus 3 (2) Germany 1 Iran 1	E. Europe 35 W. Europe 1 S. Asia 1	Merker <i>et al</i> , 2015 1 Casali <i>et al</i> , 2014 32 TB-ARC Iran 1 TB-ARC Belarus 3	Russia-Iran 19 Russia-Germany 19 Iran-Germany 19 Russia-Belarus 25 Iran-Belarus 27 Belarus-Germany 30
15	LIN-2	Beijing	0.88	52	Russia 1 Belarus 50 (27) Turkmenistan 1	E. Europe 51 Central Asia 1	Merker <i>et al</i> , 2015 2 TB-ARC Belarus 50	Russia-Belarus 20 Turkmenistan-Belarus 30 Russia-Turkmenistan 34

Continued

Table 1 Continued

MDR-defining node	Lineage	Spoligotype	Bootstrap value	Total # MDR strains	# Strains per country (# XDR)	# Strains per UN region	# Strains per dataset	min SNP dist. bt. countries
16	LIN-4	S	0.97	3	Sweden 1 Romania 2	N. Europe 1 E. Europe 2	TB-ARC Romania 2 TB-ARC Sweden 1	Sweden-Romania 21
17	LIN-2	Beijing	0.82	3	Russia 2 (1) Uzbekistan 1	E. Europe 2 Central Asia 1	Merker <i>et al</i> , 2015 3	Russia-Uzbekistan 22
18	LIN-2	Beijing	0.96	5	South Africa 4 Sweden 1 (1)	S. Africa 4 N. Europe 1	Cohen <i>et al</i> , 2015 2 TB-ARC MRC 2 TB-ARC Sweden 1	South Africa-Sweden 27
19	LIN-2	Beijing	0.88	4	Sweden 3 Iran 1	N. Europe 3 S. Asia 1	TB-ARC Iran 1 TB-ARC Sweden 3	Sweden-Iran 33
20	LIN-2	Beijing	0.95	2	Russia 1 Georgia 1	E. Europe 1 W. Asia 1	Merker <i>et al</i> , 2015 1 Casali <i>et al</i> , 2014 1	Russia-Georgia 33
21	LIN-2	Beijing	0.99	10	Sweden 2 Belarus 8 (3)	N. Europe 2 E. Europe 8	TB-ARC Sweden 2 TB-ARC Belarus 8	Sweden-Belarus 36
22	LIN-2	Beijing	1.00	7	Russia 5 Sweden 2	N. Europe 2 E. Europe 5	Casali <i>et al</i> , 2014 5 TB-ARC Sweden 2	Russia-Sweden 39
(B) Movement of MDR strains between countries within a UN geographical region. This table lists MDR clades found in a single UN geographic region, but in >1 country.								
23	LIN-4	S	1.00	37	South Africa 36 Swaziland 1	S. Africa 37	Cohen <i>et al</i> , 2015 18 TB-ARC MRC 19	South Africa-Swaziland 1
24	LIN-4	T1	1.00	22	Russia 1 Belarus 21 (3)	E. Europe 22	Casali <i>et al</i> , 2014 1 TB-ARC Belarus 21	Russia-Belarus 7
25	LIN-2	Beijing	0.99	5	Russia 4 (1) Moldova 1	E. Europe 5	TB-ARC Moldova 1 Casali <i>et al</i> , 2014 4	Russia-Moldova 9
26	LIN-4	T1	0.91	19	South Africa 18 (1) Swaziland 1	S. Africa 19	Cohen <i>et al</i> , 2015 7 TB-ARC MRC 12	South Africa-Swaziland 13
27	LIN-4	H3	0.91	14	Russia 12 Moldova 2	E. Europe 14	TB-ARC Moldova 2 Casali <i>et al</i> , 2014 12	Russia-Moldova 16
28	LIN-2	Beijing	1.00	4	UK 1 (1) Estonia 3 (1)	N. Europe 4	Bryant <i>et al</i> , 2013 3 Casali <i>et al</i> , 2014 1	UK-Estonia 16
29	LIN-2	Beijing	0.91	4	Belarus 1 Moldova 3	E. Europe 4	TB-ARC Moldova 3 TB-ARC Belarus 1	Belarus-Moldova 17
30	LIN-2	Beijing	0.83	3	Russia 1 Belarus 2 (2)	E. Europe 3	Casali <i>et al</i> , 2014 1 TB-ARC Belarus 2	Russia-Belarus 22
31	LIN-2	Beijing	0.98	2	Russia 1 (1) Moldova 1	E. Europe 2	TB-ARC Moldova 1 Casali <i>et al</i> , 2014 1	Russia-Moldova 27
32	LIN-4	H3	1.00	3	Russia 2 Belarus 1	E. Europe 3	Casali <i>et al</i> , 2014 2 TB-ARC Belarus 1	Russia-Belarus 116

For each unique MDR clade in the phylogeny, defined as the set of descendants from a single node where MDR arose in our parsimony-based analysis (see Methods section), we calculated the number of countries and UN geographical regions represented. For each MDR-defining node, the table provides lineage and spoligotype designations for descendent strains, the level of statistical confidence for each node in the phylogenetic tree (figure 3), the total number of descendent strains as well as by country, UN geographic region and dataset, and the minimum SNP distance between isolates from each country pair, defined as the minimum distance between all pairs of strains from each two countries contained in a single MDR clade. WHO high burden countries are indicated in red.

MDR, multidrug-resistant; SNP, single nucleotide polymorphism; XDR, extensively drug-resistant.

differences in the relative fitness of different MDR or XDR clades. In a recent 3-year prospective cohort study in Peru, household contacts of MDR-TB patients were approximately half as likely to develop active TB disease when compared with household contacts of drug-susceptible index cases.³⁸ While these data suggest that MDR-TB may be less fit to transmit, smaller studies have shown similar progression to active TB among household contacts of MDR-TB and drug-susceptible TB

cases³⁹ and, in a larger retrospective 4-year household contact study of individuals treated for either MDR or XDR in Peru, contacts of XDR patients developed active TB at nearly twice the rate of MDR contacts.⁴⁰ Though results of studies like these can be conflicting, our results clearly show that MDR-TB is readily transmissible. It would be of interest to determine whether globally transmissible MDR clades are endowed with special features not found in less widely disseminated MDR clades. This kind of

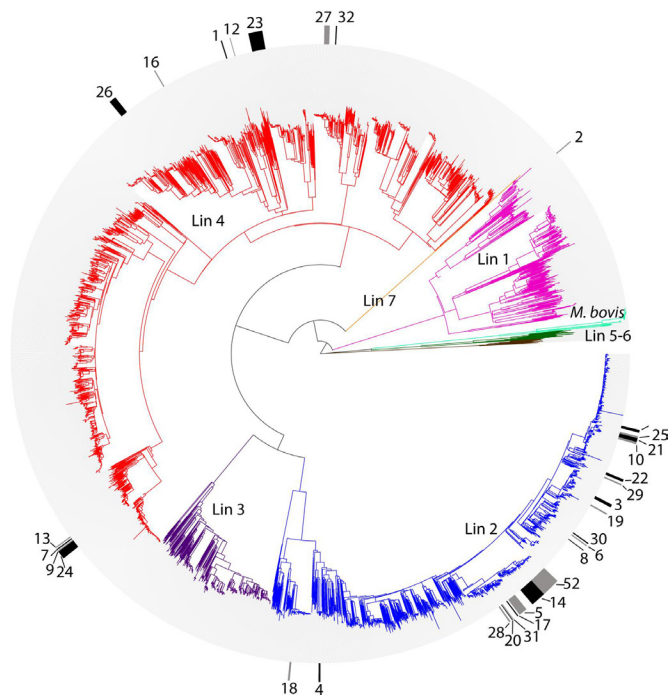


Figure 3 Phylogenetic tree of all 5310 *M. tuberculosis* isolates. To enhance visualisation of MDR clades on the phylogeny, alternating black and grey markers in the outer circle indicate strains belonging to migratory MDR clades. Clades are numbered per table 1. In the central radial phylogeny, lineages are labelled and colour-coded as follows: pink, lineage 1; blue, lineage 2; purple, lineage 3; red, lineage 4; brown, lineage 5 (*M. africanum*); dark green, lineage 6 (*M. africanum*); orange, lineage 7; light green, *M. bovis*. MDR, multidrug-resistant.

analysis would be helped by more comprehensive global databases of *M. tuberculosis* genomes.

Given that a third of the world's population is estimated to be infected with *M. tuberculosis*, undoubtedly our study, focused on only 5310 isolates, highlights only a very small portion of the international movement of MDR-TB and is skewed by larger studies. In fact, the 17 countries from which we observed migratory clade members include some of the most heavily represented countries in our global dataset, including 7 of the 10 most highly sampled countries (Russia, South Africa, UK, Netherlands, Belarus, Sweden and Uganda). Thus, it is likely that with increased sampling, we would see representatives of migratory clades in additional countries, which are currently not as well-sampled. While previous studies have focused on the global dominance of drug-resistant Beijing strains,^{10 11} we observed 12 examples of non-Beijing MDR clades that disseminated across borders. The large observed fraction of migratory strains from lineage 2 likely reflects the phylogenetic structure of our data set, which contains densely sampled lineage 2 outbreaks, providing more opportunities to observe international spread among Beijing lineages.

We did not see evidence of international movement of XDR-TB within our global dataset. This lack of observed international spread of XDR may reflect the size of our sample collection or that our study methodology was based on genotypic prediction of XDR phenotype, which is imperfect but has been recently improving for second-line drugs.⁴¹ However, it is also possible that this may be due to lower fitness or lower transmissibility of these strains. On a more optimistic note, it may be true that, at present, XDR-TB has been relatively geographically contained.

To date, we do not have evidence that XDR-TB strains would be unable to disperse globally in a similar way to MDR-TB.

A large number of migrant MDR-TB clades contained strains spanning low-burden and high-burden MDR-TB countries. It is reasonable to assume that MDR most frequently evolves de novo in high-burden countries and is 'imported' into low-incidence TB countries. There were several clades in our dataset that fit this paradigm; for example, clade 18 contained a total of five strains, four of which were isolated in South Africa and one in Sweden; and clade 22 contained a total of seven strains, five of which were isolated in Russia and two in Sweden. However, for other clades, disambiguating the most likely directionality of movement was less clear (online supplementary appendix figure 4). For example, clade 4 contained a total of five isolates, two of which were isolated in Russia, one in Moldova and two in Georgia; thus, it is not clear where MDR-level resistance emerged, where it was subsequently transmitted or in which direction-resistant strains migrated along with infected individuals. Thus, due to the complexity of MDR clades spanning multiple countries, as well as long latency periods for TB, we did not report the directionality of MDR transmission in this study. Furthermore, it is worth noting that the transmission events that resulted in the *M. tuberculosis* strain dispersal patterns depicted in figure 2 may have resulted from multiple intermediary steps—either additional geographic locations or additional human 'carriers'—that we did not capture in our dataset. Epidemiological data to confirm chains of human transmission or information regarding country of origin and travel history may make these analyses more feasible, but these data were unavailable.

A known limitation of parsimony-based analyses is that commonly occurring resistance mutations may be interpreted as having evolved once at a more ancestral position on the tree, rather than at multiple recent positions. Thus, this effect of convergent evolution could result in underestimation of the number of independent MDR clades represented in the phylogeny. However, our choice of parameters for the loss versus gain of resistance, involving a higher cost for loss events than gain events¹⁴ should minimise this effect, resulting in a conservative prediction of the total amount of MDR migration, predicting a larger number of smaller MDR clades. In agreement with this, most of the SNP distances in our dataset were remarkably small (as few as 1 SNP). One likely exception is a migrant MDR clade from our dataset that included strains from Russia and Belarus separated by 116 SNPs (table 1B). Using SNP difference as a proxy for time, this large SNP difference would place this clade as having emerged more than a hundred years ago, before there were selective antibiotic pressures for the emergence of MDR-TB, which is unlikely. Thus, in this case, it is more probable that MDR evolved convergently in multiple, more recent instances. With a denser sampling of closely related isolates on this part of the phylogeny, one would be able to disambiguate the evolutionary pattern and provide a more accurate portrayal of the geographic divergence of these strains.

Another limitation of our analysis is that, despite being the largest global dataset of *M. tuberculosis* strains compiled to date, our dataset contained strains from only 48 countries, including 2245 strains (42%) from high-burden MDR-TB countries. Additional strains from more broad geographic regions would allow us to further contextualise MDR-TB evolution and transmission globally. Similarly, only country-level geographic site of isolation was available for the majority of studied strains, which limited our ability to examine geographic transmission within national borders. In addition, if a large dataset of appropriate sample size and geography were collected in a population-based manner, it

would become possible to estimate the relative contributions of de novo versus person-to-person transmission of MDR-TB as well as to discern the directionality of transmission.

While multiple online resources such as SITVIT⁴² and SPOTCLUST⁴³ exist for interrogating geography with DNA fingerprinting data, similar platforms are in their infancy for genomic data, such as ReSeqTB (<https://platform.reseqtb.org/>).⁴⁴ However, coordinated efforts like those sponsored by Public Health England (PHE) to scale up routine whole-genome sequencing to include all prospectively identified *M. tuberculosis* clinical isolates⁴⁵ should have profound implications for TB control. Beyond drug susceptibility prediction to optimise antibiotic selection, PHE's undertaking would enable real-time epidemiology since it would include analysis of whole-genome sequencing data, combined with patient metadata including travel history. Comparing PHE data to an international strain database such as ours would enable global contextualisation of the local TB epidemic, including tracking the source of drug-resistant strains. Improved international whole genome sequence databases would inform both the scientific community and public health authorities with more up-to-date information regarding drug-resistant TB outbreaks and international movement of these strains. Inclusion of strains from parts of the globe not included in this study—such as South America, Asia and the Pacific Islands—will also be necessary to provide comprehensive global tracking.

In nearly all cases of MDR clade movement that we detected, strains from different countries were isolated and sequenced as part of separate studies, highlighting the power of combining data across diverse studies. Larger scale whole-genome sequencing studies will be needed to further elucidate global patterns of drug-resistant strain emergence and movement so that international public health efforts to contain drug-resistant TB can be optimally directed.

Contributors This study was designed and conducted by KAC, ALM and AE. Analysed the data: KAC, ALM, TA and CAD. Interpreted results: KAC, ALM. Wrote the manuscript: KAC, ALM, AE. Involved in sample acquisition and handling, including oversight of these activities: AME, BWB, SH, SBC. All authors have read the manuscript and confirm that they meet ICMJE criteria for authorship.

Funding National Institute of Allergy and Infectious Diseases Contract No: HHSN272200900018C and Grant Number U19AI110818 to ALM, TA, CAD, BWB, AE; National Heart, Blood, and Lung Institute T32HL007633 and K08HL139994 to KAC and Burroughs Wellcome Fund Career Award for Medical Scientists to KAC. This research used infrastructure resources from the Broad Institute, the Delft University of Technology, Internet2 and SURFnet (the Dutch research and education network), supported under the Enlighten Your Research Global program.

Disclaimer The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

- 1 WHO. *Global tuberculosis report 2016*. 21, 2016.
- 2 WHO. *Use of high burden country lists for TB by WHO in the post-2015 era*, 2015.
- 3 Moss AR, Alland D, Telzak E, et al. A city-wide outbreak of a multiple-drug-resistant strain of *Mycobacterium tuberculosis* in New York. *Int J Tuberc Lung Dis* 1997;1:115–21.
- 4 Rullán JV, Herrera D, Cano R, et al. Nosocomial transmission of multidrug-resistant *Mycobacterium tuberculosis* in Spain. *Emerg Infect Dis* 1996;2:125–9.
- 5 Casali N, Nikolayevskiy V, Balabanova Y, et al. Evolution and transmission of drug-resistant tuberculosis in a Russian population. *Nat Genet* 2014;46:279–86.
- 6 Zhao M, Li X, Xu P, et al. Transmission of MDR and XDR tuberculosis in Shanghai, China. *PLoS One* 2009;4:e4370.
- 7 Walker TM, Lalor MK, Broda A, et al. Assessment of *Mycobacterium tuberculosis* transmission in Oxfordshire, UK, 2007–12, with whole pathogen genome sequences: an observational study. *Lancet Respir Med* 2014;2:285–92.
- 8 Moonan PK, Teeter LD, Salcedo K, et al. Transmission of multidrug-resistant tuberculosis in the USA: a cross-sectional study. *Lancet Infect Dis* 2013;13:777–84.
- 9 Anderson LF, Tamne S, Brown T, et al. Transmission of multidrug-resistant tuberculosis in the UK: a cross-sectional molecular and epidemiological study of clustering and contact tracing. *Lancet Infect Dis* 2014;14:406–15.
- 10 De Beer JL, Kodmon C, van der Werf MJ, et al. Molecular surveillance of multi- and extensively drug-resistant tuberculosis transmission in the European Union from 2003 to 2011. *Euro Surveill* 2014;19. doi:10.2807/1560-7917.ES2014.19.11.20742.
- 11 Merker M, Blin C, Mona S, et al. Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nat Genet* 2015;47:242–9.
- 12 Williams OM, Abeel T, Casali N, et al. Fatal nosocomial MDR TB identified through routine genetic analysis and whole-genome sequencing. *Emerg Infect Dis* 2015;21:1082–4.
- 13 Coscolla M, Barry PM, Oeltmann JE, et al. Genomic Epidemiology of multidrug-resistant *Mycobacterium tuberculosis* during transcontinental spread. *J Infect Dis* 2015;212:302–10.
- 14 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.
- 15 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60.
- 16 Walker BJ, Abeel T, Shea T, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 2014;9:e112963.
- 17 Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 2009;26:1641–50.
- 18 UN. United Nations Geoscheme. Available: <http://unstats.un.org/unsd/methods/m49/m49regin.htm>
- 19 Cohen KA, Abeel T, Manson McGuire A, et al. Evolution of extensively drug-resistant tuberculosis over four decades: Whole genome sequencing and dating analysis of *Mycobacterium tuberculosis* isolates from KwaZulu-Natal. *PLoS Med* 2015;12:e1001880.
- 20 Hoffner SE. Drug-resistant *Mycobacterium tuberculosis*; some data from Sweden, Estonia and Ethiopia. *Scand J Infect Dis Suppl* 1995;98:17–18.
- 21 Wollenberg KR, Desjardins CA, Zalutskaya A, et al. Whole-genome sequencing of *Mycobacterium tuberculosis* provides insight into the evolution and genetic composition of drug-resistant tuberculosis in Belarus. *J Clin Microbiol*. In Press 2017;55.
- 22 Clark TG, Mallard K, Coll F, et al. Elucidating emergence and transmission of multidrug-resistant tuberculosis in treatment experienced patients by whole genome sequencing. *PLoS One* 2013;8:e83012.
- 23 Gagneux S, DeRiemer K, Van T, et al. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* 2006;103:2869–73.
- 24 Filliol I, Motiwala AS, Cavatore M, et al. Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. *J Bacteriol* 2006;188:759–72.
- 25 Comas I, Coscolla M, Luo T, et al. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat Genet* 2013;45:1176–82.
- 26 Brynildsrud OB, Pepperell CS, Suffys P, et al. Global expansion of *Mycobacterium tuberculosis* lineage 4 shaped by colonial migration and local adaptation. *Sci Adv* 2018;4.
- 27 Holt KE, McAdam P, Thai PVK, et al. Frequent transmission of the *Mycobacterium tuberculosis* Beijing lineage and positive selection for the EsxW Beijing variant in Vietnam. *Nat Genet* 2018;50:849–56.
- 28 Pepperell CS, Granka JM, Alexander DC, et al. Dispersal of *Mycobacterium tuberculosis* via the Canadian fur trade. *Proc Natl Acad Sci U S A* 2011;108:6526–31.
- 29 Luo T, Comas I, Luo D, et al. Southern East Asian origin and coexpansion of *Mycobacterium tuberculosis* Beijing family with Han Chinese. *Proc Natl Acad Sci U S A* 2015;112:8136–41.
- 30 Eldholm V, Pettersson JH-O, Brynildsrud OB, et al. Armed conflict and population displacement as drivers of the evolution and dispersal of *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* 2016;113:13881–6.
- 31 Liu Q, Ma A, Wei L, et al. China's tuberculosis epidemic stems from historical expansion of four strains of *Mycobacterium tuberculosis*. *Nat Ecol Evol* 2018;2:1982–92.
- 32 Stucki D, Brites D, Jeljeli L, et al. *Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sublineages. *Nat Genet* 2016;48:1535–43.
- 33 Gandhi NR, Moll A, Sturm AW, et al. Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. *Lancet* 2006;368:1575–80.

- 34 Zhao Y, Xu S, Wang L, *et al.* National survey of drug-resistant tuberculosis in China. *N Engl J Med* 2012;366:2161–70.
- 35 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.
- 36 Andrews JR, Gandhi NR, Moodley P, *et al.* Exogenous reinfection as a cause of multidrug-resistant and extensively drug-resistant tuberculosis in rural South Africa. *J Infect Dis* 2008;198:1582–9.
- 37 Guerra-Assunção JA, Crampin AC, Houben RMGJ, *et al.* Large-scale whole genome sequencing of *M. tuberculosis* provides insights into transmission in a high prevalence area. *Elife* 2015;4. doi:10.7554/eLife.05166.
- 38 Grandjean L, Gilman RH, Martin L, *et al.* Transmission of multidrug-resistant and drug-susceptible tuberculosis within households: a prospective cohort study. *PLoS Med* 2015;12:e1001843. discussion e43.
- 39 Teixeira L, Perkins MD, Johnson JL, *et al.* Infection and disease among household contacts of patients with multidrug-resistant tuberculosis. *Int J Tuberc Lung Dis* 2001;5:321–8.
- 40 Becerra MC, Appleton SC, Franke MF, *et al.* Tuberculosis burden in households of patients with multidrug-resistant and extensively drug-resistant tuberculosis: a retrospective cohort study. *Lancet* 2011;377:147–52.
- 41 Coll F, Phelan J, Hill-Cawthorne GA, *et al.* Genome-wide analysis of multi- and extensively drug-resistant *Mycobacterium tuberculosis*. *Nat Genet* 2018;50:307–16.
- 42 Demay C, Liens B, Burguière T, *et al.* SITVITWEB--a publicly available international multimer database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. *Infect Genet Evol* 2012;12:755–66.
- 43 Vitol I, Driscoll J, Kreiswirth B, *et al.* Identifying *Mycobacterium tuberculosis* complex strain families using spoligotypes. *Infect Genet Evol* 2006;6:491–504.
- 44 Farhat MR MM, Choirat C, Prasad R, *et al.* Translational genomics of tuberculosis, 2015. Available: gentb.hms.harvard.edu
- 45 Walker TM, Cruz ALG, Peto TE, *et al.* Tuberculosis is changing. *Lancet Infect Dis* 2017;17:359–61.