Transmission of multidrug-resistant *Mycobacterium tuberculosis* in Shanghai, China: a retrospective observational study using whole-genome sequencing and epidemiological investigation

Chongguang Yang, Tao Luo*, Xin Shen*, Jie Wu, Mingyu Gan, Peng Xu, Zheyuan Wu, Senlin Lin, Jiyun Tian, Qingyun Liu, ZhengAn Yuan, Jian Mei†, Kathryn DeRiemer†, Qian Gao†*

Summary

Background **Multidrug-resistance is a substantial threat to global elimination of tuberculosis. Understanding transmission patterns is crucial for control of the disease. We used a genomic and epidemiological approach to assess recent transmission of multidrug-resistant (MDR) tuberculosis and identify potential risk factors for transmission.**

Methods **We did a population-based, retrospective study of patients who tested positive for tuberculosis between Jan 1, 2009, and Dec 31, 2012, in Shanghai, China. We did variable-number-of-tandem-repeat genotyping and whole-genome sequencing of isolates. We measured strain diversity within and between genomically clustered isolates. Genomic and epidemiological data were combined to construct transmission networks.**

Findings **367 (5%) of 7982 patients with tuberculosis had MDR tuberculosis and 324 (88%) of these had isolates available for genomic analysis. 103 (32%) of the 324 MDR strains were in 38 genomic clusters that differed by 12 or fewer single nucleotide polymorphisms (SNPs), indicating recent transmission of MDR strains. Patients who had delayed diagnosis or were older than 45 years had high risk of recent transmission. 235 (73%) patients with MDR tuberculosis probably had transmission of MDR strains. Transmission network analysis showed that 33 (87%) of the 38 clusters accumulated additional drug-resistance mutations through emergence or fixation of mutations during transmission. 68 (66%) of 103 clustered MDR strains had compensatory mutations of rifampicin resistance.**

Interpretation **Recent transmission of MDR tuberculosis strains, with increasing drug-resistance, drives the MDR tuberculosis epidemic in Shanghai, China. Whole-genome sequencing can measure of the heterogeneity of drug-resistant mutations within and between hosts and help to determine the transmission patterns of MDR tuberculosis.**

Funding **National Science and Technology Major Project, National Natural Science Foundation of China, and US National Insitutes of Health.**

Lancet Infect Dis 2017; 17: 275–84

Published Online December 2, 2016 http://dx.doi.org/10.1016/S1473-3099(16)30418-2

See Comment page 238

*Contributed equally

†Contributed equally

The Key Laboratory of Medical Molecular Virology of Ministries of Education and Health, Institutes of Biomedical Sciences and Institute of Medical Microbiology, School of Basic Medical Science, Fudan University, Shanghai, China (C Yang PhD, T Luo PhD, M Gan MSc, P Xu PhD, J Tian MSc, Q Liu PhD, Q Gao PhD); **Department of Tuberculosis Control, Shanghai Municipal Centre for Disease Control and Prevention, Shanghai, China** (X Shen PhD, J Wu MPH, Z Wu MSc, S Lin MSc, ZA Yuan MSc, J Mei PhD); **School of Basic Medical Science and the West China Centre of Medical Sciences of Sichuan University, Chengdu, Sichuan, China** (T Luo); **School of Medicine, University of California, Davis, CA, USA** (K DeRiemer PhD); **and Department of Epidemiology of Microbial Diseases, School of Public Health, Yale University, New Haven, CT, USA** (C Yang)

Correspondence to: Dr Qian Gao, Key Laboratory of Medical Molecular Virology of Ministries of Education and Health, School of Basic Medical Science, Fudan University, 138 Yi Xue Yuan Road, Shanghai, China, 200032 **qiangao@fudan.edu.cn**

Introduction

The worldwide emergence of multidrug-resistant (MDR) tuberculosis threatens the global eradication of tuberculosis. An estimated 480 000 cases of MDR tuberculosis occurred worldwide in 2015, but only one in ten cases were diagnosed, treated, and cured.[1] As well as acquisition of MDR tuberculosis during treatment for tuberculosis, person-to-person transmission of MDR strains occurs and can be increased by delayed diagnosis, prolonged treatment, and unfavourable treatment outcomes in patients with MDR tuberculosis. Therefore, understanding the causes and transmission patterns of MDR tuberculosis is crucial to inform public health efforts to reduce this disease.[2]

Molecular epidemiological methods that combine epidemiological investigations and genotyping of *Mycobacterium tuberculosis* strains provide the means to assess the recent transmission (within 2–3 years) of *M tuberculosis* and risk factors for transmission.[3] However, traditional genotyping methods have limited discriminatory power and are limited in assessing homoplasy, which reduces their accuracy in identifying recent transmissions of *M tuberculosis*.[4] High-throughput whole-genome sequencing provides increased resolution and accuracy over older methods, and is a powerful tool to study the transmission of *M tuberculosis*.[4–10] Furthermore, phylogenetic networks based on whole-genome sequencing can be used to identify putative source cases, super-spreaders, and transmission directions in the absence of, or complementary to, extensive epidemiological data.[8,9] However, few studies have applied whole-genome sequencing to address recent transmission of MDR tuberculosis at the population level.[5,10]

China has a high prevalence of drug-resistant tuberculosis and the second largest number of MDR cases worldwide.[1] The high risk of incident MDR tuberculosis among patients who have previously been treated for tuberculosis has resulted in a common belief that most cases of MDR tuberculosis are due to acquired resistance, and led to the allocation of resources to improve treatment. However, many cases of MDR tuberculosis are due to

Research in context

Evidence before the study

We searched PubMed for molecular epidemiology studies of *Mycobacterium tuberculosis* that used whole-genome sequencing using the search terms "tuberculosis", "multidrug resistance", "whole genome sequencing", and "transmission". We only selected articles published in English before Oct 31, 2016. We identified 20 studies published from 2011 in which whole-genome sequencing was used to investigate transmission of *M tuberculosis* at the population level. We also searched the China Knowledge Resource Integrated Database, with the same search terms, for relevant papers published in Chinese, but did not identify any studies. In 2011, Gardy and colleagues applied whole-genome sequencing-based analysis to a large mycobacterial interspersed-repetitive-unit–variable-number tandem-repeat cluster of *M tuberculosis* and showed that whole-genome sequencing had higher resolution compared with traditional genotyping. Walker and colleagues evaluated the potential of whole-genome sequencing to delineate outbreaks of *M tuberculosis* and to measure the recent transmission of *M tuberculosis* in the UK. In a large-scale population-based study using whole-genome sequencing, Guerra-Assunção and colleagues showed that most cases of tuberculosis in a high-incidence setting in Malawi were caused by just one lineage of *M tuberculosis*. Casali and colleagues investigated the long-term evolution and endemic spread of MDR tuberculosis in a Russian population, and an additional study of an outbreak of isonazid-resistant tuberculosis in London identified a cluster of multidrug-resistant (MDR) strains, which involved rare rifampin-resistant mutations.

Added value of this study

Although several studies used whole-genome sequencing to track transmission of *M tuberculosis*, few studies used this approach to address transmission of MDR tuberculosis at the population level. To our knowledge, we describe the first population-based study combining genomics with detailed epidemiological data to identify transmission pathways of MDR tuberculosis in a region over time in China, the country with the second highest number of MDR tuberculosis cases in the world. We provide direct evidence of the emergence of diversity in *M tuberculosis* subpopulations within hosts and the fixation of specific subpopulations of *M tuberculosis* between hosts along a chain of transmission. Meanwhile, many MDR strains became more resistant during the recent transmission. Inadequate treatment was considered to be the most common way of developing MDR tuberculosis; however, our data show that most of the patients in our study population who had MDR tuberculosis had been infected with MDR strains. We also found that a majority of transmission events occurred in settings such as residential communities or complexes and related public facilities.

Implications of all the available evidence

Whole-genome sequencing provides greater precision than traditional genotyping to study recent transmission of MDR strains of *M tuberculosis*. Whole-genome sequencing is also useful to detect heterogeneity of strain diversity within and between hosts, which can be used to infer the transmission trajectory of MDR strains. Our findings suggest that strategies and interventions to halt ongoing transmission of MDR strains should be a priority for tuberculosis control programmes in China and other settings with a high burden of MDR tuberculosis.

transmission of MDR strains,[11–14] but few studies provide direct evidence of transmission of MDR tuberculosis at the population level. Thus, we hypothesised that the transmission of MDR strains has a major role in the high prevalence of MDR tuberculosis in China. We did a population-based retrospective study in Shanghai, China, using whole-genome sequencing and epidemiological investigations. We quantified the magnitude of MDR tuberculosis arising from the transmission of MDR strains, tracked transmission patterns, and identified the risk factors for transmission.

Methods

Study design and population

Shanghai is the most populous city in China, with an estimated population of 24 million, and has a reasonably well-functioning tuberculosis control programme. All patients with suspected pulmonary tuberculosis are referred to local designated hospitals, where the diagnosis is made by sputum smear and culture. Through a routine surveillance system, all cases of tuberculosis are reported to Shanghai Municipal Centre for Disease Control and Prevention (CDC).

See Online for appendix

This population-based, retrospective observational study included all patients aged 15 years or older with culture-confirmed MDR pulmonary tuberculosis who were reported by local designated hospitals in Shanghai between Jan 1, 2009, to Dec 31, 2012.

The institutional review boards of Shanghai CDC and the Institute of Biomedical Sciences, Fudan University, approved the study. Written informed consent was obtained from patients before each epidemiological investigation.

Procedures

All clinical isolates were collected at diagnosis, before starting treatment, and were sent to the Tuberculosis Reference Laboratory in Shanghai CDC for drug-susceptibility testing. We used the proportion method on Löwenstein-Jensen medium for four first-line drugs (isoniazid, rifampicin, ethambutol, and streptomycin). Pre-extensively drug-resistant and extentively drug-resistant *M tuberculosis* genotypes were defined on the basis of genomic markers associated with drug resistance (appendix p 2).[15]

Genomic DNA was extracted from cultures of single sputum specimens (one per patient) following a

previously published method.[16] Beijing strains were identified by RD105 deletion-targeted multiplex PCR-based assay and were confirmed by use of sequencing to identify specific single nucleotide polymorphisms (SNPs).[17] We did variable-number-of-tandem-repeat (VNTR) genotyping for all strains using the 9+3 loci set as described previously,[18] and then did integrated whole-genome sequencing to determine the genomic relatedness among the MDR strains that shared identical VNTR genotype patterns. Strains with cross-contamination were excluded as previously described.[16]

Genomic DNA was sequenced using Illumina Hiseq 2000 (Illumina, San Diego, CA, USA) with an expected coverage of 100 times. Paired-end reads were mapped to the reference genome H37Rv (GenBank AL123456) with Bowtie2. The SAMtools/BCFtools suite was used for calling fixed SNPs (frequency ≥95%).[19] SNPs were called at loci where the alternative basecalls were supported by at least five reads without strand bias (both strands were mapped by the reads). The fixed SNPs, excluding SNPs in the proline-glutamic acid—proline-proline-glutamic acid, proline-glutamic acid—polymorphic GC-rich sequence, and drug-resistance associated genes, were combined into a concatenated alignment (appendix p 16) that was used to construct a maximum-likelihood phylogeny using MEGA (version 5.0). The sequences were also used to generate median-joining networks for each cluster using Network (version 5.0). VarScan2 (version 2.3.6) was used to call unfixed SNPs (frequency between 5% and 95%) in genes associated with drug resistance, which represent the emergence of drug-resistant mutations within individual patients, as previously described.[20] The sequencing data were deposited in the National Center for Biotechnology Information Sequence Read Archive (appendix p 17; accession number SRP058221).

Clinical, demographic, and epidemiological records for each patient with MDR tuberculosis were obtained during routine surveillance. Diagnostic delay was defined as the time between onset of symptoms and date of diagnosis. For each patient with MDR tuberculosis that was in a genotypic cluster, we used a standardised questionnaire to obtain information about their social characteristics, history of exposure to people with active tuberculosis, and locations frequented where transmission could have occurred.

Patients with epidemiological links were defined as the following: confirmed (patients knew each other or lived at the same address or complex); probable (patients did not know each other but shared locations where transmission probably occurred, including in a neighborhood complex or street in the same district); and without links (patients did not know each other and lacked a common neighbourhood, location, or setting). Putative transmission networks were constructed on the basis of the structure of the genomic phylogeny and the epidemiological links.

Statistical analysis

We used the χ^2 test of proportions and the Wilcoxon non-parametric rank sum test to compare covariates between groups. We used univariate and multivariable logistic regression analysis to calculate the odds ratios (ORs) and 95% CIs for risk factors associated with genomic clusters. A forward, step-wise approach was used to add covariates to the multivariable model. Statistical analyses were done using Stata (version 13.1).

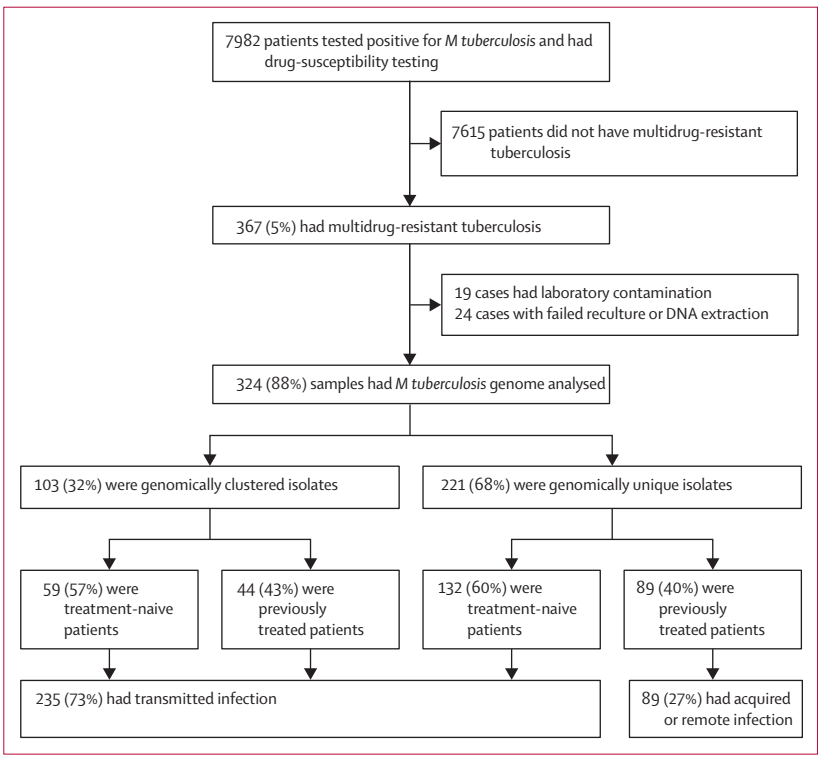

Figure 1: **Classification of multidrug-resistant tuberculosis based on treatment history and genomic analysis**

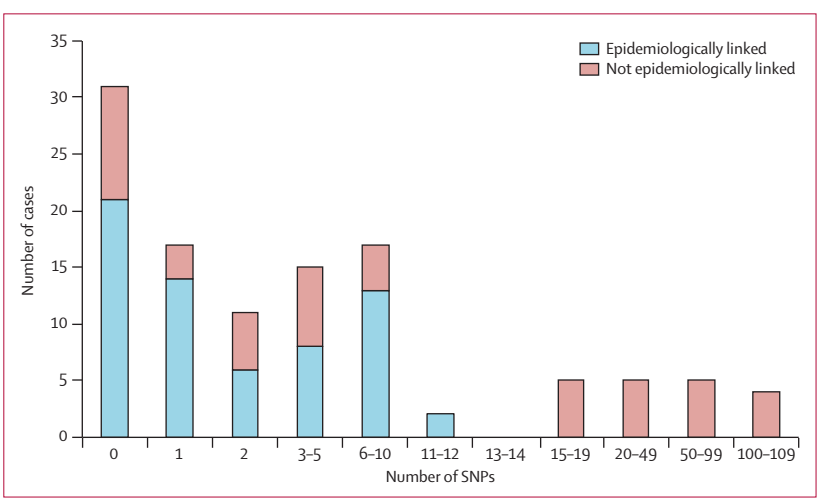

Figure 2: **Distribution of the number of SNPs in isolates from the closest multidrug-resistant tuberculosis cases within a cluster**
SNP=single nucleotide polymorphism.

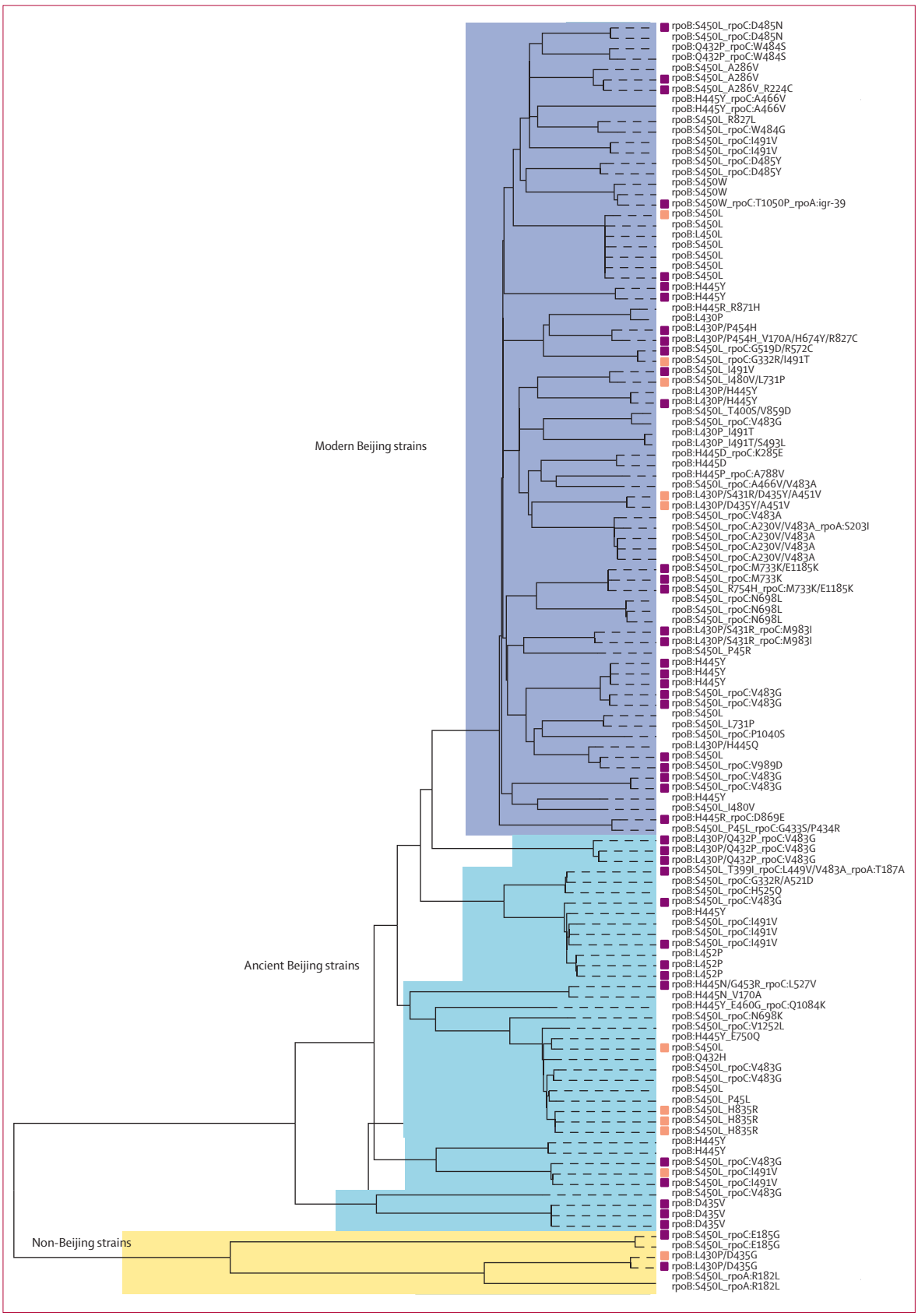

Figure 3: **Maximum-likelihood, whole-genome, single-nucleotide polymorphism-based phylogeny of multidrug-resistant isolates and mutations in RNA polymerase subunits *rpoA*, *rpoB*, and *rpoC***

Squares beside the nodes indicate pre-extensively drug-resistant (purple) and extensively drug-resistant (orange) genotypes among genomically clustered multidrug-resistant tuberculosis strains. The aminoacid change caused by mutation of *rpoA*, *rpoB*, or *rpoC* is given beside the relevant phylogeny.

Role of the funding source
The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

7982 patients tested positive for *M tuberculosis* and had their isolates sent for drug susceptibility testing during the study (figure 1). Of the 324 patients with MDR tuberculosis who had a clinical isolate suitable for analysis (figure 1; appendix p 10), 191 (59%) were treatment-naive, 238 (73%) were male, and the median age was 39 years (range 16–88).

VNTR genotyping analysis showed that 125 (39%) of 324 patients with MDR tuberculosis had isolates with genotypic patterns that were identical to at least one other patient (appendix p 10). Useable DNA was obtained and whole-genome sequencing was done in 122 (98%) of the genotypically clustered strains. After excluding SNPs in highly repetitive regions and mutations associated with drug resistance, the smallest genomic differences between any two strains that had identical genotypes ranged from 0 to 109 SNPs (figure 2). The maximum-likelihood phylogeny tree (figure 3) showed that 116 (95%) of 122 of the sequenced strains were Beijing-family strains and the predominant sublineages were so-called modern Beijing strains (78 [67%] of 116).

Epidemiological investigations were done among 112 (92%) of 122 patients with MDR tuberculosis (ten had died or were lost to follow-up). Confirmed or probable epidemiological links were identified in 64 (57%) of 112 patients (table 1). No patients with epidemiological links had strains that were separated by more than 12 SNPs (figure 2). Therefore, we defined a genomic cluster in this study as a group of strains that differed by no more than 12 SNPs. Thus, 103 (32%) of 324 strains

	Number of cases in cluster	Median age (years [IQR])	Number of patients living in the same district	Number of patients with positive sputum smear result	Number of new cases	Treatment outcome (number of patients)	Known risk factors (number of patients)	Epidemiologically linked	Nature of epidemiological link [number of patients, n=64]
Cluster 33	8	49 (29–52)	3, 4*	6	1	Cured (5), died (1), default (1), unknown (1)	Retreatment (7); game room (3)	Yes	Social (same street [2], neighbourhood street [2], and community game room [3])
Cluster 35	7	59 (47–61)	6	6	5	Cured or completed (3), died (2), failed(1), unknown (1)	Hospitalisation (5)	Yes	Nosocomial (health centre [2]), social (resident community [5])
Cluster 09†	7	53 (51–58)	7	7	5	Cured (1), died (3), default (1), unknown (2)	Drug misuse (2); game room (5)	Yes	Household [2], social (community game room [4]), drug user [1]; all lived in the same residential complex
Cluster 21	5	63 (48–65)	4	4	1	Cured (2), default (1), unknown (1), died (1)	ND	Probable	Social (neighborhood street [5])
Cluster 08	3	30 (27–53)	2	3	2	Cured (2), failed (1)	ND	Yes	Social (community, same street [2])
Cluster 22	3	37 (25–53)	3	3	3	Cured or completed (2), failed (1)	HIV infection (1)	Yes	Social (community, same street [2])
Cluster 23	3	64 (54–66)	0	3	3	Cured (1), failed (2)	ND	Unknown	Unknown
Cluster 25·2	3	37 (27–38)	3	3	2	Cured (3)	ND	Yes	Social (same street [3])
Cluster 31	3	30 (27–37)	3	3	3	Cured (1), default (1), unknown (1)	ND	Probable	Social (community, former residence place [2])
Cluster 32	3	28 (25–50)	2	3	2	Cured (2), failed (1)	ND	Yes	Social (community, known to each other [2])
Cluster 37	3	ND	2	ND	2	Cured (1), default (1), unknown (1)	ND	Yes	Social (community, residential place [2])
Cluster 38	3	61 (50–78)	3	3	3	Cured (3)	Old age (3)	Probable	Social (community, food market [3])
Other clusters‡	52	43 (27–61)	ND	43	28	Cured or completed (32), died (3), default (6), unknown (11)	ND	ND	Household [2], social (residential complex [8], community [4], neighborhood street [8])

*Four cases were from one district, and three cases were from another neighbour district. †Includes two variable-number tandem repeat clusters. ‡Patients in clusters with two cases of multidrug-resistant tuberculosis. Cured was defined as a negative sputum smear result at the end of treatment and on one other occasion among patients who had previously had a positive sputum smear result. ND=not determined.

Table 1: **Characteristics of multidrug-resistant tuberculosis genomic clusters based on whole-genome sequencing analysis (n=103)**

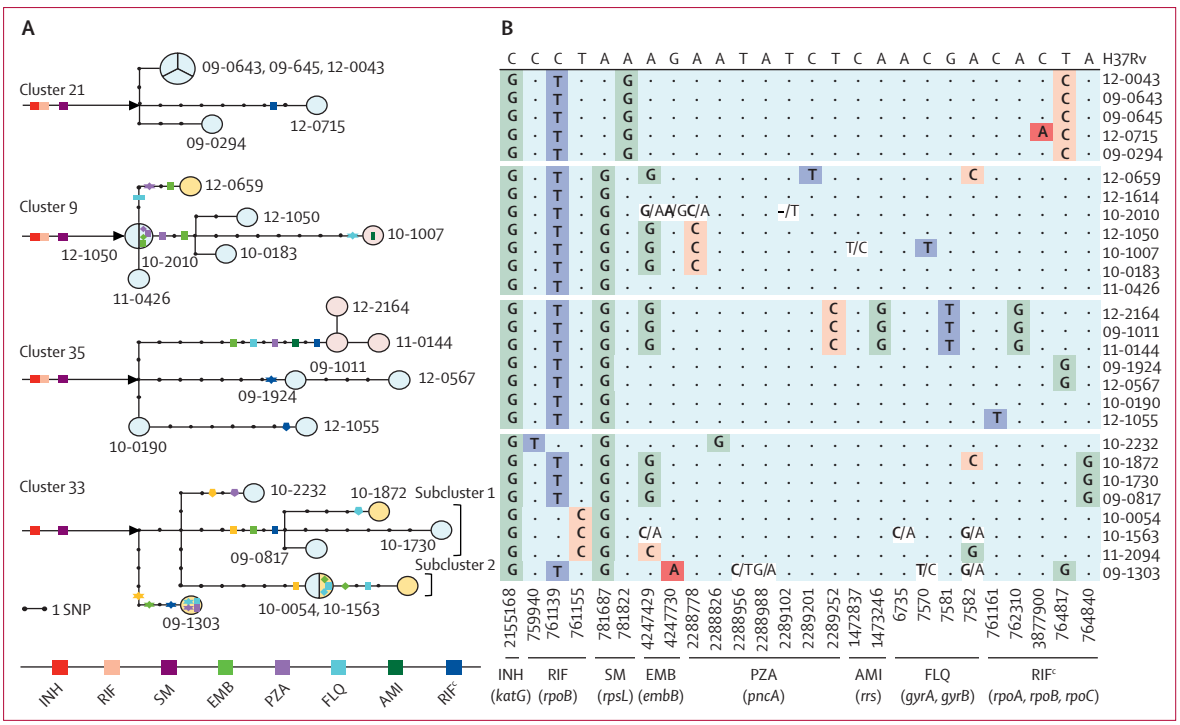

Figure 4: **Transmission networks of multidrug-resistant tuberculosis based on genetic distances and drug-resistance mutations**
(A) Median-joining networks for four main multidrug-resistant tuberculosis clusters. For each network, the arrow indicates the root and circles represent *Mycobacterium tuberculosis* isolates (numbered according to patient identification). *M tuberculosis* isolates are separated by lines with length representing genetic distance. Isolates with identical genomes are grouped in the same circle. Patients with multidrug-resistant (blue), pre-extensively drug-resistant (yellow), and extensively drug-resistant (pink) tuberculosis are shown. Different shapes of the same colour indicate different mutations for the same resistance phenotype. Heterogeneous mutations were mapped into corresponding circles to indicate emerging resistance. Fixed mutations were mapped onto branches by assuming parsimonious acquisition with no reversion. (B) Matrix of resistance mutations for clusters in (A). Fixed mutations are highlighted with colours, heterogeneous mutations have a white background and wild-type alleles are dots in a light-blue background. INH=isoniazid-resistance mutation. RIF=rifampicin-resistance mutation. SM=streptomycin-resistance mutation. EMB=ethambutol-resistance mutation. PZA=pyrazinamide-resistance mutation. AMI=amikacin-resistance mutation. FLQ=fluoroquinolone-resistance mutation. RIF^{c}=RIF compensatory.

were in 38 genomic clusters, which ranged in size from two patients (26 clusters) to eight patients (one cluster), indicating recent transmission of MDR strains. The presence of MDR tuberculosis among treatment-naive patients suggested transmission of MDR strains. If the cases of MDR tuberculosis among treatment-naive cases was combined with those in genomic clusters, up to 235 (73%; 95% CI 67·3–77·3) of 324 MDR tuberculosis cases were probably caused by transmission of MDR strains (figure 1).

The genomic clusters might represent transmission of an MDR strain, or initial transmission of a non-MDR strain that later developed multidrug resistance. To validate the transmission of an MDR strain, we constructed a median-joining network of each genomic cluster and mapped drug-resistance mutations onto the network (figure 4; appendix p 11, 19). 34 (89%) of the 38 genomic clusters had mutations associated with resistance to isoniazid and rifampicin that were consistent among all of the strains within a cluster, confirming that MDR strains were transmitted rather than having acquired resistance. Many other mutations associated with drug resistance were also consistent within all of the strains in these 34 clusters, including 16 clusters that contained strains with shared rifampicin-resistance compensatory mutations (appendix p 11). Taken together, 92% (95 of 103) of genomically clustered cases of MDR tuberculosis resulted from recent transmission of MDR strains.

On the basis of epidemiological investigations, 64 (69%) of the 93 genomically clustered cases of MDR tuberculosis had identifiable epidemiological links, whereas 29 (31%) cases were not epidemiologically linked. The most common links were living in the same residential complex or community, or on the same neighbourhood street, or using shared public facilities such as food markets (45 [70%] of 64 cases; table 1). However, household links were identified in only four patients. Table 2 shows the results of the univariate analysis of the risk factors associated with genomic clustering of MDR tuberculosis. In the multivariable model, having a diagnostic delay of at least 2 months (adjusted OR 2·02, 95% CI 1·08–3·81), being aged 45–64 years (1·90, 1·04–3·46), or being 65 years or older (2·94, 1·28–6·79) were independently associated with genomic clusters of MDR tuberculosis.

In the present study, we identified unfixed SNPs in genes associated with drug-resistance in 37 (36%) of the 103 genomically clustered MDR strains, suggesting the emergence of drug-resistance mutations within the host. By mapping these mutations onto the putative transmission networks, eight clusters showed between-patient fixation of these newly emerged mutations, evidence of the directionality of transmission (figure 4; appendix p 11). Notably, 87% (33 of 38) of the MDR tuberculosis clusters accumulated additional drug-resistance mutations along the transmission chain. 43 (42%) of the 103 genomically clustered MDR strains were pre-extensively drug resistant and 11 (11%) were extensively drug resistant (figure 3; appendix p 11).

One cluster, C-09, represents both emergence and fixation of drug-resistant mutations during transmission (figure 5). Two putative index cases, a husband (P1614) and wife (P0659), were first diagnosed with MDR tuberculosis in March, 2006, before the present study. Both putative index cases were non-compliant with their treatment and probably transmitted *M tuberculosis* to all other individuals in this cluster upon moving to a residential complex in 2008. However, whole-genome sequencing-based analysis showed existence of a sub-index case diagnosed in 2010 (P2010) with within-host emergence of a pyrazinamide-related resistance mutation (*pncA* p.Val155Gly) and an ethambutol-related resistance mutation (*embB* p.Met306Val); these two mutations were fixed during the subsequent infection of three other hosts (P1050, P0183, and P1007). The MDR strain within patient P2010 probably evolved into at least two clonal subpopulations and one of the clonal subpopulations was transmitted to the subsequent hosts, with increased drug resistance. These transmission events probably occurred in the game room of the residential complex, although patient P1007 did not share the game room.

Mutations in *rpoA*, *rpoC*, and regions outside the rifampicin-resistance determining region were associated with compensation for rifampicin resistance.[21] We investigated the occurrence of compensatory mutations in *rpoA* and *rpoC* genes in strains carrying rifampicin-resistance mutations. 37 non-synonymous SNPs in *rpoA* (four) and *rpoC* (33) were identified (appendix p 8, 15), and most (27 [73%] of 37) were in strains containing the *rpoB* p.Ser450Leu substitution. These SNPs occurred more frequently in MDR strains with the *rpoB* p.Ser450Leu mutation (64 [88%] of 73) than in MDR strains without this mutation (15 [31%] of 49, p=0·0001; appendix p 8). We identified an additional 21 non-synonymous SNPs in regions outside the rifampicin-resistance determining region of the *rpoB* gene. The 21 non-synonymous SNPs were significantly more likely to occur in strains without compensatory mutations in *rpoA* and *rpoC* than to occur in the strains with these mutations (20 [34%] of 58 *vs* four [6%] of 64, p=0·0007; figure 3; appendix p 15).

	Genomic clusters by whole-genome sequencing		OR (95% CI)†	p value
	Non-cluster (n=203)	Cluster (n=96)		
Sex				
Men	65 (32%)	21 (22%)	1·00	
Women	138 (68%)	75 (78%)	1·68 (0·95–2·97)	0·07
Age, years				
15–34	95 (47%)	28 (29%)	1·00	
35–44	32 (16%)	15 (16%)	1·59 (0·75–3·37)	0·18
45–64	60 (30%)	38 (40%)	2·15 (1·18–3·90)	0·009
≥65	16 (8%)	15 (16%)	3·18 (1·36–7·41)	0·004
Previous treatment				
No	106 (52%)	55 (57%)	1·00	
Yes	97 (48%)	41 (43%)	0·81 (0·49–1·33)	0·41
Diagnostic delay				
<2 months	175 (86%)	70 (73%)	1·00	
≥2 months	28 (14%)	26 (27%)	2·20 (1·19–4·07)	0·005
Positive sputum smear result				
No	37 (18%)	16 (17%)	1·00	
Yes	164 (81%)	80 (83%)	1·13 (0·59–2·15)	0·58
Unknown	2 (1%)	0	..	..
Treatment outcome				
Cured or treatment completed	119 (59%)	58 (60%)	1·00	
Treatment failed	11 (5%)	6 (6%)	1·11 (0·39–3·18)	0·83
Default, moved, or lost to follow-up	34 (17%)	11 (11%)	0·66 (0·31–1·41)	0·28
Died	21 (10%)	10 (10%)	0·98 (0·43–2·21)	0·96
Still on treatment or unknown	18 (9%)	11 (11%)	1·25 (0·55–2·83)	0·58
Beijing strains				
No	27 (13%)	6 (6%)	1·00	
Yes	176 (87%)	90 (94%)	2·30 (0·91–5·81)	0·06
Cavitary disease‡ (n=296)				
No	97/201 (48%)	53/95 (56%)	1·00	
Yes	104/201 (52%)	42/95 (44%)	0·74 (0·45–1·21)	0·22

Data are n (%) unless otherwise specified. OR=odds ratio. *Data were available for 299 patients. †Comparison between genomically clustered and non-clustered multidrug-resistant tuberculosis cases. ‡Cavitary disease was based on the presence or absence of a cavity on the chest radiograph at the time of diagnosis.

Table 2: **Univariable analysis of risk factors associated with multidrug-resistance in genomic clusters (n=299)***

Overall, 68 (66%, 95% CI 56·0–75·0) of the 103 genomically clustered MDR strains had compensatory mutations in *rpoA*, *rpoB*, or *rpoC* genes, with 63 (86%, 76·2–93·2) of 73 strains with the *rpoB* p.Ser450Leu resistance mutation having compensatory mutations (figure 3; appendix p 23).

Discussion

Our strategy combined traditional genotyping, whole-genome sequencing, and epidemiological investigation. Transmission of MDR strains accounted for most (73%) cases of MDR tuberculosis overall. Residential communities or complexes, and related

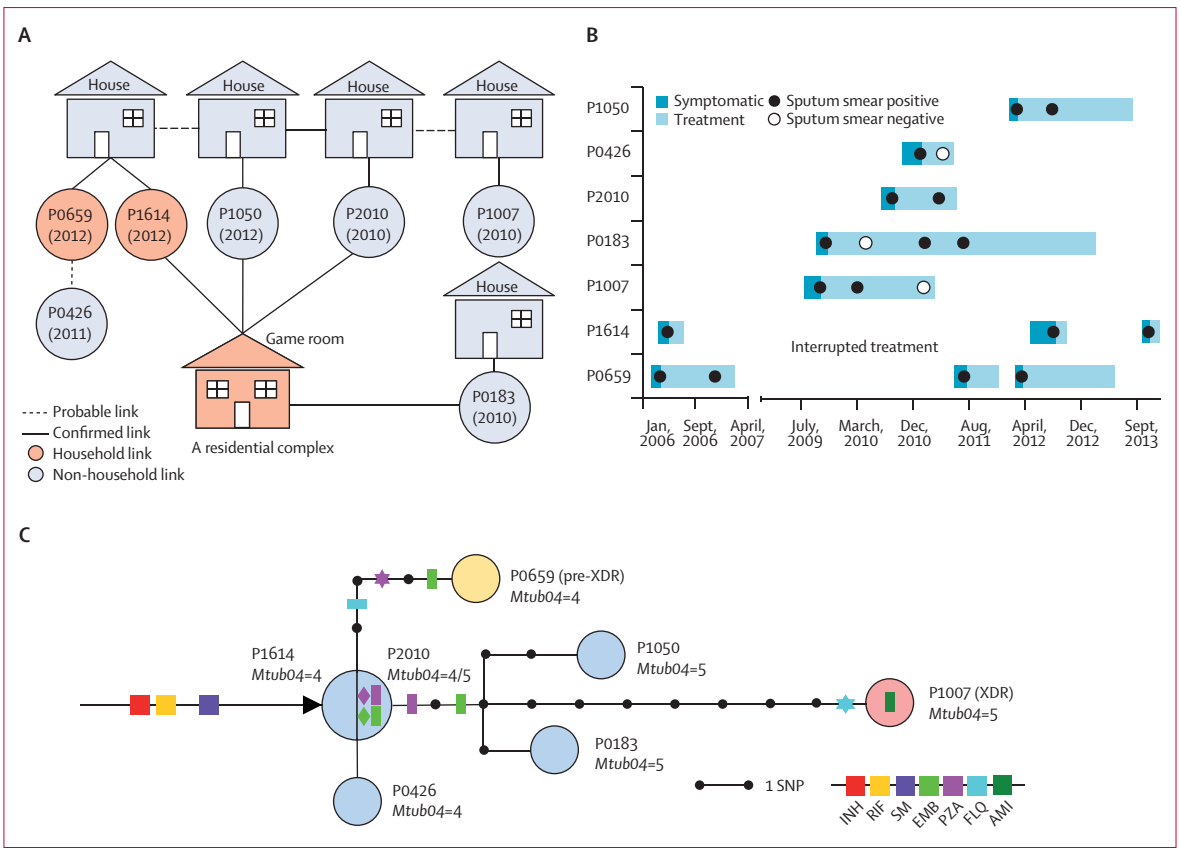

Figure 5: **Results of the detailed epidemiological investigation and genomic analysis of cluster 09**
(A) Putative transmission network based on epidemiological links. (B) Time of onset of symptoms, diagnosis, and treatment (based on Walker and colleagues' study[8]) (C) Median-joining network and drug-resistance mutations of clusters. Circles represent clusters of strains separated by no SNPs; *Mycobacterium tuberculosis* isolates are separated by lines with length representing genetic distance. Patients with multidrug-resistant (blue), pre-XDR (yellow) and XDR (pink) tuberculosis are shown. Heterogeneous mutations are mapped into corresponding circles in the median-joining networks to indicate emerging resistance. The fixed mutations are mapped onto the branches assuming parsimonious acquisition with no reversion. Different shapes of the same colour indicate different mutations for the same resistance phenotype. XDR=extensively drug-resistant. INH=isoniazid-resistance mutation. RIF=rifampin-resistance mutation. SM=streptomycin-resistance mutation. EMB=ethambutol-resistance mutation. PZA=pyrazinamide-resistance mutation. FLQ=fluoroquinolone-resistance mutation. AMI=amikacin-resistance mutation.

public facilities, were the most common transmission settings. Whole-genome sequencing analyses were highly informative to determine the transmission patterns of MDR strains and the ongoing acquisition of additional antibiotic resistance during transmission.

About a third of cases of MDR tuberculosis in our study population were attributable to recent transmission, a higher proportion than that reported in low-burden countries such as the USA or the UK.[11,13] In a multi-setting study[16] in China, 43% of MDR tuberculosis cases were in genotypic clusters and multidrug resistance was an independent risk factor for recent transmission of *M tuberculosis* strains. We might still have underestimated the real burden of MDR tuberculosis transmission; patients with genomically unique strains could also have reinfection with MDR strains. Our previous study[22] showed that 84% of drug-resistant tuberculosis cases in people with tuberculosis treatment history were attributed to exogenous reinfection with a new drug-resistant strain.

Transmission of MDR tuberculosis in other areas in China could be much more serious than in Shanghai. The 2008 national survey[23] of drug-resistant tuberculosis in China suggested that more than 80% of MDR tuberculosis cases resulted from the transmission of MDR strains. Rural areas of China have few resources to undertake culture and drug susceptibility testing, and WHO estimated[1] that less than 10% of patients with MDR tuberculosis in China were diagnosed or treated. Since most cases of MDR tuberculosis in China are undetected or inadequately treated, the transmission of these strains will result in a much more serious MDR tuberculosis epidemic than in the current setting.

Whole-genome sequencing currently offers several advantages over traditional genotyping methods to differentiate clinical *M tubercolosis* isolates, and allows for effective studies of strain transmission dynamics.[6,8,10,24–27] In this study, whole-genome sequencing provided definitive evidence of emergence and fixation of drug-resistance mutations along transmission chains.

Inadequate therapies could accelerate selection of new mutations and increase drug resistance, and these drug-resistant strains could be transmitted to others. Unfortunately, inadequate treatment of MDR tuberculosis is a problem in China, particularly in hospitals.[23,28] Therefore, treating patients with MDR tuberculosis on the basis of results of drug susceptibility testing should become a mandatory policy.

Fitness reduction due to drug-resistant mutations can be restored through compensatory evolution, which might be associated with the spread of MDR strains.[29] Most genotypically clustered MDR strains in our study had *rpoA*, *rpoB*, or *rpoC* compensatory mutations that potentially reduce the fitness cost associated with drug-resistance mutations.[5,6,21,30] The proportion of compensatory mutations found in the present study (66%) was much higher than that reported globally (20%) and in high-burden regions (31%),[21] but is similar to the high percentage observed in Samara, Russia (87%), where MDR tuberculosis is endemic.[5] One study[31] in South Africa observed that *rpoC* compensatory mutations were associated with recent transmission of drug-resistant tuberculosis. Larger, prospective studies are needed to determine the transmissibility of MDR strains with compensatory mutations.

Our study has several limitations that could lead to underestimation of the true magnitude of the transmission of MDR strains. First, the retrospective study design and the study timeframe make it possible that we did not capture all the MDR tuberculosis cases and strains in the population. Second, strains could be misclassified as unique if they were in fact clustered with strains outside the study period and geographical setting. Third, some patients died or were lost to follow-up before the epidemiological investigation was completed. We might have missed the identification of some transmission settings and epidemiological links.

Transmission of MDR strains of tuberculosis has a substantial role in the burden of MDR tuberculosis in China. Furthermore, most strains develop increased drug resistance during the expansion of MDR clusters. Interventions, such as early detection of cases, infection control, and evidence-based treatment, are urgently needed to stem the epidemic of MDR tuberculosis in China.

Contributors

CY, TL, XS, JM, KD, and QG designed the study. CY, XS, ZW, SL, and ZY did the epidemiological investigations. JW, PX, and JT did the laboratory work and generated the laboratory data. CY, TL, MG, and QL did the data analyses. CY, TL, XS, JM, KD, and QG drafted and revised the manuscript. All coauthors reviewed and approved the final manuscript before submission.

Declaration of interests

We declare no competing interests.

Acknowledgments

This study was supported by the National Science and Technology Major Project (2013ZX10004903-006); National Natural Science Foundation of China (81402727 to CY, 91631301 to QG); Ministry of Science and Technology of China (2014DFA30340); the International Postdoctoral Fellowship Program funded by China Postdoctoral Science Foundation (20150058 to CY); the GloCal Health Fellowship Program sponsored by the National Insititutes of Health Fogarty International Center and University of California Global Health Institute (R25 TW009343 to CY); National Institutes of Health grants (D43TW007887 to TL, QG, and KD and DP2OD006452 to KD); and Shanghai Municipal Commission of Health and Family Planning grants (GWTD2015S02 to XS).
We thank the tuberculosis public health teams in Shanghai Municipal Centre for Disease Control and Prevention for their support. We thank the reviewers and Ted Cohen from Yale University for their valuable comments on this manuscript

References

1 WHO. Global tuberculosis report 2016. http://www.who.int/tb/publications/global_report/en/ (accessed Oct 13, 2016).
2 Nathanson E, Nunn P, Uplekar M, et al. MDR tuberculosis—critical steps for prevention and control. *N Engl J Med* 2010; **363:** 1050–58.
3 Barnes PF, Cave MD. Molecular epidemiology of tuberculosis. *N Engl J Med* 2003; **349:** 1149–56.
4 Gardy JL, Johnston JC, Ho Sui SJ, et al. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N Engl J Med* 2011; **364:** 730–39.
5 Casali N, Nikolayevskyy V, Balabanova Y, et al. Evolution and transmission of drug-resistant tuberculosis in a Russian population. *Nat Genet* 2014; **46:** 279–86.
6 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.
7 Luo T, Yang C, Peng Y, et al. Whole-genome sequencing to detect recent transmission of *Mycobacterium tuberculosis* in settings with a high burden of tuberculosis. *Tuberculosis (Edinb)* 2014; **94:** 434–40.
8 Walker TM, Ip CLC, Harrell RH, et al. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect Dis* 2013; **13:** 137–46.
9 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.
10 Casali N, Broda A, Harris SR, Parkill J, Drobniewski F. Whole genome sequence analysis of a large isoniazid-resistant tuberculosis outbreak in London: a retrospective observational study. *PLoS Med* 2016; **13:** e1002137.
11 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.
12 Dobler CC, Korver S, Batbayar O, et al. Multidrug-resistant tuberculosis in patients for whom first-line treatment failed, Mongolia, 2010-2011. *Emerg Infect Dis* 2015; **21:** 1451–54.
13 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.
14 Kendall EA, Fofana MO, Dowdy DW. Burden of transmitted multidrug resistance in epidemics of tuberculosis: a transmission modelling analysis. *Lancet Respir Med* 2015; **3:** 963–72.
15 Flandrois JP, Lina G, Dumitrescu O. MUBII-TB-DB: a database of mutations associated with antibiotic resistance in *Mycobacterium tuberculosis*. *BMC Bioinformatics* 2014; **15:** 107.
16 Yang C, Shen X, Peng Y, et al. Transmission of *Mycobacterium tuberculosis* in China: a population-based molecular epidemiologic study. *Clin Infect Dis* 2015; **61:** 219–27.
17 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 USA* 2015; **112:** 8136–41.
18 Luo T, Yang C, Pang Y, Zhao Y, Mei J, Gao Q. Development of a hierarchical variable-number tandem repeat typing scheme for *Mycobacterium tuberculosis* in China. *PLoS One* 2014; **9:** e89726.
19 Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009; **25:** 2078–79.

20 Sun G, Luo T, Yang C, et al. Dynamic population changes in *Mycobacterium tuberculosis* during acquisition and fixation of drug resistance in patients. *J Infect Dis* 2012; **206:** 1724–33.

21 Comas I, Borrell S, Roetzer A, et al. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat Genet* 2012; **44:** 106–10.

22 Li X, Zhang Y, Shen X, et al. Transmission of drug-resistant tuberculosis among treated patients in Shanghai, China. *J Infect Dis* 2007; **195:** 864–69.

23 Zhao Y, Xu S, Wang L, et al. National survey of drug-resistant tuberculosis in China. *N Engl J Med* 2012; **366:** 2161–70.

24 Bryant JM, Harris SR, Parkhill J, et al. Whole-genome sequencing to establish relapse or re-infection with *Mycobacterium tuberculosis*: a retrospective observational study. *Lancet Respir Med* 2013; **1:** 786–92.

25 Guerra-Assuncao JA, Crampin AC, Houben RM, et al. Large-scale whole genome sequencing of *M. tuberculosis* provides insights into transmission in a high prevalence area. *Elife* 2015; **4:** e05166.

26 Kato-Maeda M, Ho C, Passarelli B, et al. Use of whole genome sequencing to determine the microevolution of *Mycobacterium tuberculosis* during an outbreak. *PLoS One* 2013; **8:** e58235.

27 Perez-Lago L, Comas I, Navarro Y, et al. Whole genome sequencing analysis of intrapatient microevolution in *Mycobacterium tuberculosis*: potential impact on the inference of tuberculosis transmission. *J Infect Dis* 2014; **209:** 98–108.

28 Xue He G, van den Hof S, van der Werf MJ, et al. Inappropriate tuberculosis treatment regimens in Chinese tuberculosis hospitals. *Clin Infect Dis* 2011; **52:** e153–56.

29 Cohen T, Murray M. Modeling epidemics of multidrug-resistant *M. tuberculosis* of heterogeneous fitness. *Nat Med* 2004; **10:** 1117–21.

30 Gagneux S, Long CD, Small PM, Van T, Schoolnik GK, Bohannan BJ. The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. *Science* 2006; **312:** 1944–46.

31 de Vos M, Muller B, Borrell S, et al. Putative compensatory mutations in the *rpoC* gene of rifampin-resistant *Mycobacterium tuberculosis* are associated with ongoing transmission. *Antimicrob Agents Chemother* 2013; **57:** 827–32.