The Lancet Regional Health - Western Pacific 8 (2021) 100106

Contents lists available at ScienceDirect

The Lancet Regional Health - Western Pacific

journal homepage: www.elsevier.com/locate/lanwpc

Research paper

A population-based genomic epidemiological study of the source of tuberculosis infections in an emerging city: Shenzhen, China

Tingting Yang[a,b], Yunxia Wang[c], Qingyun Liu[a], Qi Jiang[a,d], Chuangyue Hong[b], Likai Wu[b], Shuangjun Li[b], Chendi Zhu[a], Howard Takiff[e,f,g], Weiye Yu[b], Weiguo Tan[b,*], Qian Gao[a,b,**]

[a] Key Laboratory of Medical Molecular Virology (MOE/NHC/CAMS), School of Basic Medical Sciences, Shanghai Medical College and Shanghai Public Health Clinical Center, Fudan University, Room 201, Fuxing Building, Dongan Road No. 131, Shanghai 200032, China
[b] Shenzhen Center for Chronic Disease Control, 2021 Buxin Road, Shenzhen, China
[c] Bao'an District Hospital for Chronic Diseases Prevention and Cure, 332 Yu'an 2nd Road, Shenzhen, China
[d] School of Health Sciences, Wuhan University, 115 Donghu Road, Wuhan, China
[e] Unité de Pathogenetique Integrée Mycobacterienne, Institut Pasteur, Paris, France
[f] Laboratorio de Genética Molecular, CMBC, IVIC, Caracas, Venezuela
[g] Shenzhen Nanshan Center for Chronic Disease Control, Shenzhen, China

ARTICLE INFO

Article history:
Received 19 October 2020
Revised 13 January 2021
Accepted 27 January 2021
Available online 5 February 2021

ABSTRACT

Background: Tuberculosis (TB) in emerging cities is often a disease of recent immigrants, and understanding this epidemiology is crucial for designing effective control and prevention strategies.

Methods: We conducted a retrospective population-based genomic epidemiological study of culture-positive pulmonary TB patients diagnosed between June 2014 and November 2017 in the Bao'an District of Shenzhen, a Chinese city with dramatic recent growth. After whole genome sequencing, transmission clusters were defined as strains differing by no more than 12 SNPs.

Findings: Of 1696 culture-positive TB patients, 93•8% (1591/1696) were migrants, with 51•6% (821/1591) employed in housekeeping or unemployed. Of the 1460 migrants with known residence time, 47•7% (697/1460) developed TB within two years after arriving in Bao'an. Only 12•2% (207/1696) of Bao'an isolates were in genomic clusters, indicating that recent transmission was not the primary cause of TB in Bao'an. The isolates' median terminal branch length was 56 SNPs, more than could have accumulated since the arrival of the migrants in Bao'an. The migrants' isolates had genotypic distributions similar to those in their home provinces. One strain isolated in Bao'an belonged to a clade circulating in the patient's home province, providing further evidence that the strains were brought to Bao'an with the migrants.

Interpretation: TB in the Bao'an District is principally caused by reactivation of infections acquired by migrants in their home provinces. Nearly half developed TB within two years after arriving in Bao'an, suggesting a need for increased TB screening of migrants, especially housekeeping workers and the unemployed.

Funding: Sanming Project of Medicine in Shenzhen; National Science and Technology Major Project of China; Natural Science Foundation of China.

Research in context

Evidence before this study

We searched PubMed, the web of science and the China national knowledge infrastructure (CNKI) for reports published before December 2019 using search terms: tubercu-

* Corresponding author.
** Corresponding author at: Key Laboratory of Medical Molecular Virology (MOE/NHC/CAMS), School of Basic Medical Sciences, Shanghai Medical College and Shanghai Public Health Clinical Center, Fudan University, Room 201, Fuxing Building, Dongan Road No. 131, Shanghai 200032, China.
E-mail addresses: szmbjfk@wjw.sz.gov.cn (W. Tan), qiangao@fudan.edu.cn (Q. Gao).

https://doi.org/10.1016/j.lanwpc.2021.100106

losis, China, migrant, and urban or urbanization. We found 26 articles related to the effect of massive internal migration on the tuberculosis burden in rapidly growing cities of China. These regional studies suggested that the influx of migrants brought serious challenges for TB control and prevention in emerging cities such as Shenzhen, Suzhou, and Ningbo. Between 2001 and 2010, internal migrants constituted 80% of registered TB patients in Shenzhen. In Suzhou, from 2008–2013 the TB incidence among migrants was 21•6% higher than in the local population. In the Jiangbei District of Ningbo, migrants accounted for 50•6% of TB patients. The migrants are drawn to the new cities largely for manufacturing jobs, where they often live and work together in crowded housing and factories. It was therefore not clear whether the high TB rates were the result of active transmission in the cities, or infections that the migrants acquired in their home towns and brought with them.

Added value of this study

This is the first study to investigate the source of TB infections in an emerging city of China. We estimate that 93•8% of TB cases in the Bao'an District of Shenzhen occurred in internal migrants, nearly half of whom had lived in Bao'an for no more than 2 years. Whole genome sequencing showed that the overall clustering rate in Bao'an was only 12•2%, suggesting that local transmission is responsible for only a small portion of the TB burden in this district. Analyses of the genetic structure of the *Mycobacterium tuberculosis* strains further demonstrated that the infections were brought to Bao'an with the migrants. The isolates' median terminal branch length was 56 SNPs, indicating more strain evolution than could have accumulated during the 40 years of Shenzhen's dramatic population growth. Moreover, the migrants' isolates had genotypic distributions that were more similar to the distributions in their home provinces than in strains isolated from long-term Shenzhen residents. We also found a strain isolated from a migrant patient in Bao'an that belonged to a clade circulating in the patient's home province but not in Shenzhen.

Implication of all the available evidence

This is the first study to demonstrate that the TB burden in an emerging city of China is predominantly caused by reactivation of TB infections acquired by migrants in their home provinces. This result suggests that effective control strategies in this and similar Chinese cities should include a greater effort to screen for TB in newly arrived migrants.

1. Introduction

China's urbanization process over the past several decades has been accompanied by rapid growth of cities such as Shenzhen, Suzhou, and Xiamen, principally due to the migration of young people from rural areas attracted by employment opportunities. In 1979, as part of China's reform process and economic opening, Shenzhen was the first Chinese city designated as a special economic zone. Since then, Shenzhen has grown from a village of 30,000 inhabitants to a metropolis with a population greater than 12 million [1]. The massive influx of migrants to Shenzhen has brought serious challenges for tuberculosis (TB) control and prevention. The reported incidence of TB in Shenzhen was 46•8/100,000 in 2019, and a previous study has found that, between 2001 and 2010, internal migrants constituted 80% of registered TB patients in the city [2].

In order to design effective TB control programs in Shenzhen and other emerging cities, it is essential to know whether TB in the migrant population is primarily the result of reactivated TB infections acquired in their home provinces, or is caused by recent transmission after arriving in the city. For example, TB in foreign-born patients make up a high proportion of cases in countries with many immigrants, such as the United States (71•5%), United Kingdom (73%), Australia (88%), New Zealand (79•5%) and Canada (71%) [3–7]. Most of these cases were the result of latent infections acquired in their birthplaces that reactivated to cause active TB within two-five years after arriving in these countries [8,9]. TB prevention and control in these countries is therefore focused on screening recent immigrants from high burden countries [10]. In contrast, local recent transmission in migrants is an important contribution to TB burden in some cities. For example, in Madrid, Spain, recent transmission was responsible for 29% of the immigrant cases [11]. Transmission frequently occurred between immigrants from different countries as well as between immigrants and the local Spanish population. Therefore, in addition to TB screening of new immigrants to detect imported cases, TB control in this setting also requires strategies to interrupt local transmission [11]. A recent study in Songjiang District of Shanghai found that more than two-thirds of migrants belonging to genomic clusters were likely infected after they arrived in Songjiang [12].

The Bao'an District has the largest migrant population of the ten districts in Shenzhen, and more than 90% of new TB cases occur in migrants [13,14]. This district is therefore an informative site to study the sources of the TB burden in cities that have experienced rapid recent growth due to the influx of internal migrants. To understand the epidemiology and thereby facilitate TB control in emerging cities of China, we performed a genomic epidemiological study of TB strains isolated from cases of pulmonary tuberculosis diagnosed in Bao'an from June 2014 to November 2017.

2. Methods

2.1. Sample source

The Bao'an District is located in northwestern Shenzhen, Guangdong Province. In 2017, it had a population of 3•149 million. TB is principally diagnosed and treated at the Bao'an District Hospital for Chronic Diseases Prevention and Cure. Before 2017, only those sputum samples positive for bacilli by microscopy were cultured, but beginning in 2017, cultures were performed on sputum samples from all patients diagnosed as having TB. Positive cultures were sent to the central Shenzhen Center for Chronic Disease Control (CCDC) TB laboratory for species identification and drug susceptibility testing [15]. All patients diagnosed in Bao'an with positive TB cultures between June 2014 to November 2017 were enrolled in this study. The study protocol was approved by the Ethical Review Board of Shenzhen CCDC (No. 20180310).

2.2. Whole-genome sequencing

The *Mycobacterium tuberculosis* strains were re-cultured and sequenced as previously described [16]. A previously validated pipeline was used to identify SNPs [17], with the inferred *M. tuberculosis* complex ancestor sequence as the reference genome. In brief, the Sickle tool [18] was used for trimming whole-genome sequencing data, and sequencing reads with Phred base quality above 20 and read length longer than 30 were kept for analysis. The retained sequencing reads were mapped to the reference genome with Bowtie2 (v2•4•1) [19], and the SAMtools (v1•6)/VarScan (v2•3•6) [20,21] suite was used for SNP calling with mapping quality greater than 30. The fixed SNPs (frequency $\geq$ 75%), excluding those in drug-resistance associated genes and repetitive regions of the genome (e.g., PPE/PE-PGRS family genes, phage sequence, insertion or mobile genetic elements), were used to calculate the pairwise SNP distances. We defined genomically-clustered strains as those within a threshold distance of twelve or fewer SNPs [22]. Strains with a sequence depth less than 20X or

a genome coverage less than 95% were excluded from the analysis. The remaining strains were classified into different lineages according to Coll et al. [23]. L2 strains were classified into L2•1, L2•2 and L2•3 [24]. The L2•3 represents "modern" Beijing and others are "ancient" Beijing. L4 strains were further classified according to the previously defined sublineages [25]. Sequencing data were deposited in the Genome Sequence Archive (https://bigd.big.ac.cn/gsa) under accession number CRA003250.

The drug-resistance profile was predicted for 14 anti-TB drugs based on the mutations reported to be associated with resistance (Supplementary Table 1) [26]. Multidrug-resistant tuberculosis (MDR-TB) was defined as isolates with mutations conferring resistance to at least isoniazid and rifampicin. Pre-extensively drug-resistant (pre-XDR) TB was defined as an MDR-TB strain with mutations conferring additional resistance to fluoroquinolones (FQs) or any second-line injected drugs (SLIDs), and XDR-TB was defined as MDR-TB with mutations conferring resistance to both of these antibiotic classes.

Phylogenetic reconstructions were performed as described previously by Liu et al [17] using MEGA X [27] with the neighbor-joining method and 100 bootstraps. The phylogenetic trees were visualized with Interactive Tree of Life (https://itol.embl.de/). The terminal branch lengths (TBLs) were calculated using method described in [28].

2.3. Statistical analyses

We extracted the basic demographic and clinical data for the enrolled TB patients from the routine TB surveillance system of the Bao'an CCDC, and their social factors from a structured questionnaire routinely administered to all new TB patients. The surveyed data was entered into Epidata software (v4•6) and exported as an Excel table. Continuous data was expressed as mean (m) ± standardized deviation (sd), while categorical variables were described using proportions and interquartile ranges. Differences between groups were tested using the chi-square test. Risk factors including demographic, social and clinical information were analyzed for patients infected by genetically-clustered *M. tuberculosis* strains using univariable and multivariable logistic regressions, with the estimation of their odds ratios (ORs) and 95% confidence intervals (CIs). Factors with a *P*-value less than 0•2 in the univariable regression were included in the multivariable analysis, and were considered to be significantly associated with genomic clustering if their *P*-value were less than 0•05 in the final multivariable model. All statistical analysis was performed in SPSS V25•0 (IBM, USA). Missing data was not included in the analyses.

2.4. Role of the funding source

The sponsors were not involved 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 as well as final responsibility for the decision to submit for publication.

3. Results

3.1. General information and residence time of Bao'an TB patients

A total of 3704 patients with active pulmonary TB were registered in the Bao'an district of Shenzhen during the period from June 2014 to November 2017, of whom 1819 had positive sputum cultures. Of these 1819 patients, 123 were excluded because strains isolated at the time of their initial diagnosis could not be re-cultured from frozen stocks or the poor quality of their genomic sequences did not permit analysis (Fig. 1a). Since the population of Shenzhen is mainly composed of internal migrants, we divided the patients into local residents and internal migrants to study the TB epidemiology. The internal migrants are defined as those who were not born in Shenzhen and their home provinces could be inferred from their ID number. Of the 1696 patients whose isolates were included in the study, 93.8% (1591/1696) were internal migrants originally from 26 provinces, but most were from provinces close to Shenzhen (Fig. 1b).

The general information of the TB patients in Bao'an is described in Table 1. The average age of the 1696 patients was 34 years (range 13 to 92 years), and 65•3% (1108/1696) were male. Just over half of patients (51•5%, 874/1696) were engaged in housekeeping or were unemployed, and most of the other patients were factory workers (38•4%, 651/1696). New TB cases accounted for 90•6% (1515/1672). Migrant patients with TB (average age 34 years, range 13 to 82 years) were significantly younger than the local residents with TB (average 42 years, range 16 to 92 years) ($P < 0•0001$). The migrants also had a lower education level, with 65•7% not going beyond junior high school, compared to 37•2% of local residents with this level of education, and more migrants were factory workers (40•0% vs 13•3%) (Table 1).

To investigate how soon after arriving in Bao'an the migrants developed TB, we used their residence time at the address given when they were diagnosed. Of the 1460 migrant patients with information on residence time, 29•1% (425/1460) developed active TB within 1 year of residence, 18•6% (272/1460) within 2 years, and the percentages progressively declined with increasing residence time (Fig. 2). It is notable that nearly half (47•7%) of migrant patients developed active TB within 2 years after arriving in Bao'an.

3.2. Whole genome sequencing for genotyping and drug resistance prediction

The whole genomes of the 1696 *M. tuberculosis* strains were sequenced with an average depth of 135X (25–513X) and an average coverage of 98•1% (96•5–99•1%). Phylogenetic analysis showed that these isolates predominantly belonged to the Beijing family (73%, 1238/1696), including 66•6% (824/1238) modern and 33•4% (414/1238) ancient Beijing strains. The 27% (458/1696) non-Beijing strains included 442 strains belonging to lineage 4 (L4), 15 belonging to L1 and one strain belonging to L3 (Table 1 and Fig. 3).

Analysis of the genome sequences for mutations conferring resistance to 14 anti-tuberculosis drugs showed that there were 1364 (80•4%) pan-sensitive strains and 332 (19•6%) strains resistant to at least one anti-tuberculosis drug (Table 1). The rifampin-resistant and MDR strains accounted for 7•9% (134/1696) and 6•7% (113/1696) respectively, and the 113 MDR strains included 29 pre-XDR (25 FQ-resistant strains and four SLID-resistant strains) and nine XDR strains. No mutations associated with resistance to linezolid, clofazimine or bedaquiline were detected. Compared with isolates from new TB patients, a greater percentage of isolates from retreated patients were drug resistant (35% vs 16•7%), or MDR (20•4% vs 3•8%) (Supplementary Table 1).

3.3. Identification of genomic-clustered cases and their risk factors

We defined transmission clusters as strains that differ by 12 or fewer SNPs. In the Bao'an isolates, we identified 207 clustered strains grouped into 91 genomic clusters containing two to six strains (Fig. 3 and Supplementary Table 2). The overall clustering rate was only 12•2% (207/1696), suggesting that recent transmission makes a small contribution to the TB burden in Bao'an. Univariable logistic regression suggested a significant association of younger patient age and birth in Shenzhen with isolates belonging to a genetic cluster. However, multivariable regression revealed

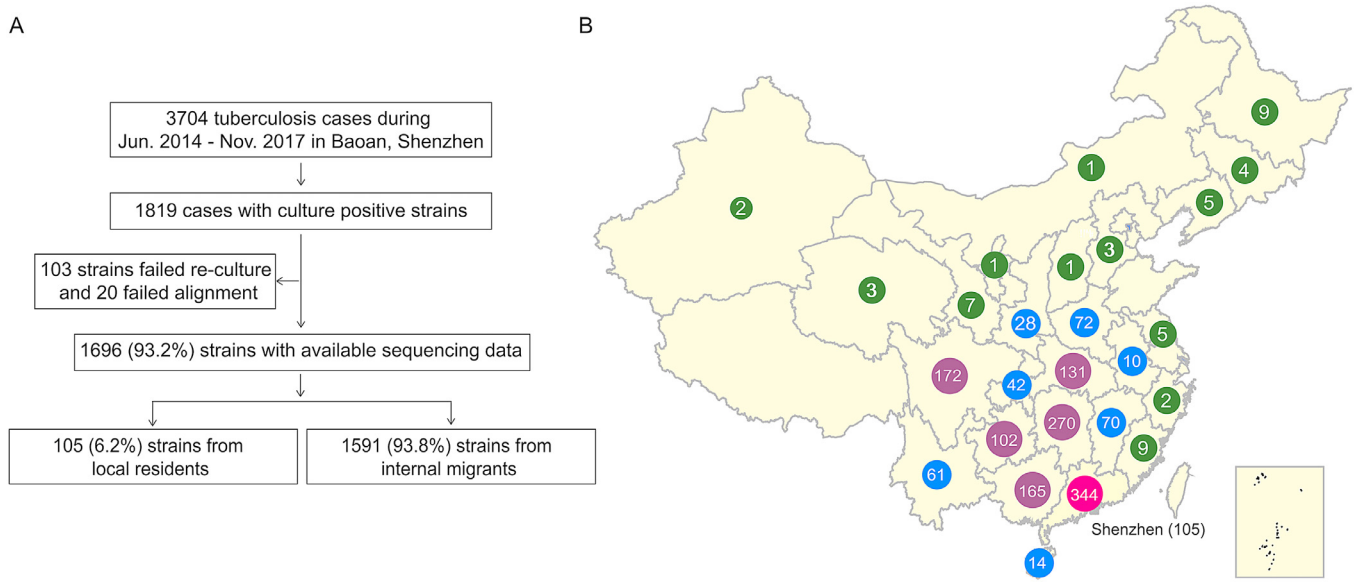

Fig. 1. Sample enrollment and geographic distribution of the home province of tuberculosis patients diagnosed in Bao'an, Shenzhen, June 2014–November 2017. (a) Flow chart of the isolates included in the study. (b) Home provinces of patients with pulmonary tuberculosis diagnosed in Bao'an. The numbers of enrolled patients who were born in the different provinces of China are indicated within the circles. Green, blue, and purple circles indicate, respectively, provinces where less than 10, between 10 and 100, and more than 100 Bao'an tuberculosis patients were born. The pink circle indicates cases from Guangdong province outside of Shenzhen.

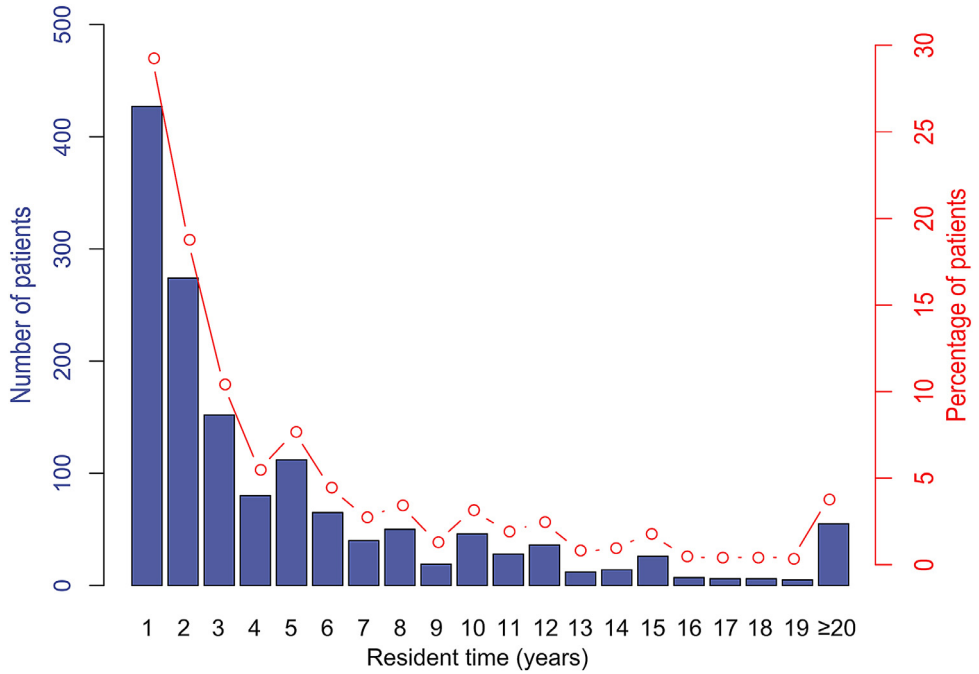

Fig. 2. The number (bars) and proportion (points) of migrant tuberculosis patients by years of residence time in Bao'an.

that residence years in Shenzhen rather than birth in Shenzhen directly increased the risk of belonging to a TB cluster. Considering the interaction of patient' age and residence years, the risk of TB transmission increased by 11•1% (4•6%–18•0%) for each additional year of residence time (Table 2), while age, mediated by residence time (Supplementary Table 3), had a weak effect on decreasing the risk of TB transmission by 0•2% (0–0•3%) (Table 2). It is worth mentioning that the clustering rate of patients living in Bao'an for more than two years (18•1%, 153/844) is twice that of patients living there for two years or less (9•8%, 70/711).

We compared the hometown origins of the clustered Bao'an migrant patients to further investigate whether strain clustering was the result of transmission within Bao'an. Of the 91 clusters, 77 accounting for 86•5% (160/185) of the clustered migrant patients, contained migrants who had come from different home provinces, suggesting that transmission had occurred within Bao'an. The other 12 clusters, comprised of a total of 25 patients, contained migrants from the same town or village, where they could have been infected with very similar local strains (Supplementary Table 4).

3.4. Genotype distributions reflect the source M. tuberculosis *strains in Bao'an*

The low percentage of clustered patients (12•2%) in Bao'an suggested that the disease in most patients likely resulted from reactivation of TB infections acquired in their hometowns. To verify

Table 1
Demographic, clinical, and bacteriological characteristics of local resident and internal migrant TB patients diagnosed in Bao'an, Shenzhen, June 2014–November 2017.

	Total (N = 1696)		Local resident (N = 105)		Internal migrant (N = 1591)		χ^2	P-value
	No.	%	No.	%	No.	%		
Demographic factors								
Gender							0•38	0•539
Female	588	34•7	33	31•4	555	34•9		
Male	1108	65•3	72	68•6	1036	65•1		
Age,years[a]							37•80	<0•0001
<24	430	25•4	14	13•3	416	26•1		
25–34	611	36	22	21•0	589	37•0		
35–44	287	16•9	25	23•8	262	16•5		
45–54	218	12•9	26	24•8	192	12•1		
≥55	149	8•8	18	17•1	131	8•2		
Occupation							94•00	<0•0001
Housekeeping or unemployed	874	51•5	53	50•5	821	51•6		
Factory worker	651	38•4	14	13•3	637	40•0		
Office worker	86	5•1	19	18•1	67	4•2		
Other	85	5	19	18•1	66	4•1		
Education[b]							51•15	<0•0001
University or above	168	9•9	23	21•9	145	9•1		
High school and technical secondary school	405	23•9	42	40•0	363	22•8		
Junior high school	909	53•6	28	26•7	881	55•4		
Primary school	148	8•7	6	5•7	142	8•9		
Illiterate or semi illiterate	27	1•6	5	4•8	22	1•4		
Residence time in Bao'an, years[c]							220•33	<0•0001
≤2	711	45•7	14	13•3	697	43•8		
3~5	343	22•1	5	4•8	338	21•2		
6~10	280	18	14	13•3	266	16•7		
>10	221	14•2	62	59•0	159	10•0		
Clinical factor								
Tuberculosis history[d]							0•00	0•952
New	1515	90•6	94	89•5	1421	89•3		
Previously treated	157	9•4	9	8•6	148	9•3		
Bacteriological factors								
Genotype							1•18	0•554
Ancient Beijing	414	24•4	21	20•0	393	24•7		
Modern Beijing	824	48•6	54	51•4	770	48•4		
Non-Beijing	458	27•0	30	28•6	428	26•9		
Drug resistance							0•32	0•853
Multi-drug resistance	113	6•7	6	5•7	107	6•7		
Other drug resistance	219	12•9	15	14•3	204	12•8		
Pan Sensitive	1364	80•4	84	80•0	1280	80•5		
Genomic clustered							5•59	0•018
Yes	207	12•2	21	20•0	186	11•7		
No	1489	87•8	84	80•0	1405	88•3		

[a] Default for one case.
[b] Default for 39 cases.
[c] Default for 141 cases.
[d] Default for 24 cases.

Table 2
Univariable and multivariable logistic regression of risk factors for clustering.

	Clustered (N = 207) No. (%); M±SD	Unique (N = 1489) No. (%); M±SD	Univariable Regression OR (95% CI)	P-value	Multivariable Regression OR (95% CI)	P-value
Male Gender	137 (66•18)	971 (65•21)	1•044 (0•768–1•419)	0•783	…	…
Age group, years[†]	32•28±11•12	34•63+13•37	0•985 (0.973–0.997)	0•016	0•993 (0•977–1•009)	0•387
Passive case finding	180 (86•96)	1339 (89•93)	0•747 (0•482–1•158)	0•192	0•636 (0•395–1•025)	0•063
Diagnostic delay >4 weeks	122 (58•94)	803 (53•93)	1•226 (0•913–1•647)	0•175	…	…
Retreated TB	22 (10•73)	135 (9•07)	1•186 (0•737–1•910)	0•482	…	…
Residence years[††]	7•13±8•33	6•09±9•34	1•011 (0•997–1•025)	0•141	1•111 (1•046–1•180)	0•001
Birth in Shenzhen	21 (10•14)	84 (5•64)	1•888 (1.143–3.120)	0•013	…	…
Beijing genotype	161 (77•8)	1076 (72•3)	1•343 (0•950–1•900)	0•095	1•348 (0•941–1•931)	0•104
Multidrug resistance	12 (5•8)	101 (6•8)	0•846 (0•456–1•567)	0•594	…	…
Age × residence time	…	…	…	…	0•998 (0•997–1•000)	0•011

Abbreviations: OR, odds ratio; CI, confidential interval; M, mean; SD, standardized deviation.
[†] Default for one case.
[††] Default for 141 cases.

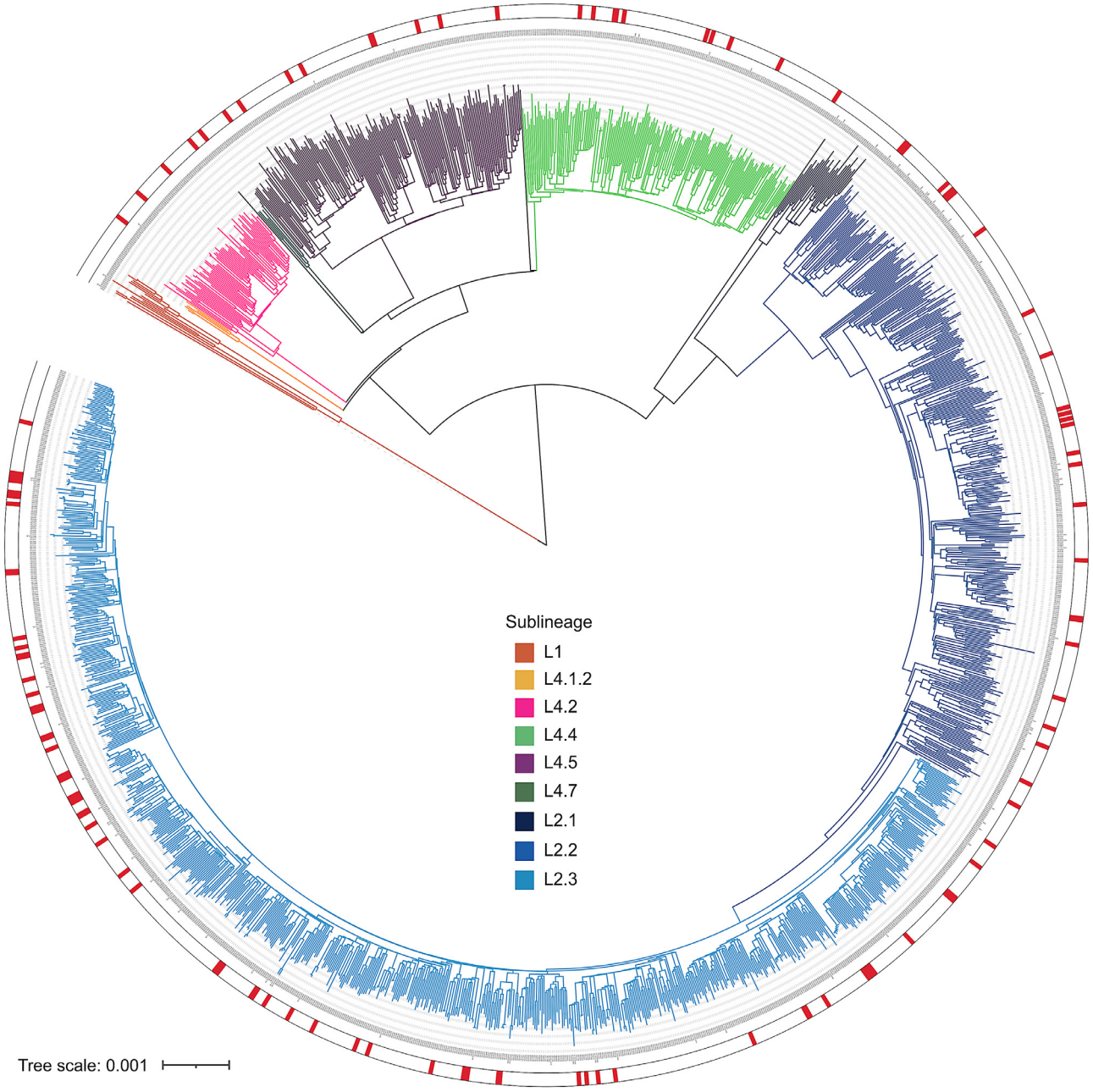

Fig. 3. The phylogenetic tree of 1696 *Mycobacterium tuberculosis* strains isolated in Bao'an. The different colors on the branches indicate different lineages and sublineages. The outside circle indicates genomically clustered strains.

this hypothesis, we analyzed branch lengths on the phylogenetic tree, where SNP differences from the nearest node reflect evolutionary time. A previous study in a setting with considerable recent transmission found short TBLs [28]. In contrast, in the Bao'an strains we observed a preponderance of long TBLs, with a median of 56 SNPs (interquartile range, 34–78 SNPs), although the median TBLs varied with the different sublineages (Fig. 4a). The modern Beijing strains (L2•3) had the shortest TBLs, with a median of 46 SNPs (interquartile range, 28–60 SNPs), while L4•5 strains showed the longest TBLs, with a median of 97 SNPs (interquartile range, 68–125 SNPs). Assuming that the mutations rate in *M. tuberculosis* is 0•3-0•5 SNPs per year per genome [29–31], it would take 92–323 years for strains of different sublineages to accumulate 46–97 SNPs, which is much longer than the 40-year history of migration into Shenzhen.

The different provinces of China have been shown to have distinct distributions of TB lineages and sublineages [17]. To see if these differences were reflected in the isolates from Bao'an, we grouped the isolates according to the patients' home provinces, excluding those in clusters, and compared their distributions of lineages and sublineages with the distributions of 3227 strains collected from 32 provinces across China [17]. This analysis found significant differences ($\chi^2 = 91•631$, $P < 0•0001$) in the genotypic distributions of the isolates that varied with the Bao'an patients' origins (Fig. 4b). Moreover, the isolates of Bao'an patients from some provinces had genotype distributions that were similar to the distributions in their home provinces (e.g. Guizhou), but significantly different from the genotype distribution in isolates from local residents ($\chi^2 = 14•697$, $P = 0•012$). This provides further evidence that most of the Bao'an patients were infected in their home provinces

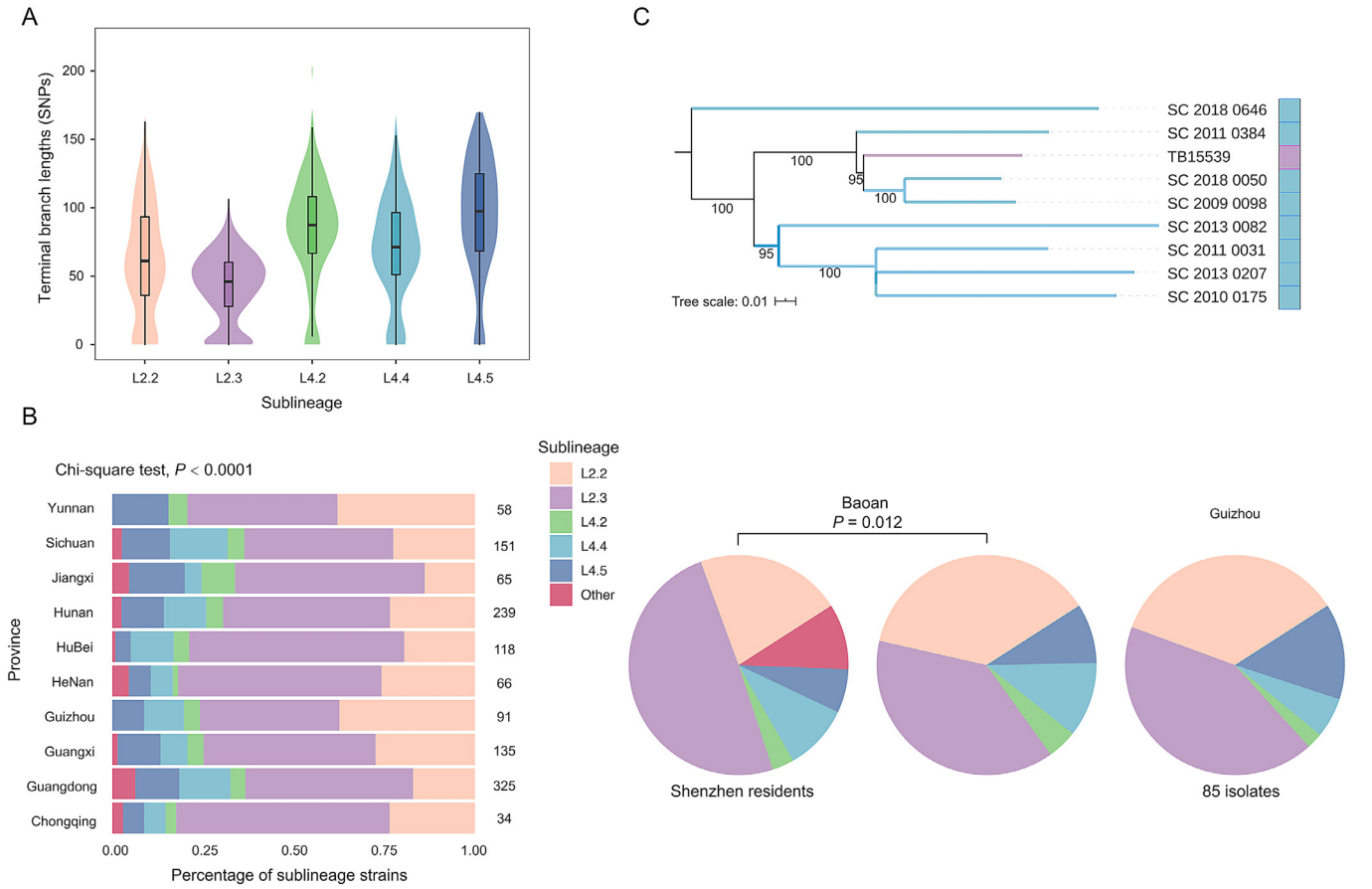

Fig. 4. The primary source of *Mycobacterium tuberculosis* in Bao'an. (a) The distribution of terminal branch lengths for different sublineages of *M. tuberculosis* in Bao'an. (b) The genotype composition of isolates from different groups of patients. There were significant differences (χ^2 = 91•631, P < 0•0001) in the genotype distribution of isolates from patients who came to Bao'an from different provinces (stack bar charts). Isolates from Bao'an patients originally from Guizhou (center pie chart) showed a significantly different genotype distribution (χ^2 = 14•697, P = 0.012) than the isolates from patients born in Shenzhen (left pie chart), but shared similar genotype distribution with isolates from TB patients in Guizhou (right pie chart). (c) The phylogenetic relationship between a strain (purple) isolated from a Bao'an patient who came from Guang'an City, Sichuan Province and eight strains (blue) isolated from patients in Wusheng County, Guang'an City, Sichuan Province. The minimal genetic distance between the Bao'an strain (TB15539) and Sichuan strains was 59 SNPs.

and then brought these strains with them to Bao'an, presumably as latent infections that reactivated to cause active disease after arriving in the city.

Finally, we took advantage of an unpublished collection of the genome sequences of 664 *M. tuberculosis* isolates from Guang'an City in Sichuan Province. A phylogenetic analysis of this collection together with the 1696 strains from Bao'an found that strain TB15539, isolated in Bao'an from a patient originally from Guang'an City, was nested within a phylogenetic subclade of the strains isolated in Guang'an City (Fig. 4c). Although it differed by 59 SNPs from the most similar strain in the Guang'an collection, it nevertheless appears to belong to a clade of strains circulating in Guang'an, suggesting that the patient acquired their infection there before coming to Bao'an.

4. Discussion

This retrospective genomic epidemiological study of isolates from pulmonary TB patients diagnosed in Bao'an, Shenzhen over three and a half years revealed that TB in this district is primarily caused by reactivation of latent infections in recent migrants, nearly half of whom developed active TB within two years after arriving in Bao'an.

Although the incidence of TB in China has been declining at about 4% per year, roughly twice the rate of global decline, tuberculosis remains an important public health problem, even in cities with modern infrastructure. With the recent urbanization and industrialization in China, there has been a large influx of migrants into rapidly growing cities where they now constitute the majority of the urban population. Although it was recognized that TB in Chinese cities is predominantly a disease of migrants, the details of the epidemiology of TB in emerging cities was not well defined. In this study we found that the percentage of TB in Bao'an attributable to recent transmission was just 12•2%, which is considerably lower than the percentage of recent transmission reported in cities such as Shanghai (32%) or New York (40%) [32,33]. Although the exact percentages may not be completely comparable, due to differences in the included samples and the methods used to assess recent transmission, the low rate nevertheless suggests that recent transmission makes a relatively small contribution to TB burden in Bao'an.

We found also ample evidence that most TB was likely due to reactivation of infections acquired by migrants in their home provinces and brought with them to Bao'an. The *M. tuberculosis* strains in Bao'an showed a median TBL of 56 SNPs, far exceeding the number of SNPs (~20) that local strains could be expected to accumulate over the 40 years of Shenzhen growth. This can be contrasted with Ho Chi Minh City, Vietnam, where recent transmission causes most TB, and the median TBL for L2 and L4 strains was only nine SNPs (interquartile range, 4–22 SNPs) (Supplementary Figure 1 and Supplementary Table 5) [28].

When the *M. tuberculosis* strains in Bao'an were considered separately according to the patients' home provinces, some of the lineage and sublineage distributions were similar to those seen in the home provinces, again suggesting that the infections were acquired before coming to Bao'an. A similar trend was seen in Italy, where significant differences in the genotype distributions were observed not only between isolates of foreign-born and local patients, but also amongst isolates of patients from different regions (i.e. Africa, Asia, Eastern Europe, Central and South America) [34]. In Finland, the genotype distribution of isolates from immigrants was also found to be similar to the distribution in strains isolated in their home countries [35]. Additionally, when we compared Bao'an strains to a collection of strains from Guang'an city in Sichuan province, we found that a strain isolated from a patient in Bao'an belonged to a clade circulating in Guang'an, the patient's hometown. If similar collections of sequenced strains from other Chinese provinces were available, it is likely that many other Bao'an isolates could be matched to clades circulating in their home regions.

In this study, 93•8% of Bao'an TB patients were migrants, nearly half of whom had lived in Bao'an for at most two years, and the percentage of cases decreased as the residence time in Bao'an increased. A similar trend emerged from a recent cohort study following immigrants to Denmark, where TB risk was highest during the first year of residence followed by a gradual decline [8]. A study in the United Kingdom found that approximately 50% of foreign-born TB cases developed the disease within the first five years after immigration [9]. Most TB develops within two year after infection [9], so it is possible that the migrants were infected shortly before coming to Shenzhen and already had subclinical disease upon arrival. It seems more probable, though, that the various types of economic and social stress experienced by the migrants led to reactivation of TB infections acquired well before coming to Shenzhen [36]. However, both recent and distant infections could be detected with routine pre-entry TB/LTBI screening and these individuals could be given appropriate preventive treatment, as is done with immigrants from high-TB-incidence countries to low-TB-incidence countries, such as the US, England, and Canada [10]. Currently, Shenzhen requires TB screening of all workers before entering employment in companies and public institutions, but there is a lack of effective screening programs for housekeeping workers and the unemployed, who represent over half of migrant TB patients.

This retrospective study had limitations. The percentage of TB due to recent transmission was probably underestimated. This is partly because the policy for most of the collection period was to culture only smear-positive sputa, while the strains responsible for smear-negative TB were never obtained. However, the comparison of basic information between smear-positive and smear-negative patients in 2017 did not find significant differences (Supplementary Table 6), suggesting that there was no obvious selection bias in the patients included in the study. However, because we only analyzed strains isolated from patients diagnosed in Bao'an, we could not detect TB transmission across different districts of the city [16]. Also, because the study lacked a comprehensive epidemiological survey of the patients, it was unable to identify geographic areas that could be hotspots for TB transmission in Shenzhen.

In conclusion, this population-based study found that the TB burden in the Bao'an District of Shenzhen, a new Chinese city that has experienced extraordinary population growth through internal migration, is predominantly caused by the endogenous reactivation of TB infections acquired by migrants in their home provinces. Nearly half of the migrant patients developed TB within two years after arrival. These results suggest that effective control strategies should include a greater effort to screen newly arrived migrants for evidence of active or latent TB, especially those working in housekeeping or unemployed. Prophylactic treatment of those found to have latent TB might be considered as a possible strategy to reduce the incidence of TB in migrants.

Contributors

T. Y., W. T., H.T. and Q. G. designed and managed the study. Y. W. and Q. J. performed the epidemiological investigation; C. H., L. W., and S. L. performed the experiments and collected the laboratory data; T. Y. and C. Z. cleaned the data and performed statistical analysis; T. Y. and Q. L. performed the sequence analysis and interpretation. T. Y., Q. J., Q. L., W. Y., H.T. and Q. G. prepared the article. All authors contributed to and gave input to the final article.

Declaration of Competing Interest

We declare no competing interest.

Acknowledgments

This work was supported by Sanming Project of Medicine in Shenzhen [grant numbers SZSM201611030], and National Science and Technology Major Project of China [grant numbers 2017ZX10201302–006 and 2018ZX10715012–005 and 2018ZX10715004]. This work also was supported by the Natural Science Foundation of China [grant numbers 91631301 and 81661128043].

Editor note: The Lancet Group takes a neutral position with respect to territorial claims in published maps and institutional affiliations.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.lanwpc.2021.100106.

References

[1] Shenzhen Statistics Bureau Shenzhen statistical yearbook 2018; 2018.
[2] Guan H, Yang Y, Tan W, Wu Q, Lv D. Effect evaluation of Shenzhen tuberculosis control and prevention programme implementation from 2001 to 2010. Chin J Antituberc 2013;29:729–37.
[3] Centers for Disease Control and Prevention. Reported tuberculosis in the United States, 2019. https://www.cdc.gov/tb/statistics/reports/2019/national_data.htm (Accessed 8 January 2021).
[4] Hanway A, Comiskey CM, Tobin K, O'Toole RF. Relating annual migration from high tuberculosis burden country of origin to changes in foreign-born tuberculosis notification rates in low-medium incidence European countries. Tuberculosis (Edinb) 2016;101:67–74.
[5] Toms C, Stapledon R, Waring J, et al. Tuberculosis notifications in Australia, 2012 and 2013. Commun Dis Intell Q Rep 2015;39:E217–35.
[6] Institute of Environmental Science and Research Ltd (ESR). Tuberculosis in New Zealand: annual report 2013. https://surv.esr.cri.nz/PDF_surveillance/AnnTBReports/ (Accessed 29 April 2020).
[7] Public Health Agency of Canada. Tuberculosis in Canada 2013: pre-release. https://www.canada.ca/en/services/health/publications/diseases-conditions.html (Accessed 29 April 2020).
[8] Kristensen KL, Ravn P, Petersen JH, et al. Long-term risk of tuberculosis among migrants according to migrant status: a cohort study. Int J Epidemiol 2020;49:776–85.
[9] Behr MA, Edelstein PH, Ramakrishnan L. Revisiting the timetable of tuberculosis. BMJ 2018;362:k2738.
[10] European Centre for Disease Prevention and Control Public health guidance on screening and vaccination for infectious diseases in newly arrived migrants within the EU/EEA. Stockholm: ECDC; 2018.
[11] Alonso Rodriguez N, Chaves F, Inigo J, et al. Transmission permeability of tuberculosis involving immigrants, revealed by a multicentre analysis of clusters. Clin Microbiol Infect 2009;15:435–42.
[12] Yang C, Lu L, Warren JL, et al. Internal migration and transmission dynamics of tuberculosis in Shanghai, China: an epidemiological, spatial, genomic analysis. Lancet Infect Dis 2018;18:788–95.
[13] Zhang J, Hu F, Wang Y, Chen W, Zheng J, Mei J. Epidemiological characteristics of tuberculosis in Bao'an district, Shenzhen, 2008-2015. China Trop Med 2018;18:128–46.

[14] Dai D-E. Study of current situation and trend of floating population in Shenzhen. Mod Bus Trade Ind 2016;37:17–18.
[15] Chinese Antituberculosis AssociationThe laboratory science procedure of diagnostic bacteriology in tuberculosis. Chin J Antituberc 1996;18.
[16] Jiang Q, Liu Q, Ji L, et al. Citywide transmission of multidrug-resistant tuberculosis under China's rapid urbanization: a retrospective population-based genomic spatial epidemiological study. Clin Infect Dis 2020;71:142–51.
[17] 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.
[18] Joshi NA, Fass JN. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files; 2011.
[19] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357–9.
[20] Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and samtools. Bioinformatics 2009;25:2078–9.
[21] Koboldt DC, Zhang Q, Larson DE, et al. Varscan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 2012;22:568–76.
[22] Yang C, Luo T, Shen X, et al. Transmission of multidrug-resistant *Mycobacterium tuberculosis* in Shanghai, China: a retrospective observational study using whole-genome sequencing and epidemiological investigation. Lancet Infect Dis 2017;17:275–84.
[23] Coll F, McNerney R, Guerra-Assuncao JA, et al. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. Nat Commun 2014;5:4812.
[24] 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.
[25] 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.
[26] Papaventsis D, Casali N, Kontsevaya I, Drobniewski F, Cirillo DM, Nikolayevskyy V. Whole genome sequencing of *Mycobacterium tuberculosis* for detection of drug resistance: a systematic review. Clin Microbiol Infect 2017;23:61–8.
[27] Kumar S, Stecher G, Li M, Knyaz C, Tamura K. Mega X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 2018;35:1547–9.
[28] 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.
[29] Walker TM, Clp CL, Harrell RH, et al. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. Lancet Infect Dis 2013;13:137–46.
[30] Ford CB, Lin PL, Chase MR, et al. Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. Nat Genet 2011;43:482–6.
[31] Bryant JM, Schurch AC, van Deutekom H, et al. Inferring patient to patient transmission of *Mycobacterium tuberculosis* from whole genome sequencing data. BMC Infect Dis 2013;13:110.
[32] Yang CG, 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.
[33] Alland D, Kalkut GE, Moss AR, et al. Transmission of tuberculosis in New York city. An analysis by DNA fingerprinting and conventional epidemiologic methods. N Engl J Med 1994;330:1710–16.
[34] Garzelli C, Lari N, Cuccu B, Tortoli E, Rindi L. Impact of immigration on tuberculosis in a low-incidence area of Italy: a molecular epidemiological approach. Clin Microbiol Infect 2010;16:1691–7.
[35] Raisanen PE, Soini H, Vasankari T, et al. Tuberculosis in immigrants in Finland, 1995-2013. Epidemiol Infect 2016;144:425–33.
[36] Salazar MA, Hu XJ. Health and lifestyle changes among migrant workers in China: implications for the healthy migrant effect. Lancet Diabetes Endo 2016;4:89–90.