Distinguishing Relapse From Reinfection With Whole-Genome Sequencing in Recurrent Pulmonary Tuberculosis: A Retrospective Cohort Study in Beijing, China

Jian Du[1†], Qing Li[1†], Min Liu[2†], Yufeng Wang[3], Zhongtan Xue[3], Fengmin Huo[4], Xuxia Zhang[1], Yuanyuan Shang[1], Shanshan Li[1*‡], Hairong Huang[4*‡] and Yu Pang[1*‡]

[1]Department of Bacteriology and Immunology, Beijing Chest Hospital, Capital Medical University/Beijing Tuberculosis and Thoracic Tumor Research Institute, Beijing, China, [2]Provincial Center for Tuberculosis Control and Prevention, Liaoning Provincial Center for Disease Control and Prevention, Shenyang, China, [3]Department of Laboratory Quality Control, Innovation Alliance on Tuberculosis Diagnosis and Treatment (Beijing), Beijing, China, [4]National Clinical Laboratory on Tuberculosis, Beijing Key Laboratory on Drug-Resistant Tuberculosis Research, Beijing Chest Hospital, Capital Medical University/Beijing Tuberculosis and Thoracic Tumor Institute, Beijing, China

Background: Tuberculosis recurrence is still a major problem for the control of tuberculosis, and the cause of the recurrence is still unclear.

Methods: We retrospectively recruited 68 pairs of samples of *Mycobacterium tuberculosis* (MTB) from recurrent TB cases in Beijing Chest Hospital between January 2008 and December 2019. The whole-genome sequencing was conducted to analyze single-nucleotide polymorphism (SNP) and to identify whether recurrent disease was due to relapse or reinfection. The BACTEC MGIT was performed to compare differences in drug susceptibility profiles between two episodes.

Results: 62 (91.2%) out of 68 confirmed recurrence were due to relapse, whereas the remaining six (8.8%) were due to reinfection. And there was a strong association between earlier relapse and underlying chronic diseases. In addition, the MTB isolates from non-diabetic patients had a higher mutation rate than those from diabetic patients. A community transmission was also identified in our cohort. Levofloxacin resistance was the most frequently observed drug resistance for 12.9% relapse cases.

Conclusion: The relapse of a previous episode in Beijing. The underlying chronic diseases are associated with an earlier TB relapse. MTB isolates were more prone to develop levofloxacin resistance than moxifloxacin resistance after FQ exposure. The patients at high-risk for relapses deserve more careful investigation.

Keywords: tuberculosis, relapse, reinfection, whole-genome sequencing, pulmonary

INTRODUCTION

Tuberculosis (TB), caused by *Mycobacterium tuberculosis* (MTB) complex, is the leading cause of death and disease from a single infectious agent and constitutes a major global public health issue (Matteelli et al., 2018; Organization, W.H, 2020a). World Health Organization guidelines recommend that patients with drug-susceptible TB should be treated with a 6-month multidrug regimen (Organization, W.H, 2020b). Although this regimen exhibits high efficacy in treatment of drug-susceptible disease, approximately 5% of these patients have a recurrence after 6 months of therapy (Lambert et al., 2003). For patients whose initial disease was multidrug-resistant, recurrence of TB was particularly common (Cox et al., 2006; Xu et al., 2018). Historically, the recurrence was attributed to endogenous relapse of disease due to incapability to differentiate isolates (Witney et al., 2017). Recent advances in genotyping technologies have allowed the discrimination of relapse with the same strain that had caused previous TB episode and reinfection by a different strain (Wells et al., 2013). Inadequate treatment is the main cause for relapse (Wells et al., 2013), whereas reinfection reveals many missed TB cases circulating in the community (Zong et al., 2018). As a consequence, efforts to distinguish relapse and reinfection are of great importance to formulate effective control strategies for TB epidemics.

Since the 1980s, a series of genomic-based methods for typing MTB strains have been developed to differentiate relapse from reinfection, including IS6110 restriction fragment length polymorphism (RFLP), spoligotyping, mycobacterial interspersed repetitive units-variable number tandem repeats (MIRU-VNTR) typing, and whole-genome sequencing (WGS) (Hermans et al., 1990; Supply et al., 2006; Bryant et al., 2013). Of these available technologies, WGS provides remarkably more information by simultaneous analysis of full genomes of multiple MTB strains, thus yielding unprecedented accuracy compared with conventional genotyping methods (Nikolayevskyy et al., 2019). Several WGS-based molecular epidemiological studies demonstrated that endogenous reactivation appeared to be the primary reason for TB recurrence in low TB-burden countries (Interrante et al., 2015), whereas reinfection may be attributed to a high proportion of TB recurrence in high TB-burden countries. These highlight the continuous interest in the geographic diversity of causes of TB recurrence of regions.

Despite great achievement in TB control and prevention, China remains one of the most high-burden countries globally. Specially, there is an unbalanced distribution of TB cases throughout the country (An et al., 2016). According to the data from nationwide survey, markedly higher incidence for TB patients is noted in the western region compared with other parts (Wang et al., 2014). In addition, the TB prevalence in rural regions is significantly higher than urban regions (Wang et al., 2014). The pre-existing disparity in TB burden has potential effect on interpretation of TB recurrences for different regions in this country. Beijing is a modernized international city in China, with an estimated TB incidence of 30.4/100,000. In spite of low incidence of TB in this municipality, more attention should be paid to patients with recurrent TB who are the source of further disease transmission in the community. However, only limited data are available on TB recurrences in our population. In order to address this concern, we retrospectively recruited 68 pairs of samples of MTB from recurrent TB cases. The WGS was conducted to identify whether recurrent disease was due to relapse or reinfection and to estimate the rate of microevolution of MTB within relapses.

MATERIALS AND METHODS

Patients and Ethics Statement

In this retrospective, observational study, we used the BioBank of Tuberculosis System, which collected standardized data of all MTB isolates from culture-positive cases, who sought healthcare in the Beijing Chest Hospital, a national TB-designated hospital in China. Eligible patients met all of the following criteria: (i) completion of anti-TB treatment following the guidelines of National Tuberculosis Control Program; (ii) successful outcomes indicated by sputum culture conversion plus clinical findings; (iii) two or more episodes of TB which occurred in the period between January 2008 and December 2019; and (iv) a minimum time interval of 12 months between two episodes (Zong et al., 2018). For patients experiencing multiple recurrences, only the first recurrent episode was included in our analysis. Data on TB cases in our cohort came from the Electronic Medical Record System in our hospital. The sociodemographic and clinical characteristics were analyzed in this study: age, sex, ethnicity, local authority of residence, TB recurrence, previous TB treatment, and site of TB disease. The study protocol was approved by the Ethics Committee of Beijing Chest Hospital, Capital Medical University and written informed consent was obtained from each participant prior to the study.

Bacteria Subculture, DNA Purification, and Species Identification

All MTB isolates were stored in 7H9 medium supplemented with 25% glycerin at −80°C refrigerator. Prior to DNA extraction and *in vitro* drug susceptibility testing (DST), the frozen bacteria were subcultured into Löwenstein Jensen (L-J) medium. After 4 weeks of incubation at 37°C, the fresh colonies were scraped from the surface of L-J medium and inactivated in a water bath at 85°C for 30 min. The bacteria in the pellet were subjected to DNA extraction by using the MasterPure™ Complete DNA and RNA Purification Kit (Epicenter, Madison, WI, United States) following the manufacturer's instructions. The quality of the extracted DNA was assessed by NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, United States). A single absorbance peak at 260 nm plus a 260/280 absorbance ratio of 1.8–2.0 indicated high-quality intact DNA. In addition, the concentration was determined by Qubit 2.0 fluorometer (Invitrogen, Thermo Fisher Scientific, United States). One microliter of purified DNA was added to reaction mixture for species identification (Zeesan Biotech, Xiamen, China). A decontamination at 50°C for 2 min was followed by initial denaturation at 95°C for 10 min, 55 cycles of denaturation at

95°C for 15 s, annealing at 57°C for 20 s, and plus a final extension at 78°C for 20 s. The amplicons were analyzed by multicolor melting curve analysis as previously reported (Xu et al., 2019; Tan et al., 2021).

Drug Susceptibility Testing

Because the previous DST results were produced by the absolution concentration method which was not endorsed by the World Health Organization (Organization, W.H, 2018), the fresh MTB cultures were collected for phenotypic DST with the BACTEC MGIT system (BD, Franklin Lakes, NJ). Briefly, the suspension was prepared to a turbidity equivalent to that of 0.5 McFarland standard; 0.5 ml of a 1:5 dilution was used for inoculation into the MGIT containing drugs, and 0.5 ml of a 1:500 dilution was used for inoculation into the drug-free control MGIT tube. After 6–14 days incubation in the MGIT 960, the results were reported automatically by comparing the number of growth units (GUs) in the control tube (GU = 400) and the number of GUs in the tubes containing drugs. The concentrations of drugs were as follows: isoniazid 0.1 μg/ml, rifampicin 1.0 μg/ml, ethambutol 5.0 μg/ml, streptomycin 1.0 μg/ml, levofloxacin 1.0 μg/ml, moxifloxacin 0.25 μg/ml, amikacin 1.0 μg/ml, and capreomycin 2.5 μg/ml. Isolates resistant to isoniazid and rifampicin were defined as multidrug-resistant tuberculosis (MDR-TB).
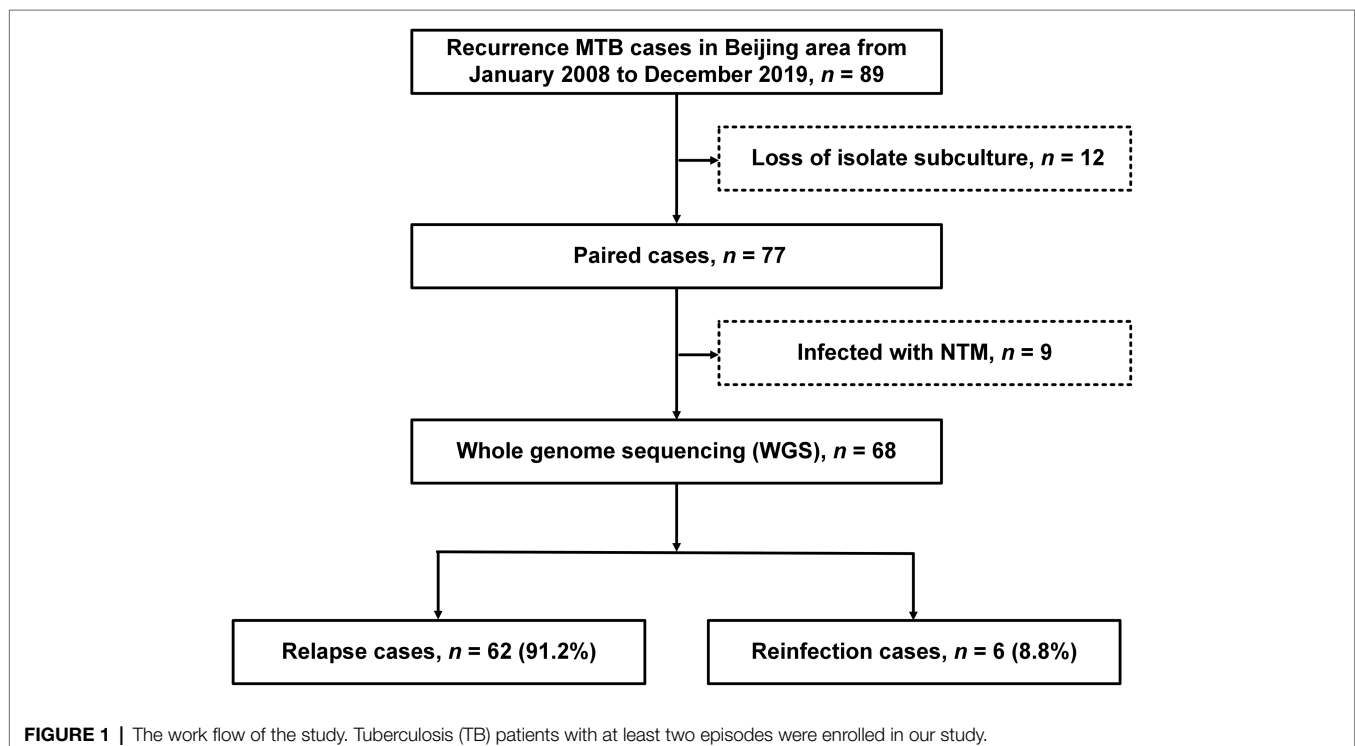
Whole-Genome Sequencing and Sequence Analysis

Whole-genome sequencing was performed using the TIANSeq direct rapid DNA library construction kit (Tiangen, China) and the NovaSeq Illumina sequencer (Illumina, United States). After quantification and quality inspection, the extracted genomic DNA was broken into nucleic acid fragments with a length of 200–300 bp by an ultrasonic interrupter (Covaris, United States) with Peak Incident Power: 175; duty factor: 10%; cycle/burst: 200; temperature: <8°C; and time: 120 s. The 5′ and 3′ protruding ends of the broken DNA fragments were flattened and connected to adaptors by T4 DNA ligase. Fragment selection and purification are performed by magnetic bead purification. An amplification method was used for enrichment and magnetic bead purification was used for purification to obtain a DNA library for sequencing. Qubit Flurometer 3.0 fluorometer (Thermo Fisher Scientific, United States) and Agilent 2100 (Agilent, United States) were used to measure the concentration and length distribution of the library, respectively. The paired-end sequencing program

TABLE 1 | Clinical data of 89 patients with recurrent pulmonary tuberculosis.

Characteristics	All patients (n = 89)
Median age (range), year	51.0 (16.0–86.0)
Male sex, n (%)	68 (76.4)
Cavity, n (%)	69 (77.5)
Complications, n (%)	
Diabetes	29 (32.6)
Malnutrition	16 (18.0)
Anemia	14 (15.7)
Autoimmune disease	4 (4.5)
Viral hepatitis	4 (4.5)
Malignant tumor	3 (3.4)
Other chronic diseases*	19 (21.4)

*Other chronic diseases include Cardiovascular and Cerebrovascular Diseases, Chronic Respiratory Diseases, Renal Failure, Liver Cirrhosis, excluding Diabetes and Malignant Tumors.
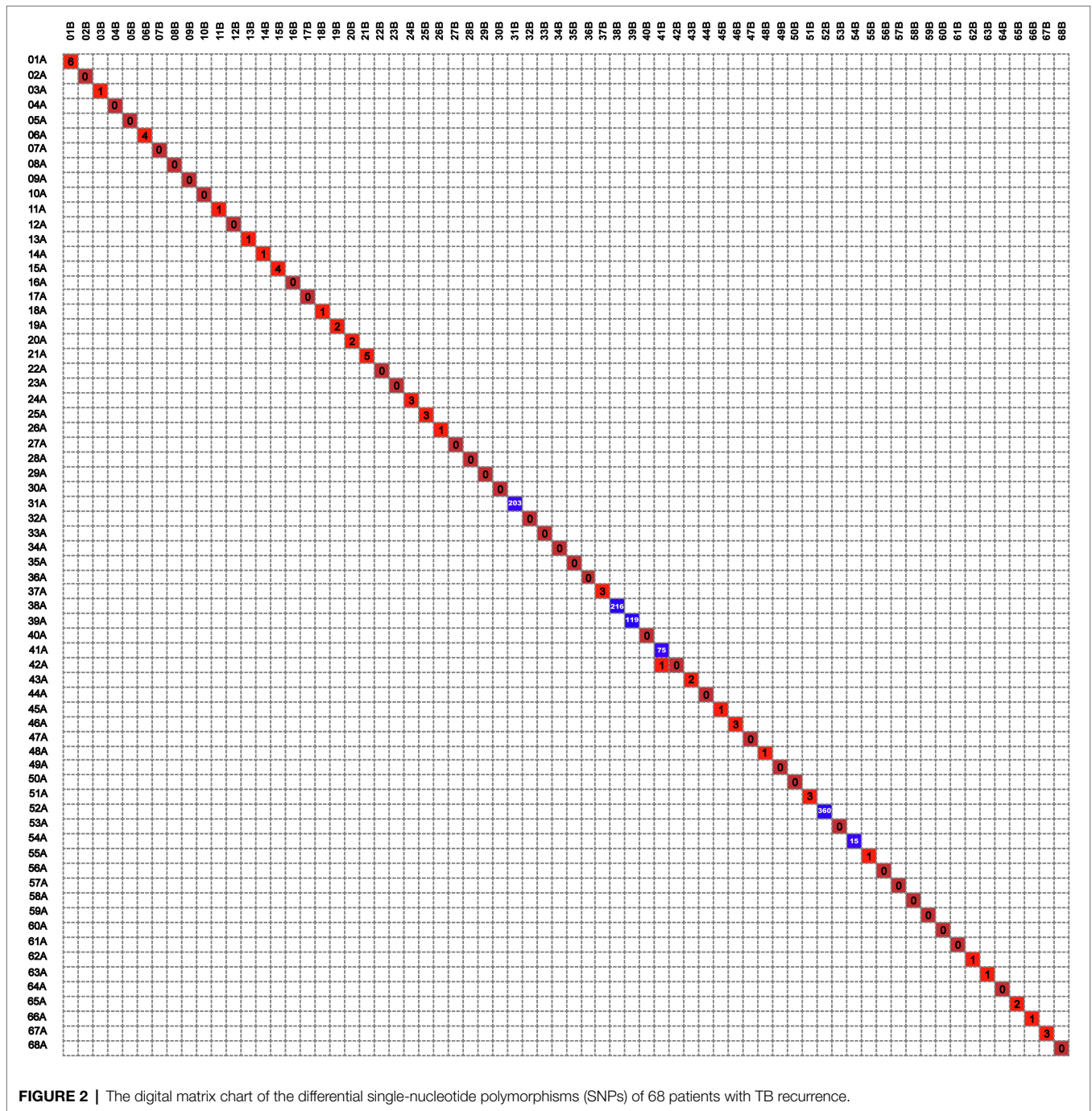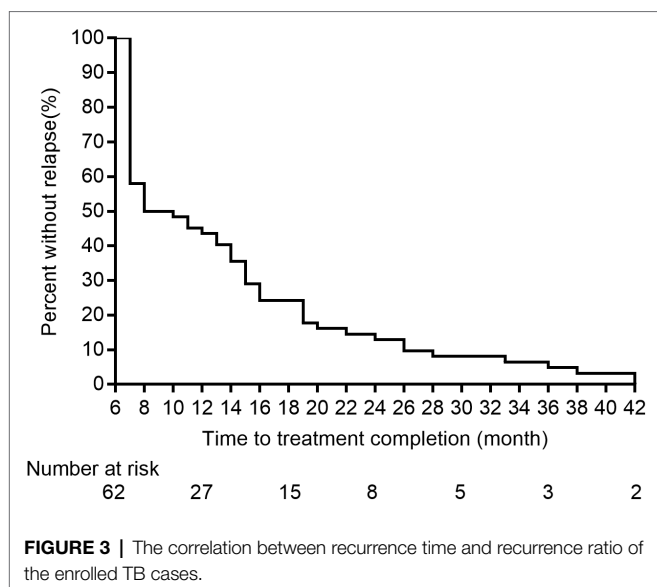

FIGURE 1 | The work flow of the study. Tuberculosis (TB) patients with at least two episodes were enrolled in our study.

FIGURE 2 | The digital matrix chart of the differential single-nucleotide polymorphisms (SNPs) of 68 patients with TB recurrence.

(PE150) for the qualified library was performed on the NovaSeq Illumina sequencing platform. FastQC quality control software was used for quality control and removal of low-quality data in Raw Data, Quality filter: q, qualified_quality_phred, set the base quality value to be no less than a qualified base, the default base quality value is 15, that is, the default base quality ≥15 is qualified base, <15 is unqualified Base. u, unqualified_percent_limit, set the allowable percentage of unqualified bases, remove this read, the default is 40, that is, when the default percentage of unqualified bases >40%,

remove the read; Q, disable_quality_filtering, set this parameter to disable the default quality filtering parameters (q, u). Length filter: l, length required, set the minimum length of read, the default is 15, that is, reads with length <15 are removed; L, length limit, sets the maximum length of read, the default is 0, there is no maximum length limit; and Low complexity filtering: Y, complexity threshold, set the complexity filter threshold of read, the default is 30. The MTB whole-genome sequence analysis platform (SAM-TB) software system was used to compare the sequencing data with MTB H37Rv

FIGURE 3 | The correlation between recurrence time and recurrence ratio of the enrolled TB cases.

whole-genome sequences and evaluate the sequencing depth and coverage. FastQC was used for quality control of SAM-TB, *via* adopting Cutadapt (v1.15) software to remove linkers and low-quality sequences. The retained high-quality sequence was aligned to the reference genome H37Rv (NC_000962.2) using BWA (v0.7.15). After removing duplicate reads with Picard (v2.0.1), we used SAMtools (v1.6)/VarScan (v2.3.6) to identify SNPs and short indels, and used Delly (v0.8.7) software to identify larger deletions (≥50 bp). By comparison with the reference genome H37Rv, a series of information, such as single-nucleotide polymorphism (SNP) and insertion–deletion (InDel) of the target genome, were obtained and a mutation matrix of all the mutation sites of each strain was generated. SNP distance matrix, cluster strains, and genetic distance were analyzed according to the mutation matrix. SNP differences were determined between the pairs of isolates using WGS. In addition, PE/PPEs/PGRs genes, insert sequences, and other highly repetitively related positions are filtered. Reads of these regions are usually mapped to the wrong genomic coordinates, which will lead to misrecognition of SNPs. Therefore, these regions should not be considered in SNP calling. and recurrence was classified as being due to relapse if there was a difference of ≤6 SNPs (Cock et al., 2010; Jiang et al., 2011; Chen et al., 2020; Shanmugam et al., 2021). A minimum spanning tree was constructed based on pairwise whole-genome SNP differences using BioNumerics (Version5. 0, Applied Maths, Sint-Martens-Latem, Belgium) software (Mazuet et al., 2017; Rouard et al., 2019). Genotypic drug resistance prediction was predicted using the TB-Profiler (Faksri et al., 2019). The average variant depth for each isolate was ≥10 and the variant calls for drug resistance were based on the algorithms of TB-Profiler. Resistance was predicted for a total of 14 anti-TB drugs, including rifampicin, isoniazide, ethambutol, pyrazinamide, and streptomycin, and the second-line drugs Linezolid, clofazimine, bedaquiline, fluoroquinolone, capreomycin, amikacin, kanamycin, PAS, and ethionamide.

Statistical Analysis

For the patients' information, Epidata 3.0 software with dual-entry method was used to establish the database, and SPSS 22.0 (SPSS Inc.)[1] statistical software was used to perform statistical analysis on the data. For continuous variables that were normally distributed, the mean ± standard deviation (SD) was used; for data that were not normally distributed, the median [interquartile range; M (IQR)] was used. Counting data were described in terms of rate or composition ratio (%). Univariate and multivariate unconditional logistic regression models were used to determine the influencing factors of patients' recurrence. Wilcoxon rank sum test was used to evaluate gene mutation rate. The Kaplan–Meier curve was generated to describe and compare the time to relapse between the different groups. $p < 0.05$ was considered statistically significant.

RESULTS

Description of the Study Population

Between January 2008 and December 2019, a total of 89 patients meeting the inclusion criteria were included in our analysis. The majority of patients was adult men with a median age of 51 years (range: 16–86 years). Of the 89 subjects in the study, diabetes ($n = 29$, 32.6%) was the most frequently observed comorbidity, followed by chronic disease ($n = 19$, 21.4%), malnutrition ($n = 16$, 18.0%), and anemia ($n = 14$, 15.7%; **Table 1**). **Figure 1** illustrates a flowchart of the enrolled samples. Among 89 recurrent TB patients, 12 (13.5%) were excluded due to subculture failure of any paired isolate. Isolates from the remaining 77 cases were analyzed by the WGS to distinguish relapse and reinfection. Of 77 cases, nine were excluded from analysis due to identification of nontuberculous mycobacteria (NTM) by WGS and multicolor melting curve analysis, leaving 68 patients with paired patients in final analysis.

Relapse and Reinfection

Lineage 2 was predominant among recurrent TB patients in the present study, accounting for 97.0% (132/136) of MTB strains, followed by Lineage 4 (3.0%, 4/136, **Supplementary Figure S1**). A comparison with SNP differences is shown in **Figure 2**. Two main groups were identified: 62 pairs of isolates had six or fewer SNP differences, and six pairs of samples had a dramatically higher number of SNP differences (range: 15–360). Ultimately, of the confirmed recurrence, 62 (91.2%) were due to relapse (**Supplementary Table 1**), whereas the remaining six (8.8%) were due to reinfection. Of 62 relapses, 35 (56.5%) occurred within 12 months of treatment completion and 54 (87.1%) within 24 months (**Figure 3**). We further assessed the time intervals to TB relapse stratified to drug susceptibility profiles and comorbidities. As shown in **Figure 4**, there was a strong association between earlier relapse and underlying chronic diseases ($p = 0.033$). In addition, no significant differences were noted between MDR- and non-MDR-TB groups.
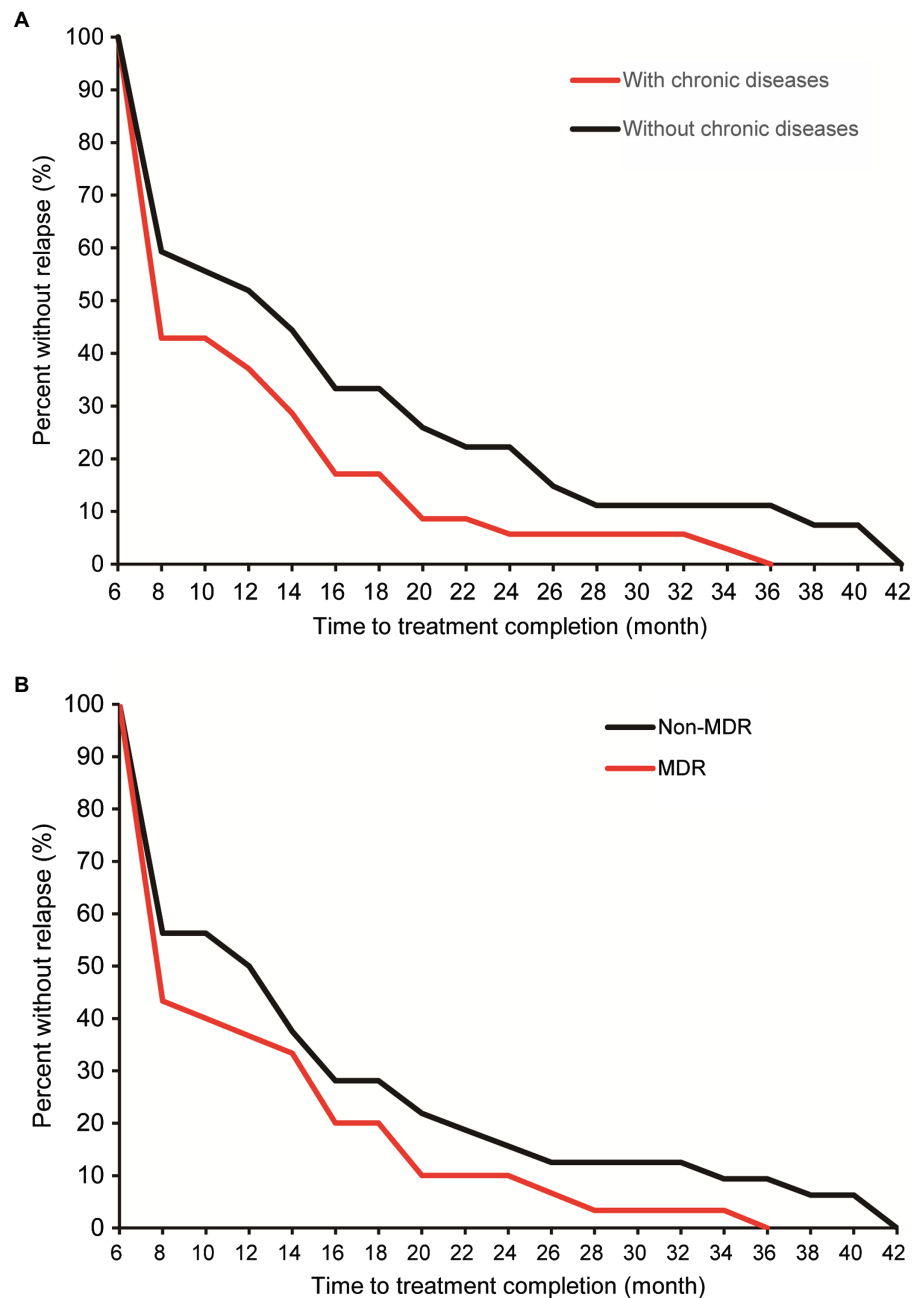
FIGURE 4 | Cumulative percent of confirmed relapse tuberculosis by comorbidity **(A)** and multidrug-resistant tuberculosis (MDR-TB; **B**).

Recurrence Due to Community Transmission

Interestingly, there was one isolate that mapped closely to other patients isolates on the tree, and this merited closer attention to see if there were nosocomial infection or community transmission. As demonstrated in **Figure 2**, the isolate of second episode of Patient 41 showed high sequence similarity to that of Patient 42, and only one SNP was found between two isolates from different individuals. However, the two isolates of patient 41 differed by 75 SNPs, suggesting recurrence due

to reinfection. We firstly reviewed their hospitalization records. Nosocomial infection was excluded since no overlapping hospitalization period occurred. Of note, both two patients lived in the neighboring blocks. Thus, we conducted interview to determine whether the community transmission could contribute to this cluster. Epidemiological data revealed that these two retired elderly individuals routinely completed morning exercises in the same park. This combined with their genotyping data supported the potential community transmission between Patient 41 and Patient 42 (**Figure 5**).

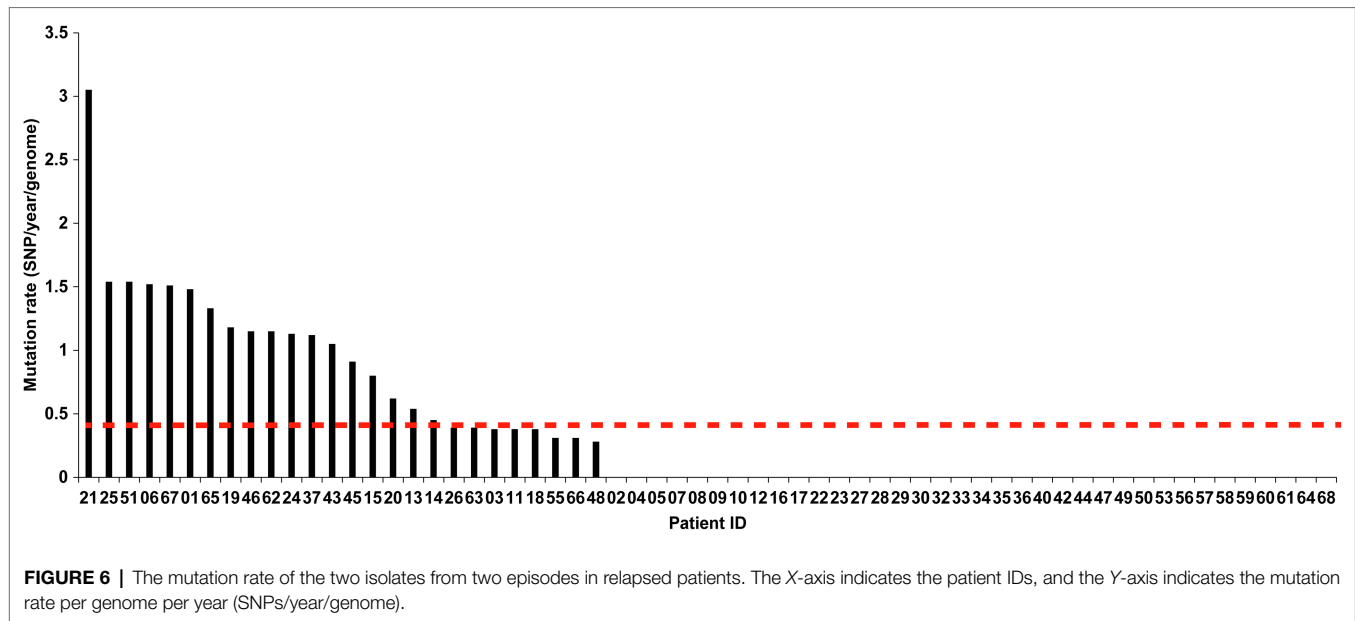**FIGURE 5 |** Transmission network based on epidemiological links.
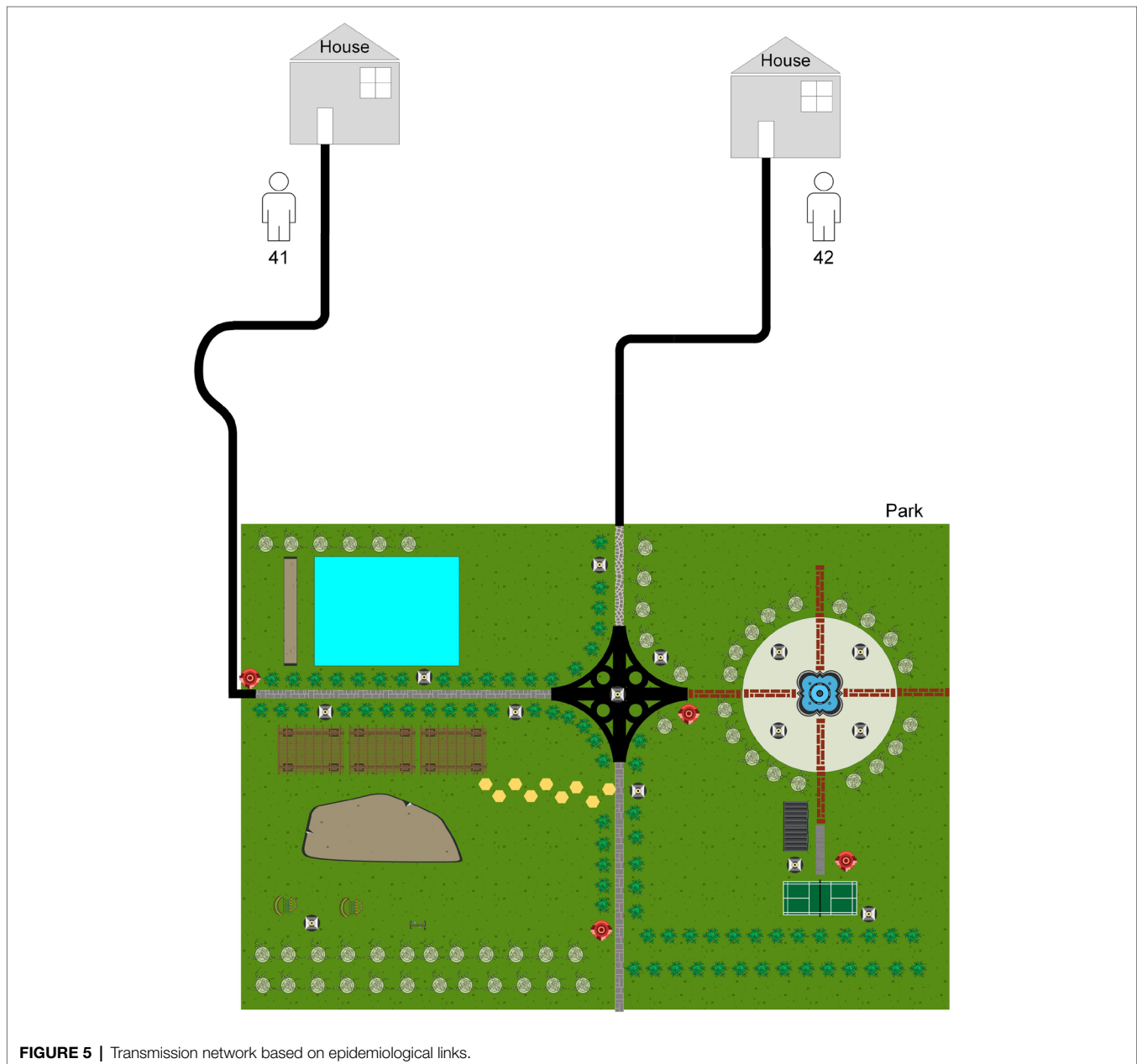
Microevolution of MTB Isolates

We further analyzed the rate of microevolution of MTB from the first and second episodes of the longitudinally sampled patients (**Figure 6**). A rate of change in DNA sequences of 0.40 SNPs per genome per year [95% confidential interval (CI) 0.24–0.56] was noted in our study. We recorded a significant higher mutation rate in non-diabetic patients (0.53 SNPs per genome per year) compared with that in diabetic patients (0.14 SNPs per genome per year, $p = 0.012$). By contrast, no difference was observed in mutation rate between MDR-TB and non-MDR-TB patients (0.39 SNPs per genome per year for MDR-TB vs. 0.42 SNPs per genome per year for non-MDR-TB, $p = 0.919$).

Emergence of Drug Resistance After Relapse

We further analyzed the differences in drug susceptibility profiles between first and second TB episode. As shown in **Figure 7**, levofloxacin resistance was the most frequently observed for 8 (12.9%) of 62 relapse cases, followed by rifampicin (9.7%, $n = 6$), streptomycin (9.7%, $n = 6$), and amikacin (9.7%, $n = 6$). For moxifloxacin, another fluoroquinolone (FQ), 2 (3.2%) paired isolates exhibited changes from sensitivity to resistance. Statistical analysis revealed that this proportion was significantly lower than that for levofloxacin ($p = 0.048$), indicating that MTB isolates were more prone to develop levofloxacin resistance than moxifloxacin resistance after FQ exposure. A deep analysis

FIGURE 6 | The mutation rate of the two isolates from two episodes in relapsed patients. The *X*-axis indicates the patient IDs, and the *Y*-axis indicates the mutation rate per genome per year (SNPs/year/genome).

was conducted to compare the drug susceptibility profiles between two episodes by WGS. As shown in **Supplementary Table 2**, FQ resistance was recorded in 22.6% (14/62) of relapse cases, followed by kanamycin (11.3%, 7/62), and pyrazinamide (8.1%, 5/62). In contrast, no bedaquiline and linezolid resistance were noted in these relapse cases.

DISCUSSION

To our knowledge, this is one of the largest studies investigating the recurrence of TB with WGS in China. The current study confirmed that the recurrence was majorly due to the relapse of a previous episode and that exogenous reinfection approximately accounted for 9% of recurrent cases. The extent to which exogenous reinfection occurs depends on the prevalence of pulmonary TB, and the greater likelihood of exogenous reinfection is associated with higher prevalence of this diseases (Bandera et al., 2001). In low-incidence countries, researchers have found the percentage of reinfection ranging from 4% in the United States and Canada to 23% in Italy (Jasmer et al., 2004; Schiroli et al., 2015). By contrast, reinfection was the major reason for the recurrent cases, ranging from 17% in Georgia to 69% in South Africa (Charalambous et al., 2008; Maghradze et al., 2019). In studies of MTB isolates circulating in China, the rate of reinfection among pulmonary TB patients exhibited great diversity, ranging from 47% in Jiangsu to 61% in Shanghai (Shen et al., 2006; Shao et al., 2021). This significant differences may be partially explained by the varying TB incidences across geographic regions and time. In addition, a molecular epidemiological study by Shen et al. (2017) revealed 42% of recurrent cases had different genotypes, indicating that the immunity conferred by initial infection is of importance to protect the individuals from reinfection by the same genotypes. Therefore, we speculate that the predominant prevalence of

Beijing genotype serves as another potential contributor to low proportion of reinfection in our observation.

Another interesting finding of our study was a community transmission of pulmonary TB *via* WGS-based clustering was identified. The neighboring blocks of residence of two patients within this cluster attracted our attention. Further epidemiological survey revealed that there was temporal-spatial overlap between Patient 41 and Patient 42 during morning exercises. On the one hand, our results underline the great importance of monitoring the recurrence to timely recognize index cases and block transmission. A cost-effectiveness concern is raised how long we should follow up to monitor the recurrence. In consistent with previous findings, approximate 56.5% of relapses occurred within 12 months of treatment completion. Therefore, the patients at high-risk for relapses deserve more careful investigation. On the other hand, exogenous reinfection in Patient 41 emphasizes the host susceptibility to tubercle bacilli. Further experimental studies are required to elucidate the molecular mechanisms in susceptible populations.

The emergence of drug resistance in MTB after relapse undermines the efficacy of subsequent anti-TB treatment. In this study, we noted that levofloxacin resistance is the most frequently observed drug resistance among second episodes. Several potential explanations could be responsible for the high levofloxacin resistance rate. First, empirical treatment with a fluoroquinolone is always provided for patients with suspected pulmonary infection due to its potent activity against pathogens in China (Zhang et al., 2014). Therefore, the abuse and misuse of FQs may play an important role in the accumulation of drug-resistant MTB in this country. In line with our findings, numerous previous reports have demonstrated dramatically increased prevalence of FQ-resistant MTB during the past decades in China (Pang et al., 2017; Huo et al., 2020). Second, previous experimental evidence showed that FQs are effective inducers of the SOS DNA repair system (Schröder et al., 2013). Consequently, the boosted hydrogen
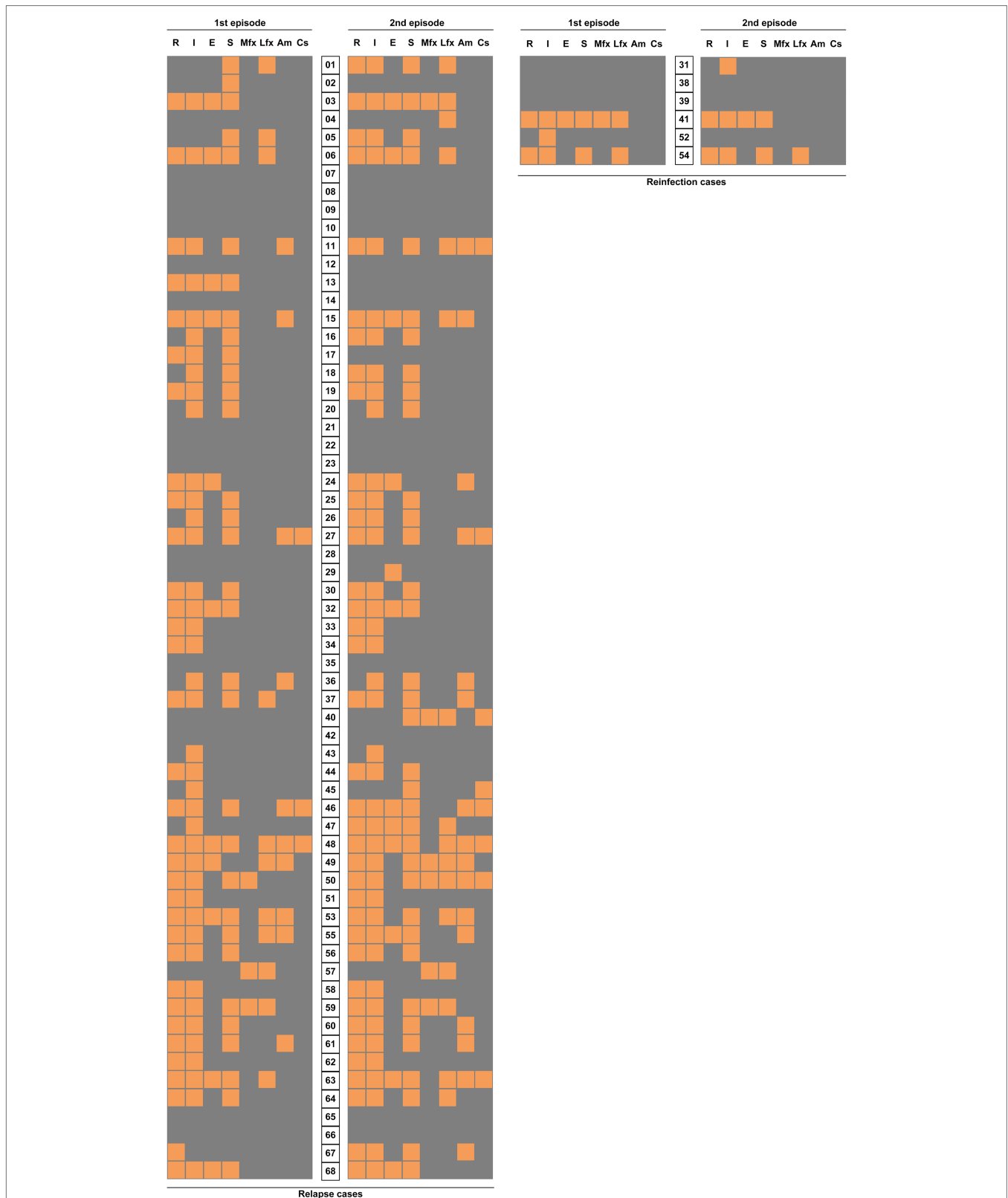
FIGURE 7 | Comparison of drug sensitivity distribution between the two TB episodes. Each grid represents the result of drug susceptibility test with isolates from two episodes of each patient. The results are shown in an orange-gray color scale, where orange represents resistance to the corresponding drug and gray represents susceptivity to the corresponding drug. R, rifampicin; I, isoniazid; E, ethambutol; S, streptomycin; Mfx, moxifloxacin; Lfx, levofloxacin; Am, amikacin; and Cs, capreomycin.

peroxide production increases the spontaneous mutation rate of tubercle bacilli, thereby resulting in the emergence of FQ resistance. Taken together, more actions are required to minimize inappropriate usage of FQs in clinical practice, which is essential for control of antibiotic-resistant bacteria.

Findings in a primate model with active and latent TB define a molecular clock rate of ∼0.3–0.5 SNPs/year (Ford et al., 2011), which are consistent with the estimated 0.3–0.5 SNP changes per year per genome from epidemiological studies (Pérez-Lago et al., 2014; Guerra-Assunção et al., 2015a). Based on our data, a slightly higher rate (∼0.5 SNPs/year) was noted in our relapse cohort. We propose two possible explanations for the diversity. First, MTB isolates from linage 2 (East Asian lineage and Beijing sublineage) had higher mutation rate than isolates from lineage 4, accounting for the optimal adaptability and an emergence ratio of drug resistance (Guerra-Assunção et al., 2015b). The predominance of Beijing genotype in our study may be associated with higher mutation rate among relapse cases. Alternatively, a recent study by Colangeli et al. (2020) demonstrated a significant difference in mutation rate between two phases of latent infection in humans, compromising of early latency with high mutation rate and late latency with slow mutation rate. Given the fact that 87.1% of patients in our study relapsed within 2 years, the tubercle bacilli were more properly be considered an incubation period of early latency, thus resulting in a higher mutation rate.

Our analysis suggested that the dormant tubercle bacilli in non-diabetic patients harbored a higher mutation rate than diabetic patients. Recently, Liu et al. identified elevated bacterial mutation rates in MTB isolates from HIV-negative but not HIV-positive individuals, indicating that the *in vivo* molecular clock for MTB was modulated by host immune (Liu et al., 2020). This phenomenon is consistent with the growing knowledge of impaired immunity in diabetic hosts, who produce decreased ROS stress compared with non-diabetic hosts (Newsholme et al., 2007; Rendra et al., 2019). This heterogeneity perhaps helps explain why isolates from non-diabetic hosts accumulate higher numbers of SNPs between the first and second episode.

We also acknowledged several obvious limitations to our study. First, despite the enrollment of all paired isolates from relapses, the relative sample size weakened the significance of our conclusion. Second, due to the predominance of relapse in our cohort, it is difficult to analyze the risk factors associated with relapse and reinfections. Third, crude mutation rates of MTB isolates were estimated in the present study using the mixed population of tubercle bacilli rather than isolation of single colonies. Therefore, this mutation rate may be underestimated. Fourth, we observed that some samples showed loss of antibiotic resistance in the second episode. In addition to the viability of the bacteria, the potential laboratory errors may be an explanation for this phenomenon despite performing quality control on each batch. Finally, the short posttreatment relapse interval may lead to the underestimation of the prevalence of reinfection among our recurrent patients considering that MTB is characterized by a low mutation rate (Eldholm and Balloux, 2016). Nevertheless, our study provides new insights to help formulate effective strategies against TB recurrence in a Chinese metropolis city with low TB incidence.

CONCLUSION

Our data demonstrate that the recurrence is majorly due to the relapse of a previous episode in Beijing. The underlying chronic diseases are associated with an earlier TB relapse. Levofloxacin resistance is the most frequently observed drug resistance among second episodes. In addition, the dormant tubercle bacilli in non-diabetic patients harbored a higher mutation rate than diabetic patients. An identified community transmission highlights that the patients at high-risk for relapses deserve more careful investigation.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committees of the Beijing Chest Hospital, Capital Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YP, HH, SL, and JD: conceptualization. QL, ZX, and JD: data curation. QL and ML: formal analysis. ML, YW, ZX, FH, XZ, and YS: methodology. YP and JD: project administration. JD, QL, YW, and ZX: writing - original draft. YP, JD, HH, ML, and SL: writing - review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at https://www.frontiersin.org/articles/10.3389/fmicb.2021.754352/full#supplementary-material

REFERENCES

An, J., Gao, M., Chu, N., Huang, H., Pang, Y., and Li, L. (2016). Transregional movement of multidrug-resistant tuberculosis in North China: an underlying threat to tuberculosis control. *Sci. Rep.* 6:29727. doi: 10.1038/srep29727

Bandera, A., Gori, A., Catozzi, L., Degli Esposti, A., Marchetti, G., Molteni, C., et al. (2001). Molecular epidemiology study of exogenous reinfection in an area with a low incidence of tuberculosis. *J. Clin. Microbiol.* 39, 2213–2218. doi: 10.1128/JCM.39.6.2213-2218.2001

Bryant, J. M., Harris, S. R., Parkhill, J., Dawson, R., Diacon, A. H., van Helden, P., et al. (2013). Whole-genome sequencing to establish relapse or re-infection with *Mycobacterium tuberculosis*: a retrospective observational study. *Lancet Respir. Med.* 1, 786–792. doi: 10.1016/S2213-2600(13)70231-5

Charalambous, S., Grant, A. D., Moloi, V., Warren, R., Day, J. H., van Helden, P., et al. (2008). Contribution of reinfection to recurrent tuberculosis in south African gold miners. *Int. J. Tuberc. Lung Dis.* 12, 942–948

Chen, X., He, G., Lin, S., Wang, S., Sun, F., Chen, J., et al. (2020). Analysis of serial multidrug-resistant tuberculosis strains causing treatment failure and within-host evolution by whole-genome sequencing. *mSphere* 5:e00884-20. doi: 10.1128/mSphere.00884-20

Cock, P. J., Fields, C. J., Goto, N., Heuer, M. L., and Rice, P. M. (2010). The sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res.* 38, 1767–1771. doi: 10.1093/nar/gkp1137

Colangeli, R., Gupta, A., Vinhas, S. A., Chippada Venkata, U. D., Kim, S., Grady, C., et al. (2020). *Mycobacterium tuberculosis* progresses through two phases of latent infection in humans. *Nat. Commun.* 11:4870. doi: 10.1038/s41467-020-18699-9

Cox, H., Kebede, Y., Allamuratova, S., Ismailov, G., Davletmuratova, Z., Byrnes, G., et al. (2006). Tuberculosis recurrence and mortality after successful treatment: impact of drug resistance. *PLoS Med.* 3:e384. doi: 10.1371/journal.pmed.0030384

Eldholm, V., and Balloux, F. (2016). Antimicrobial resistance in *Mycobacterium tuberculosis*: the odd one out. *Trends Microbiol.* 24, 637–648. doi: 10.1016/j.tim.2016.03.007

Faksri, K., Kaewprasert, O., Ong, R. T., Suriyaphol, P., Prammananan, T., Teo, Y. Y., et al. (2019). Comparisons of whole-genome sequencing and phenotypic drug susceptibility testing for *Mycobacterium tuberculosis* causing MDR-TB and XDR-TB in Thailand. *Int. J. Antimicrob. Agents* 54, 109–116. doi: 10.1016/j.ijantimicag.2019.04.004

Ford, C. B., Lin, P. L., Chase, M. R., Shah, R. R., Iartchouk, O., Galagan, J., et al. (2011). Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. *Nat. Genet.* 43, 482–486. doi: 10.1038/ng.811

Guerra-Assunção, J. A., Crampin, A. C., Houben, R. M., Mzembe, T., Mallard, K., Coll, F., et al. (2015a). Large-scale whole genome sequencing of *M. tuberculosis* provides insights into transmission in a high prevalence area. *elife* 4:e05166. doi: 10.7554/eLife.05166

Guerra-Assunção, J. A., Houben, R. M., Crampin, A. C., Mzembe, T., Mallard, K., Coll, F., et al. (2015b). Recurrence due to relapse or reinfection with *Mycobacterium tuberculosis*: a whole-genome sequencing approach in a large, population-based cohort with a high HIV infection prevalence and active follow-up. *J. Infect. Dis.* 211, 1154–1163. doi: 10.1093/infdis/jiu574

Hermans, P. W., van Soolingen, D., Dale, J. W., Schuitema, A. R., McAdam, R. A., Catty, D., et al. (1990). Insertion element IS986 from *Mycobacterium tuberculosis*: a useful tool for diagnosis and epidemiology of tuberculosis. *J. Clin. Microbiol.* 28, 2051–2058. doi: 10.1128/jcm.28.9.2051-2058.1990

Huo, F., Zhang, F., Xue, Y., Shang, Y., Liang, Q., Ma, Y., et al. (2020). Increased prevalence of levofloxacin-resistant *Mycobacterium tuberculosis* in China is associated with specific mutations within the gyrA gene. *Int. J. Infect. Dis.* 92, 241–246. doi: 10.1016/j.ijid.2020.01.021

Interrante, J. D., Haddad, M. B., Kim, L., and Gandhi, N. R. (2015). Exogenous reinfection as a cause of late recurrent tuberculosis in the United States. *Ann. Am. Thorac. Soc.* 12, 1619–1626. doi: 10.1513/AnnalsATS.201507-429OC

Jasmer, R. M., Bozeman, L., Schwartzman, K., Cave, M. D., Saukkonen, J. J., Metchock, B., et al. (2004). Recurrent tuberculosis in the United States and Canada: relapse or reinfection? *Am. J. Respir. Crit. Care Med.* 170, 1360–1366. doi: 10.1164/rccm.200408-1081OC

Jiang, L., Schlesinger, F., Davis, C. A., Zhang, Y., Li, R., Salit, M., et al. (2011). Synthetic spike-in standards for RNA-seq experiments. *Genome Res.* 21, 1543–1551. doi: 10.1101/gr.121095.111

Lambert, M. L., Hasker, E., Van Deun, A., Roberfroid, D., Boelaert, M., and Van der Stuyft, P. (2003). Recurrence in tuberculosis: relapse or reinfection? The lancet. *Infect. Dis.* 3, 282–287. doi: 10.1016/s1473-3099(03)00607-8

Liu, Q., Wei, J., Li, Y., Wang, M., Su, J., Lu, Y., et al. (2020). *Mycobacterium tuberculosis* clinical isolates carry mutational signatures of host immune environments. *Sci. Adv.* 6:eaba4901. doi: 10.1126/sciadv.abe3717

Maghradze, N., Jugheli, L., Borrell, S., Tukvadze, N., Aspindzelashvili, R., Avaliani, Z., et al. (2019). Classifying recurrent *Mycobacterium tuberculosis* cases in Georgia using MIRU-VNTR typing. *PLoS One* 14:e0223610. doi: 10.1371/journal.pone.0223610

Matteelli, A., Rendon, A., Tiberi, S., Al-Abri, S., Voniatis, C., Carvalho, A. C. C., et al. (2018). Tuberculosis elimination: where are we now? *Eur. Respir. Rev.* 27:180035. doi: 10.1183/16000617.0035-2018

Mazuet, C., Legeay, C., Sautereau, J., Bouchier, C., Criscuolo, A., Bouvet, P., et al. (2017). Characterization of clostridium baratii type F strains responsible for an outbreak of botulism linked to beef meat consumption in France. *PLoS Curr.* 9:e1-16. doi: 10.1371/currents.outbreaks.6ed2fe754b58a5c42d0c3 3d586ffc606

Newsholme, P., Haber, E. P., Hirabara, S. M., Rebelato, E. L., Procopio, J., Morgan, D., et al. (2007). Diabetes associated cell stress and dysfunction: role of mitochondrial and non-mitochondrial ROS production and activity. *J. Physiol.* 583, 9–24. doi: 10.1113/jphysiol.2007.135871

Nikolayevskyy, V., Niemann, S., Anthony, R., van Soolingen, D., Tagliani, E., Ködmön, C., et al. (2019). Role and value of whole genome sequencing in studying tuberculosis transmission. *Clin. Microbiol. Infect.* 25, 1377–1382. doi: 10.1016/j.cmi.2019.03.022

Organization, W.H (2018). *Technical Manual for Drug Susceptibility Testing of Medicines Used in the Treatment of Tuberculosis*. Geneva: World Health Organization.

Organization, W.H (2020a). *Global tuberculosis report 2020*. Geneva: World Health Organization.

Organization, W.H (2020b). *WHO Consolidated Guidelines on Tuberculosis: Module 4: Treatment - Drug-Resistant Tuberculosis Treatment*. Geneva: World Health Organization.

Pang, Y., Zhang, Z., Wang, Y., Wang, S., Song, Y., Zhao, B., et al. (2017). Genotyping and prevalence of pyrazinamide- and moxifloxacin-resistant tuberculosis in China, 2000 to 2010. *Antimicrob. Agents Chemother.* 61:e02170-16. doi: 10.1128/AAC.02170-16

Pérez-Lago, L., Comas, I., Navarro, Y., González-Candelas, F., Herranz, M., Bouza, E., et al. (2014). Whole genome sequencing analysis of intrapatient microevolution in *Mycobacterium tuberculosis*: potential impact on the inference of tuberculosis transmission. *J. Infect. Dis.* 209, 98–108. doi: 10.1093/infdis/jit439

Rendra, E., Riabov, V., Mossel, D. M., Sevastyanova, T., Harmsen, M. C., and Kzhyshkowska, J. (2019). Reactive oxygen species (ROS) in macrophage activation and function in diabetes. *Immunobiology* 224, 242–253. doi: 10.1016/j.imbio.2018.11.010

Rouard, C., Bourgeois-Nicolaos, N., Rahajamanana, L., Romain, O., Pouga, L., Derouin, V., et al. (2019). Evaluation of an 'all-in-one' seven-day whole-genome sequencing solution in the investigation of a *Staphylococcus aureus* outbreak in a neonatal intensive care unit. *J. Hosp. Infect.* 102, 297–303. doi: 10.1016/j.jhin.2019.01.029

Schiroli, C., Carugati, M., Zanini, F., Bandera, A., Di Nardo Stuppino, S., Monge, E., et al. (2015). Exogenous reinfection of tuberculosis in a low-burden area. *Infection* 43, 647–653. doi: 10.1007/s15010-015-0759-9

Schröder, W., Goerke, C., and Wolz, C. (2013). Opposing effects of aminocoumarins and fluoroquinolones on the SOS response and adaptability in *Staphylococcus aureus*. *J. Antimicrob. Chemother.* 68, 529–538. doi: 10.1093/jac/dks456

Shanmugam, S., Bachmann, N. L., Martinez, E., Menon, R., Narendran, G., Narayanan, S., et al. (2021). Whole genome sequencing based differentiation between re-infection and relapse in Indian patients with tuberculosis recurrence, with and without HIV co-infection. *Int. J. Infect. Dis.* doi: 10.1016/j.ijid.2021.03.020

Shao, Y., Song, H., Li, G., Li, Y., Li, Y., Zhu, L., et al. (2021). Relapse or re-infection, the situation of recurrent tuberculosis in eastern China. *Front. Cell. Infect. Microbiol.* 11:638990. doi: 10.3389/fcimb.2021.638990

Shen, G., Xue, Z., Shen, X., Sun, B., Gui, X., Shen, M., et al. (2006). The study recurrent tuberculosis and exogenous reinfection, Shanghai, China. *Emerg. Infect. Dis.* 12, 1776–1778. doi: 10.3201/eid1211.051207

Shen, X., Yang, C., Wu, J., Lin, S., Gao, X., Wu, Z., et al. (2017). Recurrent tuberculosis in an urban area in China: relapse or exogenous reinfection? *Tuberculosis* 103, 97–104. doi: 10.1016/j.tube.2017.01.007

Supply, P., Allix, C., Lesjean, S., Cardoso-Oelemann, M., Rüsch-Gerdes, S., Willery, E., et al. (2006). Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis. J. Clin. Microbiol.* 44, 4498–4510. doi: 10.1128/JCM.01392-06

Tan, Y., Deng, Y., Yan, X., Liu, F., Tan, Y., Wang, Q., et al. (2021). Nontuberculous mycobacterial pulmonary disease and associated risk factors in China: a prospective surveillance study. *J. Infect.* 83, 46–53. doi: 10.1016/j.jinf.2021.05.019

Wang, L., Zhang, H., Ruan, Y., Chin, D. P., Xia, Y., Cheng, S., et al. (2014). Tuberculosis prevalence in China, 1990–2010; a longitudinal analysis of national survey data. *Lancet* 383, 2057–2064. doi: 10.1016/S0140-6736(13)62639-2

Wells, W. A., Boehme, C. C., Cobelens, F. G., Daniels, C., Dowdy, D., Gardiner, E., et al. (2013). Alignment of new tuberculosis drug regimens and drug susceptibility testing: a framework for action. *Lancet Infect. Dis.* 13, 449–458. doi: 10.1016/S1473-3099(13)70025-2

Witney, A. A., Bateson, A. L., Jindani, A., Phillips, P. P., Coleman, D., Stoker, N. G., et al. (2017). Use of whole-genome sequencing to distinguish relapse from reinfection in a completed tuberculosis clinical trial. *BMC Med.* 15:71. doi: 10.1186/s12916-017-0834-4

Xu, Y., Liang, B., Du, C., Tian, X., Cai, X., Hou, Y., et al. (2019). Rapid identification of clinically relevant *mycobacterium* species by multicolor melting curve analysis. *J. Clin. Microbiol.* 57:e01096-18. doi: 10.1128/JCM.01096-18

Xu, C., Pang, Y., Li, R., Ruan, Y., Wang, L., Chen, M., et al. (2018). Clinical outcome of multidrug-resistant tuberculosis patients receiving standardized second-line treatment regimen in China. *J. Infect.* 76, 348–353. doi: 10.1016/j.jinf.2017.12.017

Zhang, Z., Lu, J., Wang, Y., Pang, Y., and Zhao, Y. (2014). Prevalence and molecular characterization of fluoroquinolone-resistant *Mycobacterium tuberculosis* isolates in China. *Antimicrob. Agents Chemother.* 58, 364–369. doi: 10.1128/AAC.01228-13

Zong, Z., Huo, F., Shi, J., Jing, W., Ma, Y., Liang, Q., et al. (2018). Relapse versus reinfection of recurrent tuberculosis patients in a national tuberculosis specialized hospital in Beijing, China. *Front. Microbiol.* 9:1858. doi: 10.3389/fmicb.2018.01858