Whole Genome Sequence-Based Analyses of Drug Resistance Characteristics, Genetic Diversity, and Transmission Dynamics of Drug-Resistant \textit{Mycobacterium tuberculosis} in Urumqi City

Jiandong Yang\textsuperscript{1}, Yaoqin Lu\textsuperscript{2}, Yanggui Chen\textsuperscript{2}, Yida Wang\textsuperscript{1}, Kai Wang\textsuperscript{3}

\textsuperscript{1}School of Public Health, Xinjiang Medical University, Urumqi, 830017, People’s Republic of China; \textsuperscript{2}Department of Tuberculosis Control and Prevention, Urumqi Center for Disease Control and Prevention, Urumqi, 830026, People’s Republic of China; \textsuperscript{3}State Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence Diseases in Central Asia, School of Public Health, Xinjiang Medical University, Urumqi, 830017, People’s Republic of China

Correspondence: Kai Wang, State Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence Diseases in Central Asia, School of Public Health, Xinjiang Medical University, Urumqi, 830017, People’s Republic of China, Email wangkaimath@sina.com

Objective: This study aims to analyze the drug resistance spectrum, genetic diversity, and transmission dynamics to provide a basis for the prevention and control of drug-resistant (DR) tuberculosis (TB) epidemics.

Methods: This retrospective study is based on routine national drug resistance surveillance. The demographic, epidemiological, and clinical information on DR-TB patients from 2016 to 2021 was collected and used for phenotypic drug susceptibility testing and whole-genome sequencing.

Results: It was indicated that L2.2.1 was the dominant lineage in Urumqi. The drug resistance spectrum in Urumqi was narrow, which means more drug combinations can be used for clinical treatment. Furthermore, mutations identification of drug-resistance gene \textit{katG}, \textit{rpoB}, \textit{embA/B}, \textit{rrs}, \textit{rpsL}, \textit{eis}, \textit{gyrA/B}, \textit{folC} and \textit{tryA} are important for clinical drug use. However, mutations in cross-resistance genes \textit{rrs} have limited guidance for clinical selection of KM, CPM and AK. Moreover, there is an increased risk of cluster transmission of DR-TB, and the difference in clustering rate among L2, L3, and L4 was not statistically significant ($\chi^2 = 2.6410, p = 0.2670$).

Conclusion: In the Urumqi, DR-TB has a complex prevalence state, a narrow drug resistance spectrum, and a high clustering rate and burden of drug resistance. To reduce the burden of DR-TB, related research should be strengthened, and the development of prevention, control, and treatment strategies should be accelerated.

Keywords: \textit{Mycobacterium tuberculosis}, resistance, whole genome sequencing, transmission dynamics, Urumqi

Background

Tuberculosis (TB) has been a threat to public health for decades and prevailed even during the COVID-19 pandemic. According to the WHO Global Tuberculosis Report 2020, there were approximately 9.96 million new cases, and the population with latent TB infection was approximately 2 billion.\footnote{2020 Global Tuberculosis Report, World Health Organization} In the past four years, the COVID-2019 pandemic had varying impacts on the global TB control system, with access to TB services particularly affected, especially multidrug-resistant TB (MDR-TB) and extensively drug-resistant (DR) TB.\footnote{MDR-TB and DR-TB: A Global Failure to Meet Treatment Targets, World Health Organization} DR-TB is widely recognized as a challenging and burdensome disease that poses a significant obstacle to TB control.\footnote{Global Tuberculosis Report 2022} The Global Tuberculosis Report 2022 indicated a substantial increase in the global burden of DR-TB by 2021, with only one-third of these patients receiving treatment, which has a success rate of merely 60\%.\footnote{Global Tuberculosis Report 2022} This alarming trend undermines the WHO’s objective of eliminating TB by 2035. Therefore, it is essential to prioritize DR-TB prevention and control in the post-pandemic era.

The use of whole genome sequencing (WGS) in genomic research of pathogenic microorganisms has continued to advance.\footnote{Whole Genome Sequencing: A Research Tool for Global Tuberculosis Control, World Health Organization} In the field of TB research, the identification and publication of the whole genome sequence of \textit{Mycobacterium tuberculosis}
**Information and Methods**

**Data Source and Database Establishment**

This retrospective study was conducted based on routine national drug resistance surveillance in Urumqi. The demographic, epidemiological, and clinical information of the patients from 2016 to 2021 was acquired from the National Drug Resistance Surveillance Database (NDRS, [https://www.carss.cn/](https://www.carss.cn/)). Moreover, drug resistance and WGS data were acquired through Phenotypic Drug Susceptibility Testing (Phenotypic DST) and high-throughput sequencing. Appendix S1 shows detailed information on the data collection process.

**Experimental Methodology**

This research followed an experimental method solid Lowenstein-Jensen (L-J) medium and identification medium to obtain MTB strains, and Phenotypic DST was done using the resistance ratio method on L-J media. The sputum samples were fully digested with 2–4 volumes of 4% NaOH depending on the viscosity of the sputum. The fully digested specimens were inoculated onto solid L-J medium, P-Nitrobenzoic acid (PNB) and Thiophene-2-carboxylic acid hydrazide (TCH) identification medium slants and incubated at 37°C. When cultures were positive and MTB grown to logarithmic phase, the samples will be tested for Phenotypic DST according to standardized operating procedures. MTB DNA was extracted by the manual membrane adsorption column method. Then, using Illumina platforms with PE150 strategy in Novogene Bioinformatics, a high-throughput sequencing of MTB was performed. Further detailed information on the experimental methodology, experimental process, and related data is present in Appendix S2.

**Data Analysis**

The FastQC (v0.11.9), Trimmomatic (v0.39), Snippy (v4.6.0), GATK4 (v4.0.4.0), Vcftools, IQ-TREE (v2.2.2.7), TB-Profiler (v4.4.2), and RStudio Desktop (v2022.07.1–554) software were used for biological information analysis (including quality control, assembly, calling SNP, annotation, build evolutionary tree and so on) of the genomic data. Furthermore, for statistical descriptions and statistical tests, the RStudio Desktop v2022.07.1–554 was employed. Additionally, the SaTScan v10.1 was utilized for Spatial-Temporal scanning analysis. Test level $\alpha = 0.05$, and the
difference was statistically significant when \( p < 0.05 \). Appendix S3 contains detailed information on the software, analysis methods, and parameter settings.

**Results**

**Demographic and Clinical Characteristics**
A total of 207 DR-TB patients were included in the study. Most of the included participants were male (61.84%), under the age of 60 years (55.07%), lived in urban areas (87.44%), had non-farmer occupations (87.44%), and not floating population (lived Urumqi more than or equal to 6 months) (75.85%). It is worth noting that 82.61% of the patients were receiving their first treatment, and half of the patients (52.66%) had delayed visits. The proportion of positive sputum smear and GeneXpert were 82.61 and 93.72%, respectively. Additionally, one patient was diagnosed with tuberculous meningitis (0.48%), and 2.42 and 13.53% of patients self-reported co-existing HIV/AIDS and diabetes, respectively. Appendix S4 represents detailed information on the included participants.

**Phenotypic Resistance Profiles**
The DST was conducted on 207 strains of DR-MTB, which indicated that Streptomycin (STR) had the highest resistance rate (50.72%), followed by isoniazid (INH) (48.31%), whereas kanamycin (KM) showed the least resistance (3.86%) among anti-TB drugs. In terms of drug resistance type, isoniazid resistance (Hr-TB) had the highest percentage (28.02%), followed by mono-resistant TB (MR-TB) at 24.64%, and then poly-resistant TB (PR-TB) at 1.93%. The combination analysis of 11 anti-TB drugs indicated 55 drug combinations in the resistance spectrum, accounting for 2.69% (55/2047) of the total number of combinations. Mono-resistance to STR had the highest percentage (20.77%), followed by the combinations INH+STR and Ofloxacin (OFL) + Levofloxacin (LFX) + Moxifloxacin (MFX) (both 7.73%). Mono-resistance to INH ranked third (63.76%). The detailed information is illustrated in Appendices S5 and S6.

**Molecular Resistance Characterization**
The drug resistance genes mutational patterns indicated relatively low (< 35.00%) mutation rates of single resistance genes. Furthermore, a large number of joint mutation phenomena were observed among resistance genes. Moreover, genome-wide association analysis indicated that Amikacin (AK), KM, and Capreomycin (CPM) shared the drug resistance gene, \( rrs \). Therefore, \( rrs \) gene mutations may affect the phenotypic resistance of AK, KM, and CPM. Appendix S7 contains data analysis results and a detailed description of the results.

**Comparison of Phenotypic Resistance and Molecular Mutation**
It was observed that most of the mutated resistance genes were statistically significant in biological information (\( p < 0.05 \)); however, the distribution of most of these mutants between the phenotypically resistant and sensitive groups was less statistically different (\( p > 0.05 \)). Furthermore, the differences were only observed in the distribution of mutations in the KatG gene of INH, \( rpoB \) of Rifampicin (R), \( embA/B \) of Ethambutol (EMB), \( rrs \) and \( rpsL \) of STR, \( rrs \) of AK and CPM, \( rrs \) and \( eis \) of KM, \( gyrA \) of OFL and LFX, \( gyrA/B \) of MFX, as well as in the \( folC \) and \( rryA \) genes of Para-aminosalicylic-acid (PAS) between the phenotypically resistant and sensitive groups (\( p < 0.05 \)). Moreover, the diagnostic test analysis indicated that the sensitivity and specificity of the above genes in predicting phenotypic resistance are generally high, although there are differences. Among the aforementioned genes, the highest diagnostic index was the prediction of resistance to fluoroquinolones by \( gyrA \) gene mutations, followed by the resistance to INH by \( katG \) gene mutations, and then of R by \( rpoB \) gene mutations. More interestingly, it was observed that \( rrs \) gene mutations are predictive of resistance to STR, AK, CPM, and KM, and the diagnostic power is relatively high. Furthermore, after mutated genes with differences in distribution between phenotypically resistant and sensitive groups were compared with the 2nd edition of the WHO mutation catalog, it is noteworthy that the variant C1402A of cross-resistance gene \( rrs \) of AK, CPM, and KM in sample056 and the variant Thr308Pro of \( katG \) in sample028 were not found in the second edition of the WHO mutation catalog. Appendix S8 contains data analysis results and a detailed description of the results.
Phylogenetic Analysis

Population genetic structure analysis showed that locally prevalent strains belong to lineage 2 (L2), L3, and L4, among which the Beijing family was the dominant strain (68.12%). Furthermore, it was indicated that only STR and OFL of the 11 anti-TB drugs have differences in resistance distribution among L2, L3, and L4 ($p < 0.05$). The STR had the highest resistance in the L3, whereas OFL had the highest resistance in the L2.

Following the study result by Wencong et al, genomic transmission clusters were defined using 12 SNPs as a cutoff value in our research. The clustering analysis identified 14 clusters (C1 to C14) with a clustering rate of 29.47%, and the larger cluster had 21 MTB strains. The spatiotemporal analysis of clustered strains revealed that 60.66% had spatiotemporal correlation. Among the clustered strains, Hr-TB strains accounted for the highest proportion (27.87%), followed by MR-TB (24.59%), and then Other-resistant TB (Other-DR) (16.39%). Additionally, no statistical difference was observed in the clustering rate among the three lineages ($\chi^2 = 2.6410, p = 0.2670$). Moreover, according to the epidemiological investigation, the strains in cluster C14 spanned five years in time (from 2017 to 2021). In addition, 11 patients were infected with non-clustered MTB, and they were migrants. Further detailed information on the results and data analysis results are presented in Table 1, Appendix S9 and Figures 1–3.

Discussion

This study found that the distribution of DR patients in Urumqi has both similarities and differences with other countries and regions. Furthermore, male patients and those with delayed hospital visits were more likely to develop DR than female patients or those who had not delayed hospital visits, which is consistent with the findings of Denis Okethwangu et al. Contrary to the findings of Olanrewaju Oladimeji and Wenlong Zhang, this research indicated that > 60 years old, non-farmers, and non-migrant individuals have a higher proportion of DR-TB. Some scholars believe that newly treated patients are more likely to develop drug resistance than re-treated patients, whereas others believe that the opposite is true. This phenomenon might be influenced by regional differences and drug resistance transmission patterns. According to the Urumqi Municipal Bureau of Statistics’ 2022 Urumqi National Economic and Social Development Statistical Bulletin, Urumqi’s urbanization rate is as high as 96.5% and has been steadily increasing. As urbanization progresses, the range of available job opportunities to the rural population is increasing, decreasing the proportion of farmers in the overall population. Consequently, the prevalence of TB and DR-TB among farmers has decreased. This highlights the social nature of TB and suggests that changes in social conditions can impact its epidemiological characteristics. The elevated percentage of DR-TB among ≥ 60-year-old individuals, the non-mobile population, and patients receiving initial treatment might be linked with the high DR-TB cluster rate in Urumqi, which was found to be as high as 29.47%. Previous research has shown that the basic regeneration number (R0) of TB was highest in Xinjiang (11.38), whereas it was 5.46 in Urumqi. This infers that DR strain transmission might replace main factors, including poor compliance and irregular drug intake, affecting the DR-TB epidemic in Urumqi. This highlights the higher transmission of DR-TB in Urumqi and also explains the high proportion of first-received treatment among DR patients. These results emphasize the significance of controlling the infection source and inhibiting the transmission pathway to prevent the DR-TB spread from patients to non-infected health people in Urumqi.

Another interesting discovery is that only 55 of the 2047 combinations of 11 anti-TB drugs appeared in the Urumqi resistance spectrum, accounting for 2.69%. However, research by Junli Yi et al showed that the number of drug resistance spectrum combinations accounted for 19.11% of the total number of combinations. This phenomenon shows that anti-TB

<table>
<thead>
<tr>
<th>Clustered</th>
<th>Unclustered</th>
<th>$\chi^2$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lineage 2</td>
<td>44</td>
<td>28.21</td>
<td>112</td>
</tr>
<tr>
<td>Lineage 3</td>
<td>9</td>
<td>45.00</td>
<td>11</td>
</tr>
<tr>
<td>Lineage 4</td>
<td>8</td>
<td>25.81</td>
<td>23</td>
</tr>
</tbody>
</table>
drugs in Urumqi have a relatively narrow resistance spectrum. A narrow drug resistance spectrum indicates relatively concentrated types of anti-TB drug resistance. Furthermore, such a spectrum means that clinicians can have more choices when using a combination of anti-TB drugs to treat DR-TB patients, which is good news. Moreover, the resistance spectrum of anti-TB drugs in Urumqi is specific compared to other regions. Some studies show the resistance rates of four first-line anti-TB drugs in northern China were STR > R > INH > EMB, and the combinations with the highest proportion of anti-TB drug resistance in MR-TB, PR-TB, and MDR-TB patients are STR, INH+STR, and INH+R, respectively. However, here, the resistance rates of the four first-line anti-TB drugs followed the order of STR > INH > R > EMB. Additionally, the same resistance rates were observed for INH + STR and OFL + LFX + MFX, and the rate of INH resistance is higher than R. Other studies have shown that patients with Hr-TB have a higher cure rate but also a higher resistance than R resistance, which is not exciting for Hr-TB patients. Therefore, although the cure rate of Hr-TB is high, it should not be underestimated, and long-term Hr-TB patient follow-up should be performed after successful treatment. It is noteworthy that the proportion of isoniazid-sensitive strains among R-resistant isolates is relatively high, suggesting that the detection of R resistance might not be a reliable indicator of predicting DR-TB. Therefore, new approaches should be considered, as highlighted by the findings of Nasiri et al.
Figure 2 Maximum-likelihood tree of 61 resistant tuberculosis strains within 14 clusters and their phenotypic drug-resistant profiles. The red lines indicate boundaries of individual clusters. C1-C14 represents the first to fourteenth clusters, respectively. H, R, E, S, OAK, KM, CPM, PAS, LFX and MFX represent anti-TB drugs Isoniazid, Rifampicin, Ethambutol, Streptomycin, Ofloxacin, Amikacin, Kanamycin, Capreomycin, Para-aminosalicylic-acid, Levofloxacin and Moxifloxacin, respectively. Isoniazid-resistant tuberculosis (Hr-TB) is TB that is resistant to isoniazid but sensitive to rifampicin; Multidrug-resistant tuberculosis (MDR-TB) is defined as resistance to rifampicin and isoniazid; Mono-resistant tuberculosis (MR-TB) is defined as resistance to one kind of first-line anti-TB drug; Pre-extensively drug-resistant tuberculosis (Pre-XDR-TB) is TB that is resistant to rifampicin and any fluoroquinolone (a class of second-line anti-TB drug); Poly-resistant tuberculosis (PR-TB) is TB that is resistant to at least of first-line anti-TB drug (but not include simultaneous resistance to isoniazid and rifampicin); Rifampicin-resistant tuberculosis (RR-TB) is TB that is resistant to rifampicin (No matter what of resistance to other anti-TB drugs); Extensively drug-resistant tuberculosis (XDR-TB) is TB that is resistant to rifampicin, plus any fluoroquinolone, plus at least one of the drugs bedaquiline and linezolid; Other-resistant tuberculosis (Other-DR) is defined as resistance combinations that do not fall within the seven definitions above.

Figure 3 Spatiotemporal scanning area aggregation diagram. The red circle is the first gathering area; The blue circle is the second gathering area; The position circled by the green circle is the only distant point in this study. The horizontal coordinate represents longitude and the vertical coordinate represents latitude.
This research also indicated that drug-resistance gene mutation sites are unique, and mutations at unconventional mutation sites were also observed, such as in the \textit{katG}, \textit{inhA}, \textit{emb}, and \textit{rrs} genes. This suggests that the genetic mutation of DR MTB is highly prevalent in Urumqi. However, it is still uncertain whether these unconventional locus mutations can cause resistance to the corresponding anti-TB drugs and the mechanism of resistance. Therefore, further research, investigation, and validation of these results are necessary.

A comparison between molecular mutations and phenotypic resistance was conducted to assist doctors in promptly and accurately determining patients’ resistance status. Although the diagnostic index of each anti-TB drug resistance gene mutation was not very high, these mutations still guide the clinical diagnosis of DR-TB. Consistent with this research, some studies have indicated that KM, CPM, and AK share resistance genes.\textsuperscript{27} According to this investigation, KM, CPM, and AK shared the drug resistance gene, \textit{rrs}, and the gene mutation has the same ability to predict the phenotypic resistance of these drugs. Therefore, the other two drugs must be used cautiously as long as drug resistance is detected for any of KM, CPM, and AK. However, it also reminds us that mutations in cross-resistance genes \textit{rrs} have limited guidance for clinical selection of KM, CPM and AK. In addition, our results showed that the variant C1402A of cross-resistance gene \textit{rrs} of AK, CPM, and KM in sample056 and the variant \textit{Thr308Pro} of \textit{katG} in sample028 were not found in the second edition of the WHO mutation catalog.\textsuperscript{28} However, since only one sample in each case had this mutation, we cannot be sure whether the mutation is a chance phenomenon. Therefore, more samples and further research are still needed to prove whether there is a real association between this mutation and drug resistance.

The high cluster rate indicates that direct DR MTB infection has a significant role in the prevalence of DR-TB in Urumqi. Furthermore, the delayed visits to hospitals contribute to the spread of DR-TB. Hence, it is crucial to diagnose and treat DR-TB patients at an early stage. The literature indicates that DR-TB patients who undergo prolonged treatment have an extended infection period, which increases the risk of DR-TB transmission to the healthy population,\textsuperscript{29} which is contradictory to this study. Currently, Urumqi employs a centralized isolation strategy for TB treatment, which effectively prevents its transmission from patients to healthy individuals. Therefore, DR-TB transmission from primary patients who are not timely diagnosed and treated to the surrounding healthy population may significantly spread DR-TB in Urumqi. The COVID-19 pandemic has severely compromised the TB prevention and control service system due to strict infectious disease prevention and control policies. This might have contributed to the drug resistance development in MTB and created favorable conditions for DR-TB spread.\textsuperscript{30} Spatial-temporal scanning analysis revealed that most of the areas with clustered DR-TB patients were within a 20 km radius of the geographical center of the cluster. This data further validates the hypothesis that the majority of DR-TB cases in Urumqi result from direct infection with DR MTB strains. Interestingly, the patients within the same cluster had different phenotypic resistance, which might reflect the molecular characteristics diversity of resistance to anti-TB drugs. This is related to the fact that Urumqi has a high population mobility because of the overland Silk Road trade. Some studies show that the frequent movement of the population helps gene flow and polymorphism,\textsuperscript{31} consistent with the data of this study. Moreover, the clustering rates were not significantly different between L2, L3, and L4. Some studies have shown a high level of transmission in L2 and resistance in L4.\textsuperscript{32,33} However, the significant difference in drug resistance between MTB lineages has not been indicated.\textsuperscript{17} Therefore, this study supports the latter view, but the reasons for this phenomenon need further exploration.

**Conclusion**

In summary, to reduce the burden of DR-TB in Urumqi, the progress in relevant research and the development of prevention, control, and treatment strategies should be accelerated. Furthermore, because of cross-resistance, if resistance to KM, CPM, and AK is detected, the other two non-resistant drugs should be cautiously used. Also need to note that the mutations in cross-resistance genes \textit{rrs} have limited guidance for clinical selection of KM, CPM and AK.

**Data Sharing Statement**

Data that support the findings of this study are available from the National Microbiology Data Center (NMDC) (https://nmdc.cn/resource/genomics/project/detail/NMDC10018716, Project number: NMDC10018716) upon reasonable request and with permission from the corresponding author.


Ethics Statement

National drug-resistant surveillance (DRS) was ethically approved by the Ethics Committee of the Chinese Center for Disease Control and Prevention since the first national survey in 2007. Ethics approval of the present study was skipped because all the isolates used were from previous DRS studies, demographic characteristics were extracted from previous data sets, and no additional data and specimens were collected. Each patient signed an informed consent form during the routine DRS.

Funding

This study was supported by the the State Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence Diseases in Central Asia Fund (SKL-HIDCA-2022-JH2), Special Project for Healthy Young medical in Xinjiang Uygur Autonomous Region of China (WJWY-202109), Biosafety strategic defense integration construction innovation team (2022TSYCTD0015), Youth science and technology innovation talent of Tianshan Talent Training Program in Xinjiang (Grant No.: 2022TSYCCX0099), and Science and Technology project of the Health Commission in Urumqi (Grant No.: 202349).

Disclosure

All authors declared no competing interests in this work.

References


