ORIGINAL RESEARCH

Molecular and Transmission Characteristics of *Mycobacterium Tuberculosis* Strains Among College Students in Beijing, China

Xiaolong Cao (ID)[1,2,*], Xinyue Li[1,3,*], Zexuan Song[4], Ping He[5], Ruiqing Zhang[1,3], Chong Teng[6], Qian Sun[2], Xue Wang[2], Bing Zhao[3], Zhiguo Zhang[7], Yanlin Zhao[3]

[1]National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, 102206, People's Republic of China; [2]Outpatient Department, Beijing Changping Institute for Tuberculosis Prevention and Treatment, Beijing, 102200, People's Republic of China; [3]National Tuberculosis Reference Laboratory, Chinese Center for Disease Control and Prevention, Beijing, 102206, People's Republic of China; [4]Department of Clinical Laboratory, Children's Hospital, Capital Institute of Pediatrics, Beijing, 100020, People's Republic of China; [5]Center for Infection Biology, School of Medicine, Tsinghua University, Beijing, 100084, People's Republic of China; [6]Tuberculosis Prevention and Control Department, Beijing Dongcheng District Center for Disease Control and Prevention, Beijing, 100050, People's Republic of China; [7]Hospital Management Office, Beijing Changping Mental Health Care Hospital, Beijing, 102202, People's Republic of China

*These authors contributed equally to this work

Correspondence: Zhiguo Zhang, Beijing Changping Mental Health Care Hospital, Changping District, Beijing, 102202, People's Republic of China, Email zhgzh8689@126.com; Yanlin Zhao, National Tuberculosis Reference Laboratory, Chinese Center for Disease Control and Prevention, Changping District, Beijing, 102206, People's Republic of China, Email zhaoyl@chinacdc.cn

Background: College students are a crucial link in curbing the epidemic. The aim of this study is to analyze the genetic diversity and drug resistance of *Mycobacterium tuberculosis* strains in college students with tuberculosis in Beijing, revealing the lineage structure and transmission patterns specific to this group.

Methods: This study used the hospital's electronic management system to screen for tuberculosis among college students in Changping District, Beijing, from January 2004 to December 2023. Socio-demographic and clinical data were collected, and whole-genome sequencing was performed on culture-positive isolates. Isolates with a genetic distance of less than 12 SNPs were grouped into the same genomic cluster. The TB Profiler software predicted drug resistance mutations, and categorical data were analyzed using Chi-square or Fisher's exact tests.

Results: Among the 1436 college students with tuberculosis, a total of 153 isolates successfully underwent whole-genome sequencing. The results showed that about one-third (49/153) of the isolates carried one or more drug resistance genes, with more than half (26/49) associated with first-line anti-tuberculosis drugs. However, encouragingly, the incidence of drug-resistant tuberculosis showed a significant downward trend, with statistical significance ($p<0.05$). Lineage 2 (86.3%, 132/153) was the predominant genotype, with the Beijing genotype (90.1%, 120/153) being the most common, while the isolation of Lineage 3 in a student from Xinjiang. Sixteen college student isolates clustered, and all of which were Beijing genotype. Transmission within the same campus showed characteristics of short clustering time.

Conclusion: The drug resistance rate among college students is relatively high, however it shows a declining trend. School tuberculosis infections could stem not only from within-campus transmission but also necessitate consideration of spatial and cross-regional spread possibilities.

Keywords: *Mycobacterium tuberculosis*, whole genome sequencing, lineage, transmission, drug resistance, college students

Introduction

Tuberculosis (TB) is a type of ancient and tenacious infectious disease caused by *Mycobacterium tuberculosis* (MTB), remains one of the significant challenges faced by the global public health community today. Before the COVID-19 pandemic, TB was the leading cause of death from a single infectious disease in humans.[1–3] According to the 2023 Global Tuberculosis Report, there were approximately 10.6 million new TB cases and 1.3 million deaths worldwide in 2022. Over the past three years, the incidence rate of TB has increased by 3.9%. China ranks third globally in new TB

cases, following India and Indonesia,[4] and bears a significant economic and prevention burden.[5,6] Previous studies have shown that children and adolescents are key populations for the occurrence and latent infection of tuberculosis (LTBI), accounting for 10% of the total reported cases globally.[1,4] Previous domestic research indicates that high schools are the most common locations for TB outbreaks, likely due to factors such as large population density, close contact among students, and the relatively enclosed environments of classrooms and dormitories.[7,8] Other countries around the world are also facing the same challenge of TB outbreaks in schools.[9–11] Therefore, a comprehensive understanding of molecular and transmission characteristics of MTB among student population can accelerate the achievement of the World Health Organization's (WHO) goal to eliminate TB.

The premise of curbing TB spread is early detection and intervention. However, traditional epidemiological investigations mainly rely on the medical history of close contacts, making it challenging to trace the transmission chain of TB and thereby affecting the accurate and complete reconstruction of its transmission history.[1,12] Whole genome sequencing (WGS) technology can effectively address this issue. With the development of WGS technology, it is now widely used in TB basic research, clinical work and routine monitoring, which includes epidemiological investigations, predicting drug resistance, and revealing the evolutionary patterns of strains.[13–15] Compared to other pathogens, MTB accumulates mutations at a slower rate. It is estimated that it acquires mutations at a rate of 0.2–0.5 single nucleotide polymorphisms (SNPs) per year on average. Studies have indicated that the mutation rate may be associated with the infected population and specific strain typing.[16] The threshold of SNPs differences can be used to determine whether it is a recent transmission. Previous studies have indicated that isolates with a 2-year period with ≤ 10 SNPs are considered to be recent transmissions and should be prioritized for epidemiological investigation.[16,17] WGS can also be used for the analysis of drug resistance in MTB, including the degree and types of resistance. The burden of rifampicin-resistant tuberculosis (RR-TB), multidrug-resistant tuberculosis (MDR-TB), and extensively drug-resistant tuberculosis (XDR-TB) is significant in our country. Delayed diagnosis is a major cause of treatment failure and even death.[18,19] Therefore, WGS technology can provide critical information about mutations that confer drug resistance, which is of significant guidance for the early clinical development of individualized treatment plans.

College students, as a significant mobile population within the country, engage in large-scale migration activities annually during winter and summer vacations. This undoubtedly poses challenges for TB prevention and control efforts at educational institutions.[20] Previous studies have indicated that the incidence of TB is associated with geographical conditions and economic development. Notably, population migration can exacerbate the cross-regional transmission of tuberculosis.[21,22] However, there is a notable scarcity of research within the country on the pathways of TB transmission among college students and the strain characteristics specific to this demographic. This study is a retrospective investigation that selects TB patients from universities and higher vocational colleges in Changping District as research subjects. Located in the northwest region of Beijing, Changping District serves as a significant hub for higher education in the city, hosting numerous renowned universities and higher vocational colleges. This study employs whole-genome sequencing (WGS) and other methodologies to systematically analyze tuberculosis strains among college students in Changping District, Beijing, over the past 20 years. The objective is to elucidate the molecular and transmission characteristics of *Mycobacterium tuberculosis* within this specific population group and to gain a deeper understanding of TB characteristics in young people. Through these research findings, the effectiveness of TB prevention and control in schools can be significantly enhanced, providing a scientific basis for the development of targeted prevention measures.

Materials and Methods
Sample and Information Collection

This research was a retrospective study based on routine work of the Changping Institute for TB Prevention and Treatment in Beijing. Utilizing the electronic patient management system of the hospital, TB cases among universities and higher vocational colleges students were identified from January 2004 to December 2023. We collected socio-demographic information such as gender, age, native place and school name, in addition to acquiring sputum test results like smear, culture, and Gene X-pert. Through information sorting and screening, the research population was identified. The research process was shown in Figure 1.
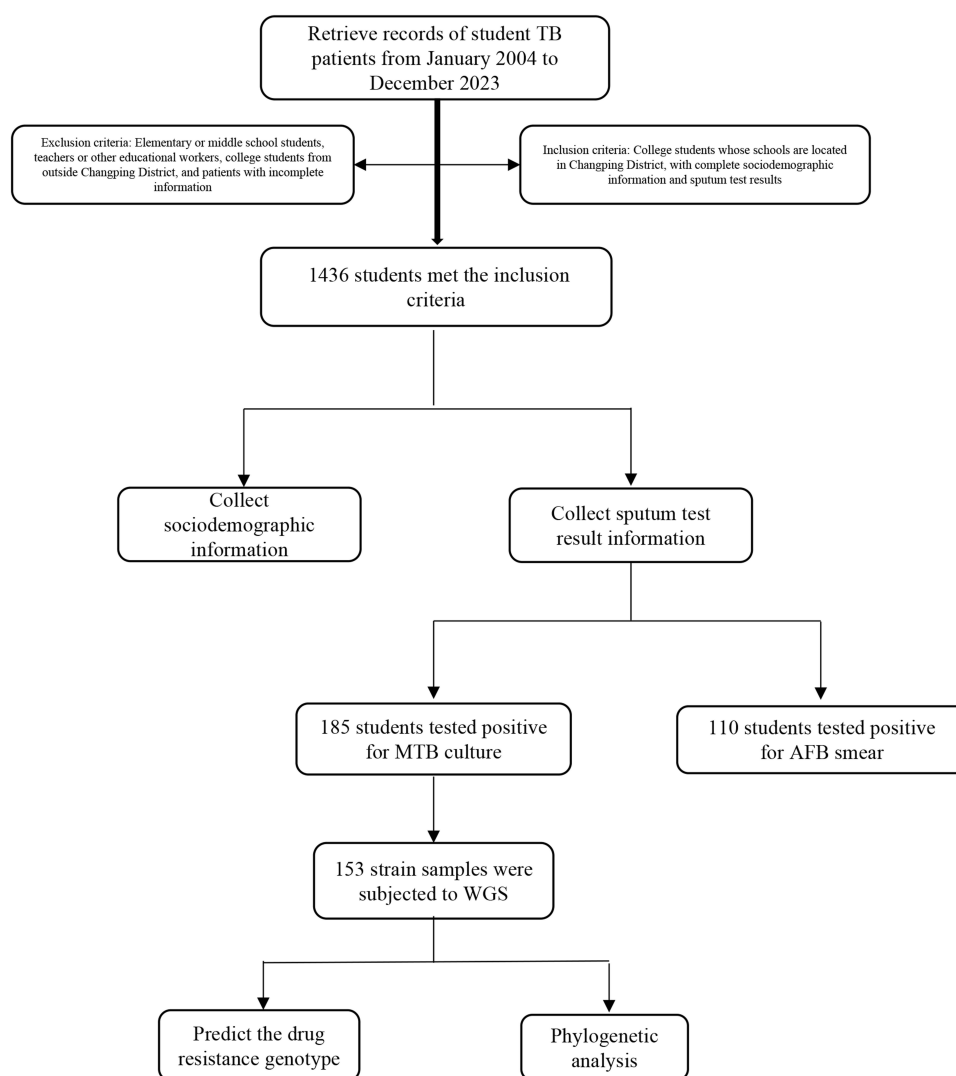
Figure 1 The flow chart of the study showing the study process and participants.

Inclusion Criteria

University or higher vocational college students, whose schools are located in Changping District, with complete socio-demographic information and sputum test results.

Exclusion Criteria

Elementary or middle school students, teachers or other educational workers, students from schools not located in Changping District, and patients with incomplete information will be excluded from the analysis.

DNA Extraction and Sequencing

All genomic DNA used for sequencing originated from inactivated samples provided by the Changping Institute for TB Prevention and Treatment in Beijing. In reference to previous study, we used the Cetyltrimethylammonium Bromide (CTAB) method for the extraction of genomic DNA.[23] After quality assessment (concentration, purity and integrity) of the genomic DNA to be sequenced, the sequencing libraries were prepared according to the protocol provided by Illumina Nextera kit manufacturer. WGS was performed using the Illumina HiSeq PE150 technology. All whole-genome sequencing steps were completed by Annoroad Gene Technology company. (Beijing, China).

 501

Phylogenetic Analysis

The overall quality of the sequence reads was verified using FastQC (v0.11.8), and the confirmed paired-end reads were filtered through Trimmomatic with default parameters and a minimum Phred Quality score set to 20.[3,24] Following the approach of the earlier study, the sequencing data were compared to the H37Rv (NC_000962.3) reference genome using BWA-MEM (v0.7.17), subsequent to the alignment, SAMtools (v1.3.1) and GATK (v3.8.0) were employed to pinpoint variations, which encompassed single nucleotide polymorphisms and insertions/deletions (indels).[6] During the variant filtering process, it is imperative to ensure that each variant is coverage depth at least 10 ×, with a quality score no less than Q20, and the allele frequency must exceed 75%.

In the process of constructing a phylogenetic tree, all SNPs located in regions known to be associated with drug resistance, mobile genetic elements, and PE or PPE family genes were first excluded. Then, a concatenated SNP sequence alignment was created in at least 95% of the isolates. A maximum likelihood phylogenetic tree was constructed using IQ-TREE, and a bootstrap test with 1000 bootstraps was conducted to assess the reliability of the tree. The visualization and editing of the phylogenetic tree were completed using the ChiPlot.[13] Based on previous studies, isolates with a pairwise genetic distance of less than 12 SNPs were classified into the same genomic cluster.[25]

Lineage and Antimicrobial Resistance Prediction

The fast-lineage-caller (v0.3) tool was used to determine the lineage and sublineage of each strain.[13] In predicting the drug resistance of strains, the TB Profiler software (https://github.com/jodyphelan/TBProfiler) was utilized. This software can identify genetic variations related to drug resistance and accordingly infer the drug-resistant genotypes of the strains.[26]

Statistical Analysis

Data organization and descriptive analysis were performed using Microsoft Excel 2013. Categorical data analysis was conducted through Chi-square tests or Fisher's exact tests, as appropriate. All statistical analyses were executed using SPSS 18.0 software (SPSS Inc., Chicago, Illinois). Statistical significance was determined with $P < 0.05$.

Results

General Population Characteristics

In this study, we screened student TB patients from universities and higher vocational colleges in Changping District from 2004 to 2023. A total of 2010 patients were identified. However, due to missing background information or changes in diagnosis for some patients, a total of 574 students were excluded from the study. As a result, the remaining 1436 students were included in subsequent research analysis. Among the 1436 students included in the study, there were 887 males (61.8%, 887/1436) and 549 females (38.2%, 549/1436). In this group of students, 996 had negative etiological test results and were clinically diagnosed cases (69.4%, 996/1436). The remaining 440 students had positive etiological test results and were classified as etiologically diagnosed cases (30.6%, 440/1436). Among these 440 students with positive etiological tests, 185 were sputum culture-positive (42.0%, 185/440), and 110 were sputum smear-positive (25.0%, 110/440). This study successfully completed WGS of 153 isolate samples from students, whose household registrations span all 26 provinces in China and include one international student from the Central African Republic. The students involved are associated with 21 universities and 23 higher vocational colleges in Changping District, and the isolated strains cover a timeline from 2005 to 2023 (Table 1).

Mutated Resistance Genes Associated with Different Drugs

Based on the WGS technology, the genotypic drug resistance mutations of the population strains were predicted. Nearly one-third (32.0%, 49/153) of the strains exhibited genotypic resistance to one or multiple antituberculosis drugs. There was a total of 11 strains classified as RR-TB/MDR-TB/Pre-XDR-TB types, with isoniazid resistant tuberculosis (HR-TB) and other genotypic drug-resistant strains accounting for 15 and 23 isolates, respectively (Figure 2). Over half of the DR-TB strains exhibited mutations in one or more first-line anti-tuberculosis drugs (53.1%, 26/49). Among the 11 rifampicin-resistant strains, most had *rpoB* gene mutations (90.9%, 10/11), while one strain (CP10628) had dual mutations in both

Table 1 General Characteristics of Whole Genome Sequenced Strains

Variable	Total (N=153), N (%)
Gender	
Male	93 (60.8%)
Female	60 (39.2%)
School category	
University	106 (69.3%)
Higher vocational college	47 (30.7%)
Number of isolates in different periods	
2004–2008	42 (27.5%)
2009–2013	31 (20.3%)
2014–2018	59 (38.6%)
2019–2023	21 (13.7%)
Place of Origin	
Domestic (26 provinces)	152 (99.3%)
Internation (the Central African Republic)	1 (0.7%)

rpoB and *rpoC* genes. Isoniazid had the highest resistance (46.9%, 23/49), with *katG* gene mutations being the most common (69.6%, 16/23), followed by *inhA* gene mutations (21.7%, 5/23). One strain (CP10668) had dual mutations (*inhA* and *katG*), and another (CP10628) showed multi-gene mutations involving *inhA, katG*, and *ahpC* genes. Ethambutol resistance was found in 7 strains with *embB* gene mutations, while pyrazinamide resistance occurred in 5 strains with *pncA* gene mutations (Table S1).

Streptomycin had the highest mutation rate (59.2%, 29/49), with *rpsL* gene mutations being predominant (51.7%, 15/29), followed by *rrs* gene mutations (31.0%, 9/29). Mutations in the *gid* gene were found in 3 strains, and 2 strains had dual gene mutations involving both *rpsL* and *rrs* genes, as well as *rpsL* and *gid* genes. For fluoroquinolones, 7 strains exhibited resistance, with *gyrA* gene mutations in 6 strains (85.7%) and *gyrB* gene mutations in 1 strain (14.3%). As second-line anti-tuberculosis injectable aminoglycoside drugs, there were mutations in 3 strains on the *rrs* gene. No resistance-related mutations were found to new anti-tuberculosis drugs like Bedaquiline, Delamanid, Pretomanid, Linezolid, and Clofazimine (Table S1).

Temporal Distribution of Resistance

To investigate the frequency characteristics of drug-resistant mutation genes in this population, the observation period was segmented into four five-year intervals: 2004–2008, 2009–2013, 2014–2018, and 2019–2023. Across these periods, a total of 42, 31, 59, and 21 strains were collected respectively. The incidence rates of drug-resistant mutation genes for each time interval stood at 50.0% (21/42), 29.0% (9/31), 25.4% (15/59), and 19.0% (4/21) respectively (Figure 3A). The incidence of genotype resistance showed a downward trend, with $p < 0.05$ indicating statistical significance (Table 2). The incidence of RR-TB/MDR-TB/Pre-XDR-TB also merits attention. During the period from 2009 to 2013, compared to the period from 2004 to 2008, these incidence rates showed a rapid increase. However, they gradually declined thereafter. Specifically, the incidence rates for each time interval were 4.8% (2/42), 9.7% (3/31), 8.5% (5/59), and 4.8% (1/21) (Figure 3B).

Phylogenetic Analysis

To deeply analyze the population structure, genetic diversity, and transmission patterns of MTB isolates from students, a phylogenetic tree was constructed using the maximum likelihood method based on SNPs.[27] The strains included in this study displayed genetic diversity, encompassing Lineage 2, Lineage 3, and Lineage 4 genotypes (Figure 2). Of all the
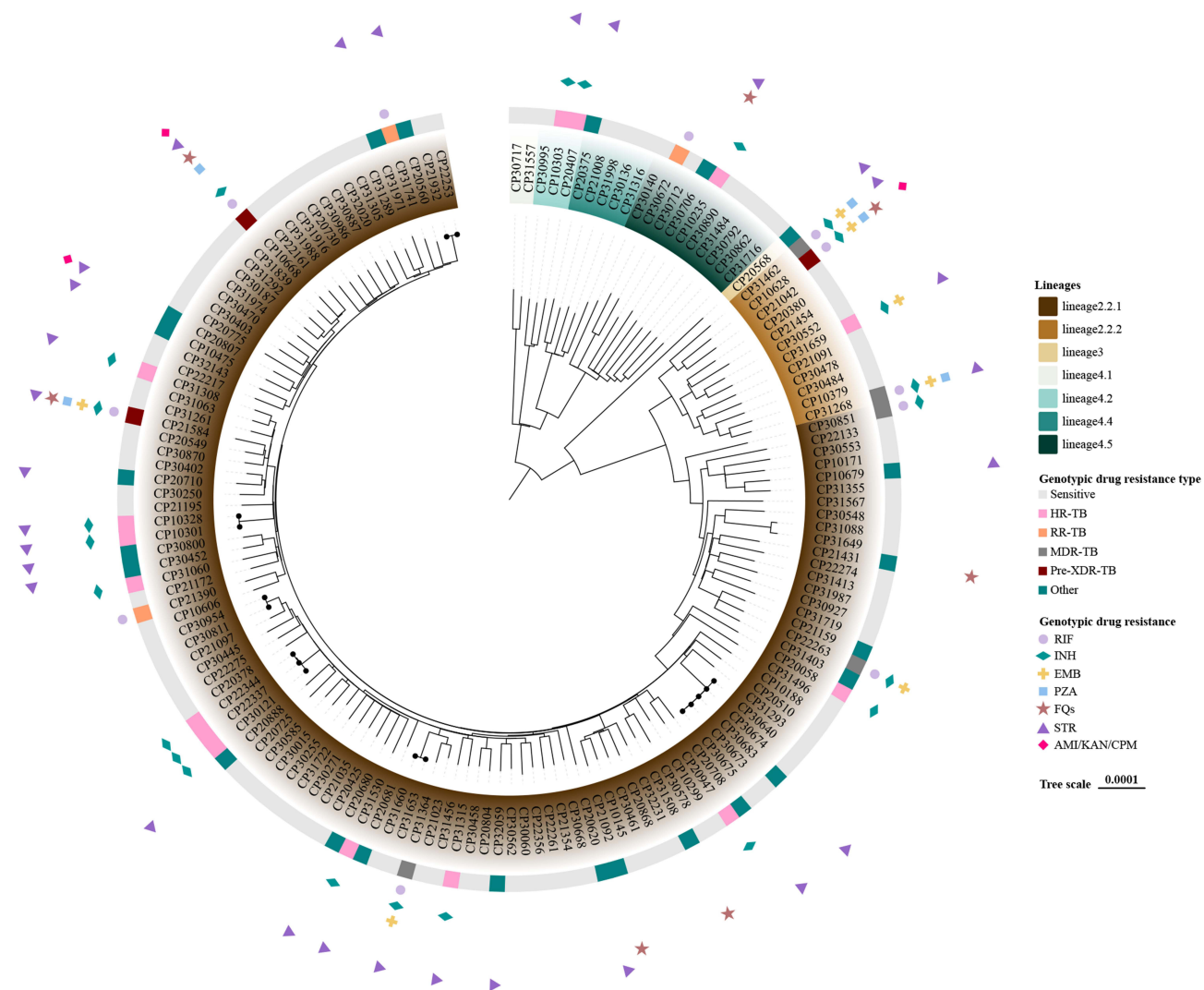
Figure 2 The maximum likelihood tree of 153 Mycobacterium tuberculosis strains from the college student population.
Notes: The lineage, genotype drug resistance types, and an overview of drug resistance mutation genes are displayed in the diagram; potential transmission clusters defined by no more than 12 SNPs are indicated with black circles at the tips of the branches; the scale represents the genetic distance proportional to the total number of single nucleotide polymorphisms.

strains analyzed, Lineage 2 (132) was predominant, with the Beijing genotype (sublineage 2.2.1, 120) being the most common. In addition, there were 12 strains belonging to sublineage 2.2.2. Following by Lineage 4, comprising a total of 20 strains, which further divided into four sublineages: sublineage 4.1 (2), sublineage 4.2 (3), sublineage 4.4 (5), and sublineage 4.5 (10). Additionally, one strain isolated from a student registered in Xinjiang was classified as Lineage 3, the study found no statistically significant difference in the distribution of lineages among college student populations. In addition, the genotype-resistant strains in this population were predominantly Lineage 2 (85.7%, 42/49), with over 90% of RR-TB/MDR-TB/Pre-XDR-TB strains belonging to Lineage 2 (10/11), the frequency of drug resistance genes did not show a statistically significant difference across lineages.

Transmission Analysis

Based on the threshold values of the 12 SNPs and the background information of the strains, we identified a cluster with an onset interval of less than 2 years and below 12 SNPs, this study identified 6 clusters comprising 16 clustered strains, with a clustering rate of 10.5% (16/153) (Figure 2). The sizes of the clusters ranged from 2 to 5 strains. Upon retrieving the patient's medical information, four of the clusters were indicated to be transmissions within a campus, while two
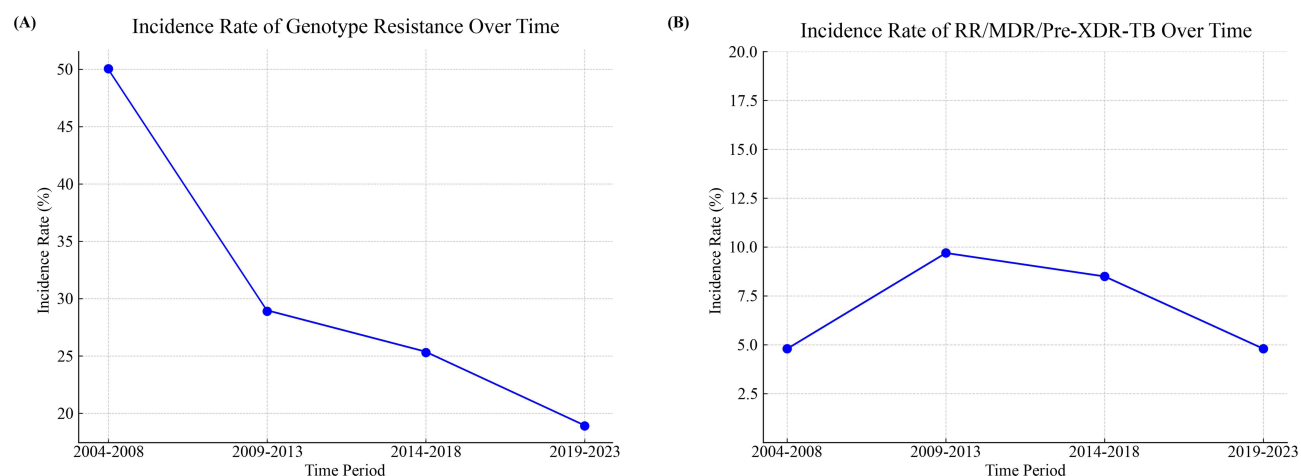
Figure 3 Trends in the incidence of drug-resistant and RR/MDR/Pre-XDR-TB strains. Left (**A**) shown incidence rate of genotype resistance over time, and right (**B**) shown incidence rate of RR/MDR/Pre-XDR-TB over time.

others involved transmissions between different schools. Regarding the transmission between different schools, combining background information of the students reveals that one cluster is related to the proximity of students' home addresses to their schools, while another cluster is related to the close distance between two schools. Both distances are less than 5 kilometers. The longest interval for clusters within the same campus did not exceed nine months, while the shortest was less than one month. The clustering times between different schools were all over one year. It is noteworthy that all the clustered strains were of the Beijing genotype, accounting for 100% (16/16) of all clustered strains, with a clustering rate of 13.3% (16/120) for the Beijing genotype (Table 3).

Discussion

China's vast population and the varying TB incidence rates across regions highlight the need for effective intervention strategies. Targeting key populations, such as college students, is essential for controlling the epidemic. College students, known for their unique social interactions and close contact behaviors, are a significant focus for TB prevention. Notably, internal migration among this group contributes to the rising TB burden on campuses.

This study included a total of 1436 student TB patients, with males being more susceptible to infection. This finding aligns with the results of earlier studies and may be associated with unhealthy lifestyle habits such as smoking.[28,29] Nearly 70% of the patients were diagnosed through clinical diagnosis, but this method may lead to delayed diagnosis, increasing the risk of disease progression and transmission. Etiological diagnosis is the gold standard for TB diagnosis, especially sputum culture, which is crucial for disease diagnosis, drug resistance analysis, and epidemiological investigation. However, traditional culture methods are time-consuming, necessitating the urgent development of rapid and accurate diagnostic tools to achieve early diagnosis and guide treatment. Previous research has indicated that student patients often overlook symptoms of tuberculosis such as coughing and expectoration, which can similarly result in delayed diagnosis.[1] Consequently, disseminating knowledge on TB control is crucial for campus prevention and control

Table 2 Characteristics of the Incidence of Genotypic Drug-Resistant Strains in Different Periods

Time Period	Number of Genotypic Drug-Resistant Strains	Total Number of Strains	Incidence of Genotypic Drug-Resistant Strains	Chi-Square Test
				p
2004–2008	21	42	50.0%	0.027
2009–2013	9	31	29.0%	
2014–2018	15	59	25.4%	
2019–2023	4	21	19.0%	

Table 3 Characteristics of Cluster Strains

Cluster of Strains	Isolated Strain	Same School (Yes or No)	Time of Diagnosis	Lineage	Sublineage
Cluster 1	CP10301	Yes	2005.08	Lineage2	Lineage2.2.1
	CP10328		2005.11	Lineage2	Lineage2.2.1
Cluster 2	CP21932	No	2010.10	Lineage2	Lineage2.2.1
	CP22253		2011.10	Lineage2	Lineage2.2.1
Cluster 3	CP22337	No	2012.01	Lineage2	Lineage2.2.1
	CP22344		2012.02	Lineage2	Lineage2.2.1
	CP30121		2013.04	Lineage2	Lineage2.2.1
Cluster 4	CP30640	Yes	2015.03	Lineage2	Lineage2.2.1
	CP30673		2015.03	Lineage2	Lineage2.2.1
	CP30674		2015.03	Lineage2	Lineage2.2.1
	CP30675		2015.03	Lineage2	Lineage2.2.1
	CP30683		2015.03	Lineage2	Lineage2.2.1
Cluster 5	CP30811	Yes	2015.10	Lineage2	Lineage2.2.1
	CP30954		2016.06	Lineage2	Lineage2.2.1
Cluster 6	CP31653	Yes	2019.04	Lineage2	Lineage2.2.1
	CP31660		2019.06	Lineage2	Lineage2.2.1

efforts. The diagnosis of TB in students must be made with caution, as a confirmed diagnosis typically requires suspension from studies for treatment. This not only affects their academic progress but also represents a significant psychological blow to the students. Therefore, while pursuing rapid diagnosis, attention should also be paid to the psychological well-being of student patients.

The spread of RR-TB/MDR-TB/ XDR-TB poses the greatest threat to TB control efforts. To effectively tackle this issue, rapid and accurate diagnosis is crucial. WGS can identify resistance-conferring mutations through reliable online resources.[3] The results of this study indicated that approximately one-third of the strains carried one or more drug resistance genes, with over half being associated with first-line anti-tuberculosis drugs, demonstrating a high level of drug resistance.[30] This finding required significant attention. Additionally, 7.2% of the strains were classified as RR-TB/MDR-TB/Pre-XDR-TB, which was roughly consistent with the national average (7.42%).[17] Multiple studies have shown that over 90% of RR-TB strains exhibit mutations in the *rpoB* gene's RRDR region, highlighting RRDR mutations as a dominant factor in detecting rifampicin resistance.[31–33] Consistent with these findings, all RR-TB/MDR-TB/Pre-XDR-TB strains in this study had *rpoB* gene mutations in the RRDR, with the most common being *rpoB*S450L (63.6%, 7/11). Furthermore, one Pre-XDR-TB strain exhibited mutations not only in the *rpoB* gene but also in the *rpoC* gene, which was consistent with earlier research conclusions and demonstrated the predictive advantages of WGS technology for drug resistance.[34] In the analysis of isoniazid resistance mutations, the *katG*315 gene mutation was dominant. Previous research indicated that the *katG*315 mutation accounted for 64% of phenotypic isoniazid resistance, consistent with the findings in this study,[35] where over 60% (14/23) of strains exhibited the *katG*315 gene mutation. Notably, *katG* gene mutations demonstrated a high level of INH resistance.[6] Additionally, the drug resistance mutations of isoniazid exhibited genetic diversity, manifesting in various combinations of resistance genes. In this study, seven isolates showed *embB* gene mutations, which can be present in both ethambutol-resistant and susceptible strains. Additionally, 100% (5/5) of pyrazinamide resistance was caused by *pncA* gene mutations, suggesting regional variations in pncA gene mutation patterns.[3,36]

Streptomycin had the highest mutation rate among anti-tuberculosis drugs, with *rpsL* L43A (51.7%, 15/29) being the main mutation in college student strains. Although streptomycin is now a second-line treatment due to resistance and adverse reactions, it remains valuable for MDR-TB. The study showed that DR-TB patients had high resistance to fluoroquinolone antibiotics, with *gyrA* gene mutations (85.7%, 6/7) being dominant, including D94G, A90V, and A90A each at 33.3% (2/6). The fluoroquinolone resistance rate among RR-TB/MDR-TB/Pre-XDR-TB strains (27.3%, 3/11)

was significantly higher than for other types of resistance (2.8%, 4/142), indicating a need for attention as high resistance could affect treatment plans for these patients. Fortunately, new drugs such as bedaquiline and delamanid were fundamental to the treatment regimens for RR-TB/MDR-TB/ XDR-TB patients, and this study did not detect any mutations in these drugs' resistance genes, which was advantageous for the selection of treatment regimens.

The results of this study showed that over the past 20 years, with a five-year interval, there was a noticeable downward trend in the incidence of drug-resistant strains, which might have been related to effective control measures for TB. Particularly during the period from 2019 to 2023, amidst the COVID-19 pandemic, personal protective measures effectively curtailed the spread of TB. However, unlike the declining trend in the incidence of drug-resistant tuberculosis, the occurrence rate of RR-TB/MDR-TB/Pre-XDR-TB strains began to decline starting from 2009 to 2013. The accuracy of the study results may have been affected by the limited number of cases, necessitating validation with a large number of clinical strains. Additionally, in this study, Lineage 2 genotype strains dominated both resistant strains (85.7%, 42/49) and RR-TB/MDR-TB/Pre-XDR-TB strains (90.9%, 10/11). Research indicated that Lineage 2 was closely associated with drug resistance, which might have been due to different selective pressures faced by strains of various lineages.[37]

This study revealed the dynamic of MTB in the college student population through WGS, finding that Lineage 2, especially the Beijing genotype, was the predominant genotype in this group, followed by Lineage 4. Among all the isolates, there was also one strain from Lineage 3, which was isolated from a student from the Xinjiang region. Research suggested that strains from Lineage 3 were mainly concentrated in the Xinjiang region of China,[38] indicating that the tuberculosis infections in college students might have originated from their place of origin. Furthermore, studies have shown that Lineage 2 had a higher transmissibility compared to Lineage 4, and was capable of frequently spreading across regions. This aligns with the characteristic of internal migration within the college student population, further explaining the widespread transmission of the Beijing genotype among this group.[14] The terminal branch lengths and the distance from nodes to tips of the Beijing genotype were shorter, indicating that the isolates of this genotype had a lower mutation rate, which was also consistent with its higher transmissibility.[14]

This study analyzed 153 isolates from college students, of which 16 exhibited clustering, and all clustered isolates belonged to sublineage 2.2.1. The research indicated that Lineage 2 had a higher clustering rate than Lineage 4, and the Beijing genotype was significantly associated with clustering. These findings were consistent with this result.[37] While there was no universally recognized gold standard for defining recent transmission, a general consensus was that infections occurring within a 2-year period with $\leq$ 10 SNPs indicated recent transmission.[12,16] In this research, the maximum clustering duration was 15 months, and the minimum was less than 1 month, with SNPs among clustered isolates ranging from 0 to 3. Consequently, it was determined that these six clusters constituted recent transmission. Among all clusters, Cluster 1, Cluster 4, Cluster 5, and Cluster 6 were associated with on-campus outbreaks, with clustering times less than 3 months, highlighting a feature of brief cluster formation. Conversely, Cluster 2 and Cluster 3 implicated transmission across different schools. Our examination showed that the schools of the two students in Cluster 2 were separated by less than 3 kilometers, whereas student CP30121 in Cluster 3 was less than 5 kilometers distant from another student's school. We hypothesized that these two cases of cross-campus transmission might have happened by accidental spatial dissemination. This also implied that the origins of tuberculosis outbreaks in schools might not stem from within the schools but could be due to external infections.

This study had several limitations. First, strain samples were limited, resulting in inadequate representation of the region and constraining the investigation of transmission patterns among college students. Second, all strains were inactivated, precluding phenotypic drug susceptibility testing and thereby hindering the evaluation of the accuracy of genotype resistance mutations. Lastly, the lack of comprehensive epidemiological data on the students could have led to an underestimation of the incidence rate of TB outbreaks in schools. Despite some limitations, this study provides valuable insights for schools and governments in formulating TB control measures. The findings indicate a high level of drug-resistant gene mutations in *Mycobacterium tuberculosis*, highlighting the need to be vigilant about antibiotic misuse among students and strengthen relevant regulatory measures. Additionally, further exploration into the trends of drug-resistant genes related to anti-tuberculosis drugs and the mechanisms of unknown drug-resistant gene mutations is essential to more effectively address the challenges posed by tuberculosis.

Conclusions

In summary, this study systematically analyzed TB patients among the college student population over the past 20 years using WGS and data compilation. The findings revealed a high frequency of drug resistance gene mutations, particularly to isoniazid and streptomycin. Encouragingly, the frequency of these mutations is declining. No resistance was found to newer drugs like bedaquiline and delamanid. Campus outbreaks result from both internal and external transmission, underscoring the need for targeted prevention and control strategies to contain and reduce further spread within school communities.

Data Sharing Statement

All data supporting the findings of this study are included in the article. Sequencing reads have been submitted to the Sequence Read Archive (SRA) under the accession number PRJNA1173257.

Ethics Statement

All participants involved in this study voluntarily participated with informed consent. The collected strains and patient information were strictly confidential. The research methods and protocols of this study were conducted in accordance with relevant guidelines and regulations (Declaration of Helsinki) and have been approved by the Ethics Committee of Beijing Changping Tuberculosis Prevention and Treatment Research Institute (CJ-2023-002).

Acknowledgments

Funding

Disclosure

The authors declare no conflicts of interest in this work.

References

1. Xu ZH, Liu HC, Liu YP, et al. Whole-Genome Sequencing and Epidemiological Investigation of Tuberculosis Outbreaks in High Schools in Hunan, China. *Infect Drug Resist*. 2022;15:5149–5160. doi:10.2147/IDR.S371772
2. Song ZX, Liu CF, He WC, et al. Insight into the drug-resistant characteristics and genetic diversity of multidrug-resistant Mycobacterium tuberculosis in China. *Microbiol Spectr*. 2023;11(5):e01324. doi:10.1128/spectrum.01324-23.
3. He WC, Tan YH, Liu CF, et al. Drug-Resistant Characteristics, Genetic Diversity, and Transmission Dynamics of Rifampicin-Resistant in Hunan, China, Revealed by Whole-Genome Sequencing. *Microbiol Spectr*. 2022;10(1). doi:10.1128/spectrum.01543-21.
4. World Health Organization. *Global Tuberculosis Report 2023*. Geneva: World Health Organization; 2023.
5. Jiang Q, Liu HC, Liu QY, et al. The Evolution and Transmission Dynamics of Multidrug-Resistant Tuberculosis in an Isolated High-Plateau Population of Tibet, China. *Microbiol Spectr*. 2023;11(2). doi:10.1128/spectrum.03991-22.
6. Wang JY, Yu CC, Xu YN, et al. Analysis of Drug-Resistance Characteristics and Genetic Diversity of Multidrug-Resistant Tuberculosis Based on Whole-Genome Sequencing on the Hainan Island, China. *Infect Drug Resist*. 2023;16:5783–5798. doi:10.2147/IDR.S423955
7. Cao DM, Zhang ZG, Yang Z, et al. The association between tuberculin skin test result and active tuberculosis risk of college students in Beijing, China: a retrospective cohort study. *Bmc Infect Dis*. 2019;19(1):19. doi:10.1186/s12879-018-3624-5
8. Pan DX, Lan RS, Graviss EA, et al. Adolescent tuberculosis associated with tuberculosis exposure in classrooms and dorm rooms in Guangxi, China. *Int J Infect Dis*. 2019;78:8–14. doi:10.1016/j.ijid.2018.09.019
9. Stosic M, Playsa D, Mavroeidi N, et al. Tuberculosis outbreak among high school students in Novi Pazar, Serbia 2016: a retrospective-cohort study. *J Infect Dev Countr*. 2019;13(2):101–110. doi:10.3855/jidc.10952
10. Ilic M, Spahic S, Spahic M, et al. Tuberculosis outbreak in a grammar school, Serbia, 2016. *Ann I Super Sanita*. 2019;55(1):55–58.
11. Cegolon L, Mastrangelo G, Gentili D, et al. Tuberculosis in schools: an outbreak in northeastern Italy and some key health protection interventions. *Croat Med J*. 2021;62(1):90–94. doi:10.3325/cmj.2021.62.90
12. Sobkowiak B, Romanowski K, Sekirov I, et al. Comparing transmission reconstruction models from whole genome sequence data. *Epidemiol Infect*. 2023;151:e105.
13. Song ZX, He WC, Cao XL, et al. The Recent Transmission and Associated Risk Factor of in Golmud City, China. *Infect Drug Resist*. 2024;17:417–425. doi:10.2147/IDR.S437026

14. Li YF, Yang Y, Kong XL, et al. Transmission dynamics and phylogeography of in China based on whole-genome phylogenetic analysis. *Int J Infect Dis*. 2024;140:124–131. doi:10.1016/j.ijid.2023.10.015

15. Zhao JN, Qian CY, Jiang YQ, et al. Drug-Resistant Characteristics, Genetic Diversity, and Transmission Dynamics of Multidrug-Resistant in Jiangxi, China. *Infect Drug Resist*. 2024;17:2213–2223. doi:10.2147/IDR.S460267

16. Nelson KN, Talarico S, Poonja S, et al. Mutation of and Implications for Using Whole-Genome Sequencing for Investigating Recent Tuberculosis Transmission. *Front Public Health*. 2022;9:1. doi:10.3389/fpubh.2021.790544.

17. Liu DX, Huang F, Li YR, et al. Transmission characteristics in Tuberculosis by WGS: nationwide cross-sectional surveillance in China. *Emerg Microbes Infec*. 2024;13(1):2348505.

18. Li MJ, Zhang YY, Wu ZY, et al. Transmission of fluoroquinolones resistance among multidrug-resistant tuberculosis in Shanghai, China: a retrospective population-based genomic epidemiology study. *Emerg Microbes Infec*. 2024;13(1):2302837.

19. Zhang MW, Lu YW, Zhu YL, et al. Whole-Genome Sequencing to Predict Drug Resistance: a Retrospective Observational Study in Eastern China. *Antibiotics-Basel*. 2023;12(8):1.

20. Yu SS, Pan Y, Chen QP, et al. Analysis of the epidemiological characteristics and influencing factors of tuberculosis among students in a large province of China, 2008-2018. *Sci Rep*. 2024;14(1):20472.

21. Zhou ZL, Yi HM, Zhou QR, et al. Evolution and epidemic success of Mycobacterium tuberculosis in eastern China: evidence from a prospective study. *Bmc Genomics*. 2023;24(1):241. doi:10.1186/s12864-023-09312-6

22. Guo H, An J, Li S, et al. Transmission and resistome of extremely drug-resistant tuberculosis in Beijing, China: a retrospective population-based epidemiological study. *J Infect Public Health*. 2023;16(8):1193–1200. doi:10.1016/j.jiph.2023.05.020

23. Che Y, Lin Y, Yang TC, et al. Evaluation of whole-genome sequence to predict drug resistance of nine anti-drugs and characterize resistance genes in clinical rifampicin-resistant isolates from Ningbo, China. *Front Public Health*. 2022;10:956171.

24. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–2120. doi:10.1093/bioinformatics/btu170

25. Yang CG, Luo T, Shen X, et al. Transmission of multidrug-resistant in Shanghai, China: a retrospective observational study using whole-genome sequencing and epidemiological investigation. *Lancet Infect Dis*. 2017;17(3):275–284. doi:10.1016/S1473-3099(16)30418-2

26. Maladan Y, Krismawati H, Wahyuni T, et al. The whole-genome sequencing in predicting drug updates susceptibility and resistance in Papua, Indonesia. *Bmc Genomics*. 2021;22(1). doi:10.1186/s12864-021-08139-3.

27. Aung HL, Nyunt WW, Fong Y, et al. Genomic Profiling of Mycobacterium tuberculosis Strains, Myanmar. *Emerg Infect Dis*. 2021;27(11):2847–2855. doi:10.3201/eid2711.210726

28. Luo D, Yu SM, Huang YY, et al. Recent Transmission and Prevalent Characterization of the Beijing Family in Jiangxi, China. *Pol J Microbiol*. 2022;71(3):371–380. doi:10.33073/pjm-2022-033

29. Cao XL, Song ZX, He WC, et al. Tuberculosis screening characteristics amongst freshmen in Changping District, Beijing, China. *Bmc Infect Dis*. 2023;23(1). doi:10.1186/s12879-023-08802-y.

30. Fan YF, Liu DX, Chen YW, et al. Inferring Drug Resistance and Transmission using Whole-genome Sequencing in a High TB-burden Setting in China. *Biomed Environ Sci*. 2024;37(2):157–169. doi:10.3967/bes2023.116

31. Tseng ST, Tai CH, Li CR, et al. The mutations of katG and inhA genes of isoniazid-resistant Mycobacterium tuberculosis isolates in Taiwan. *J Microbiol Immunol Infect*. 2015;48(3):249–255. doi:10.1016/j.jmii.2013.08.018

32. Chen L, Gan X, Li N, et al. rpoB gene mutation profile in rifampicin-resistant Mycobacterium tuberculosis clinical isolates from Guizhou, one of the highest incidence rate regions in China. *J Antimicrob Chemother*. 2010;65(6):1299–1301. doi:10.1093/jac/dkq102

33. Tang K, Sun H, Zhao Y, et al. Characterization of rifampin-resistant isolates of Mycobacterium tuberculosis from Sichuan in China. *Tuberculosis*. 2013;93(1):89–95. doi:10.1016/j.tube.2012.10.009

34. Siu GK, Zhang Y, Lau TC, et al. Mutations outside the rifampicin resistance-determining region associated with rifampicin resistance in Mycobacterium tuberculosis. *J Antimicrob Chemother*. 2011;66(4):730–733. doi:10.1093/jac/dkq519

35. Seifert M, Catanzaro D, Catanzaro A, et al. Genetic mutations associated with isoniazid resistance in Mycobacterium tuberculosis: a systematic review. *PLoS One*. 2015;10(3):e0119628. doi:10.1371/journal.pone.0119628

36. He W, Tan Y, Liu C, et al. Drug-Resistant Characteristics, Genetic Diversity, and Transmission Dynamics of Rifampicin-Resistant Mycobacterium tuberculosis in Hunan, China, Revealed by Whole-Genome Sequencing. *Microbiol Spectr*. 2022;10(1):e0154321.

37. Li YF, Kong XL, Song WM, et al. Genomic analysis of lineage-specific transmission of multidrug resistance tuberculosis in China. *Emerg Microbes Infec*. 2024;13(1):2294858.

38. Liu QY, Ma AJ, Wei LH, et al. China's tuberculosis epidemic stems from historical expansion of four strains. *Nat Ecol Evol*. 2018;2(12):1982–1992. doi:10.1038/s41559-018-0680-6