*JAC Antimicrob Resist*
https://doi.org/10.1093/jacamr/dlaf064

JAC-Antimicrobial Resistance

# Genomic characterization of multidrug-resistant tuberculosis in Shanghai, China: antibiotic resistance, virulence and transmission

**Yang Guo[1,2], Jing Lu[3], Peipei Jin[4], Zhipeng Qiu[5], Fangyou Yu[6]†, Yan Zhu[7]† and Jiayuan Huang[1]*†**

*[1]School of Medicine, Shenzhen Campus of Sun Yat-Sen University, Shenzhen, Guangdong 518107, China; [2]State Key Laboratory of Respiratory Disease, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, Guangdong 510120, China; [3]Institute of Infectious Disease, The Second Affiliated Hospital of Tianjin Medical University, Tianjin 300222, China; [4]Department of Laboratory Medicine, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 201801, China; [5]Independent Researcher, Shenzhen, Guangdong 518000, China; [6]Department of Laboratory Medicine, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai 200433, China; [7]Systems Biology Center, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China*

*Corresponding author. E-mail: huangjy389@mail.sysu.edu.cn
†Senior authors (equally contributed).

*Received 23 January 2025; accepted 10 April 2025*

**Objectives:** Whole-genome sequencing (WGS) was employed to investigate antibiotic resistance, virulence and transmission profiles of multidrug-resistant tuberculosis (MDR-TB) isolates from Shanghai, China.

**Methods:** A total of 306 MDR-TB clinical isolates were collected from Shanghai Pulmonary Hospital and underwent phenotypic drug susceptibility testing (DST) for common anti-TB drugs and WGS. Combined 778 published bacterial sequences, we performed phylogenetic analysis, resistance and virulence gene identification to understand the genetic relationships and resistance mechanisms among those strains.

**Results:** WGS determination, supported by DST, revealed high resistance rates for isoniazid (83.66%) and rifampicin (90.20%) among the MDR-TB isolates. Key resistance-associated mutations included *katG* Ser315Thr for isoniazid, *rpoB* mutations for rifampicin, and *embB* Met306Val for ethambutol. WGS demonstrated >90% concordance with culture-based DST for most drugs, except ethambutol that showed a 76.80% concordance. Analyses of virulence factors and phylogenetics revealed the genetically homogeneous, endemic MDR-TB population in Shanghai, with no evidence of recent transmission.

**Conclusions:** This study highlights the genetic homogeneity and endemic nature of MDR-TB in Shanghai, providing insights into key resistance mechanisms of TB.

## Introduction

Tuberculosis (TB) is a significant global health threat, with >8.2 million new TB cases reported in 2023 according to the World Health Organization (WHO).[1] This makes the highest number of cases ever recorded and a substantial increase from 7.5 million cases in 2022, positioning TB as the leading cause of death from infectious diseases ahead of COVID-19. Efforts to control the disease are further complicated by multidrug-resistant TB (MDR-TB), which is resistant to at least isoniazid and rifampicin.[1] Nearly half a million drug-resistant cases of MDR-TB are predominantly reported in high-burden nations, with India accounting for 26% of the global TB burden, followed by Indonesia (10%), China (6.8%), Philippines (6.8%) and Pakistan (6.3%).[2,3] Treatment for MDR-TB is time and cost intensive, and often associated with significant toxicity, with success rates of only 56% for MDR-TB and 39% for extensively drug-resistant TB (XDR-TB). These outcomes highlight the urgent necessity for rapid and accurate detection of drug resistance to prevent the further dissemination of XDR-TB.[2]

Drug susceptibility testing (DST) remains the gold standard for diagnosing drug-resistant TB.[3] However, its application is hampered by significant limitations, including a lengthy turnaround time of 4–6 weeks due to the slow growth rate of *Mycobacterium tuberculosis*, restricted drug coverage and stringent biosafety requirements.[4] To address these challenges,

molecular-based DST methods, such as the WHO-recommended GeneXpert MTB/RIF assay, have offered a faster alternative by detecting rifampicin resistance-associated *rpoB* mutations.[5] Despite their speed, these molecular assays are limited by their relative narrow diagnostic scope, the potential to overlook heteroresistance and the risk of misclassifying synonymous or silent mutations as indicators of drug resistance.[6] WGS has emerged as a transformative approach in pathogen diagnostics, providing high-resolution genotyping and precise pathogen identification.[7,8] Unlike traditional DST methods, WGS provides a comprehensive analysis of MTB genomic DNA to prediction of drug resistance mechanisms across multiple antimicrobials.[9] While WGS has primarily been applied in economically developed regions with low TB burden, its potential in determining resistance to first-line drugs is increasingly being recognized.[10]

In this study, we harness WGS to characterize the drug resistance profile of MTB in Shanghai, a city in China with a high TB burden, and critically assess the efficacy in determining drug resistance. This approach addresses the growing recognition of WGS as a vital tool for understanding MTB evolution and resistance mechanisms, as well as its potential in distinguishing between relapse and reinfection in prolonged TB cases.[11] By integrating WGS data with classical epidemiological interviews, we aim to provide an in-depth understanding of MTB transmission and resistance, which are essential for shaping public health strategies aimed to control outbreaks and reducing the TB incidence. Ultimately, this analysis contributes to global efforts in combating MDR-TB by advancing our understanding of resistance and transmission dynamics.

## Materials and methods

### Study design and data collection

From January 2013 to June 2019, a total of 306 MDR-TB clinical isolates were obtained from Shanghai Pulmonary Hospital, and stored at −80°C. These isolates underwent DST for three first-line anti-tuberculosis drugs: isoniazid, rifampicin and ethambutol, and three second-line drugs: fluoroquinolones, streptomycin and amikacin, as previously reported.[12,13] Data were electronically obtained from hospital medical and pharmacy records, then manually summarized by coordinating physicians for further analysis.

### Ethics

This study was approved by the Ethical Committee of Shanghai Pulmonary Hospital and Ruijin Hospital (Approval ID 2018-015-1).

### MDR-TB identification

Clinical MDR-TB isolates were identified using matrix assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) with Bruker BioTyper system (microflex LT; Bruker Daltonik GmbH, Bremen, Germany) at Shanghai Pulmonary Hospital. For analysis, bacterial colonies were placed on a stainless-steel MALDI target plate and covered with 1 μL of α-cyano-4-hydroxycinnamic acid matrix solution. The target plate was then air-dried for 10 minutes before insertion into the mass spectrometer. Spectral data were acquired in the linear positive ion mode, covering a mass-to-charge (*m/z*) ratio from 2000 to 20 000, using a nitrogen laser at 337 nm wavelength.[13] Each sample was analysed with 240 laser shots per spectrum, and spectra were compared against the Bruker reference database for species identification. System performance was verified using H37Rv as a reference strain for quality control prior to analysing the samples.[9,14]

### Antimicrobial susceptibility testing

In this study, six anti-tuberculosis drugs were tested using the rapid DST technology (BACTEC MGIT 960 system: BACTEC MGIT, Becton Dickinson, Cockeysville, MD, USA) in the clinical laboratory at Shanghai Pulmonary Hospital, an ISO 15189 accredited laboratory specialized in MTB detection. The laboratory participates in the inter-laboratory quality assessment of the National Tuberculosis Reference Laboratory and the Tuberculosis Prevention Alliance for DST every year and passes the test. The protocol involved vortexing 5 mL of a bacterial solution with glass beads in a centrifuge tube for 30 s, followed by the addition of 800 μL of the solution to 7-mL BBL MGIT tubes. The first tube served as the control, with subsequent tubes dedicated to the specific agents: the three first-line and the three second-line anti-tuberculosis drugs previously mentioned. Subsequently, 100 μL of the corresponding drug was added to the respective tubes, and 50 μL of diluted bacterial suspension was added to the control tube. Furthermore, 500 μL of bacterial solution was added to other tubes, and the turbidity of the bacterial solution was adjusted to the 0.5 McFarland standard. The tubes were incubated in the BACTEC MGIT 960 system, which facilitated automatic result interpretation and allowed for the retesting of samples with reported incorrect drug susceptibilities. The quality control strains used were the MTB H37Rv.[9] All steps were performed by trained and specialized technicians in a biosafety cabinet following the Clinical and Laboratory Standards Institute (CLSI) and the WHO guidelines.

### WGS sequencing, raw data processing and assembly

Frozen isolates were subcultured on Löwenstein–Jensen Medium, and genomic DNA was extracted for sequencing using the cetyltrimethylammonium bromide (CTAB) method of DNA purification[6,14] at Shanghai Pulmonary Hospital. Each extracted DNA was quantified by a Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA). Sequencing library preparations were constructed following the manufacturer's protocol (Illumina TruSeq DNA Nano Library Prep Kit). Then libraries with different indices were multiplexed and loaded on an Illumina HiSeq instrument according to manufacturer's instructions (Illumina, San Diego, CA, USA). Sequencing was carried out using a 2×150 paired-end (PE) configuration by Uni-medica Inc. (Shenzhen, Guangdong, China). The sequence reads were aligned to the reference strain MTB H37Rv (GenBank accession no. NC000962.3). Drug resistance and strain-type profiles directly from raw sequences were reported directly from raw sequences using an online TBProfiler tool (https://github.com/jodyphelan/TBProfiler).[15] The corresponding IDs of the sequencing samples are listed in Table S3 (available as Supplementary data at *JAC-AMR* Online). Raw sequencing reads were processed to obtain valid reads for further analysis. First, raw reads were processed using Fastp v.0.23.4 for adapter removing, quality assessment and low-quality reads trimming with default parameters.[10]

### Phylogenetic analysis

To elucidate the evolutionary relationships among 306 distinct MTB strains, a phylogenetic tree was meticulously assembled utilizing the RAxML (random accelerated maximum likelihood) software.[15] First, single nucleotide polymorphisms (SNPs) were identified across the genomes of the 306 stains, with a specific focus on high-confidence variants. SNPs located within PE/PPE regions, as well as those associated with insertion elements and repetitive DNA sequences, were deliberately excluded from the analysis. This exclusion was designed to mitigate any potential interference from genetic variations that do not contribute to evolutionary divergence, such as those that may arise due to insertional events or genomic instability. The remaining SNPs were then used to generate a

phylogenetic tree, employing RAxML with the GTR+G evolutionary model[16,17] to maximize the accuracy of the tree construction. The final tree was visualized using SplitsTree software[18] to clearly delineate the phylogenetic relationships among the strains, providing insights into their evolutionary history and potential transmission patterns.

### Resistance and virulence analysis

All contigs and other sequences were analysed to predict susceptibility to six anti-tuberculosis drugs. WGS was utilized to determine drug resistance with the TBProfiler.[19] Additionally, antimicrobial resistance, virulence and resistance genes were identified using the VFDB[20] (Virulence Factor Database) and the CARD[21] (Comprehensive Antibiotic Resistance Database), both of which are databases provided by Abricate v.1.0.1. The outcomes of the resistance and virulence analyses were integrated using the 'Abricate -summary' function, which facilitated the calculation of identity with the genes in the database.

### Average nucleotide identity (ANI) analysis

For ANI calculations, we employed SKani,[22] which estimates sequence identity using a sparse *k*-mer chaining approach. Genomes were fragmented into 20-kb nonoverlapping chunks, and a sparse set of *k*-mers was used to seed chains against a reference genome. Orthologous chains were extracted, and ANI was estimated for each chunk, with a mean ANI output and a learned debiasing step applied. SKani was executed with the following parameters: SKani dist genome1.fa genome2.fa -o output_file.txt, which facilitated rapid querying against a database of >65 000 prokaryotic genomes within seconds and 6 GB of memory.

### Statistical analysis

Statistical analysis was performed using MedCalc software.[23] Key performance metrics, including true positives, false positives, true negatives and false negatives, were calculated in relation to WGS. Concordance, sensitivity and specificity were determined and compared, with a 95% confidence interval established for all statistical assessments.

## Results

### WGS determination of in antimicrobial resistance of MDR-TB

WGS determination of drug susceptibility using TBProfiler were complemented by DST for six anti-TB drugs. The minimum inhibitory concentration (MIC) results and phenotypic data are detailed in Table 1 and Table S1. As shown in Table 1, R/R indicates the rate of resistance for both MIC phenotypic and TBProfiler predicted values. Accordingly, for the two first-line anti-tuberculosis drugs, the clinical resistance rates for isoniazid and rifampicin have reached alarmingly high levels, with R/R values of 76.45% and 85.95%, respectively. Additionally, the second-line drug streptomycin exhibits a concerning resistance rate, with an R/R value of 65.36% (Table 1). In contrast, the tested strains showed relatively high susceptibility to ethambutol, fluoroquinolones and amikacin, with resistance rates remaining below 40%.

Table 1 shows the drug resistance profile of all 306 clinical MTB isolates based on culture-based DST and WGS. Overall, 93.14% (285/306) were resistant to at least one drug, with only 21 isolates being susceptible to all six drugs. Specifically, 90.85% (278/306) were resistant to at least one first-line drug, and 77.12% (236/306) were resistant to at least one second-line drug. The total resistance rates were as follows: isoniazid (256, 83.66%), rifampin (276, 90.20%), ethambutol (118, 38.56%), fluoroquinolones (125, 40.85%), streptomycin (210, 68.63%) and amikacin (34, 11.11%). For monoresistance, the number of rifampin-resistant isolates was the highest (11, 3.59%). Additionally, 83.01% (254/306) of TB cases were diagnosed as MDR-TB, becoming the main focus of this study, and 11.11% (34/306) of isolates were diagnosed as XDR-TB, while 6.21% (19/306) of isolates were diagnosed as TDR-TB (Table 1).

**Table 1.** Antimicrobial agent susceptibilities of 306 *Mycobacterium tuberculosis* isolates

| Antimicrobial agent | Count of resistant isolates based on DST | Percentage of resistant isolates based on DST (%) | Count of resistant isolates based on WGS | Percentage of resistant isolates based on WGS (%) |
|---|---|---|---|---|
| Isoniazid[a] | 256 | 83.66 | 241 | 78.76 |
| Rifampicin[a] | 276 | 90.20 | 263 | 85.95 |
| Ethambutol[a] | 118 | 38.56 | 177 | 57.84 |
| Fluoroquinolones[b] | 125 | 40.85 | 120 | 39.22 |
| Streptomycin[b] | 210 | 68.63 | 205 | 66.99 |
| Amikacin[b] | 34 | 11.11 | 27 | 8.82 |

The MIC thresholds for DST were determined based on references.[22–30] The specific MIC thresholds are as follows: Isoniazid (S ≤ 100 mg/L, R ≥ 1000 mg/L); Rifampicin (S ≤ 1000 mg/L, R ≥ 4000 mg/L); Ethambutol (S ≤ 2000 mg/L, R ≥ 8000 mg/L); Fluoroquinolones (S ≤ 250 mg/L, R ≥ 1000 mg/L); Streptomycin (S ≤ 2000 mg/L, R ≥ 8000 mg/L); and Amikacin (S ≤ 2000 mg/L, R ≥ 8000 mg/L).
[a]First-line drugs for MDR-TB.
[b]Second-line drugs for MDR-TB.

### Drug resistance-related mutation gene patterns

Quality control of the sequencing data for the 306 MTB isolates showed that an average genome sequencing depth ranging from 51.67 to 238.11 (mean 166.29). The genome 10× coverage was from 98.69 to 99.95 (mean 99.19), and the genome 1× coverage was from 99.12 to 99.99 (mean 99.38).

Genes and mutations associated with resistance to the anti-tuberculosis drugs in MDR-TB were identified by WGS (Table 2 and Table S2). A total of 242 isolates (79.08%) had mutations in genes associated with resistance to isoniazid, including *katG* ($n = 195$), the promoter of *fabG1* ($n = 40$), *ahpC* ($n = 6$) and *inhA* ($n = 1$). Among isolates, 24 strains had more than one isoniazid resistance-associated mutation. The most common mutations were *katG* Ser315Thr ($n = 184$) and *fabG1* −15C>T ($n = 34$). Rifampicin resistance-associated mutations were identified in 274 isolates (89.54%), all had mutations in the *rpoB* gene. In addition, there were 11 isolates with *rpoC* mutations in addition to *rpoB* mutations. Also, 18 mutations in the *rpoB* gene were not in the 81-bp *rpoB* rifampicin resistance-determining region (RRDR), precisely Ala286Val ($n = 5$), Val170Phe ($n = 3$), Ile480Val ($n = 3$), Glu761Asp ($n = 2$), Arg552Cys ($n = 1$), Ile491Phe ($n = 1$), Glu460Gly ($n = 1$), Pro483Leu ($n = 1$) and Thr400Ala ($n = 1$). Among them, 15 isolates had mutations in RRDR, while three

**Table 2.** Genetic variations associated with resistance to the six anti-tuberculosis drugs in MDR-TB

| Antimicrobial agent | Mutation gene | Mutation type | No. of isolates | Mutation rate (%) |
|---|---|---|---|---|
| Isoniazid | *ahpC* | *ahpC c.-48G>A* | 2 | 0.83 |
| | | *ahpC c.-52C>T* | 1 | 0.41 |
| | | *ahpC c.-54C>T* | 1 | 0.41 |
| | | *ahpC c.-74G>A* | 1 | 0.41 |
| | | *ahpC c.-81C>T* | 1 | 0.41 |
| | *fabG1* | *fabG1 c.-8T>C* | 6 | 2.48 |
| | | *fabG1 c.-15C>T* | 34 | 14.05 |
| | *inhA* | *inhA p.Ile21Val* | 1 | 0.41 |
| | *katG* | *katG c.-10A>C* | 11 | 4.55 |
| | | *katG p.Ser315Thr* | 184 | 76.03 |
| Rifampicin | *rpoB* | *rpoB p.Ala286Val* | 5 | 1.82 |
| | | *rpoB p.Asp435Ala* | 10 | 3.65 |
| | | *rpoB p.Gln432Lys* | 8 | 2.92 |
| | | *rpoB p.Val170Phe* | 3 | 1.09 |
| | | *rpoB p.Ile480Val* | 3 | 1.09 |
| | | *rpoB p.Ile491Val* | 1 | 0.36 |
| | | *rpoB Glu761Asp* | 2 | 0.73 |
| | | *rpoB Glu460Asp* | 1 | 0.36 |
| | | *rpoB Pro483Leu* | 1 | 0.36 |
| | | *rpoB Thr400Ala* | 1 | 0.36 |
| | | *rpoB Arg552Cys* | 1 | 0.36 |
| | | *rpoB p.Ser450Leu* | 229 | 89.60 |
| | *rpoC* | *rpoC p.Asp747Ala* | 5 | 2.23 |
| | | *rpoC p.Leu527Val* | 2 | 0.74 |
| | | *rpoC p.Gly332Arg* | 2 | 0.74 |
| Ethambutol | *embA* | *embA c.-11C>A* | 2 | 1.18 |
| | | *embA c.-12C>T* | 4 | 2.23 |
| | | *embA c.-16C>T* | 12 | 6.70 |
| | *embB* | *embB p.Asp354Ala* | 14 | 7.82 |
| | | *embB p.Gly406Asp* | 42 | 23.46 |
| | | *embB p.Met306Val* | 69 | 38.55 |
| | | *embB p.Met306Ile* | 36 | 20.11 |
| Fluoroquinolones | *gyrA* | *gyrA p.Ala90Val* | 17 | 16.50 |
| | | *gyrA p.Asp94Gly* | 73 | 70.87 |
| | | *gyrA p.His70Arg* | 1 | 0.97 |
| | | *gyrA p.Ser91Pro* | 4 | 3.88 |
| | *gyrB* | *gyrB p.Glu501Asp* | 3 | 2.91 |
| | | *gyrB p.Arg446His* | 2 | 1.94 |
| | | *gyrB p.Ile486Leu* | 2 | 1.94 |
| | | *gyrB p.Asn499Thr* | 1 | 0.97 |
| Streptomycin | *gid* | *gid c.353 354insG* | 2 | 0.98 |
| | | *gid c.115 115del* | 1 | 0.49 |
| | *rpsL* | *rpsL p.Lys43Arg* | 172 | 84.31 |
| | *rrs* | *rrs r.514a>c* | 17 | 8.33 |
| | | *rrs r.517c>t* | 12 | 5.88 |
| Amikacin | *rrs* | *rrs r.1401a>g* | 23 | 69.70 |

isolates (Ile491Phe, Glu460Gly, Thr400Ala) had mutations outside RRDR. There were 179 strains with ethambutol genetic resistance: 18 isolates had *embA* mutations and 161 isolates had *embB* mutations. Among them, 10 isolates had both *embA* and *embB* mutations, and five had mutations at two different *embB* sites. The most common mutations were *embB* Met306Val and Met306Ile, accounting for 38.55% (69/179) and 20.11% (36/179), respectively. For fluoroquinolones, 95 isolates (92.23%) showed mutations in the *gyrA* gene, with the mutation Asp94Gly being the most common one ($n=73$), and rare mutations at codon 70 and codon 89 in three isolates. Mutations in *gyrB* were found in eight isolates. Streptomycin resistance was associated with mutations in the *gid*, *rpsL* and *rrs* genes, occurring in 3 (1.47%), 172 (84.31%) and 29 (14.22%) isolates,

**Table 3.** WGS compared with DST in drug resistance diagnosis of MDR-TB

| Antimicrobial agent | True positive | False positive | False negative | True negative | Concordance (%) | Sensitivity (%) | Specificity (%) |
|---|---|---|---|---|---|---|---|
| Isoniazid | 237 | 4 | 19 | 46 | 92.48 | 92.58 | 92.00 |
| Rifampicin | 263 | 0 | 13 | 30 | 95.75 | 95.29 | 100.00 |
| Ethambutol | 112 | 65 | 6 | 123 | 76.80 | 94.92 | 65.43 |
| Fluoroquinolones | 113 | 7 | 12 | 174 | 93.79 | 90.40 | 96.13 |
| Streptomycin | 200 | 5 | 10 | 91 | 95.10 | 95.24 | 94.79 |
| Amikacin | 26 | 1 | 8 | 271 | 97.06 | 76.47 | 99.63 |

respectively. Additionally, all 23 amikacin-resistant isolates exhibited mutations in the *rrs* gene, specifically the *1401A>G* mutation.

### *Consistency comparison of phenotypic and genotypic DST*

Using the culture-based DST as the gold standard, we evaluated the capability of WGS to determine MDR-TB drug resistance for six drugs. True positives, false positives, false negatives, true negatives, concordance, sensitivity and specificity for each drug are shown in Table 3, and parameters and formulas involved in the study are shown in Table S4. For 306 isolates tested, we found an average concordance of 91.83% across all six drugs, ranging from 76.80% (ethambutol) to 97.06% (amikacin), an average sensitivity of 90.82%, ranging from 76.47% (amikacin) to 95.29% (rifampicin) and an average specificity of 91.33%, ranging from 65.43% (ethambutol) to 100.00% (rifampicin). The drug resistance predictive values of isoniazid, rifampin, fluoroquinolones, streptomycin and amikacin were all high, all of which exceeded 90%.

Seventy-two isolates showed drug resistance-associated mutations despite being classified as susceptible by culture-based DST. Among these isolates, mutations associated with isoniazid resistance were identified in *katG*, specifically at Ser140Asn ($n=2$), Tyr155Cys ($n=1$), Ser140Asn (along with *fabG1*-15C>T, $n=1$); for ethambutol at *embA*-11C>A ($n=1$), *embA*-12C>T ($n=3$), *embA*-16C>T ($n=3$), *embA*-16C>G ($n=1$), *embB* Asp1024Asn ($n=3$), *embB* Asp328Tyr ($n=1$), *embB* Asp354Ala ($n=3$), *embB* Gln497Arg ($n=3$), *embB* Gly406Ala ($n=3$), *embB* Gly406Asp ($n=8$), *embB* Gly406Ser ($n=3$), *embB* His1002Arg ($n=2$), *embB* Met306Ile ($n=15$), *embB* Met306Leu ($n=3$), *embB* Met306Val ($n=10$), *embB* Ser347Ile ($n=1$), *embB* Tyr334His ($n=1$), *embA*-16C>T+*embB* Gln497Arg ($n=1$), *embB* Gly406Asp+*embB* Gln497Lys ($n=1$) and *embB* Met306Ile+*embB* Gly406Asp ($n=1$); for ofloxacin at *gyrB* Arg446Cys ($n=1$), *gyrB* Ala504Val ($n=1$); for streptomycin at *gid* 115del ($n=1$), *rpsL* Lys88Arg ($n=1$), *rrs* 514A>C ($n=1$) and *rrs* 51C>T ($n=1$).

Conversely, 33 isolates that were drug-resistant according to culture-based DST ($n=14$ for isoniazid, $n=8$ for rifampicin, $n=3$ for ethambutol, $n=9$ for ofloxacin, $n=6$ for streptomycin and $n=9$ for amikacin) showed no known drug resistance mutations using WGS.

### *Phylogenetic classification of drug resistance*

After rigorous quality control that excluded 23 isolates, the remaining 283 MTB isolates were genetically analysed alongside 778 publicly available strains from NCBI repository,[13] uncovering close phylogenetic relationships among strains of diverse geographic origins. Figure 1 presents the phylogenetic tree of these 1061 MTB isolates, delineating drug resistance profiles, Hb type classification of strains based on the genetic characteristics and lineages pinpointed by 48 495 high-confidence SNPs. A genetic distance threshold of 12 SNPs or fewer was established to signify recent transmission events.[24] The findings disclosed no evidence of recent transmission among the strains from this collection, underscoring an evolutionary relationship between the isolates gathered here and those historically documented in China.

### *Genetic homogeneity and endemicity of MDR-TB strains*

The SKanI analysis of the current study demonstrates exceedingly high genetic identity among Shanghai MDR-TB isolates, with ANI values predominantly exceeding 99.95% when compared to historical strains from China (Figure 2a–d). This significant genetic correlation underscores the endemism of the disease and suggests that the strains sequenced in this study can find corresponding prototypes in past outbreaks within the country. The homogeneity observed implies a clonal expansion of a resistant lineage, indicating that our isolates are not novel imports but rather part of an established, circulating population.

### *Comparative genomic analysis of virulence factors and resistance genes*

The genomic analysis identified 68 virulence factors with the identity of >99.95%, highlighting their conservation among MDR-TB isolates (Figure 3a). Conversely, based on CARD database alignment, 11 distinct resistance genes were identified: *APH(3′)-Ie*, *AAC(2′)-Ib*, *RbpA*, *Erm.37*, *mfpA*, *embB*, *efpA*, *rpoB*, *tsnr*, *murA* and *mtrA* (Figure 3b). These results indicated a higher frequency of resistance genes, reflecting the genetic landscape shaped by antibiotic selection pressure. The prevalence underscored the genetic basis of drug resistance in MDR-TB and the necessity for genomic-informed strategies to address these resistance mechanisms.

## Discussion

WGS has emerged as a powerful tool capable of rapidly providing comprehensive genetic drug resistance profiles and has demonstrated robust performance in determining drug resistance, particularly for first-line drugs.[25] In this study, we leveraged WGS to determine drug resistance in 306 clinical MDR-TB isolates from

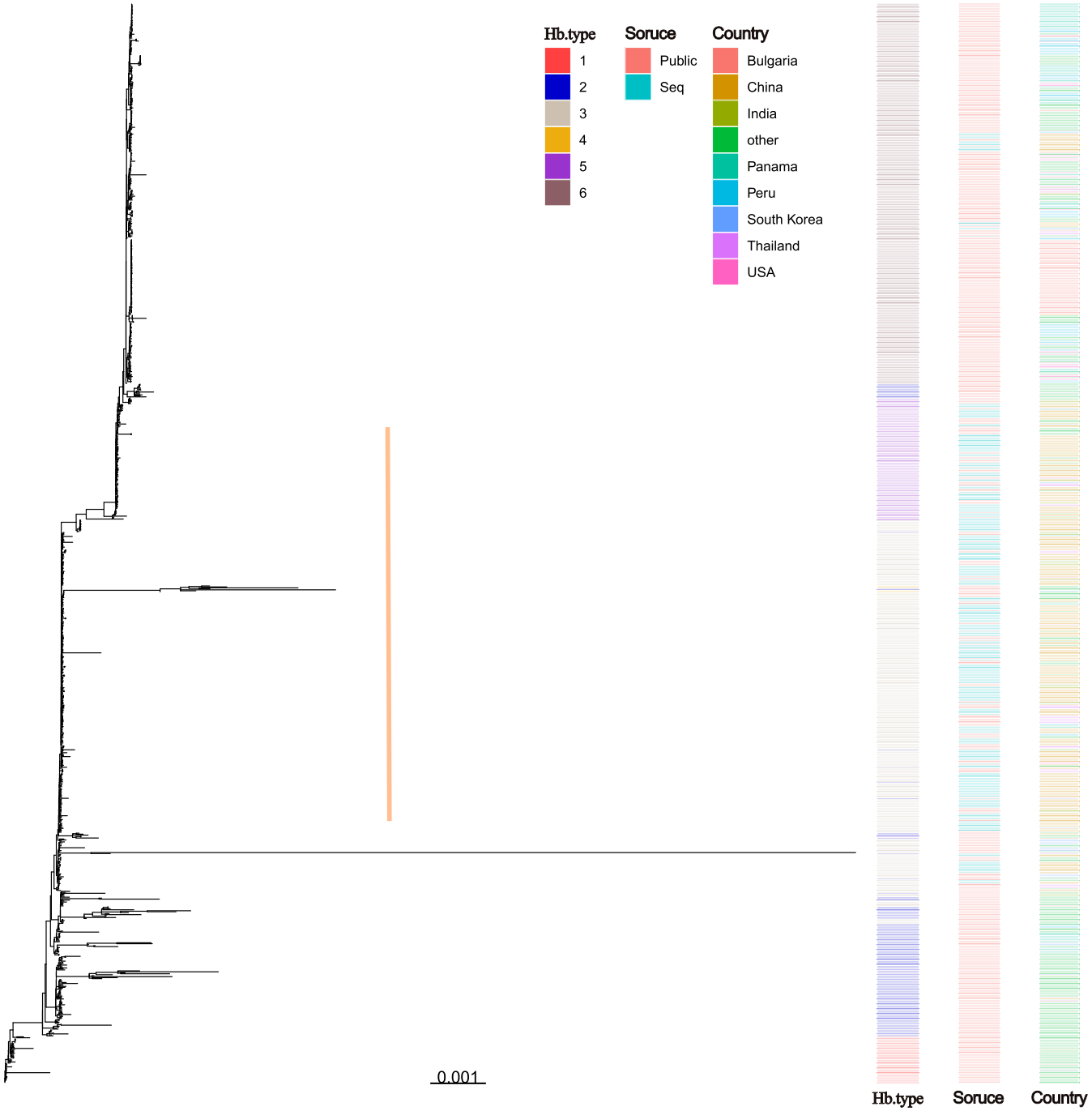


**Figure 1.** Phylogenetic tree of 1061 MTB isolates indicating Hb type and lineages with 48 495 high-confidence SNPs.

Shanghai and analysed their phylogenetic relationship combining with 778 strains from public database MDR-TB. This study highlights the efficacy of WGS in accurately determining resistance profiles, and underscores its significance in guiding tailored treatment strategies and addressing the growing challenge of MDR-TB in high-burden regions.

Our study corroborates previous research by confirming the *katG* Ser315Thr mutation as a key determinant of isoniazid resistance.[26] This mutation was observed in 76.03% of our samples, and it is associated with a high-level isoniazid resistance in 94.36% of cases, a finding that is consistent with global trends. The prevalence of this mutation, along with those in *fabG1*, which together were identified in 97.11% of our genotypic resistant isolates, underscores their predictive value for isoniazid resistance. Notably, the *katG* Ser315Thr mutation emerged as the most prevalent, highlighting its critical role in conferring resistance to

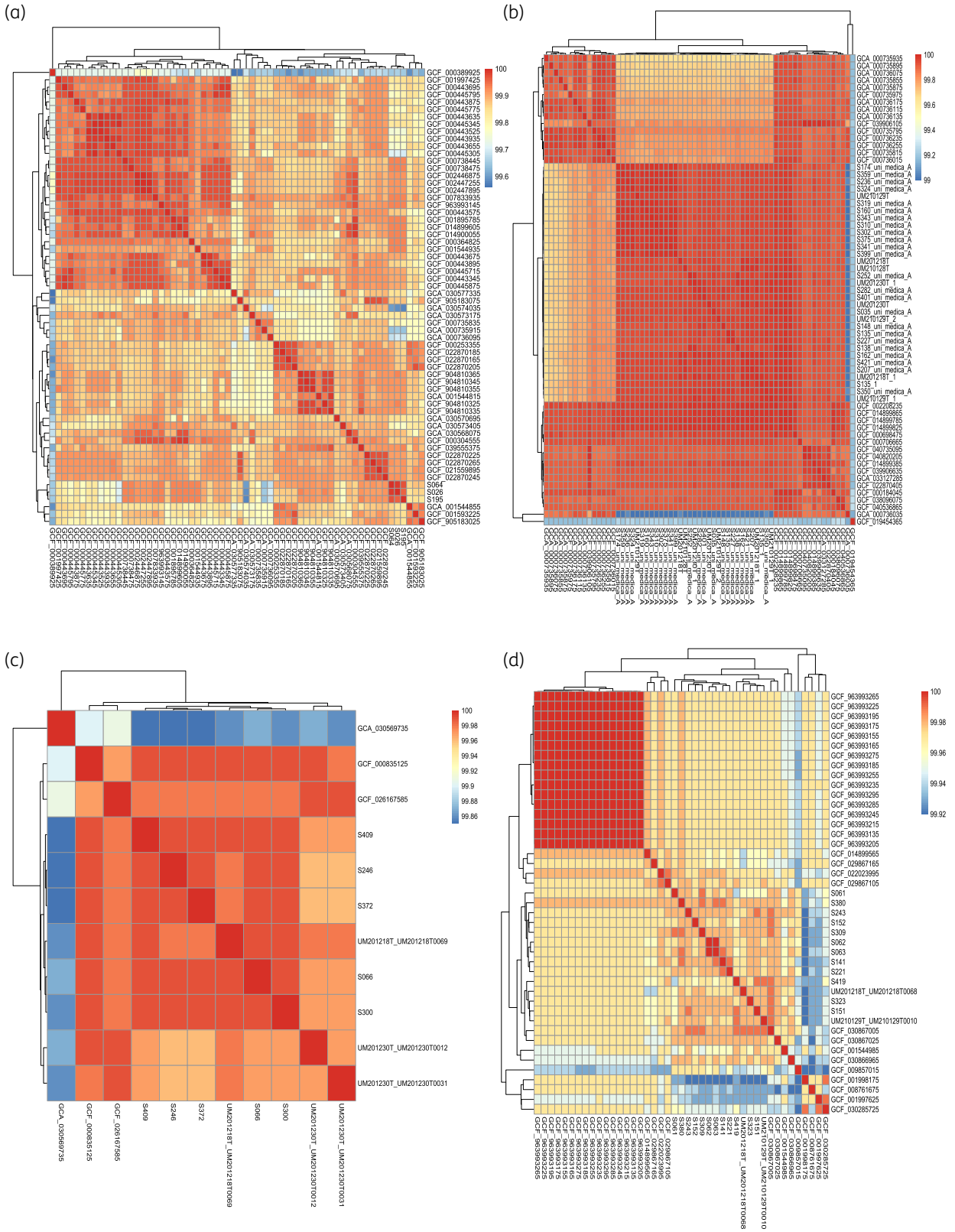


**Figure 2.** ANI-based phylogenetic relationships of MDR-TB in Shanghai. The sequence consistency results for SKani_g7 group (a), SKani_g10 group (b), SKani_g17 group (c), and SKani_g24 group (d) are represented.

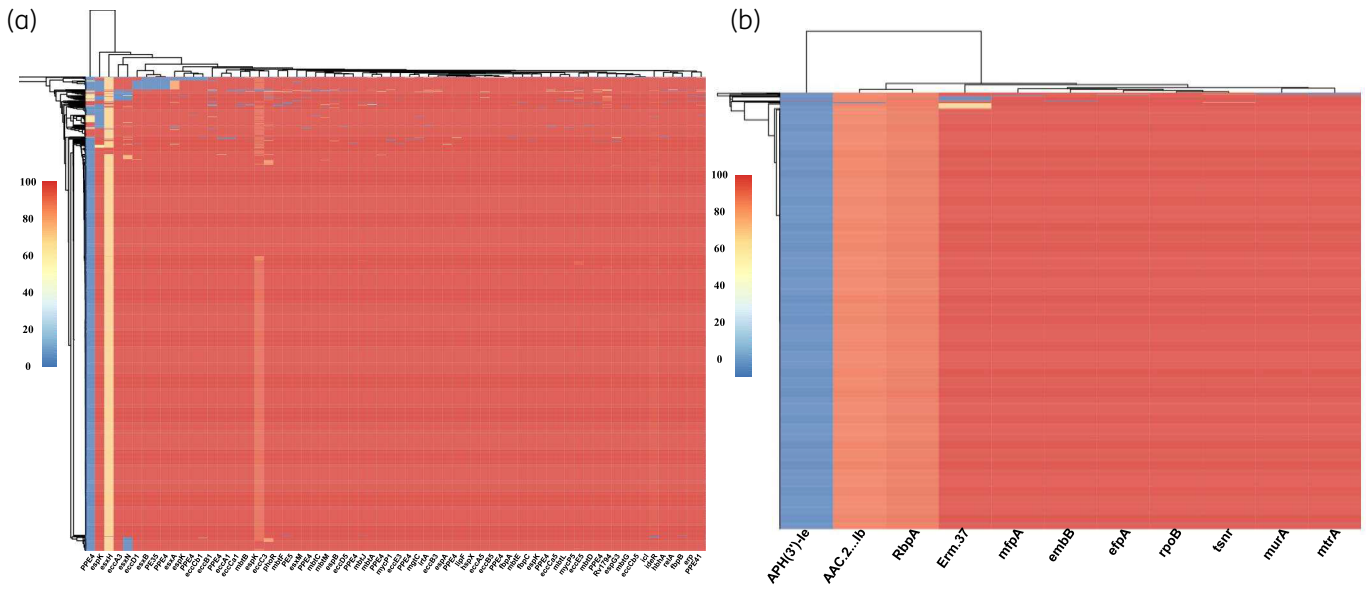


**Figure 3.** MDR-TB genomic profile: virulence factors and antibiotic resistance. (a) The genomic analysis results identifying 68 conserved virulence factors in MDR-TB isolates, with an identity of >99.95% compared to the reference sequences in the VFDB. (b) The identification of 11 distinct resistance genes in MDR-TB isolates, aligned with the CARD.

isoniazid. Consistent with existing literature, our findings indicate that over 96% of rifampicin resistance is attributable to mutations within the 81-base pair region of the *rpoB* gene, known as the RRDR.[27] However, our analysis extends these findings by identifying three rifampicin-resistant isolates with mutations outside the RRDR, a revelation that underscores the limitations of molecular DST, such as the GeneXpert MTB/RIF.[28] Specifically, the three RRDR-external missense mutations—Ile491Phe (Ile → Phe), Glu460Gly (Glu → Gly) and Thr400Ala (Thr → Ala)—may incorrectly classify isolates with RRDR-external mutations as susceptible, suggesting a risk of false negatives in the detection of rifampicin resistance.[29] Furthermore, our findings indicated that synonymous or silent mutations within the RRDR could also result in false-positive resistance readings, adding another layer of complexity to the accurate detection of rifampicin resistance. For ethambutol, *embB* codon 306 mutations were the most common in clinical isolates resistant to ethambutol. The most frequent mutations were *embB* Met306Val and Gly406Asp, with mutation rates of 38.55% and 23.46%, respectively. The primary mechanism of MTB resistance to fluoroquinolones is chromosomal mutations located in the quinolone resistance-determining regions of *gyrA*[30,31] or *gyrB*.[32,33] In our study, the most common mutations in fluoroquinolone-resistant isolates were at positions 90 and 94 of *gyrA*, with fewer mutations at codons 70, 89 and 91, consistent with other studies. Consistent with other studies, there are three streptomycin-related resistance genes, namely *rrs*, *rpsL* and *gid*, encoding 16S rRNA,[34] ribosomal protein S12[35] and 16S rRNA-specific methyltransferase,[36] respectively. In streptomycin-resistant isolates, mutations in *rpsL* were the most frequent and associated with high-level streptomycin resistance, followed by *rrs*, associated with moderate-level streptomycin resistance. Mutations in *gid* were rare, occurring in only three streptomycin genotypic resistant isolates. It is believed that mutations in *gid* are associated with low-level streptomycin resistance. However, it must be considered that this may be related to the small number of *gid* mutant strains.[37] For the 23 amikacin genotypic resistant isolates, all harboured the *1401a>g* mutation in *rrs*, which is associated with the global resistance of 70%–80% of MTB isolates to amikacin and kanamycin[38,39] and 60% resistance to kanamycin. Our study not only reaffirms the established associations between specific genetic mutations and drug resistance in MTB, but also reveals new insights, particularly regarding the detection limitations of current molecular DST tests and the need for a more nuanced understanding of resistance mechanisms to guide treatment strategies effectively.

In this study, WGS demonstrated high sensitivity and specificity, exceeding 90%, for determining resistance to isoniazid, rifampicin, fluoroquinolones and streptomycin, with a remarkable concordance of >92% with culture-based DST for these drugs, as corroborated by SKani and phylogenetic analyses. However, for ethambutol, the concordance was significantly lower at 76.80%, suggesting a need for improved specificity in WGS determination.[40] This discrepancy may be attributed to the inherent limitations of phenotypic DST for ethambutol and the complex interplay of resistance mechanisms, including efflux pumps,[41,42] which are not fully understood.

## Conclusion

In conclusion, our study elucidates the complex resistance profiles and virulence patterns of MDR-TB in Shanghai, providing critical insights into the genetic determinants of drug resistance and the transmission dynamics of resistant strains.

(i) A comprehensive WGS analysis of MTB drug resistance was conducted, revealing detailed resistance mechanisms and transmission patterns.

(ii) High prevalence of key mutations (e.g. *katG* Ser315Thr for isoniazid, *rpoB* for rifampicin) informs tailored diagnostic tests and treatment strategies.
(iii) WGS demonstrated >90% concordance with culture-based DST for most drugs, except ethambutol, which showed a 76.80% concordance.
(iv) Phylogenetic analysis revealed a genetically homogeneous and endemic MDR-TB population in Shanghai, with no evidence of recent transmission.

## Acknowledgements

We thank the medical and laboratory staff for their valuable support and collaboration in this research.

## Funding

This project was funded by the National Natural Science Foundation of China (32300043), the Open Project of State Key Laboratory of Respiratory Disease (SKLRD-OP-202408), Postdoctoral Fellowship Program of CPSF(GZC20230604) and the National Key Laboratory of Respiratory Disease's Annual Independent Project Fund for 2025 (SKLRD-Z-202518).

## Transparency declarations

The authors declare no conflict of interest.

### *Author contributions*

J.H., Y.Z. and F.Y. conceptualized and supervised this study. Y.G., P.J., J.L. and Z.Q. performed experiments and analysed data. Y.G. and J.H. wrote the manuscript. All authors revised the manuscript and approve the submission.

## Data availability

All sequencing data have been deposited with the National Center for Biotechnology Information (NCBI) under BioProject Accession ID PRJNA1225886. The relevant data analysis code has been uploaded to GitHub (https://github.com/huangjy389/MTB_genomics).

## Supplementary data

Tables S1–S4 are available as Supplementary data at *JAC-AMR* Online.

## References

**1** World Health Organization. Geneva, 2024. Global tuberculosis report 2024, Licence: CC BY-NC-SA 3.0 IGO. https://iris.who.int/bitstream/handle/10665/379339/9789240101531-eng.pdf?sequence=1

**2** Harichander S, Wiafe E, Mensah KB *et al*. The incidence of TB and MDR-TB in pediatrics and therapeutic options: a systematic review. *Syst Rev* 2022; **11**: 157. https://doi.org/10.1186/s13643-022-02023-1

**3** Chin KL, Anibarro L, Chang ZY *et al*. Impacts of MDR/XDR-TB on the global tuberculosis epidemic: challenges and opportunities. *Curr Res Microb Sci* 2024; **7**: 100295. https://doi.org/10.1016/j.crmicr.2024.100295

**4** Pfyffer GE, Wittwer F. Incubation time of mycobacterial cultures: how long is long enough to issue a final negative report to the clinician? *J Clin Microbiol* 2012; **50**: 4188–9. https://doi.org/10.1128/JCM.02283-12

**5** Hussain S, Sultan S, Riaz S *et al*. Synergy of Xpert (MTB/RIF) and line probe assay for detection of rifampicin resistant strains of *Mycobacterium tuberculosis*. *J Infect Dev Ctries* 2024; **18**: 1241–8. https://doi.org/10.3855/jidc.18945

**6** Luukinen B, Aittoniemi J, Miikkulainen-Lahti T *et al*. Evaluation of the STANDARD M10 MDR-TB and MTB/NTM assays for the detection of *Mycobacterium tuberculosis*, rifampicin and isoniazid resistance, and non-tuberculous mycobacteria in a low-incidence setting. *J Clin Microbiol* 2024; **62**: e0040224. https://doi.org/10.1128/jcm.00402-24

**7** Tang C, Wu L, Li M *et al*. High-throughput nanopore targeted sequencing for efficient drug resistance assay of *Mycobacterium tuberculosis*. *Front Microbiol* 2024; **15**: 1331656. https://doi.org/10.3389/fmicb.2024.1331656

**8** Li Y, Shao Y, Li Y *et al*. Association between toxin-antitoxin system mutations and global transmission of MDR-TB. *BMC Infect Dis* 2024; **24**: 1250. https://doi.org/10.1186/s12879-024-10142-4

**9** Wu X, Gao R, Shen X *et al*. Use of whole-genome sequencing to predict *Mycobacterium tuberculosis* drug resistance in Shanghai, China. *Int J Infect Dis* 2020; **96**: 48–53. https://doi.org/10.1016/j.ijid.2020.04.039

**10** Chen S, Zhou Y, Chen Y *et al*. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018; **34**: i884–90. https://doi.org/10.1093/bioinformatics/bty560

**11** The CRyPTIC Consortium. A data compendium associating the genomes of 12,289 *Mycobacterium tuberculosis* isolates with quantitative resistance phenotypes to 13 antibiotics. *PLoS Biol* 2022; **20**: e3001721. https://doi.org/10.1371/journal.pbio.3001721

**12** Pei S, Song Z, Yang W *et al*. The catalogue of *Mycobacterium tuberculosis* mutations associated with drug resistance to 12 drugs in China from a nationwide survey: a genomic analysis. *Lancet Microbe* 2024; **5**: 100899. https://doi.org/10.1016/S2666-5247(24)00131-9

**13** Sanchini A, Lanni A, Giannoni F *et al*. Exploring diagnostic methods for drug-resistant tuberculosis: a comprehensive overview. *Tuberculosis* 2024; **148**: 102522. https://doi.org/10.1016/j.tube.2024.102522

**14** Sharma D, Lata M, Singh R *et al*. Cytosolic proteome profiling of aminoglycosides resistant *Mycobacterium tuberculosis* clinical isolates using MALDI-TOF/MS. *Front Microbiol* 2016; **7**: 1816. https://doi.org/10.3389/fmicb.2016.01816

**15** Verboven L, Phelan J, Heupink TH *et al*. TBProfiler for automated calling of the association with drug resistance of variants in *Mycobacterium tuberculosis*. *PLoS ONE* 2022; **17**: e0279644. https://doi.org/10.1371/journal.pone.0279644

**16** Harrison LB, Kapur V, Behr MA. An imputed ancestral reference genome for the *Mycobacterium tuberculosis* complex better captures structural genomic diversity for reference-based alignment workflows. *Microb Genom* 2024; **10**: 001165. https://doi.org/10.1099/mgen.0.001165

**17** Liu D, Huang F, Zhang G *et al*. Whole-genome sequencing for surveillance of tuberculosis drug resistance and determination of resistance level in China. *Clin Microbiol Infect* 2022; **28**: 731.e9–e15. https://doi.org/10.1016/j.cmi.2021.09.014

**18** Manson AL, Cohen KA, Abeel T *et al*. Genomic analysis of globally diverse *Mycobacterium tuberculosis* strains provides insights into the emergence and spread of multidrug resistance. *Nat Genet* 2017; **49**: 395–402. https://doi.org/10.1038/ng.3767

**19** Phelan JE, O'Sullivan DM, Machado D *et al*. Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs. *Genome Med* 2019; **11**: 41. https://doi.org/10.1186/s13073-019-0650-x

**20** Liu B, Zheng D, Zhou S *et al*. VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Res* 2022; **50**(D1): D912–7. https://doi.org/10.1093/nar/gkab1107

**21** Alcock BP, Raphenya AR, Lau TTY *et al*. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database.

*Nucleic Acids Res* 2020; **48**(D1): D517–25. https://doi.org/10.1093/nar/gkz935

**22** Nayfach S, Ondov BD, Belbasi M *et al*. Skani enables accurate and efficient genome comparison for modern metagenomic datasets. *Nat Methods* 2023; **20**: 1633–4. https://doi.org/10.1038/s41592-023-02019-2

**23** Schoonjans F, Zalata A, Depuydt CE *et al*. MedCalc: a new computer program for medical statistics. *Comput Methods Programs Biomed* 1995; **48**: 257–62. https://doi.org/10.1016/0169-2607(95)01703-8

**24** Walker TM, Ip CL, Harrell RH *et al*. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect Dis* 2013; **13**: 137–46. https://doi.org/10.1016/S1473-3099(12)70277-3

**25** Parsaei M, Spotin A, Matini M *et al*. Prevalence of toxoplasmosis in patients infected with tuberculosis; a sero-molecular case-control study in northwest Iran. *Comp Immunol Microbiol Infect Dis* 2022; **81**: 101720. https://doi.org/10.1016/j.cimid.2021.101720

**26** Reingewertz TH, Meyer T, McIntosh F *et al*. Differential sensitivity of mycobacteria to isoniazid is related to differences in *KatG*-mediated enzymatic activation of the drug. *Antimicrob Agents Chemother* 2020; **64**: e01899-19. https://doi.org/10.1128/AAC.01899-19

**27** Andriani RAL, Ahmad Z. The effect of *KatG* S315t *M. tuberculosis* gene mutation on conversion rate of MDR-TB patients with shorter-course treatment. *Biosci Med J Biomed Transl Res* 2020; **4**: 1–7. https://doi.org/10.32539/bsm.v4i4.158

**28** Cao X, Zhan Q, Guo Y *et al*. Genotypic characterization of *katG*, *inhA*, and *ahpC* in isoniazid-resistant *Mycobacterium tuberculosis* clinical isolates in Shanghai, China. *Jundishapur J Microbiol* 2020; **12**: e95713. https://doi.org/10.5812/jjm.95713

**29** Li MC, Lu J, Lu Y *et al*. *rpoB* mutations and effects on rifampin resistance in *Mycobacterium tuberculosis*. *Infect Drug Resist* 2021; **14**: 4119–28. https://doi.org/10.2147/IDR.S333433

**30** Ma P, Luo T, Ge L *et al*. Compensatory effects of *M. tuberculosis rpoB* mutations outside the rifampicin resistance-determining region. *Emerg Microbes Infect* 2021; **10**: 743–52. https://doi.org/10.1080/22221751.2021.1908096

**31** McAlister AJ, Driscoll J, Metchock B. DNA sequencing for confirmation of rifampin resistance detected by Cepheid Xpert MTB/RIF assay. *J Clin Microbiol* 2015; **53**: 1752–3. https://doi.org/10.1128/JCM.03433-14

**32** Xia H, Song Y, Zheng Y *et al*. Detection of *Mycobacterium tuberculosis* rifampicin resistance conferred by borderline *rpoB* mutations: Xpert MTB/RIF is superior to phenotypic drug susceptibility testing. *Infect Drug Resist* 2022; **15**: 1345–52. https://doi.org/10.2147/IDR.S358301

**33** Li J, Gao X, Luo T *et al*. Association of *gyrA/B* mutations and resistance levels to fluoroquinolones in clinical isolates of *Mycobacterium tuberculosis*. *Emerg Microbes Infect* 2014; **3**: e19. https://doi.org/10.1038/emi.2014.21

**34** Alifano P, Palumbo C, Pasanisi D *et al*. Rifampicin-resistance, *rpoB* polymorphism and RNA polymerase genetic engineering. *J Biotechnol* 2015; **202**: 60–77. https://doi.org/10.1016/j.jbiotec.2014.11.024

**35** Zhang J, Xu J, Lei H *et al*. The development of variation-based rifampicin resistance in *Staphylococcus aureus* deciphered through genomic and transcriptomic study. *J Hazard Mater* 2023; **442**: 130112. https://doi.org/10.1016/j.jhazmat.2022.130112

**36** Watanabe S, Matsumura K, Iwai H *et al*. A mutation in the 16S rRNA decoding region attenuates the virulence of *Mycobacterium tuberculosis*. *Infect Immun* 2016; **84**: 2264–73. https://doi.org/10.1128/IAI.00417-16

**37** Chu T, Weng X, Law COK *et al*. The ribosomal maturation factor P from *Mycobacterium smegmatis* facilitates the ribosomal biogenesis by binding to the small ribosomal protein S12. *J Biol Chem* 2019; **294**: 372–8. https://doi.org/10.1074/jbc.RA118.002298

**38** Wu X, Tan G, Yang J *et al*. Prediction of *Mycobacterium tuberculosis* drug resistance by nucleotide MALDI-TOF-MS. *Int J Infect Dis* 2022; **121**: 47–54. https://doi.org/10.1016/j.ijid.2022.04.061

**39** Cohen KA, Stott KE, Munsamy V *et al*. Evidence for expanding the role of streptomycin in the management of drug-resistant *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* 2020; **64**: e00860-20. https://doi.org/10.1128/AAC.00860-20

**40** Penn-Nicholson A, Georghiou SB, Ciobanu N *et al*. Detection of isoniazid, fluoroquinolone, ethionamide, amikacin, kanamycin, and capreomycin resistance by the Xpert MTB/XDR assay: a cross-sectional multicentre diagnostic accuracy study. *Lancet Infect Dis* 2022; **22**: 242–9. https://doi.org/10.1016/S1473-3099(21)00452-7

**41** Quan TP, Bawa Z, Foster D *et al*. Evaluation of whole-genome sequencing for mycobacterial species identification and drug susceptibility testing in a clinical setting: a large-scale prospective assessment of performance against line probe assays and phenotyping. *J Clin Microbiol* 2018; **56**: e01480-17. https://doi.org/10.1128/JCM.01480-17

**42** Srivastava S, Musuka S, Sherman C *et al*. Efflux-pump-derived multiple drug resistance to ethambutol monotherapy in *Mycobacterium tuberculosis* and the pharmacokinetics and pharmacodynamics of ethambutol. *J Infect Dis* 2010; **201**: 1225–31. https://doi.org/10.1086/651377