



## OPEN ACCESS

## EDITED BY

Michael Marceau,  
Université Lille Nord de France, France

## REVIEWED BY

Dora-Luz Flores,  
Universidad Autónoma de Baja California,  
Mexico  
Hung Nguyen Van,  
National Lung Hospital, Vietnam

## \*CORRESPONDENCE

Yang Che  
✉ 13805876046@163.com  
Yi Chen  
✉ penicillion@163.com

<sup>†</sup>These authors have contributed equally to this work

RECEIVED 05 June 2025

ACCEPTED 22 July 2025

PUBLISHED 07 August 2025

## CITATION

Fang Q, Li X, Lu Y, Gao J, Wu Y, Chen Y and Che Y (2025) Whole-genome sequencing and machine learning reveal key drivers of delayed sputum conversion in rifampicin-resistant tuberculosis.  
*Front. Cell. Infect. Microbiol.* 15:1641385.  
doi: 10.3389/fcimb.2025.1641385

## COPYRIGHT

© 2025 Fang, Li, Lu, Gao, Wu, Chen and Che. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Whole-genome sequencing and machine learning reveal key drivers of delayed sputum conversion in rifampicin-resistant tuberculosis

Qing Fang<sup>1†</sup>, Xiangchen Li<sup>2,3†</sup>, Yewei Lu<sup>3</sup>, Junshun Gao<sup>3</sup>, Yvette Wu<sup>4</sup>, Yi Chen<sup>5\*</sup> and Yang Che<sup>5\*</sup>

<sup>1</sup>Departments of Pulmonary Medicine, The First Affiliated Hospital of Ningbo University, Ningbo, Zhejiang, China, <sup>2</sup>Jiaxing Key Laboratory of Clinical Laboratory Diagnostics and Translational Research, Affiliated Hospital of Jiaxing University, Jiaxing, Zhejiang, China, <sup>3</sup>Cosmos Wisdom Mass Spectrometry Center of Zhejiang University Medical School, Hangzhou, Zhejiang, China, <sup>4</sup>Fountain Valley School, Colorado Springs, CO, United States, <sup>5</sup>Institute of Tuberculosis Prevention and Control, Ningbo Municipal Center for Disease Control and Prevention, Ningbo, Zhejiang, China

Rifampicin-resistant tuberculosis (RR-TB) remains a major global health challenge, with delayed sputum culture conversion (SCC) predicting poor treatment outcomes. This study integrated whole-genome sequencing (WGS) and machine learning to identify clinical and genomic determinants of SCC failure in 150 RR-TB patients (2019–2023). Phenotypic and genotypic analysis revealed high rates of isoniazid resistance (74.0%) and *rpoB* mutations (97.3%, predominantly Ser450Leu), with 90% of strains belonging to Lineage 2 (Beijing family). While 64.7% achieved 2-month SCC, 18.0% remained culture-positive at 6 months. Univariate analysis linked 2-month SCC failure to smear positivity, resistance to isoniazid, amikacin, capreomycin, and levofloxacin, and pre-XDR-TB status, though only smear positivity (aOR=2.41, P=0.008) and levofloxacin resistance (aOR=2.83, P=0.003) persisted as independent predictors in multivariable analysis. A Random Forest model achieved robust prediction of SCC failure (AUC: 0.86 ± 0.06 at 2 months; 0.76 ± 0.10 at 6 months), identifying levofloxacin resistance (feature importance: 6.37), *embB*\_p.Met306Ile (5.94), and smear positivity (5.12) as top 2-month predictors, while *katG*\_p.Ser315Thr (4.85) and *gyrA*\_p.Asp94Gly (3.43) dominated 6-month predictions. These findings underscore smear positivity, levofloxacin resistance, and specific resistance mutations as critical drivers of SCC failure, guiding targeted RR-TB treatment strategies.

## KEYWORDS

rifampicin-resistant tuberculosis, whole-genome sequencing, sputum culture conversion, machine learning, drug resistance mutations

## Introduction

Tuberculosis (TB) remains a major global health challenge, with an estimated 10 million new cases and 1.5 million deaths annually (Bagcchi, 2023). Rifampicin-resistant tuberculosis (RR-TB) is defined as TB resistant to at least rifampicin (RIF). The emergence of drug-resistant TB, particularly RR-TB, poses a significant threat to global TB control efforts. RR-TB, which includes multidrug-resistant TB (MDR-TB) and extensively drug-resistant TB (XDR-TB), is associated with higher mortality rates, prolonged treatment durations, and increased healthcare costs. In 2020, almost half a million individuals developed RR-TB, contributing to an estimated 6.9 million disability-adjusted life years (DALYs) (Menzies et al., 2023). The global burden of RR-TB underscores the urgent need for improved diagnostic and treatment strategies to address this growing public health crisis.

Sputum culture conversion (SCC), defined as the absence of *Mycobacterium tuberculosis* (MTB) growth in sputum culture, is a key indicator of treatment efficacy in TB management (Holtz et al., 2006). In RR-TB, SCC is often delayed due to drug resistance, highlighting the need to identify factors influencing conversion to optimize treatment outcomes. Delayed SCC is associated with treatment failure, relapse, and poor prognosis, underscoring its prognostic significance (Wenlu et al., 2024). Clinically, SCC is typically assessed at two critical time points: 2 and 6 months after treatment initiation. While 2-month SCC reflects early bacterial suppression, its predictive value for long-term outcomes is limited, especially in patients with comorbidities such as HIV (Kurbatova et al., 2015). In contrast, 6-month SCC serves as a reliable predictor of sustained bacterial clearance and treatment success, providing critical guidance for clinical decision-making and outcome prediction (Meyvisch et al., 2018).

Prior studies have explored demographic, clinical, and resistance-related factors influencing SCC in TB, yet research specific to RR-TB remains limited (Liu et al., 2018). The impact of genetic mutations on SCC outcomes in RR-TB is particularly underexplored. Emerging evidence suggests that specific MTB mutations, such as those in *inhA* and *katG*, are associated with increased second-line drug resistance and delayed SCC, highlighting the critical role of genetic factors in treatment response (Click et al., 2020). Whole-genome sequencing (WGS) provides a robust tool for comprehensive mutation detection and phylogenetic lineage classification, which may affect bacterial fitness and therapeutic outcomes (Meehan et al., 2019; He et al., 2020). Moreover, machine learning (ML) algorithms can effectively integrate high-dimensional clinical and genomic data to predict treatment outcomes with precision and prioritize key predictors, complementing traditional statistical approaches (Chafai et al., 2024). However, the combined application of WGS and machine learning to identify SCC determinants in RR-TB remains in its early stages, with the impact of specific resistance mutations on SCC timing still poorly understood (Kurbatova et al., 2015). Addressing these gaps is crucial for developing targeted interventions to enhance RR-TB treatment success.

This study aims to elucidate the clinical and genetic determinants of SCC in RR-TB using a WGS-based approach. We analyzed clinical data and isolates from 150 RR-TB patients diagnosed between January 2021 and September 2023, assessing the association of demographic, clinical, and microbiological characteristics with 2- and 6-month SCC outcomes. Phenotypic and genotypic resistance profiles were characterized through drug susceptibility testing and WGS, with a focus on identifying specific resistance mutations linked to SCC. By integrating clinical, genomic, and machine-learning-driven insights, this study seeks to uncover factors contributing to delayed SCC, inform precision treatment strategies, and ultimately advance RR-TB management and control.

## Materials and methods

### Study design and sample enrollment

The study was a retrospective study that included all the culture-positive patients diagnosed with RR-TB at local TB dispensaries in Ningbo, China from Jan 1, 2021, to Dec 31, 2023. Patients aged above 18 years with sputum culture-positive, pulmonary, RR-TB were assessed for eligibility, and those with consent to standardized RR-TB regimen were included. Patients were excluded if they were pregnant or infected with HIV, hepatitis B or C virus, or refused to participate. Records related to demographics, clinical and microbiology were retrieved from the national TB information management system.

### Main definitions

The response to treatment in this study was evaluated by: 2-month sputum culture conversion as a marker of early treatment response, 6-month culture conversion (previously reported to be predictive of treatment outcome). Sputum culture conversion was defined as two consecutive negative cultures of samples taken at least 30 days apart (Mirzayev et al., 2021). This definition was strictly maintained for all timepoint evaluations (2/6 months).

### Drug susceptibility test

Drug susceptibility tests of four first-line anti-TB drugs and five second-line drugs were conducted on solid media (Lowenstein-Jensen) based on WHO recommendations (Bagcchi, 2023). The drug concentrations are isoniazid (INH) 0.2 µg/ml, rifampicin (RIF) 40 µg/ml, ethambutol (EMB) 2 µg/ml, streptomycin (SM) 4 µg/ml, levofloxacin (LVX) 2 µg/ml, amikacin (AM) 30 µg/ml, capreomycin (CM) 40 µg/ml, prothionamide (PTO) 40 µg/ml, and para-aminosalicylic acid (PAS) 1 µg/ml. H37RV strains were used as a reference for quality control.

## WGS and bioinformatics analysis

MTB culture products were inactivated, and genomic DNA was isolated using a bacterial DNA extraction kit (QIAGEN Inc., Dusseldorf, Germany), according to the manufacturer's instructions. The isolated and purified DNA products were transported via a cold chain to a sequencing facility. The purified genomic DNA was quantified using a TBS-380 fluorometer (Turner BioSystems Inc., Sunnyvale, CA, USA) to ensure that the DNA met the quality requirements for library preparation, sequencing, and detection. At least 1 µg of genomic DNA per sample was used as the input material for DNA sample preparation. The DNA samples were treated and fragmented to a size of ~400 bp. Sequencing libraries were generated using the NEXTflex<sup>TM</sup> Rapid DNA-Seq Kit. The prepared library was multiplexed and loaded on Illumina NovaSeq 6000 PE150 system (San Diego, CA92122, USA). Sequencing was carried out using a 2×150 paired-end (PE) configuration. Raw sequencing data were processed using fastp (v0.20.1) to remove adapter sequences and filter out low-quality bases (Chen et al., 2018). High-quality sequence data were then input into Kraken v2 for species identification, and samples identified as other species or with an MTB proportion below 90% were rejected as contaminated samples (Wood et al., 2019). Finally, the sequencing data from the remaining samples were aligned to the H37Rv reference genome (NC\_000962.3) using BWA (v0.7.17) (Li, 2013). Samples with an average sequencing depth > 20× and average genome coverage >95% were selected for subsequent data analysis. The SAMtools/BCftools suite was used for calling fixed (frequency ≥90%) SNPs at loci where the alternate alleles were supported by at least five reads (including both forward and reverse reads) (Danecek et al., 2021).

## Resistance mutation analysis

Clean sequencing data were input into the local version of TB-Profiler (v6.5.0) with its strict implementation of the V2 catalog (database name: who) to identify the genotype of resistance-associated mutations and detect the resistance profile of 15 anti-TB drugs, including Amikacin (AM), Bedaquiline (BDQ), Capreomycin (CM), Clofazimine (CFZ), Delamanid (DLM), Ethambutol (EMB), Ethionamide (ETO), Isoniazid (INH), Kanamycin, Levofloxacin (LVX), Linezolid (LZD), Moxifloxacin (MPX), Pyrazinamide (PZA), Rifampicin (RIF), Streptomycin (SM). Mutations with a frequency of less than 10% were excluded. WGS-based drug susceptibility testing (DST) results were determined by assessing the presence or absence of resistance-associated mutations in a WHO-recommended database, which classified mutations into Tier 1 (those most likely to confer resistance) and Tier 2 (genes with a reasonable pretest probability of resistance) (Walker et al., 2022). Hetero-resistance was defined based on the frequency of resistant alleles in the sequence reads <99% in this study.

## Phylogeny construction

The fixed SNPs, excluding those in SNPs in PE/PPE areas, insertion elements, repetitive regions, and drug resistance-associated genes, were combined into a concatenated alignment (Luo et al., 2014). Maximum-likelihood (ML) phylogenetic trees were inferred from the concatenated alignment using IQ-Tree v2 (Nguyen et al., 2015). The best-scoring ML tree were rooted using *M. canettii* (RefSeq: NC\_015848.1) as the outgroup and visualized with the Interactive Tree of Life (iTOL) (Letunic and Bork, 2021).

## Association analysis

Statistical analyses were performed in R package gtsummary (Sjoberg et al., 2021). To compare multiple categorical variables, binary logistic regression was employed where appropriate, with results reported as odds ratios (ORs) and 95% confidence intervals (CIs). Variables showing a significant association ( $P < 0.05$ ) in univariate logistic regression were included in the multivariate analysis. Forward stepwise logistic regression was then performed to identify whether these statistically significant covariates were independently associated with SCC.

## Machine learning analysis

A Random Forest (RF) model was implemented in R v4.0 with a fixed seed of 27 to identify genetic mutations and clinical factors associated with delayed SCC at 2 and 6 months in RR-TB. RF was chosen for its ability to handle high-dimensional, nonlinear data, robustness to overfitting, and capacity to rank feature importance, making it well-suited for integrating clinical and genomic predictors compared to other models like logistic regression or support vector machines. Mutations in DR related genes and clinical metadata from 150 isolates were merged. Clinical and phenotypic DST features were encoded, and mutations imputed with 0. Mutations in ≥5% of samples were selected. Class imbalance was addressed using SMOTE (DMwR package) (Kumari et al., 2020). The RF model was trained using the caret package's train function (randomForest method, 100 trees) with an 80/20 train-test split and 5-fold cross-validation (ROC-AUC). Hyperparameter tuning was performed with tuneLength = 3, optimizing the number of variables sampled at each split (mtry). Feature importance in the RF models was assessed using mean decrease in accuracy (MDA), with standard deviations estimated from 100 bootstrap iterations to quantify variability. Test-set AUC and its standard deviation were estimated using 100 bootstrap iterations. Feature importance and ROC curves were visualized using ggplot2 (Villanueva and Chen, 2019).

## Results

### Study population and characteristics of RR-TB patients

Between January 2021 and September 2023, a total of 2,704 patients were diagnosed with sputum culture-positive tuberculosis, among whom 225 were identified as having RR-TB. From 177 initially enrolled patients, 48 were excluded: 11 lost to follow-up pre-treatment, 22 declined treatment, 2 died pre-enrollment, and 13 transferred. During WGS processing, 27 additional cases were excluded: 17 from culture failure, six due to contamination (<90% MTBC sequences), and four confirmed as NTM infections (three *M. intracellulare*, one *M. abscessus*) by Kraken2 analysis. Ultimately, 150 patients with successful WGS results were included in the final analysis (Figure 1). Among the 150 RR-TB patients, 97 patients (64.7%) achieved SCC after 2 months of treatment. By 6 months of treatment, the number of patients who successfully achieved SCC increased to 123 (82.0%).

As shown in Table 1, the mean age of patients was 45 years, and 117 (78.0%) were male. A total of 72 patients (48.0%) were smear-positive. Retreatment cases accounted for 44.0% ( $n = 66$ ). Underlying health conditions were reported in 62 patients (42.7%), with diabetes (53.2%, 33/62) and hypertension (35.5%, 22/62) being the most common. Additionally, 24 patients were smokers, and 21 reported alcohol consumption. Migrant individuals comprised 46.0% ( $n = 60$ ) of the cohort. Regarding

occupation, the largest groups were workers (32.7%,  $n = 49$ ), unemployed individuals (29.3%,  $n = 44$ ), and farmers (20.0%,  $n = 30$ ).

### Phenotypic and genotypic profiles of RR-TB isolates

Excluding RIF, phenotypic DST revealed that resistance rates for the remaining eight anti-tuberculosis drugs ranged from highest to lowest as follows (Table 1): isoniazid (INH), 74.0% (111/150); streptomycin (SM), 62.0% (93/150); levofloxacin (LVX), 36.0% (54/150); ethambutol (EMB), 33.3% (50/150); para-aminosalicylic acid (PAS), 16.7% (25/150); prothionamide (PTO), 15.3% (23/150); amikacin (AM), 8.7% (13/150); and capreomycin (CM), 6.7% (10/150).

A total of 133 mutations were identified by WGS across 19 antibiotic resistance genes related to RIF, INH, EMB, PZA, ETO, SM, PAS, aminoglycosides, and fluoroquinolones resistance (Supplementary Table S1). Genomic analysis revealed RIF-resistance-associated mutations in 98.7% (148/150) of isolates, predominantly in *rpoB* (146 strains), with rare *rpoC* mutations (2 strains). The most frequent *rpoB* mutation was Ser450Leu, accounting for 58.2% (85/146) of cases, followed by Leu452Pro (12.3%, 18/146), with other low-frequency mutations (e.g., Leu430Pro, Asp435Ala). Additionally, two strains carried *rpoC* mutations (Ile435Thr, Phe452Ser). Beyond RIF, mutations conferring resistance to other drugs were detected, including

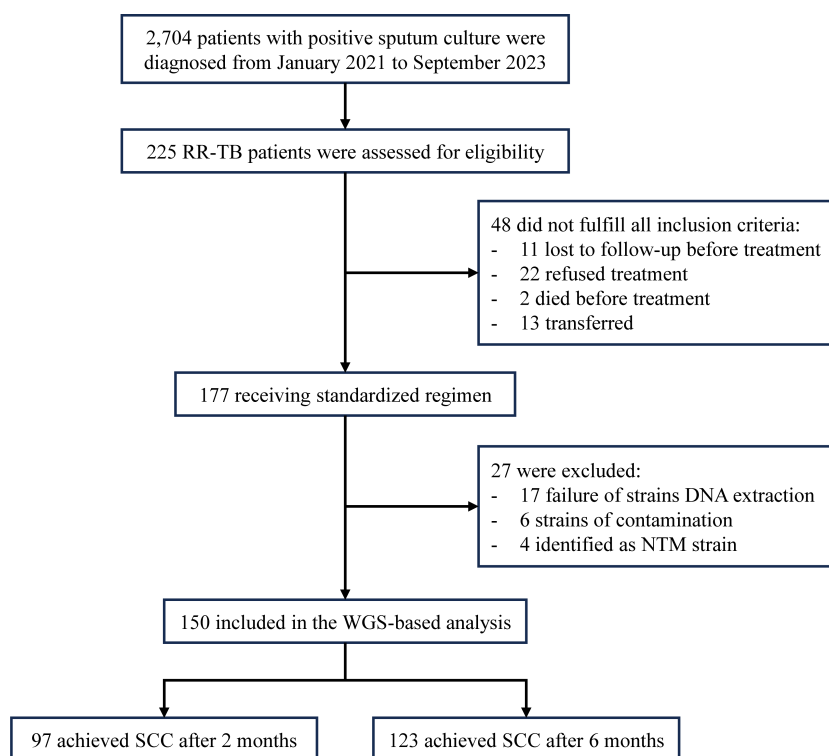


FIGURE 1  
Flowchart of this study.

TABLE 1 Demographic and clinical characteristics of patients with RR-TB in this study.

| Characteristics            | Number        |
|----------------------------|---------------|
| Age, mean ± SD             | 45 ± 16.1     |
| Male sex                   | 114 (76.0%)   |
| Smear-positive             | 72 (48.0%)    |
| Medical history            | 62 (41.3%)    |
| Smoking                    | 33 (22.0%)    |
| Drunk                      | 21 (14.0%)    |
| Retreatment                | 66 (44.0%)    |
| Migrant                    | 69 (46.0%)    |
| Occupation                 |               |
| Farmer                     | 30 (20.0%)    |
| Unemployed                 | 44 (29.0%)    |
| Worker                     | 50 (33.0%)    |
| Other                      | 26 (17.0%)    |
| Phenotypic drug resistance |               |
| INH                        | 111.0 (74.0%) |
| SM                         | 93.0 (62.0%)  |
| EMB                        | 50.0 (33.3%)  |
| AM                         | 13.0 (8.7%)   |
| CM                         | 10.0 (6.7%)   |
| LVX                        | 54.0 (36.0%)  |
| PAS                        | 25.0 (16.7%)  |
| PTO                        | 23.0 (15.3%)  |

*embB*\_p.Met306Ile (EMB), *katG*\_p.Ser315Thr (INH), *gyrA*\_p.Asp94Gly (fluoroquinolones), and *rrs*\_n.1401A>G (aminoglycosides), reflecting the multidrug-resistant nature of the cohort. Based on WGS results and WHO's updated definitions, the 150 RR-TB strains were further categorized into 65 MDR-TB, 51 pre-XDR-TB, and 34 RR-TB cases. These findings highlight the genetic diversity of resistance in RR-TB and underscore the utility of WGS in identifying critical mutations driving treatment challenges.

To elucidate the evolutionary relationships of the 150 RR-TB strains, we constructed a ML phylogenetic tree based on concatenated sequences from non-redundant SNP loci (Figure 2). Genotyping analysis revealed that 90.0% (135/150) of the strains were classified into lineage 2 (L2), with the majority falling under the L2.2.1 ( $n = 125$ ) and L2.2.2 ( $n = 9$ ) sublineages (Beijing family), accounting for a total of 134 strains. The remaining 10.0% (15/150) belonged to lineage 4 (L4), comprising sublineages L4.4 ( $n = 7$ ), L4.5 ( $n = 6$ ), and L4.2 ( $n = 2$ ), respectively.

## Demographic and clinical factors influencing the SCC in RR-TB

Univariate analysis of 150 RR-TB patients identified factors associated with the SCC failure at 2- and 6-month post-treatment initiation (Table 2). At 2-month, SCC failure showed significant associations (  $OR = 3.15$ , 95 %  $CI$ : 1.58 – 6.46,  $P = 0.001$ ) with smear-positive status, phenotypic resistance to (INH  $OR = 3.24$ , 95 %  $CI$ : 1.38 – 8.55,  $P = 0.011$ ), AM (  $OR = 7.29$ , 95 %  $CI$ : 2.11 – 33.7,  $P = 0.004$ ), CM (  $OR = 8.44$ , 95 %  $CI$ : 2.02 – 57.5,  $P = 0.009$ ), LVX (  $OR = 4.53$ , 95 %  $CI$ : 2.23 – 9.45,  $P < 0.001$ ), and genotypic pre-XDR-TB (  $OR = 4.70$ , 95 %  $CI$ : 1.80 – 13.5,  $P = 0.002$ ). However, by 6-month, only smear-positive status and LVX resistance remained significantly associated with delayed conversion (  $OR = 3.25$ , 95 %  $CI$ : 1.39 – 7.86,  $P = 0.007$ ), indicating that other resistance profiles became less influential over time.

To identify independent predictors of delayed SCC, we constructed multivariable logistic regression models. For 2-month conversion failure, the model was adjusted for smear-positive status, phenotypic resistance to INH, AM, CM and LVX, as well as genotypic pre-XDR-TB. For 6-month failure, adjustments included smear-positive status and LVX resistance. After adjustment, smear positivity (  $aOR = 3.32$ , 95 %  $CI$ : 1.54 – 7.44,  $P = 0.003$ ) and LVX resistance (  $aOR = 6.09$ , 95 %  $CI$ : 1.31 – 35.0,  $P = 0.026$ ) were significantly associated with 2-month SCC failure. Other factors in the 2-month model, including INH, AM, CM resistance, and genotypic pre-XDR-TB were not statistically significant. For 6-month SCC failure, smear positivity (  $aOR = 2.90$ , 95 %  $CI$ : 1.19 – 7.61,  $P = 0.023$ ) and LVX resistance (  $aOR = 3.01$ , 95 %  $CI$ : 1.27 – 7.40,  $P = 0.013$ ) remained significant predictors. The strong association of smear positivity and LVX resistance with SCC failure suggested that higher bacterial burden and fluoroquinolone resistance significantly impede early and sustained treatment response.

## ML analysis of delayed SCC

The RF model demonstrated robust performance in predicting SCC failure at both 2 and 6 months among patients with RR-TB, yielding ROC-AUC values of  $0.86 \pm 0.06$  and  $0.763 \pm 0.103$ , respectively, based on 5-fold cross-validation (Figures 3A, B). Feature importance analysis measured as MDA, highlighted key determinants of delayed SCC with variability estimated from 100 bootstrap iterations (Figures 3C, D). For the 2-month endpoint, the most influential predictors included LVX resistance (  $MDA = 6.37 \pm 1.05$ ), AM resistance (  $MDA = 5.95 \pm 0.46$ ), and the *embB*\_p.Met306Ile mutation (EMB resistance,  $MDA = 5.94 \pm 0.47$ ), followed by *rrs*\_n.1401A>G (AM resistance,  $MDA = 5.83 \pm$

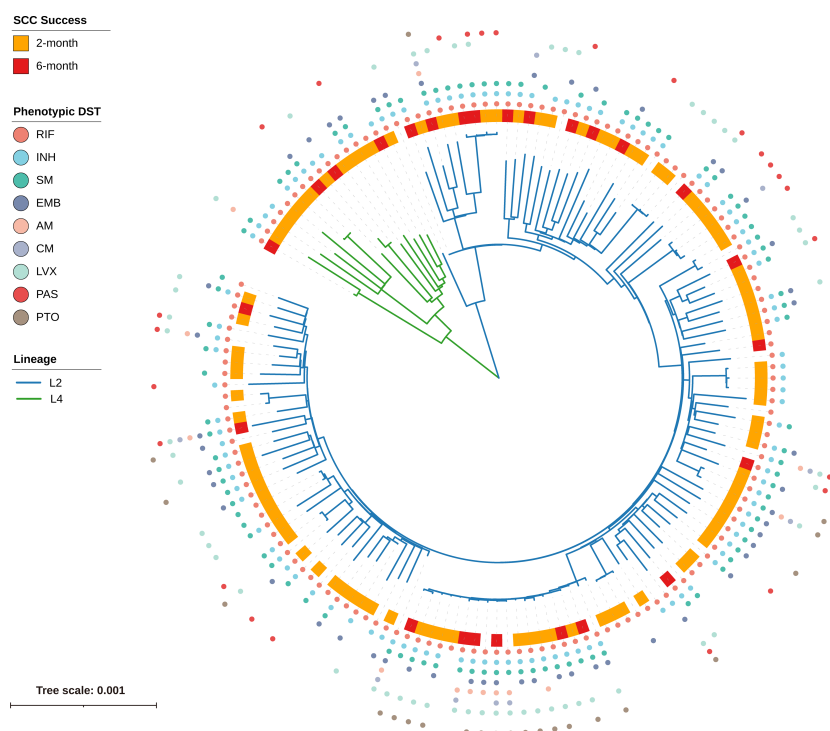


FIGURE 2

Phylogenetic tree and phenotypic DST profile of 150 RR-TB strains. The maximum-likelihood phylogenetic tree is based on concatenated SNP loci, with branch lengths representing the number of nucleotide substitutions per site. Branch colors represent lineages: blue for L2 and green for L4. The outer ring indicates sputum culture conversion (SCC) status: orange for SCC achieved at 2 months, red for SCC at 6 months, and white for no SCC by 6 months. The outermost colored dots denote phenotypic resistance to nine anti-TB drugs (RIF, INH, SM, EMB, AM, CM, LVX, PAS, PTO), with each drug assigned a unique color as indicated in the legend.

0.30) and INH resistance ( $MDA = 5.61 \pm 0.58$ ). At 6 months, the top contributors were LVX resistance ( $MDA = 5.46 \pm 1.33$ ), baseline sputum smear positivity ( $MDA = 5.00 \pm 1.50$ ), and the *katG*\_p.Ser315Thr mutation (INH resistance,  $MDA = 4.85 \pm 0.92$ ), with genotypic DR type ( $MDA = 4.16 \pm 0.96$ ) and *gyrA*\_p.Asp94Gly (fluoroquinolone resistance,  $MDA = 3.43 \pm 0.62$ ) also playing significant roles. Notably, *katG*\_p.Ser315Thr and *gyrA*\_p.Asp94Gly emerged as stronger predictors at the 6-month mark, underscoring their association with prolonged treatment failure due to INH and fluoroquinolone resistance, respectively.

## Discussion

This study provides a comprehensive analysis of clinical and genomic determinants of delayed SCC in RR-TB, combining WGS and ML to unravel critical predictors of treatment failure. Our findings underscore the persistent challenge of delayed SCC in RR-TB, with 35.3% of patients failing to achieve conversion at 2 months and 18.0% remaining culture-positive at 6 months. These rates align with prior reports of poor SCC outcomes in RR-TB cohorts, particularly in high-burden settings, but extend existing knowledge by dissecting the interplay of phenotypic resistance, genetic mutations, and bacterial lineage dynamics.

The association of smear positivity with delayed SCC at both time points reinforces its role as a biomarker of high bacterial burden and poor treatment response. This aligns with studies by Kurbatova et al, who demonstrated that smear grade correlates with prolonged culture positivity in MDR-TB (Kurbatova et al., 2015). However, our work uniquely identifies LVX resistance as a persistent independent predictor of SCC failure, even after adjusting for confounding resistance profiles. This finding is critical in light of global shifts toward shorter, fluoroquinolone-intensive regimens for RR-TB (Pranger et al., 2019). Our results corroborate earlier evidence that fluoroquinolone resistance undermines bactericidal activity, but further highlight that its impact persists beyond the early treatment phase, likely due to compensatory mutations in gyrase genes (e.g., *gyrA*\_p.Asp94Gly) that enhance fitness under drug pressure (Pantel et al., 2016; Diriba et al., 2022).

The diminishing significance of aminoglycoside resistance (AM/CM) in multivariable models at 6 months contrasts with its prominence at 2 months. This temporal divergence may reflect the delayed sterilizing effects of second-line injectables, which are often prioritized in early-phase regimens but phased out later (Conradie et al., 2022). Similarly, the loss of association between pre-XDR-TB and SCC failure in adjusted analyses suggests that resistance complexity alone may not drive outcomes if core drugs (e.g.,

TABLE 2 Risk factors for SCC failure at 2 and 6 months: univariable and multivariable regression analysis.

|                      | Univariable (2-month) |                 |                     |         | Multivariable (2-month) |                     |         | Univariable (6-month) |                     |         | Multivariable (2-month) |                     |         |
|----------------------|-----------------------|-----------------|---------------------|---------|-------------------------|---------------------|---------|-----------------------|---------------------|---------|-------------------------|---------------------|---------|
|                      | N                     | OR <sup>1</sup> | 95% CI <sup>1</sup> | p-value | AOR <sup>1</sup>        | 95% CI <sup>1</sup> | p-value | OR <sup>1</sup>       | 95% CI <sup>1</sup> | p-value | AOR <sup>1</sup>        | 95% CI <sup>1</sup> | p-value |
| Sex                  | 150                   |                 |                     |         |                         |                     |         |                       |                     |         |                         |                     |         |
| female               |                       | —               | —                   |         |                         |                     |         | —                     | —                   |         |                         |                     |         |
| male                 |                       | 1.33            | 0.60, 3.06          | 0.5     |                         |                     |         | 2.93                  | 0.94, 12.9          | 0.10    |                         |                     |         |
| Age                  | 150                   |                 |                     |         |                         |                     |         |                       |                     |         |                         |                     |         |
| <60 y                |                       | —               | —                   |         |                         |                     |         | —                     | —                   |         |                         |                     |         |
| >60 y                |                       | 1.38            | 0.55, 3.34          | 0.5     |                         |                     |         | 2.18                  | 0.76, 5.80          | 0.13    |                         |                     |         |
| Sputum smear         | 150                   |                 |                     |         |                         |                     |         |                       |                     |         |                         |                     |         |
| Negative             |                       | —               | —                   |         | —                       | —                   |         | —                     | —                   |         | —                       | —                   |         |
| Positive             |                       | 3.15            | 1.58, 6.46          | 0.001   | 3.32                    | 1.54, 7.44          | 0.003   | 3.14                  | 1.31, 8.12          | 0.013   | 2.90                    | 1.19, 7.61          | 0.023   |
| Smoking              | 150                   |                 |                     |         |                         |                     |         |                       |                     |         |                         |                     |         |
| No                   |                       | —               | —                   |         |                         |                     |         | —                     | —                   |         |                         |                     |         |
| Yes                  |                       | 1.06            | 0.46, 2.34          | 0.9     |                         |                     |         | 2.06                  | 0.80, 5.08          | 0.12    |                         |                     |         |
| Drunk                | 150                   |                 |                     |         |                         |                     |         |                       |                     |         |                         |                     |         |
| No                   |                       | —               | —                   |         |                         |                     |         | —                     | —                   |         |                         |                     |         |
| Yes                  |                       | 1.15            | 0.43, 2.94          | 0.8     |                         |                     |         | 2.06                  | 0.67, 5.73          | 0.2     |                         |                     |         |
| Registered residence | 150                   |                 |                     |         |                         |                     |         |                       |                     |         |                         |                     |         |
| Local                |                       | —               | —                   |         |                         |                     |         | —                     | —                   |         |                         |                     |         |
| Migrant              |                       | 1.36            | 0.69, 2.67          | 0.4     |                         |                     |         | 0.93                  | 0.39, 2.14          | 0.9     |                         |                     |         |
| Occupation           | 150                   |                 |                     |         |                         |                     |         |                       |                     |         |                         |                     |         |
| unemployed           |                       | —               | —                   |         |                         |                     |         | —                     | —                   |         |                         |                     |         |
| worker               |                       | 0.56            | 0.23, 1.32          | 0.2     |                         |                     |         | 1.01                  | 0.33, 3.13          | >0.9    |                         |                     |         |
| farmer               |                       | 0.96            | 0.37, 2.48          | >0.9    |                         |                     |         | 2.64                  | 0.88, 8.32          | 0.086   |                         |                     |         |
| other                |                       | 0.76            | 0.27, 2.07          | 0.6     |                         |                     |         | 0.44                  | 0.06, 2.01          | 0.3     |                         |                     |         |
| TB history           | 150                   |                 |                     |         |                         |                     |         |                       |                     |         |                         |                     |         |
| No                   |                       | —               | —                   |         |                         |                     |         | —                     | —                   |         |                         |                     |         |
| Yes                  |                       | 1.96            | 1.00, 3.89          | 0.052   |                         |                     |         | 1.76                  | 0.76, 4.15          | 0.2     |                         |                     |         |
| Treatment history    | 150                   |                 |                     |         |                         |                     |         |                       |                     |         |                         |                     |         |
| Initial treatment    |                       | —               | —                   |         |                         |                     |         | —                     | —                   |         |                         |                     |         |

(Continued)

TABLE 2 Continued

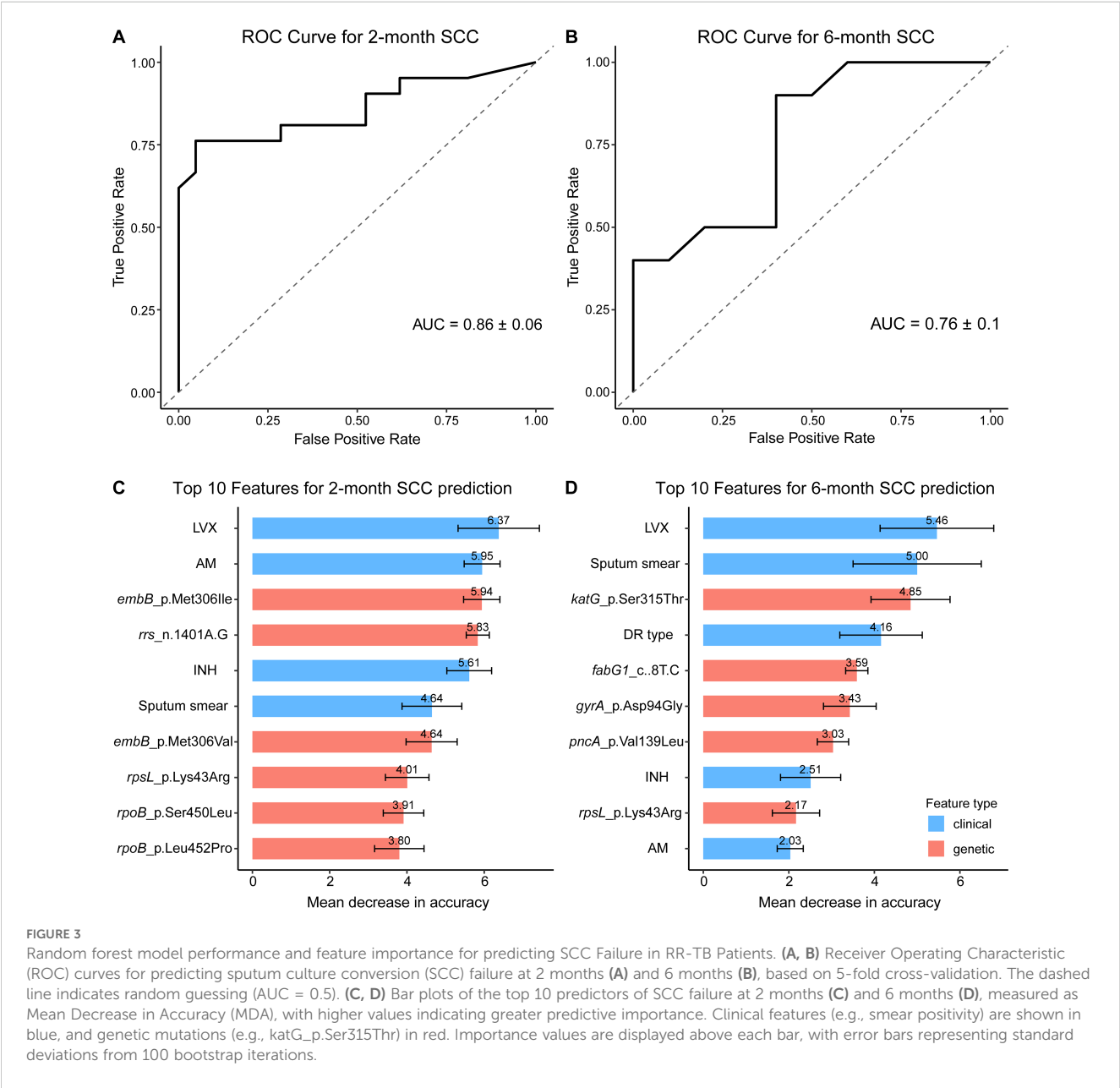
|                   | Univariable (2-month) |                 |                     |                  | Multivariable (2-month) |                     |              | Univariable (6-month) |                     |              | Multivariable (2-month) |                     |              |
|-------------------|-----------------------|-----------------|---------------------|------------------|-------------------------|---------------------|--------------|-----------------------|---------------------|--------------|-------------------------|---------------------|--------------|
|                   | N                     | OR <sup>1</sup> | 95% CI <sup>1</sup> | p-value          | AOR <sup>1</sup>        | 95% CI <sup>1</sup> | p-value      | OR <sup>1</sup>       | 95% CI <sup>1</sup> | p-value      | AOR <sup>1</sup>        | 95% CI <sup>1</sup> | p-value      |
| Treatment history | 150                   |                 |                     |                  |                         |                     |              |                       |                     |              |                         |                     |              |
| retreatment       |                       | 1.96            | 1.00, 3.89          | 0.052            |                         |                     |              | 1.76                  | 0.76, 4.15          | 0.2          |                         |                     |              |
| INH               | 150                   |                 |                     |                  |                         |                     |              |                       |                     |              |                         |                     |              |
| S                 |                       | —               | —                   |                  | —                       | —                   |              | —                     | —                   |              |                         |                     |              |
| R                 |                       | 3.24            | 1.38, 8.55          | <b>0.011</b>     | 4.02                    | 0.89, 23.4          | 0.088        | 1.68                  | 0.63, 5.33          | 0.3          |                         |                     |              |
| SM                | 150                   |                 |                     |                  |                         |                     |              |                       |                     |              |                         |                     |              |
| S                 |                       | —               | —                   |                  |                         |                     |              | —                     | —                   |              |                         |                     |              |
| R                 |                       | 1.49            | 0.74, 3.05          | 0.3              |                         |                     |              | 0.72                  | 0.31, 1.70          | 0.4          |                         |                     |              |
| EMB               | 150                   |                 |                     |                  |                         |                     |              |                       |                     |              |                         |                     |              |
| S                 |                       | —               | —                   |                  |                         |                     |              | —                     | —                   |              |                         |                     |              |
| R                 |                       | 1.75            | 0.87, 3.53          | 0.12             |                         |                     |              | 1.22                  | 0.50, 2.87          | 0.7          |                         |                     |              |
| AM                | 150                   |                 |                     |                  |                         |                     |              |                       |                     |              |                         |                     |              |
| S                 |                       | —               | —                   |                  | —                       | —                   |              | —                     | —                   |              |                         |                     |              |
| R                 |                       | 7.29            | 2.11, 33.7          | <b>0.004</b>     | 1.43                    | 0.25, 9.37          | 0.7          | 3.27                  | 0.92, 10.8          | 0.055        |                         |                     |              |
| CM                | 150                   |                 |                     |                  |                         |                     |              |                       |                     |              |                         |                     |              |
| S                 |                       | —               | —                   |                  | —                       | —                   |              | —                     | —                   |              |                         |                     |              |
| R                 |                       | 8.44            | 2.02, 57.5          | <b>0.009</b>     | 3.64                    | 0.51, 31.6          | 0.2          | 2.07                  | 0.42, 8.05          | 0.3          |                         |                     |              |
| LVX               | 150                   |                 |                     |                  |                         |                     |              |                       |                     |              |                         |                     |              |
| S                 |                       | —               | —                   |                  | —                       | —                   |              | —                     | —                   |              | —                       | —                   |              |
| R                 |                       | 4.53            | 2.23, 9.45          | <b>&lt;0.001</b> | 6.09                    | 1.31, 35.0          | <b>0.026</b> | 3.25                  | 1.39, 7.86          | <b>0.007</b> | 3.01                    | 1.27, 7.40          | <b>0.013</b> |
| PAS               | 150                   |                 |                     |                  |                         |                     |              |                       |                     |              |                         |                     |              |
| S                 |                       | —               | —                   |                  |                         |                     |              | —                     | —                   |              |                         |                     |              |
| R                 |                       | 2.30            | 0.96, 5.56          | 0.060            |                         |                     |              | 2.04                  | 0.72, 5.38          | 0.2          |                         |                     |              |
| PTO               | 150                   |                 |                     |                  |                         |                     |              |                       |                     |              |                         |                     |              |
| S                 |                       | —               | —                   |                  |                         |                     |              | —                     | —                   |              |                         |                     |              |
| R                 |                       | 1.86            | 0.75, 4.58          | 0.2              |                         |                     |              | 2.34                  | 0.81, 6.27          | 0.10         |                         |                     |              |
| Lineage           | 150                   |                 |                     |                  |                         |                     |              |                       |                     |              |                         |                     |              |
| L2                |                       | —               | —                   |                  |                         |                     |              | —                     | —                   |              |                         |                     |              |
| L4                |                       | 0.64            | 0.17, 1.98          | 0.5              |                         |                     |              | 0.00                  |                     | >0.9         |                         |                     |              |

(Continued)

TABLE 2 Continued

|                   | Univariable (2-month) |                 |                     |              | Multivariable (2-month) |                     |         | Univariable (6-month) |                     |         | Multivariable (2-month) |                     |         |
|-------------------|-----------------------|-----------------|---------------------|--------------|-------------------------|---------------------|---------|-----------------------|---------------------|---------|-------------------------|---------------------|---------|
|                   | N                     | OR <sup>1</sup> | 95% CI <sup>1</sup> | p-value      | AOR <sup>1</sup>        | 95% CI <sup>1</sup> | p-value | OR <sup>1</sup>       | 95% CI <sup>1</sup> | p-value | AOR <sup>1</sup>        | 95% CI <sup>1</sup> | p-value |
| Genotypic DR type | 150                   |                 |                     |              |                         |                     |         |                       |                     |         |                         |                     |         |
| RR-TB             |                       | —               | —                   |              | —                       | —                   |         | —                     | —                   |         |                         |                     |         |
| MDR-TB            |                       | 1.48            | 0.56, 4.21          | 0.4          | 0.54                    | 0.09, 2.95          | 0.5     | 0.56                  | 0.17, 1.90          | 0.3     |                         |                     |         |
| Pre-XDR-TB        |                       | 4.70            | 1.80, 13.5          | <b>0.002</b> | 0.29                    | 0.03, 2.20          | 0.3     | 1.77                  | 0.62, 5.52          | 0.3     |                         |                     |         |

<sup>1</sup>OR, Odds Ratio; CI, Confidence Interval; AOR, Adjusted Odds Ratio.  
Significant results (P < 0.05) are shown in bold.



bedaquiline, linezolid) retain efficacy—a hypothesis supported by recent trials (Trevisi et al., 2023).

By integrating WGS with ML, this study advances beyond traditional regression approaches, capturing nonlinear interactions among predictors. The Random Forest model's high accuracy (ROC-AUC: 0.86 at 2 months) outperforms prior SCC prediction tools reliant on clinical variables alone (Kurbatova et al., 2015). Importantly, our ML framework prioritized mutations (e.g., *rrs\_n.1401A>G* for aminoglycoside resistance) that are rarely assessed in phenotypic DST but critically influence SCC. This supports the WHO's push for expanded genetic DST in RR-TB management (World Health Organization, 2021).

Notably, ML-driven feature importance analysis uncovered mutation-specific temporal effects. While *embB\_p.Met306Ile* dominated 2-month SCC failure—potentially by compromising EMB's role in early bacterial suppression—*katG\_p.Ser315Thr* and *gyrA\_p.Asp94Gly* emerged as stronger predictors at 6 months. This aligns with mechanistic studies showing that *katG* mutations confer enduring isoniazid resistance via catalase-peroxidase inactivation (Ando et al., 2010), while *gyrA* mutations stabilize DNA gyrase under prolonged fluoroquinolone exposure (Aldred et al., 2014). Such findings underscore the need for dynamic, mutation-adjusted treatment strategies.

Phylogenetic analysis revealed that 90.0% of the isolates belonged to the L2 Beijing family and 10.0% to L4, consistent with the epidemiological profile of RR-TB in Eastern China (Yang et al., 2017). While no significant lineage-specific SCC effects were detected in our analysis, their potential impact cannot be entirely ruled out (Liu et al., 2020). Previous studies suggest that Beijing strains may exhibit higher drug resistance acquisition due to increased fitness and compensatory mutations, such as those in *rpoC* and *katG*, which mitigate the fitness cost of rifampicin resistance (Nguyen et al., 2025). Additionally, our ML analysis prioritized specific mutations (e.g., *katG\_p.Ser315Thr*, *gyrA\_p.Asp94Gly*) over lineage itself, suggesting that mutation-driven resistance profiles may outweigh lineage-specific effects in determining SCC outcomes. Future studies with more balanced lineage representation are needed to quantify the direct impact of lineages on SCC and explore interactions with resistance mutations.

These findings have immediate clinical implications. First, rapid detection of LVX resistance through genotypic assays should guide regimen selection, avoiding empiric fluoroquinolone use in high-resistance settings (Pillay et al., 2022). Second, smear-positive patients require intensified monitoring, with early escalation to novel agents (e.g., pretomanid) if SCC is delayed (Dooley et al., 2023). Third, lineage-specific therapeutic approaches may be warranted, particularly for L2 strains harboring *katG/embB* mutations that may benefit from adjunctive therapies (Gupta et al., 2020).

Study limitations include potential selection bias from excluding 75 of 225 RR-TB cases (33%) due to objective reasons (48 cases, e.g., loss to follow-up, refusal, death, transfer) or

laboratory issues (27 cases, e.g., failed isolate recovery, contamination, nontuberculous mycobacteria), which may bias delayed SCC findings by altering the impact of predictors like smear positivity or resistance mutations. Due to incomplete records in the national TB information management system, socioeconomic factors (e.g., income, education, housing), treatment adherence, and prior anti-tuberculosis drug exposure were excluded, potentially biasing our models toward microbiological and genetic predictors by overestimating the impact of resistance mutations and smear positivity on SCC failure, particularly in retreatment or socially vulnerable patients (Nidoi et al., 2021). It should be noted that China's free treatment policy and adjustment for proxy variables (migrant status/occupation) likely mitigated this bias (Long et al., 2021). The predominance of L2 strains may limit generalizability to settings where other lineages dominate. Future studies should integrate standardized socioeconomic assessments and objective adherence monitoring.

In conclusion, our study demonstrates that delayed sputum culture conversion in RR-TB is driven by a complex interplay of clinical, microbiological, and genetic factors, with smear positivity and LVX resistance emerging as persistent predictors of poor treatment response. By integrating WGS with ML, we identified key resistance mutations (e.g., *katG\_p.Ser315Thr*, *gyrA\_p.Asp94Gly*) that exert time-dependent effects on SCC outcomes, providing novel insights into the dynamic nature of drug resistance in RR-TB. These findings underscore the critical need for rapid genotypic DST to guide personalized regimen selection, particularly in high-burden settings where fluoroquinolone resistance and Beijing lineage strains predominate. Our results support the incorporation of WGS-based resistance profiling into clinical decision-making to optimize RR-TB management, while highlighting the potential of machine learning to improve outcome prediction. Future studies should explore the functional mechanisms underlying these mutation-specific effects and validate our model in diverse epidemiological settings to advance precision medicine approaches for drug-resistant tuberculosis.

## Data availability statement

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

## Ethics statement

This study was approved by the Ethics Committee of the Ningbo Municipal Center for Disease Control and Prevention. All eligible participants who agreed to participate in the program and

signed an informed consent form were required to complete a questionnaire and provide sputum specimen for subsequent studies. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

## Author contributions

QF: Data curation, Methodology, Writing – original draft, Writing – review & editing. XL: Conceptualization, Formal Analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. YL: Investigation, Software, Writing – original draft. JG: Project administration, Visualization, Writing – review & editing. YW: Resources, Software, Writing – original draft. YiC: Supervision, Writing – review & editing. YaC: Conceptualization, Project administration, Supervision, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was supported by Zhejiang Provincial Natural Science Foundation of China under Grant No. LTGY23H190001, Ningbo Public Welfare Science and Technology Program Project (No.2024S040), Ningbo Top Medical and Health Research Program (No.2023020713). The funder had no role in study design, data collection, analysis, interpretation of data and writing the manuscript.

## References

- Aldred, K. J., Kerns, R. J., and Osheroff, N. (2014). Mechanism of quinolone action and resistance. *Biochemistry* 53, 1565–1574. doi: 10.1021/bi5000564
- Ando, H., Kondo, Y., Suetake, T., Toyota, E., Kato, S., Mori, T., et al. (2010). Identification of *katG* Mutations Associated with High-Level Isoniazid Resistance in *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* 54, 1793–1799. doi: 10.1128/AAC.01691-09
- Bagcchi, S. (2023). WHO's global tuberculosis report 2022. *Lancet Microbe* 4, e20. doi: 10.1016/S2666-5247(22)00359-7
- Chafai, N., Bonizzi, L., Botti, S., and Badaoui, B. (2024). Emerging applications of machine learning in genomic medicine and healthcare. *Crit. Rev. Clin. Lab. Sci.* 61, 140–163. doi: 10.1080/10408363.2023.2259466
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560
- Click, E. S., Kurbatova, E. V., Alexander, H., Dalton, T. L., Chen, M. P., Posey, J. E., et al. (2020). Isoniazid and rifampin-resistance mutations associated with resistance to second-line drugs and with sputum culture conversion. *J. Infect. Dis.* 221, 2072–2082. doi: 10.1093/infdis/jiaa042
- Conradie, F., Bagdasaryan, T. R., Borisov, S., Howell, P., Mikiashvili, L., Ngubane, N., et al. (2022). Bedaquiline–pretomanid–linezolid regimens for drug-resistant tuberculosis. *N Engl. J. Med.* 387, 810–823. doi: 10.1056/NEJMoa2119430
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *Gigascience* 10, giab008. doi: 10.1093/gigascience/giab008
- Diriba, G., Alemu, A., Tola, H. H., Yenew, B., Amare, M., Eshetu, K., et al. (2022). Pre-extensively drug-resistant tuberculosis among multidrug-resistant tuberculosis

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcimb.2025.1641385/full#supplementary-material>

### SUPPLEMENTARY TABLE 1

Frequency of drug resistance mutations in 150 RR-TB strains among successful and failed cases of 2-month SCC and 6-month SCC.

patients in Ethiopia: a laboratory-based surveillance study. *IJID regions* 5, 39–43. doi: 10.1016/j.ijregi.2022.08.012

Dooley, K. E., Hendricks, B., Gupta, N., Barnes, G., Narunsky, K., Whitelaw, C., et al. (2023). Assessing pretomanid for tuberculosis (APT), a randomized phase 2 trial of pretomanid-containing regimens for drug-sensitive tuberculosis: 12-week results. *Am. J. Respir. Crit. Care Med.* 207, 929–935. doi: 10.1164/rccm.202208-1475OC

Gupta, A., Sinha, P., Nema, V., Gupta, P. K., Chakraborty, P., Kulkarni, S., et al. (2020). Detection of Beijing strains of MDR *M. tuberculosis* and their association with drug resistance mutations in *katG*, *rpoB*, and *embB* genes. *BMC Infect. Dis.* 20, 752. doi: 10.1186/s12879-020-05479-5

He, G., Li, Y., Chen, X., Chen, J., and Zhang, W. (2020). Prediction of treatment outcomes for multidrug-resistant tuberculosis by whole-genome sequencing. *Int. J. Infect. Dis.* 96, 68–72. doi: 10.1016/j.ijid.2020.04.043

Holtz, T. H., Sternberg, M., Kammerer, S., Laserson, K. F., Riekstina, V., Zarovska, E., et al. (2006). Time to sputum culture conversion in multidrug-resistant tuberculosis: predictors and relationship to treatment outcome. *Ann. Intern. Med.* 144, 650–659. doi: 10.7326/0003-4819-144-9-200605020-00008

Kumari, C., Abulaish, M., and Subbarao, N. (2020). Using SMOTE to deal with class-imbalance problem in bioactivity data to predict mTOR inhibitors. *SN Comput. Sci.* 1, 150. doi: 10.1007/s42979-020-00156-5

Kurbatova, E. V., Cegielski, J. P., Lienhardt, C., Akksilp, R., Bayona, J., Becerra, M. C., et al. (2015). Sputum culture conversion as a prognostic marker for end-of-treatment outcome in patients with multidrug-resistant tuberculosis: a secondary analysis of data from two observational cohort studies. *Lancet Respir. Med.* 3, 201–209. doi: 10.1016/S2213-2600(15)00036-3

- Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*. doi: 10.48550/arXiv.1303.3997
- Liu, Q., Lu, P., Martinez, L., Yang, H., Lu, W., Ding, X., et al. (2018). Factors affecting time to sputum culture conversion and treatment outcome of patients with multidrug-resistant tuberculosis in China. *BMC Infect. Dis.* 18, 114. doi: 10.1186/s12879-018-3021-0
- Liu, Q., Wang, D., Martinez, L., Lu, P., Zhu, L., Lu, W., et al. (2020). Mycobacterium tuberculosis Beijing genotype strains and unfavourable treatment outcomes: a systematic review and meta-analysis. *Clin. Microbiol. Infection* 26, 180–188. doi: 10.1016/j.cmi.2019.07.016
- Long, Q., Guo, L., Jiang, W., Huan, S., and Tang, S. (2021). Ending tuberculosis in China: health system challenges. *Lancet Public Health* 6, e948–e953. doi: 10.1016/S2468-2667(21)00203-6
- Luo, T., Yang, C., Peng, Y., Lu, L., Sun, G., Wu, J., et al. (2014). Whole-genome sequencing to detect recent transmission of *Mycobacterium tuberculosis* in settings with a high burden of tuberculosis. *Tuberculosis* 94, 434–440. doi: 10.1016/j.tube.2014.04.005
- Meehan, C. J., Goig, G. A., Kohl, T. A., Verboven, L., Dippenaar, A., Ezewudo, M., et al. (2019). Whole genome sequencing of *Mycobacterium tuberculosis*: current standards and open issues. *Nat. Rev. Microbiol.* 17, 533–545. doi: 10.1038/s41579-019-0214-5
- Menzies, N. A., Allwood, B. W., Dean, A. S., Dodd, P. J., Houben, R. M. G. J., James, L. P., et al. (2023). Global burden of disease due to rifampicin-resistant tuberculosis: a mathematical modeling analysis. *Nat. Commun.* 14, 6182. doi: 10.1038/s41467-023-41937-9
- Meyvisch, P., Kambili, C., Andries, K., Lounis, N., Theeuwes, M., Dannemann, B., et al. (2018). Evaluation of six months sputum culture conversion as a surrogate endpoint in a multidrug resistant-tuberculosis trial. *PLoS One* 13, e0200539. doi: 10.1371/journal.pone.0200539
- Mirzayev, F., Viney, K., Linh, N. N., Gonzalez-Angulo, L., Gegia, M., Jaramillo, E., et al. (2021). World Health Organization recommendations on the treatment of drug-resistant tuberculosis 2020 update. *Eur. Respir. J.* 57, 2003300. doi: 10.1183/13993003.03300-2020
- Nguyen, Q. H., Nguyen, T. V. A., and Bañuls, A.-L. (2025). Multi-drug resistance and compensatory mutations in *Mycobacterium tuberculosis* in Vietnam. *Trop. Med. Int. Health* 30, 426–436. doi: 10.1111/tmi.14104
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Nidoi, J., Muttamba, W., Walusimbi, S., Imoko, J. F., Lochoro, P., Ichtho, J., et al. (2021). Impact of socio-economic factors on Tuberculosis treatment outcomes in north-eastern Uganda: a mixed methods study. *BMC Public Health* 21. doi: 10.1186/s12889-021-12056-1
- Pantel, A., Petrella, S., Veziris, N., Matrat, S., Bouige, A., Ferrand, H., et al. (2016). Description of compensatory *gyrA* mutations restoring fluoroquinolone susceptibility in *Mycobacterium tuberculosis*. *J. Antimicrobial Chemotherapy* 71, 2428–2431. doi: 10.1093/jac/dkw169
- Pillay, S., Steingart, K., Davies, G., Chaplin, M., De Vos, M., Schumacher, S., et al. (2022). Xpert MTB/XDR for detection of pulmonary tuberculosis and resistance to isoniazid, fluoroquinolones, ethionamide, and amikacin. *Cochrane Database Systematic Rev.* doi: 10.1002/14651858.CD014841.pub2
- Pranger, A. D., van der Werf, T. S., Kosterink, J. G. W., and Alffenaar, J. W. C. (2019). The role of fluoroquinolones in the treatment of tuberculosis in 2019. *Drugs* 79, 161–171. doi: 10.1007/s40265-018-1043-y
- Sjöberg, D., Whiting, K., Curry, M., Lavery, J. A., and Larmarange, J. (2021). Reproducible summary tables with the gtsummary package. *R J.* 13, 570. doi: 10.32614/RJ-2021-053
- Trevisi, L., Hernán, M. A., Mitnick, C. D., Khan, U., Seung, K. J., Rich, M. L., et al. (2023). Effectiveness of bedaquiline use beyond six months in patients with multidrug-resistant tuberculosis. *Am. J. Respir. Crit. Care Med.* 207, 1525–1532. doi: 10.1164/rccm.202211-2125OC
- Villanueva, R. A., and Chen, Z. J. (2019). Ggplot2: elegant graphics for data analysis (2nd ed.). *Measurement: Interdiscip. Res. Perspect.* 17, 160–167. doi: 10.1080/15366367.2019.1565254
- Walker, T. M., Miotto, P., Köser, C. U., Fowler, P. W., Knaggs, J., Iqbal, Z., et al. (2022). The 2021 WHO catalogue of *Mycobacterium tuberculosis* complex mutations associated with drug resistance: a genotypic analysis. *Lancet Microbe* 3, e265–e273. doi: 10.1016/S2666-5247(21)00301-3
- Wenlu, Y., Xia, Z., Chuntao, W., Qiaolin, Y., Xujue, X., Rong, Y., et al. (2024). Time to sputum culture conversion and its associated factors among drug-resistant tuberculosis patients: a systematic review and meta-analysis. *BMC Infect. Dis.* 24, 169. doi: 10.1186/s12879-024-09009-5
- Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20, 257. doi: 10.1186/s13059-019-1891-0
- World Health Organization. (2021). “WHO consolidated guidelines on tuberculosis,” in *Module 3: Diagnosis - rapid diagnostics for tuberculosis detection*. (Geneva: World Health Organization).
- Yang, C., Luo, T., Shen, X., Wu, J., Gan, M., Xu, P., et al. (2017). Transmission of multidrug-resistant *Mycobacterium tuberculosis* in Shanghai, China: a retrospective observational study using whole-genome sequencing and epidemiological investigation. *Lancet Infect. Dis.* 17, 275–284. doi: 10.1016/S1473-3099(16)30418-2