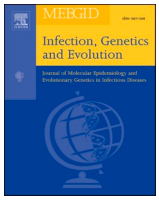
Infection, Genetics and Evolution 132 (2025) 105763

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid



# Characterisation of *M. tuberculosis* isolates obtained from Tamil Nadu prevalence survey by whole genome sequencing analysis

Priya Rajendran [a,1], Prathiksha Giridaran [a,b,1], Silla Varghese Thomas [a], Navinkumar Nagaraj [a], Kannan Thiruvengadam [a], Golla Radhika [a], Roja Samyuktha [a], Sriram Selvaraju [a], Asha Frederick [c], Chandrasekaran Padmapriyadarsini [a], Sivakumar Shanmugam [a,*]

[a] ICMR, National Institute for Research in Tuberculosis, India
[b] ICMR, National Institute of Epidemiology, India
[c] Directorate of Medical and Rural Health Services, Chennai, India

ARTICLE INFO

*Keywords:*
*M. tuberculosis*
Whole genome sequencing
Phenotypic DST
Genotypic DST
Mutations
Lineage distribution

ABSTRACT

Recent advances in whole genome sequencing have facilitated the understanding of drug resistance patterns and lineage distribution of *M. tuberculosis* worldwide. In this study, we aimed to determine the genetic diversity of MTB isolates from presumptive pulmonary TB patients obtained from a state prevalence survey. A total of 124 isolates were available for further characterisation, out of which 71 (57.2 %) and 47 (37.9 %) were subjected to sequencing and phenotypic DST, respectively. The phenotypic resistance profile revealed 3 isolates with multidrug resistance and 3 with mono-INH resistance. Out of 71 isolates, sequencing data were available for 61 (85.9 %), where the lineage distribution and drug resistance profile were analysed in comparison with phenotypic DST results. All the mutations were significant, accounting for one or the other resistance pattern. The concordance between pDST and gDST for the drugs was above 90 % except for ETH (77 %) and INH (87 %). The phylogenetic analysis of the lineage distribution revealed three clusters with MDR isolates belonging to lineage 1 and lineage 3. While lineage 2 is more frequently associated with MDR distribution both in India and worldwide, we did not find any lineage 2 MDR-TB isolates in our study. The use of WGS analysis improved our understanding of the genetic characteristics of MTB and its correlation with DR-TB transmission.

## 1. Introduction

Tuberculosis (TB) caused by *M. tuberculosis* (MTB) remains the leading cause of mortality worldwide, accounting for 1.26 million deaths and an estimated 10.8 million cases (WHO, 2024). Moreover, an increase in MTB drug resistance adds to the burden by delaying or hindering successful treatment. Multi-drug-resistant (MDR) TB, that is resistant to the two first-line drugs isoniazid and rifampicin, and extremely drug-resistant XDR (TB), a MDR TB that is resistant to a fluoroquinolone, and bedaquiline and/or linezolid, is becoming a major challenge in recent years. Early diagnosis and treatment of MDR TB and XDR TB are crucial for TB elimination, and this will demand an intensive network for screening, diagnosis, treatment, and follow-up (Dheda et al., 2024).

India is one of the countries with the highest TB burden, accounting for 25 % of total TB cases globallywith an estimated TB incidence of 2.77 million in 2022 (Mandal et al., 2023). In addition, India has seen an increase in people with MDR-TB over the years, with an estimate of around 410,000 persons in 2022. However, a significantly smaller number of patients (175,650) only received the diagnosis and began treatment (Vishwakarma et al., 2023). Prompt diagnosis, consistent epidemiological surveillance, and insights on the genetic diversity of MTB isolates in diverse settings are the major factors that could help in eliminating DR TB. Due to the limitation of its slow growth, culture-based phenotypic drug susceptibility test (pDST), despite being the gold standard, is slowly being replaced by molecular methods for DRTB surveillance (Meehan et al., 2019). With the advantage of providing exhaustive information on sequences and enabling drug resistance prediction, WGS serves as a promising tool in DRTB diagnosis (Papaventsis et al., 2017). The increased variability in the lineage distribution of MTB

* Corresponding author at: Department of Bacteriology, ICMR, National Institute for Research in Tuberculosis, Chetpet, Chennai 31, India.
*E-mail address:* shanmugam.sk@icmr.gov.in (S. Shanmugam).
[1] Both authors made equal contributions.

https://doi.org/10.1016/j.meegid.2025.105763
Received 3 February 2025; Received in revised form 9 May 2025; Accepted 9 May 2025
Available online 10 May 2025

and its association with drug resistance indicates that simultaneous detection of both characteristics is crucial for region or country-based DR-TB management (Oppong et al., 2019). In this study, we aimed to carry out molecular characterisation of MTB isolates obtained from a subnational prevalence survey conducted during 2022–2024 from several districts of Tamil Nadu that mostly had a pDST done for first- and second-line drugs.

## 2. Methods

The culture isolates subjected to characterisation in this study were from a state-wide community-based cross-sectional survey. The study details, including the design, results, and analysis of this survey, are published elsewhere (Prathiksha et al., 2024). In brief, individuals (≥15 years) in 180 clusters (comprising rural and urban clusters of Madurai, Chennai, Tiruvallur, Trichy, Sivagangai, Namakkal, Villupuram, Puducherry etc.,) with a sample size of 144,000 across Tamil Nadu, India, were screened for pulmonary symptoms and underwent chest X-ray (CXR). Participants who had symptoms suggestive of TB, history of TB, on TB treatment, and/or abnormal CXR were tested for MTB using cartridge-based nucleic acid amplification test (CBNAAT), smear microscopy (SM), and liquid culture (LC). From all patients, two consecutive samples were collected, with one sample for CBNAAT at the field and one sample for SM and LC at the reference lab. The patient who was positive for CBNAAT in the field setting was requested to provide a third sample for repeat CBNAAT testing, along with SM and LC at the reference lab. Results of these testing methods have been published earlier (Prathiksha et al., 2024). Among 244 microbiologically confirmed TB patients by either of the methods (SM, NAAT, culture), around 124 culture isolates representing one patient were further characterised to study the phenotypic and genotypic drug susceptibility pattern.

### 2.1. Phenotypic drug susceptibility testing

Mycobacterial Growth Indicator Tubes (MGIT) in the BD *BACTEC MGIT* 960 system (BD, Franklin Lakes, NJ, USA) were used to perform pDST for the following drugs based on the suggested critical concentration (CC) by the **WHO (2018)**. The chosen critical concentration was 0.1 μg/ml for isoniazid (INH); 0.25 and 1.0 μg/ml for moxifloxacin (MFX); 0.5 μg/ml for rifampicin (RIF) 1.0 μg/ml for streptomycin (STM), levofloxacin (LFX), linezolid (LZD), and amikacin (AMK); 2.5 μg/ml for kanamycin (KAN) and capreomycin (CPR); 4.0 μg/ml for P-amino salicylic acid (PAS); 5.0 μg/ml for ethambutol (EMB) and, ethionamide (ETH) and 100 μg/ml for pyrazinamide (PZA).

### 2.2. Whole genome sequencing

Using Qubit, dsDNA Assay Kits, and Nanodrop (Thermo Fisher Scientific, Waltham, MA, United States), the quality and quantity of the DNA were measured. Further, DNA libraries were set up through NexteraXT DNA Library Preparation and Index Kits (Illumina, San Diego, CA). After normalizing in equimolar concentrations, the Bioanalyzer 2100 System (Agilent Technologies, Santa Clara, CA, USA), which measured the average library size at 850 bp, was loaded for WGS (Miseq Reagent Kit v3; Illumina, San Diego, CA, USA). The Miseq sequencer (Illumina, San Diego, CA, USA) was used to perform 2 × 251 paired-end read sequencing cycles.

### 2.3. Bioinformatics methodology

#### 2.3.1. Data analysis

The sequencing runs produced 251 bp paired-end reads. The outcome data were in FastQ format for both forward and reverse sequences. These were mapped to the H37Rv reference genome using the NIRT CAMRespred Bioinformatics tool. Genome drug resistance was predicted for anti-TB drugs using this tool.

#### 2.3.2. Sequence-based resistance prediction

Reads that were at least 30 bp long (150 bp read length) and minimum base quality of 20 were filtered using Trimmomatic v0.36 (Bolger et al., 2014). Contamination with other species was checked using Kraken v1.0 (Wood and Salzberg, 2014). Reads were mapped to the H37Rv reference genome (NC_000916.3) using BWA v-0.7.12 with default parameters. Mapping of indels was corrected using Picard v2.2.4 and GATK v3.5. Variants were identified using SAMtools v1.3.1 with default parameters. Variants were filtered based on the following metrics: base quality >50, mapping quality >30, read depth > 5, and at least one read mapping in either direction. Variants supported by >80 % of the mapped reads were classified as homozygous sites, and those with <80 % mapped reads were classified as heterozygous sites. Variants were compared to a database of resistance-conferring variants generated by combining reports from previous studies. Lineages were predicted using both SNP-based variants and region of difference (RD) analysis using the tool RD-analyzer (Coll et al., 2014). Repeat phenotypic testing or sequencing was not performed in the event of discrepancies.

#### 2.3.3. Genomic analyses

A pseudo-genome was generated for each isolate by substituting the nucleotide base in the H37Rv reference genome sequence with the variants detected using a Python script. Repetitive regions were masked using bedtools v2.27.1. SNP-sites v2.5.1 were used to identify variable sites from the concatenated alignment of pseudo genomes (Page et al., 2016). The output generated was then used to identify pairwise SNP differences using SNP-dists v0.6.2. The phylogenetic tree was generated from SNP-sites output using RAxML with a GTR-GAMMA model and 1000 bootstrap replications.

## 3. Results

Among 124 MTB isolates, 71 (57.2 %) isolates were subjected to whole-genome Sequencing (WGS), and high-quality WGS data were available for 61 isolates. Results of pDST were available for 51 (41.1 %) isolates, out of which 49 (39.5 %) had genotypic data as well (Fig. 1). Genotypic DST (gDST) data alone were available for 12 isolates. Analysis of the lineage distribution of 61 isolates showed that Lineage 1 (L1 - East African Indian (EAI)) was the predominant (46/61 (75 %)), and sublineage 1.1.2 (37/46 (80.4 %) was common among them. Lineage 3 (L3 -Central Asian Strain (CAS) was the second dominant (6/61 (10 %)), followed by Lineage 4 (5/61 (8 %)) (L4 - Euro American (EA) and Lineage 2 (4/61(7 %)) (L2 -East Asian, subfamily Beijing).

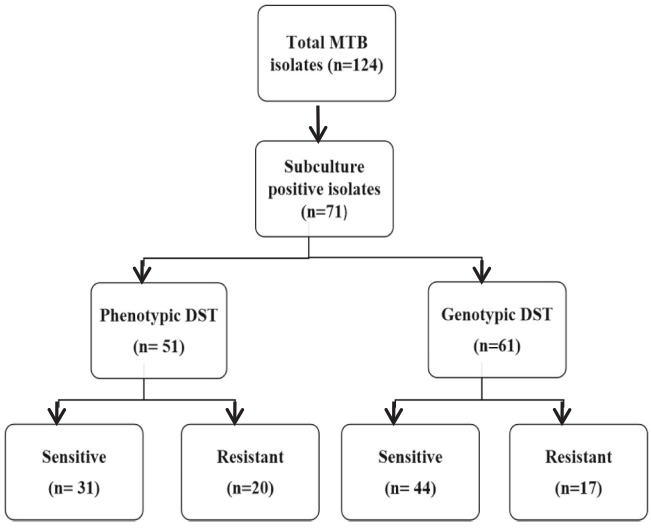



**Fig. 1.** Number of MTB isolates obtained in the survey and their resistance pattern.

### 3.1. Phenotypic DST (pDST)

The pDST revealed 31 (60 %) isolates as pan-susceptible, and 20 (39.2 %) as resistant to at least one drug. The resistance pattern revealed MonoINH resistance in 3 (5.8 %) and MDR in 3 isolates (5.8 %). Resistance to other drugs in different combinations included ETH in 12 (23.5 %), PZA in 7(13.7 %), STM in 5(9.8 %), LFX and OFX in 3 (5.8 %), and MOX and PAS in 2(3.9 %) isolates. The resistance pattern of all 49 isolates with pDST and gDST results is summarized in Table 1.

### 3.2. Genotypic DST (gDST)

Out of 61 isolates, 17 (28 %) were found to have significant mutations in drug-resistant genes, and 44 (72 %) had no mutations.

**Table 1**
*M. tuberculosis* phenotypic drug susceptibility testing results.

| ID | Sub-lineage | STM | INH | RIF | EMB | KAN | ETH | OF | AMK | CPR | PAS | LEV | MOX 0.25 | MOX 1 | PZA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TN 038 | 1.1.2 | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| TN 8380 | 1.1.2 | S | R | S | S | S(d) | S | S | S(d) | S(d) | R | R | R | S | S |
| TN 2266 | 1.1.2 | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| TN 005 | 1.1.3 | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| TN 2003 | 1.1.2 | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| TN 2586 | 3 | R | R | R | S(d) | S | S | S | S | S | S | S | S | S | S |
| TN 7437 | 1.1.2 | S | S | S | S | S | R(d) | R(d) | S | S | S | S | S | S | R(d) |
| TN 5686 | 1.1.3 | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| TN 704 | 1.1.2 | S | S | S | S | S | R(d) | S | S | S | S | S | S | S | S |
| TN 044 | 1.1.2 | S | S | S | S | S(d) | S | S | S(d) | S(d) | S | S | S | S | S |
| TN 374 | 1.1.2 | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| TN 3744 | 1.1.2 | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| TN 7644 | 1.1.2 | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| TN 3246 | 1.1.2 | S | R | S | S | S | R | S | S | S | S | S | S | S | S |
| TN 2331 | 1.1.2 | R | R | R | S(d) | S | R(d) | S | S | S | S | S | S | S | R(d) |
| TN 2138 | 1.1.2 | R(d) | S | S | S | S | S | S | S | S | S | S | S | S | S |
| TN 1409 | 1.2.2 | R | R | S(d) | S | S | R | S | S | S | S | S | S | S | S |
| TN 1978 | 1.1.2 | S | S | S | S | S | S | S | NA | | | | | | |
| TN 2072 | 1.1.2 | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| TN 1594 | 1.1.2 | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| TN 2292 | 1.1.2 | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| TN 7956 | 1.1.2 | S | R(d) | S | S | S | R(d) | S | S | S | S | R(d) | S | S | R(d) |
| TN 1909 | 1.1.2 | S(d) | S(d) | S | S | S | S(d) | S | S | S | S | S | S | S | S |
| TN 2939 | 1.1.2 | S | S(d) | S | S(d) (g3) | S | S | S | S | S | S | S | S | S | S |
| TN 5149 | 1.1.3 | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| TN 8341 | 4.4.1.2 | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| TN 3432 | 1.1.2 | NA | | S | NA | | | | | | | | | | |
| TN RR87 | 1.1.2 | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| TN 2002 | 4.1.2 | S | R | S(d) | S | S | S(d) | R | S | S | S | R | R | R | R |
| TN 2267 | 1.1.2 | S | R | S | S | S | R(d) | S | S | S | S | S | S | S | R |
| TN 2784 | 1.1.2 | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| TN 8268 | 3 | S | R(d) | S | S | S | S | S | S | S | S | S | S | S | S |
| TN 2176 | 1.2.2 | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| TN 944 | 1.1.2 | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| TN 8239 | 1.2.2 | NA | | S | NA | | | | | | | | | | |
| TN 1617 | 2.2.2 | S | R | S(d) | S | S | S | S | S | S | S | S | S | S | S |
| TN 3803 | 1.2.2 | S | S(d) | S | S | S | R(d) | S | S | S | S | S | S | S | R |
| TN 481 | 1.1.2 | S | S | S | S | S | R(d) | S | S | S | S | S | S | S | R(d) |
| TN 7438 | 1.1.2 | S | R(d) | R(d) | S | S | R(d) | S | S | S | S | S | S | S | S |
| TN 457 | 1.2.2 | S(d) (g3) | R | S | S | S | R | S | S | S | S | S | S | S | S |
| TN 3200 | 1.1.2 | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| TN 182 | 3.1.2.1 | S | R | S | S(d) | S | S | S | S | S | S | S | S | S | S |
| TN 317 | 2.2.1 | R | R | S | S | S | S | S | S | S | S(d) | S | S | S | S |
| TN 8258 | 4.8 | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| TN 1476 | 1.1.2 | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| TN 5140 | 1.1.2 | S | S | S | S | S | R(d) | R(d) | S | S | R(d) | S | R(d) | S | S |
| NPS 1403 | 1.1.2 | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| NPS 2366 | 1.1.2 | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| NPS2511 | 1.1.2 | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| | | STM | INH | RIF | EMB | KAN | ETH | OF | AMK | CPR | PAS | LEV | MOX (0.25) | MOX (1) | PZA |
| Phenotypic/genotypic match, n/tested (%) | | 46/49 (93.8 %) | 43/49 (87.7 %) | 45/49 (91.8 %) | 45/49 (91.8 %) | 47/49 (95.9 %) | 38/49 (77.5 %) | 47/49 (95.9 %) | 47/49 (95.9 %) | 47/49 (95.9 %) | 41/49 (83.6 %) | 48/49 (97.9 %) | 48/49 (97.9 %) | 49/49 (100 %) | 44/49 (89.7 %) |

Result description: R: Resistant; S: sensitive; NA: not available.
Discordance with WGS-based genotypic drug susceptibility results observed is indicated by letter (d).
When the evaluation of agreement between phenotypic/genotypic drug susceptibility testing results was based on grade 3 mutations (i.e., "uncertain significance") are indicated by letter (g3); the genetic DST result was assumed to be "resistant".
Abbreviations: NA, not available; AMK, amikacin; CPR, capreomycin; EMB, ethambutol; ETH, ethionamide; KAN, kanamycin; MOX, moxifloxacin; RIF, rrifampicin; INH, isoniazid STM, streptomycin; PZA, pyrazinamide; LEV, levofloxacin; OF, ofloxacin; PAS, Para-salicylic acid.

Genotypic resistance patterns revealed resistance to INH in 14 (23 %) isolates, RIF and EMB in seven (12 %) isolates each. In addition, 6(10 %) and 5(8 %) isolates showed resistance to STM and ETH, respectively. Resistance towards injectable drugs was observed in 2(7 %) isolates, and one isolate showed resistance to fluoroquinolones and PAS each (Fig. 2). When the frequency of mutations was analysed, higher numbers were observed in INH (*katG*, *inhA*, *fabG1*-promoter, *ahpc*-promoter, and *ndh)* 17 (27.8 %), followed by RIF (*rpoB* and *rpoC)* resistant isolates (11 (18.03 %)). The drug resistance-associated genetic variants (RAVs) identified in gDST, along with their pDST correlation, are summarized in Table 2.

### 3.3. Comparison of phenotypic and genotypic DST

When the phenotypic and genotypic DST results were compared for 13 drugs, discordance was observed only in a few isolates. Except for 77 % and 87 % concordance in ETH and INH, respectively, pDST of all other drugs had concordance above 90 % with gDST (Fig. 3).

### 3.4. Rifampicin, isoniazid, and ethionamide

The RAVs associated with INH and RIF resistance that were detected in genotypically resistant isolates had 87 % and 92 % concordance, respectively, with pDST. Eight RAVs were detected in *inhA*, *fabG1*, *ahpc*-promoter, *ndh,* and *katG* genes in seven genomic positions in 14 isolates, and among them, 11 isolates were resistant to INH by pDST. In addition to these 11 isolates, three others were resistant to pDST but sensitive to gDST. When the discrepancy of three isolates that were resistant by gDST and sensitive by pDST were analysed, mutations were seen in *fabG* (S/315/T), in ahpC (C/52/T), and *ndh* (R/268/H), respectively. RIF resistance analysis revealed five RAVs in four genomic positions in 7 isolates, whereas by pDST, only 2 were resistant, three were sensitive, and results were not available for 2 isolates. When the discrepant isolates were analysed, mutations were seen in position D/435/Y (rpoB) in two isolates and D/435/V (rpoB) in one isolate. One isolate that was resistant to pDST did not harbour any mutation in sequence analysis. Among 12 phenotypic ETH-resistant isolates, only 3 isolates had resistance by gDST with RAVs in the *fabG1* and *inhA gene* promoters. In addition to these three, two had RAVs in *ethR* and *inhA* each, accounting for a total of 5 ETH-resistant isolates by gDST, resulting in 77 % concordance between pDST and gDST.

### 3.5. Ethambutol, pyrazinamide, and streptomycin

A total of 7 isolates were resistant by gDST, where RAVs in the embB gene were detected in 6 isolates and *embC* gene in one isolate, with mutation in three genomic positions in the former and one in the latter. Out of these 7 isolates, 4 were sensitive to pDST, and the results of the other three isolates were not available. These 4 discrepant samples had mutations in G/406/S in two isolates, D/1024/N and M/306/I in one isolate each. The overall concordance rate was found to be 92 % between pDST and gDST. Seven samples were resistant to PZA by pDST, out of which only 2 were resistant by gDST. RAVs in the *pncA* gene accounting for PZA resistance were observed in these 2 isolates in 3 different positions (H/71/Q in one and P/54/S and D/49/G in another isolate). A total of 2 RAVs in 6 isolates associated with STM resistance were seen with mutation in *rpsL* (K/43/R) in five isolates and *gid* (L/79/S) in one isolate. Among these 6 gDST resistant isolates, 2 were sensitive, 3 were resistant to pDST, and the result of one isolate was not available.

### 3.6. Aminoglycosides and fluoroquinolones

All the isolates were sensitive to Aminoglycosides (KAN, CPR, and AMK) by pDST, with a concordance rate of 96 % between pDST and gDST. Two samples that were resistant to pDST also had a mutation in *rrs* at C/1402/A. In pDST of fluoroquinolones, resistance to OFX and LFX was seen in one isolate each. In three isolates, resistance with different combinations (OFX + MOX (0.25), LFX + MOX (0.25), OFX + MOX + LFX) was detected. The mutation associated with these resistance patterns was observed in the *gyrA* gene at D/94/G for one isolate, accounting for 96–100 % concordance.

### 3.7. Lineage-specific analysis

Mutations accounting for resistance and their distribution across different lineages were demonstrated through a Venn plot. A total of 16 mutations out of 17 were confined to L1, and one mutation was common among L2 and L3. Lineage 4 had 14 mutations, including 11 specific mutations and three mutations shared with L3(2) and L2(1). Similarly, L3 contained 6 mutations, with 2 specific mutations, two shared with L4, and one shared with L2 and L1 each. The least number of mutations (4) was observed in L2, with one specific mutation and other shared mutations.

Cluster analysis revealed a total of three transmission clusters, where

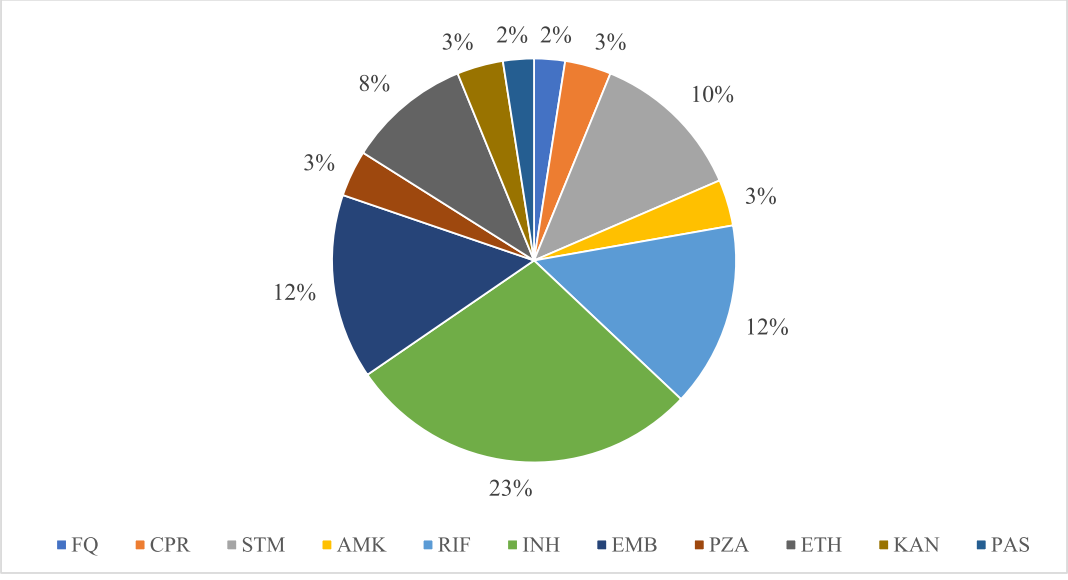



**Fig. 2.** Isolates showing resistance to drugs by sequencing.

**Table 2**
List of drug resistance mutations detected in *M. tuberculosis* isolates.

| ID | lineage | Drug | Genome position | Locus Tag | Gene | Mutation | Mutation grading | pDST resistant | Match between pDST/ gDST data |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | ETH, INH, STM | |
| | | ETH | 1,673,425 | | fabG1 | | | sensitive | No |
| | Sub lineage | INH | 1,673,425 | | fabG1 | C/15/T | Associated with | | |
| TN1909 | 1.1.2 | STM | 781,687 | | rpsL | K/43/R | resistance | | |
| | Sub lineage | AMK, KAN, | | | | C/1402/ | Associated with | | |
| TN 044 | 1.1.2 | CPR | 1,473,247 | | rrs | A | resistance | No | No |
| | Sub lineage | AMK, KAN, | | | | C/1402/ | Associated with | | |
| TN 8380 | 1.1.2 | CPR | 1,473,247 | | rrs | A | resistance | No | No |
| | Sub lineage | | | | | | Associated with | ETH and INH | |
| TN 3246 | 1.1.2 | ETH, INH | 1,673,425 | | fabG1 | C/−15/T | resistance | resistance | Yes (ETH, INH) |
| | | ETH, INH | 1,673,432 | | inhA | T/−8/C | | | |
| | | INH | 1,674,481 | | katG | S/94/A | | | |
| | Sub lineage | RIF | 761,110 | | rpoB | D/435/V | Associated with | INH Resistant | Yes (INH) |
| TN 1409 | 1.2.2 | INH | 2,102,240 | | ndh | R/268/H | resistance | RIF, ETH Sensitive | No (RIF and ETH) |
| | Sub lineage | PZA | 2,289,029 | | pncA | H/71/Q | Associated with | PZA Resistant | |
| TN 3803 | 1.2.2 | INH | 2,102,240 | | ndh | R/268/H | resistance | INH Sensitive | Yes (PZA) No (INH) |
| | | | | | | D/1024/ | | | |
| | Sub lineage | EMB | 4,249,583 | | embB | N | Associated with | | |
| TN_2939 | 1.1.2 | INH | 2,726,141 | | ahpc | C/−52/T | resistance | Sensitive | No |
| | | ETH, INH | 1,673,425 | | inhA | C/−15/T | | STM Sensitive | |
| | Sub lineage | INH | 2,102,240 | | ndh | R/268/H | Associated with | ETH, INH | No (STM) Yes (ETH, |
| TN 457 | 1.2.2 | STM | 4,407,967 | | gid | L/79/S | resistance | Resistant | INH) |
| | Sub lineage | RIF | 761,109 | | rpoB | D/435/Y | Associated with | RIF Sensitive | No (RIF) |
| TN 1617 | 2.2.2 | INH | 2,155,168 | | katG | S/315/T | resistance | INH Resistant | Yes (INH) |
| | | STM | 781,687 | | rpsL | K/43/R | | STM, INH | |
| | Sub lineage | INH | 2,155,168 | | katG | S/315/T | Associated with | Resistant | Yes (STM, INH) |
| TN317 | 2.2.1 | PAS | 2,747,471 | | folC | I/43/T | resistance | PAS Sensitive | No (PAS) |
| | Sub lineage | EMB | 4,247,431 | | embB | M/306/I | Associated with | EMB Sensitive | Yes (INH) |
| TN182 | 3.1.2.1 | INH | 2,155,168 | | katG | S/315/T | resistance | INH Resistant | No (EMB) |
| | | | | | | | Associated with | | |
| | | RIF | 761,155 | | rpoB | S/450/L | resistance | | |
| | | RIF | 764,703 | | rpoC | K/445/R | Uncertain significance | | |
| | | EMB | 4,247,729 | | embB | G/406/S | | | |
| | | STM | 7,816,687 | | rpsL | K/43/R | | | |
| TN 2324 | Sublineage 3 | INH | 2,155,168 | | katG | S/315/T | | NA | No |
| | | | | | | | Associated with | | |
| | | RIF | 761,155 | | rpoB | S/450/L | resistance | | |
| | | RIF | 764,703 | | rpoC | K/445/R | Uncertain significance | | |
| | | EMB | 4,247,729 | | embB | G/406/S | | RIF, STM, INH | |
| | | STM | 7,816,687 | | rpsL | K/43/R | | Resistant | Yes (RIF, STM, INH) No |
| TN 2331 | Sublineage 3 | INH | 2,155,168 | | katG | S/315/T | | EMB Sensitive | (EMB) |
| | | | | | | | Associated with | | |
| | | RIF | 761,155 | | rpoB | S/450/L | resistance | | |
| | | RIF | 764,703 | | rpoC | K/445/R | Uncertain significance | | |
| | | EMB | 4,247,729 | | embB | G/406/S | | RIF, STM, INH | |
| | | STM | 7,816,687 | | rpsL | K/43/R | | Resistant | Yes (RIF, STM, INH) No |
| TN 2586 | Sublineage 3 | INH | 2,155,168 | | katG | S/315/T | | EMB Sensitive | (EMB) |
| | | RIF | 761,109 | | rpoB | D/435/Y | | | |
| | | ETH | 4,327,876 | | ethR | F/110/L | | | |
| | | PZA | 2,289,082 | | pncA | P/54/S | | PZA, FLQ Resistant | |
| | Sub lineage | PZA | 2,289,096 | | pncA | D/49/G | Associated with | RIF and ETH | Yes (PZA, FLQ) |
| TN 2002 | 4.1.2 | FLQ | 7582 | | gyrA | D/94/G | resistance | Sensitive | No (ETH, RIF) |
| | | | | | | | Interim association with | | |
| | | INH | 1,673,432 | | inhA | T/−8/G | resistance | | |
| | Sub lineage | | | | | | Not associated with | | |
| TN 6252 | 4.3.3 | EMB | 4,242,182 | | embC | A/774/S | resistance | NA | No |
| | | | | | | | Associated with | | |
| | | RIF | 761,155 | | rpoB | S/450/L | resistance | | |
| | | RIF | 765,463 | | rpoC | N/698/K | Uncertain significance | | |
| | | | | | | | Associated with | | |
| | Sub lineage | EMB | 4,247,431 | | embB | M/306/I | resistance | | |
| 492 | 4.1.2.1 | INH | 2,155,214 | | katG | W/300/G | Uncertain significance | NA | No |

Mutation grading was assigned based on the WHO mutation catalog and the Institutional pipeline (NIRTCAMRespread).
For aminoglycosides and ethionamide/isoniazid mutations, pDST data are drug-specific.
Abbreviations: pDST, phenotypic drug susceptibility testing; gDST, genotypic drug susceptibility testing.
NA: not available AMK, amikacin; CPR, capreomycin; EMB, ethambutol; ETH, ethionamide; KAN, kanamycin; MOX, moxifloxacin; RIF, rrifampicin; INH, Isoniazid STM, streptomycin; PZA, pyrazinamide; LEV, levofloxacin; OF, ofloxacin; PAS, Para-salicylic acid

two clusters belonged to L1 and one cluster belonged to L3. In L1, one transmission cluster was observed between two districts, Tiruvallur and Vellore (TN 1594 and TN 944), and another within Tiruvallur district (TN 7437 and TN 7438 (MDR)). In L2, transmission was observed between 3 isolates having the same mutation pattern (TN 2586 (MDR), TN 2331(MDR), TN 2324), of which two were from Krishnagiri and one from Tiruvallur. A phylogenetic tree representing the transmission cluster and lineage distribution is depicted in Fig. 4, and the GIS

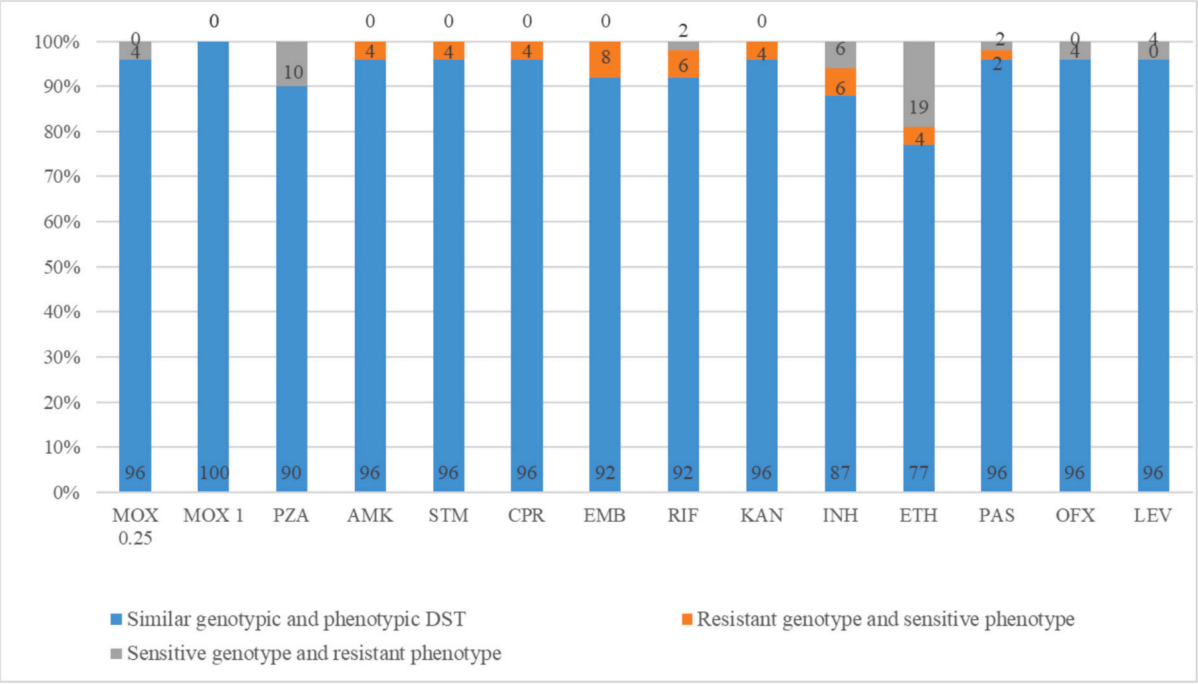



**Fig. 3.** Graph plot demonstrating the phenotypic and genotypic DST concordance.

mapping between these clusters is shown in Fig. 5.

### 3.8. Association between treatment outcome and resistance pattern

Among 61 isolates subjected to sequencing, we had gender details of 53 patients, of which 42 (79.2 %) were male and 11 (20.7 %) were female. Among these 61 patients, 21 (39.6 %) got cured of which 6 were resistant to either two or more drugs, 9 (16.9 %) completed treatment of which 3 were resistant either two or more drugs (Table 3), 2 (3.7 %) died due to TB, 19 (43.9 %) had no clinical data available and contact details of 2(3.7 %) patients were not available. When the association between resistance pattern and treatment outcome was analysed, no significance was observed ($p$ value = 0.956).

## 4. Discussion

In this study, we characterised the MTB isolates obtained from presumptive TB patients in different districts of Tamil Nadu, South India, to understand their lineage distribution and drug resistance pattern. When the lineage distribution was analysed, L1 (75 %) was most common, followed by L3 (10 %), and these findings are in alignment with previous Indian studies (Shanmugam et al., 2022; Manson et al., 2017a). However, while the Beijing lineage (L2) is most commonly associated with drug resistance in these studies, in our study, there were few L2 isolates, and L1 was predominantly associated with drug resistance (8/17). Studies from other parts of the country and the world also demonstrate that L2 is more commonly associated with drug resistance (Dixit et al., 2025; Lagutkin et al., 2022; Hakamata et al., 2020; Atavliyeva et al., 2024) and are considered to be the main carriers of MDR and XDR TB in Central Asia (Keikha and Majidzadeh, 2021). The study by Shanmugam et al, although published from the Tamil Nadu region, the isolates included in the study are from other parts of the country as well (Shanmugam et al., 2022). In contrast, the isolates included in our study are only from districts of Tamil Nadu, and hence, the distinct finding of L1's association with drug resistance needs further investigation and monitoring.

The agreement between pDST and gDST was found to be 92 % for RIF and 87 % for INH in our study, while 100 % concordance is generally seen in most of the studies published so far (Consortium et al., 2018; Liu et al., 2022; Kabahita et al., 2022). The INH resistance analysis revealed, S315T mutation in *katG,* which is more common and is in agreement with earlier findings from India (Shanmugam et al., 2022; Manson et al., 2017b; Chatterjee et al., 2017). The mutations seen in *fabG* (C/15/T), ahpC (C/52/T), and *ndh* (R/268/H) in three discrepant isolates (gDST resistant, pDST sensitive) are reported to be associated with low level or uncertain or interim significance, and hence the discrepancy is self-explanatory (Pei et al., 2024; *Indian Catalogue of Mycobacterium Tuberculosis Mutations and their Association with Drug Resistance, Version* 2, 2024). Moreover, these mutations have been reported to be associated with resistance only in rare instances (Vilcheze and Jacobs Jr., 2014; Liu et al., 2024). The mutation S/450/L at the *rpoB* gene was more common in rifampicin resistance, which is consistent with earlier findings (Shanmugam et al., 2022; Rao et al., 2023). However, this mutation was associated with L3($n$ = 3) and L4 ($n$ = 1) in our study, whereas in earlier studies it was found to be common in L2 (Shanmugam et al., 2022) and L1 (Rao et al., 2023). The mutations (D/435/Y and D/435/V) in discrepant (pDST sensitive, gDST resistant) isolates were found to be associated with resistance as per the WHO mutation catalog. However, the D/435/Y mutation was not found to be associated with phenotypic resistance in a study (Li et al., 2021) similar to our finding, where the isolate was sensitive to RIF at 0.5 and 1.0μg/ml. One pDST RIF-resistant isolate not harbouring any significant mutation in sequence analysis indicates the possibility of an unidentified mutation or technical error.

The overall concordance rate of 77 % between pDST and gDST was found for ETH in our study, where the RAVs (*fabG1*, *ethR*, and *inhA gene* promoter) were found only in five isolates, while 12 were resistant by pDST. This discrepant result could be explained by already existing dilemmas with ethionamide susceptibility testing of MTB, where the overlap of MIC distribution between resistant and susceptible isolates is still being researched (Ushtanit et al., 2022; Varma-Basil and Prasad, 2015). It's been reported that isolates with mutations in *fabG1* and *ethR* specific for ETH could still give pDST results as sensitive, and treatment is found to be effective in such isolates (Song et al., 2021). This explains the discrepancy observed in two isolates with RAVS *fabG1* and *eth* that were sensitive to pDST in our study. Similar discordant or moderate agreement results in other anti-TB drugs, PZA, STM, and EMB, have

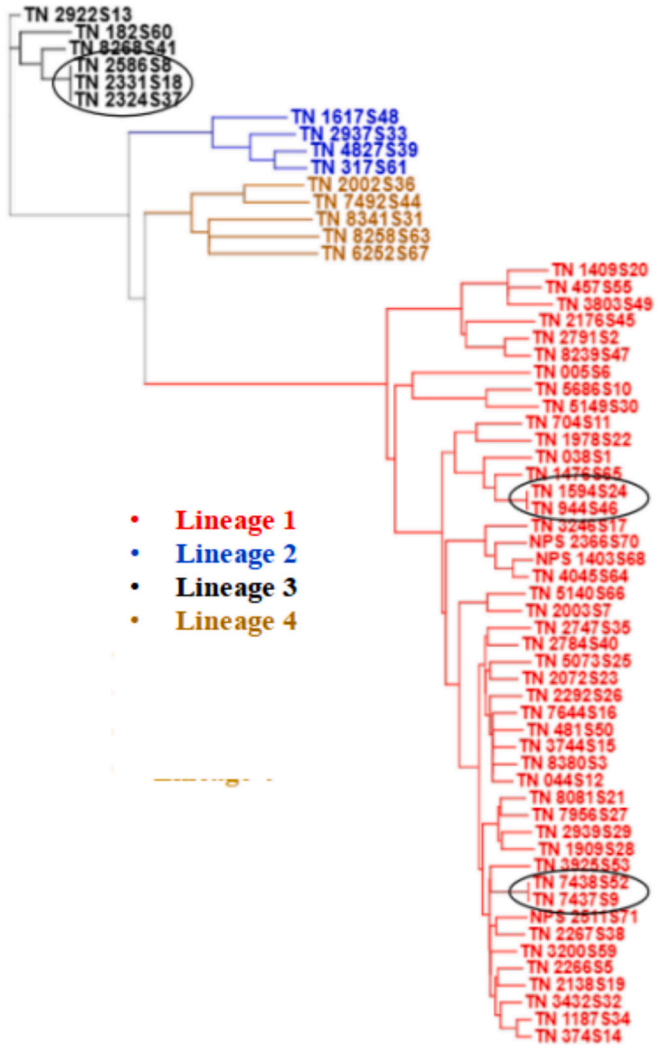



**Fig. 4.** Phylogenetic tree depicting lineage distribution and transmission clusters.

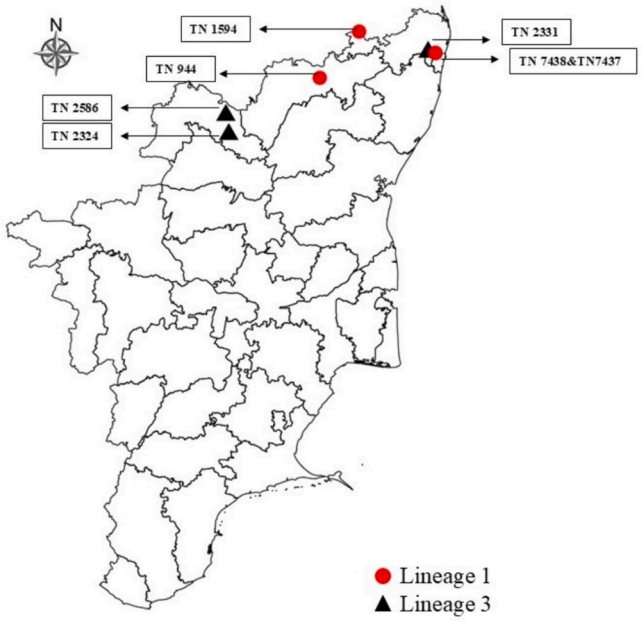



**Fig. 5.** GIS mapping of the isolates clustering together in lineage 1 and lineage 3.

**Table 3**
Association between treatment outcome and resistance pattern of MTB isolates.

| Treatment outcome | Lab Number | Resistance pattern | |
|---|---|---|---|
| | | Phenotypic | Genotypic |
| Cured | TN8380 | INH, PAS, LEV and MOX | KAN, AMK, CAP |
| | TN3246 | INH, ETH | INH, ETH |
| | TN1909 | No resistance detected | INH, ETH, and STM |
| | TN1409 | 1NH, ETH, and STM | INH, ETH, and RIF |
| | TN3803 | ETH | INH, PZA |
| | TN182 | INH | INH, EMB |
| Completed treatment | TN044 | No resistance detected | KAN, AMK, CPR |
| | TN7492 | Not available | INH, RIF, EMB |
| | TN2939 | No resistance detected | INH, EMB |

already been reported as common (Macedo et al., 2018; Korhonen et al., 2022) and this could be the reason for genotypic resistant isolates being identified as sensitive by phenotypic testing in our study. In the fluoroquinolone resistance pattern, we could see a mutation at *gyrA* for one isolate, while a total of six isolates were resistant to fluoroquinolone by pDST. This could be due to the lower sensitivity of the sequencing method reported for FQ, which ranged from 50 to 87.8 % in earlier studies (Chen et al., 2017; Moure et al., 2013; Sayadi et al., 2020).

The cluster analysis observed in two clusters of L1 between Tiruvallur and Vellore (TN 1594 and TN 944), within Tiruvallur district (TN 7437 and TN 7438 (MDR), and in another cluster of L3 between Krishnagiri and Tiruvallur is alarming. While the strains of L2 were reported to be associated with outbreaks of MDR-TB and characterised by increased transmissibility and hypervirulence (Shanmugam et al., 2022; Merker et al., 2018; Sethi et al., 2020), in our study, we could see that L3 was involved in MDR transmission (2/3). The transmission was observed between 3 isolates having the same mutation pattern, of which two were from Krishnagiri (TN 2586 (MDR), TN 2324) and one from Tiruvallur (TN 2331(MDR)). Interestingly, the distance between Tiruvallur and Krishnagiri, being 233 kms, MDR isolated from these two districts and clustering together is of utmost importance and needs intense monitoring. However, we could not get the travel history of these two patients to prove our hypothesis of transmission. A study from North India has reported an XDR isolate belonging to L3 (Sethi et al., 2020), and this indicates the need for monitoring L3 isolates in the future to prevent the emergence of XDR strains. One MDR isolate clustering with another susceptible isolate within Tiruvallur district, belonging to L1, is also not commonly reported and needs further research. The insignificant association between resistance pattern and treatment outcome could be explained due to the limited availability of the data. Moreover, there was already a poor correlation between the phenotypic and genotypic DST for some of the isolates. Hence, it would be more useful to analyze the association of treatment outcome with drug resistance mutation along with their phenotype data in a large population. Currently, we are conducting a drug resistance surveillance study throughout India that compares the phenotypic and genotypic resistance patterns of MTB isolates. Such a large-scale study with a varied geographic representation would open doors for unresolved mysteries in terms of transmission dynamics and the association of mutation with drug resistance. Moreover, a simultaneous MIC-based study for the antituberculosis drugs, including the newer drugs, is also being conducted. This will help in determining the critical concentration for resistance determination, and the discordance between phenotypic and genotypic DST could be reduced.

## 5. Limitations of the study

One of the key limitations of this study is the poor availability of comprehensive clinical data for many patients, which poses a significant challenge in establishing robust associations between drug resistance patterns and clinical outcomes. This data gap can lead to potential biases or misinterpretations, particularly when attempting to link resistance profiles to treatment success, failure, or relapse. Consequently, our findings should be interpreted with caution, and future studies with more extensive and standardized clinical datasets are warranted to validate and strengthen these associations.

## 6. Conclusion

In summary, the sequence analysis of the isolates from a prevalence survey conducted in Tamilnadu, facilitated detailed characterisation of the isolates in terms of drug resistance, lineage distribution, and their transmission dynamics and paved the way for further research. The concordance between pDST and gDST for the drugs was above 90 % except for ETH and INH, suggesting that WGS could serve as a promising approach for drug resistance prediction. Lineage 3 and lineage 1 replacing lineage 2 in MDR association in our study needs an extensive research of lineage distribution in Tamil Nadu, since earlier studies in India have reported lineage 2 more commonly associated with MDR.

## CRediT authorship contribution statement

**Priya Rajendran:** Writing – original draft, Formal analysis, Conceptualization. **Prathiksha Giridaran:** Writing – review & editing, Funding acquisition, Data curation. **Silla Varghese Thomas:** Formal analysis. **Navinkumar Nagaraj:** Formal analysis, Data curation. **Kannan Thiruvengadam:** Formal analysis, Data curation. **Golla Radhika:** Methodology, Investigation. **Roja Samyuktha:** Methodology, Investigation. **Sriram Selvaraju:** Supervision, Data curation. **Asha Frederick:** Supervision, Data curation. **Chandrasekaran Padmapriyadarsini:** Supervision, Project administration. **Sivakumar Shanmugam:** Writing – review & editing, Supervision, Project administration.

## Ethical statement

The survey was approved by the Institutional Ethics Committee of ICMR- National Institute for Research in Tuberculosis and all other participating institutes (Approval number: 334/NIRT-IEC/2018 dated 26th November 2018). All Participants above 18 years provided a written informed consent, and for participants from 15 years to 18 years, written informed assent and parents'/ legally authorized representatives' consent were obtained.

## Funding

The survey was funded by the Ministry of Health and Family Welfare, Government of India (F.No.5/8/5NTBP/Main File/ECD 1/2018, dt 18.12.2018, awarded to SS). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper titled "**Characterisation of *M. tuberculosis* isolates obtained from Tamil Nadu prevalence survey by whole genome sequencing analysis**".

## Data availability

The authors confirm that the data supporting the findings of this study titled "**Characterisation of *M. tuberculosis* isolates obtained from Tamil Nadu prevalence survey by whole genome sequencing analysis"** are available within the article and further information if required shall be provided.

The datasets presented in this study can be found in online repositories. The names of the repository and accession number(s) can be found below:

https://www.ncbi.nlm.nih.gov/, PRJNA1256943.

## References

Atavliyeva, S., Auganova, D., Tarlykov, P., 2024. Genetic diversity, evolution and drug resistance of *Mycobacterium tuberculosis* lineage 2. Front. Microbiol. 15, 1384791.
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30 (15), 2114–2120.
Chatterjee, A., et al., 2017. Whole genome sequencing of clinical strains of *Mycobacterium tuberculosis* from Mumbai, India: a potential tool for determining drug-resistance and strain lineage. Tuberculosis (Edinb.) 107, 63–72.
Chen, J., et al., 2017. Early detection of multidrug- and pre-extensively drug-resistant tuberculosis from smear-positive sputum by direct sequencing. BMC Infect. Dis. 17 (1), 300.
Coll, F., et al., 2014. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. Nat. Commun. 5, 4812.
Consortium, C.R., et al., 2018. Prediction of susceptibility to first-line tuberculosis drugs by DNA sequencing. N. Engl. J. Med. 379 (15), 1403–1415.
Dheda, K., et al., 2024. Multidrug-resistant tuberculosis. Nat. Rev. Dis. Primers 10 (1), 22.
Dixit, A., et al., 2025. Drug resistance and epidemiological success of modern *Mycobacterium tuberculosis* lineages in western India. J. Infect. Dis. 231 (1), 84–93.
Hakamata, M., et al., 2020. Higher genome mutation rates of Beijing lineage of *Mycobacterium tuberculosis* during human infection. Sci. Rep. 10 (1), 17997.
Indian Catalogue of *Mycobacterium Tuberculosis* Mutations and their Association with Drug Resistance, Version 2, 2024.
Kabahita, J.M., et al., 2022. First report of whole-genome analysis of an extensively drug-resistant *Mycobacterium tuberculosis* clinical isolate with bedaquiline, linezolid and clofazimine resistance from Uganda. Antimicrob. Resist. Infect. Control 11 (1), 68.
Keikha, M., Majidzadeh, M., 2021. Beijing genotype of *Mycobacterium tuberculosis* is associated with extensively drug-resistant tuberculosis: a global analysis. New Microbes. New Infect. 43, 100921.
Korhonen, V., et al., 2022. Multidrug-resistant tuberculosis in Finland: treatment outcome and the role of whole-genome sequencing. ERJ Open Res. 8 (4).
Lagutkin, D., et al., 2022. Genome-wide study of drug-resistant *Mycobacterium tuberculosis* and its intra-host evolution during treatment. Microorganisms 10 (7).
Li, M.C., et al., 2021. rpoB mutations and effects on rifampin resistance in *Mycobacterium tuberculosis*. Infect. Drug Resist. 14, 4119–4128.
Liu, D., et al., 2022. Whole-genome sequencing for surveillance of tuberculosis drug resistance and determination of resistance level in China. Clin. Microbiol. Infect. 28 (5), 731e9–731e15.
Liu, D., et al., 2024. Characterization of isoniazid resistance and genetic mutations in isoniazid-resistant and rifampicin-susceptible *Mycobacterium tuberculosis* in China. Infect. Med. (Beijing) 3 (3), 100129.
Macedo, R., et al., 2018. Dissecting whole-genome sequencing-based online tools for predicting resistance in *Mycobacterium tuberculosis*: can we use them for clinical decision guidance? Tuberculosis (Edinb.) 110, 44–51.
Mandal, S., Rao, R., Joshi, R., 2023. Estimating the burden of tuberculosis in India: a modelling study. Indian J. Community Med. 48 (3), 436–442.
Manson, A.L., et al., 2017a. Genomic analysis of globally diverse *Mycobacterium tuberculosis* strains provides insights into the emergence and spread of multidrug resistance. Nat. Genet. 49 (3), 395–402.
Manson, A.L., et al., 2017b. *Mycobacterium tuberculosis* whole genome sequences from southern India suggest novel resistance mechanisms and the need for region-specific diagnostics. Clin. Infect. Dis. 64 (11), 1494–1501.
Meehan, C.J., et al., 2019. Whole genome sequencing of *Mycobacterium tuberculosis*: current standards and open issues. Nat. Rev. Microbiol. 17 (9), 533–545.
Merker, M., et al., 2018. Compensatory evolution drives multidrug-resistant tuberculosis in Central Asia. Elife 7.
Moure, R., et al., 2013. Detection of streptomycin and quinolone resistance in *Mycobacterium tuberculosis* by a low-density DNA array. Tuberculosis (Edinb.) 93 (5), 508–514.
Oppong, Y.E.A., et al., 2019. Genome-wide analysis of *Mycobacterium tuberculosis* polymorphisms reveals lineage-specific associations with drug resistance. BMC Genomics 20 (1), 252.
Page, A.J., et al., 2016. SNP-sites: rapid, efficient extraction of SNPs from multi-FASTA alignments. Microb. Genom. 2 (4), e000056.
Papaventsis, D., et al., 2017. Whole genome sequencing of *Mycobacterium tuberculosis* for detection of drug resistance: a systematic review. Clin. Microbiol. Infect. 23 (2), 61–68.

Pei, S., et al., 2024. The catalogue of *Mycobacterium tuberculosis* mutations associated with drug resistance to 12 drugs in China from a nationwide survey: a genomic analysis. Lancet Microbe 5 (11), 100899.
Prathiksha, G., et al., 2024. Programmatic implications of a sub-national TB prevalence survey in India. Int. J. Tuberc. Lung Dis. 28 (7), 348–353.
Rao, M., et al., 2023. Lineage classification and antitubercular drug resistance surveillance of *Mycobacterium tuberculosis* by whole-genome sequencing in southern India. Microbiol. Spectr. 11 (5), e0453122.
Sayadi, M., et al., 2020. Genotypic and phenotypic characterization of *Mycobacterium tuberculosis* resistance against fluoroquinolones in the northeast of Iran. BMC Infect. Dis. 20 (1), 390.
Sethi, S., et al., 2020. Elucidation of drug resistance mutations in *Mycobacterium tuberculosis* isolates from North India by whole-genome sequencing. J. Glob. Antimicrob. Resist. 20, 11–15.
Shanmugam, S.K., et al., 2022. *Mycobacterium tuberculosis* lineages associated with mutations and drug resistance in isolates from India. Microbiol. Spectr. 10 (3), e0159421.
Song, Y., et al., 2021. The value of the inhA mutation detection in predicting Ethionamide resistance using melting curve technology. Infect. Drug Resist. 14, 329–334.
Ushtanit, A., et al., 2022. Molecular determinants of Ethionamide resistance in clinical isolates of *Mycobacterium tuberculosis*. Antibiotics (Basel) 11 (2).
Varma-Basil, M., Prasad, R., 2015. Dilemmas with ethionamide susceptibility testing of *Mycobacterium tuberculosis*: a microbiologist & physician's nightmare. Indian J. Med. Res. 142 (5), 512–514.
Vilcheze, C., Jacobs Jr., W.R., 2014. Resistance to isoniazid and Ethionamide in mycobacterium tuberculosis: genes, mutations, and causalities. Microbiol. Spectr. 2 (4). MGM2-0014-2013.
Vishwakarma, D., et al., 2023. Multi-drug resistance tuberculosis (MDR-TB) challenges in India: a review. Cureus 15 (12), e50222.
WHO, 2024. Global Tuberculosis Report. WHO, Geneva, Switzerland.
Wood, D.E., Salzberg, S.L., 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 15 (3), R46.