



# OPEN Genomic decoding of drug-resistant tuberculosis transmission in Thailand over three decades

Naphatcha Thawong<sup>1,2,7</sup>, Prapaporn Srilohasin<sup>3</sup>, Jody E. Phelan<sup>1</sup>, Worawich Phornsiricharoenphant<sup>4</sup>, Sissades Tongsim<sup>4</sup>, Prapat Suriyaphol<sup>3</sup>, Therdsak Prammananan<sup>4</sup>, Kiattichai Faksri<sup>5</sup>, Waritta Sawaengdee<sup>2</sup>, Linfeng Wang<sup>1</sup>, Woranich Hinthong<sup>1</sup>, Martin L. Hibberd<sup>1</sup>, Susana Campino<sup>1</sup>, Sukanya Wattanapokayakit<sup>2</sup>, Surakameth Mahasirimongkol<sup>2,7</sup>, Angkana Chaiprasert<sup>3,7</sup>✉ & Taane G. Clark<sup>1,6,7</sup>✉

Thailand has a high burden of tuberculosis, with control efforts hindered by drug-resistant *Mycobacterium tuberculosis* (Mtb). The increasing use of whole-genome sequencing (WGS) of Mtb offers valuable insights for clinical management and public health surveillance. WGS can be used to profile drug resistance, identify circulating sub-lineages, and trace transmission pathways or outbreaks. We analysed WGS data from 2,005 Mtb isolates collected across Thailand from 1994–2020, including 816 retrieved and 1,189 newly sequenced samples, with most isolates being multidrug-resistant (MDR-TB). Most isolates are lineage two strains (78.3%), primarily the Beijing sub-lineage (L2.2.1). Drug resistance profiling revealed substantial isoniazid and rifampicin resistance, and 67.3% classified as MDR-TB. Phenotypic and genotypic drug susceptibility testing showed high concordance (91.1%). Clustering analysis identified 206 transmission clades (maximum size 288), predominantly with MDR-TB, especially in Central and Northeastern regions. One cluster (n = 22) contains the *ddn* Gly81Ser mutation, linked to delamanid resistance, with some members pre-dating drug roll-out. In the largest cluster (n = 288), containing isolates spanning two decades, we applied transmission reconstruction methods to estimate a mutation rate of  $1.1 \times 10^{-7}$  substitutions per site per year. Overall, this study demonstrates the value of WGS in uncovering TB transmission and drug resistance, offering key data to inform better control strategies in Thailand and elsewhere.

**Keywords** *Mycobacterium tuberculosis*, Tuberculosis, Transmission, Drug resistance, Genomics, Thailand

Tuberculosis (TB), caused by the bacterium *Mycobacterium tuberculosis* (Mtb), remains a persistent and formidable global health challenge, causing 10.8 M cases and 1.1 M deaths in 2023 alone. Despite significant advancements in medical science and public health initiatives<sup>1</sup>, the WHO South-East Asia region (WHO SEA), which is home to around one-fourth of the world's population, has more than 45% burden of annual TB incidence. Thailand had amongst the highest rates of TB incidence (157/100 K population, ~113 K cases) in 2023, with an increase by 4.4% from 2022, and combined with prevalent drug resistance, will make World Health Organization (WHO) “End TB” targets difficult to achieve. Resistance levels to frontline rifampicin (RR-TB) and isoniazid (HR-TB) drugs, together called multi-drug resistance (MDR-TB), are high across WHO SEA (172 K cases; 8.4/100 K population). Worryingly, extensively drug-resistant TB strains (XDR; MDR + fluoroquinolones + group A resistance) and pre-XDR (MDR + fluoroquinolones) forms exist, providing a progression of resistance and limiting treatment options. In response, recent WHO guidelines suggest a 6-month regimen comprising bedaquiline, pretomanid, linezolid and moxifloxacin (BPaLM) to reduce both treatment duration and non-compliance levels due to drug toxicity. However, bedaquiline resistance is increasing, and with

<sup>1</sup>Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK. <sup>2</sup>Department of Medical Sciences, Ministry of Public Health, Nonthaburi, Thailand. <sup>3</sup>Faculty of Medicine, Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand. <sup>4</sup>Ministry of Higher Education, Science, Research and Innovation, National Science and Technology Development Agency, Bangkok, Thailand. <sup>5</sup>Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand. <sup>6</sup>Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK. <sup>7</sup>Naphatcha Thawong, Prapaporn Srilohasin, Surakameth Mahasirimongkol, Angkana Chaiprasert, Taane G. Clark contributed equally to this article. ✉email: Angkana.cha@mahidol.ac.th; taane.clark@lshtm.ac.uk

anti-TB drugs being costly with long durations (>6-months), personalised treatment based on Mtb resistance knowledge is crucial<sup>2</sup>.

Successful worldwide efforts to decrease TB burden have focused on advanced algorithms for early diagnosis, appropriate therapy choice and active case finding. However, these approaches have not been applied systematically across WHO SEA<sup>3</sup>. Whilst diagnostics endorsed for TB and drug resistance detection (e.g., Xpert MTB/RIF, XDR) as part of the WHO's "End TB" strategy are rapid compared to laboratory phenotypic drug susceptibility tests (pDSTs), they are costly and do not capture all genetic mutations required for precise management of advanced drug resistance forms. For example, the new Xpert XDR cartridge will not detect bedaquiline, linezolid, clofazimine and delamanid resistance, and does not cover some useful first-line drugs (e.g., rifabutin, ethambutol). Recent successes in TB treatment decision-making in developed countries have been led by advances in next generation sequencing technologies (NGS; e.g., Illumina, Oxford Nanopore Technologies (ONT))<sup>4</sup>, with increasing opportunities to use these directly from sputum or DNA from limited Mtb culture (MGIT), in near real time, at decreasing costs<sup>5</sup>. Whole genome (WGS) and targeted gene amplicon (AMP-SEQ) sequencing data generated using NGS can be used to profile Mtb for drug resistance and strain-types (sub-lineages), as well as infer transmission events or outbreaks through sequence similarity<sup>5–7</sup>, facilitated through advances in health informatics (e.g., TB-Profiler software)<sup>6</sup>.

In UK, the UK Health Security Agency (UKHSA) and some hospitals now use NGS-based characterisation as a clinical standard of care for TB management<sup>8</sup>. Thailand is seeking to adopt NGS as part of clinical care and surveillance, with government investment in genomics capacity. Recently, the WHO released the "genotype to phenotype" interpretation of NGS data<sup>7,9</sup>, but this approach needs to be adopted by National Tuberculosis Control Programmes (NTPs). NTP Guidelines developed in Thailand recommend using Mtb WGS to investigate cases involving clusters of TB patients or suspected outbreaks<sup>10</sup>, but to date the sample sizes have been small<sup>11–13</sup>. Challenges to wider adoption include limited funding, the need for sustained capacity strengthening in bioinformatics and laboratory infrastructure, and the complexity of data analysis and interpretation. Addressing these barriers is critical for integrating WGS into routine TB control efforts and realising its full potential for surveillance and outbreak investigation in the region. Genomics studies of Mtb from Thailand have revealed a dominance of lineage 2 strains<sup>11</sup>, and interactions with host genetics<sup>14</sup>. Our study analysed a large Mtb sample set from Thailand (n = 2,005; spanning 1994 to 2020), including 1,189 newly sequenced isolates, most isolates in the dataset were classified as MDR-TB, aims to identify mutations associated with drug resistance and uncover evidence of transmission. By establishing a baseline analysis of circulating strains, we seek to assist infection control teams embarking on using WGS to assist clinical and surveillance activities.

## Methods

### Sequence data

This study includes 2,005 *M. tuberculosis* whole-genome sequences obtained through convenience sampling in Thailand, predominantly from patients with drug-resistant TB, including MDR-TB and XDR-TB forms. Of these, 1,189 (59.3%) were recently gathered from Siriraj Hospital (years 2017–2020), a specialist TB hub for nationwide care, and 816 were retrieved from the European Nucleotide Archive (ENA) database (1994–2016)<sup>11,15</sup>. All sequence data were generated using the Illumina sequencing platform (see Table S1 for ENA accession numbers). Metadata, when available, included the year of collection, location, and phenotypic drug susceptibility tests (pDSTs) results. The pDSTs were conducted as part of routine TB laboratory processes using the standard agar proportion method on Lowenstein-Jensen medium. The number of isolates tested varied depending on the study, and the drugs included in this analysis were isoniazid (INH), rifampicin (RIF), ethambutol (EMB), streptomycin (STR), levofloxacin (LFX), moxifloxacin (MXF), amikacin (AMK), kanamycin (KAN), and linezolid (LZD). Pre-XDR and XDR have been defined under new WHO definitions<sup>16</sup>. For some isolates where XDR-TB could not be confirmed due to missing linezolid (LZD) and bedaquiline (BDQ) data, we classified them as pre-extensively drug-resistant-plus TB (Pre-XDR+ TB). The new study was approved by the Mahidol University ethics committee (SIRB 510/2561, MOPH EC 16/2562, Makarak EC30/2561).

### Bioinformatic and phylogenetic analysis

All raw sequence files were aligned to the H37Rv reference genome (accession number NC\_000962.3) using bwa-mem software. High-quality single nucleotide polymorphism (SNP) and insertion/deletion (indel) variants were called using samtools and GATK software and merged into a variant FASTA file using fastq2matrix. To ensure high-quality variant calls, we applied the following quality control criteria to all samples:  $\geq 90\%$  of reads mapped, median sequencing depth  $\geq 30$ -fold, and  $< 10\%$  of the genome with coverage below tenfold, based on summary statistics from the bamstat output. Additionally, variants were retained only if the alternate allele frequency was  $\geq 75\%$ , and *pe/ppe* gene regions were not excluded, but results from these regions were interpreted with caution due to known mapping challenges<sup>17</sup>. TB-Profiler software (v6.2.2) was used to determine isolate sub-lineages and genotypic drug resistance across 18 antibiotics<sup>6</sup>. The TB-Profiler drug resistance database incorporates mutations identified in the WHO catalogue as being associated with drug resistance in the *Mycobacterium tuberculosis* complex<sup>9</sup>. Phylogenetic trees were constructed from the multi-sample FASTA file, using the maximum likelihood method implemented in IQ-TREE software (v2.2.6) with a set of 75,060 genome-wide SNPs. The trees were subsequently visualised using iTOL (v6). The geographic distribution of isolates was mapped using QGIS software (v3.36.0-Maidenhead). For the transmission analyses, pairwise SNP distances were calculated using the snp-dists tool within the fastq2matrix pipeline. A Gaussian Mixture Model (GMM) was applied to the distribution of pairwise SNP distances between isolates to identify natural groupings in the data. This unsupervised clustering approach enabled the determination of an appropriate SNP distance cut-off for defining putative transmission clusters, accounting for the population structure and time span of our dataset. This method avoids the need for an arbitrary SNP threshold and allows for context-specific

identification of transmission links. Timed phylogenetic trees were constructed for the largest clusters using BEAST2 (v2.7.6)<sup>18</sup>, applying a strict molecular clock, a coalescent constant population model, and a General Time Reversible (GTR) substitution model. The Markov chain Monte Carlo (MCMC) analysis was run for 100 M iterations, sampling every 10,000<sup>th</sup> step. Temporal signals were assessed using TempEst, which showed a positive correlation between genetic divergence and sampling time and a moderate root-to-tip correlation ( $R^2 = 0.5241$ ). Analysis settings were based on approaches described previously<sup>19</sup>. A resulting maximum clade credibility (MCC) tree was annotated using Tree annotate software (v2.7.6). The Transphylo package (v1.4.10) was used to determine the direction of any transmission events using the default settings. The transmission graph was inferred and formatted using the tgv web tool<sup>20</sup>. The frequency of SNP and indel variants is compared to a global dataset ( $n = 50,723$ ) curated in the TB-Profiler database.

### Statistical analysis

Logistic regression models were employed to determine associations between transmissibility outcomes (whether an isolate was part of a transmission cluster), lineage and genotypic drug resistance profiles. Additionally, SNP loci associated with transmissibility were assessed using linear mixed models implemented in the Genome-wide Efficient Mixed Model Association (GEMMA) software (v0.98.5). This approach, related to genome-wide association study (GWAS) methods, accounted for the kinship matrix, incorporating SNPs, lineage, and drug resistance to control for relatedness among isolates<sup>21</sup>. Associations between geographical and SNPs distances of isolates were assessed using non-parametric Kruskal–Wallis tests and Spearman's correlations. Statistical analyses were performed in R software (4.4.1).

## Results

### Study population

A total of 2,005 Mtb samples from Thailand, collected between 1994 and 2020, were included in this study (Table 1). A high number of Mtb isolates (838/2005, 41.8%) were collected between 2013 and 2017, spanning 4 regions of Thailand (Central, Northeastern, Northern, Southern). Most samples originated from the Central region (1299/2005, 64.8%), with Kanchanaburi province contributing the largest number (558/2005, 27.8%). This dataset includes a high proportion of drug-resistant TB cases. Based on Thailand's national TB reports, 6,992 laboratory-confirmed MDR/RR-TB cases were reported between 2014 and 2020<sup>22</sup>. Our dataset includes 594 MDR/RR-TB isolates from the same period, covering approximately 8.5% of the national burden during those years. The majority of Mtb strain-types were lineage L2 (East-Asian; 1569/2005, 78.3%), followed by L1 (Indo-Oceanic, 324/2005, 16.2%), L4 (Euro-American, 110/2005, 5.5%) and L3 (East-African-Indian, 2/2005, 0.1%). The most prevalent sub-lineage, L2.2.1 (Beijing strain-type) accounted for most isolates (1317/2005, 65.7%). Across the 2,005 isolates, 75,060 high-confidence SNPs were identified, and the resulting phylogenetic tree revealed the expected grouping of isolates by (sub-)lineage (Fig. 1). A principal component analysis (PCA) using the SNPs supported the lineage-based clustering patterns in the phylogenetic tree (Figure S1).

### Genotypic drug resistance

Genotypic drug resistance profiling (Table 1) revealed that 88.4% of isolates (1772/2005) were resistant to at least one anti-TB drug, with the highest resistance observed for isoniazid (1686/2005, 84.1%), rifampicin (1693/2005, 84.4%), and streptomycin (1212/2005, 60.5%). WGS-based profiling indicated that 67.3% (1349/2005) of isolates were classified as MDR-TB, 14.6% (293/2005) as pre-XDR-TB, 11.6% (233/2005) as sensitive, and 0.2% (5/2005) as XDR-TB (Table 1). The geographic distribution of lineages and genotypic drug resistance categories across Thailand, suggested that MDR-TB was present across most regions (Fig. 2).

The most common mutations linked to drug resistance were *katG* Ser315Thr (71.6%, isoniazid), *rpsL* Lys43Arg (46.4%, streptomycin), *rpoB* Ser450Leu (46.0%, rifampicin), *ethA* c.639\_640delGT (26.6%, ethionamide), *embB* (Gly406Asp) (16.0%, ethambutol), *pncA* Ile31Thr (16.0%, pyrazinamide), and *embB* Met306Val (14.4%, ethambutol). All mutations have been observed in the TB-Profiler database, which includes global isolates ( $N = 50,723$ ). However, there were differences in allele frequencies. For example, the *pncA* Ile31Thr mutation, linked to pyrazinamide resistance, was over 288 times more frequent than in the global dataset. Similarly, the *ethA* (c.639\_640delGT) mutation, associated with ethionamide resistance, occurred at a frequency more than 140 times higher than in the global collection (Table 2). A detailed catalogue of identified mutations associated with drug resistance is presented (Table S2). The degree of resistance to drugs may be underestimated, especially as functional genotype–phenotype mutations are being found and may need to be included in updates to databases. For example, the *ddn* Gly81Ser mutation has recently been associated with delamanid resistance<sup>23</sup>. This mutation was identified in 22 isolates (1.1%), including 6 Mtb isolates from Kanchanaburi (2005–2017) in the ENA database<sup>11,15</sup>, and 16 isolates collected between 2007 and 2014 from Kanchanaburi, Bangkok, Ratchaburi, Suphan Buri, and Chachoengsao - all located in or around the central region of Thailand.

### Concordance with phenotypic DST

Phenotypic DST assays were performed on a subset of the samples ( $n = 1,826$ ) (Table S3). Isoniazid exhibited the highest resistance rate at 95.6% (1745/1826), followed by rifampicin at 94.7% (1730/1826), streptomycin at 61.8% (1113/1801), ethambutol at 44.0% (727/1653), and levofloxacin at 12.2% (199/1602). All other resistance rates were below 10%, with moxifloxacin at 8.6% (95/1101), kanamycin at 7.5% (127/1687), amikacin at 6.9% (81/1,172), and linezolid at 0.4% (4/1,097). Overall, 1,522 isolates (75.9%) were classified as MDR-TB, followed by 201 isolates (10.0%) with Pre-XDR+ TB, and 59 isolates (2.9%) being pan-sensitive. XDR-TB was identified in 3 isolates (0.2%) (Table S3).

The overall concordance rate between phenotypic DSTs and genotypic drug resistance was 91.1% (11,522/12,640 tests). Concordance exceeded 88% for all drugs except ethambutol (71.7%), with rates of 96.1%

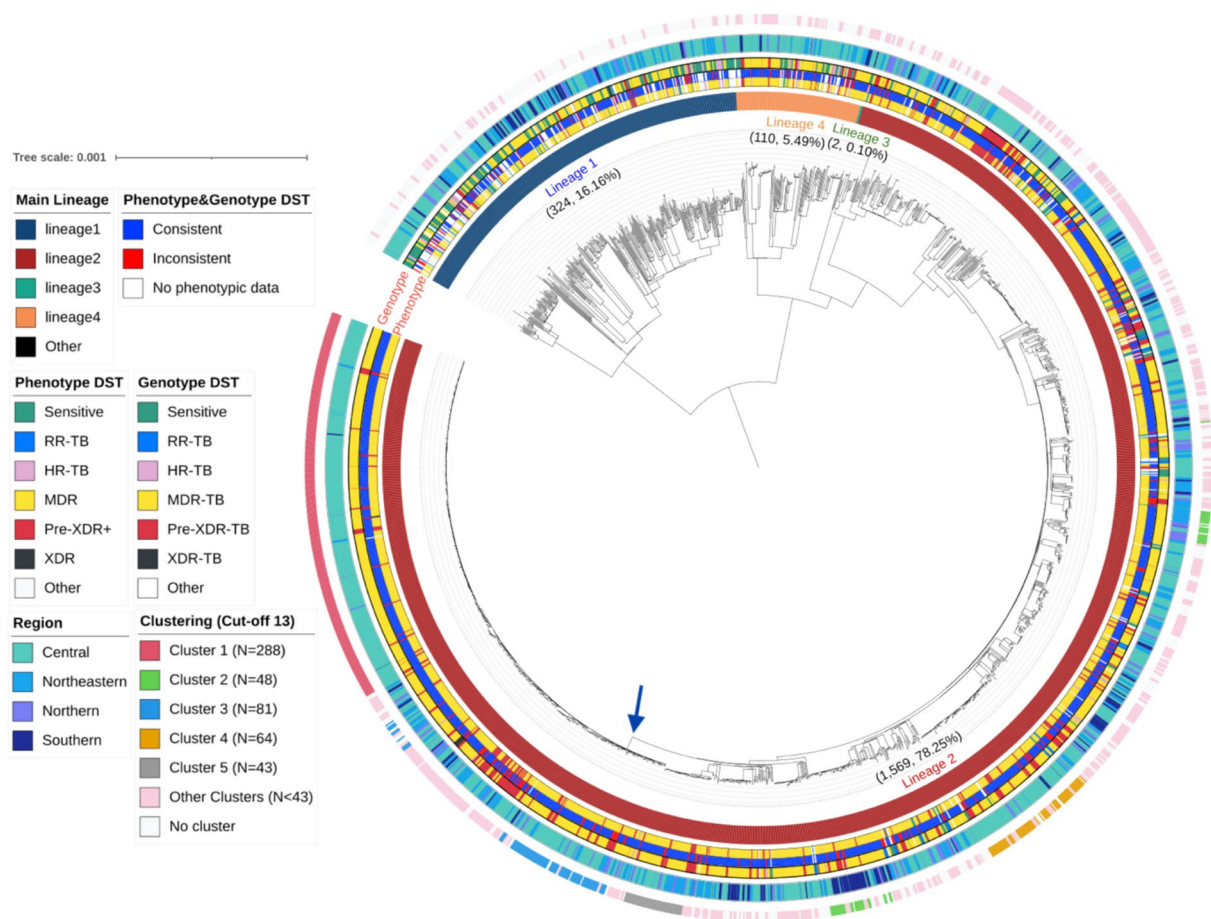
| Characteristics         |                   | N    | %    |
|-------------------------|-------------------|------|------|
| Year                    | 1994–2002         | 45   | 2.2  |
|                         | 2003–2007         | 378  | 18.9 |
|                         | 2008–2012         | 592  | 29.5 |
|                         | 2013–2017         | 838  | 41.8 |
|                         | 2018 +            | 152  | 7.6  |
| Region                  | Central           | 1299 | 64.8 |
|                         | Northeastern      | 379  | 18.9 |
|                         | Northern          | 145  | 7.2  |
|                         | Southern          | 182  | 9.1  |
| Province                | Kanchanaburi      | 558  | 27.8 |
|                         | Bangkok           | 297  | 14.8 |
|                         | Nakhon Ratchasima | 108  | 5.4  |
|                         | Buri Ram          | 104  | 5.2  |
|                         | Other             | 938  | 46.8 |
| Drug resistance         | Sensitive         | 233  | 11.6 |
|                         | RR-TB             | 39   | 1.9  |
|                         | HR-TB             | 53   | 2.6  |
|                         | MDR-TB            | 1349 | 67.3 |
|                         | Pre-XDR TB        | 293  | 14.6 |
|                         | XDR-TB            | 5    | 0.3  |
|                         | Other             | 33   | 1.7  |
| Anti-TB drug resistance | INH               | 1686 | 84.1 |
|                         | RIF               | 1693 | 84.4 |
|                         | EMB               | 1095 | 54.6 |
|                         | STR               | 1212 | 60.5 |
|                         | PZA               | 812  | 40.5 |
|                         | LFX               | 305  | 15.2 |
|                         | MXF               | 305  | 15.2 |
|                         | AMK               | 108  | 5.4  |
|                         | KAN               | 128  | 6.4  |
|                         | ETO               | 875  | 43.6 |
|                         | BDQ               | 14   | 0.7  |
|                         | LZD               | 0    | 0    |
| Lineage                 | L2                | 1569 | 78.2 |
|                         | L1                | 324  | 16.2 |
|                         | L4                | 110  | 5.5  |
|                         | L3                | 2    | 0.1  |

**Table 1.** Demographic data and genotypic drug-resistant Profiles of 2,005 TB cases, Thailand, 1994–2020. RR-TB Rifampicin-Resistant tuberculosis; HR-TB isoniazid-mono-resistant; MDR-TB, Multidrug-Resistant; Pre-XDR TB Pre-Extensively Drug-Resistant; XDR-TB Extensively Drug-Resistant; INH Isoniazid; RIF Rifampicin; EMB Ethambutol; STM Streptomycin; PZA Pyrazinamide; LFX, Levofloxacin; MXF, Moxifloxacin; AMK Amikacin, KAN Kanamycin; ETO Ethionamide; PAS Para-aminosalicylic acid; CAP Capreomycin.

for isoniazid, 96.9% for rifampicin, 97.5% for amikacin, 91.9% for levofloxacin, 89.7% for streptomycin, 96.8% for kanamycin, and 88.6% for moxifloxacin (Table S4). Assuming phenotypic DST results as the gold standard, the sensitivity and specificity for isoniazid and rifampicin, which are critical for defining MDR-TB, were approximately 96%. The remaining drugs demonstrated high levels of sensitivity and specificity, ranging from 70 to 99%, with ethambutol showing a notably lower specificity of 60.4%.

Despite the potential for DST errors, we identified 88 mutations that may account for cases of phenotypic resistance with genotypic susceptibility. These include 24 mutations linked to isoniazid (*katG* 22, *aphC* 2), 7 to rifampicin (*rpoB* 5, *rpoC* 2), 16 to ethambutol (*embA* 7, *embB* 8, *embC* 1), and 19 to streptomycin (*gid* 11, *rpsL* 2, *rrs* 6) (Table S5). Some isolates with known drug resistance mutations also harboured additional mutations that might act as compensatory mechanisms (e.g., *rpoB* Lys37Arg, *rpoB* Met655Thr; *rpoC* Gly1198Ser). At the composite resistance level (e.g., MDR-TB), the concordance between phenotypic DSTs and genotypic profiles was high at 87.4% (1589/1819) (Table S6). Most discordances (181/230, 78.7%) were linked to variations in MDR-TB classification, particularly cases where genotypic profiling identified mutations meeting the new definition of (pre-)XDR-TB (114/181). Instances of discordance where DST indicated susceptibility, but genotypic profiles suggested resistance (8/59) likely reflected phenotypic errors. These cases included several well-established



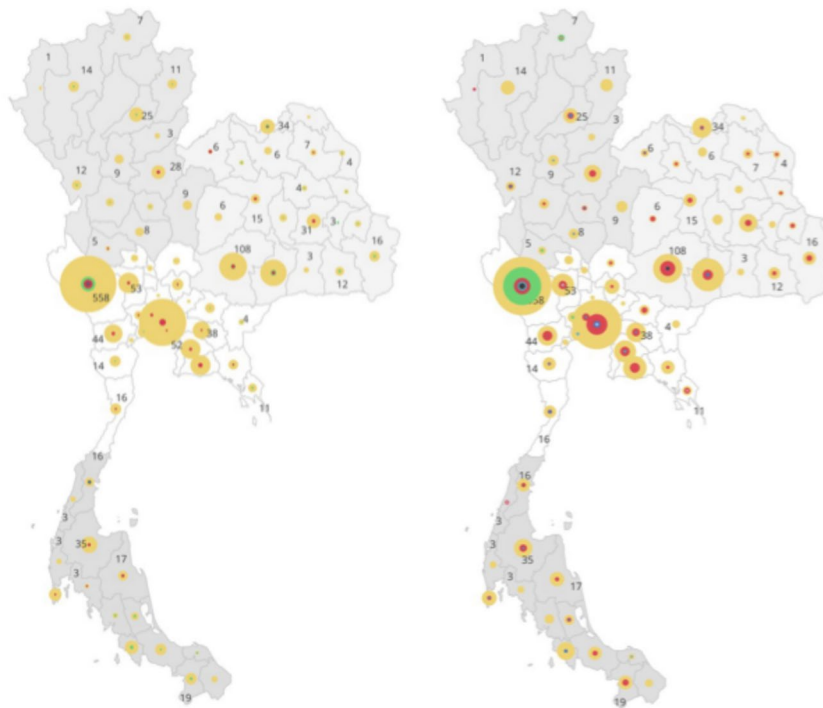


**Fig. 1.** Phylogenetic Tree of 2,005 Mtb isolates from Thailand, 1994–2020. The scale bar indicates 0.01 substitutions per site SNP. The outer ring highlights isolates within the transmission cluster, defined by an SNP distance cut-off of 13. Genotypic DSTs refer to WGS-based drug resistance profiles. Blue arrow indicates a cluster of highly similar Mtb isolates within lineage L2 (n = 531).

resistance mutations, such as *embB* Gly406Asp (ethambutol), *gyrA* Asp94Ala (moxifloxacin), and *eis* c.-10G > A (kanamycin) (Table S7).

### Transmission cluster analysis

The phylogenetic tree revealed clusters of Mtb isolates within the same lineage that were highly similar, and we sought to investigate whether these are indicative of transmission events (Fig. 1). An analysis of the pairwise SNP distance distribution across 2,005 isolates (median 452 SNPs; range: 0 to 2,369 SNPs) (Figure S2) suggested a cut-off of 13 SNPs to define potential transmission links. At this cut-off, we identified 206 transmission clusters (total n = 1,265 (63.1%); median size: 2, range: 2–288), including 190 clusters with 2–9 isolates, 15 clusters with 10–100 isolates, and one large cluster containing 288 isolates (Figure S3; Table S8). Using a 5-SNP cutoff found a similar number of transmission clusters (201). As expected, each cluster was homogenous for sub-lineage. The most prevalent sub-lineage was L2.2.1 (131/206 clusters; total n = 1,317), followed by L1.1.1 (10/206 clusters; total n = 159). The 1,265 Mtb isolates in clusters were identified in every region of Thailand, with the majority classified as MDR-TB and pre-XDR-TB (1,178/1,265, 93.1%). The clusters span several years (median 2; range 1–18 years), with a notable presence from 2005 to 2017. Within the set of 16 clusters comprising 10 or more isolates (Table S8), distinct geographical distributions were observed. For example, one cluster includes 22 isolates, all classified as L2.2.1 and pre-XDR-TB, originating from the central region, specifically Kanchanaburi, Ratchaburi, Suphanburi, and Bangkok, with samples collected between 2005 and 2017, and having the *ddn* Gly81Ser mutation. Another cluster consists of 48 isolates classified as L2.1.1, predominantly from southern Thailand, with 43 (93.8%) identified as MDR-TB; these samples were collected from 2001 to 2020. Additionally, a cluster of L4.2.2 contains 12 isolates, mainly from the central region, with 9 (75%) classified as MDR-TB and 3 (25%) as pre-XDR-TB, with samples collected from 2003 to 2017. Overall, most clusters included isolates from both the newly sequenced and older public datasets (Figure S4), highlighting the persistence of lineages L1, L2, and L4 over time. To further explore the geographical spread of these clusters, we assessed the correlation between geographic and SNP distances (Figure S5). The analysis showed a small but significant positive trend (Kruskal–Wallis  $P < 0.001$ ) with a weak positive correlation (Spearman's rho = 0.046).



**Fig. 2.** Geographic Distribution of Mtb Lineages and Genotypic Drug resistance profiles in Thailand (N = 2,005 isolates). The size of the diagrams corresponds to the number of isolates; however, the size of smaller diagrams has been adjusted to enhance visibility, as some provinces reported very few TB cases.

A logistic regression analytical approach was used to identify factors associated with an increased risk of being in a transmission cluster. Compared to lineage L1, being an isolate in L2 (odds ratio [OR] 7.41, 95% CI [5.34, 10.29],  $p < 0.001$ ) or L4 (OR 2.03, 95% CI [1.23, 3.36],  $p < 0.05$ ) had increased odds. Whilst, compared to sensitive strains, those with MDR-TB (OR 5.82, 95% CI [3.93, 8.61],  $p < 0.001$ ) or pre-XDR-TB (OR 10.98, 95% CI [6.72, 17.94],  $p < 0.001$ ) were significantly associated with an increased risk of clustering. Further, those isolates from Central and Northeastern regions were more likely to cluster compared to the Northern region (OR > 2.5,  $p < 0.001$ ) (Table S9).

Through a genome-wide analysis using common SNPs (minor allele frequency (MAF) > 0.01), we identified 4 loci that were significantly associated with being in transmission clusters, located in the genes *katG* (Rv1908c), *folC* (Rv2447c), *ppe8* (Rv0355c), and *folK* (Rv3606c) ( $P < 0.001$ ) (Table S10). Among loci with odds ratios greater than one, we detected the *katG* Ser315Thr mutation - the most common INH resistance mutation<sup>24</sup> as most clusters are linked to at least MDR-TB (OR 7.94, 95% CI [6.39, 9.87],  $p < 0.00001$ ). The *folC* (Rv2447c) S150G mutation, involved in the production of para-aminobenzoic acid (pABA) in the folate biosynthetic pathway<sup>25</sup>, was found in lineages L2 and L4 (OR 6.5, 95% CI [3.13, 13.51],  $p < 0.00001$ ). Additionally, the *ppe8* (Rv0355c) V2309I mutation was also associated with clustering (OR 2.81, 95% CI [1.05, 8.28],  $p < 0.05$ ), this gene linked to adaptations in response to host defence mechanisms<sup>26</sup>. These mutations were observed across multiple clusters, including *katG* Ser315Thr (147 clusters), *folC* S150G (6 clusters), and *ppe8* V2309I (5 clusters) (Figure S4). Like *folC*, the *folK* (Rv3606c) locus is associated with folate metabolism. While the *folK* Gln47His mutation has an odds ratio below one, there is no strong evidence of compensatory effects (Figure S4).

### Investigation of the largest cluster

The time-calibrated phylogenetic tree of the largest cluster was analysed using BEAST2 software. This cluster comprises 288 isolates, all belonging to sublineage L2.2.1, with 96.9% originating from the central region and 93.8% classified as MDR-TB. The geographic distribution is shown in Figure S6, and additional details for clusters with  $\geq 10$  isolates are provided in Table S8. This analysis estimated the mutation rate to be  $1.10 \times 10^{-7}$  substitutions per site per year (95% highest posterior density (HPD) interval:  $9.89 \times 10^{-8}$  to  $1.22 \times 10^{-7}$ ), which corresponds to  $\sim 0.48$  substitutions per genome per year (HPD: 0.44 to 0.53), and in keeping with estimates for L2 from other settings<sup>27,28</sup>. The estimated time of the most recent common ancestor is 1989 (95% HPD: 1984–1994), which is plausible given the known roll-out of the anti-TB drugs. All parameters had an effective sample size score greater than 200. The TransPhylo package was used to reconstruct the transmission history among cases, offering detailed insights into the source of infections within the studied population. Using maximum a posteriori estimates, we identified the most probable transmission pathways (Fig. 3), which revealed the evolution of Mtb, including from MDR-TB to pre-XDR-TB, through time. Based on this reconstruction, the model inferred a total of 420 transmission events within the largest cluster, of which 132 (31.4%) were attributed to unsampled individuals. These unsampled cases likely represent individuals who were infected but

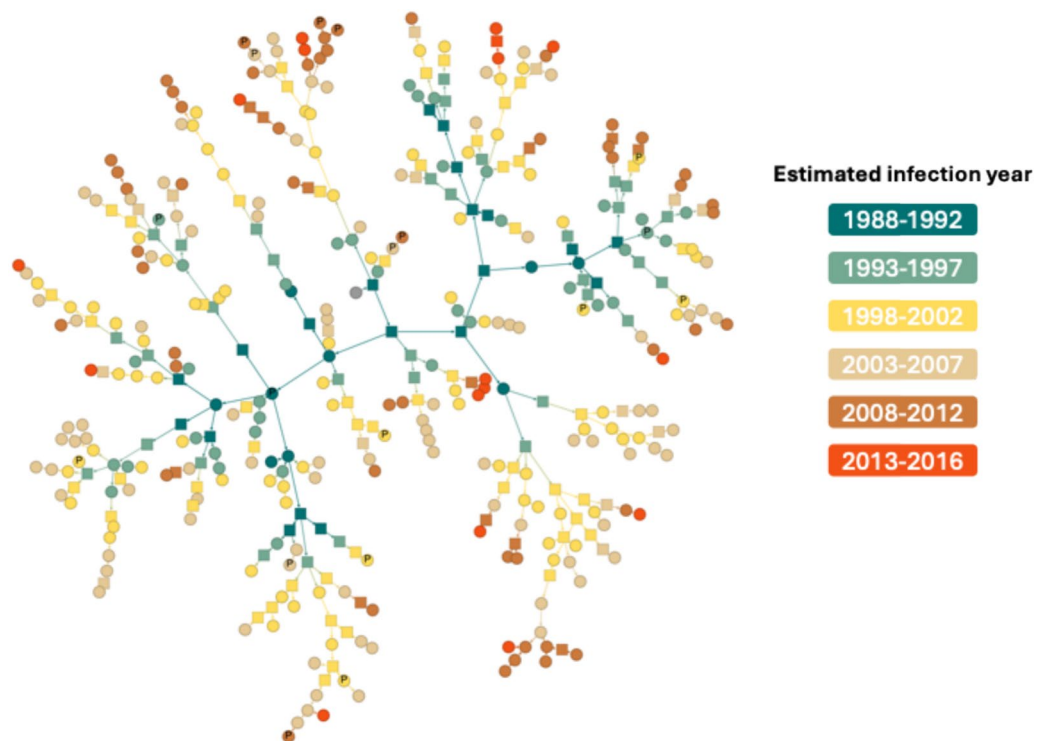
| Anti-TB drugs | Gene name   | Changes        | Our study<br>N = 2,005 | Our study<br>(%) | Global (%)<br>N = 50,723 <sup>1</sup> |
|---------------|-------------|----------------|------------------------|------------------|---------------------------------------|
| INH           | <i>katG</i> | Ser315Thr      | 1436                   | 71.62            | 26.55                                 |
|               | <i>inhA</i> | c.-777C > T    | 182                    | 9.08             | 4.06                                  |
|               | <i>inhA</i> | Ser94Ala       | 29                     | 1.45             | 0.62                                  |
|               | <i>katG</i> | Ser315Asn      | 25                     | 1.25             | 0.53                                  |
| RIF           | <i>rpoB</i> | Ser450Leu      | 923                    | 46.03            | 18.80                                 |
|               | <i>rpoB</i> | His445Tyr      | 209                    | 10.42            | 1.21                                  |
|               | <i>rpoB</i> | Asp435Val      | 114                    | 5.69             | 2.30                                  |
|               | <i>rpoB</i> | His445Asp      | 107                    | 5.34             | 1.04                                  |
| EMB           | <i>embB</i> | Gly406Asp      | 320                    | 15.96            | 0.78                                  |
|               | <i>embB</i> | Met306Val      | 261                    | 13.02            | 7.73                                  |
|               | <i>embB</i> | Met306Ile      | 236                    | 11.77            | 5.94                                  |
| STR           | <i>rpsL</i> | Lys43Arg       | 931                    | 46.43            | 13.49                                 |
|               | <i>rpsL</i> | Lys88Arg       | 74                     | 3.69             | 2.55                                  |
|               | <i>rrs</i>  | n.514A > C     | 34                     | 1.70             | 2.53                                  |
|               | <i>gid</i>  | c.102delG      | 30                     | 1.50             | 1.25                                  |
| PZA           | <i>pncA</i> | Ile31Thr       | 289                    | 14.41            | 0.05                                  |
|               | <i>pncA</i> | Ile90Ser       | 31                     | 1.55             | 0.05                                  |
|               | <i>pncA</i> | c.-11A > G     | 24                     | 1.20             | 0.90                                  |
| LFX           | <i>gyrA</i> | Asp94Gly       | 119                    | 5.94             | 4.30                                  |
|               | <i>gyrA</i> | Ala90Val       | 88                     | 4.39             | 3.15                                  |
|               | <i>gyrA</i> | Asp94Ala       | 39                     | 1.95             | 1.18                                  |
| MFX           | <i>gyrA</i> | Asp94Gly       | 119                    | 5.94             | 4.30                                  |
|               | <i>gyrA</i> | Ala90Val       | 88                     | 4.39             | 3.15                                  |
|               | <i>gyrA</i> | Asp94Ala       | 39                     | 1.95             | 1.18                                  |
| AMK           | <i>rrs</i>  | n.1401A > G    | 98                     | 4.89             | 4.71                                  |
|               | <i>eis</i>  | c.-14C > T     | 8                      | 0.40             | 0.48                                  |
| KAN           | <i>rrs</i>  | n.1401A > G    | 98                     | 4.89             | 4.71                                  |
|               | <i>eis</i>  | c.-10G > A     | 17                     | 0.85             | 0.72                                  |
|               | <i>eis</i>  | c.-14C > T     | 8                      | 0.40             | 0.48                                  |
| ETO           | <i>ethA</i> | c.639_640delGT | 534                    | 26.63            | 0.18                                  |
|               | <i>inhA</i> | c.-777C > T    | 182                    | 9.08             | 4.06                                  |
|               | <i>inhA</i> | Ser94Ala       | 29                     | 1.45             | 0.62                                  |
| PAS           | <i>folC</i> | Ser150Gly      | 91                     | 4.54             | 0.31                                  |
|               | <i>thyX</i> | c.-16C > T     | 76                     | 3.79             | 0.75                                  |
|               | <i>folC</i> | Glu40Gly       | 29                     | 1.45             | 0.17                                  |
| CAP           | <i>rrs</i>  | n.1401A > G    | 98                     | 4.89             | 4.71                                  |
|               | <i>tlyA</i> | c.52_53dupCG   | 3                      | 0.15             | < 0.001                               |

**Table 2.** Summary of Anti-TB Drug Resistance Mutations in 2,005 Isolates. There could be > 1 mutation associated with resistance to a single drug. <sup>1</sup>The global percentage of anti-TB drug resistance mutations identified by TB-Profler database; INH, Isoniazid; RIF, Rifampicin; EMB, Ethambutol; STR, Streptomycin; PZA, Pyrazinamide; LFX, Levofloxacin; MFX, Moxifloxacin; AMK, Amikacin, KAN, Kanamycin; ETO, Ethionamide; PAS, Para-aminosalicylic acid; CAP, Capreomycin.

not captured in our dataset due to limitations in retrospective sampling, diagnostic coverage, or sequencing availability. (Figure S7). By applying a probability threshold of > 0.2, the large cluster was subdivided into 83 smaller sub-clusters. This breakdown facilitated more detailed investigation of each sub-cluster, allowing us to visualise potential transmission pathways (i.e., who-infected-whom), integrate drug resistance profiles and collection years, and assess epidemiological plausibility using available metadata to guide further contact investigations by infection control teams (Figure S8). We also compared the Thai clusters to a reference set of 154 representative global Beijing strains<sup>29</sup>, revealing notable similarities (Figure S9).

## Discussion

The WHO classifies Thailand as a high-burden TB country. This study presents a WGS analysis of 2,005 Mtb isolates from Thailand, spanning three decades (1994–2020), to examine circulating strains, drug resistance profiles, and transmission dynamics. The dataset includes isolates from previous studies<sup>11,15</sup> and new data from Siriraj Hospital, with most samples originating from cases involving drug resistant TB, of which over 60% classified as MDR-TB. The frequency distribution of randomly recruited Mtb isolates in Thailand predominantly



**Fig. 3.** The maximum a posteriori (MAP) transmission tree generated from TransPhylo results, showing the transmission dynamics of TB cases. Nodes represent hosts, coloured according to the estimated infection year intervals: 1988 (1988–1992), 1993 (1993–1997), 1998 (1998–2002), 2003 (2003–2007), 2008 (2008–2012), and 2013 (2013–2016). Squares denote unsampled cases, while grey nodes represent the initial infection source. Nodes labelled with “P” represent Pre-XDR-TB cases, while unlabelled circles indicate MDR-TB cases.

comprises L1 and L2<sup>13</sup>. Our findings reveal a high prevalence of L2, particularly the Beijing sub-lineage (L2.2.1). This aligns with prior studies reporting that L2 is the main lineage related to clusters of outbreaks and drug-resistant isolates in Thailand, while L1 is more prevalent in general *Mtb* cases<sup>13</sup>. Although our comparative analysis with global MDR clusters is limited, our findings indicate that the dominant MDR clusters in Thailand align with the global expansion of L2, sharing similar underlying resistance mutations. A more detailed phylogenetic and cluster comparison with global MDR genomes<sup>6</sup> would be valuable in future studies to explore this further. The geographic patterns observed suggest that specific lineages may be driving local transmission, especially in densely populated regions such as the Central area.

The genotypic drug resistance profiles generated by TB-Profiler software had high concordance with phenotypic DST, confirming its reliability. Since most isolates were MDR-TB, resistance rates detected through both phenotypic DST and genotypic resistance profiling were highest for isoniazid and rifampicin. Discrepancies between phenotypic and genotypic approaches may arise due to DST laboratory errors, limitations in WGS's ability to capture certain mutations, or gaps in the drug resistance mutation database used for annotation. We identified 88 low-frequency mutations that explained phenotypic resistance and genotypic sensitivity inconsistencies, including 66 linked to four drugs: isoniazid (24), rifampicin (7), ethambutol (16), and streptomycin (19). These mutations, confirmed using phenotypic DST, were not lineage-specific and occurred at low frequencies in a dataset of 50,723 global *Mtb* samples. As such, they should be considered for inclusion in the TB-Profiler database. At the DR profile level, samples initially classified as MDR-TB by phenotypic DST were re-categorised genotypically in over 4.9% of cases to a lower resistance level (sensitive, RR-TB, or HR-TB) and in 7.5% of cases to a higher resistance level (pre-XDR-TB or XDR-TB). This reclassification is a crucial consideration, as treatment for each type of TB is specific, with distinct drug regimens and durations<sup>30</sup>. Inappropriate treatment could lead to further complications. A particularly concerning finding was the presence of the *ddn* Gly81Ser mutation, which is associated with delamanid resistance<sup>23</sup>, even in isolates predating the drugs rollout. Intrinsic resistance to delamanid and pretomanid - both sharing a similar mechanism of action - could undermine the effectiveness of the latest BPaLM regimen, especially as bedaquiline resistance continues to rise. Although this mutation may be lineage-associated, its functional impact on drug resistance has been experimentally validated, supporting its biological relevance. These findings emphasise the importance of using WGS data to identify mutations potentially associated with drug resistance, even in cases where phenotypic DST results indicate sensitivity.

The phylogenetic analysis of *Mtb* isolates revealed distinct clusters within the same lineage, suggesting potential transmission chains. Using a pairwise SNP distance cut-off of 13, we identified 206 transmission clusters comprising 1,265 isolates. These clusters ranged from small groups (190 clusters with 2–9 isolates) to



larger networks (16 clusters with 10–100 isolates), including one exceptionally large cluster of 288 isolates, all highly homogenous for sub-lineage. The presence of transmission clusters in every region of Thailand, coupled with a substantial proportion of MDR-TB and pre-XDR-TB cases, underscores the widespread nature of drug-resistant TB across the country. Clusters spanned several years, with the majority arising between 2005 and 2017, reflecting ongoing transmission. Due to the convenient sampling method, it is possible that other isolates existed outside the given collection years. Clusters with higher percentages in more recent years (e.g., 2012–2017) may indicate ongoing transmission. It suggests that certain strains have persisted over time, reflecting continuous transmission of similar strains across years.

The study found that lineage, drug resistance, and geographic factors were significantly associated with the transmission dynamics of Mtb. Lineage L2 exhibited the highest transmissibility and adaptability, showing the strongest odds of clustering. Although less frequently involved in outbreaks than L2, L4 still presented an increased risk of clustering<sup>31</sup>, with both being modern strains. Drug resistance, particularly MDR-TB and pre-XDR-TB, further increased the risk of clustering. Geographic factors also played a key role, with isolates from the Central and Northeastern regions having higher risk of clustering. These findings can inform the design of tailored interventions aimed at high-risk areas and populations.

Using a GWAS approach, three loci in *katG*, *folC*, and *ppe8* were identified as associated with clustering or transmissibility, suggesting a potential increase in Mtb fitness and transmission. These loci have not been reported in similar analyses in Asia with comparable lineages<sup>32,33</sup>. Both the *katG* and *folC* genes are associated with drug resistance<sup>25,32</sup>. However, studies suggest that the increased transmission risk of drug-resistant tuberculosis is driven more by delays in diagnosis and treatment than by the inherent transmissibility of resistant strains<sup>34</sup>. The *ppe8* locus exhibits complexity, with deletions at the gene's start in L1 strains and within the encompassed RD304 region in L2 strains<sup>4</sup>. Although Mtb GWAS can be used to identify associations between SNPs and transmissibility, some suggest incorporating host genomics into the analysis, such as through genome-to-genome approaches, for more comprehensive results<sup>14</sup>.

We identified a large cluster of 288 isolates and employed time-calibrated phylogenetic analysis using BEAST2 to reconstruct the outbreak's transmission history from WGS data. We subsequently used TransPhylo to infer transmission patterns within the studied population. It enabled us to estimate the number of unsampled cases, revealing a hidden burden of TB transmission and suggesting that undetected infections may sustain ongoing transmission chains. To improve confidence in our inferences, we filtered out transmission probabilities below 0.2, reflecting the varied, convenience-based nature of sampling. Consequently, we could more reliably focus on smaller transmission clusters. Occasionally, we observed transmission patterns that appeared counterintuitive, such as transmission from Pre-XDR-TB to MDR-TB cases or from more recent years to earlier years (e.g., 2008 to 2006). These anomalies may reflect unsampled transmission links or uncertainties in determining the actual timing of infection. Specifically, our analysis relied on the year of sample collection rather than the precise year of infection, which can introduce discrepancies. By continually integrating updated WGS data, we can enhance the monitoring of patients and their transmission networks, thereby strengthening efforts to contain and prevent the spread of transmission clusters.

While our study shows that WGS is useful for studying drug resistance and TB transmission, several challenges remain before its routine implementation in TB control programs. These include the high cost of sequencing, delays in turnaround time in clinical settings, and the requirement for specialised personnel and infrastructure to support genomic data analysis and interpretation. Furthermore, effective integration into national TB surveillance systems necessitates standardised protocols, real-time data sharing, and clear guidance on how to act on genomic findings. Despite these challenges, WGS provides valuable insights by identifying undetected cases, transmission clusters, and resistance-associated mutations. These insights can support more targeted contact investigations, track the spread of high-risk strains, and inform treatment strategies in regions with ongoing transmission. Future efforts should aim to enhance WGS accessibility, streamline analytical workflows, and integrate host and pathogen genomic data to improve our understanding of TB transmission and support precision public health interventions.

In conclusion, this study provides a comprehensive analysis of WGS data from Mtb isolates in Thailand, offering valuable insights into circulating strains, drug resistance profiles, and transmission dynamics, particularly among drug-resistant TB cases. The findings highlight the critical role of integrating genomic and epidemiological data to deepen our understanding of Mtb and to guide targeted interventions for TB control. However, WGS alone cannot address all TB-related challenges. Sustained efforts combining WGS with strengthened public health strategies, enhanced surveillance systems, and multifaceted interventions are essential to effectively combat TB in Thailand and beyond.

## Data availability

Raw sequencing data are available in the European Nucleotide Archive (ENA). A complete list of accession numbers is provided in Table S1.

Received: 29 March 2025; Accepted: 5 August 2025

Published online: 13 August 2025

## References

- Salina, E. G. *Mycobacterium tuberculosis* Infection: Control and Treatment. *Microorganisms* **11**, 1057 (2023).
- WHO. Global tuberculosis report 2023. Geneva: World Health Organization, 2023 <https://iris.who.int/handle/10665/373828>.
- Maung, H. M. W. et al. Geno-spatial distribution of *mycobacterium tuberculosis* and drug resistance profiles in Myanmar-Thai Border Area. *Trop. Med. Infect. Dis.* **5**, 153 (2020).

4. Thorpe, J. et al. Multi-platform whole genome sequencing for tuberculosis clinical and surveillance applications. *Sci. Rep.* **14**, 5201 (2024).
5. Dippenaar, A. et al. Nanopore Sequencing for *Mycobacterium tuberculosis*: A critical review of the literature, new developments, and future opportunities. *J. Clin. Microbiol.* **60**, e0064621 (2022).
6. Phelan, J. E. et al. Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs. *Genome Med.* **11**, 41 (2019).
7. WHO. The use of next-generation sequencing technologies for the detection of mutations associated with drug resistance in *Mycobacterium tuberculosis* complex: technical guide. Geneva: World Health Organization, <https://iris.who.int/handle/10665/274443> (2018).
8. Satta, G. et al. *Mycobacterium tuberculosis* and whole-genome sequencing: how close are we to unleashing its full potential?. *Clin. Microbiol. Infect.* **24**, 604–609 (2018).
9. WHO. Catalogue of mutations in *Mycobacterium tuberculosis* complex and their association with drug resistance, 2nd ed. Geneva: World Health Organization, <https://iris.who.int/handle/10665/374061> (2023).
10. Division of Tuberculosis. National Tuberculosis Control Programme Guideline, Thailand 2021. Bangkok, (2021).
11. Nonghanphithak, D. et al. Clusters of drug-resistant *mycobacterium tuberculosis* detected by whole-genome sequence analysis of nationwide sample, Thailand, 2014–2017. *Emerg. Infect. Dis.* **27**, 813–822 (2021).
12. Faksri, K. et al. Comparisons of whole-genome sequencing and phenotypic drug susceptibility testing for *Mycobacterium tuberculosis* causing MDR-TB and XDR-TB in Thailand. *Int. J. Antimicrob. Agents* **54**, 109–116 (2019).
13. Miyahara, R. et al. Risk for Prison-to-Community Tuberculosis Transmission, Thailand, 2017–2020. *Emerg. Infect. Dis.* **29**, 477–483 (2023).
14. Phelan, J. et al. Genome-wide host-pathogen analyses reveal genetic interaction points in tuberculosis disease. *Nat. Commun.* **14**, 549 (2023).
15. Srihohasin, P. et al. Genomic evidence supporting the clonal expansion of extensively drug-resistant tuberculosis bacteria belonging to a rare proto-Beijing genotype. *Emerg. Microbes. Infect.* **9**, 2632–2641 (2020).
16. World Health Organization. Meeting report of the WHO expert consultation on the definition of extensively drug-resistant tuberculosis, 27–29 October 2020 (World Health Organization, 2021).
17. Gómez-González, P. J. et al. Functional genetic variation in *pe/ppe* genes contributes to diversity in *Mycobacterium tuberculosis* lineages and potential interactions with the human host. *Front. Microbiol.* **14**, 1244319 (2023).
18. Bouckaert, R. et al. BEAST 2: A software platform for bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537 (2014).
19. Sobkowiak, B. et al. Bayesian reconstruction of *Mycobacterium tuberculosis* transmission networks in a high incidence area over two decades in Malawi reveals associated risk factors and genomic variants. *Microb. Genom.* **2020**, 6. <https://doi.org/10.1099/mgen.0.000361> (2020).
20. Phelan, J. E. et al. TGV: suite of tools to visualize transmission graphs. *NAR Genom Bioinform* **2024**, 6. <https://doi.org/10.1093/nargab/lqae158> (2024).
21. Coll, F. et al. Genome-wide analysis of multi- and extensively drug-resistant *Mycobacterium tuberculosis*. *Nat. Genet.* **50**, 307–316 (2018).
22. Department of Disease Control, Ministry of Public Health, Thailand. Thailand's Sixth Joint International Tuberculosis Monitoring Mission Report, 2016–2020. 1st ed. Nonthaburi: Bureau of General Communicable Diseases; 2021. 60 p. ISBN: 978–616–11–4820–1.
23. Kim, S. et al. Functional analysis of genetic mutations in *ddn* and *fbtA* linked to delamanid resistance in rifampicin-resistant *Mycobacterium tuberculosis*. *Tuberculosis* <https://doi.org/10.1016/j.tube.2025.102630> (2025).
24. Ando, H. et al. Identification of *katG* mutations associated with high-level isoniazid resistance in *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* **54**, 1793–1799 (2010).
25. Wang, R., Li, K., Yu, J., Deng, J. & Chen, Y. Mutations of *folC* cause increased susceptibility to sulfamethoxazole in *Mycobacterium tuberculosis*. *Sci. Rep.* **2021**, 11. <https://doi.org/10.1038/s41598-020-80213-4> (2021).
26. Chitwood, M. H. et al. The recent rapid expansion of multidrug resistant Ural lineage *Mycobacterium tuberculosis* in Moldova. *Nat. Commun.* **2024**, 15. <https://doi.org/10.1038/s41467-024-47282-9> (2024).
27. Torres Ortiz, A. et al. Genomic signatures of pre-resistance in *Mycobacterium tuberculosis*. *Nat. Commun.* **2021**, 12. <https://doi.org/10.1038/s41467-021-27616-7> (2021).
28. Rutaihw, L. K. et al. Multiple introductions of *Mycobacterium tuberculosis* Lineage 2-Beijing into Africa over centuries. *Front. Ecol. Evol.* **2019**, 7. <https://doi.org/10.3389/fevo.2019.00112> (2019).
29. Merker, M. et al. Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nat. Genet.* **47**, 242–249 (2015).
30. Alsayed, S. S. R. & Gunosewoyo, H. Tuberculosis: Pathogenesis, Current Treatment Regimens and New Drug Targets. *Int. J. Mol. Sci.* **2023**, 24. <https://doi.org/10.3390/ijms24065202> (2023).
31. Holt, K. E. et al. Frequent transmission of the *Mycobacterium tuberculosis* Beijing lineage and positive selection for the *EsxW* Beijing variant in Vietnam. *Nat. Genet.* **50**, 849–856 (2018).
32. Wang, L. et al. Whole genome sequencing analysis of *Mycobacterium tuberculosis* reveals circulating strain types and drug-resistance mutations in the Philippines. *Sci. Rep.* **2024**, 14. <https://doi.org/10.1038/s41598-024-70471-x> (2024).
33. Napier, G. et al. Characterisation of drug-resistant *Mycobacterium tuberculosis* mutations and transmission in Pakistan. *Sci. Rep.* **2022**, 12. <https://doi.org/10.1038/s41598-022-11795-4> (2022).
34. Atre, S. R. et al. Tuberculosis Pathways to Care and Transmission of Multidrug Resistance in India. *Am. J. Respir. Crit. Care Med.* **205**, 233–241 (2022).

## Acknowledgements

The study was funded by MRC – NSTDA – Newton (ref. MR/R020973/1; P-18-50228). NT is funded by a Thailand government Ministry of Public Health PhD scholarship. LW is funded by a BBSRC LIDO PhD studentship. WH is a recipient of a Chulabhorn Royal Academy fellowship. TGC and SC are funded by the UKRI (BBSRC BB/X018156/1; MRC MR/X005895/1; EPSRC EP/Y018842/1)). NT, WS, SW and SM were funded by the Department of Medical Sciences (Thailand Ministry of Public Health). The Thailand Health Systems Research Institute provided funding through HSRI 64-179 (NT, WS, SW and SM) and HSRI 65-113 (KF, ST, AC, TP, PS, WP). KF, ST and AC is funded by the National Research Council of Thailand (NRCT) (NRC MHESI 483/2563). The funders had no role in study design, data collection and analysis, decision to publish, or the preparation of the manuscript.

## Author contributions

SM, AC, PSu, KF, and TGC conceived and supervised the project. PS, TP, KF, PSu, and AC conducted sample processing and DNA extraction. PS performed bacterial culture. SC and PS conducted sequencing. NT and WP

performed bioinformatics and statistical analyses under the supervision of PS, JEP, SC, SW, SM, AC, and TGC. NT, PS, JEP, WS, MLH, SC, SW, SM, AC, and TGC contributed to data interpretation. NT drafted the manuscript, with input from all authors. All authors reviewed, edited, and approved the final manuscript. NT, PS, SM, AC, and TGC compiled the final submission.

### Funding

Ministry of Public Health, PhD Scholarship, Thailand Health Systems Research Institute, HSRI 65-113, HSRI 65-113, HSRI 65-113, HSRI 65-113, HSRI 65-113, HSRI 65-113, HSRI 64-179, HSRI 64-179, HSRI 64-179, HSRI 65-113, National Research Council of Thailand, NRC MHESI 483/2563, NRC MHESI 483/2563, NRC MHESI 483/2563, Chulabhorn Royal Academy, Fellowship, UKRI, BBSRC BB/X018156/1; MRC MR/X005895/1; EPSRC EP/Y018842/1, BBSRC BB/X018156/1; MRC MR/X005895/1; EPSRC EP/Y018842/1, MRC – NSTDA – Newton, ref. MR/R020973/1; P-18-50228, ref. MR/R020973/1; P-18-50228.

### Declarations

#### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-15093-7>.

**Correspondence** and requests for materials should be addressed to A.C. or T.G.C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025