Diversified Lineages and Drug-Resistance Profiles of Clinical Isolates of *Mycobacterium tuberculosis* Complex in Malaysia

Mohd Nur Noorizhab Fakhruzzaman1, Norzuliana Zainal Abidin2, Zirwatul Adilah Aziz2, Wai Feng Lim3, Johari James Richard3, Mohd Noordin Noorliza3, Mat Hussin Hani4, Rusli Norhayati4, Abu Bakar Zamzurina4, Md. Yusof Farida Zuraihana4, Muhammad Jamari Hisyam1, Lay Kek Teh1, Mohd Nor Norazmi1, Mohd Salleh Zaki1

1Faculty of Pharmacy, Integrative Pharmacogenomics Institute, Universiti Teknologi MARA, 2Faculty of Applied Sciences, Integrative Pharmacogenomics Institute, Universiti Teknologi MARA, 3Integrative Pharmacogenomics Institute, Universiti Teknologi MARA, Puncak Alam, 4School of Health Sciences, Universiti Sains Malaysia, Kubang Kerian, Kelantan, Malaysia

**Abstract**

**Background:** Tuberculosis (TB) is still a major health problem in Malaysia with thousands of cases reported yearly. This is further burdened with the emergence of multidrug-resistant TB (MDR-TB). Whole-genome sequencing (WGS) provides high-resolution molecular epidemiological data for the accurate determination of *Mycobacterium tuberculosis* complex (MTBC) lineages and prediction of the drug-resistance patterns. This study aimed to investigate the diversity of MTBC in Malaysia in terms of lineage and drug-resistance patterns of the clinical MTBC isolates using WGS approach. **Methods:** The genomes of 24 MTBC isolated from sputum and pus samples were sequenced. The phenotypic drug susceptibility testing (DST) of the isolates was determined for ten anti-TB drugs. Bioinformatic analysis comprising genome assembly and annotation and single-nucleotide polymorphism (SNP) analysis in genes associated with resistance to the ten anti-TB drugs were done on each sequenced genome. **Results:** The draft assemblies covered an average of 97% of the expected genome size. Eleven isolates were aligned to the Indo-Oceanic lineage, eight were East-Asian lineage, three were East African-Indian lineage, and one was of Euro-American and Bovis lineages, respectively. Twelve of the 24 MTBC isolates were phenotypically MDR *M. tuberculosis*: one is polyresistance and another one is monoresistance. Twenty-six SNPs across nine genes associated with resistance toward ten anti-TB drugs were detected where some of the mutations were found in isolates that were previously reported as pan-susceptible using DST. A haplotype consisting of 65 variants was also found among the MTBC isolates with drug-resistance traits. **Conclusions:** This is the first effort done in Malaysia to utilize 24 genomes of the local clinical MTBC isolates. The high-resolution molecular epidemiological data obtained provide valuable insights into the mechanistic and epidemiological qualities of TB within the vicinity of Southeast Asia.

**Keywords:** Multidrug resistance, *Mycobacterium tuberculosis*, next-generation sequencing-based drug susceptibility testing screening, phenotypic drug susceptibility testing screening, whole-genome sequence

**Introduction**

Tuberculosis (TB) accounts for 1.3 million deaths in 2017 making it one of the top ten causes of death and the leading cause of death due to a single infection agent. In 2017, ten million people were estimated to develop TB worldwide where 5% (558,000) of the cases were expected to be caused by multidrug-resistant *Mycobacterium tuberculosis* (MDR-Mtb) strain. In Malaysia, the annual government health census reported that the incidence rate for TB was 81.30/100,000 population in 2017, with 26,422 patients diagnosed with TB. Another disturbing worldwide phenomenon is that the difference between the numbers of reported TB cases and the number of successful TB treatments is still relatively big. This phenomenon is believed to be linked to the emergence...
of MDR-TB and extensively drug-resistant TB (XDR-TB) where the success rate of treatment on MDR-TB was reported to be low (55% globally) with 8.5% of the estimated MDR-TB cases due to XDR-TB.\(^1\) For the record, MDR-TB is caused by \textit{M. tuberculosis} complex (MTBC) which developed resistance toward isoniazid (INH) and rifampicin (RIF), whereas XDR-TB is caused by MTBC-resistant against INH, RIF, fluoroquinolones, and one of the injectable second-line anti-TB drugs such as amikacin (AMK), capreomycin (CAP), and kanamycin (KAN).\(^2\)

In general, \textit{Mycobacterium} spp. acquire drug-resistant phenotypes following mutation on certain genes. So far, up to 42 genes within strains of \textit{M. tuberculosis} have been identified to be highly linked to resistance toward 15 anti-TB drugs. The gold standard method which involves culturing the bacteria on Löwenstein–Jensen (LJ) medium followed by biochemical drug susceptibility testing (DST) on the same medium could take up to 2 months for the final report to be released by the laboratory.\(^3\) Application of PCR-based molecular DST kits like the GeneXpert (Cepheid Inc., USA), MTBDRplus and MTBDRsl (Hain Lifescience, Denmark), INNO-LiPA RifTB (Innogenetics, Belgium) and AID TB Resistance (AID Diagnostika, Denmark) which detect the presence of the pathogen, as well as testing for resistance to a few anti-TB drugs managed to reduce the analysis turnaround time to two or four days.\(^4\)

Even so, most of the available PCR-based molecular DST kits determine the resistance pattern of the pathogen based on a few selected mutations at \textit{rpoB}, \textit{katG}, \textit{inhA}, \textit{gyrA}, \textit{rrs}, \textit{embB}, and \textit{rpsL} genes, a mere seven of 42 genes that were found to be highly linked to drug resistance. This would pose the risk of having false-negative results on the resistant strains which in return lead to the prescription of hepatotoxic drugs such as INH and pyrazinamide (PZA) to patients with little or no benefit at all. Therefore, there are practically no rapid methods that can accurately identify drug-resistant MTB for clinical use. Perhaps, one of the effective approaches to control the emergence of MDR-TB cases is to profile and document the genomic structures of the MTB and avoid the emergence of resistant cases through high-throughput sequencing approach to profile the drug-resistance patterns of the pathogen.

Whole-genome sequencing (WGS) analysis is able to provide deep resolution molecular epidemiological data which not only accurately identify the pathogen that caused TB but also all the possible mutations that confer the drug resistance. In 2016, a study done by Wellcome Trust/University of Oxford (and collaborators) has not only proven that WGS can be integrated into the routine TB diagnostic workflows, but also the data regarding the pathogen identity and its drug-resistance patterns can be generated within 9 days (compared to 2 months using the conventional workflow) and at cheaper (if not similar) price than the current workflow.\(^5\) In Malaysia, genome-based TB studies are still lacking as most of the molecular studies used previously were PCR-based tools. Several studies on the genomes of local isolates of \textit{M. tuberculosis} had been published;\(^6\) however, most of the studies only focused on a single genome which limited the epidemiological quality of the data.

**Methods**

**Ethical consideration**

The study was approved by the Research Ethics Committee of Universiti Teknologi MARA, Malaysia (Approval No: REC/271/17), and the National Medical Research Registration by the Ministry of Health, Malaysia (Approval No: NMRR-18-833-40380).

**Mycobacterium tuberculosis complex isolates**

A total of 24 MTBC isolates were collected from a total of 24 TB patients in 2017. Nineteen isolates were cultured from pulmonary (sputum) and five from extrapulmonary (pus) clinical specimens. Patients’ identities were kept confidential. All specimens were cultured on solid LJ medium and MGIT 960 system media according to the manufacturer’s instructions (Becton Dickinson, Sparks, MD, USA).

**Phenotypic drug susceptibility testing**

MTBC isolates that were successfully grown on the LJ medium were tested against anti-TB drugs using automated liquid method based on absolute concentration method or proportion method (BACTEC MGIT 960 TB detection system).\(^6,11\) The drugs tested were streptomycin (STM), INH, RIF, ethambutol (EMB), PZA, AMK, CAP, KAN, ofloxacin (OFX), and moxifloxacin (MXF). However, phenotypic DST testing against PZA, AMK, CAP, KAN, OFX, and MXF was only done on isolates that were resistant toward STM, INH, RIF, and EMB. \textit{M. tuberculosis} H37R\(_v\) strain was used as the internal control. The concentrations of the drugs used in both methods were as prescribed in the standard WHO guideline of drug sensitivity test for MTB.\(^12\)

**Genomic DNA extraction**

Suspensions of MTBC isolates were prepared by inoculating the colonies of the respective MTBC isolates from the LJ culture with 500 µL of 10-mM Tris-HCl buffer (pH 8.0). The MTBC suspensions were deactivated by incubation at 95°C for 10 min. Next, extraction and purification of genomic DNA of each MTBC isolates were done using an in-house modified chloroform extraction method. Quantification and quality assessment of each purified DNA were performed using Qubit fluorometer with hsDNA Assay Kit (Thermo Fisher Scientific), NanoDrop 2000c spectrophotometry (Thermo Fisher Scientific), and agarose gel electrophoresis. MTBC genomic DNA samples with the purity of 1.8–2.0 (OD\(_{260/280}\)) and 2.0–2.2 (OD\(_{260/230}\)) at concentration >50 ng/µL were used for the subsequent WGS.

**Whole-genome sequencing, assembly, and annotation of the Mycobacterium tuberculosis complex genomes**

The genomes of the 24 MTBC isolates were sequenced using Illumina Miseq genome sequencer with a 2 × 250 paired-end sequencing strategy. Libraries of 150-bp paired-end samples were prepared using Nextera XT DNA Sample
Preparation Kit (Illumina, USA), according to the manufacturer’s protocol. The quality of the.fastq files was assessed using the FastQC software followed by removal of adaptor sequences and trimming of reads with Phred score below Q30 using the Cutadapt software.\(^{[13,14]}\) De novo genome assemblies were done using the Velvet assembler together with the VelvetOptimiser tool.\(^{[15,16]}\) Resulting contigs were mapped and sorted against the *M. tuberculosis* H37Rv (GenBank accession number: NC_000962) genome as reference using the Mauve software.\(^{[17]}\) *In silico* gap filling was done using the Medusa. All the assembled draft genomes were submitted to the Rapid Annotation Using Subsystem Technology (RAST) server for genome annotation.\(^{[18,19]}\) Trimmed.fastq files of each MTBC isolates were submitted to TB Profiler and PhyResSE annotation servers to detect their spoligotypes and SNPs associated with resistance toward INH, RIF, EMB, STM, PZA, and fluoroquinolones (FLQ).\(^{[20,21]}\) Using the spoligotype profile of each MTBC, a neighbor-joining (N-J) phylogenetic tree was constructed against 85 references *Mycobacterium* spp. strains [Figure 1] using a web-based tool available at www.miru-vntrplus.org.\(^{[22]}\) Haplotype of drug-resistant MTBC was analyzed using Snippy, a variant calling software (https://github.com/tseemann/snippy).

**RESULTS**

*Mycobacterium tuberculosis* complex isolates

The phenotypic DST tests showed that 10 of the 24 MTBC isolates were pan-susceptible isolates, whereas the rest were at least resistant toward RIF. Detailed information on the drug-resistance profiles of each MTBC from the phenotypic DST assays is summarized in Table 1.

### Genomic properties, lineages, and phylogeny of each *Mycobacterium tuberculosis* complex genome

Overall, the sizes of the assembled draft genomes range between 4.2 and 4.35 Mbp with an average genome completion of 97% and a sequencing coverage of 107X (ranging from 61X to 159X). Based on the TB Profiler, 23 MTBC isolates were *M. tuberculosis* and 1 was *Mycobacterium bovis*. In terms of lineage, 11 (45.8%) isolates belong to the Indo-Oceanic lineage, eight (33.3%) were East-Asian lineage, three (12.5%) were East-African-Indian lineage, and the remaining one (4.2%) was of Euro-American and Bovis lineages, respectively. The distribution of lineages and drug-resistance profiles of each MTBC isolate is shown in Table 2. In addition, phylogenetic analysis revealed that the 24 MTBC isolates were clustered into 5 main groups involving 12 strains of *Mycobacterium* spp., as listed in Table 3 and illustrated in Figure 1.

### Predicted drug resistance and concordance with drug susceptibility testing

The drug-resistance patterns of the MTBC isolates which were predicted using TB Profiler and PhyResSE were compared against the results of phenotypic DST, as summarized in Table 4. Overall, there were a few discrepancies of MTB sensitivities toward the drugs. For INH, 11 of the 24 isolates

---

**Table 1: Lineage, spoligotype and phenotypic drug resistant pattern of the *Mycobacterium tuberculosis* complex isolates**

| Isolate | Resistant pattern | Spoligotype | Lineage | Clade |
|---------|-------------------|-------------|---------|-------|
| R10     | SIRZ              | 407765460412251 | Indo-Oceanic (lineage 1) | EAI   |
| R12     | IR                | 0000000000000371 | East-Asian (lineage 2) | Beijing |
| R13     | IR                | 0000000000000171 | East-Asian (lineage 2) | Beijing |
| R14     | IRE               | 00000000000000371 | East-Asian (lineage 2) | Beijing |
| R15     | SIREZ             | 00000000000000371 | East-Asian (lineage 2) | Beijing |
| R17     | SIRZOM            | 000000000000000000171 | East-Asian (lineage 2) | Beijing |
| R19     | IRE               | 41375730000000000071 | East-Asian (lineage 2) | Beijing |
| R1      | R                 | 637757633360731 | East-Asian (lineage 2) | Beijing |
| R20     | SIR               | 0000000000000000000241 | East-Asian (lineage 2) | Beijing |
| R21     | SIREZ             | 4007644400000051 | East-Asian (lineage 2) | Beijing |
| R23     | REZ               | 67777774341271771 | East-Asian (lineage 1) | EAI-2-Manila |
| R2      | IR                | 75677775411771 | East-Asian (lineage 1) | EAI6-BGD1 |
| R6      | IRE               | 0000000000000000000371 | East-Asian (lineage 2) | Beijing |
| R7      | IRZ               | 776000777413731 | East-Asian (lineage 1) | EAI1-SOM |
| S10     | None              | 177177377012631 | East-Asian (lineage 1) | EAI |
| S11     | None              | 6377554724111771 | East-Asian (lineage 1) | EAI-2-Manila |
| S13     | None              | 70375574000000371 | East-Asian (lineage 1) | EAI-2-Manila |
| S1      | None              | 677727743413671 | East-Asian (lineage 1) | EAI-2-Manila |
| S2      | None              | 672277743012671 | East-Asian (lineage 1) | EAI-2-Manila |
| S3      | None              | 0000000000000000000371 | East-Asian (lineage 2) | Beijing |
| S5      | None              | 67353757411771 | East-Asian (lineage 1) | EAI |
| S6      | None              | 755765753412461 | East-Asian (lineage 1) | EAI6-BGD1 |
| S7      | None              | 442767777013071 | East-Asian (lineage 1) | EAI3-IND |
| S8      | None              | 636351677777600 | *Mycobacterium bovis* | BOV |

S: Streptomycin, I: Isoniazid, R: Rifampicin, E: Ethambutol, Z: Pyrazinamide, O: Ofloxacin, M: Moxifloxacin, EAI: East African-Indian, CAS: Central Asia, SOM: Somalia, BOV: Bovis, T: T*
were predicted to be resistant compared to 12 using the phenotypic DST. For RIF, 13 of the isolates were predicted to be resistant compared to 14 when using phenotypic DST. As for EMB, 7 isolates were predicted to be resistant, whereas only 6 were found to be resistant by phenotypic DST. For STM, both next-generation sequencing (NGS) and phenotypic DST detected five of the isolates to be resistant. Overall, 26 single-nucleotide polymorphisms (SNPs) of the nine genes linked to drug-resistance traits were detected. However, allele \( \text{kasA} \ p.\text{Gly312Ser} \) was only detected by TB Profiler [Table 5].
Common variants of *Mycobacterium tuberculosis* complex isolates with different drug-resistance profiles

Comparative genome analysis of the genomes of the MTBC isolates against the *M. tuberculosis* H37Rv reference revealed 7576 variants common to all of the MTBC isolates with 4195 of them being nonsynonymous variants. The effects of the nonsynonymous variants can be classified into seven classes, as listed in Table 5. In addition, the study also identified a drug-resistant haplotype among the MTBC with drug-resistance phenotypes. The haplotype consists of 65 variants where they were found in 65 MTBC genes. The majority of the variants identified in the haplotype were involved in the PE-PRGS and PPE genes (24 genes) followed by those without any Clusters of Orthologous Groups classification (17 genes) and also hypothetical proteins (13 genes). The remaining common variants are found on the *EspB*, *GlnA*, *PstP*, *PpsD*, transcriptional regulator, and transmembrane protein genes, as summarized in Figure 2.

The 65 variants that make up the drug-resistant haplotype were also cross-checked against the Genome-based Mycobacterium tuberculosis Variants database, hosted by the Theodosius Dobzhansky Center for Genome Bioinformatics (St. Petersburg) for any unreported variant, which provides detailed information on *M. tuberculosis* genetic variations associated with phylogeographic distribution, drug resistance, and clinical outcome of TB.[23] Eleven of 65 variants were unique to this study, as listed in Table 6.

**DISCUSSION**

The motivation of the current study is to establish a preliminary knowledge base on the distribution of the types of MTBC isolates in Malaysia. Such knowledge base is important not only for epidemiological study for infection control purpose but also to assist in providing effective antibiotic therapy. Although the genotypic diversity of MTBC in Malaysia has been reported by Ismail *et al.* and by Dale *et al.* in 2014 and 1999, respectively, these studies used conventional approaches which resulted in low-throughput finding.[24,25] Moreover, it is worth mentioning that the combination of phenotypic and genotypic approaches as used in the present study could promote better insights into the drug-resistance profiles, informed therapeutic efforts, and clearer epidemiological picture of TB in Malaysia. For the purpose, this study involved WGS of 24 MTBC isolates which were isolated from patients diagnosed as TB in 2017. In addition, each isolate has also undergone *in vitro* DST assays to profile their resistance patterns against 10 drugs used in anti-TB therapy (STM, INH, RIF, EMB, PZA, CAP, AMK, KAN, OFX, and MXF).

**Figure 2:** Gene products of drug-resistance haplotype
Table 4: Identified variants associated with drug resistance in the genomes of the 24 clinical Mycobacterium tuberculosis complex isolates

| ID | Phe-DST  | Drug-resistance mutation(s) |
|----|----------|-----------------------------|
|    |          | I  | R     | E     | S     | Z     | F     |
| S1 | None     | -  | -     | -     | -     | -     | -     |
| S2 | None     | -  | -     | -     | -     | -     | -     |
| S3 | None     | -  | -     | -     | -     | -     | -     |
| S4 | None     | -  | -     | -     | -     | -     | -     |
| S5 | None     | -  | -     | -     | -     | -     | -     |
| S6 | None     | -  | -     | -     | -     | -     | -     |
| S7 | None     | -  | -     | -     | -     | -     | -     |
| S8 | None     | -  | -     | -     | -     | -     | -     |
| S9 | None     | -  | -     | -     | -     | -     | -     |
| S10| None     | -  | -     | -     | -     | -     | -     |
| S11| None     | -  | -     | -     | -     | -     | -     |
| S12| None     | -  | -     | -     | -     | -     | -     |
| R1*| R        | -  | -     | -     | -     | -     | -     |
| R2 | IR       | fabG1—15C>T<sup>TP</sup> | rpoB Asp435Val<sup>TP</sup> | -     | -     | -     | -     |
| R6*| IREZQ    | katG Asn138Asp<sup>TP</sup> | rpoB Ser450Leu<sup>TP</sup> | embB Gly406Asp<sup>TP</sup> | -     | -     | -     |
| R7*| IRZ      | katG Ser315Thr<sup>TP</sup> | rpoB Ser450Leu<sup>TP</sup> | embB Met306Val<sup>TP</sup> | rpsL Lys43Arg<sup>TP</sup> | pncA Leu85Pro<sup>TP</sup> | -     |
| R10*| IRSZ    | fabG1—8T>C<sup>TP</sup> | rpoB Ser445Arg<sup>TP</sup> | -     | -     | -     | -     |
| R12| IR       | katG Gln525Pro<sup>TP</sup> | rpoB Asp435Val<sup>TP</sup> | -     | -     | -     | -     |
| R13*| IR      | -  | -     | -     | -     | -     | -     |
| R14*| IREZ    | katG Ser315Thr<sup>TP</sup> | rpoB Ser450Leu<sup>TP</sup> | embB Gly406Cys<sup>TP</sup> | -     | pncA Val139Gly<sup>TP</sup> | -     |
| R15*| IRESZ   | katG Ser315Thr<sup>TP</sup> | rpoB Ser450Leu<sup>TP</sup> | embB Met306Val<sup>TP</sup> | rpsL Lys43Arg<sup>TP</sup> | -     | -     |
| R17*| IRSQ    | katG Ser315Thr<sup>TP</sup> | rpoB Ser450Leu<sup>TP</sup> | embB Met306Val<sup>TP</sup> | rpsL Lys43Arg<sup>TP</sup> | pncA Leu27Pro<sup>TP</sup> | gyrA Asp94Gly<sup>TP</sup> |
| R19| IRE      | katG Ser315Thr<sup>TP</sup> | rpoB His445Leu<sup>TP</sup> | embB Met306Val<sup>TP</sup> | -     | -     | -     |
| R20| IRS      | katG Ser315Thr<sup>TP</sup> | rpoB His445Leu<sup>TP</sup> | -     | rpsL Lys43Arg<sup>TP</sup> | -     | -     |
| R21| IRESZ    | katG Ser315Thr<sup>TP</sup> | rpoB His445Thr<sup>TP</sup> | embB Asp354Ala<sup>TP</sup> | rrs A906Gl<sup>TP</sup> | -     | -     |
| R23*| REZ     | -  | -     | -     | -     | -     | -     |

Remarks - *Isolate with discrepancy between Phe- and NGS-DST results. DST: Drug susceptibility test, Phe-DST: Phenotypic DST, T.P: Also found by TB Profiler and/or PhyResSE, respectively, -: No drug-resistance mutation was found, None: Susceptible to all six drugs, I: Isoniazid, R: Rifampicin, E: Ethambutol, S: Streptomycin, Z: Pyrazinamide, NGS: Next-generation sequencing, F: Fluoroquinolones

Table 5: Distribution of nonsynonymous variants found in all Mycobacterium tuberculosis complex isolates and the effect of each variant

| Number of variant | Effect                        | Description                                                                 |
|------------------|-------------------------------|-----------------------------------------------------------------------------|
| 11022            | Missense Variant              | Variant that causes a codon and amino acid change                           |
| 203              | Stop gained                   | Variant causes a stop codon                                                |
| 63               | Stop_lost and splice_region_variant | Variant causes a stop codon to be mutated into a nonstop codon            |
|                  |                               | A sequence variant in which a change has occurred within the region of the spicite, either within 1-3 bases of the exon or 3-8 bases of the intron |
| 8                | Splice_region_variant and stop_retained_variant | Variant causes a stop codon to be mutated into another stop codon               |
|                  |                               | A sequence variant in which a change has occurred within the region of the spicite, either within 1-3 bases of the exon or 3-8 bases of the intron |
| 16               | Start_lost                    | Variant causes a start codon to be mutated into a nonstart codon            |
| 10               | Initiator_codon_variant       | Variant causes a start codon to be mutated into another start codon       |

From the results of genome annotations using the RAST server, 23 of the 24 MTBC isolates were identified as strains of *M. tuberculosis*, whereas the remaining 1 isolate was identified as *M. bovis*. Both species of *Mycobacterium* were known to cause TB where *M. tuberculosis* is mostly associated with TB among humans, whereas *M. bovis* infection is more common among the bovines. However, zoonotic transmissions of *M. bovis* have been reported before, where transmission from cattle to humans was once common before the practice of routine pasteurization of cow milk. The symptoms of TB caused by these two *Mycobacterium* species almost resembled each other. In this study, *M. bovis* was isolated from a pus
sample of a 14-month-old baby. The common causes of *M. bovis* infection in humans are mostly from consumption of improperly pasteurized or contaminated dairy product. In certain regions of Malaysia, raw cow or goat milks are still sold to consumers, increasing the risk of getting *M. bovis* TB.

As for TB caused by *M. tuberculosis* infection, spoligotyping analysis done on the draft genomes revealed four lineages of *M. tuberculosis* – the East African-Indian (EAI), Beijing/East Asia (BEI), Delhi/Central Asia, and European/American strains. Such diversity is expected in Malaysia for its geographical feature which is the center of the Southeast Asian region and also its role as one of the main economic hubs of the region attracting immigrants from neighboring countries. Nonetheless, from the current study data, the EAI and BEI spoligotypes showed predominance over the other strains. This finding is consistent with a previous genotypic diversity study from Malaysia in 2014. Almost similar trends can also be seen in other neighboring countries around the Southeastern, such as Indonesia, Singapore, Thailand, Philippines, and Cambodia where genetic diversities of the MTBC are being driven by the EAI and Beijing lineages. There was no specific pattern of preference in terms of state, gender, age, and age groups, implying that most patients are randomly infected by any of the *M. tuberculosis* strains.

Phylogenetic analysis using the spoligotype data was also performed in this study. In the analysis, the 24 spoligotyped MTBC isolates were matched against 85 reference genomes of *Mycobacterium* spp. to construct a phylogenetic tree, as shown in Figure 1. In the figure, the 24 MTBC isolates were distributed into five groups as implied by their lineages. Interestingly, R7 and R20 isolates were found to have unexpected phylogenetic properties. The former, while grouped into the clade of Beijing lineage, was positioned on a branch that was completely separated from the rest of *M. tuberculosis* strains within the same node. From the length of the branch which houses the R7, it seemed like its evolutionary pathway has undergone less genetic mutations compared to the other isolates that reside on the same node of the phylogenetic tree. As for R20 isolate, phylogenetically, it resembled *Mycobacterium canetti* and *Mycobacterium microti*. Both are members of the MTBC which can cause TB in humans even though the cases are rare. *M. microti* infection is more common among voles and other mammals rather than humans. As for *M. canetti* or also known as “smooth tuberculosis bacilli,” its natural reservoir and route of transmission are still unknown. However, TB caused by *M. canetti* infections were peculiarly linked to patients with reported contacts to the Horn of Africa. Having such unique spoligotype may suggest that both R7 and R20 isolates are novel strains of *M. tuberculosis*, but additional analysis needs to be done to support the inference.

Another goal of this study was also to investigate the diversity of the drug-resistance patterns among the local MTBC isolates. Overall, 58.33% of the isolates (14/24) were found to be at least resistant to RIF. Moreover, 12 of the 14 isolates were MDR-Mtb and one of them was a poly-resistant strain. Overall, the MTBC isolates in this study showed resistant phenotypes toward seven of the ten tested drugs, in which the remaining three drugs were injectable aminoglycosides. The current study employed two approaches in detecting drug-resistance patterns among the MTBC isolates – the phenotypic DST and NGS-based DST. Taking that into consideration, a discrepancy of DST profiles between the two approaches was expected, where 11 were found in the DST testing using EMB and 4 in that of PZA. For DST using EMB, phenotypic DST reported that six MTBC isolates were EMB resistant, whereas NGS-based DST predicted that seven of the MTBC isolates were EMB resistant due to mutations that were associated with EMB resistance. To date, the performance of phenotypic DST for EMB is still plagued with unreliability issues. This is due to the characteristics of the drug which include bacteriostatic, reduced activity in *in vitro* testing, and minimum inhibitory concentration range which can barely differentiate between the susceptible and resistant isolates of *M. tuberculosis*. It is also worth mentioning that like EMB, phenotypic DST for PZA is also a very challenging task. PZA is a prodrug which requires low pH to be activated, but a low pH condition is still difficult to control in *in vitro* testing which may lead to false-negative result.

| Number | Position in genome | Locus tag | Gene name | Effect |
|--------|-------------------|-----------|-----------|--------|
| 1      | 212040            | Rv0180c   | -         | Missense variant c.211G>A p.Gly71Ser |
| 2      | 1864254           | Rv1651c   | PE_PGRS30 | Missense variant c.1129G>C p.Gly377Arg |
| 3      | 2115572           | -         | -         | No annotation data available |
| 4      | 2128794           | Rv1878    | glnA3     | Missense variant c.737T>C p.Leu258Pro |
| 5      | 2165123           | Rv1917c   | PPE34     | Missense variant c.2189G>C p.Gly730Ala |
| 6      | 2430017           | -         | -         | No annotation data available |
| 7      | 3136406           | Rv2828A   | -         | Missense variant c.194C>T p.Pro65Leu |
| 8      | 3551250           | -         | -         | No annotation data available |
| 9      | 3766873           | Rv3350    | PPE56     | Missense variant c.230G>A p.Gly77Asp |
| 10     | 4388761           | Rv3903c   | -         | Missense variant c.1672G>C p.Pro558Ser |
| 11     | 4388850           | Rv3903c   | -         | Missense variant c.1583G>C p.Pro528Leu |

Table 6: List of unreported variants which were common in the drug resistant-Myobacterium tuberculosis complex isolates in this study
Next, four discrepancies in PZA DST were contributed by the inability of NGS-based DST to detect mutations that conferred PZA resistant as observed in phenotypic DST. The R10, R15, R21, and R23 were phenotypically resistant to PZA, but no PZA-resistant-conferring mutations were detected in NGS-based DST. In *M. tuberculosis*, it is well known that mutations in *pncA* gene conferred resistant toward PZA. However, a clear hotspot region is yet to be identified where resistant-conferring mutation spread along the entire gene including the promoter region.[33] Therefore, it is possible that the library of resistant-conferring mutations used in TB Profiler and PhyRes software did not include the mutations that cause PZA-resistant in R10, R15, R21, and R23 isolates which lead to the discrepancies. The same explanation also goes to the discrepancies found in RIF, STM, and INH-DST results as both were caused by the inability of NGS-based DST to detect mutations that conferred resistant toward the drugs.

The sequencing of *M. tuberculosis* genomes allows genes that are significantly linked to drug resistance to be identified. However, the aspect of interaction between genes was often not taken into consideration when investigating the genetic factor of drug resistance among *M. tuberculosis*. In fact, there are evidence of the emergence of transmissible drug-resistant *M. tuberculosis* which resulted from multitude of additional mutations that interact between each other.[34] Consistent with the finding of the current study, one of the genes that may interact with genes associated with resistance to anti-TB drugs such as INH and RIF are the *pe* and *ppe* genes. In this study, eight *pe* and four *ppe* genes that contain mutations were common in all drug-resistant MTBC isolates where mutations in PE-PGRS 30, PPE 34, and PPE 56 genes were not previously reported. In the genetics of mycobacteriology, *pe* and *ppe* genes were found mostly in slow-growing pathogenic mycobacteria which were thought to influence the function and immunopathogenicity of *M. tuberculosis*. [35]

Even though the two gene families were more linked to the bacterial pathogenesis than drug resistance, certain mutations in *ppe* genes were found to be in paired with mutations in genes that involved in anti-TB drug activities such as katG, rpoB, embA and embB genes in a study involving the genomes of 208 drug-resistant *M. tuberculosis*. [34] Thus, mutations in *ppe* genes may also relate to drug resistance in *M. tuberculosis*. In this study, as all drug-resistant *M. tuberculosis* isolates were at least resistant to RIF, and since the *rpoB* mutation was detected, the *rpoB*-PPE34, *rpoB*-PPE54, *rpoB*-PPE56, and *rpoB*-PPE66 gene pairs could play possible roles in the resistance toward RIF.

**Conclusions**

This study reports the diversity of MTBC lineages and mutations found within the genomes of 24 clinical MTBC isolates with varying drug-resistance profiles. The study was the first of its kind in Malaysia that provided high-throughput analysis on multiple strains of MTBC at rapid pace thanks to the application of WGS. Thus, WGS is an indispensable tool for future epidemiological and drug-resistance study on MTBC. Especially in Malaysia where TB is still a major problem, the need for a high-throughput investigation approach is crucial to gain insights into the mechanistic and epidemiologic aspects of TB.

**Acknowledgement**

The study was approved by the Director General of Health, Malaysia for this publication. Study ethics was approved by the Universiti Teknologi MARA (UiTM) Research Ethics Committee and the National Medical Research Registration (NMRR) of Ministry of Health, Malaysia (Approval No. NMRR-18-833-40380, dated 21/02/2019).

**Financial support and sponsorship**

The project is funded by the Ministry of Higher Education Malaysia (MOE) (Grant no. 600-RMI/LGCS 5/3 (3/2016) and Integrative Pharmacogenomics Institute (iPROMISE) internal grant (Fund ID. 241910/2017/MZS/10).

**Conflicts of interest**

There are no conflicts of interest.

**References**

1. World Health Organization. WHO Global Report, Global Tuberculosis Report 2018. Geneva: World Health Organization; 2018.

2. World Health Organization. Companion Handbook to the WHO Guidelines for the Programmatic Management of Drug-Resistant Tuberculosis. World Health Organization; 2014.

3. Ministry of Health (MOH). Health Facts 2017. MOH/S/RAN/47.17(AR). Malaysia: Planning Informatics Centre; 2017.

4. Lee RS, Behr MA. The implications of whole-genome sequencing in the control of tuberculosis. Ther Adv Infect Dis 2016;3:47-62.

5. Pankhurst LJ, Del Ojo Elias C, Votintseva AA, Walker TM, Cole K, Davies J, et al. Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: A prospective study. Lancet Respir Med 2016;4:49-58.

6. Philip N, Rodrigues KE, William T, John DV. Whole genome sequencing of *Mycobacterium tuberculosis* SB24 isolated from Sabah, Malaysia. Genom Data 2016;9:137-9.

7. Issa R, Seradja VH, Abdullah MK, Abdul H. Annotated sequence of *Mycobacterium tuberculosis* MTBR3/09 isolated from a sputum sample in Malaysia. Genome Announce 2016;4. pii: e00517-16.

8. Halim MZ, Jaafar MM, Teh LK, Ismail MI, Lee LS, Ngeow YF, et al. Genome sequencing and annotation of multidrug resistant *Mycobacterium tuberculosis* (MDR-TB) PR10 strain. Genom Data 2016;7:245-6.

9. Kuan CS, Chan CL, Yew SM, Toh YF, Khoo JS, Chong J, et al. Genome analysis of the first extensively drug-resistant (XDR) *Mycobacterium tuberculosis* in Malaysia provides insights into the genetic basis of its biology and drug resistance. PLoS One 2015;10:e0131694.

10. Canetti G, Froman S, Grosset J, Hauduroy P, Langerova M, Mahler HT, et al. *Mycobacteria*: laboratory methods for testing drug sensitivity and resistance. Bull World Health Organ 1963;29:565-78.

11. World Health Organization. Treatment of Tuberculosis: Guidelines for National Programmes. 3rd ed. Geneva: World Health Organization; 2003.

12. World Health Organization. Technical Manual for Drug Susceptibility Testing of Medicines Used in the Treatment of Tuberculosis. World Health Organization; 2018.

13. Andrews S. Fast QC: A quality control tool for high throughput sequence data. Version 0.11.3; 2010. Available from: http://www.bioinformatics. babraham.ac.uk/?projects/fastqc/. [Last accessed on 2019 Apr 25].

14. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. Eur Mosq Bull Net J 2011;17:10-2.
15. Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 2008;18:821-9.

16. Gladman S, Seemann T. VelvetOptimiser: automate Your Velvet Assemblies. Version 2.2.6; 2008. Available from: https://github.com/tseemann/VelvetOptimiser. [Last accessed on 2019 April 25].

17. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. Genome Res 2004;14:1394-403.

18. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res 2014;42:D206-14.

19. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. Genome Res 2004;14:1394-403.

20. Coll F, McNerney R, Preston MD, Guerra-Assunção JA, Warry A, Hill-Cawthorne G, et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. Genome Med 2015;7:51.

21. Feuerriegel S, Schleusener V, Beckert P, Kohl TA, Miotto P, Cirillo DM, et al. PhyResSE: A Web tool delineating Mycobacterium tuberculosis antibiotic resistance and lineage from whole-genome sequencing data. J Clin Microbiol 2015;53:1908-14.

22. Coll F, McNerney R, Preston MD, Guerra-Assunção JA, Warry A, Hill-Cawthorne G, et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. Genome Med 2015;7:51.

23. Chernyaeva EN, Shulgina MV, Rotkevich MS, Dobrynin PV, Simonov SA, Shitikov EA, et al. Genome-wide Mycobacterium tuberculosis variation (GMTV) database: A new tool for integrating sequence variations and epidemiology. BMC Genomics 2014;15:308.

24. Ismail F, Couvin D, Farakhin I, Abdul Rahman Z, Rastogi N, Suraiya S, et al. Study of Mycobacterium tuberculosis complex genotypic diversity in Malaysia reveals a predominance of ancestral East-African-Indian lineage with a Malaysia-specific signature. PLoS One 2014;9:e114832.

25. Dale JW, Nor RM, Ramayah S, Tang TH, Zainuddin ZF. Molecular epidemiology of tuberculosis in Malaysia. J Clin Microbiol 1999;37:1265-8.

26. Zhang J, Heng S, Le Moulec S, Refregier G, Gicquel B, Sola C, et al. A first assessment of the genetic diversity of Mycobacterium tuberculosis complex in Cambodia. BMC Infect Dis 2011;11:42.

27. Lim LK, Sng LH, Win W, Chee CB, Hsu LY, Mak E, et al. Molecular epidemiology of Mycobacterium tuberculosis complex in Singapore, 2006-2012. PLoS One 2013;8:e84487.

28. Disratthakit A, Meada S, Prammananan T, Thaipsuttikul I, Doi N, Chaiprasert A. Genotypic diversity of multidrug-, quinolone- and extensively drug-resistant Mycobacterium tuberculosis isolates in Thailand. Infect Genet Evol 2015;32:432-9.

29. Lisdawati V, Puspandari N, Rif’ati L, Soekarno T, Melatiwati, M, Syamsidar, K, et al. Molecular epidemiology study of Mycobacterium tuberculosis and its susceptibility to anti-tuberculosis drugs in Indonesia. BMC Infect Dis 2015;15:366.

30. Roa MB, Tablizo FA, Morado EK, Cunanan LF, Uy ID, Ng KC. Whole-genome sequencing and single nucleotide polymorphisms in multidrug-resistant clinical isolates of Mycobacterium tuberculosis from the Philippines. J Glob Antimicrob Resist 2018;15:239-45.

31. Bozid F, Brégeon F, Poinçon L, Weber P, Drancourt M, Canaan S, et al. Mycobacterium canetti infection of adipose tissues. Front Cell Infect Microbiol 2017;7:189.

32. Madison B, Robinson-Dunn B, George I, Gross W, Lipman H, Metchock B, et al. Multicenter evaluation of ethambutol susceptibility testing of Mycobacterium tuberculosis by agar proportion and radiometric methods. J Clin Microbiol 2002;40:3976-9.

33. Cirillo DM, Miotto P, Tortoli E. Evolution of phenotypic and molecular drug susceptibility testing. Adv Exp Med Biol 2017;1019:221-46.

34. Cui ZJ, Yang QY, Zhang HY, Zhu Q, Zhang QY. Bioinformatics identification of drug resistance-associated gene pairs in Mycobacterium tuberculosis. Int J Mol Sci 2016;17. pii: E1417.

35. Brennan MJ. The enigmatic PE/PPE multigene family of mycobacteria and tuberculosis vaccination. Infect Immun 2017;85. pii: e00969-16.