Draft Genome Sequences of Local Clinical Isolates of Drug-Resistant and Drug-Sensitive Mycobacterium tuberculosis

Norzuliana Zainal Abidin,a Mohd Nur Fakhruzzaman Noorizhab,a,b Lay Kek Teh,a,b Wai Feng Lim,a Noorliza Mohd Noordin,c Zirwatul Adilah Aziz,c Zamzurina Abu Bakar,d Ahmad Izuauuddin Ismail,e Norazmi Mohd Nor,f Mohd Zaki Salleh,a

aIntegrative Pharmacogenomics Institute, Universiti Teknologi MARA Selangor Branch, Selangor, Malaysia
bFaculty of Pharmacy, Universiti Teknologi MARA Selangor Branch, Selangor, Malaysia
cNational Public Health Laboratory, Selangor, Malaysia
dInstitute of Respiratory Medicine, Ministry of Health, Kuala Lumpur, Malaysia
eFaculty of Medicine, Universiti Teknologi MARA Sungai Buloh Branch, Selangor, Malaysia
fSchool of Health Sciences, Health Campus, Universiti Sains Malaysia, Kelantan, Malaysia

ABSTRACT In the battle against tuberculosis (TB), plasticity of the Mycobacterium tuberculosis genome is believed to contribute to the pathogen’s virulence and drug resistance. Here, we report 10 draft genome sequences of clinical M. tuberculosis isolated in Malaysia as the basis for understanding the genome plasticity of the M. tuberculosis isolates.

Mycobacterium tuberculosis is a member of the family Mycobacteriaceae and is the causative agent for tuberculosis (TB). Challenges in TB treatment and prevention to curb the spread of TB became complicated due to drug-resistant TB (DR-TB). According to the 2019 Global TB report released by the World Health Organization (WHO), the prevalence rate of TB in Malaysia is at 92 per 100,000 people, with an estimation of 1.2% for multidrug resistant/rifampicin-resistant TB (MDR/RR-TB) (1).

Here, we report five drug-resistant and five drug-susceptible clinical M. tuberculosis isolates collected from pulmonary TB patients at the local hospital in Malaysia before they started on anti-TB treatment. Sputum specimens were collected by clinicians at the respective hospitals and sent to the National Public Health Laboratory (NPHL) for culture and isolation of pure M. tuberculosis; isolates were subsequently tested for their sensitivity and resistance toward anti-TB drugs. Pure M. tuberculosis isolates were cultured on Lowenstein-Jensen medium (L-J) and incubated at 37°C for 3 to 5 weeks. The phenotypic drug susceptibility testing (Phe-DST) was done using the absolute concentration method or proportion method (Bactec MGIT 960 TB detection system) according to the standard protocols (Becton, Dickinson, Sparks, MD, USA) (2–4). Each M. tuberculosis isolate was first tested for susceptibility against the first-line anti-TB drugs. Then, isolates that were resistant toward isoniazid and rifampicin were further tested against the second-line anti-TB drugs. However, isolates which were sensitive toward first-line anti-TB drugs were not subjected to the Phe-DST of the second-line anti-TB drugs. The isolates were heat deactivated and sent to iPROMISE UiTM for whole-genome sequencing.

The genomic DNA of each M. tuberculosis isolate was extracted using the conventional chloroform extraction method (5). According to the manufacturer’s instructions, DNA libraries of each M. tuberculosis isolate were prepared using the Nextera DNA Flex library preparation kit (Illumina, Inc., San Diego, CA) and sequenced using a next-generation sequencer (Illumina, Inc., San Diego, CA). Adaptor sequences and low-quality reads of the paired-end read sequences were trimmed using the Cutadapt tool version 3.4 with Phyton version 2.7.12 (6). The quality of the trimmed reads was assessed using the FastQC software version 0.11.3 with the default parameters (Abraham Bioinformatics, Cambridge, UK).
### TABLE 1 Details of the assembled *M. tuberculosis* draft genome sequences

<table>
<thead>
<tr>
<th>Isolate*</th>
<th>Accession no. by database</th>
<th>Total no. of reads</th>
<th>Length (bp [%])</th>
<th>No. of scaffolds</th>
<th>N$_{50}$ value* (bp)</th>
<th>G+C (%)</th>
<th>Cov.×</th>
<th>Total no. of variants</th>
<th>Lineage*</th>
<th>Octal code of <em>M. tuberculosis</em> spoligotype patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>MR0031</td>
<td>JADPQQ00000000000 SRR14270082 SAMN16624767</td>
<td>2,662,964</td>
<td>4,392,022 (99.65)</td>
<td>2</td>
<td>4,389,866</td>
<td>65</td>
<td>79</td>
<td>2,067</td>
<td>L1</td>
<td>774377777413011</td>
</tr>
<tr>
<td>MR0021</td>
<td>JADPQR00000000000.1 SRR14270083 SAMN16624766</td>
<td>2,792,316</td>
<td>4,375,917 (99.71)</td>
<td>4</td>
<td>4,373,967</td>
<td>65</td>
<td>84</td>
<td>1,443</td>
<td>L2</td>
<td>000000000003771</td>
</tr>
<tr>
<td>MR2407</td>
<td>JADPQP00000000000 SRR14270081 SAMN16624768</td>
<td>2,787,620</td>
<td>4,380,639 (99.73)</td>
<td>6</td>
<td>4,377,722</td>
<td>65</td>
<td>83</td>
<td>1,425</td>
<td>L2</td>
<td>000000000003771</td>
</tr>
<tr>
<td>MR2972</td>
<td>JADPQQ00000000000 SRR14270080 SAMN16624769</td>
<td>2,832,572</td>
<td>4,391,115 (99.68)</td>
<td>9</td>
<td>4,384,447</td>
<td>65</td>
<td>85</td>
<td>2,123</td>
<td>L2</td>
<td>700000007403771</td>
</tr>
<tr>
<td>MR1481</td>
<td>JADPQN00000000000 SRR14270079 SAMN16624770</td>
<td>2,611,384</td>
<td>4,365,058 (99.67)</td>
<td>9</td>
<td>4,362,970</td>
<td>65</td>
<td>78</td>
<td>1,353</td>
<td>L2</td>
<td>000000000003771</td>
</tr>
<tr>
<td>MS2854</td>
<td>JADPQM00000000000 SRR14270078 SAMN16624771</td>
<td>2,685,229</td>
<td>4,370,561 (99.70)</td>
<td>5</td>
<td>4,369,052</td>
<td>65</td>
<td>80</td>
<td>1,402</td>
<td>L2</td>
<td>000000000003771</td>
</tr>
<tr>
<td>MS2947</td>
<td>JADPQL00000000000 SRR14270077 SAMN16624772</td>
<td>2,736,312</td>
<td>4,401,789 (99.67)</td>
<td>2</td>
<td>4,392,317</td>
<td>65</td>
<td>81</td>
<td>2,089</td>
<td>L3</td>
<td>777777750003331</td>
</tr>
<tr>
<td>MS3002</td>
<td>JADPQK00000000000 SRR14270076 SAMN16624773</td>
<td>2,696,572</td>
<td>4,366,050 (99.69)</td>
<td>7</td>
<td>4,364,323</td>
<td>65</td>
<td>81</td>
<td>1,383</td>
<td>L2</td>
<td>000000000003771</td>
</tr>
<tr>
<td>MS3207</td>
<td>JADPQJ00000000000 SRR14270075 SAMN16624774</td>
<td>2,800,646</td>
<td>4,376,767 (99.68)</td>
<td>4</td>
<td>4,373,296</td>
<td>65</td>
<td>84</td>
<td>819</td>
<td>L2</td>
<td>000000000003771</td>
</tr>
<tr>
<td>MS3244</td>
<td>JADPQI00000000000 SRR14270074 SAMN16624775</td>
<td>2,750,191</td>
<td>4,361,988 (99.70)</td>
<td>5</td>
<td>4,359,624</td>
<td>65</td>
<td>83</td>
<td>1,397</td>
<td>L2</td>
<td>000000000003771</td>
</tr>
</tbody>
</table>

*a* MS, isolates sensitive toward the first-line anti-TB drugs; MR, isolates resistant toward any of the anti-TB drugs tested.

*b* Percentage of identity when mapped against the reference genome calculated using.

*c* N$_{50}$ value, sequence length of the shortest contigs at 50% of the total genome length.

*d* Cov., depth of coverage.

*e* L1, lineage 1, East African-Indian strain; L2, lineage 2, Beijing strain; L3, lineage 3, Central Asian/Delhi strain.
average, more than 80% of the trimmed reads have Phred scores of $>-30$, with an average coverage depth of $81 \times$ and a G+C content of 65%.

An average of more than 99% genome completion was achieved from the de novo assembly of the *M. tuberculosis* genome sequences done using the VelvetOptimizer tool version 1.2.10 (7), and the contigs were scaffolded using the multidraft-based scaffolder (MeDuSa) version 1.6 that mapped the query sequences against reference genome *M. tuberculosis* H37Rv strain (accession number NC_018143) (8). Variants of the *M. tuberculosis* genomes were called using the Snippy tool released under the GPL (version 2) (https://github.com/tseemann/snippy). Table 1 shows the details of the assembled *M. tuberculosis* draft genome sequences and the total number of variants called per genome. In this study, default parameters were used for all the online Web-based tools. *M. tuberculosis* isolates were classified into different *M. tuberculosis* lineages based on the octal code of the *M. tuberculosis* spoligotype patterns using the SpoTyping tool version 2.0 (9) (Table 1).

**Data availability.** Draft genome sequences of the local clinical *M. tuberculosis* isolates were deposited in the National Center of Biotechnology Information (NCBI) with BioProject number PRJNA674940. The accession numbers of each *M. tuberculosis* complex (MTBC) genome are included in Table 1.

**ACKNOWLEDGMENTS**

The study was approved by the Director General of Health, Malaysia, for publication. Ethical approvals were obtained from the Universiti Teknologi MARA (UiTM) Research Ethics Committee (approval no. REC/271/17) and the National Medical Research Registration (NMRR) of Ministry of Health, Malaysia (approval no. NMRR-18-833-40380). This project was funded by the Long-Term Research Grant Scheme (LRGS), Ministry of Higher Education Malaysia (MOHE) [grant no. 600-RMI/LRGS 5/3 (3/2016)].

**REFERENCES**