

# Complete Genome Sequence of *Mycobacterium tuberculosis* Clinical Isolate Spoligotype SIT745/EAI1-MYS

S. Suraiya,<sup>a</sup> N. Semail,<sup>a</sup> M. F. Ismail,<sup>a</sup> J. M. Abdullah<sup>b</sup>

Medical Microbiology Department, School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian, Kelantan, Malaysia<sup>a</sup>; Center for Neuroscience Services and Research (P3Neuro) Health Campus, Universiti Sains Malaysia, Kubang Kerian, Kelantan, Malaysia<sup>b</sup>

All authors contributed equally to this work.

*Mycobacterium tuberculosis* is known to cause pulmonary and extrapulmonary tuberculosis. This organism showed special phylogeographical specificity. Here, we report the complete genome sequence of *M. tuberculosis* clinical isolate spoligotype SIT745/EAI1-MYS, which was isolated from a Malaysian tuberculosis patient.

Received 9 March 2016 Accepted 8 April 2016 Published 19 May 2016

**Citation** Suraiya S, Semail N, Ismail MF, Abdullah JM. 2016. Complete genome sequence of *Mycobacterium tuberculosis* clinical isolate spoligotype SIT745/EAI1-MYS. *Genome Announc* 4(3):e00323-16 doi:10.1128/genomeA.00323-16.

**Copyright** © 2016 Suraiya et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to S. Suraiya, [ssuraiya@usm.my](mailto:ssuraiya@usm.my).

Tuberculosis (TB) remains a major global public health problem. TB is an ancient infectious disease caused by *Mycobacterium tuberculosis*. It ranks as the second leading cause of death from an infectious disease worldwide after human immunodeficiency virus (HIV) (1).

Recently, spoligotyping of *M. tuberculosis* isolated in Malaysia by our research group (2) highlighted that the East-African-Indian (EAI) lineage is the most prevalent lineage in Malaysia. Deeper analysis revealed that there is a phylogeographical specificity of SIT745/EAI1-MYS for Malaysia, and there is probable ongoing evolution with locally evolved strains sharing a specific signature characterized by the absence of spacers 37, 38, and 40 in the spoligotyping results.

*M. tuberculosis* strain SIT745/EAI1-MYS was isolated from a tuberculosis patient who presented to Hospital Universiti Sains Malaysia (HUSM), a tertiary teaching hospital located in the northeastern region of Malaysia.

Genomic DNA was extracted from a 3-week-old culture by a genomic DNA extraction kit (Qiagen, Hilden, Germany). The purified genomic DNA was subjected to whole-genome shotgun sequencing on an Illumina MiSeq (Illumina, Inc., USA) platform with 2 × 101-bp read length. The raw reads were trimmed using Trimmomatic (3) and Scythe (<https://github.com/vsbuffalo/scythe>). SGA (4) was used for error correction. The trimmed reads were subjected to *de novo* assembly with IDBA-UD 1.0.9 (5). Gene prediction was performed with the prokaryotic gene prediction algorithm Prodigal (version 2.60) (6), while rRNAs were predicted with RNAmmer (7). tRNA prediction was performed using Aragorn (8). Subsequently, the genome sequence was annotated with BLASTn (9, 10) against the Swiss-Prot database.

The *M. tuberculosis* SIT745/EAI1-MYS genome, consisting of 4,371,919 bases, was obtained at 194× coverage, and its G+C content is 65.5%. The total number of protein-coding genes is 4,105.

The sequenced genome of *M. tuberculosis* SIT745/EAI1-

MYS serves as a foundation through which we can better understand the *M. tuberculosis* species that has a specific occurrence in Malaysia.

**Nucleotide sequence accession numbers.** This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession no. [LUDZ000000000](https://www.ncbi.nlm.nih.gov/nuclseq/LUDZ000000000). The version described in this paper is version LUDZ01000000.

## ACKNOWLEDGMENT

This work was supported by Research University grant RUT 1001/PPSP/853001 from Universiti Sains Malaysia.

## FUNDING INFORMATION

This work, including the efforts of Siti Suraiya, was funded by Universiti Sains Malaysia (USM) (RUT 1001/PPSP/853001).

## REFERENCES

1. WHO. 2013. Global tuberculosis report 2013. World Health Organization, Geneva, Switzerland. [http://apps.who.int/iris/bitstream/10665/91355/1/9789241564656\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/91355/1/9789241564656_eng.pdf).
2. Ismail F, Couvin D, Farakhin I, Abdul Rahman Z, Rastogi N, Suraiya S. 2014. 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* 9:e114832. <http://dx.doi.org/10.1371/journal.pone.0114382>.
3. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <http://dx.doi.org/10.1093/bioinformatics/btu170>.
4. Simpson JT, Durbin R. 2010. Efficient construction of an assembly string graph using the FM-index. *Bioinformatics* 26:i367–i373. <http://dx.doi.org/10.1093/bioinformatics/btq217>.
5. Peng Y, Leung HC, Yiu SM, Chin FY. 2012. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28:1420–1428. <http://dx.doi.org/10.1093/bioinformatics/bts174>.
6. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <http://dx.doi.org/10.1186/1471-2105-11-119>.
7. Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery

- DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35:3100–3108. <http://dx.doi.org/10.1093/nar/gkm160>.
8. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 32:11–16. <http://dx.doi.org/10.1093/nar/gkh152>.
9. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [http://dx.doi.org/10.1016/S0022-2836\(05\)80360-2](http://dx.doi.org/10.1016/S0022-2836(05)80360-2).
10. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <http://dx.doi.org/10.1093/nar/gkh340>.