Whole-Genome Sequencing of a *Mycobacterium tuberculosis* Strain Belonging to Lineage 1 (Indo-Oceanic) and the East African Indian Spoligotype, Isolated in Jazan, Saudi Arabia

[ID] Aymen Abdelhaleem,[a] Almonther Hershan,[b] Pradeep Agarwal,[c] Abdullah Farasani,[a,d] Shaheed V. Omar,[e,f] [ID] Arshad Ismail,[e]
[ID] Mushal Allam[e]

[a]Medical Research Center, Jazan University, Jazan, Saudi Arabia
[b]Department of Medical Microbiology and Parasitology, College of Medicine, Jeddah University, Jeddah, Saudi Arabia
[c]Jazan Chest Diseases Hospital, Abu Arish, Saudi Arabia
[d]Department of Medial Laboratory Technology, Applied Medical College, Jazan University, Jazan, Saudi Arabia
[e]National Institute for Communicable Diseases, National Health Laboratory Service, Johannesburg, South Africa
[f]Department of Molecular Medicine and Haematology, School of Pathology, Faculty of Health Sciences, University of Witwatersrand, Johannesburg, South Africa

ABSTRACT We describe here the draft genome sequence of AY1MRC, a *Mycobacterium tuberculosis* strain belonging to lineage 1 (Indo-Oceanic) and the East African Indian spoligotype, isolated from a patient with tuberculosis in Jazan, Saudi Arabia.

Tuberculosis (TB), for which the etiological agent is the *Mycobacterium tuberculosis* complex (MTBC), has been considered a global public health emergency since 1931 (1). According to the World Health Organization (WHO), Saudi Arabia has a moderate TB infection rate (2). However, studies have indicated the presence of multidrug-resistant (MDR) and extensively drug-resistant (XDR) TB strains belonging to different lineages circulating in Saudi Arabia in the past 2 decades (3–6). Here, we report the genome sequence of an MTBC strain belonging to the East African Indian spoligotype in lineage 1 (Indo-Oceanic), isolated from a pulmonary TB patient in Jazan, southwest Saudi Arabia.

Mycobacterium tuberculosis strain AY1MRC was isolated from a sputum sample obtained from a 23-year-old male suspected of having pulmonary TB in April 2016. The presence of the bacterium was confirmed through direct smear molecular techniques focused on IS*6110* insertion and *Hsp65* genetic targets, the GeneXpert MTB/RIF assay, and a culture isolated using Lowenstein-Jensen (LJ) medium (7). A single-colony growth of *M. tuberculosis* complex observed on the LJ medium was used for DNA extraction and purposed for whole-genome sequencing. The genomic DNA was extracted using a Qiagen minikit (Germany) according to the manufacturer's instructions. The integrity of the extracted DNA was examined by the Genova Plus spectrophotometer (Bibby Scientific, USA) and validated by agarose gel electrophoresis. A paired-end library (2 × 150 bp) was prepared using a TruSeq Nano DNA kit (Illumina, USA) and sequenced on a NovaSeq 6000 instrument (Illumina). The sequencing of AY1MRC produced 22,777,886 raw reads, which were quality trimmed using fastp v0.20.1 (quality [Q], 20 on 90% of bases on each read; read length, 15 bases) (8) and Trimmomatic v0.36 (Q, 20 on 90% of bases on each read; read length, 15 bases) (9), *de novo* assembled using SPAdes v3.13.0 (10) with the parameter "only-error-correction," and evaluated using QUAST v5.0.2 (11). The assembled genome yielded 43 contiguous sequences longer than 1,000 bp covering 4,386,080 bp, with a G+C content of 65.5%, an N_{50} value of 215,026 bp, and a maximum contig length of 395,194 bp. The assembled genome was deposited in GenBank and annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (12). Among the 4,144 genes predicted by PGAP, 3,963 were protein-

coding genes, 130 were pseudogenes, and 51 were RNAs (3 rRNAs [5S, 16S, and 23S], 45 tRNAs, and 3 noncoding RNAs [ncRNAs]).

The sequencing data were further uploaded to the TB-Profiler webserver (http://tbdr.lshtm.ac.uk/) (13, 14) to predict lineage and drug resistance. AY1MRC was predicted to belong to the Indo-Oceanic lineage (lineage 1.1.2) and wild-type spoligotypes East Africa India 3 and East Africa India 5. Susceptibility to rifampin, isoniazid, aminoglycosides, fluoroquinolones, bedaquiline, delamanid, ethambutol, ethionamide, streptomycin, pyrazinamide, and linezolid was inferred from the genomic data. All software and tools were run with default parameters, unless otherwise indicated.

Ethical clearance was obtained from the Jazan University Standing Committee for Biomedical Research Ethics with approval number 2198/60.

Data availability. This whole-genome sequence project has been deposited in DDBJ/ENA/GenBank with the BioProject number PRJNA587526 and BioSample number SAMN12785870 under the accession number WMCK00000000. The described version is WMCK00000000.1. The raw reads have been submitted under the SRA accession number SRR12001309.

REFERENCES

1. World Health Organization. 2013. Using the Xpert MTB/RIF assay to detect pulmonary and extrapulmonary tuberculosis and rifampicin resistance in adults and children: expert group meeting report: 2013. World Health Organization, Geneva, Switzerland. https://apps.who.int/iris/handle/10665/112659.
2. World Health Organization. 2018. Global tuberculosis report. World Health Organization, Geneva, Switzerland.
3. Varghese B, Enani M, Alrajhi A, Al Johani S, Albarak A, Althawadi S, Elkhizzi N, AlGhafli H, Shoukri M, Al-Hajoj S. 2018. Impact of Mycobacterium tuberculosis complex lineages as a determinant of disease phenotypes from an immigrant rich moderate tuberculosis burden country. Respir Res 19:259. https://doi.org/10.1186/s12931-018-0966-x.
4. Al-Orainey I, Alhedaithy MA, Alanazi AR, Barry MA, Almajid FM. 2013. Tuberculosis incidence trends in Saudi Arabia over 20 years: 1991-2010. Ann Thorac Med 8:148–152. https://doi.org/10.4103/1817-1737.114303.
5. Al-Hajoj SAM, Zozio T, Al-Rabiah F, Mohammad V, Al-Nasser M, Sola C, Rastogi N. 2007. First insight into the population structure of Mycobacterium tuberculosis in Saudi Arabia. J Clin Microbiol 45:2467–2473. https://doi.org/10.1128/JCM.02293-06.
6. Al-Ghafli H, Kohl TA, Merker M, Varghese B, Halees A, Niemann S, Al-Hajoj S. 2018. Drug-resistance profiling and transmission dynamics of multidrug-resistant Mycobacterium tuberculosis in Saudi Arabia revealed by whole genome sequencing. Infect Drug Resist 11:2219–2229. https://doi.org/10.2147/IDR.S181124.
7. Abdelhaleem AA, Hershan AA, Agarwal PK. 2017. Diagnostic accuracy of IS6110 insertion gene, Hsp65, and Xpert MTB/RIF for rapid diagnosis of pulmonary tuberculosis. J Tuberc Res 5:1–12. https://doi.org/10.4236/jtr.2017.51001.
8. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
9. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
10. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
11. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
12. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.
13. Coll F, McNerney R, Preston MD, Guerra-Assunção JA, Warry A, Hill-Cawthorne G, Mallard K, Nair M, Miranda A, Alves A, Perdigão J, Viveiros M, Portugal I, Hasan Z, Hasan R, Glynn JR, Martin N, Pain A, Clark TG. 2015. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. Genome Med 7:51. https://doi.org/10.1186/s13073-015-0164-0.
14. Phelan JE, O'Sullivan DM, Machado D, Ramos J, Oppong YEA, Campino S, O'Grady J, McNerney R, Hibberd ML, Viveiros M, Huggett JF, Clark TG. 2019. Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs. Genome Med 11:41. https://doi.org/10.1186/s13073-019-0650-x.