



Genome Sequencing of Polydrug-, Multidrug-, and Extensively Drug-Resistant *Mycobacterium tuberculosis* Strains from South India

Sivakumar Shanmugam,^a Narender Kumar,^b Dina Nair,^a Mohan Natrajan,^a Srikanth Prasad Tripathy,^a Sharon J. Peacock,^{b,d} Soumya Swaminathan,^c Uma Devi Ranganathan^a

^aICMR–National Institute for Research in Tuberculosis, Chennai, India

^bDepartment of Clinical Medicine, University of Cambridge, Cambridge, United Kingdom

^cWorld Health Organization, Geneva, Switzerland

^dLondon School of Hygiene and Tropical Medicine, London, United Kingdom

ABSTRACT The genomes of 16 clinical *Mycobacterium tuberculosis* isolates were subjected to whole-genome sequencing to identify mutations related to resistance to one or more anti-*Mycobacterium* drugs. The sequence data will help in understanding the genomic characteristics of *M. tuberculosis* isolates and their resistance mutations prevalent in South India.

Tuberculosis (TB) is the leading cause of death due to a single infectious agent worldwide. As per a World Health Organization (WHO) report, an estimated 10 million people developed TB in 2017 (1). The emergence of strains of multidrug-resistant tuberculosis (MDR-TB) that are resistant to the most effective drugs (rifampin and isoniazid) poses a threat to TB control programs. These constitute 82% of the 558,000 rifampin-resistant cases reported in 2017 worldwide. Among the MDR-TB cases, 8.5% are identified as extensively drug-resistant TB (XDR-TB) and have additional resistance to one of the fluoroquinolones and aminoglycosides. India alone contributed to 2,650 notified cases of XDR-TB in 2017 (1, 2). Early and accurate diagnosis of MDR and XDR-TB followed by appropriate treatment is of critical importance if the number of new cases is to be reduced worldwide. Here, we describe the use of whole-genome sequencing (WGS) technology to define the genetic mutations in isolates from a part of the world that lacks such data and which could contribute to the genetic prediction of a tailored treatment regimen for TB patients in the future.

We report the whole-genome sequence of 16 *Mycobacterium tuberculosis* isolates from patients in Chennai, India. These have a range of drug resistance patterns that categorize them as MDR, polydrug-resistant (PDR), or XDR. In order to obtain the mycobacterial DNA for WGS, *M. tuberculosis* strains were isolated from sputum samples of pulmonary tuberculosis patients treated at NIRT, Chennai and grown on Lowenstein-Jensen (LJ) medium, and colonies were scraped from the 3+ growth on the medium into Tris-ethylenediaminetetraacetic acid (TE) buffer (3–5) followed by genomic DNA extraction with a cetyltrimethylammonium bromide-sodium chloride (CTAB–NaCl) extraction method (6). Drug susceptibility of the *M. tuberculosis* isolates was tested with an automated MGIT 960 instrument as per the manufacturer's protocol (7, 8). DNA purity and concentration were determined with Nanodrop (9) and Qubit 3.0 fluorometers, respectively (10). Library preparation was performed with the NEXTFLEX DNA library kit as defined by the manufacturer's protocol (11). Briefly, genomic DNA was sheared to generate fragments of approximately 200 to 600 bp. The fragment size distribution was checked using an Agilent Bioanalyzer (Agilent Technologies) according to the manufacturer's instructions (12). These adapter ligated fragments were subjected

Citation Shanmugam S, Kumar N, Nair D, Natrajan M, Tripathy SP, Peacock SJ, Swaminathan S, Ranganathan UD. 2019. Genome sequencing of polydrug-, multidrug-, and extensively drug-resistant *Mycobacterium tuberculosis* strains from South India. *Microbiol Resour Announc* 8:e01388-18. <https://doi.org/10.1128/MRA.01388-18>.

Editor J. Cameron Thrash, University of Southern California

Copyright © 2019 Shanmugam 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 Uma Devi Ranganathan, krumadevi@gmail.com.

Received 1 November 2018

Accepted 11 February 2019

Published 21 March 2019

TABLE 1 Summary of sequence details, drug resistance category, and lineage prediction for *M. tuberculosis* isolates

NCBI identification no. ^a	Assembly accession no.	Coverage depth (×)	Read length (bp)	GC content (%)	N ₅₀ value (bp)	No. of contigs >500	Genome size (bp)	Reference coverage (H37Rv) (%)	No. of tRNAs	No. of rRNAs	No. of coding DNA sequences	Lineage
NIRTX002_PDR	SDOU00000000	74	151	65.58	90,381	103	4,220,229	99.3	44	0	3,997	1
NIRTX003_XDR	SDOT00000000	131	151	65.58	98,656	107	4,375,554	99.2	45	3	4,139	3
NIRTX004_XDR	SDOS00000000	79	100	65.35	57,187	136	4,233,503	96.3	45	2	4,112	4
NIRTX005_XDR	SDOR00000000	80	150	65.37	60,784	140	4,262,450	96.8	45	4	4,163	4
NIRTX006_XDR	SDOQ00000000	75	76	65.59	98,740	114	4,350,302	98.6	45	3	4,119	2
NIRTX007_PDR	SDOP00000000	85	76	65.58	85,411	133	4,389,631	99.5	45	3	4,145	1
NIRTX008_XDR	SDOO00000000	52	76	65.67	114,871	89	4,300,325	97.5	45	1	4,068	4
NIRTX009_XDR	SDON00000000	91	76	65.59	108,252	91	4,379,992	99.3	45	3	4,103	1
NIRTX010_PDR	SDOM00000000	69	76	65.61	72,683	127	4,329,145	98.1	45	3	4,118	2
NIRTX011_XDR	SDOL00000000	60	76	65.59	115,334	96	4,384,136	99.4	45	3	4,113	1
NIRTX015_MDR	SDOK00000000	198	76	65.59	105,690	109	4,366,330	99.0	45	3	4,113	1
NIRTX016_MDR	SDOJ00000000	71	76	65.59	104,750	99	4,378,706	99.3	45	3	4,123	1
NIRTX019_MDR	SDOI00000000	53	76	65.56	97,386	98	4,370,427	99.1	45	3	4,113	1
NIRTX022_MDR	SDOH00000000	174	151	65.61	94,182	101	4,362,069	98.9	45	3	4,136	4
NIRTX024_MDR	SDOG00000000	68	151	65.56	105,728	92	4,383,766	99.4	45	3	4,152	1
NIRTX025_MDR	SDOF00000000	92	151	65.54	25,670	323	4,380,819	99.9	49	8	4,295	1

^a MDR, multidrug-resistant TB; XDR, extensively drug-resistant TB; PDR, polydrug-resistant TB.

to 10 rounds of PCR with primers provided in the NEXTFLEX DNA sequencing kit. The libraries were quantified using quantitative PCR (qPCR, KAPA library quantification kit, Kappabiosystems) normalized to 1 nM and denatured with 0.1 N NaOH before sequencing with MiSeq v2 cluster chemistry (Illumina). Raw sequence reads were filtered with Trimmomatic v0.36 (13) (quality >20) and assembled with SPAdes v3.11.0 (14) with default parameters, and the contigs larger than 500 bp were filtered for further analysis. The filtered reads were also aligned to H37Rv (GenBank accession number [NC_000962](#)) using bwa v0.7.12 (15). The alignment obtained was processed for correction of alignments at indels by the local realignment with a combination of Picard v2.2.4 (<http://broadinstitute.github.io/picard/>) and GATK v3.5 (16). The variants were identified with SAMtools v1.3.1 (17) and were filtered based on the following parameters: base quality greater than 50, depth greater than 20, mapping quality greater than 30, and alternate allele proportion greater than 75%. The variants were further annotated using a custom Python script (https://github.com/kumarnaren/mtb_vcf_annotator).

Table 1 lists the characteristics of the 16 genomes that were sequenced. The complete length of the assembled genomes ranged from 4,220,229 to 438,963 bp, with coverage depth ranging from 52× to 198×. Examination of the genetic variants revealed the presence of the well-known mutations S315T in *katG* and S450L in *rpoB* (18, 19) to be the predominant mutations in isoniazid- and rifampin-resistant isolates, respectively. Isolates were assigned to lineages 1 to 4 based on the presence or absence of phylogenetic single nucleotide polymorphisms (SNPs) reported by Coll et al. (18). Lineage 1 was dominant in our collection of 16 isolates, which is consistent with our previous reports (3, 5) based on spoligotyping and WGS of drug-susceptible clinical isolates of *M. tuberculosis*.

The sequenced genomes of these isolates will contribute toward an understanding of the genomic characteristics of strains of *M. tuberculosis* that are prevalent in India, including the genetic mutations that cause resistance. In particular, these data add to knowledge about isolates from South India and provide a resource for future studies of *M. tuberculosis* in the region.

Data availability. The whole-genome sequence reads of the *Mycobacterium tuberculosis* isolates reported here are deposited in the Sequence Read Archive (SRA) under accession number [PRJNA492975](#). This whole-genome shotgun project (assembly) has been deposited in GenBank under the accession numbers [SDOF00000000](#) to [SDOU00000000](#). The versions described in this paper are accession numbers SDOF01000000 to SDOU01000000.

ACKNOWLEDGMENTS

This publication presents research supported by the MRC-DBT-funded partnership between the National Institute for Research in Tuberculosis, Chennai, India (Indian Council of Medical Research, New Delhi 5/2-8/LDCE/2014 for S.K., Department of Biotechnology [BT/IN/DBT-MRC (UK)/12/SS/2015-2016] for D.N., M.N., S.P.T., S.S., and U.D.R.) and the University of Cambridge (UK Medical Research Council [MR/N501864/1] for N.K. and S.P.).

REFERENCES

1. WHO. 2017. WHO: global tuberculosis report 2017. WHO, Geneva, Switzerland.
2. RNTCP. 2018. India TB report 2018. RNTCP, New Delhi, India.
3. Manson AL, Abeel T, Galagan JE, Sundaramurthi JC, Salazar A, Gehrman T, Shanmugam SK, Palaniyandi K, Narayanan S, Swaminathan S, Earl AM. 2017. *Mycobacterium tuberculosis* whole genome sequences from Southern India suggest novel resistance mechanisms and the need for region-specific diagnostics. *Clin Infect Dis* 64:1494–1501. <https://doi.org/10.1093/cid/cix169>.
4. Martin RS, Sumarah RK, Robart EM. 1975. Comparison of four culture media for the isolation of *Mycobacterium tuberculosis*: a 2-year study. *J Clin Microbiol* 2:438–440.
5. Shanmugam S, Selvakumar N, Narayanan S. 2011. Drug resistance among different genotypes of *Mycobacterium tuberculosis* isolated from patients from Tiruvallur, South India. *Infect Genet Evol* 11:980–986. <https://doi.org/10.1016/j.meegid.2011.03.011>.
6. Baess I. 1974. Isolation and purification of deoxyribonucleic acid from mycobacteria. *Acta Pathol Microbiol Scand B* 82:780–784. <https://doi.org/10.1111/j.1699-0463.1974.tb02375.x>.
7. Kam KM, Sloutsky A, Yip CW, Bulled N, Seung KJ, Zignol M, Espinal M, Jun SC, Kim SJ. 2010. Determination of critical concentrations of second-line anti-tuberculosis drugs with clinical and microbiological relevance. *Int J Tuberc Lung Dis* 14:282–288.
8. Siddiqi S, Ahmed A, Asif S, Behara D, Javaid M, Jani J, Jyoti A, Mahatre R, Mahto D, Richter E, Rodrigues C, Visalakshi P, Rüscher-Gerdes S. 2012. Direct drug susceptibility testing of *Mycobacterium tuberculosis* for rapid detection of multidrug resistance using the Bactec MGIT 960 system: a multicenter study. *J Clin Microbiol* 50:435–440. <https://doi.org/10.1128/JCM.05188-11>.
9. Desjardins P, Conklin D. 2010. NanoDrop microvolume quantitation of nucleic acids. *J Vis Exp* 45:e2565. <https://doi.org/10.3791/2565>.
10. Nakayama Y, Yamaguchi H, Einaga N, Esumi M. 2016. Pitfalls of DNA quantification using DNA-binding fluorescent dyes and suggested solutions. *PLoS One* 11:e0150528. <https://doi.org/10.1371/journal.pone.0150528>.
11. Xia S, Huang C-C, Le M, Dittmar R, Du M, Yuan T, Guo Y, Wang Y, Wang X, Tsai S, Suster S, Mackinnon AC, Wang L. 2015. Genomic variations in plasma cell free DNA differentiate early stage lung cancers from normal controls. *Lung Cancer* 90:78–84. <https://doi.org/10.1016/j.lungcan.2015.07.002>.
12. Brena RM, Auer H, Kornacker K, Hackanson B, Raval A, Byrd JC, Plass C. 2006. Accurate quantification of DNA methylation using combined bisulfite restriction analysis coupled with the Agilent 2100 Bioanalyzer platform. *Nucleic Acids Res* 34:e17. <https://doi.org/10.1093/nar/gnj017>.
13. 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>.
14. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski 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>.
15. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
16. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303. <https://doi.org/10.1101/gr.107524.110>.
17. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
18. Coll F, McNeerney R, Guerra-Assunção JA, Glynn JR, Perdigão J, Viveiros M, Portugal I, Pain A, Martin N, Clark TG. 2014. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat Commun* 5:4812. <https://doi.org/10.1038/ncomms5812>.
19. Zeng X, Kwok JS-L, Yang KY, Leung KS-S, Shi M, Yang Z, Yam W-C, Tsui SK-W. 2018. Whole genome sequencing data of 1110 *Mycobacterium tuberculosis* isolates identifies insertions and deletions associated with drug resistance. *BMC Genomics* 19:365. <https://doi.org/10.1186/s12864-018-4734-6>.