



# Complete Genome Sequence of a *Mycobacterium tuberculosis* Strain Belonging to the East African-Indian Family in the Indo-Oceanic Lineage, Isolated in Hanoi, Vietnam

Takayuki Wada,<sup>a,b</sup> Minako Hijikata,<sup>b</sup> Shinji Maeda,<sup>c</sup> Nguyen Thi Le Hang,<sup>d</sup> Pham Huu Thuong,<sup>e</sup> Nguyen Phuong Hoang,<sup>f</sup> Nguyen Van Hung,<sup>g</sup> Naoto Keicho<sup>b,h</sup>

Department of International Health, Institute of Tropical Medicine, Nagasaki University, Nagasaki, Japan<sup>a</sup>; Department of Pathophysiology and Host Defense, Research Institute of Tuberculosis, Kiyose, Tokyo, Japan<sup>b</sup>; Hokkaido Pharmaceutical University School of Pharmacy, Sapporo, Hokkaido, Japan<sup>c</sup>; NCGM-BMH Medical Collaboration Center, Hanoi, Vietnam<sup>d</sup>; Hanoi Lung Hospital, Hanoi, Vietnam<sup>e</sup>; Department of Microbiology, Hanoi Lung Hospital, Hanoi, Vietnam<sup>f</sup>; Department of Microbiology, National Lung Hospital, Hanoi, Vietnam<sup>g</sup>; National Center for Global Health and Medicine, Tokyo, Japan<sup>h</sup>

**ABSTRACT** The East African-Indian (EAI) family of *Mycobacterium tuberculosis* is an endemic group mainly observed in Southeast Asia. Here, we report the complete genome sequence of an *M. tuberculosis* strain isolated as a member of the EAI family in Hanoi, Vietnam, a country with a high incidence of tuberculosis.

The incidence of tuberculosis in Vietnam was estimated to be 128 per 100,000 individuals in 2015 (1). In Hanoi, the capital city of the country, a population-based study to determine the genetic background of *Mycobacterium tuberculosis* was conducted by our group from 2007 to 2009, and 465 clinical strains were isolated (2). Although Beijing family strains belonging to the East Asian lineage (lineage 2) (3, 4) were predominant (58.5%) in Hanoi, the East African-Indian (EAI) family belonging to an ancestral Indo-Oceanic lineage (lineage 1), isolated from Southeast Asian countries and other regions (5), has also been isolated (17.6%) in the same area (2). In Vietnam, the influence of EAI family strains may exceed that of Beijing family strains in rural areas, indicating that the EAI family may have been indigenous to this region before the Beijing family started to spread (6, 7).

Here, we describe the complete genome sequence of HN-024, an EAI4-VNM strain of *M. tuberculosis*, defined by *in silico* spoligotyping. This strain was isolated from a 58-year-old Vietnamese woman before initial treatment for tuberculosis in 2007 (2). The strain showed no resistance to rifampin, isoniazid, streptomycin, ethambutol, or pyrazinamide by drug susceptibility testing and a pyrazinamidase assay. A PacBio RS II instrument (Pacific Biosciences, CA) was used to determine the complete genome sequence. Long-read sequences were obtained (399.8 Mb, with 4,821 bp as the average length of read inserts) and assembled with Hierarchical Genome Assembly Process (HGAP) version 3. Consequently, a single contig over 4.4 Mb in length was assembled. The contig was polished by a mapping analysis of 300-bp paired-end reads obtained by MiSeq sequencing (Illumina, CA), using CLC Genomics Workbench (Qiagen, CA, USA). After polishing, two regions, including repetitive structures by remapping analysis, remained, and their nucleotide sequences were directly determined by Sanger sequencing for confirmation.

Received 26 April 2017 Accepted 28 April 2017 Published 15 June 2017

**Citation** Wada T, Hijikata M, Maeda S, Hang NTL, Thuong PH, Hoang NP, Hung NV, Keicho N. 2017. Complete genome sequence of a *Mycobacterium tuberculosis* strain belonging to the East African-Indian family in the Indo-Oceanic lineage, isolated in Hanoi, Vietnam. *Genome Announc* 5:e00509-17. <https://doi.org/10.1128/genomeA.00509-17>.

**Copyright** © 2017 Wada 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 Naoto Keicho, [nkeicho-ty@umin.ac.jp](mailto:nkeicho-ty@umin.ac.jp).

Finally, the complete genome sequence was 4,399,916 bp, with 65.6% G+C content. These values were similar to previously reported complete genomes, such as those of H37Rv (a reference strain, with 4,411,532 bp and 65.6% G+C content) (8) and EAI5 (a member of the EAI family, with 4,391,174 bp and 65.6% G+C content) (9). Prior to submission to DDBJ, D-FAST (10), a pipeline, including annotation by PROKKA version 1.11 (11), was applied to the genome sequence, which predicted 4,027 protein-coding regions and 52 tRNAs. To improve the compatibility and usability of the new sequence, the original annotation by D-FAST was modified to include locus tags and gene names of the H37Rv genome (GenBank accession no. AL123456.3), according to reciprocal BLASTP best hits by standalone BLAST+ (version 2.2.29) (12).

Because the genome sequence reported in this study was obtained from a typical Vietnamese EAI strain, it could serve as a reference for comparative genomics of *M. tuberculosis* and may provide clues to elucidate the history of the spread of EAI family strains in Vietnam.

**Accession number(s).** This whole-genome sequencing study has been deposited at DDBJ/ENA/GenBank under the accession number [AP018033](https://doi.org/10.1093/nar/gkz180). The version described in this paper is the first version.

## ACKNOWLEDGMENTS

We thank Daisuke Motooka, Research Institute for Microbial Diseases Osaka University, who provided great support for PacBio data collection.

This work was supported by the Japan Initiative for Global Research Network on Infectious Diseases (J-GRID), from the Ministry of Education, Culture, Sport, Science, & Technology in Japan (MEXT), and the Japan Agency for Medical Research and Development (AMED).

## REFERENCES

- World Health Organization. 2016. Global tuberculosis report 2016 Annex 2. Country profiles for 30 high-burden countries. World Health Organization, Geneva, Switzerland.
- Maeda S, Hang NT, Lien LT, Thuong PH, Hung NV, Hoang NP, Cuong VC, Hijikata M, Sakurada S, Keicho N. 2014. *Mycobacterium tuberculosis* strains spreading in Hanoi, Vietnam: Beijing sublineages, genotypes, drug susceptibility patterns, and host factors. *Tuberculosis* 94:649–656. <https://doi.org/10.1016/j.tube.2014.09.005>.
- European Concerted Action on New Generation Genetic Markers and Techniques for the Epidemiology and Control of Tuberculosis. 2006. Beijing/W genotype *Mycobacterium tuberculosis* and drug resistance. *Emerg Infect Dis* 12:736–743. <https://doi.org/10.3201/eid1205.050400>.
- Glynn JR, Whiteley J, Bifani PJ, Kremer K, van Soolingen D. 2002. Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: a systematic review. *Emerg Infect Dis* 8:843–849. <https://doi.org/10.3201/eid0805.020002>.
- Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, Nicol M, Niemann S, Kremer K, Gutierrez MC, Hilty M, Hopewell PC, Small PM. 2006. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* 103:2869–2873. <https://doi.org/10.1073/pnas.0511240103>.
- Buu TN, van Soolingen D, Huyen MN, Lan NT, Quy HT, Tiemersma EW, Kremer K, Borgdorff MW, Cobelens FG. 2012. Increased transmission of *Mycobacterium tuberculosis* Beijing genotype strains associated with resistance to streptomycin: a population-based study. *PLoS One* 7:e42323. <https://doi.org/10.1371/journal.pone.0042323>.
- Nguyen VA, Choisy M, Nguyen DH, Tran TH, Pham KL, Thi Dinh PT, Philippe J, Nguyen TS, Ho ML, Van Tran S, Bañuls AL, Dang DA. 2012. High prevalence of Beijing and EAI4-VNM genotypes among *M. tuberculosis* isolates in northern Vietnam: sampling effect, rural and urban disparities. *PLoS One* 7:e45553. <https://doi.org/10.1371/journal.pone.0045553>.
- Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE III, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393:537–544. <https://doi.org/10.1038/31159>.
- Al Rashdi ASA, Jadhav BL, Deshpande T, Deshpande U. 2014. Whole-genome sequencing and annotation of a clinical isolate of *Mycobacterium tuberculosis* from Mumbai, India. *Genome Announc* 2(2):e00154-14. <https://doi.org/10.1128/genomeA.00154-14>.
- Tanizawa Y, Fujisawa T, Kaminuma E, Nakamura Y, Arita M. 2016. DFAST and DAGA: Web-based integrated genome annotation tools and resources. *Biosci Microbiota Food Health* 35:173–184. <https://doi.org/10.12938/bmfh.16-003>.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>.