

Draft Genome Sequence of the *Mycobacterium tuberculosis* Clinical Isolate C2, Belonging to the Latin American–Mediterranean Family

Yu-Chieh Liao,^b Yih-Yuan Chen,^a Hsin-Hung Lin,^b Jia-Ru Chang,^a Ih-Jen Su,^a Tsi-Shu Huang,^c Horng-Yunn Dou^a

National Institute of Infectious Diseases and Vaccinology, National Health Research Institutes, Zhunan, Miaoli, Taiwan^a; Division of Biostatistics and Bioinformatics, Institute of Population Health, National Health Research Institutes, Zhunan, Miaoli, Taiwan^b; Department of Microbiology, Kaohsiung Veterans General Hospital, Kaohsiung, Taiwan^c

Tuberculosis remains a major infectious disease in Taiwan. Here we present the draft genome sequence of the *Mycobacterium tuberculosis* C2 strain, belonging to the Latin American–Mediterranean lineage. The draft genome sequence comprises 4,453,307 bp with a G+C content of 65.6%, revealing 4,390 coding genes and 45 tRNA genes.

Received 13 May 2014 Accepted 16 May 2014 Published 5 June 2014

Citation Liao Y-C, Chen Y-Y, Lin H-H, Chang J-R, Su I-J, Huang T-S, Dou H-Y. 2014. Draft genome sequence of the *Mycobacterium tuberculosis* clinical isolate C2, belonging to the Latin American–Mediterranean family. *Genome Announc.* 2(3):e00536-14. doi:10.1128/genomeA.00536-14.

Copyright © 2014 Liao et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Horng-Yunn Dou, hydou@nhri.org.tw.

Tuberculosis is a leading notifiable communicable disease in Taiwan. Based on molecular methods, six distinct clades, Beijing, Haarlem, East-African Indian (EAI), Latin American–Mediterranean (LAM), U, and T lineages, were isolated. Epidemiological surveillance indicated that the Beijing lineage was the most prevalent strain in Taiwan, followed by EAI and Haarlem strains (1, 2). The percentages of isolated LAM strains were around 1% to 7% (2, 3).

Here we report the complete genome sequence of a clinical isolate of *Mycobacterium tuberculosis* strain C2, isolated at Kaohsiung Veterans General Hospital, Taiwan, from human sputum. Sputum microscopy and culture samples from the patient tested positive for tuberculosis. The genomic DNA of isolate C2 was extracted from cultured cells as described previously (4), and was sequenced using a MiSeq platform (Illumina, San Diego, CA, USA). Strain C2 was also analyzed by spoligotyping (337777607760771) and mycobacterial interspersed repetitive unit-variable number tandem repeat (MIRU-VNTR) typing (2241243261532241144132250) and was identified as a member of the LAM family. The LAM lineage was one of six major clades of *M. tuberculosis* strains identified in Taiwan. LAM originating in Europe and America may have first occurred during the Portuguese reign in the 16th century and been passed on to the natives of Taiwan.

This study was approved by the Human Ethics Committee of the National Health Research Institutes, Taiwan (code EC1010804-E). Because of the retrospective nature, routine collection of clinical data in daily practice, and dislinkage of personal information, the requirement to obtain informed consent was waived by our institutional review board.

A total of 8,259,434 paired-end reads of 251 bp in length, with an average insert size of 289 bp, were produced. The sequencing reads were trimmed and discarded by limiting the quality score to 0.05 and permitting at most two ambiguous nucleotides in the minimum length of 50 bp. Five different genome assemblies generated separately by using Abyss v1.3.4 (5), Edena v3.130110 (6), SPAdes v2.5.0 (7), and Velvet v1.2.09

(8), were subsequently integrated into an assembly by using CISA (9). The assembly resulted in 85 contigs (>200 bp) at N_{50} value 168,084 bp and a total assembly of 4,453,307 bp, with a G+C content of 65.6%. The draft genome was annotated using the RAST server (10) resulting in a total of 4,390 coding genes and 45 tRNA genes.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under accession no. [JHAD010000000](https://www.ncbi.nlm.nih.gov/nuccore/JHAD010000000). The version described in this paper is the first version, JHAD01000000.

ACKNOWLEDGMENTS

This work was sponsored by grants from National Health Research Institutes, Taiwan (IV-103-SP-06 and PH103-PP-05). The computational facilities for this work were partly supported by the National Science Council, Taiwan (Taiwan Bioinformatics Institute Core Facility [NSC102-2319-B-400-001]).

We also thank the mycobacteriology laboratory of Kaohsiung Veterans General Hospital for providing bacterial isolates.

REFERENCES

1. Chen YY, Chang JR, Huang WF, Kuo SC, Su IJ, Sun JR, Chiueh TS, Huang TS, Chen YS, Dou HY. 2012. Genetic diversity of the *Mycobacterium tuberculosis* Beijing family based on SNP and VNTR typing profiles in Asian countries. *PLoS One* 7:e39792. <http://dx.doi.org/10.1371/journal.pone.0039792>.
2. Huang SF, Su WJ, Dou HY, Feng JY, Lee YC, Huang RM, Lin CH, Hwang JJ, Lee JJ, Yu MC. 2012. Association of *Mycobacterium tuberculosis* genotypes and clinical and epidemiological features—a multi-center study in Taiwan. *Infect. Genet. Evol.* 12:28–37. <http://dx.doi.org/10.1016/j.meegid.2011.10.001>.
3. Dou HY, Tseng FC, Lu JJ, Jou R, Tsai SF, Chang JR, Lin CW, Miu WC, Su IJ. 2008. Associations of *Mycobacterium tuberculosis* genotypes with different ethnic and migratory populations in Taiwan. *Infect. Genet. Evol.* 8:323–330. <http://dx.doi.org/10.1016/j.meegid.2008.02.003>.
4. Chen YY, Chang JR, Huang WF, Kuo SC, Yeh JJ, Lee JJ, Jang CS, Sun JR, Chiueh TS, Su IJ, Dou HY. 2014. Molecular epidemiology of *Mycobacterium tuberculosis* in aboriginal peoples of Taiwan, 2006–2011. *J. Infect.* 68:332–337. <http://dx.doi.org/10.1016/j.jinf.2013.12.004>.

5. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res.* 19:1117–1123. <http://dx.doi.org/10.1101/gr.089532.108>.
6. Hernandez D, François P, Farinelli L, Osterås M, Schrenzel J. 2008. De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. *Genome Res.* 18:802–809. <http://dx.doi.org/10.1101/gr.072033.107>.
7. 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. <http://dx.doi.org/10.1089/cmb.2012.0021>.
8. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18:821–829. <http://dx.doi.org/10.1101/gr.074492.107>.
9. Lin SH, Liao YC. 2013. CISA: contig integrator for sequence assembly of bacterial genomes. *PLoS One* 8:e60843. <http://dx.doi.org/10.1371/journal.pone.0060843>.
10. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 9:75. <http://dx.doi.org/10.1186/1471-2164-9-75>.