

# Draft Genome Sequences of the *Mycobacterium tuberculosis* Clinical Strains A2 and A4, Isolated from a Relapse Patient in Taiwan

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

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

**The recurrence rate of *Mycobacterium tuberculosis* in Taiwan is 3%. Here, we present the draft genome sequences of *M. tuberculosis* strains A2 and A4 from a relapse patient. The draft genome sequences comprise 4,443,031 bp and 4,487,096 bp, revealing 4,220 and 4,143 coding sequences for A2 and A4, respectively, as well as 49 tRNA genes for the both isolates.**

Received 12 June 2014 Accepted 8 August 2014 Published 4 September 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 sequences of the *Mycobacterium tuberculosis* clinical strains A2 and A4, isolated from a relapse patient in Taiwan. *Genome Announc*. 2(5):e00672-14. doi:10.1128/genomeA.00672-14.

**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](mailto:hydou@nhri.org.tw).

Resulting in approximately 9.4 million new cases and 1.7 million deaths each year, tuberculosis (TB) remains a global health concern (1). Although anti-TB drugs are effective at treating patients with good compliance, the recurrence rate in Taiwan is still around 3% (2). Several host factors, such as HIV infections and ethnicity, predispose cured patients to relapse (3). In addition, bacterial factors may also play a role. Therefore, we aim to compare the genomes of the relapsed isolates with those with the same genotype. We hope to identify the virulence gene that contributes to relapse. Relapse cases were defined as patients who were cured by a full course of anti-TB drug treatment and then had a recurrence of *Mycobacterium tuberculosis* infection during a follow-up period, with bacteriological confirmation. The fine changes in the genomic sequence may help *M. tuberculosis* evade the immune system. Therefore, an analysis of different stages of mycobacterial genome variation will also help pinpoint how isolated strains evade host immunity.

In this study, we present the completed genome sequence of the clinical isolates *M. tuberculosis* A2 and A4, isolated at Kaohsiung Veterans General Hospital, Taiwan, from a human sputum sample of a relapsed patient, which was confirmed at different stages to be tuberculosis. Strains A2 and A4 were also analyzed by spoligotyping (77777777720771) and mycobacterial interspersed repetitive unit-variable-number tandem-repeat (MIRU-VNTR) typing (314222325163223234233262), identifying them as belonging to the Haarlem lineage and confirming that the two strains indeed belong to the same clone.

This study was approved by the Human Ethics Committee of the National Health Research Institutes, Taiwan (code EC1010804-E). Because of the retrospective nature and deidentification of personal information of the subjects, the requirement of obtaining informed consent was waived by our institutional review board.

Genomic DNA of both isolates was extracted from cultured cells as described previously (4) and sequenced by the MiSeq platform (Illumina, USA) to generate 8,282,272 and 11,387,732

paired-end reads for A2 and A4, respectively. The sequencing reads (length, 251 bp) were trimmed and discarded by limiting the quality score at 0.05 and permitting two ambiguous nucleotides at most in the minimum length of 50 bp. The filtered reads were assembled into contigs using *de novo* assemblers, including ABySS 1.3.4 (5), Edena version 3.130110 (6), SPAdes 2.5.0 (7), and Velvet 1.2.09 (8). The contigs from the assemblies were then integrated by CISA (9), resulting in final sets of 190 and 166 contigs for A2 and A4, respectively. Please note that because an extra-large genome size (>7 Mb) was produced by SPAdes in assembling A2, we excluded the assembly of SPAdes from the contig integration of A2 using CISA. The draft genome assemblies of 4,443,031 bp and 4,487,096 bp were annotated by Prokka (10) to contain 4,220 and 4,143 coding sequences for A2 and A4, respectively, as well as 49 tRNA genes for both isolates.

**Nucleotide sequence accession numbers.** The whole-genome shotgun projects have been deposited at GenBank under accession numbers [JNGF00000000](https://www.ncbi.nlm.nih.gov/nuccore/JNGF00000000) and [JNNW00000000](https://www.ncbi.nlm.nih.gov/nuccore/JNNW00000000) for A2 and A4, respectively. The versions described in this paper are in the first versions, [JNGF01000000](https://www.ncbi.nlm.nih.gov/nuccore/JNGF01000000) and [JNNW01000000](https://www.ncbi.nlm.nih.gov/nuccore/JNNW01000000). The BioProject ID is PRJNA248335.

## 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 ([NSC103-2319-B-400-001]) and VYM Genome Research Center, National Yang-Ming University.

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

## REFERENCES

2010. WHO global tuberculosis control report 2010. Summary. *Cent. Eur. J Public. Health* 18:237.
- Wang JY, Lee LN, Lai HC, Hsu HL, Liaw YS, Hsueh PR, Yang PC. 2007. Prediction of the tuberculosis reinfection proportion from the local incidence. *J. Infect. Dis.* 196:281–288. <http://dx.doi.org/10.1086/518898>.

3. Crofts JP, Andrews NJ, Barker RD, Delpech V, Abubakar I. 2010. Risk factors for recurrent tuberculosis in England and Wales, 1998–2005. *Thorax* 65:310–314. <http://dx.doi.org/10.1136/thx.2009.124677>.
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. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <http://dx.doi.org/10.1093/bioinformatics/btu153>.