AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements

Draft Genome Sequence of *Mycobacterium tuberculosis* KT-0184, Isolated in South Korea

Taesoo Kwon,[a,b] Seung Jung Han,[d] Won Gi Yoo,[c] Mi-ran Yun,[b] Sanghyun Lee,[b] Jong Seok Lee,[e] Dae-Won Kim[b]

School of Biological Sciences, Seoul National University, Seoul, Republic of Korea[a]; Division of Biosafety Evaluation and Control, Korea National Institute of Health, Korea Centers for Disease Control and Prevention, Chungbuk, Republic of Korea[b]; Department of Environmental Medical Biology, Chung-Ang University, Seoul, Republic of Korea[c]; Department of Microbiology and Institute for Immunology and Immunological Diseases, Brain Korea 21 Plus Project for the Medical Sciences, Yonsei University College of Medicine, Seoul, Republic of Korea[d]; International Tuberculosis Research Center, 236 Gaposunhwan-ro, Masanhappo-gu, Changwon, Republic of Korea[e]

T.K. and S.J.H. contributed equally to the work.

Here, we describe the draft genome sequence of *Mycobacterium tuberculosis* KT-0184, from the Beijing family. This genome will provide insight into the evolution and adaptation of *M. tuberculosis* KT-0184 in human hosts.

Received 24 December 2015 **Accepted** 4 January 2016 **Published** 18 February 2016

Citation Kwon T, Han SJ, Yoo WG, Yun M-R, Lee S, Lee JS, Kim D-W. 2016. Draft genome sequence of *Mycobacterium tuberculosis* KT-0184, isolated in South Korea. Genome Announc 4(1):e01755-15. doi:10.1128/genomeA.01755-15.

Address correspondence to Jong Seok Lee, cosmosljs@gmail.com, or Dae-Won Kim, todaewon@gmail.com.

Tuberculosis (TB) is the leading cause of death worldwide, and almost one-third of the world's population is infected with *Mycobacterium tuberculosis* (1). In addition, TB from the Beijing family has been found globally and is a major health problem in South Korea (2, 3). Here, we report the draft genomic sequence of *M. tuberculosis* strain KT-0184, which was isolated from a South Korean patient. Based on spoligotyping, *M. tuberculosis* KT-0184 belongs to Beijing family *M. tuberculosis* and was susceptible to first-line anti-TB drugs. *M. tuberculosis* KT-0184 was isolated from the sputum from a patient with active pulmonary TB at Masan National Hospital (MNH) in South Korea. Genomic DNA was isolated from *M. tuberculosis* KT-0184 grown in 7H9 broth (Difco Laboratories, Detroit, MI, USA) supplemented with 10% (vol/vol) oleic acid-albumin-dextrose-catalase (OADC) (Becton, Dickinson, Sparks, MD, USA) for 1 month at 37°C, as previously described.

We constructed a paired-end sequencing library using the Nextera sample preparation kit (Illumina, San Diego, CA, USA). The insert size of the paired-end sequencing library was 500 bp, and the DNA sequencing platform used was Illumina MiSeq. We produced 6,626,240 reads from the whole-genome sequencing, and the coverage was 360.52×. We assembled the reads into 117 contigs with the CLC Genomics Workbench program (version 7.5; CLC bio) (4). The N_{50} size of the contigs was 86,782 bp. The KT-0184 strain has a genome of 4,368,202 bp, with 65.6% G+C content. After the contigs were entered into the National Center of Biotechnology Information submission portal, KT-0184 was confirmed to belong to the Beijing family (CCDC5079 RefSeq genome accession no. NC_017523.1 [5]), with 98% sequence identity. We identified 4,099 putative open reading frames (ORFs) with Glimmer (version 3.02) (6), 45 tRNAs with tRNAscan-SE (7), and 3 rRNA genes with RNAmmer (8).

A total of 2,863 genes were assigned to Clusters of Orthologous Groups (COG) functional categories, including 121 genes (4.23%) classified as being involved in cell wall/membrane/envelope biogenesis, and 256 genes (8.94%) classified as being involved in lipid transport and metabolism as the most abundant, except for general function genes (407 genes [14.22%]). This abundance is considered to be related to the fact that *M. tuberculosis* has >100 outer membrane proteins and uses lipids to construct an outer membrane (9). On the other hand, we identified 2,889 single-nucleotide variants (SNVs) and 267 indels with reference to the *M. tuberculosis* H37Rv genome (accession no. NC_000962) with GATK (version 3.2.2) (10). Among the SNVs, 77.6% (2,241) are in ORFs, and the rest are in intergenic regions. Of the indels, 58.5% (93/159) of the insertions and 66.7% (72/108) of the deletions are in ORFs, and the rest are in intergenic regions.

This genome will provide insight into the evolution and adaptation of *M. tuberculosis* KT-0184 in human hosts.

Nucleotide sequence accession number. This whole-genome shotgun project has been deposited in DDBJ/EMBL/GenBank under the accession no. JUEX00000000.

ACKNOWLEDGMENT

This work was supported by the Korea National Institute of Health (NIH 4800-4847-311).

FUNDING INFORMATION

MOHW | Korea National Institute of Health (KNIH) provided funding to Taesoo Kwon, Seung Jung Han, Won Gi Yoo, Mi-ran Yun, Sanghyun Lee, Jong Seok Lee, and Dae-Won Kim under grant number NIH 4800-4847-311.

REFERENCES

1. **Zumla A, George A, Sharma V, Herbert RH, Baroness Masham of Ilton, OxleyA, Oliver M.** 2015. The WHO 2014 global tuberculosis report—further to go. Lancet Glob Health 3:e10–e12.
2. **Kim SJ, Bai GH, Lee H, Kim HJ, Lew WJ, Park YK, Kim Y.** 2001. Transmission of Mycobacterium tuberculosis among high school students in Korea. Int J Tuberc Lung Dis 5:824–830.
3. **Park YK, Kang H, Lim JK, Ha JS, Cho JO, Bai GH.** 2006. Analysis of

DNA fingerprints of Mycobacterium tuberculosis isolates from patients registered at health center in Gyeonggi province in 2004. Tuberc Resper Dis **60:**290–296.

4. **Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, Li S, Yang H, Wang J, Wang J.** 2010. *De novo* assembly of human genomes with massively parallel short read sequencing. Genome Res **20:**265–272. http://dx.doi.org/10.1101/gr.097261.109.
5. **Zhang Y, Chen C, Liu J, Deng H, Pan A, Zhang L, Zhao X, Huang M, Lu B, Dong H, Du P, Chen W, Wan K.** 2011. Complete genome sequences of Mycobacterium tuberculosis strains CCDC5079 and CCDC5080, which belong to the Beijing family. J Bacteriol **193:** 5591–5592. http://dx.doi.org/10.1128/JB.05452-11.
6. **Salzberg SL, Delcher AL, Kasif S, White O.** 1998. Microbial gene identification using interpolated Markov models. Nucleic Acids Res **26:** 544–548. http://dx.doi.org/10.1093/nar/26.2.544.
7. **Schattner P, Brooks AN, Lowe TM.** 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res **33:**W686–W689. http://dx.doi.org/10.1093/nar/gki366.
8. **Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res **35:**3100–3108. http://dx.doi.org/10.1093/nar/gkm160.
9. **Niederweis M, Danilchanka O, Huff J, Hoffmann C, Engelhardt H.** 2010. Mycobacterial outer membranes: in search of proteins. Trends Microbiol **18:**109–116. http://dx.doi.org/10.1016/j.tim.2009.12.005.
10. **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. http://dx.doi.org/10.1101/gr.107524.110.