AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements

Draft Genome Sequence of *Mycobacterium tuberculosis* KT-0133, 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, Changwon, Republic of Korea[e]

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

Here, we present the draft genome sequence of *Mycobacterium tuberculosis* KT-0133, which belongs to the Korean-Beijing family. This sequence will provide a new perspective on the evolution and accommodation of *M. tuberculosis* KT-0133 in human hosts.

Received 21 December 2015 **Accepted** 22 December 2015 **Published** 11 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-0133, isolated in South Korea. Genome Announc 4(1):e01731-15. doi:10.1128/genomeA.01731-15.

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

Tuberculosis (TB) remains a public health problem in South Korea, and its causative agent, *Mycobacterium tuberculosis*, infects about one-third of the world's population. According to a World Health Organization survey, there were approximately 8.8 million new cases of TB causing nearly 1.1 million deaths in 2010 (1). Among the TB strains, the Beijing lineage is a cause of major concern worldwide (2, 3), owing to its global spread, involvement in several TB outbreaks, and association with rapid disease progression. Despite its importance, genome information and reports on Beijing TB are relatively sparse.

M. tuberculosis KT-0133 was identified as a member of the Korean-Beijing family MTB according to spoligotyping, and was sensitive to first-line anti-TB drugs. Strain KT-0133 was isolated from the sputum of a retreatment TB patient at Masan National Hospital in South Korea. *M. tuberculosis* KT-0133 was grown in 7H9 broth (Difco Laboratories, USA) supplemented with 10% (vol/vol) oleic acid-albumin-dextrose-catalase (OADC; Becton, Dickinson, USA) for 1 month at 37°C, and the genomic DNA was isolated as previously described.

Using the Nextera sample preparation kit (Illumina, USA), a paired-end sequencing library was constructed for the Illumina MiSeq platform, and the insert size was 500 bp. A total of 4,655,846 reads were produced from the whole-genome sequencing with 259.66-fold coverage. The reads were assembled into 119 contigs with the CLC Genomics Workbench version 7.5 program (CLCbio, USA) (4), and the N_{50} size was 99,151. After the assembly, we estimated a genome size of 4,365,676 bp with a 65.6% G + C content for the KT-0133 strain. Using Glimmer version 3.02 (5), 4,114 putative open reading frames (ORFs) were identified. With respect to RNA genes, 45 tRNAs and 3 rRNAs were identified with tRNAScan-SE (6) and RNAMMER (7), respectively. Among the 4,114 ORFs, 2,788 genes could be assigned to COG functional categories; 249 (8.93%) genes were classified to lipid transport and metabolism-related genes, which was the third-ranked category for abundance. This high abundance is likely due to the fact that *M. tuberculosis* has more than 100 outer membrane proteins, and lipids are required to construct the outer membrane (8). Moreover, 2,852 single nucleotide variants (SNVs) and 254 indels were identified with reference to the *M. tuberculosis* H37Rv genome (accession no. NC_000962) with GATK version 3.2.2 (9). Among the SNVs, 78.7% (2,244) were from ORFs, with the rest in intergenic regions. Of the identified indels, 59.5% (91/153) of the insertions and 67.3% (68/101) of the deletions were in ORFs, with the rest identified to intergenic regions.

This genome will provide a new perspective on the evolution and accommodation of *M. tuberculosis* KT-0133 in human hosts.

Nucleotide sequence accession number. This whole-genome shotgun project has been deposited in the DDBJ/EMBL/GenBank under the accession number JUFG00000000.

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 4800-4847-311.

REFERENCES

1. **WHO.** 2011. Global tuberculosis control: WHO Report 2011. WHO, Geneva, Switzerland. http://apps.who.int/iris/bitstream/10665/44728/1/9789241564380_eng.pdf.
2. **Bifani PJ, Mathema B, Kurepina NE, Kreiswirth BN.** 2002. Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. Trends Microbiol **10**:45–52. http://dx.doi.org/10.1016/S0966-842X(01)02277-6.
3. **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. http://dx.doi.org/10.3201/eid0805.020002.

4. **Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, 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. **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.
6. **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.
7. **Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, 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.
8. **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.
9. **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.