

Draft Genome Sequence of *Mycobacterium tuberculosis* KT-0204, Isolated in South Korea

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

Division of Biosafety Evaluation and Control, Korea National Institute of Health, Korea Centers for Disease Control and Prevention, Chungbuk, Republic of Korea^a; School of Biological Sciences, Seoul National University, Seoul, 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, Masanhappo-gu, Changwon, Republic of Korea^e

M.-R.Y. and S.J.H. contributed equally to the work.

Here, we describe the draft genome sequence of *Mycobacterium tuberculosis* KT-0204, non-Beijing family. This sequence will reveal genes related to the evolution and adaptation of *M. tuberculosis* KT-0204 in human hosts.

Received 10 November 2015 Accepted 18 November 2015 Published 4 February 2016

Citation Yun M-R, Han SJ, Yoo WG, Kwon T, Lee S, Lee JS, Kim D-W. 2016. Draft genome sequence of *Mycobacterium tuberculosis* KT-0204, isolated in South Korea. *Genome Announc* 4(1):e01519-15. doi:10.1128/genomeA.01519-15.

Copyright © 2016 Yun 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 Jong Seok Lee, cosmosljs@gmail.com, or Dae-Won Kim, todaewon@gmail.com.

Tuberculosis (TB) is one of the main infectious causes of death worldwide, with more than 9 million new cases of active disease diagnosed every year resulting in nearly 2 million deaths (World Health Organization, 2014) (1). Although Beijing family strains were assumed to be the predominant strains causing active TB in China, different sublineages of non-Beijing strains have been detected in the countries and regions adjacent to Mainland China (2–5). Thus, non-Beijing strains may also contribute to the high TB burden in Asia.

Mycobacterium tuberculosis KT-0204 is a member of the non-Beijing family MTB according to spoligotyping, and was found to be susceptible to first-line anti-TB drugs. *M. tuberculosis* KT-0204 was isolated from the sputum of a retreatment TB patient at Masan National Hospital (MNH) in South Korea. *M. tuberculosis* KT-0204 was 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, and the genomic DNA was isolated.

A paired-end sequencing library was constructed with a 500-bp insert size using the NexTera sample preparation kit (Illumina, San Diego, CA, USA) for the Illumina-MiSeq platform. We produced a total of 5,088,454 reads from whole-genome sequencing with 311.14 fold coverage. We assembled the reads into 115 contigs with the CLC Genomics Workbench program (CLCbio, version 7.5) (6). The N_{50} size of the contigs was 102,547. After the assembly, we found that the KT-0204 strain has a genome of 4,377,418 bp with a 65.57% G+C content. From the assembled contigs, we identified 4,149 putative open reading frames (ORFs) with Glimmer (version 3.02) (7). To identify RNA genes, we used tRNAScan-SE (8) and RNAMMER (9), and 45 tRNAs and 3 rRNA genes were identified. From the 4,114 ORFs, we could assign clusters of orthologous group (COG) functional categories for 2,899 genes. The majority of genes were assigned to the COG functional categories cell wall/membrane/envelope biogenesis-related genes (126 genes, 4.35%) and lipid transport and metabolism-related

genes (253 genes, 8.73%), except for general function genes (412 genes, 14.21%). *M. tuberculosis* has more than 100 outer membrane proteins and lipids required to construct the outer membrane (10), which explains why lipid transport and metabolism-related genes are particularly abundant in KT-0204. Moreover, we identified 1,936 single nucleotide variants (SNVs) and 176 indels against the *M. tuberculosis* H37Rv genome (accession number NC_000962) with GATK (11) (version 3.2.2). Among the identified SNVs, 78.7% (1,401) were in ORFs. For indels, 59.5% (58/118) of the insertions and 67.3% (37/58) of the deletions were in ORFs.

This genome will provide insight into the evolution and adaptation of *M. tuberculosis* KT-0204 in human hosts.

Nucleotide sequence accession number. This whole-genome shotgun project has been deposited in the DDBJ/EMBL/GenBank under the accession number [JUFW00000000](https://www.ncbi.nlm.nih.gov/nuccore/JUFW00000000).

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 Mi-ran Yun, Taesoo Kwon, Sanghyun Lee, and Dae-Won Kim under grant number NIH 4800-4847-311.

REFERENCES

- World Health Organization. 2014. Global tuberculosis report 2014, p 08. World Health Organization, Geneva, Switzerland.
- Arora J, Singh UB, Suresh N, Rana T, Porwal C, Kaushik A, Pande JN. 2009. Characterization of predominant *Mycobacterium tuberculosis* strains from different subpopulations of India. *Infect Genet Evol* 9:832–839. <http://dx.doi.org/10.1016/j.meegid.2009.05.008>.
- Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, Al-Hajjaj SA, Allix C, Aristimuño L, Arora J, Baumanis V, Binder L, Cafrune P, Cataldi A, Cheong S, Diel R, Ellermeier C, Evans JT, Fauville-Dufaux M, Ferdinand S, Garcia de Viedma D, Garzelli C, Gazzola L, Gomes

- HM, Gutierrez MC, Hawkey PM, van Helden PD, Kadival GV, Kreiswirth BN, Kremer K, Kubin M, Kulkarni SP, Liens B, Lillebaek T, Ho ML, Martin C, Martin C, Mokrousov I, Narvskaia O, Ngeow YF, Naumann L, Niemann S, Parwati I, Rahim Z, Rasolofo-Razanamparany V, Rasolonavalona T, Rossetti ML, Rusch-Gerdes S, Sajduda A, Samper S, Shemyakin IG. 2006. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol* 6:23. <http://dx.doi.org/10.1186/1471-2180-6-23>.
4. Singh UB, Arora J, Suresh N, Pant H, Rana T, Sola C, Rastogi N, Pande JN. 2007. Genetic biodiversity of *Mycobacterium tuberculosis* isolates from patients with pulmonary tuberculosis in India. *Infect Genet Evol* 7:441–448. <http://dx.doi.org/10.1016/j.meegid.2007.01.003>.
 5. Tanveer M, Hasan Z, Siddiqui AR, Ali A, Kanji A, Ghebremicheal S, Hasan R. 2008. Genotyping and drug resistance patterns of *M. tuberculosis* strains in Pakistan. *BMC Infect Dis* 8:171. <http://dx.doi.org/10.1186/1471-2334-8-171>.
 6. 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>.
 7. 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>.
 8. 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>.
 9. 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>.
 10. 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>.
 11. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky 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>.