Complete Genome Sequence of the Clinical Beijing-Like Strain *Mycobacterium tuberculosis* 323 Using the PacBio Real-Time Sequencing Platform

Juan Germán Rodríguez,[a] Camilo Pino,[b] Andreas Tauch,[c] Martha Isabel Murcia[a]

Departamento de Microbiología, Facultad de Medicina, Grupo MICOBAC-UN, Universidad Nacional de Colombia, Sede Bogotá, Colombia[a]; Facultad de Ingeniería, Grupo BioLISI, Universidad Nacional de Colombia, Sede Bogotá, Colombia[b]; Institut für Genomforschung und Systembiologie, Centrum für Biotechnologie (CeBiTec), Universität Bielefeld, Bielefeld, Germany[c]

We report here the whole-genome sequence of the multidrug-resistant Beijing-like strain *Mycobacterium tuberculosis* 323, isolated from a 15-year-old female patient who died shortly after the initiation of second-line drug treatment. This strain is representative of the Beijing-like isolates from Colombia, where this lineage is becoming a public health concern.

Address correspondence to Martha Isabel Murcia, mimurciaa@unal.edu.co.

In recent years, strains of the *Mycobacterium tuberculosis* Beijing lineage have become successful pathogens worldwide due to their capacity to spread in a population. In some cases, these strains have been associated with multidrug-resistant tuberculosis outbreaks (1). In Colombia, the Beijing lineage has been reported since 1998 (2), and recent evidence suggests that strains of this lineage have an increased incidence and are associated with multidrug-resistant tuberculosis outbreaks (3, 4). We therefore sequenced the complete genome of an *M. tuberculosis* Beijing-like strain with the internal nomenclature isolate 323 to gain insights into its genetic characteristics and develop new diagnostic tools to improve the epidemiological control of this dangerous *M. tuberculosis* lineage.

Spoligotyping was used to confirm the assignment of isolate 323 to the Beijing-like lineage of *M. tuberculosis* (5). It was classified as spoligotype international type 190 (SIT190), as previously reported (3). Genomic DNA was extracted from isolate 323, according to the protocol provided by the Institute for Genomic Sciences, University of Maryland (standard operating procedures [SOP] manual 26, version 2.01). The DNA was sequenced using the PacBio RS II sequencer. The hierarchical genome assembly process from the Pacific Biosciences single-molecule real-time (SMRT) Analysis tool kit was used (6) to obtain a *de novo* assembly. The original reads were aligned to the trimmed assembly (7) using BLASR (8). The results showed an assembly with a single contig and a sequence length of 4,411,216 bp, with an average coverage depth of 298.65×. The N_{50} contig length was 8,627 bp, and the mean read length was 5,971 bp. The assembled sequence was annotated using the GenDB annotation system for prokaryotic genomes (9). The genome sequence has a mean G+C content of 65.6%, 4,237 protein-coding regions, 45 tRNA genes, and 3 rRNA genes. Taking the genome of *M. tuberculosis* H37Rv as a reference sequence (10), the assembly has an average identity of 99.88%, with 3,507 total variants, comprising 2,143 single nucleotide polymorphisms (SNPs), 207 multiple nucleotide polymorphisms, 588 insertions, and 569 deletions. Among the SNPs, 1,787 were detected in coding regions, of which 743 are synonymous, 1,026 are missense, and 18 are nonsense substitutions.

SNPs known to confer resistance to rifampin (S450L in *rpoB*), isoniazid (S315T and R463L in *katG* and D229G in *accD6*), ethambutol (M306V in *embB*), pyrazinamide (T142A in *pncA*), and streptomycin (K43R in *rpsL* and E92D in *gidB*) were observed and confirm the multidrug-resistant phenotype of isolate 323. Moreover, a novel mutation was observed in the *embC* gene (deletion of C2972) coding for ethambutol resistance. Regarding second-line antituberculous drugs, a known mutation in the *gyrA* gene (D94G) and the novel mutations E21Q and D668G were detected. SNPs known to confer resistant to amikacin/kanamycin/capreomycin (A1401G in *rrs*) and a novel mutation in the *eis* gene (deletion of CA804) were observed, suggesting that this strain is extensively drug resistant. This is the first report of whole-genome sequencing of a multidrug-resistant *M. tuberculosis* Beijing-like isolate circulating in Colombia. With this genome project, we demonstrate that PacBio sequencing technology is a reliable method for generating finished microbial genomes of *M. tuberculosis* isolates.

Nucleotide sequence accession number. The genome sequence of the *M. tuberculosis* Beijing-like isolate 323 has been deposited in GenBank under the accession no. CP010873.

ACKNOWLEDGMENTS

REFERENCES

1. Merker M, Blin C, Mona S, Duforet-Frebourg N, Lecher S, Willery E, Blum MG, Rüsch-Gerdes S, Mokrousov I, Aleksic E, Allix-Béguec C,

Antierens A, Augustynowicz-Kopeć E, Ballif M, Barletta F, Beck HP, Barry CE, III, Bonnet M, Borroni E, Campos-Herrero I, Cirillo D, Cox H, Crowe S, Crudu V, Diel R, Drobniewski F, Fauville-Dufaux M, Gagneux S, Ghebremichael S, Hanekom M, Hoffner S, Jiao WW, Kalon S, Kohl TA, Kontsevaya I, Lillebæk T, Maeda S, Nikolayevskyy V, Rasmussen M, Rastogi N, Samper S, Sanchez-Padilla E, Savic B, Shamputa IC, Shen A, Sng LH, Stakenas P, Toit K, Varaine F, Vukovic D, Wahl C, Warren R, Supply P, Niemann S, Wirth T. 2015. Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. Nat Genet **47**:242–249. http://dx.doi.org/10.1038/ng.3195.

2. Laserson KF, Osorio L, Sheppard JD, Hernández H, Benitez AM, Brim S, Woodley CL, Hazbón MH, Villegas MV, Castaño MC, Henriquez N, Rodriguez E, Metchock B, Binkin NJ. 2000. Clinical and programmatic mismanagement rather than community outbreak as the cause of chronic, drug-resistant tuberculosis in Buenaventura, Colombia, 1998. Int J Tuberc Lung Dis **4**:673–683.

3. Murcia MI, Manotas M, Jiménez YJ, Hernández J, Cortès MI, López LE, Zozio T, Rastogi N. 2010. First case of multidrug-resistant tuberculosis caused by a rare "Beijing-like" genotype of *Mycobacterium tuberculosis* in Bogotá, Colombia. Infect Genet Evol **10**:678–681. http://dx.doi.org/10.1016/j.meegid.2010.03.010.

4. Ferro BE, Nieto LM, Rozo JC, Forero L, van Soolingen D. 2011. Multidrug-resistant *Mycobacterium tuberculosis*, southwestern Colombia. Emerg Infect Dis **17**:1259–1262. http://dx.doi.org/10.3201/eid1707.101797.

5. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, van Embden J. 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol **35**:907–914.

6. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods **10**:563–569. http://dx.doi.org/10.1038/nmeth.2474.

7. Treangen TJ, Sommer DD, Angly FE, Koren S, Pop M. 2011. Next generation sequence assembly with AMOS. Curr Protoc Bioinformatics Chapter 11:Unit 11.8. http://dx.doi.org/10.1002/0471250953.bi1108s33.

8. Chaisson MJ, Tesler G. 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. BMC Bioinformatics **13**:238. http://dx.doi.org/10.1186/1471-2105-13-238.

9. Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, Clausen J, Kalinowski J, Linke B, Rupp O, Giegerich R, Pühler A. 2003. GenDB—an open source genome annotation system for prokaryote genomes. Nucleic Acids Res **31**:2187–2195. http://dx.doi.org/10.1093/nar/gkg312.

10. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE, III, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature **393**:537–544. http://dx.doi.org/10.1038/31159.