




Annotated Genome Sequences of 16 Lineage 4 *Mycobacterium tuberculosis* Strains from Guatemala

Joseph W. Saelens,^a Dalia Lau-Bonilla,^b Anneliese Moller,^b Ana M. Xet-Mull,^a Narda Medina,^b Brenda Guzmán,^b Maylena Calderón,^b Raúl Herrera,^b Jason E. Stout,^c Eduardo Arathoon,^{d,e} Blanca Samayoa,^{b,f}  David M. Tobin^a

^aDepartment of Molecular Genetics and Microbiology and Center for Host-Microbial Interactions, Duke University Medical Center, Durham, North Carolina, USA

^bAsociación de Salud Integral, Guatemala City, Guatemala

^cDivision of Infectious Diseases and International Health, Department of Medicine, Duke University Medical Center, Durham, North Carolina, USA

^dHospital General San Juan de Dios, Guatemala City, Guatemala

^eClínica Familiar Luis Ángel García, Guatemala City, Guatemala

^fFacultad de Ciencias Químicas y Farmacia, Universidad de San Carlos de Guatemala, Guatemala City, Guatemala

ABSTRACT Whole-genome sequencing has resulted in new insights into the phylogeography of *Mycobacterium tuberculosis*. However, only limited genomic data are available from *M. tuberculosis* strains in Guatemala. Here we report 16 complete genomes of clinical strains belonging to the Euro-American lineage 4, the most common lineage found in Guatemala and Central America.

Genome sequencing has revealed much about the phylogeography of *Mycobacterium tuberculosis*, wherein discrete genetic lineages of this human pathogen associate with specific regions of the world. While many isolates have been sequenced from a variety of locales, limited genomic sequence data are available regarding *M. tuberculosis* strains in Guatemala. Previously, we reported on the presence of East Asian lineage 2 strains in Guatemala in an urban setting (1). However, throughout Central and South America, Euro-American lineage 4 strains are most common (2). Here we report 16 complete genomes of Euro-American lineage 4 strains. Isolates were collected at the Clínica Familiar Luis Ángel García (CFLAG), an HIV-specialized clinic associated with the Hospital General San Juan de Dios in Guatemala City.

Cultures were grown on Lowenstein-Jensen medium. Genomic DNA was extracted using a modified protocol of the ArchivePure DNA cell/tissue purification kit (5 Prime GmbH, Germany). Spoligotyping was carried out using the spoligotyping kit and protocol from Isogen Biosolutions (Ocimum Biosolutions Ltd., India), identifying the isolates as lineage 4 strains. Paired-end 50-bp reads were sequenced on the Illumina HiSeq 2500 platform to depths ranging from 327- to 627-fold coverage. Reads were aligned against the H37Rv reference genome (GenBank accession number NC_000962) using Burrows-Wheeler alignment (3). Variants were called with SAMtools (4) and filtered with VarScan (5) for a minimum read depth of 10, a consensus quality score of 20, and a minimum variant frequency of 0.75. Single nucleotide polymorphisms (SNPs) adjacent to indels and within repetitive regions of the genome were discarded. Neighbor-joining and maximum-likelihood methods of phylogeny construction based on genome-wide SNPs placed the Guatemalan isolates among known lineage 4 strains, confirming the spoligotyping results. Consensus sequences for each isolate were generated using BCFtools (6), and gene annotations were added by the NCBI Prokaryotic Genome Annotation Pipeline.

Received 18 January 2018 **Accepted** 29 January 2018 **Published** 15 February 2018

Citation Saelens JW, Lau-Bonilla D, Moller A, Xet-Mull AM, Medina N, Guzmán B, Calderón M, Herrera R, Stout JE, Arathoon E, Samayoa B, Tobin DM. 2018. Annotated genome sequences of 16 lineage 4 *Mycobacterium tuberculosis* strains from Guatemala. *Genome Announc* 6:e00024-18. <https://doi.org/10.1128/genomeA.00024-18>.

Copyright © 2018 Saelens 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 Blanca Samayoa, investigacion@asi.org.gt, or David M. Tobin, david.tobin@duke.edu.

J.W.S. and D.L.-B. contributed equally to this article.

TABLE 1 Genome sequence data for 16 lineage 4 *M. tuberculosis* strains from Guatemala

Strain	GenBank accession no.	Sequencing depth (×)	Genome size (bp)
GG-111-10	CP025593	539.6	4,411,563
GG-5-10	CP025594	496.0	4,411,442
GG-20-11	CP025595	403.2	4,411,504
GG-27-11	CP025596	565.5	4,411,443
GG-36-11	CP025597	557.2	4,411,469
GG-37-11	CP025598	530.7	4,411,526
GG-45-11	CP025599	464.7	4,411,469
GG-77-11	CP025600	627.0	4,411,508
GG-90-10	CP025601	522.9	4,411,602
GG-109-10	CP025602	550.3	4,411,463
GG-121-10	CP025603	526.2	4,411,510
GG-129-11	CP025604	525.2	4,411,413
GG-134-11	CP025605	551.0	4,411,399
GG-137-10	CP025606	475.6	4,411,446
GG-186-10	CP025607	327.6	4,411,478
GG-229-10	CP025608	557.2	4,411,519

Accession number(s). The genome sequences of the *M. tuberculosis* isolates reported here have been deposited in GenBank under the accession numbers listed in Table 1.

ACKNOWLEDGMENTS

This work was supported by the Asociación de Salud Integral (Guatemala), a National Science Foundation Graduate Research Fellowship (J.W.S.), and a Whitehead Scholar Award (D.M.T.) and approved by the Institutional Ethics Committee of the Hospital San Juan de Dios (Guatemala).

Patient information was collected and maintained exclusively in Guatemala; genome sequencing and analysis of deidentified samples were performed at Duke.

REFERENCES

- Saelens JW, Lau-Bonilla D, Moller A, Medina N, Guzmán B, Calderón M, Herrera R, Sisk DM, Xet-Mull AM, Stout JE, Arathoon E, Samayoa B, Tobin DM. 2015. Whole genome sequencing identifies circulating Beijing-lineage *Mycobacterium tuberculosis* strains in Guatemala and an associated urban outbreak. *Tuberculosis (Edinb)* 95:810–816. <https://doi.org/10.1016/j.tube.2015.09.001>.
- Stucki D, Brites D, Jeljeli L, Coscolla M, Liu Q, Trauner A, Fenner L, Rutaihua L, Borrell S, Luo T, Gao Q, Kato-Maeda M, Ballif M, Egger M, Macedo R, Mardassi H, Moreno M, Tudo Vilanova G, Fyfe J, Globan M, Thomas J, Jamieson F, Guthrie JL, Asante-Poku A, Yeboah-Manu D, Wampande E, Ssengooba W, Joloba M, Henry Boom W, Basu I, Bower J, Saraiva M, Vaconcellos SEG, Suffys P, Koch A, Wilkinson R, Gail-Bekker L, Malla B, Ley SD, Beck HP, de Jong BC, Toit K, Sanchez-Padilla E, Bonnet M, Gil-Brusola A, Frank M, Penlap Beng VN, Eisenach K, Alani I, Wangui Ndung'u P, Revathi G, Gehre F, Akter S, Ntoumi F, Stewart-Isherwood L, Ntinginya NE, Rachow A, Hoelscher M, Cirillo DM, Skenders G, Hoffner S, Bakonyte D, Stakenas P, Diel R, Crudu V, Moldovan O, Al-Hajj S, Otero L, Barletta F, Jane Carter E, Diero L, Supply P, Comas I, Niemann S, Gagneux S. 2016. *Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sublineages. *Nat Genet* 48:1535–1543. <https://doi.org/10.1038/ng.3704>.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstein GM, Wilson RK, Ding L. 2009. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* 25:2283–2285. <https://doi.org/10.1093/bioinformatics/btp373>.
- Danecek P, McCarthy SA. 2017. BCFtools/csq: haplotype-aware variant consequences. *Bioinformatics* 33:2037–2039. <https://doi.org/10.1093/bioinformatics/btx100>.