Functional Analysis Using Whole-Genome Sequencing of a Drug-Sensitive *Mycobacterium tuberculosis* Strain from Peru

D. Tarazona, V. Borda, M. Galarza, J. C. Agapito, H. Guio

Laboratorio de Biotecnología y Biología Molecular, Instituto Nacional de Salud, Lima, Peru; Asociación Latinoamericana de Biotecnología (ALBIOTEC), Lima, Peru; Unidad de Biotecnología Molecular-Laboratorios de Investigación y Desarrollo de Ciencia y Tecnología (LID), Universidad Peruana Cayetano Heredia, Lima, Peru

We report the whole-genome sequence of a Latin American-Mediterranean (LAM) lineage drug-sensitive *Mycobacterium tuberculosis* strain from Peru, INS-SEN. The functional analysis revealed more mutations in secondary metabolite biosynthesis, transport, and catabolism (clusters of orthologous groups [COG] category Q) than for other LAM-sensitive strains. This study contributes to the understanding of the genomic diversity of drug-sensitive *M. tuberculosis*.

In 2012, there were an estimated 8.6 million new cases of tuberculosis (TB) worldwide. In Peru, the incidence rate for TB was 95 cases/100,000 people, of which 96% of cases were drug-sensitive TB (1). It has been reported that in Peru there is a high diversity of *Mycobacterium tuberculosis* lineages, including Latin American-Mediterranean (LAM) (23.8%), Haarlem (23.8%), T (22.3%), and Beijing (9.3%) (3). We performed whole-genome sequencing and analysis to investigate the genetic diversity and phylogeny relationships of a drug-sensitive strain of *M. tuberculosis*, INS-SEN.

INS-SEN was isolated from Lima, Peru. The establishment of this strain’s lineage was based on 24 mycobacterial interspersed repetitive unit-variable number of tandem repeat (MIRU-VNTR) loci (4) and by single-nucleotide polymorphisms (SNPs) based on phylogeny (5). The genomic DNA of INS-SEN was sequenced to 1,406× coverage, which consisted of 61,422,158 paired-end reads, using the Illumina HiSeq 2000 sequencer machine. Then, the genomic sequence was assembled with BWA v 0.5.9-r16 (6), using the H37Rv genome (AL123456.3) as a reference, producing 18 contigs. The genomic sequence was annotated with the Rapid Annotations using Subsystem Technology (RAST) server (7) and Prokaryotic Genome Annotation Pipeline (PGAAP). A polymorphism study of the INS-SEN genome was carried out by comparative analysis against the genome of the drug-sensitive strain KZN 4207 and H37Rv. Additionally, INS-SEN showed more mutations in category Q than the strain KZN 4207. It is possible that the organization of SNPs in INS-SEN may have a role in adaptation to its environment.

INS-SEN had more SNPs in PPE associated with antigenic variation (11) in category N and in PE-PGRS associated with antigenic variation and immune evasion (12) in category M than the strains KZN 4207 and H37Rv. Additionally, INS-SEN showed more mutations in category Q than the strain KZN 4207. It is possible that the organization of SNPs in INS-SEN may have a role in adaptation to its environment.

**Nucleotide sequence accession numbers.** This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number JAQH00000000. The version described in this paper is JAQH01000000.

**ACKNOWLEDGMENT**

This study was supported by the Peruvian National Institute of Health.

**REFERENCES**

1. WHO. 2013. Global tuberculosis report 2013. World Health Organization, Geneva, Switzerland.
2. Reference deleted.
3. Taype CA, Agapito JC, Accinelli RA, Espinoza JR, Godreuil S, Goodman SJ, Batulis AL, Shaw MA. 2012. Genetic diversity, population structure and...
drug resistance of Mycobacterium tuberculosis in Peru. Infect. Genet. Evol. 12:577–585. http://dx.doi.org/10.1016/j.meegid.2012.02.002.

4. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüss-Gerdes S, Willery E, Savine E, de Haas P, van Deutekom H, Roring S, Bifani P, Kurepina N, Kreiswirth B, Sola C, Rastogi N, Vatin V, Gutierrez MC, Fauville M, Niemann S, Skuce R, Kremer K, Locht C, van Soolingen D. 2006. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of Mycobacterium tuberculosis. J. Clin. Microbiol. 44:4498–4510. http://dx.doi.org/10.1128/JCM.01392-06.

5. Filliol I, Motiwala AS, Cavatore M, Qi W, Hazbón MH, Bobadilla del Valle M, Fyfe J, Garcia-García L, Rastogi N, Sola C, Zoio T, Guerrero MI, León A, Crabtree J, Angiulio S, Eisenach KD, Durmaz B, Joloba MI, Rendón A, Sifuentes-Osorio J, Ponce de León A, Cave MD, Fleischmann R, Whittam TS, Alland D. 2006. Global phylogeny of Mycobacterium tuberculosis based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. J. Bacteriol. 188:759–772. http://dx.doi.org/10.1128/JB.188.2.759-772.2006.

6. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. http://dx.doi.org/10.1093/bioinformatics/btp324.

7. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JK, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 42:D206–D214. http://dx.doi.org/10.1093/nar/gkt1226.

8. loeger TR, Koo S, No EG, Chen X, Larsen MH, Jacobs WR, Jr, Pillay M, Sturm AW, Sacchettini JC. 2009. Genome analysis of multi- and extensively drug-resistant tuberculosis from KwaZulu-Natal, South Africa. PLoS One 4:e7778. http://dx.doi.org/10.1371/journal.pone.0007778.

9. Song J, Xu Y, White S, Miller KW, Wolinsky M. 2005. SNPsFinder—a web-based application for genome-wide discovery of single nucleotide polymorphisms in microbial genomes. Bioinformatics 21:2083–2084. http://dx.doi.org/10.1093/bioinformatics/bti176.

10. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4:41. http://dx.doi.org/10.1186/1471-2105-4-41.

11. Zheng H, Lu L, Wang B, Pu S, Zhang X, Zhu G, Shi W, Zhang L, Wang H, Wang S, Zhao G, Zhang Y. 2008. Genetic basis of virulence attenuation revealed by comparative genomic analysis of Mycobacterium tuberculosis strain H37Ra versus H37Rv. PLoS One 3:e2375. http://dx.doi.org/10.1371/journal.pone.0002375.

12. Koh KW, Soh SE, Seah GT. 2009. Strong antibody responses to Mycobacterium tuberculosis PE-PGRS62 protein are associated with latent and active tuberculosis. Infect. Immun. 77:3337–3343. http://dx.doi.org/10.1128/IAI.01175-08.