Whole-Genome Sequences of *Mycobacterium bovis* Strain MbURU-001, Isolated from Fresh Bovine Infected Samples

Moira Lasserre,[a] Luisa Berná,[a] Gonzalo Greif,[a] Florencia Díaz-Viraqué,[a] Gregorio Iraola,[b] Hugo Naya,[b] Miguel Castro-Ramos,[c] Arturo Juambeltz,[c] Carlos Robello[a,d]

Unidad de Biología Molecular, Institut Pasteur de Montevideo, Montevideo, Uruguay[a]; Unidad de Bioinformática, Institut Pasteur de Montevideo, Uruguay[b]; Departamento de Bacteriología, División de Laboratorios Veterinarios (DI.LA.VE.), Montevideo, Uruguay[c]; Departamento de Bioquímica, Facultad de Medicina, Universidad de la República, Montevideo, Uruguay[d]

M.L. and L.B. contributed equally to this work.

Bovine tuberculosis in cattle has a high incidence in Uruguay, where it is considered a disease of national importance. We present the genome sequence of *Mycobacterium bovis* strain MbURU-001, isolated from pectoral lymph nodes of a bovine host from a cattle farm.

Address correspondence to Carlos Robello, robello@pasteur.edu.uy.

Mycobacterium bovis is the etiological agent of bovine tuberculosis, an infectious, chronic disease considered one of the most important in dairy cattle. *M. bovis* has a wide host spectrum; besides cattle, it affects other mammal species, including humans (1) and many other wild animals, which act as a reservoir of the disease and hamper its eradication (2, 3). Moreover, bovine tuberculosis is a zoonotic risk of worldwide socioeconomic importance. In Uruguay, bovine tuberculosis is considered a disease of national importance that has shown a growing incidence in the last few years, and its high prevalence in Uruguay still does not have a clear epidemiological explanation. To gain better knowledge about the epidemiology of this disease in Uruguay and characterize its diversity, it is necessary to compare national strains to available sequences of this species to understand their biology. For these reasons, we sequenced and annotated the complete genome of *M. bovis* MbURU-001, a strain isolated from the pectoral lymph nodes of an infected cow from a cattle farm in Uruguay.

Sequencing of this strain was performed at the Institut Pasteur de Montevideo on an Illumina MiSeq platform from a paired-end library (2 × 75 cycles). Velvet (4) was used to perform a *de novo* assembly, followed by SPAdes (5). Next, CISA (6) software was used to integrate the results of the assemblies. The assembly was further improved using PAGIT toolkit (7) with the reference genome of *M. bovis* AF2122/97 (GenBank accession number NC_002945). The final assembly was automatically annotated with RAST (8). Single nucleotide polymorphisms (SNPs) versus *M. bovis* AF2122/97 were identified through VarScan (9) and the GATK pipeline (10), and snpEff (11) software was used to annotate the effects of these variations.

A total of 2,390,334 paired-end reads were assembled into 100 contigs, with an average coverage of 82-fold. This resulted in a genome size of ~4.30 Mb with a GC content of 65.5%. The genome of the *M. bovis* MbURU-001 strain contains 4,289 predicted coding sequences (CDSs), 1 rRNA operon, and 45 tRNA genes. A total of 596 SNPs and 46 indels were identified in this strain when it was compared to the reference. These SNPs affect 517 unique genes, of which 450 code for proteins. Fourteen genes were truncated by the introduction of stop codons, including one involved in lipid metabolism and a putative RNA methyltransferase; 275 genes had missense mutations, many involved in biosynthesis of secondary metabolites, antibiotics, and metabolism in different environments; and 12 genes were affected by frameshifts and involved in oxidative metabolism and translation. All 14 open reading frames (ORFs) from the PhiRv1 prophage were absent from this strain.

Further comparative genomics between this and other local and global sequenced strains may provide genotype-phenotype associations that might explain the diversity of phenotypes locally found, as well as provide more insights regarding *M. bovis* epidemiology.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number LFGY00000000. The version described in this paper is version LFGY01000000 and consists of sequences LFGY01000001 to LFGY01000100.

REFERENCES

1. **Pérez-Lago L, Navarro Y, García-de-Viedma D.** 2014. Current knowledge and pending challenges in zoonosis caused by *Mycobacterium bovis*: a review. Res Vet Sci **97**:S94–S100. http://dx.doi.org/10.1016/j.rvsc.2013.11.008.
2. **Delahay RJ, Cheeseman CL, Clifton-Hadley RS.** 2001. Wildlife disease

reservoirs: the epidemiology of *Mycobacterium bovis* infection in the European badger (*Meles meles*) and other British mammals. Tuberculosis (Edinb) **81**:43–49. http://dx.doi.org/10.1054/tube.2000.0266.

3. **Delahay RJ, Smith GC, Barlow AM, Walker N, Harris A, Clifton-Hadley RS, Cheeseman CL.** 2007. Bovine tuberculosis infection in wild mammals in the South-West region of England: a survey of prevalence and a semi-quantitative assessment of the relative risks to cattle. Vet J **173:** 287–301. http://dx.doi.org/10.1016/j.tvjl.2005.11.011.

4. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res **18**:821–829. http://dx.doi.org/10.1101/gr.074492.107.

5. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.** 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **19**:455–477. http://dx.doi.org/10.1089/cmb.2012.0021.

6. **Lin S, Liao Y.** 2013. CISA: contig integrator for sequence assembly of bacterial genomes. PLoS One **8**:e60843. http://dx.doi.org/10.1371/journal.pone.0060843.

7. **Swain MT, Tsai IJ, Assefa SA, Newbold C, Berriman M, Otto TD.** 2012. A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. Nat Protoc **7**:1260–1284. http://dx.doi.org/10.1038/nprot.2012.068.

8. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O.** 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics **9**:75. http://dx.doi.org/10.1186/1471-2164-9-75.

9. **Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstock GM, Wilson RK, Ding L.** 2009. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. BioInformatics **25**:2283–2285. http://dx.doi.org/10.1093/bioinformatics/btp373.

10. **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.

11. **Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM.** 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain *w1118; iso-2; iso-3*. Fly (Austin) **6**(2): 80–92. http://dx.doi.org/10.4161/fly.19695.