Updated Reference Genome Sequence and Annotation of Mycobacterium bovis AF2122/97

Author(s):
Malone, Kerri M.; Farrell, Damien; Stuber, Tod P.; Schubert, Olga T.; Aebersold, Ruedi; Robbe-Austerman, Suelee; Gordon, Stephen V.

Publication date:
2017-04-06

Permanent link:
https://doi.org/10.3929/ethz-b-000228047

Originally published in:
Genome Announcements 5(14), https://doi.org/10.1128/genomeA.00157-17

AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

PROKARYOTES

Updated Reference Genome Sequence and Annotation of *Mycobacterium bovis* AF2122/97

Kerri M. Malone,[a] Damien Farrell,[a] Tod P. Stuber,[b] Olga T. Schubert,[c]* Ruedi Aebersold,[c] Suelee Robbe-Austerman,[b] Stephen V. Gordon[a,d,e,f]

School of Veterinary Medicine, University College Dublin, Dublin, Ireland[a]; Diagnostic Bacteriology Laboratory, National Veterinary Services Laboratories, Ames, Iowa, USA[b]; Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland[c]; School of Medicine, University College Dublin, Dublin, Ireland[d]; School of Biomolecular and Biomedical Science, University College Dublin, Dublin, Ireland[e]; UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Dublin, Ireland[f]

ABSTRACT We report here an update to the reference genome sequence of the bovine tuberculosis bacillus *Mycobacterium bovis* AF2122/97, generated using an integrative multiomics approach. The update includes 42 new coding sequences (CDSs), 14 modified annotations, 26 single-nucleotide polymorphism (SNP) corrections, and disclosure that the RD900 locus, previously described as absent from the genome, is in fact present.

Received 9 February 2017 **Accepted** 10 February 2017 **Published** 6 April 2017

Citation Malone KM, Farrell D, Stuber TP, Schubert OT, Aebersold R, Robbe-Austerman S, Gordon SV. 2017. Updated reference genome sequence and annotation of *Mycobacterium bovis* AF2122/97. Genome Announc 5:e00157-17. https://doi.org/10.1128/genomeA.00157-17.

Address correspondence to Stephen V. Gordon, stephen.gordon@ucd.ie.

* Present address: Olga T. Schubert, Department of Human Genetics, University of California Los Angeles, Los Angeles, California, USA.

Mycobacterium bovis, the causative agent of bovine tuberculosis (bTB), is the most widely studied animal-adapted member of the *Mycobacterium tuberculosis* complex (MTBC). bTB exacts a tremendous global economic toll through productivity losses and disease control costs, while zoonotic transmission of *M. bovis* infection is a threat to human health (1–6).

M. bovis AF2122/97 was the first *M. bovis* strain to be sequenced and serves as the reference genome (7), and it was last updated in 2003. An updated reference *M. bovis* genome sequence will provide an essential resource for the tuberculosis (TB) research community and serve as a basis for comparative studies into animal- and human-adapted MTBC members.

To update the *M. bovis* AF2122/97 genome and annotation, a low-passage-number stock taken from the original *M. bovis* AF2122/97 seed stock was resequenced and reannotated using a combination of DNA sequencing, RNA sequencing, and proteomics data. All nucleic acid and protein samples were derived from exponentially grown cultures.

Short-read DNA sequencing libraries were prepared using the Nextera XT DNA library preparation kit (Illumina) and sequenced on the MiSeq system (Illumina), generating 250-bp paired-end reads that were trimmed using Sickle ($Q > 30$), with 60× reference coverage (8). For PacBio RSII sequencing, enzymatically extracted DNA was prepared using a large-insert library (6 kb to 8kb) size selection (9). Two single-molecule real-time (SMRT) cells were used for an output of 542,585,804 bases, a mean read length of 8,141, and 86× reference coverage. DNA sequencing data sets were analyzed using a combination of *de novo* assembly [short reads, SOAP*denovo* (10); long reads, Canu (11)] and nucleotide variant identification methods [short reads, Stampy, SAMtools, and VCFtools (12–14); long reads, Pilon (15); and MUMmer (16)]. This allowed both an update of the genome nucleotide sequence and the identification of genomic regions that had been misassembled, or missed entirely, in the original sequencing project. Reannotation of the *M. bovis* AF2122/97 genome was achieved by automatic annotation transfer from *M. tuberculosis* H37Rv (17) and a proteogenomic analysis using both *M. bovis* AF2122/97 shotgun tandem mass spectrometry (MS/MS), sequential

window acquisition of all theoretical mass spectra (SWATH MS) data sets, and *M. tuberculosis* H37Rv SWATH MS data sets (18).

Overall, 26 single nucleotide polymorphisms were identified. Strikingly, the large sequence polymorphism RD900, originally described as being deleted from *M. bovis* 2122/97 (19), was found to be present; recombination between repeat structures flanking the RD900 locus in clones used for the original shotgun sequencing genome project may have led to the loss of RD900. Furthermore, 42 novel coding sequences were identified, while 14 existing annotations were modified.

Accession number(s). This whole-genome shotgun project had been deposited in DDBJ/ENA/GenBank under the accession no. LT708304. SWATH MS data can be found on PeptideAtlas under identifier PASS00932.

ACKNOWLEDGMENTS

The work was supported by funding from the Department for Agriculture, Food and the Marine (MycobactDiagnosis 11/RD/EMIDA/1), Science Foundation Ireland (08/IN.1/B2038), SystemsX.ch (through the project TbX), and a research grant from Institut Mérieux.

We thank Gerard Cagney and Kieran Wynne for shotgun mass spectrometric analysis of *M. bovis* samples.

REFERENCES

1. De Garine-Wichatitsky M, Caron A, Kock R, Tschopp R, Munyeme M, Hofmeyr M, Michel A. 2013. A review of bovine tuberculosis at the wildlife-livestock-human interface in sub-Saharan Africa. Epidemiol Infect 141:1342–1356. https://doi.org/10.1017/S0950268813000708.
2. Malama S, Muma J, Munyeme M, Mbulo G, Muwonge A, Shamputa IC, Djønne B, Godfroid J, Johansen TB. 2014. Isolation and molecular characterization of *Mycobacterium tuberculosis* from humans and cattle in Namwala District, Zambia. Ecohealth 11:564–570. https://doi.org/10.1007/s10393-014-0940-0.
3. Kazoora HB, Majalija S, Kiwanuka N, Kaneene JB. 2016. Knowledge, attitudes and practices regarding risk to human infection due to *Mycobacterium bovis* among cattle farming communities in western Uganda. Zoonoses Publ Health 63:616–623. https://doi.org/10.1111/zph.12273.
4. Khattak I, Mushtaq MH, Ahmad MU, Khan MS, Haider J. 2016. Zoonotic tuberculosis in occupationally exposed groups in Pakistan. Occup Med (Lond) 66:371–376. https://doi.org/10.1093/occmed/kqw039.
5. Müller B, Dürr S, Alonso S, Hattendorf J, Laisse CJ, Parsons SD, van Helden PD, Zinsstag J. 2013. Zoonotic *Mycobacterium bovis*-induced tuberculosis in humans. Emerg Infect Dis 19:899–908. https://doi.org/10.3201/eid1906.120543.
6. Abernethy DA, Upton P, Higgins IM, McGrath G, Goodchild AV, Rolfe SJ, Broughan JM, Downs SH, Clifton-Hadley R, Menzies FD, de la Rua-Domenech R, Blissitt MJ, Duignan A, More SJ. 2013. Bovine tuberculosis trends in the UK and the Republic of Ireland, 1995–2010. Vet Rec 172:312. https://doi.org/10.1136/vr.100969.
7. Garnier T, Eiglmeier K, Camus JC, Medina N, Mansoor H, Pryor M, Duthoy S, Grondin S, Lacroix C, Monsempe C, Simon S, Harris B, Atkin R, Doggett J, Mayes R, Keating L, Wheeler PR, Parkhill J, Barrell BG, Cole ST, Gordon SV, Hewinson RG. 2003. The complete genome sequence of *Mycobacterium bovis*. Proc Natl Acad Sci U S A 100:7877–7882. https://doi.org/10.1073/pnas.1130426100.
8. Joshi NA, Fass JN. 2011. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files. https://github.com/najoshi/sickle.
9. van Soolingen D, Hermans PW, de Haas PE, Soll DR, van Embden JD. 1991. Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. J Clin Microbiol 29:2578–2586.
10. Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, Huang W, He G, Gu S, Li S, Zhou X, Lam TW, Li Y, Xu X, Wong GK, Wang J. 2014. SOAP*denovo*-Trans: *de novo* transcriptome assembly with short RNA-Seq reads. Bioinformatics 30:1660–1666. https://doi.org/10.1093/bioinformatics/btu077.
11. Koren S, Walenz B, Berlin K, Miller JR, Phillippy AM. 2016. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. bioRxiv. https://doi.org/10.1101/071282.
12. Lunter G, Goodson M. 2011. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res 21:936–939. https://doi.org/10.1101/gr.111120.110.
13. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing Subroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
14. Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27:2987–2993. https://doi.org/10.1093/bioinformatics/btr509.
15. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng QD, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
16. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol 5:R12. https://doi.org/10.1186/gb-2004-5-2-r12.
17. Otto TD, Dillon GP, Degrave WS, Berriman M. 2011. RATT: rapid annotation transfer tool. Nucleic Acids Res 39:e57. https://doi.org/10.1093/nar/gkq1268.
18. Schubert OT, Mouritsen J, Ludwig C, Röst HL, Rosenberger G, Arthur PK, Claassen M, Campbell DS, Sun Z, Farrah T, Gengenbacher M, Maiolica A, Kaufmann SH, Moritz RL, Aebersold R. 2013. The Mtb proteome library: a resource of assays to quantify the complete proteome of *Mycobacterium tuberculosis*. Cell Host Microbe 13:602–612. https://doi.org/10.1016/j.chom.2013.04.008.
19. Bentley SD, Comas I, Bryant JM, Walker D, Smith NH, Harris SR, Thurston S, Gagneux S, Wood J, Antonio M, Quail MA, Gehre F, Adegbola RA, Parkhill J, de Jong BC. 2012. The genome of *Mycobacterium africanum* West African 2 reveals a lineage-specific locus and genome erosion common to the *M. tuberculosis* complex. PLoS Negl Trop Dis 6:e1552. https://doi.org/10.1371/journal.pntd.0001552.