Complete Genome Sequence of *Mycobacterium orygis* Strain 51145

Syed Beenish Rufai,[a,b,c] Fiona McIntosh,[b,c] Indira Poojary,[d] Shubhada Chothe,[d] Aswathy Sebastian,[d] Istvan Albert,[e] Craig Praul,[d] Manigandan Venkatesan,[f] Gibarni Mahata,[d] Hindol Maity,[d,g] Premanshu Dandapat,[h] Joy Sarojini Michael,[f] Robab Katani,[c,i] Vivek Kapur,[c,i] [ID] Marcel A. Behr[a,b,c]

[a]Department of Microbiology and Immunology, McGill University, Montreal, Quebec, Canada

[b]McGill International TB Centre, Montreal, Quebec, Canada

[c]Infectious Diseases and Immunity in Global Health Program, Research Institute of the McGill University Health Centre, Montréal, Canada

[d]Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, Pennsylvania, USA

[e]Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, Pennsylvania, USA

[f]Department of Clinical Microbiology, Christian Medical College Vellore, Vellore, India

[g]Cisgen Biotech Discoveries, Chennai, India

[h]Indian Veterinary Research Institute, Eastern Region Station, Kolkata, West Bengal, India

[i]Department of Animal Science and the Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, Pennsylvania, USA

ABSTRACT We report the complete 4,352,172-bp genome sequence of *Mycobacterium orygis* strain 51145 assembled into a single circular chromosome. Comparative genomic analyses with other lineages of the *Mycobacterium tuberculosis* complex can provide insights into the biology, evolution, and epidemiology of this important group of pathogenic mycobacteria.

Within the *Mycobacterium tuberculosis* complex, *M. orygis* has received increasing recognition. Veterinary cases have been detected in a wide range of hosts, including oryxes, cattle, antelopes, and rhesus monkeys, and human cases have been reported mainly in patients of South Asian origin (1–6). Its detection among cattle in South Asia supports the possibility of a livestock reservoir, but the prevalence of *M. orygis* in these hosts is poorly understood (5, 6). Defining the genetic landscape of *M. orygis* is essential for understanding the biology and epidemiology of this human and veterinary pathogen.

M. orygis isolate 51145 was recovered from a human diagnosed with tuberculous meningitis in 1997 and initially classified by the Laboratoire de Santé Publique du Québec as *M. caprae* based on phenotypic characteristics (MPT70 positive, pyrazinamide sensitive). In 2005, the sample was referred to the McGill University Health Centre, Montreal, Canada, where it was definitively identified, based on specific genomic deletions, as "oryx bacillus," the name given for *M. orygis* at that time (7, 8). To characterize the genome, *M. orygis* 51145 cultures were revived from frozen glycerol stocks and grown to mid-log phase in Middlebrook 7H9 (Difco, Becton Dickinson) with oleic acid, albumin, dextrose, and catalase (OADC) enrichment and polymyxin B, amphotericin B, nalidixic acid, trimethoprim, and azlocillin (PANTA). DNA extraction was performed using a MagAttract high-molecular-weight (HMW) DNA kit (Qiagen, Inc., Valencia, CA), and the quality of DNA was checked using a Quant-iT Pico Green double-stranded DNA (dsDNA) assay kit (Thermo Fisher Scientific) (9).

Whole-genome sequencing was performed using PacBio single-molecule real-time (SMRT) sequencing technology at the Penn State Genomics Core Facility, University Park, Pennsylvania. Briefly, DNA was sheared to an average size of 10 kb using a Covaris g-TUBE for preparation of SMRTbell libraries without further DNA size selection. The template was prepared using the SMRTbell Express template prep kit 2.0 according to the

manufacturer's protocol (PacBio, Menlo Park, CA). Sequencing was performed on a PacBio Sequel instrument using v3.0 chemistry and a 20-hour movie time. Circular consensus sequences (CCS) were generated from subreads with at least three full passes to achieve Q20 read accuracy. A total of 25,135 CCS were generated with average and maximum read lengths of 2,263 bp and 11,835 bp, respectively. CCS reads were corrected, trimmed, filtered, and *de novo* assembled (genome Size = 5 m and min Overlap Length = 100) using Canu v2.0, which generated 15 contigs (8). Assembled contigs were merged and circularized with Circlator v1.5.5 using default parameters, generating two scaffolds (10). The remaining gap was closed by reference-guided alignment of subreads to the *M. bovis* genome (AF2122) and identification of subreads that bridged the two scaffolds, resulting in a single contig of 4,352,172 bp. Annotation was performed using the NCBI Prokaryotic Genome Annotation Pipeline (11). The *M. orygis* 51145 chromosome size is 4,352,172 bp with a GC content of 65.6%. It encodes 4,039 genes and 45 tRNA and 3 rRNA sequences.

Data availability. The *Mycobacterium orygis* 51145 genome was deposited in GenBank under the accession number CP063804 and BioProject accession number PRJNA672657. The version outlined here is the first version of the accession number. CCS raw reads and subreads have been made available in the Sequence Read Archive database under SRA accession numbers SRR12966358 and SRR12916665.

REFERENCES

1. Duffy SC, Srinivasan S, Schilling MA, Stuber T, Danchuk SN, Michael JS, Venkatesan M, Bansal N, Maan S, Jindal N, Chaudhary D, Dandapat P, Katani R, Chothe S, Veerasami M, Robbe-Austerman S, Juleff N, Kapur V, Behr MA. 2020. Reconsidering *Mycobacterium bovis* as a proxy for zoonotic tuberculosis: a molecular epidemiological surveillance study. Lancet Microbe 1:e66–e73. https://doi.org/10.1016/S2666-5247(20)30038-0.

2. Dawson KL, Bell A, Kawakami RP, Coley K, Yates G, Collins DM. 2012. Transmission of *Mycobacterium orygis* (*M. tuberculosis* complex species) from a tuberculosis patient to a dairy cow in New Zealand. J Clin Microbiol 50:3136–3138. https://doi.org/10.1128/JCM.01652-12.

3. Refaya AK, Kumar N, Raj D, Veerasamy M, Balaji S, Shanmugam S, Rajendran A, Tripathy SP, Swaminathan S, Peacock SJ, Palaniyandi K. 2019. Whole-genome sequencing of a *Mycobacterium orygis* strain isolated from cattle in Chennai, India. Microbiol Resour Announc 8:e01080-19. https://doi.org/10.1128/MRA.01080-19.

4. van Ingen J, Rahim Z, Mulder A, Boeree MJ, Simeone R, Brosch R, van Soolingen D. 2012. Characterization of *Mycobacterium orygis* as *M. tuberculosis* complex subspecies. Emerg Infect Dis 18:653–655. https://doi.org/10.3201/eid1804.110888.

5. Parsons SDC. 2020. *Mycobacterium orygis*: a zoonosis, zooanthroponosis, or both? Lancet Microbe 1:e240. https://doi.org/10.1016/S2666-5247(20)30142-7.

6. Duffy SC, Srinivasan S, Michael JS, Behr MA, Kapur V. 2020. *Mycobacterium orygis*: a zoonosis, zooanthroponosis, or both?—Authors' reply. Lancet Microbe 1:e241. https://doi.org/10.1016/S2666-5247(20)30145-2.

7. Saïd-Salim B, Mostowy S, Kristof AS, Behr MA. 2006. Mutations in *Mycobacterium tuberculosis* Rv0444c, the gene encoding anti-SigK, explain high level expression of MPB70 and MPB83 in *Mycobacterium bovis*. Mol Microbiol 62:1251–1263. https://doi.org/10.1111/j.1365-2958.2006.05455.x.

8. Mostowy S, Inwald J, Gordon S, Martin C, Warren R, Kremer K, Cousins D, Behr MA. 2005. Revisiting the evolution of *Mycobacterium bovis*. J Bacteriol 187:6386–6395. https://doi.org/10.1128/JB.187.18.6386-6395.2005.

9. Singer VL, Jones LJ, Yue ST, Haugland RP. 1997. Characterization of PicoGreen reagent and development of a fluorescence-based solution assay for double-stranded DNA quantitation. Anal Biochem 249:228–238. https://doi.org/10.1006/abio.1997.2177.

10. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. Genome Biol 16:294. https://doi.org/10.1186/s13059-015-0849-0.

11. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.