AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequence of *Mycolicibacterium hassiacum* DSM 44199

Mercedes Sánchez,[a] Alba Blesa,[b] Eva Sacristán-Horcajada,[a] José Berenguer[a]

[a]Centro de Biología Molecular Severo Ochoa, Universidad Autónoma de Madrid-Consejo Superior de Investigaciones Científicas, Madrid, Spain
[b]Universidad Francisco de Vitoria, Madrid, Spain

ABSTRACT *Mycolicibacterium hassiacum* is the most thermophilic of all the mycobacteria. A partial sequence based on Illumina technology of around 5 Mbp was published in 2012. Here, we report the 5,269,097-bp complete genome sequence assembled into a single circular chromosome.

Mycolicibacterium hassiacum (1), described in 1997 as *Mycobacterium hassiacum* (2), is characterized by its ability to grow at temperatures up to 65°C (2). Its fast growth and cell yield and the conservation and thermostability of most genus-specific enzymes (3, 4) make this strain an attractive laboratory mycobacterial model. Also, its isolation from clinical samples supports the conservation of virulence traits of its pathogenic relatives (5). *Mycolicibacterium hassiacum* DSM 44199 was donated by N. Empadinhas, whose group carried out a partial sequencing of the strain (6). In that work, the sequence was divided into 169 contigs, with an estimated genome size of 5 Mbp. Here, we present the whole-genome sequence obtained by PacBio technology, with further refinement using original Illumina reads. The strain was grown at 50°C under stirring (200 rpm) in Middlebrook 7H9 broth (BD Difco) with 0.15% (vol/vol) Tween 80 (Sigma-Aldrich) and 0.2% (vol/vol) glycerol as a carbon source. The cells grown up to an optical density at 600 nm (OD_{600}) of 0.8 were harvested and washed by centrifugation, and the total DNA was isolated by a combination of mechanical breaking with glass beads (Sigma-Aldrich) and phenol-chloroform extraction (7). Twenty-kilobase DNA fragments were selected using BluePippin equipment (Sage Science), and a library was prepared according to Pacific Biosciences protocols. The Pacific Biosciences RS II equipment with P6-C4 chemistry (polymerase attachment by binding P6 kit) was used, with a recording time of 360 min.

A total of 69,152 reads, with an average length of 17,024 bp and a mean coverage of 223-fold, were obtained. The HGAP version 3 software (SMRT Analysis software version 2.3.0; Pacific Biosciences) was used with its default settings for the assembly. This produced a single linear contig of 5,282,087 bp. Its circularization was carried out using the Minimus2 software and further corrected with the RS_Resequencing.1 software (SMRT Analysis version 2.3.0). A single circular contig of 5,268,611 bp was obtained after this procedure.

The PROKKA software version 1.12 (8) was employed on this circular chromosome for automatic coding sequence (CDS) annotation, with parameters set for gene code 11 and using *Mycobacterium* genus-specific databases for assignment of putative function. A manual inspection of the annotation obtained revealed that many conserved proteins within the *Mycobacterium* genus appeared to be truncated due to the inclusion of an additional base in some homopolymers of the assembled genome, including proteins essential for viability, such as the β'-subunit of the RNA polymerase, the chromosome partition protein Smc, the FtsK ATPase involved in DNA segregation, and the RecO protein involved in DNA repair. Due to this, the PacBio-based sequence was

Citation Sánchez M, Blesa A, Sacristán-Horcajada E, Berenguer J. 2019. Complete genome sequence of *Mycolicibacterium hassiacum* DSM 44199. Microbiol Resour Announc 8:e01522-18. https://doi.org/10.1128/MRA.01522-18.

Editor John J. Dennehy, Queens College

Address correspondence to José Berenguer, jberenguer@cbm.csic.es.

Received 23 November 2018
Accepted 17 December 2018
Published 24 January 2019

further refined by using the Illumina reads available online (6) using the PacBio-utilities and Pilon (version 1.22) programs (9), configured to fix only the indels detected. This allowed us to correct 483 positions in the sequence assembled, corresponding to homopolymers for which PacBio reads included an additional base. Once corrected, the initially truncated essential genes were translated to proteins of the size and sequence homology expected for their homologues. The final genome of 5,269,097 bp encodes 5,333 putative proteins, 53 tRNAs, and 2 rRNA clusters. The mean GC content was 69.29% (34.55% Cs and 34.74% Gs), with a coding density of 91.5%.

The KEGG Automatic Annotation Server and its internal BlastKOALA annotation tool (10) led to the identification of 236 pathways, of which 61 contained more than 10 genes, including 20 genes putatively involved in the metabolism of steroids (KEGG steroid degradation pathway 00984). As this bacterium metabolizes different kinds of sterols from the environment, the identification of enzymes of this pathway constitutes a first step toward the isolation of specific mutants for relevant biotechnological bioconversions involving these molecules.

Data availability. The raw reads and the annotated assembly of the whole genome of *Mycolicibacterium hassiacum* have been deposited in NCBI under the accession number GCA_900603025.

ACKNOWLEDGMENTS

This work was supported by grants RTC2014-14391-1 and BIO2016-77031-R from the Spanish Ministry of Economy and Competitiveness and grant 685474 from the H2020 research and innovation program of the European Union. An institutional grant from Fundación Ramón Areces to the Centro de Biología Molecular Severo Ochoa (CBMSO) is acknowledged.

The sequencing service was provided by the Norwegian Sequencing Centre (www.sequencing.uio.no), a national technology platform hosted by the University of Oslo and supported by the "Functional Genomics" and "Infrastructure" programs of the Research Council of Norway and the Southeastern Regional Health Authorities. The next-generation sequencing (NGS) data analysis service was provided by the Genomics and NGS Core Facility at the Centro de Biología Molecular Severo Ochoa (CBMSO, CSIC-UAM) which is part of the CEI UAM+CSIC, Madrid, Spain (http://www.cbm.uam.es/genomica).

REFERENCES

1. Gupta RS, Lo B, Son J. 2018. Phylogenomics and comparative genomic studies robustly support division of the genus *Mycobacterium* into an emended genus *Mycobacterium* and four novel genera. Front Microbiol 9:67. https://doi.org/10.3389/fmicb.2018.00067.
2. Tortoli E, Reischl U, Besozzi G, Emler S. 1998. Characterization of an isolate belonging to the newly described species *Mycobacterium hassiacum*. Diagn Microbiol Infect Dis 30:193–196. https://doi.org/10.1016/S0732-8893(97)00242-3.
3. Masood R, Sharma YK, Venkitasubramanian TA. 1985. Metabolism of mycobacteria. J Biosci 7:421–431. https://doi.org/10.1007/BF02716802.
4. Donova MV. 2017. Steroid bioconversions. Humana Press, New York, NY, p 1–13.
5. Schröder K-H, Naumann L, Kroppenstedt RM, Reischl U. 1997. Mycobacterium hassiacum sp. nov., a new rapidly growing thermophilic mycobacterium. Int J Syst Bacteriol 47:86–91. https://doi.org/10.1099/00207713-47-1-86.
6. Tiago I, Maranha A, Mendes V, Alarico S, Moynihan PJ, Clarke AJ, Macedo-Ribeiro S, Pereira PJB, Empadinhas N. 2012. Genome sequence of *Mycobacterium hassiacum* DSM 44199, a rare source of heat-stable mycobacterial proteins. J Bacteriol 194:7010–7011. https://doi.org/10.1128/JB.01880-12.
7. Jacobs WR, Jr, Kalpana GV, Cirillo JD, Pascopella L, Snapper SB, Udani RA, Jones W, Barletta RG, Bloom BR. 1991. Genetic systems for mycobacteria. Methods Enzymol 204:537–555. https://doi.org/10.1016/0076-6879(91)04027-L.
8. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
9. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, 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.
10. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 428:726–731. https://doi.org/10.1016/j.jmb.2015.11.006.