




Draft Genome Sequence of *Mycolicibacterium smegmatis* VKM Ac-1171 Contains Full Set of Sterol Catabolic Genes

 Dmitry V. Dovbnya,^a Eugeny Y. Bragin,^a Tanya V. Ivashina,^a Marina V. Donova^a

^aG. K. Skryabin Institute of Biochemistry and Physiology of Microorganisms, Pushchino Center for Biological Research, Russian Academy of Sciences, Pushchino, Russia

ABSTRACT *Mycolicibacterium smegmatis* VKM Ac-1171 is a saprotrophic bacterium that was isolated several decades ago and is deposited in microbial collections around the world. We report here a draft genome sequence of the strain. Annotation of the genome revealed the presence of a complete set of genes related to the sterol catabolic pathway.

M*ycolicibacterium smegmatis* VKM Ac-1171 (NCIMB 8548, CCM 2067) was originally deposited in ATCC decades ago as *Mycobacterium butyricum* 362 and later re-identified as *Mycobacterium smegmatis* (1). Recently, *Mycobacterium smegmatis* and closely related fast-growing species have been reclassified as *Mycolicibacterium* based on their phylogenomic differences from the “tuberculosis-like” clade (2, 3). The strain has previously been used in various studies as a nonpathogenic model microorganism (4–9), to validate the DNA isolation method (10), but its genome sequence was hitherto unreported. Here, we present a draft genome of *M. smegmatis* VKM Ac-1171.

The strain was obtained from the All-Russian Collection of Microorganisms VKM (<http://www.vkm.ru>) and cultured aerobically at 37°C to early stationary phase in MYCB broth (11) supplemented with 15 g/L Tween 80 and 15 g/L glycine.

Genomic DNA was extracted as described (12) with modifications. Briefly, cells from 10 mL broth were subjected to sequential treatment with lysozyme (20 min, 37°C), SDS, proteinase K (1.0 h, 56°C), and RNase A (30 min, 37°C). Then, the DNA was purified with phenol-chloroform.

The Illumina sequencing library construction was made by KAPA DNA library preparation kit for Illumina and KAPA dual-indexed adapter kit (Kapa Biosystems). Genome sequencing was performed by Illumina HiSeq 2000 with HiSeq SBS kit v3. For adapter and quality trimming, Trimmomatic 0.39 (13) with the settings ILLUMINACLIP:TruSeq3-PE:2:30:10:2, LEADING:3, TRAILING:3, MINLEN:50, and a self-written program in Perl language (<https://github.com/BraginE/bioinfo>) were applied. *De novo* genome assembly was made with the Ray 2.3.1 program (14); the k-mer length was 31. Genome was annotated with Prokaryotic Genome Annotation Pipeline (PGAP) (15). For average nucleotide identity (ANI), the ANI calculator (16) was applied. Default parameters were used for all software unless mentioned otherwise.

Sequencing resulted in 19,143,437 paired-end reads (2 × 100). The genome assembly generated 96 contigs with 7,600,730-bp total length (genome coverage, 44×; N_{50} length, 199,025 bp; GC content, 67.5%).

Among *M. smegmatis* strains with known genome sequences, Ac-1171 showed the highest similarity to *M. smegmatis* Nishi, whereas the ANI value between Ac-1171 and *M. smegmatis* mc² 155 was lower (Table 1).

The size of the Ac-1171 genome is approximately 600,000 bp bigger than the genomes of other *M. smegmatis* strains. The Ac-1171 genome contains 7,163 protein-coding genes, 57 RNA-coding genes (2, 2, 2, 48 and 3 genes coding for 5S rRNA, 16S rRNA, 23S rRNA, tRNA, and noncoding RNA, respectively), and 167 pseudogenes. The

Editor Frank J. Stewart, Montana State University

Copyright © 2022 Dovbnya et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Dmitry V. Dovbnya, anagoge@rambler.ru.

The authors declare no conflict of interest.

Received 28 July 2022

Accepted 28 October 2022

Published 10 November 2022

TABLE 1 Average nucleotide identities between genomes of VKM Ac-1171 and other *Mycolicibacterium smegmatis* strains

Strain	GenBank accession no.	Genome length (bp)	ANI value
<i>M. smegmatis</i> mc ² 155	CP009494	6,988,269	98.94
<i>M. smegmatis</i> strain Rabinowitchi	CP080272	7,061,747	99.01
<i>M. smegmatis</i> strain Nishi	CP080273	7,010,278	99.09
<i>M. smegmatis</i> strain Jucho	CP080274	6,895,172	98.95

strain Ac-1171 possesses a complete set of key genes of steroid catabolism, thus suggesting the ability for the full sterol degradation.

Modification of sterol catabolic pathways in some species of *Mycolicibacterium*, such as *M. neoaurum* (17), *M. fortuitum* (18), and *M. smegmatis* mc² 155 (19), has become the basis for production of pharmaceutical steroid precursors. The strain *M. smegmatis* VKM Ac-1171 is promising for the engineering of novel microbial producers for steroid biotechnology.

Data availability. The genome sequences have been deposited in NCBI GenBank database under accession number [JAMZOD000000000](https://www.ncbi.nlm.nih.gov/assembly/JAMZOD000000000/). The BioSample and BioProject accession numbers are [SAMN28113943](https://www.ncbi.nlm.nih.gov/biosample/SAMN28113943/) and [PRJNA835822](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA835822/), respectively. The draft genome raw data are available in the Sequence Read Archive (SRA) under accession number [SRR19122810](https://www.ncbi.nlm.nih.gov/sra/SRR19122810/).

ACKNOWLEDGMENT

The research was supported by RSF (project No. 21-64-00024). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

REFERENCES

- Bojalil LF, Cerbon J, Trujillo A. 1962. Adansonian classification of mycobacteria. *J Gen Microbiol* 28:333–346. <https://doi.org/10.1099/00221287-28-2-333>.
- 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>.
- Oren A, Garrity G. 2020. List of new names and new combinations previously effectively, but not validly, published. *Int J Syst Evol Microbiol* 70: 1443–1446. <https://doi.org/10.1099/ijsem.0.003991>.
- White AJ, Snow GA. 1969. Isolation of mycobactins from various mycobacteria. The properties of mycobactins S and H. *Biochem J* 111:785–792. <https://doi.org/10.1042/bj1110785>.
- Ratledge C, Ewing M. 1996. The occurrence of carboxymycobactin, the siderophore of pathogenic mycobacteria, as a second extracellular siderophore in *Mycobacterium smegmatis*. *Microbiology* 142:2207–2212. <https://doi.org/10.1099/13500872-142-8-2207>.
- Adilakshmi T, Ayling PD, Ratledge C. 2000. Mutational analysis of a role for salicylic acid in iron metabolism of *Mycobacterium smegmatis*. *J Bacteriol* 182:264–271. <https://doi.org/10.1128/JB.182.2.264-271.2000>.
- Nagachar N, Ratledge C. 2010. Roles of *trpE2*, *entC* and *entD* in salicylic acid biosynthesis in *Mycobacterium smegmatis*: salicylate biosynthesis in *M. smegmatis*. *FEMS Microbiol Lett* 308:159–165. <https://doi.org/10.1111/j.1574-6968.2010.02004.x>.
- Dulger B, Ugurlu E, Aki C, Suerdem TB, Camdeviren A, Tazeler G. 2005. Evaluation of antimicrobial activity of some endemic *Verbascum*., *Sideritis*., and *Stachys*. species from Turkey. *Pharm Biol* 43:270–274. <https://doi.org/10.1080/13880200590928861>.
- Yildiz M, Ünver H, Dülger B, Erdener D, Ocak N, Erdönmez A, Durlu TN. 2005. Spectroscopic study, antimicrobial activity and crystal structures of N-(2-hydroxy-5-nitrobenzalidene)4-aminomorpholine and N-(2-hydroxy-1-naphthylidene)4-aminomorpholine. *J Mol Struct* 738:253–260. <https://doi.org/10.1016/j.molstruc.2004.10.029>.
- Talip AB, Snelling WJ, Sleator RD, Lowery C, Dooley JSG. 2018. A rapid and sensitive system for recovery of nucleic acids from *Mycobacteria* sp. on archived glass slides. *BMC Microbiol* 18:196. <https://doi.org/10.1186/s12866-018-1335-0>.
- Dovbnya D, Khomutov S, Kollerov V, Donova MV. 2017. Obtaining of 11 α -hydroxyandrost-4-ene-3,17-dione from natural sterols, p 259–269. *In* Barredo J-L, Herráiz I (ed), *Microbial steroids*. Springer New York, New York, NY.
- Bragin E, Shtratnikova V, Dovbnya DV, Schelkunov MI, Pekov Y, Malakho SG, Egorova OV, Ivashina TV, Sokolov SL, Ashapkin VV, Donova MV. 2013. Comparative analysis of genes encoding key steroid core oxidation enzymes in fast-growing *Mycobacterium* spp. strains. *J Steroid Biochem Mol Biol* 138:41–53. <https://doi.org/10.1016/j.jsbmb.2013.02.016>.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Boisvert S, Raymond F, Godzaridis É, Laviolette F, Corbeil J. 2012. Ray Meta: scalable de novo metagenome assembly and profiling. *Genome Biol* 13:R122. <https://doi.org/10.1186/gb-2012-13-12-r122>.
- Tatsova T, DiCuccio M, Badretdin A, Chetverin 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>.
- Yoon S-H, Ha S, Lim J, Kwon S, Chun J. 2017. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* 110:1281–1286. <https://doi.org/10.1007/s10482-017-0844-4>.
- Zhao A, Zhang X, Li Y, Wang Z, Lv Y, Liu J, Alam M, Xiong W, Xu J. 2021. *Mycolicibacterium* cell factory for the production of steroid-based drug intermediates. *Biotechnol Adv* 53:107860. <https://doi.org/10.1016/j.biotechadv.2021.107860>.
- Bragin EY, Shtratnikova VY, Schelkunov MI, Dovbnya DV, Donova MV. 2019. Genome-wide response on phytosterol in 9-hydroxyandrostenedione-producing strain of *Mycobacterium* sp. VKM AC-1817D. *BMC Biotechnol* 19:39. <https://doi.org/10.1186/s12896-019-0533-7>.
- Galán B, Uthia I, García-Fernández E, Martínez I, Bahillo E, de la Fuente JL, Barredo JL, Fernández-Cabezón L, García JL. 2017. *Mycobacterium smegmatis* is a suitable cell factory for the production of steroidic synthons. *Microb Biotechnol* 10:138–150. <https://doi.org/10.1111/1751-7915.12429>.