

Complete Genome Sequence of Sterol-Transforming *Mycobacterium neoaurum* Strain VKM Ac-1815D

Victoriya Y. Shtratnikova,^{a,c} Eugeny Y. Bragin,^b Dmitry V. Dovbnaya,^b Yury A. Pekov,^c Mikhail I. Schelkunov,^c Nicolai Strizhov,^b Tanya V. Ivashina,^b Vasily V. Ashapkin,^d Marina V. Donova^b

Center of Innovations and Technologies Biologically Active Compounds and Their Applications, Moscow, Russia^a; Institute of Biochemistry and Physiology of Microorganisms, Russian Academy of Sciences, Pushchino, Russia^b; Department of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia^c; Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia^d

***Mycobacterium neoaurum* strain VKM Ac-1815D produces 4-androstene-3,17-dione as a major compound from phytosterols. Here, we report the complete genome sequence of the strain. The genome consists of a single circular 5,438,190-bp chromosome, with a G+C content of 66.88%, containing 5,318 putative open reading frames (ORFs), 46 tRNAs, and 6 rRNAs. Arrays of cholesterol metabolism genes are randomly clustered throughout the chromosome.**

Received 10 December 2013 Accepted 19 December 2013 Published 16 January 2014

Citation Shtratnikova VY, Bragin EY, Dovbnaya DV, Pekov YA, Schelkunov MI, Strizhov N, Ivashina TV, Ashapkin VV, Donova MV. 2014. Complete genome sequence of sterol-transforming *Mycobacterium neoaurum* strain VKM Ac-1815D. *Genome Announc.* 2(1):e01177-13. doi:10.1128/genomeA.01177-13.

Copyright © 2014 Shtratnikova et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](http://creativecommons.org/licenses/by/3.0/).

Address correspondence to Nicolai Strizhov, strizhov@ibpm.pushchino.ru.

A number of mycobacterial strains with an origin in soil effectively oxidize cholesterol and phytosterols (1, 2). *Mycobacterium* sp. strain VKM Ac-1815D is of particular interest because it is capable of synthesizing the valuable steroid precursor 4-androstene-3,17-dione as a major product from sitosterol (3). Based on phylogenetic analysis of 16S rRNA sequences (4) and chemotaxonomy, the strain was reclassified as *Mycobacterium neoaurum* VKM Ac-1815D.

Two genomic DNA fragment libraries of the strain were generated and read on a Genome Analyzer IIX (paired-end 72-nucleotide reads) and on HiSeq 2000 (paired-end 100-nucleotide reads) as per the protocols of the manufacturer (Illumina), with 500 to 800× coverage. The short-read library contained DNA fragments of 400-bp mean length prepared with a DNA sample preparation kit (New England BioLabs) after digestion by NEB-Next dsDNA Fragmentase. The mate-pair library with 2,000- to 10,000-bp-long fragments was created with the techniques recommended by the supplier of the Nextera mate-pair sample preparation kit (Illumina).

The short reads and mate-pair reads obtained in a FASTQ Illumina 1.5+ format were used for the *de novo* genome assembly with Velvet 1.2 (5) and SPAdes 2.5 (6). Contig sequences up to 575 kbp in length were produced (30 to 40 contigs depending on the assembler used). The circularization of the genome, gaps in scaffolds, and ambiguities were resolved by primer walking to the edges of the contigs, followed by Sanger sequencing of the PCR-amplified DNA fragments. All generated reads were assembled in a single sequence representing the complete bacterial chromosome. Genome assembly validation was performed by mapping of the reads with the Burrows-Wheeler Aligner (BWA) program (<http://bio-bwa.sourceforge.net/>).

The genome consists of a single circular 5,438,190-bp chromosome with a G+C content of 66.88% and encodes 46 tRNAs and 6 rRNAs (5S, 23S, and 16S in two copies each located in two rRNA

loci). Annotation of the genome by the Internet service xBASE (<http://www.xbase.ac.uk/annotation>) revealed 5,318 open reading frames (ORFs) with *Mycobacterium tuberculosis* H37Rv used as a reference strain and 5,192 ORFs with *Mycobacterium smegmatis* mc² 155 as a reference strain; i.e., the numbers of putative ORFs in the complete genome are greater than those found in the draft sequence (4). A preliminary analysis of the genome sequence shows that arrays of genes involved in cholesterol metabolism (side chain degradation, central pathway, and transport) are distributed throughout the bacterial chromosome, randomly located in different clusters.

The reported genome sequence will contribute not only to the elucidation of cholesterol metabolism in actinobacteria but also to the development of new biotechnological applications for mycobacterial conversion of important bioactive steroids.

Nucleotide sequence accession number. The complete genome sequence has been deposited in GenBank under the accession no. [CP006936](https://ncbi.nlm.nih.gov/nucl/CP006936).

ACKNOWLEDGMENTS

The work was partly supported by the Russian Foundation for Basic Research (grants no. RFBR 13-04-01551-a and 12-08-00895-a).

Maria Logacheva and Alexey Penin from the Laboratory of Evolutionary Genomics, Department of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia, are greatly acknowledged for sequencing with HiSeq 2000.

REFERENCES

- García JL, Uhía I, Galán B. 2012. Catabolism and biotechnological applications of cholesterol degrading bacteria. *Microb. Biotechnol.* 5:679–699. <http://dx.doi.org/10.1111/j.1751-7915.2012.00331.x>.
- Donova MV, Egorova OV. 2012. Microbial steroid transformations: current state and prospects. *Appl. Microbiol. Biotechnol.* 94:1423–1447. <http://dx.doi.org/10.1007/s00253-012-4078-0>.
- Donova MV, Gulevskaya SA, Dovbnaya DV, Puntus IF. 2005. *Mycobac-*

- terium* sp. mutant strain producing 9 α -hydroxyandrostenedione from sitosterol. *Appl. Microbiol. Biotechnol.* 67:671–678. <http://dx.doi.org/10.1007/s00253-004-1808-y>.
4. Bragin EY, Shtratnikova VY, Dovbnya DV, Schelkunov MI, Pekov YA, 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. <http://dx.doi.org/10.1016/j.jsbmb.2013.02.016>.
 5. 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>.
 6. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev M, 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>.