



**HAL**  
open science

## Complete genome sequence of *Mycobacterium* sp. Strain 3519A

Amar Bouam, Anthony Levasseur, Maryline Bonnet, Laurence Borand,  
Charles van Goethem, Michel Drancourt, Sylvain Godreuil

► **To cite this version:**

Amar Bouam, Anthony Levasseur, Maryline Bonnet, Laurence Borand, Charles van Goethem, et al..  
Complete genome sequence of *Mycobacterium* sp. Strain 3519A. *Genome Announcements*, 2018, 6  
(4), pp.1565 - 1582. 10.1128/genomeA.01565-17 . hal-01889914

**HAL Id: hal-01889914**

**<https://hal.science/hal-01889914v1>**

Submitted on 6 Jan 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



# Complete Genome Sequence of *Mycobacterium* sp. Strain 3519A

Amar Bouam,<sup>a</sup> Anthony Levasseur,<sup>a</sup> Maryline Bonnet,<sup>b</sup> Laurence Borand,<sup>c</sup> Charles Van Goethem,<sup>d</sup> Michel Drancourt,<sup>a</sup> Sylvain Godreuil<sup>e</sup>

<sup>a</sup>Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes CNRS 6 UMR 7278, IRD198, INSERM U1095, Institut Hospitalier Universitaire Méditerranée-Infection, Aix-Marseille Université, Marseille, France

<sup>b</sup>UMI233-TransVIHMI/INSERM U1175, Institut de Recherche pour le Développement (IRD) and University of Montpellier, Montpellier, France

<sup>c</sup>Epidemiology and Public Health, Institut Pasteur du Cambodge, Phnom Penh, Cambodia

<sup>d</sup>Département Biopathologie Cellulaire et Tissulaire des Tumeurs, Centre Hospitalier Universitaire (CHU) de Montpellier, Montpellier, France

<sup>e</sup>Laboratoire de Bactériologie, MIVEGEC, Centre Hospitalier Universitaire (CHU) de Montpellier, Université de Montpellier, Montpellier, France

**ABSTRACT** *Mycobacterium* sp. strain 3519A is a nontuberculous mycobacterium isolated from sputum from a Cambodian patient with a pulmonary infection. We report here the first complete 7.3-Mbp-long genome sequence of *Mycobacterium* sp. 3519A with 66.35% GC content, encoding 7,029 protein-coding genes, 50 tRNAs, and 5 rRNA genes.

*Mycobacterium* sp. strain 3519A is a rapidly growing scotochromogenic acid-fast bacillus isolated from the sputum of a Cambodian patient with a pulmonary infection. We sequenced and analyzed the whole-genome sequence of strain 3519A in order to describe its genomic content and to determine its phylogenetic relationships for facilitating its detection and identification.

Strain 3519A was cultured on Middlebrook 7H11 agar supplemented with 10% (vol/vol) oleic acid-albumin-dextrose-catalase (Becton, Dickinson, Sparks, MD, USA). The total DNA of strain 3519A<sup>T</sup> was extracted on the EZ1 biorobot (Qiagen) with an EZ1 DNA tissue kit under a 50- $\mu$ l elution volume. Total DNA was quantified by a Qubit assay with a high-sensitivity kit (Life Technologies, Inc., Carlsbad, CA, USA) to be 36.2 ng/ $\mu$ l. Genomic DNA was sequenced on a MiSeq platform (Illumina, Inc., San Diego, CA, USA) using both paired-end and mate pair techniques. The index representation for strain 3519A was determined to be 5.51%. A total of 1,116,202 paired-end reads were filtered per the read qualities. The reads were then assembled using the SPAdes software (1). Contigs obtained were combined by use of SSPACE (2) and assisted by manual finishing and GapFiller (3). This yielded a 7,306,349-bp draft genome sequence with a 66.35% GC content, composed of 14 scaffolds and 14 contigs. Open reading frames (ORFs) were predicted using Prodigal (4) with default parameters. Functional annotation was achieved using a BLASTP search against the GenBank database (*E* value, 0.001; coverage, 0.7 $\times$ ; 30% identity) (5) and the Clusters of Orthologous Groups (COGs) database (6). In the case of no hit being found, a second round was done against the NCBI nonredundant protein sequence (NR) database using BLASTP with an *E* value of 1  $\times$  0.001, coverage of 0.7 $\times$ , and 30% identity. Noncoding genes and miscellaneous features were predicted using RNAMmer (7), ARAGORN (8), Rfam (9), Pfam (10), and Infernal (11). Of the 7,084 predicted genes, 7,029 were protein-coding genes and 55 were RNAs (namely, 3 5S rRNAs, 1 16S rRNA, 1 23S rRNA, and 50 tRNAs). A total of 5,355

**Received** 14 December 2017 **Accepted** 21 December 2017 **Published** 25 January 2018

**Citation** Bouam A, Levasseur A, Bonnet M, Borand L, Van Goethem C, Drancourt M, Godreuil S. 2018. Complete genome sequence of *Mycobacterium* sp. strain 3519A. Genome Announc 6:e01565-17. <https://doi.org/10.1128/genomeA.01565-17>.

**Copyright** © 2018 Bouam 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 Michel Drancourt, [Michel.Drancourt@univ-amu.fr](mailto:Michel.Drancourt@univ-amu.fr).

genes (76.18%) were assigned putative functions (by a COG or NR BLAST search), 164 genes were identified as ORFans (ORFs with no detectable homology to other ORFs in the database) (2.33%), and 1,295 genes (18.42%) were annotated as hypothetical proteins. The genome of strain 3519A was further incorporated into *in silico* DNA-DNA hybridization (DDH) (12) with genomes selected on the basis of 16S rRNA gene sequence proximity. The DDH values were estimated using the Genome-to-Genome Distance Calculator (GGDC) version 2.0 online tool (13). This analysis yielded a DDH value of 22.60% with *Mycobacterium rutilum* DSM 45405 (GenBank accession number NZ\_LT629971), 22.40% with *Mycobacterium holsaticum* M7 (MIGZ00000000), 21.70% with *Mycobacterium rhodesiae* NBB3 (NC\_016604), 21.20% with *Mycobacterium chubuense* NBB4 (CP003053), 20.90% with *Mycobacterium mageritense* DSM 44476 (CCBF000000000), and 20% with *Mycobacterium pallens* PYR15 (CP023435).

These data indicate that *Mycobacterium* sp. strain 3519A is related to the *Mycobacterium fortuitum* complex of mycobacteria.

**Accession number(s).** The *Mycobacterium* sp. strain 3519A genome sequence has been deposited at EMBL under the accession number [OESG00000000](https://www.ebi.ac.uk/EMBL/nuclseq/EMBL/acc/OESG00000000).

## REFERENCES

1. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
2. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27:578–579. <https://doi.org/10.1093/bioinformatics/btq683>.
3. Boetzer M, Pirovano W. 2012. Toward almost closed genomes with GapFiller. *Genome Biol* 13:R56. <https://doi.org/10.1186/gb-2012-13-6-r56>.
4. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>.
5. Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. 2012. GenBank. *Nucleic Acids Res* 40:D48–D53. <https://doi.org/10.1093/nar/gkr1202>.
6. Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* 28:33–36. <https://doi.org/10.1093/nar/28.1.33>.
7. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35:3100–3108. <https://doi.org/10.1093/nar/gkm160>.
8. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 32:11–16. <https://doi.org/10.1093/nar/gkh152>.
9. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. 2003. Rfam: an RNA family database. *Nucleic Acids Res* 31:439–441. <https://doi.org/10.1093/nar/gkg006>.
10. Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD. 2012. The Pfam protein families database. *Nucleic Acids Res* 40:D290–D301. <https://doi.org/10.1093/nar/gkr1065>.
11. Nawrocki EP, Kolbe DL, Eddy SR. 2009. Infernal 1.0: inference of RNA alignments. *Bioinformatics* 25:1335–1337. <https://doi.org/10.1093/bioinformatics/btp157>.
12. Richter M, Rosselló-Móra R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A* 106:19126–19131. <https://doi.org/10.1073/pnas.0906412106>.
13. Auch AF, von Jan M, Klenk HP, Göker M. 2010. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci* 2:117–134. <https://doi.org/10.4056/sigs.531120>.