

Draft Genome Sequence of *Mycobacterium triplex* DSM 44626

Mohamed Sassi, Olivier Croce, Catherine Robert, Didier Raoult, Michel Drancourt

Aix Marseille Université, URMITE, Marseille, France

We announce the draft genome sequence of *Mycobacterium triplex* strain DSM 44626, a nontuberculosis species responsible for opportunistic infections. The genome described here is composed of 6,382,840 bp, with a G+C content of 66.57%, and contains 5,988 protein-coding genes and 81 RNA genes.

Received 7 May 2014 Accepted 13 May 2014 Published 29 May 2014

Citation Sassi M, Croce O, Robert C, Raoult D, Drancourt M. 2014. Draft genome sequence of *Mycobacterium triplex* DSM 44626. *Genome Announc.* 2(3):e00499-14. doi: 10.1128/genomeA.00499-14.

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

Address correspondence to Michel Drancourt, michel.drancourt@univ-amu.fr.

Mycobacterium triplex was described on the basis of a unique mycolic-acid pattern and a distinctive 16S rRNA gene hypervariable region (1). Phylogenetic analysis confirmed close relationships with *Mycobacterium simiae*, *Mycobacterium genavense*, *Mycobacterium lentiflavum*, and *Mycobacterium sherrisii* (2, 3). *M. triplex* has been initially isolated from sputum (4), lymph nodes (5), and cerebrospinal fluid specimens (1, 6). Only six cases of infection have been reported in immunocompromised patients, including HIV-infected patients (1) and a liver transplant patient (7).

We analyzed the whole-genome sequence of *M. triplex* to facilitate the elucidation of its relationships within the *M. simiae* complex and the design of tools for its advanced detection and identification.

Genomic DNA was isolated from *M. triplex* strain DSMZ 44626 grown in MGIT Middlebrook liquid culture (Becton Dickinson, Sparks, MD) at 37°C. DNA was sequenced using three high-throughput next-generation sequencing (NGS) technologies: Roche 454 (Roche Diagnostics Corporation, Indianapolis, IN) (8), SOLiD version 4 (Life Technologies, Carlsbad, CA), and MiSeq Illumina (Illumina, Inc., San Diego, CA). Two Roche 454 libraries, a 3.6-kb paired-end and a 1.6-kb shotgun XL+, were constructed, loaded on a picotiter (PTP) plate, and sequenced with the Roche-GS FLX Titanium sequencing kit XLR70. The two runs yielded 90.13 Mb with 265,087 passed filters and an average 348-bp length. The bar-coded paired-end SOLiD library generated 1,338,576 reads of 50- × 35-bp-length reads. Finally, a paired-end Nextera library, fragmented at 942 bp and sequenced on MiSeq in 2 × 151 bp, yielded 269,610 reads with an indexing of 2.15% on the flow cell.

The reads that issued from these three sequencing technologies were first assembled separately. The 454 reads were assembled into contigs and scaffolds using Newbler version 2.8 (Roche). Illumina reads, trimmed using Trimmomatic (9), were assembled using the Spades software (10, 11). Contigs obtained were combined by using SSPACE (12) and Opera v. 1.2 (13) software; GapFiller v. 1.10 (14) helped to reduce the set. Some manual refinements using CLC Genomics v. 5 software (CLC bio, Aarhus, Denmark) improved the genome.

These analyses yielded four scaffolds of 21 contigs containing

6,379,625 bp and an estimated size, including gaps, of 6,382,840 bp, for a 66.57% G+C content. Noncoding genes and miscellaneous features were predicted using RNAmmer (15), ARAGORN (16), Rfam (17), and PFAM (18). Open reading frames (ORFs) were predicted using Prodigal (19) and functional annotation was achieved using BLASTP against the GenBank database (20) and the Clusters of Orthologous Groups (COGs) database (21, 22). The genome was shown to encode at least 81 predicted RNAs, including 3 rRNAs in a single operon, 58 tRNAs, 1 transfer-messenger RNA (tmRNA), and 19 miscellaneous RNAs. Also, 5,988 genes represented a coding capacity of 5,867,145 bp for a 91.9% coding percentage. Among these genes, 765 (12.77%) were found to encode putative proteins and 1,011 (16.88%) were assigned hypothetical proteins. Moreover, 5,940 genes matched a least one sequence in the COGs database with BLASTP default parameters.

Nucleotide sequence accession numbers. The *M. triplex* strain DSM 44626 genome sequence has been deposited at ENA under the accession numbers [CCAU010000001](https://ena.ebi.ac.uk/ena/record/CCAU010000001) to [CCAU010000021](https://ena.ebi.ac.uk/ena/record/CCAU010000021).

ACKNOWLEDGMENT

This study was financially supported by URMITE, IHU Méditerranée Infection, Marseille, France.

REFERENCES

1. Floyd MM, Guthertz LS, Silcox VA, Duffey PS, Jang Y, Desmond EP, Crawford JT, Butler WR. 1996. Characterization of an SAV organism and proposal of *Mycobacterium triplex* sp. nov. *J. Clin. Microbiol.* 34: 2963–2967.
2. Cingolani A, Sanguinetti M, Antinori A, Larocca LM, Ardito F, Posteraro B, Federico G, Fadda G, Ortona L. 2000. Brief report: disseminated mycobacteriosis caused by drug-resistant *Mycobacterium triplex* in a human immunodeficiency virus-infected patient during highly active antiretroviral therapy. *Clin. Infect. Dis.* 31:177–179. <http://dx.doi.org/10.1086/313903>.
3. Van Ingen J, Tortoli E, Selvarangan R, Coyle MB, Crump JA, Morrissey AB, Dekhuijzen PNR, Boeree MJ, van Soolingen D. 2011. *Mycobacterium sherrisii* sp. nov., a slow-growing nonchromogenic species. *Int. J. Syst. Evol. Microbiol.* 61:1293–1298. <http://dx.doi.org/10.1099/ijs.0.024752-0>.
4. Piersimoni C, Zitti P, Mazzarelli G, Mariottini A, Nista D, Zallocco D. 2004. *Mycobacterium triplex* pulmonary disease in immunocompetent host. *Emerg. Infect. Dis.* 10:1859–1862. <http://dx.doi.org/10.3201/eid1010.040402>.

5. Hazra R, Floyd MM, Sloutsky A, Husson RN. 2001. Novel mycobacterium related to *Mycobacterium triplex* as a cause of cervical lymphadenitis. *J. Clin. Microbiol.* 39:1227–1230. <http://dx.doi.org/10.1128/JCM.39.4.1227-1230.2001>.
6. Zeller V, Nardi AL, Truffot-Pernot C, Sougakoff W, Stankoff B, Katlama C, Bricaire F. 2003. Disseminated infection with a mycobacterium related to *Mycobacterium triplex* with central nervous system involvement associated with AIDS. *J. Clin. Microbiol.* 41:2785–2787. <http://dx.doi.org/10.1128/JCM.41.6.2785-2787.2003>.
7. Hoff E, Sholtis M, Procop G, Sabella C, Goldfarb J, Wyllie R, Cunningham R, Stockman L, Hall G. 2001. *Mycobacterium triplex* infection in a liver transplant patient. *J. Clin. Microbiol.* 39:2033–2034. <http://dx.doi.org/10.1128/JCM.39.5.2033-2034.2001>.
8. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhiyani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380. <http://dx.doi.org/10.1371/journal.pone.0058088>.
9. Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, Usadel B. 2012. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res.* 40(Web Server Issue): W622–W627. <http://dx.doi.org/10.1093/nar/gks540>.
10. Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Prjibelski AD, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, Clingenpeel SR, Woyke T, McLean JS, Lasken R, Tesler G, Alekseyev MA, Pevzner PA. 2013. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J. Comput. Biol.* 20: 714–737. <http://dx.doi.org/10.1089/cmb.2013.0084>.
11. 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. <http://dx.doi.org/10.1089/cmb.2012.0021>.
12. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding preassembled contigs using SSPACE. *Bioinformatics* 27: 578–579. <http://dx.doi.org/10.1093/bioinformatics/btq683>.
13. Gao S, Sung WK, Nagarajan N. 2011. Opera: reconstructing optimal genomic scaffolds with high-throughput paired-end sequences. *J. Comput. Biol.* 18:1681–1691. <http://dx.doi.org/10.1089/cmb.2011.0170>.
14. Boetzer M, Pirovano W. 2012. Toward almost closed genomes with GapFiller. *Genome Biol.* 13:R56. <http://dx.doi.org/10.1186/gb-2012-13-6-r56>.
15. 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. <http://dx.doi.org/10.1093/nar/gkm160>.
16. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32:11–16. <http://dx.doi.org/10.1093/nar/gkh152>.
17. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. 2003. Rfam: an RNA family database. *Nucleic Acids Res.* 31:439–441. <http://dx.doi.org/10.1093/nar/gkg006>.
18. 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. <http://dx.doi.org/10.1093/nar/gkr717>.
19. 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. <http://dx.doi.org/10.1186/1471-2105-11-119>.
20. Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. 2012. GenBank. *Nucleic Acids Res.* 40:D48–D53. <http://dx.doi.org/10.1093/nar/gkr1202>.
21. 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. <http://dx.doi.org/10.1093/nar/28.1.33>.
22. Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. *Science* 278:631–637. <http://dx.doi.org/10.1126/science.278.5338.631>.