AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

PROKARYOTES

Draft Genome Sequence of *Mycobacterium colombiense*

Amar Bouam, Catherine Robert, Anthony Levasseur, Michel Drancourt

Aix Marseille Univiversité, CNRS, IRD, INSERM, AP-HM, URMITE, IHU - Méditerranée Infection, Marseille, France

ABSTRACT *Mycobacterium colombiense* is a rapidly growing mycobacterium initially isolated from the blood of an HIV-positive patient in Colombia. Its 5,854,893-bp draft genome exhibits a G+C content of 67.64%, 5,233 protein-coding genes, and 54 predicted RNA genes.

Received 2 February 2017 **Accepted** 3 February 2017 **Published** 6 April 2017

Citation Bouam A, Robert C, Levasseur A, Drancourt M. 2017. Draft genome sequence of *Mycobacterium colombiense*. Genome Announc 5:e00119-17. https://doi.org/10.1128/genomeA.00119-17.

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

Mycobacterium colombiense is an acid-fast, nonmotile, rod-shaped mycobacterium that grows in 3 weeks, producing rough, nonpigmented colonies. It was initially isolated from the blood of four HIV-coinfected patients in Colombia in 1995 (1). Further isolates have been made from diseased lymph nodes and respiratory and stool specimens (2–4), as well as from the skin biopsy of a 17-year-old boy with disseminated cutaneous infection (5). *M. colombiense* was also isolated from hospital water in dental units. To our knowledge, it has never been isolated from animals. We performed the whole-genome sequencing of *M. colombiense* CSURP297 in order to describe its genomic content and to determine its phylogenetic relationships for facilitating the detection and identification of this species.

M. colombiense CSURP297 (Collection de Souches de l'Unité des Rickettsies, Marseille, France) was cultured in MGIT Middlebrook liquid culture (Becton, Dickinson, Le Pont-de-Claix, France) at 37°C in a 5% CO_2 atmosphere. *M. colombiense* CSURP297 genomic DNA was sequenced by Illumina MiSeq runs (Illumina Inc, San Diego, CA, USA) with the mate-pair strategy using the Nextera Mate Pair sample prep kit (Illumina). The index representation for *M. colombiense* CSURP297 was determined to be 14.21%. A total of 1,092,357 paired reads were filtered per the read qualities. These reads were trimmed using Trimmomatic (6) and then assembled into scaffolds using SPAdes version 3.5 (7, 8) before manual finishing. SSPACE version 2 (9) and Opera version 2 (10) were used to combine the contigs helped by GapFiller version 1.10 (11). This yielded a draft genome consisting of 14 scaffolds composed of 123 contigs, for a total of 5,854,893 bp and a G+C content of 67.64%. Noncoding genes and miscellaneous features were predicted using RNAmmer (12), ARAGORN (13), Rfam (14), PFAM (15), and Infernal (16). Coding DNA sequences (CDSs) were predicted using Prodigal (17), and functional annotation was achieved using BLASTp against the GenBank database (18) and the Clusters of Orthologous Groups (COG) database (19, 20). The genome was shown to encode 54 predicted RNAs, including one each of the 5S rRNA, 16S rRNA, and 23S rRNA genes and 51 tRNAs. A total of 3,923 genes (74.97%) were assigned a putative function; 64 genes were identified as ORFans (1.22%); and 1,088 genes (20.79%) were annotated as hypothetical proteins. The *M. colombiense* CSURP297 genome was further incorporated into *in silico* DNA-DNA hybridization (DDH) (21) with reference genomes selected based on 16S rRNA gene proximity; DDH values were estimated using the GGDC version 2.0 online tool (22). This analysis yielded 31.95% ± 3.46 similarity with *M. intracellulare* ATCC 13950, 30.45% ± 3.46 with *M. avium* 104, 23.15% ± 3.32 with *M. szulgai* strain ACS1160, 22.8% ± 3.39 with *M. haemophilum* DSM 44634, 22.6% ± 3.39 with *M. tuberculosis* H37Rv, 22.55% ± 3.32 with *M. caprae* strain Allgaeu, and 22.15% ±

3.32 with *M. marinum* M, confirming at the genome level the taxonomic assignment of *M. colombiense* into the *M. avium* complex.

Accession number(s). The *M. colombiense* CSURP297 genome sequence has been deposited at EMBL under the accession number FUEH00000000.

ACKNOWLEDGMENT

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

REFERENCES

1. Murcia MI, Tortoli E, Menendez MC, Palenque E, Garcia MJ. 2006. *Mycobacterium colombiense* sp. nov., a novel member of the *Mycobacterium avium* complex and description of MAC-X as a new ITS genetic variant. Int J Syst Evol Microbiol 56:2049–2054. https://doi.org/10.1099/ijs.0.64190-0.
2. Esparcia O, Navarro F, Quer M, Coll P. 2008. Lymphadenopathy caused by *Mycobacterium colombiense*. J Clin Microbiol 46:1885–1887. https://doi.org/10.1128/JCM.01441-07.
3. Despierres L, Cohen-Bacrie S, Richet H, Drancourt M. 2012. Diversity of *Mycobacterium avium* subsp. *hominissuis* mycobacteria causing lymphadenitis, France. Eur J Clin Microbiol Infect Dis 31:1373–1379. https://doi.org/10.1007/s10096-011-1452-2.
4. Barretto AR, Felício JS, Sales LH, Yamada ES, Lopes ML, da Costa AR. 2016. A fatal case of pulmonary infection by *Mycobacterium colombiense* in Para State, Amazon region, Brazil. Diagn Microbiol Infect Dis 85:344–346. https://doi.org/10.1016/j.diagmicrobio.2016.02.011.
5. Gao W, Chen H, Jiang H, Wang Q, Tang M, Wang HS. 2014. Disseminated cutaneous infection caused by *Mycobacterium colombiense*. Acta Derm Venereol 94:727–728. https://doi.org/10.2340/00015555-1828.
6. 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:W622–W627. https://doi.org/10.1093/nar/gks540.
7. 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. https://doi.org/10.1089/cmb.2013.0084.
8. 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 singlecell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
9. 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.
10. Gao S, Sung WK, Nagarajan N. 2011. Opera: reconstructing optimal genomic scaffolds with high-throughput paired-end sequences. J Comput Biol 18:1681–1691. https://doi.org/10.1089/cmb.2011.0170.
11. 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.
12. 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.
13. 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.
14. 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.
15. Punta M, Coggill 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.
16. 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.
17. 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.
18. 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.
19. 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.
20. Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. Science 278:631–637. https://doi.org/10.1126/science.278.5338.631.
21. 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.
22. 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.