

Draft Genome Sequence of *Mycobacterium europaeum* Strain CSUR P1344

Michael Phelippeau, Olivier Croce, Catherine Robert, Didier Raoult, Michel Drancourt

Aix Marseille Université, URMITE, UMR CNRS 7278, IRD 198, INSERM 1095, Faculté de Médecine, Marseille, France

We report the draft genome sequence of *Mycobacterium europaeum* strain CSUR P1344, a slowly growing mycobacterium of the *Mycobacterium simiae* complex and opportunistic respiratory tract colonizer and pathogen. This genome of 6,152,523 bp exhibits a 68.18% G+C content, encoding 5,814 predicted proteins and 74 RNAs.

Received 15 June 2015 Accepted 16 June 2015 Published 23 July 2015

Citation Phelippeau M, Croce O, Robert C, Raoult D, Drancourt M. 2015. Draft genome sequence of *Mycobacterium europaeum* strain CSUR P1344. *Genome Announc* 3(4): e00816-15. doi:10.1128/genomeA.00816-15.

Copyright © 2015 Phelippeau 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 europaeum was described in 2011 on the basis of five isolates made from the respiratory tract and lymph nodes of patients in Italy, Greece, and Sweden (1). Two further cases were reported in the respiratory tract of Iranian patients (2). Recently, we acquired one *M. europaeum* isolate from the respiratory tract of a patient with a history of HIV-HCV coinfection and acute flu (3), and one further isolate from the sinus surgical drainage of a patient with sinus abscess. Both isolates were identified on the basis of a 100% similarity with the *rpoB* partial sequence (1, 4). To date, only 20 partial sequences of the 16S rRNA, 16S–23S rRNA intergenic spacers—*hsp65*, *tuf*, *sodA*, *hsp65*, *groEL2*, and *rpoB*—are available for *M. europaeum*. We thus performed whole-genome sequencing of the *M. europaeum* CSUR P1344 isolate (3) in order to increase knowledge and gain insights into this organism, which is potentially involved in the contamination and/or infection of airways.

Genomic DNA was isolated from *M. europaeum* strain CSUR P1344 and grown onto Middlebrook 7H10 solid medium (Becton, Dickinson, Le Pont-de-Claix, France) at 37°C in an atmosphere enriched with 5% CO₂. It was sequenced by parallel paired-end and mate pair high-throughput MiSeq technology (Illumina Inc, San Diego, CA, USA). Starting from 1.5 μg genomic (gDNA) fragmented with an optimal size at 6 kb, the mate pair Nextera library was constructed with an input of 1 ng of gDNA. Both libraries were sequenced on MiSeq in 2 × 251 bp, and final sequencing produced a total of 1,047,703 reads.

The whole set of reads was trimmed using Trimmomatic (5) and then assembled through the assembler software Spades (6, 7). Contigs were combined together by SSPACE (8), Opera (9) helped by GapFiller (10), and homemade tools in Python to refine the set. Finally, the draft genome of *M. europaeum* CSUR P1344 consists of 13 contigs without gaps, containing 6,152,523 bp. The G+C content of this genome is 68.18%.

Noncoding genes and miscellaneous features were predicted using RNAmmer (11), ARAGORN (12), Rfam (13), Pfam (14), and Infernal (15). Coding DNA sequences (CDSs) were predicted using Prodigal (16), and functional annotation was achieved using BLAST+ (17) and HMMER3 (18) against the UniProtKB data-

base (19). The genome was shown to encode at least 74 predicted RNAs, including 3 rRNAs, 50 tRNAs, 1 tmRNA, and 20 miscellaneous RNAs. A total of 5,814 identified genes yielded a coding capacity of 5,560,314 bp (coding percentage: 90.37%). Among these genes, 295 (5.07%) were found to be putative proteins and 1,061 (18.25%) were assigned as hypothetical proteins. Moreover, 4,005 genes matched at least one sequence in the Clusters of Orthologous Groups (COGs) database (20, 21) with BLASTp default parameters.

Nucleotide sequence accession numbers. The *M. europaeum* CSUR P1344 strain genome sequence has been deposited with its annotations at EMBL under the accession numbers CTEC01000001 to CTEC01000013.

ACKNOWLEDGMENT

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

REFERENCES

- Tortoli E, Böttger EC, Fabio A, Falsen E, Gitti Z, Grottola A, Klenk HP, Mannino R, Mariottini A, Messinò M, Pecorari M, Rumpianesi F. 2011. *Mycobacterium europaeum* sp. nov., a scotochromogenic species related to the *Mycobacterium simiae* complex. *Int J Syst Evol Microbiol* 61: 1606–1611. <http://dx.doi.org/10.1099/ijs.0.025601-0>.
- Pourahmad F, Shojaei H, Heidarieh P, Khosravi A, Hashemi A. 2012. Report of two cases of *Mycobacterium europaeum* from Iran. *Jpn J Infect Dis* 65:539–541. <http://dx.doi.org/10.7883/yoken.65.539>.
- Phelippeau M, Delord M, Drancourt M, Brouqui P. 2014. Respiratory tract isolation of *Mycobacterium europaeum* following influenza infection in an immunocompromised patient: a case report. *J Med Case Rep* 8:463. <http://dx.doi.org/10.1186/1752-1947-8-463>.
- Adékambi T, Colson P, Drancourt M. 2003. *rpoB*-based identification of nonpigmented and late-pigmenting rapidly growing mycobacteria. *J Clin Microbiol* 41:5699–5708. <http://dx.doi.org/10.1128/JCM.41.12.5699-5708.2003>.
- 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(W1):W622–W627. <http://dx.doi.org/10.1093/nar/gks540>.
- Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Pribelski 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>.
7. 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 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>.
 8. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27: 578–579. <http://dx.doi.org/10.1093/bioinformatics/btq683>.
 9. 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>.
 10. Boetzer M, Pirovano W. 2012. Toward almost closed genomes with Gap-Filler. *Genome Biol* 13:R56. <http://dx.doi.org/10.1186/gb-2012-13-6-r56>.
 11. 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>.
 12. 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>.
 13. 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>.
 14. 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/gkr1065>.
 15. Nawrocki EP, Kolbe DL, Eddy SR. 2009. Infernal 1.0: inference of RNA alignments. *Bioinformatics* 25:1335–1337. <http://dx.doi.org/10.1093/bioinformatics/btp157>.
 16. 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>.
 17. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <http://dx.doi.org/10.1186/1471-2105-10-421>.
 18. Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* 7:e1002195. <http://dx.doi.org/10.1371/journal.pcbi.1002195>.
 19. The UniProt Consortium. 2011. Ongoing and future developments at the universal protein resource. *Nucleic Acids Res* 39:D214–D219. <http://dx.doi.org/10.1093/nar/gkq1020>.
 20. 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>.
 21. 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>.