Draft Genome Sequence of *Mycobacterium lentiflavum* CSUR P1491

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 announce the draft genome sequence of *Mycobacterium lentiflavum* strain CSUR P1491, a nontuberculous mycobacterium responsible for opportunistic potentially life-threatening infections in immunocompromised patients. The genome described here comprises a 6,818,507-bp chromosome exhibiting a 65.75% G+C content, 6,354 protein-coding genes, and 75 RNA genes.

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

The initial description of *Mycobacterium lentiflavum* was based on three isolates belonging to a single DNA homology group [1]. Further phylogenetic analyses found >99.5% similarity in the 16S rRNA gene sequence with *Mycobacterium simiae* and *Mycobacterium genavense* [1]. This observation qualified *M. lentiflavum* as a member of the *Mycobacterium simiae* complex. This environmental organism [2] has been isolated from human specimens, including an intervertebral disc [1], respiratory tract samples and gastric/gut aspirates [1, 3], lymph nodes [4], feces [5], and bone marrow [5]. *M. lentiflavum* emerges in cystic fibrosis patients [6] and was responsible for the death of one heart-transplanted immunocompromised patient [5]. Routine identification of this fastidious mycobacterium can be achieved by partial *rpoB* gene sequencing [7].

We sequenced the whole genome of *M. lentiflavum* to exact the phylogenetic relationships within the *M. simiae* complex, design tools for its genotyping, and disclose any genotypic pattern possibly related to host-mycobacterium relationships.

Genomic DNA was extracted from *M. lentiflavum* strain CSUR P1491 (isolated from the sputum of a cystic fibrosis patient) grown on Middlebrook 7H10 solid medium at 37°C in a 5% CO_2 atmosphere for 3 weeks. Genomic DNA was sequenced on Illumina MiSeq throughout three runs, including one paired-end Nextera and two mate pair libraries, in a 2 × 250-bp run for each barcoded library. The whole set of reads was trimmed using Trimmomatic [8], then assembled through the assembler software SPAdes v3.5 [9, 10]. Contigs obtained were combined together by SSPACE v2 [11], Opera v2 [12] helped by GapFiller v1.10 [13], and homemade tools to refine the set.

The draft genome of *M. lentiflavum* consists of five contigs without gaps, containing 6,818,507 bp. The G+C content of this genome is 65.75%. Noncoding genes and miscellaneous features were predicted using RNAmmer [14], ARAGORN [15], Rfam [16], Pfam [17], and Infernal [18]. Coding DNA sequences (CDSs) were predicted using Prodigal [19] and functional annotation was achieved using BLAST+ [20] and HMMER3 [21] against the UniProtKB database [22]. This genome is predicted to encode at least 75 predicted RNAs, including 3 rRNAs, 52 tRNAs, 1 transfer-messenger RNA (tmRNA), and 19 miscellaneous RNAs. A total of 6,354 genes were identified, representing a coding capacity of 6,173,541 bp (coding percentage, 90.54%). Whereas 4,369 genes matched a least one sequence in the Clusters of Orthologous Groups (COGs) database [23, 24, 25] with BLASTP default parameters, 1,150 (18.1%) were assigned as hypothetical proteins, and 332 (5.23%) genes were founded as encoding putative protein. No genes encoding antibiotic resistance were found, but bleomycin resistance is encoded.

In silico DNA-DNA hybridization (DDH) was performed with the genome-to-genome distances between pairs of two closed genomes [25]. The genome of *M. lentiflavum* was locally aligned 2-by-2 using the BLAT algorithm [26, 27] against *M. simiae* and *M. genavense*. The DDH was estimated from a generalized linear model using a specific distance formula [28], resulting in following values: 27.5% (±2.43) with *M. simiae* and 34.6% (±2.47) with *M. genavense*. These data confirm *M. lentiflavum* as a unique species clearly distinct from *M. simiae* and *M. genavense*.

Nucleotide sequence accession numbers. The *M. lentiflavum* strain CSUR P1491 genome sequence has been deposited at EMBL under the accession numbers CTEE01000001 to CTEE01000005.

REFERENCES

1. **Springer B, Wu WK, Bodmer T, Haase G, Pfyffer GE, Kroppenstedt RM, Schröder KH, Emler S, Kilburn JO, Kirschner P, Telenti A, Coyle MB, Böttger EC.** 1996. Isolation and characterization of a unique group of slowly growing mycobacteria: description of *Mycobacterium lentiflavum* sp. nov. J Clin Microbiol **34:**1100–1107.

2. **Marshall HM, Carter R, Torbey MJ, Minion S, Tolson C, Sidjabat HE, Huygens F, Hargreaves M, Thomson RM.** 2011. *Mycobacterium lentiflavum* in drinking water supplies, Australia. Emerg Infect Dis **17:**395–402. http://dx.doi.org/10.3201/eid1703.090948.

3. **Nagata N, Honda M, Kobayakawa M, Maeda S, Sakurai T, Akiyama J, Gotoda T, Oka S, Uemura N.** 2011. *Mycobacterium lentiflavum* ileitis using aspirated intestinal fluid during endoscopy in HIV-infected patient. Dig Endosc **23:**271–272. http://dx.doi.org/10.1111/j.1443-1661.2010.01091.x.

4. **Haase G, Kentrup H, Skopnik H, Springer B, Böttger EC.** 1997. *Mycobacterium lentiflavum*: an etiologic agent of cervical lymphadenitis. Clin Infect Dis **25:**1245–1246. http://dx.doi.org/10.1086/516958.

5. **Thomas G, Hraiech S, Dizier S, Weiller PJ, Ene N, Serratrice J, Secq V,**

Phelippeau et al.

Ambrosi P, Drancourt M, Roch A, Papazian L. 2014. Disseminated *Mycobacterium lentiflavum* responsible for hemophagocytic lymphohistiocytosis in a man with a history of heart transplantation. J Clin Microbiol **52:**3121–3123. http://dx.doi.org/10.1128/JCM.00758-14.

6. Satana D, Erkose-Genc G, Tamay Z, Uzun M, Guler N, Erturan Z. 2014. Prevalence and drug resistance of mycobacteria in Turkish cystic fibrosis patients. Ann Clin Microbiol Antimicrob **13:**28. http://dx.doi.org/10.1186/1476-0711-13-28.

7. Adékambi T, Colson P, Drancourt M. 2003. *rpoB*-based identification of nonpigmenting and late-pigmented rapidly growing mycobacteria. J Clin Microbiol **41:**5699–5708. http://dx.doi.org/10.1128/JCM.41.12.5699-5708.2003.

8. 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. http://dx.doi.org/10.1093/nar/gks540.

9. 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.

10. 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.

11. 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.

12. 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.

13. 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.

14. 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.

15. 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.

16. 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.

17. 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. http://dx.doi.org/10.1093/nar/gkr1065.

18. 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.

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. 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.

21. Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol **7:**e1002195. http://dx.doi.org/10.1371/journal.pcbi.1002195.

22. 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.

23. 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.

24. 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.

25. Richter M, Rosselló-Móra R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci USA **106:**19126–19131. http://dx.doi.org/10.1073/pnas.0906412106.

26. Kent WJ. 2002. BLAT—the blast-like alignment tool. Genome Res **12:**656–664. http://dx.doi.org/10.1101/gr.229202.

27. 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. http://dx.doi.org/10.4056/sigs.531120.

28. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics **14:**60. http://dx.doi.org/10.1186/1471-2105-14-60.