genomeA
Journals.ASM.org

Draft Genome Sequence of *Mycobacterium vulneris* DSM 45247^T

Olivier Croce, Catherine Robert, Didier Raoult, Michel Drancourt

Aix Marseille Université, URMITE, Marseille, France

We report the draft genome sequence of *Mycobacterium vulneris* DSM 45247^T strain, an emerging, opportunistic pathogen of the *Mycobacterium avium* complex. The genome described here is composed of 6,981,439 bp (with a G+C content of 67.14%) and has 6,653 protein-coding genes and 84 predicted RNA genes.

Received 7 April 2014 **Accepted** 14 April 2014 **Published** 8 May 2014

Citation Croce O, Robert C, Raoult D, Drancourt M. 2014. Draft genome sequence of *Mycobacterium vulneris* DSM 45247^T. Genome Announc. 2(3):e00370-14. doi:10.1128/genomeA.00370-14.

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

Mycobacterium vulneris is a nontuberculous mycobacterium recently individualized among the *Mycobacterium avium* complex (1). The name was given after the initial isolation of the organism from a dog-bite wound discharge; a second isolate was made from a diseased lymph node in a 2-year-old child (1, 2). However, no further isolates have been obtained, so the clinical spectrum of this pathogen as well as its reservoir and sources are not yet known. *M. vulneris* is a mycobacterium previously referred to as *M. avium* sequevar Q (1, 2). *M. vulneris* was shown to be more closely related to *Mycobacterium colombiense*, another member of the *M. avium* complex (1, 3). Its precise relationships with several species recently described in the *M. avium* complex remain unknown (4).

In order to further decrypt the phylogenetic relationships of *M. vulneris* within the *M. avium* complex, we sequenced the whole genome of the *M. vulneris* DSM 45247^T strain.

Genomic DNA was isolated from an *M. vulneris* DSM 45247^T strain grown on MGIT Middlebrook broth (Becton Dickinson, Sparks, MD) at 37° C. Genomic DNA of *M. vulneris* was sequenced on MiSeq Technology (Illumina, Inc., San Diego, CA) by using paired-end and mate-pair applications in parallel, in a 2- × 250-bp run for each bar-coded library. On each flow cell, The index representation for *M. vulneris* was determined to be 5.46% and 7.71%, respectively, on each flow cell. The total 1,697,812 reads were filtered according to the read qualities.

The whole set of reads was trimmed using Trimmomatic (5), and then assembled with the assembler software Spades v 3.0 (6, 7). Contigs obtained were combined together by SSPACE v 2.0 (8) and Opera software v 1.4 (9) and helped by GapFiller v 1.10 (10) to reduce the set. For some manual refinements we used the CLC Genomics v 6 software (CLC bio, Aarhus, Denmark) and homemade tools. The completed draft genome sequence of *M. vulneris* consists of four contigs without gaps, containing 6,981,439 bp and 67.14% G+C content.

Noncoding genes and miscellaneous features were predicted using RNAmmer (11), ARAGORN (12), Rfam (13), and PFAM (14). Open reading frames (ORFs) were predicted using Prodigal (15), and functional annotation was achieved using BLASTP against the GenBank database (16) and the Clusters of Orthologous Groups (COGs) database (17, 18). The genome was shown to encode at least 84 predicted RNAs, including 7 rRNAs, 58 tRNAs, 1 transfer-messenger RNA (tmRNA), and 18 miscellaneous RNAs. A total of 6,653 genes were also identified, representing a coding capacity of 6,470,571 bp and a 92.6% coding percentage. Whereas 6,608 genes matched a least one sequence in the COGs database when BLASTP default parameters were used, 881 (13.24%) genes encoded putative proteins and 1,051 (15.8%) genes were assigned as hypothetical proteins.

Nucleotide sequence accession numbers. The *Mycobacterium vulneris* strain DSM 45247^T genome sequence has been deposited at EMBL under the accession numbers CCBG010000001 through CCBG010000004.

ACKNOWLEDGMENT

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

REFERENCES

1. **Van Ingen J, Boeree MJ, Kösters K, Wieland A, Tortoli E, Dekhuijzen PN, van Soolingen D.** 2009. Proposal to elevate *Mycobacterium avium* complex ITS sequevar MAC-Q to *Mycobacterium vulneris* sp. nov. Int. J. Syst. Evol. Microbiol. **59:**2277–2782. http://dx.doi.org/10.1099/ijs.0.008854-0.
2. **Mijs W, de Haas P, Rossau R, van der Laan T, Rigouts L, Portaels F, van Soolingen D.** 2002. Molecular evidence to support a proposal to reserve the designation *Mycobacterium avium* subsp. avium for bird-type isolates and "*M. avium* subsp. *hominis suis*" for the human/porcine type of *M. avium*. Int. J. Syst. Evol. Microbiol. **52**(Pt 5):1505–1518. http://dx.doi.org/10.1099/ijs.0.02037-0.
3. **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**(Pt 9):2049–2054. http://dx.doi.org/10.1099/ijs.0.64190-0.
4. **Ben Salah I, Cayrou C, Raoult D, Drancourt M.** 2009. *Mycobacterium marseillense* sp. nov., *Mycobacterium timonense* sp. nov. and *Mycobacterium bouchedurhonense* sp. nov., members of the *Mycobacterium avium* complex. Int. J. Syst. Evol. Microbiol. **59**(Pt 11):2803–2808. http://dx.doi.org/10.1099/ijs.0.010637-0.
5. **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.
6. **Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Prjibelski AD, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanaus-**

kas 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, 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.
8. **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.
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 GapFiller. 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, 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.
15. **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.
16. **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.
17. **Tatusov RL, Galperin MY, Natale DA, Koonin EV.** 2000. The COG database: a tool for genomoe-scale analysis of protein functions and evolution. Nucleic Acids Res. **28:**33–36. http://dx.doi.org/10.1093/nar/28.1.33.
18. **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.