Draft Genome Sequence of *Mycobacterium bohemicum* Strain DSM 44277^T

Shady Asmar, Michael Phelippeau, Catherine Robert, Olivier Croce, Michel Drancourt

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

The *Mycobacterium bohemicum* strain is a nontuberculosis species mainly responsible for pediatric cervical lymphadenitis. The draft genome of *M. bohemicum* DSM 44277^T comprises 5,097,190 bp exhibiting a 68.64% G+C content, 4,840 protein-coding genes, and 75 predicted RNA genes.

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

Mycobacterium bohemicum is a slow-growing, scotochromogenic mycobacterium characterized by biochemical inertia and a unique 16S RNA gene sequence (1). Although the first strain was isolated from the sputum of a 53-year-old patient with Down's syndrome (1), most *M. bohemicum* isolates have been collected from pediatric cervical lymphadenitis (2–5). Furthermore, *M. bohemicum* has also been isolated from animal and inanimate environmental sources (6).

To facilitate the development of advanced molecular tools for the detection and identification of this species, whole-genome sequencing of *M. bohemicum* DSM 44277^T originally isolated from a sputum specimen was performed. Genomic DNA was isolated from *M. bohemicum* strain DSM 44277^T grown in MGIT Middlebrook liquid culture (Becton Dickinson, Le Pont-de-Claix, France) at 37°C in a 5% CO_2 atmosphere. *M. bohemicum* genomic DNA was then sequenced in 2 Illumina MiSeq runs (Illumina, Inc., San Diego, CA, USA) using a 5-kb mate-paired library and one Roche 454 run using a 5-kb paired-end library (Roche, Basel, Switzerland). Reads from the 454 sequencer were assembled using Newbler software (Roche) and reads from Illumina were trimmed using Trimmomatic (7), and assembled using SPAdes v3.5 (8, 9). Contigs were combined together by SSPACE v2 (10), Opera v2 (11) helped by GapFiller v1.10 (12), and homemade tools in Python to refine the set. Finally, the draft genome of *M. bohemicum* DSM 44277^T consists of 18 scaffolds and 31 contigs containing 5,097,190-bp. The G+C content of this genome is 68.64%. Noncoding genes and miscellaneous features were predicted using RNAmmer (13), ARAGORN (14), Rfam (15), Pfam (16), and Infernal (17). Coding DNA sequences (CDSs) were predicted using Prodigal (18) and functional annotation was achieved using BLAST+ (19) and HMMER3 (20) against the UniProtKB database (21). The genome was shown to encode at least 75 predicted RNAs including three rRNAs, 57 tRNAs, one transfer-messenger RNA (tmRNA), and 14 miscellaneous RNAs. A total of 4,840 identified genes yielded a coding capacity of 4,578,093 bp (coding percentage, 89.82%). Among these genes, 255 (5.27%) were found to be putative proteins and 785 (16.22%) were assigned as hypothetical proteins. Moreover, 3,383 genes matched a least one sequence in the Clusters of Orthologous Groups database (22, 23) with BLASTp default parameters. *In silico* DNA-DNA hybridization (DDH) (24) was performed with nine reference genomes selected on the basis of their 16S rRNA gene proximity with *M. bohemicum*. The *M. bohemicum* genome was locally aligned 2-by-2 using the BLAT algorithm (25, 26) against each of the nine selected genomes, and DDH values were estimated from a generalized linear model (27). The DDH was of 27.7% ($\pm$ 2.43) with *Mycobacterium avium* 104, 27.1% ($\pm$ 2.42) with *Mycobacterium indicus pranii* MTCC 9506 and *Mycobacterium intracellulare* MOTT-02, 24% ($\pm$ 2.39) with *Mycobacterium kansasii* ATCC 12478, 23.9% ($\pm$ 2.39) with *Mycobacterium tuberculosis* H37Rv, 23.3% ($\pm$ 2.38) with *Mycobacterium ulcerans* Agy99, 23.2% ($\pm$ 2.38) with *Mycobacterium liflandii* 128FXT and *Mycobacterium asiaticum* DSM 44297, and 23.1% ($\pm$ 2.37) with *Mycobacterium marinum* M. These data confirm that *M. bohemicum* is a unique species more closely related to the *M. avium* complex.

Nucleotide sequence accession numbers. The *M. bohemicum* strain DSM 44277^T genome sequence has been deposited at EMBL under the accession numbers CSTD01000001 to CSTD01000018.

ACKNOWLEDGMENT

REFERENCES

1. **Reischl U, Emler S, Horak Z, Kaustova J, Kroppenstedt RM, Lehn N, Naumann L.** 1998. *Mycobacterium bohemicum* sp. nov., a new slow-growing scotochromogenic mycobacterium. Int J Syst Bacteriol **48:** 1349–1355. http://dx.doi.org/10.1099/00207713-48-4-1349.
2. **Tortoli E, Bartoloni A, Manfrin V, Mantella A, Scarparo C, Böttger E.** 2000. Cervical lymphadenitis due to *Mycobacterium bohemicum*. Clin Infect Dis **30:**210–211. http://dx.doi.org/10.1086/313600.
3. **Palca A, Aebi C, Weimann R, Bodmer T.** 2002. *Mycobacterium bohemicum* cervical lymphadenitis. Pediatr Infect Dis J **21:**982–984. http://dx.doi.org/10.1097/00006454-200210000-00022.
4. **Schulzke S, Adler H, Bär G, Heininger U, Hammer J.** 2004. *Mycobacterium bohemicum*—a cause of paediatric cervical lymphadenitis. Swiss Med Wkly **17:**221–222.
5. **Huber J, Richter E, Binder L, Maa M, Eberl R, Zenz W.** 2008. *Myco-*

bacterium bohemicum and cervical lymphadenitis in children. Emerg Infect Dis **14**:1158–1159. http://dx.doi.org/10.3201/eid1407.080142.

6. **Torkko P, Suomalainen S, Iivanainen E, Suutari M, Paulin L, Rudbäck E, Tortoli E, Vincent V, Mattila R, Katila M-**. 2001. Characterization of *Mycobacterium bohemicum* isolated from human, veterinary, and environmental sources. J Clin Microbiol **39**:207–211. http://dx.doi.org/10.1128/JCM.39.1.207-211.2001.

7. **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.

8. **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.

9. **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.

10. **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.

11. **Gao S, Sung W, 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.

12. **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.

13. **Lagesen K, Hallin P, Rodland EA, Staerfeldt H-, 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.

14. **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.

15. **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.

16. **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 ELL, 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.

17. **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.

18. **Hyatt D, Chen G, 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.

19. **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.

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

21. **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.

22. **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.

23. **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.

24. **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. http://dx.doi.org/10.1073/pnas.0906412106.

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

26. **Auch AF, Von Jan M, Klenk H-P, 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.

27. **Meier-Kolthoff JP, Auch AF, Klenk H, 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.