Draft Genome Sequence of *Mycobacterium neoaurum* Strain DSM 44074^T

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

Unité de Recherche sur les Maladies Infectieuses et Tropicales Émergentes, CNRS, UMR 7278, IRD 198, Faculté de Médecine, Aix-Marseille Université, Marseille, France

We report the draft genome sequence of *Mycobacterium neoaurum* strain DSM 44074^T, a nontuberculosis species responsible for opportunistic infections in immunocompromised patients. The strain described here is composed of 5,536,033 bp, with a G+C content of 66.24%, and carries 5,274 protein-coding genes and 72 RNA genes.

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

Mycobacterium neoaurum is an environmental, rapidly growing mycobacterium that is rarely reported as an opportunistic pathogen responsible for bloodstream infections, cutaneous infections, and pneumonia in immunocompromised patients ([1], [2]). To complete a genome sequence derived from an environmental Russian isolate ([3]) we sequenced the whole genome of *M. neoaurum* DSM 44074^T, a strain confirmed as a valuable representative of the species ([1]).

Genomic DNA was isolated from *M. neoaurum* strain DSM 44074^T grown on MGIT broth medium (Becton Dickinson, Le Pont-de-Claix, France) at 37°C in an atmosphere enriched with 5% CO_2. Genomic DNA was then sequenced using two high-throughput next-generation sequencing technologies: Roche 454 ([4]) and MiSeq Illumina (Illumina Inc., San Diego, CA). A 5-kb paired-end library was constructed, loaded on a PTP plate, and sequenced with the Roche-GS FLX Titanium sequencing kit XLR70, which produced 157,956 reads. Illumina sequencing was performed using two mate-pair Nextera libraries, sequenced on MiSeq in 2 × 250 bp. The DNA fragments ranged in size from 1 to 10 kb, and final sequencing produced a total of 145,157 reads. Reads from various sequencing technologies were first assembled separately. Reads from Roche 454 sequencing technologies were assembled into contigs and scaffolds using Newbler version 2.8 (Roche, 454 Life Sciences). Illumina reads were trimmed using Trimmomatic ([5]) and then assembled with Spades software ([6], [7]) while contigs generated from Roche 454 were added. Contigs obtained were combined by SSPACE ([8]) and Opera software ([9]) and then combined by GapFiller version 1.10 ([10]) to reduce the set. Some manual refinements using CLC Genomics version 7 software (CLC bio, Aarhus, Denmark) and some homemade Python scripts improved the genome. Finally, the draft genome of *M. neoaurum* consists of 10 scaffolds of 45 contigs containing 5,504,703 bp and has an estimated size of 5,536,033 bp, including gaps. The G+C content of this genome is 66.24%.

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 database ([19]). The genome was shown to encode at least 72 predicted RNAs, including 5 rRNAs, 49 tRNAs, 1 transfer-messenger RNA, and 17 miscellaneous RNAs. A total of 5,274 identified genes have a coding capacity of 5,112,765 bp (coding percentage: 92.35%), including 239 (4.53%) genes found encoding putative proteins and 822 (15.59%) assigned as encoding hypothetical proteins. Moreover, 5,220 genes matched at least one sequence in the Clusters of Orthologous Groups (COG) database ([20], [21]) with BLASTp default parameters.

This report illustrates the genomic variability within *M. neoaurum* ([3]), a feature on which to base further evaluations.

Nucleotide sequence accession numbers. The *M. neoaurum* DSM 44074^T strain genome sequence has been deposited at EMBL under the accession numbers LK021337 to LK021346. The whole-genome shotgun master numbers are CCDR010000001 to CCDR010000045.

REFERENCES

1. **Simmon KE, Low YY, Brown-Elliott BA, Wallace RJ, Jr, Petti CA.** 2009. Phylogenetic analysis of *Mycobacterium aurum* and *Mycobacterium neoaurum* with redescription of *M. aurum* culture collection strains. Int. J. Syst. Evol. Microbiol. **59**:1371–1375. http://dx.doi.org/10.1099/ijs.0.007799-0.
2. **Kumar A, Pazhayattil GS, Das A, Conte HA.** 2014. *Mycobacterium neoaurum* causing prosthetic valve endocarditis: a case report and review of the literature. Braz. J. Infect. Dis. **18**:235–237. http://dx.doi.org/10.1016/j.bjid.2013.05.012.
3. **Shtratnikova VY, Bragin EY, Dovbnya DV, Pekov YA, Schelkunov MI, Strizhov N, Ivashina TV, Ashapkin VV, Donova MV.** 2014. Complete genome sequence of sterol-transforming *Mycobacterium neoaurum* strain VKM Ac-1815D. Genome Announc. 2(1):e01177-13. http://dx.doi.org/10.1128/genomeA.01177-13.
4. **Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR,**

Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437**:376–380. http://dx.doi.org/10.1038/nature03959.

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, 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, 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 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 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 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.

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. 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.