

Genome Sequence of *Legionella massiliensis*, Isolated from a Cooling Tower Water Sample

Isabelle Pagnier, Olivier Croce, Catherine Robert, Didier Raoult, Bernard La Scola

Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, CNRS, UM63, CNRS7278, INSERM U1095, IRD 198, Faculté de Médecine, Aix-Marseille Université, Marseille, France

We present the draft genome sequence of *Legionella massiliensis* strain LegA^T, recovered from a cooling tower water sample, using an amoebal coculture procedure. The strain described here is composed of 4,387,007 bp, with a G+C content of 41.19%, and its genome has 3,767 protein-coding genes and 60 predicted RNA genes.

Received 10 September 2014 Accepted 11 September 2014 Published 16 October 2014

Citation Pagnier I, Croce O, Robert C, Raoult D, La Scola B. 2014. Genome sequence of *Legionella massiliensis*, isolated from a cooling tower water sample. *Genome Announc.* 2(5):e01068-14. doi:10.1128/genomeA.01068-14.

Copyright © 2014 Pagnier et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](http://creativecommons.org/licenses/by/3.0/).

Address correspondence to Bernard La Scola, bernard.la-scola@univ-amu.fr.

Legionella massiliensis strain LegA^T was isolated from an environmental water sample from a cooling tower located in South France, using an amoebal coculture procedure (1). This is a Gram-negative and Gimenez-positive bacillus classified in the genus *Legionella*. Based on the sequencing of the complete 16S rRNA gene, using BLASTn (2), the closest related species is *Legionella birming-hamensis* (accession no. NR044953), with 96.7% similarity. Regarding the macrophage infectivity potentiator (*mip*) gene, strain LegA^T exhibits 78% similarity with *Legionella feeleii* (accession no. FJ009368). The strain *L. massiliensis* LegA^T is unable to grow on Columbia with 5% sheep blood agar and is able to grow after 3 days on buffered charcoal yeast extract (BCYE) medium under a 5% CO₂ atmosphere. The bacteria showed negative reactions for oxidase, cefinase, and gelatinase. *L. massiliensis* LegA^T is deposited in the DSMZ (DSM 24804^T) and CSUR (SCUR P146^T) culture collections.

We therefore sequenced the whole genome of *L. massiliensis* LegA^T in order to determine the phylogenetic relationships with closely related *Legionella* species. The DNA genome was sequenced using two high-throughput next-generation sequencing (NGS) technologies: Roche 454 (3) and MiSeq Illumina (Illumina, Inc., San Diego, CA). A library of 5-kb paired-ends was constructed, loaded on a PicoTiterPlate (PTP), and sequenced with the Roche-GS FLX Titanium sequencing kit XLR70. MiSeq Illumina sequencing was performed using two applications, paired-end and mate-pair Nextera libraries, in a 2 × 250 bp run for each barcoded library.

The reads from various sequencing technologies were first assembled separately. The reads from 454 sequencing were assembled into contigs and scaffolds using Newbler version 2.8 (Roche, 454 Life Sciences). The Illumina reads were trimmed using Trimmomatic (4) and then assembled with the SPAdes software (5, 6) while adding contigs generated from Roche 454. The obtained contigs were combined by the SSPACE (7) and Opera softwares (8) and helped by GapFiller (9) to reduce the set. Some manual refinements using CLC Genomics software (CLC bio, Aarhus, Denmark) and homemade tools improved the genome. Finally,

the draft genome of *L. massiliensis* LegA^T consists of 8 contigs without gaps, containing 4,387,007 bp and a G+C content of 41.19%.

Noncoding genes and miscellaneous features were predicted using RNAmmer (10), ARAGORN (11), Rfam (12), Pfam (13), and Infernal (14). Coding DNA sequences (CDSs) were predicted using Prodigal (15), and functional annotation was achieved using BLAST+ (16) and HMMER3 (17) against the UniProtKB database (18). The genome was shown to contain at least 60 predicted RNAs, including 7 rRNAs, 40 tRNAs, 1 transfer-messenger RNA (tmRNA), and 12 miscellaneous RNAs. A total of 3,767 genes were also identified, representing a coding capacity of 3,811,533 bp (coding percentage, 86.88%). Among these genes, 174 (4.62%) were founded as putative proteins and 1,326 (35.2%) were assigned as hypothetical proteins. Moreover, 2,405 genes matched a least one sequence in the Clusters of Orthologous Groups (COGs) database (19, 20) with BLASTp default parameters.

Nucleotide sequence accession numbers. The *L. massiliensis* strain LegA^T genome sequence has been deposited at EMBL under the accession numbers [CCVW01000001](https://www.ebi.ac.uk/ena/record/CCVW01000001) to [CCVW01000008](https://www.ebi.ac.uk/ena/record/CCVW01000008).

ACKNOWLEDGMENT

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

REFERENCES

- Campocasso A, Boughalmi M, Fournous G, Raoult D, La Scola B. 2012. Description of two new *Legionella* species isolated from environmental water samples. *Int. J. Syst. Evol. Microbiol.* 62:3003–3006. <http://dx.doi.org/10.1099/ijs.0.037853-0>.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410. [http://dx.doi.org/10.1016/S0022-2836\(05\)80360-2](http://dx.doi.org/10.1016/S0022-2836(05)80360-2).
- 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 W, 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, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380. <http://dx.doi.org/10.1038/nature03959>.
4. 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>.
 5. 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>.
 6. 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>.
 7. 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>.
 8. 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>.
 9. 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>.
 10. 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>.
 11. 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>.
 12. 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>.
 13. 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 ELL, Eddy SR, Bateman AFinn RD. 2012. The Pfam protein families database. *Nucleic Acids Res.* 40:D290–D301. <http://dx.doi.org/10.1093/nar/gkr1065>.
 14. 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>.
 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. 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>.
 17. Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput. Biol.* 7:e1002195. <http://dx.doi.org/10.1371/journal.pcbi.1002195>.
 18. 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>.
 19. 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>.
 20. Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. *Science* 278:631–637.