Genome Sequence of Diplorickettsia massiliensis, an Emerging Ixodes ricinus-Associated Human Pathogen

Mano J Mathew, G. Subramanian, Thi Tien Nguyen, Catherine Robert, Oleg Mediannikov, Pierre-Edouard Fournier, Didier Raoult

▶ **To cite this version:**

Genome Sequence of *Diplorickettsia massiliensis*, an Emerging *Ixodes ricinus*-Associated Human Pathogen

Mano J. Mathew, Geetha Subramanian, Thi-Tien Nguyen, Catherine Robert, Oleg Mediannikov, Pierre-Edouard Fournier, and Didier Raoult

Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UMR CNRS 6236—IRD 198, Faculté de Médecine, Aix-Marseille Université, Marseille, France

Diplorickettsia massiliensis is a gammaproteobacterium in the order *Legionellales* and an agent of tick-borne infection. We sequenced the genome from strain 20B, isolated from an *Ixodes ricinus* tick. The genome consists of a 1,727,973-bp chromosome but no plasmid and includes 2,269 protein-coding genes and 42 RNA genes, including 3 rRNA genes.

*D*iplorickettsia massiliensis was first isolated from *Ixodes ricinus* ticks collected in Slovakia in 2006 (6). This Gram-negative bacillus is classified within the family *Coxiellaceae* in the order *Legionellales*. It is strictly intracellular and is mainly grouped by pairs inside vacuoles of eukaryotic cells (6). In a large serosurvey of patients with suspected tick-borne infections, three patients were found to exhibit a specific seroconversion to *D. massiliensis*, and the bacterium was also PCR amplified from blood from one of these patients (9). This study demonstrated that *D. massiliensis* was a human pathogen.

Genomic DNA isolated from *D. massiliensis* strain 20B grown in XTC-2 cells was pyrosequenced using the 454 GS FLX titanium platform (Roche, Branford, CT) (5) and assembled using the Newbler software (Roche). A total of 90,909 reads were obtained. The gaps between contigs were closed using PCR amplification and sequencing with specifically designed primers. The draft genome of *D. massiliensis* 20B, consisting of seven contigs, contained 1,727,973 bp with a G+C content of 38.9%. Potential coding sequences (CDSs) were predicted using Prodigal (http://prodigal.ornl.gov/) with default parameters, but the predicted open reading frames (ORFs) were excluded if they were spanning a sequencing gap region. Assignment of protein functions was performed by comparison with sequences in GenBank, Clusters of Orthologous Groups (COGs), and Pfam databases using BLASTP (1, 2, 8, 11). Of the 2,269 CDSs that were identified, representing a coding capacity of 1,378,587 bp (79.7% of the complete genome), 1,380 were assigned to COGs (10). Using SignalP v4.0 (7), we identified 57 signal peptide cleavage sites. Using TMHMM v2.0 (3), 376 proteins exhibited transmembrane helices. Using BLASTN and tRNAscan-SE (4), the genome was shown to contain 42 RNA genes, including three rRNA genes and 39 tRNA genes.

When compared to closely related gammaproteobacteria, *D. massiliensis*, with 1.7 Mb, had a bigger genome than *Rickettsiella grylli*, with 1.4 Mb (GenBank accession number AAQJ00000000) but smaller than *Coxiella burnetii* strain CbuK_Q154, with 2.0 Mb (CP001020). However, *D. massiliensis* had more metabolism-related genes (501 genes) than *Rickettsiella grylli* (360) and *Coxiella burnetii* (459); it also had more genes involved in energy production and conversion (109 versus 75 and 84, respectively) and more genes involved in translation, ribosomal structure, and biogenesis (170 versus 134 and 135, respectively).

Further analysis of the *D. massiliensis* genome will be conducted to identify the genes linked to pathogenesis and its specific evolutionary mechanisms.

Nucleotide sequence accession numbers. The *Diplorickettsia massiliensis* 20B whole-genome shotgun (WGS) project has been assigned the project accession number AJGC00000000 in GenBank. This version of the project (01) has been assigned the accession number AJGC01000000 and consists of sequences AJGC01000001 to AJGC01000006.

ACKNOWLEDGMENT

REFERENCES

1. **Altschul SF, et al.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389–3402.
2. **Aziz RK, et al.** 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics **9**:75. doi:10.1186/1471-2164-9-75.
3. **Krogh A, Larsson B, von Heijne G, Sonnhammer EL.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305**:567–580.
4. **Lowe TM, Eddy SR.** 1997. t-RNAscan-SE: a program for improved detection of transfer RNA gene in genomic sequence. Nucleic Acids Res. **25**:955–964.
5. **Margulies M, et al.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437**:376–380.
6. **Mediannikov O, Sekeyova Z, Birg ML, Raoult D.** 2010. A novel obligate intracellular gamma-proteobacterium associated with ixodid ticks, Diplorickettsia massiliensis, gen. nov., sp. nov. PLoS One **5**:e11478. doi:10.1371/journal.pone.0011478.
7. **Petersen TN, Brunak S, von Heijne G, Nielsen H.** 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods **8**:785–786.
8. **Punta M, et al.** 2012. The Pfam protein families database. Nucleic Acids Res. **40**:D290–D301.
9. **Subramanian G, et al.** 2012. Diplorickettsia massiliensis as a human pathogen. Eur. J. Clin. Microbiol. Infect. Dis. **31**:365–369.
10. **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.
11. **Tatusov RL, Koonin EV, Lipman DJ.** 1997. A genomic perspective on protein families. Science **278**:631–637.