**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.

Diplorickettsia 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 tRNA genes and 39 rRNA 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 AAQ100000000) but smaller than *Coxiella burnetii* strain ChuK_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 GJGC00000000 in GenBank. This version of the project (01) has been assigned the accession number GJGC01000000 and consists of sequences GJGC01000001 to GJGC01000006.

**ACKNOWLEDGMENT**

This work did not benefit from any external funding.

**REFERENCES**


Received 22 March 2012 Accepted 4 April 2012

Address correspondence to Didier Raoult, Didier.Raoult@gmail.com.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.00448-12