Genome Sequence of *Rickettsia conorii* subsp. *caspia*, the Agent of Astrakhan Fever

Erwin Sentausa, Khalid El Karkouri, Catherine Robert, Didier Raoult, and Pierre-Edouard Fournier

Unité de Recherche sur les Maladies Infectieuses et Tropicale Emergentes, UMR CNRS 7278, IRD 190, IFR48, Institut Hospitalo-Universitaire Méditerranée Infection, Faculté de Médecine, Aix-Marseille Université, Marseille, France

*Rickettsia conorii* subsp. *caspia* is the agent of Astrakhan fever, a spotted fever group rickettsiosis endemic to Astrakhan, Russia. The present study reports the draft genome of *Rickettsia conorii* subsp. *caspia* strain A-167.

Astrakhan fever was first reported in the 1970s in Astrakhan, a region of Russia located by the Caspian Sea (16, 17). In the early 1990s, the agent of this eruptive summer disease was isolated and described as a *Rickettsia conorii*-like bacterium named Astrakhan fever rickettsia, and the dog tick *Rhipicephalus pumilio* was suspected to be its vector (5, 6, 17). In 2001, the bacterium was also detected in *Rhipicephalus sanguineus* collected from dogs and an asymptomatic individual in Kosovo (7), whereas a similar bacterium was isolated from a patient with a febrile eruption who was returning from Chad (8). In 2005, Astrakhan fever rickettsia was proposed to be classified as a subspecies of *R. conorii* under the name *R. conorii* subsp. *caspia* (18). To date, the genomes of *R. conorii* subsp. *caspia* and *R. conorii* subsp. *indica* have been sequenced (13, 15). Here, we report the genome of *R. conorii* subsp. *caspia* strain A-167.

Genomic DNA of *R. conorii* subsp. *caspia* strain A-167 grown in Vero cells was pyrosequenced using the 454 GS FLX titanium platform (Roche, Branford, CT) (12) and assembled using Newbler 2.3 software (Roche). Potential coding sequences (CDSs) were predicted using AMIGene (3), and split genes or nonpredicted genes were detected and corrected manually where appropriate using Artemis (4) and BLASTN (1). Assignment of protein functions was performed by searching against the RickBase, GenBank, and Pfam databases using BLASTP (1, 2, 14), while ribosomal RNAs, tRNAs, and other RNAs were identified using BLASTN, tRNAscan-SE v.1.21 (19), and RNAmmer 1.2 (9).

The draft genome of *R. conorii* subsp. *caspia* consists of 25 contigs ranging in size from 1,481 to 123,719 bases (approximately 20-fold genome coverage), resulting in a total genome of 1,260,331 bases. We could not detect the presence of any plasmid in the genome. The GC content of 33% and the predicted total complement of 1,210 genes (1,636 open reading frames [ORFs]) are in the range of those for genomes from spotted fever group *Rickettsiae*. Among these genes, 820 (67.8%) are complete genes, 229 (18.9%) are split into two to 12 ORFs, and 78 (6.5%) are present only as fragments. Of the 1,210 genes, 859 (71.0%) were assigned putative functions, 229 (18.9%) are split into two to 12 ORFs, and 78 (6.5%) are present only as fragments. Of the 1,210 genes, 859 (71.0%) were assigned putative functions, 229 (18.9%) are split into two to 12 ORFs, and 78 (6.5%) are present only as fragments.

Orthologous genes between *R. conorii* subsp. *caspia*, *R. conorii* subsp. *indica*, and *R. conorii* subsp. *indica* were identified using OrthoMCL (10) with a BLASTP E value cutoff of 1 x 10^-5 and the default Markov cluster (MCL) inflation parameter of 1.5. The genomes of these bacteria were almost perfectly syntenic. However, we were not able to detect the gene for the chitin binding domain in either *R. conorii* subsp. *caspia* or *R. conorii* subsp. *indica*, while it was present as a split gene in *R. conorii* subsp. *caspia*.

**ACKNOWLEDGMENT**

This research did not benefit from any external funding.

**REFERENCES**