

Genome Sequence of the Asiatic Species *Borrelia persica*

Haitham Elbir,^a Pär Larsson,^b Johan Normark,^a Mukunda Upreti,^a Edward Korenberg,^c Christer Larsson,^a Sven Bergström^a

Department of Molecular Biology, Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå Centre for Microbial Research (UCMR), Umeå University, Umeå, Sweden^a; Division of CBRN Security and Defence, FOI Swedish Defence Research Agency, Umeå, Sweden^b; Gamaleya Research Institute for Epidemiology and Microbiology, Russian Academy of Medical Sciences, Moscow, Russian Federation^c

We report the complete genome sequence of *Borrelia persica*, the causative agent of tick-borne relapsing fever borreliosis on the Asian continent. Its genome of 1,784,979 bp contains 1,850 open reading frames, three ribosomal RNAs, and 32 tRNAs. One clustered regularly interspaced short palindromic repeat (CRISPR) was detected.

Received 5 December 2013 Accepted 7 December 2013 Published 9 January 2014

Citation Elbir H, Larsson P, Normark J, Upreti M, Korenberg E, Larsson C, Bergström S. 2014. Genome sequence of the Asiatic species *Borrelia persica*. *Genome Announc.* 2(1): e01127-13. doi:10.1128/genomeA.01127-13.

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

Address correspondence to Sven Bergström, sven.bergstrom@molbiol.umu.se.

Borrelia persica is the agent of relapsing fever (RF) in central Asia and Middle Eastern countries, but lack of awareness about the disease probably has led to an underestimation of its prevalence (1). Until now, *B. persica* has eluded *in vitro* cultivation, impeding its full characterization and *in vitro* experiments.

The *B. persica* strain No14 was isolated from *Ornithodoros papillipes* collected near the city of Khiva, Uzbekistan. The strain was isolated and propagated at the Laboratory of Vectors of Infections, Gamaleya Research Institute, Moscow, Russia, by passaging in guinea pigs (2). It was initially kept alive in a tick-guinea pig cycle and in 1988 it was stored at -70°C . After thawing in 2008, the strain was refractory to growth *in vitro* and was therefore propagated in CB17/Icr-Prkdc^{scid}/IcrIcoCrl mice (Taconic, Denmark) and subsequently successfully grown at 37°C in freshly made Barbour-Stoenner-Kelly II (BSK-II) medium with 1.4% (wt/vol) gelatin and 10% (vol/vol) rabbit serum (3). DNA was extracted using the Wizard genomic DNA purification kit (Promega Biotech AB, Sweden) and sequenced using an Illumina HiSeq 2000 sequencer. *De novo* assembly of reads was performed with the ABySS 1.3.4 assembler, and open reading frames (ORFs) were identified with Prodigal (4). tRNAs were predicted with the Aragorn software (5), and ribosomal RNAs were predicted with RNAmmer. The genes were further annotated using BLAST against the NCBI nonredundant database. The functional categorization of ORFs was performed using online RPS-BLAST against the cluster of orthologous groups (COG) (6) and Pfam databases (7). PHAST (8) and the Prophage Finder software were used for bacteriophage detection (9). Clustered regularly interspaced short palindromic repeats (CRISPRs) were predicted using the CRISPRfinder server (<http://crispr.u-psud.fr/Server/>). To estimate the similarity with published RF *Borrelia* genomes (10, 11), the average nucleotide identity (ANI) was calculated as described in Konstantinidis et al. (12).

A total of 1,795,421 bp of sequence was generated from 15,331,580 reads with 826-fold coverage and a G+C content of 28.7%. The scaffolds were ordered, resulting in a linear chromosome with a size of 923,419 bp and unclosed plasmids with a size of 872,002 bp. Gene finding resulted in 1,850 ORFs, three ribosomal RNAs, and 32 transfer RNAs, which is similar to the genomes of

other RF species (10, 11). The protein-coding regions represent 81% of the *B. persica* genome. Among the predicted genes, 43.4% were assigned a COG function. One CRISPR with three spacers was found. Comparing the chromosome of *B. persica* to the previously published genomes of *Borrelia duttonii*, *Borrelia recurrentis*, and *Borrelia crocidurae* revealed a high colinearity, with an average nucleotide identity of 89%.

Several genome sequences of African RF *Borrelia* spp. have been published (10, 11). The added *B. persica* genome sequence is of great importance as the first representative of Asian tick-borne RF borreliosis. An analysis of its genome sequence will provide further insights into its functional genomics and host adaptation. It will also facilitate future molecular and medical research on *B. persica*, a bacterium causing human disease, about which very little is known.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. [AYOT00000000](https://www.ncbi.nlm.nih.gov/nuccore/AYOT00000000). The version described in this paper is version AYOT01000000.

ACKNOWLEDGMENTS

We thank Ingela Nilsson for technical assistance.

This study was supported by Swedish Research Council grant no. 07922 to S.B. and a postdoctoral grant to H.E. from MIMS.

REFERENCES

1. Barmaki A, Rafinejad J, Vatandoost H, Telmadarraiy Z, Mohtarami F, Leghaei Sh, Oshaghi M. 2010. Study on presence of *Borrelia persica* in soft ticks in Western Iran. *Iran J. Arthropod. Borne Dis.* 4:19–25.
2. Kryuchevnikov VN, Korenberg EI, Sherbakov SV. 1982. Experience on the identifying of strains differences in relapsing fever pathogen. *Med. Parasitol. Parasit. Dis.* 4:74–79. (In Russian.)
3. Barbour AG. 1984. Isolation and cultivation of Lyme disease *Spirochetes*. *Yale J. Biol. Med.* 57:521–525.
4. 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>.
5. 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>.

6. 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>.
7. Sonnhammer EL, Eddy SR, Durbin R. 1997. Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins* 28:405–420. [http://dx.doi.org/10.1002/\(SICI\)1097-0134\(199707\)28:3<405::AID-PROT10>3.0.CO;2-L](http://dx.doi.org/10.1002/(SICI)1097-0134(199707)28:3<405::AID-PROT10>3.0.CO;2-L).
8. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. *Nucleic Acids Res.* 39(Web Server issue):W347–W352. <http://dx.doi.org/10.1093/nar/gkr485>.
9. Bose M, Barber RD. 2006. Prophage Finder: a prophage loci prediction tool for prokaryotic genome sequences. *In Silico Biol.* 6:223–227.
10. Lescot M, Audic S, Robert C, Nguyen TT, Blanc G, Cutler SJ, Wincker P, Couloux A, Claverie JM, Raoult D, Drancourt M. 2008. The genome of *Borrelia recurrentis*, the agent of deadly louse-borne relapsing fever, is a degraded subset of tick-borne *Borrelia duttonii*. *PLoS Genet.* 4:e1000185. <http://dx.doi.org/10.1371/journal.pgen.1000185>.
11. Elbir H, Gimenez G, Robert C, Bergström S, Cutler S, Raoult D, Drancourt M. 2012. Complete genome sequence of *Borrelia crocidurae*. *J. Bacteriol.* 194:3723–3724. <http://dx.doi.org/10.1128/JB.00118-12>.
12. Konstantinidis KT, Tiedje JM. 2005. Towards a genome-based taxonomy for prokaryotes. *J. Bacteriol.* 187:6258–6264. <http://dx.doi.org/10.1128/JB.187.18.6258-6264.2005>.