Complete Genome Sequence of *Francisella tularensis* subsp. *holarctica* Strain A271_1 (FDC408), Isolated from a Eurasian Beaver (*Castor fiber*)

David Sundell,[a] Ingrid Uneklint,[a] Caroline Öhrman,[a] Emelie Salomonsson,[a] Linda Karlsson,[a] Stina Bäckman,[a] Jonas Näslund,[a] Andreas Sjödin,[a] Mats Forsman,[a] Sandra Appelt,[b] Oliver Drechsel,[c] Daniela Jacob,[b] Klaus Heuner,[d] Kerstin Myrtennäs[a]

[a]Division of CBRN Defence and Security, Swedish Defence Research Agency (FOI), Umeå, Sweden
[b]Centre for Biological Threats and Special Pathogens (ZBS2), Robert Koch Institute (RKI), Berlin, Germany
[c]Methodology and Research Infrastructure Bioinformatics (MF1), Robert Koch Institute (RKI), Berlin, Germany
[d]Cellular Interactions of Bacterial Pathogens (ZBS2), Robert Koch Institute (RKI), Berlin, Germany

David Sundell and Ingrid Uneklint contributed equally to this work. Author order was determined by relative contribution.

ABSTRACT Here, we report the complete genome sequence of *Francisella tularensis* subsp. *holarctica* strain A271_1, isolated from a Eurasian beaver (*Castor fiber*) in 2012 in the Berlin/Brandenburg region, Germany.

Tularemia is a zoonosis caused by *Francisella tularensis* (1). Susceptibility to this disease differs between the Eurasian beaver (*Castor fiber*) and the American beaver (*Castor canadensis*), where the latter is more sensitive (2–4).

Here, we present the complete genome sequence of *F. tularensis* subsp. *holarctica* strain A271_1. For further information about the isolation, cultivation, draft genome sequence, and further characterization of this strain, see reference 5. Prior to sequencing, the strain was cultured from the original glycerol stock (−80°C), and DNA was extracted as previously described (5). Human serum (Seraclot, PAN-Biotech, Germany) was spiked with DNA and diluted 1:10 in phosphate-buffered saline (PBS) to an unmeasurable genomic DNA concentration on both a Qubit fluorometer (Thermo Fisher Scientific, USA) and a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific). DNA was extracted using MagAttract (Qiagen, Germany) and amplified using a REPLI-g midikit (Qiagen) to a concentration of 1,518 ng/μl. The first sequencing library was prepared with the Nextera XT DNA library prep kit v3 (Illumina, Inc., USA) and sequenced on a MiSeq instrument (Illumina, Inc.) using a MiSeq reagent kit v3 (600 cycles). After endonuclease treatment of the amplified DNA with T7 endonuclease I (New England Biolabs, USA), the second sequencing library was prepared using a ligation sequencing kit (SQK-LSK108) with the native barcoding expansion kit (EXP-NBD103) and sequenced using a MinION FLO-MIN107 flow cell with R9.4 chemistry (Oxford Nanopore Technologies, Ltd., UK). No DNA size selection or fragmentation was performed prior to sequencing. The sequencing generated 4,275,968 trimmed paired-end reads (300 bp) and 424,221 long reads (median read length, 1,763 bp; read length N_{50}, 4,027 bp). Illumina reads were trimmed using Trimmomatic v0.36 (LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15, MINLEN:36), removing the low-quality bases at the ends of reads (6). Nanopore reads were base called using Albacore v2.1.3 (Oxford Nanopore Technologies, Ltd.), and adapters were trimmed with Porechop v0.2.3_seqan2.1.1 (7). One 1,862,390-bp linear sequence was generated with Unicycler v0.4.7, using both short Illumina reads and long Nanopore reads as input (8). The coverage of the 30-kb *Francisella* pathogenicity island (FPI) was twice as high as the average coverage (>100×) of the genome, indicating a duplicated region. This finding was expected since the FPI is duplicated in all known genomes of this

subspecies (9). The second copy of the FPI was added, guided by the complete genome of strain FSC200 (GenBank accession no. CP003862.1) (10). By mapping Illumina reads using Bowtie 2 v2.3.4.3 (10) and Nanopore reads longer than 10 kbp using minimap2 v2.15 (11) to the complete genome of A271_1 with two FPI copies (12), we could confirm having both long and short reads spanning all four junctions of the FPIs with over 10× coverage without any introduced single nucleotide polymorphisms (SNPs) (13). Finally, the genome was rotated to start with *dnaA* using Circlator v1.5.5 (14). Annotation was performed using the NCBI Prokaryotic Genome Annotation Pipeline (22 January 2020) (15, 16). We confirmed with CanSNPer2 (17) (https://github.com/FOI-Bioinformatics/CanSNPer2) that the strain is a member of the B.12 subclade B.75. Default parameters were used for all software unless otherwise specified.

The 1,893,679-bp circular chromosome of A271_1 has a GC content of 32.16% with 1,591 protein-coding sequences, 280 pseudogenes, 10 rRNAs, 38 tRNAs, and 4 noncoding RNAs. The new assembly contained 11 single base alternates and one single base insertion, compared to the draft genome (accession no. GCA_002102455.1). The changes were identified by dnadiff v1.3 and were carefully verified with the read data using IGV v2.3. All differences were located near insertion elements (IS) and within 218 bp from the contig ends, regions known to be problematic for assemblies based solely on Illumina sequences.

Data availability. The complete genome sequence of strain A271_1 (BioProject accession no. PRJNA602379) has been deposited in GenBank under the accession no. CP048229.1 and the SRA accession no. SRR11853651 (Oxford Nanopore) and SRR11853652 (Illumina). The version described in this paper is the second version. The first version of the A271_1 (FDC408) genome sequence (BioProject accession no. PRJNA285142) was a draft genome sequence and has the GenBank and SRA accession no. GCA_002102455.1 and SRR5051916, respectively.

REFERENCES

1. Ellis J, Oyston PCF, Green M, Titball RW. 2002. Tularemia. Clin Microbiol Rev 15:631–646. https://doi.org/10.1128/cmr.15.4.631-646.2002.
2. Sjöstedt A. 2007. Epidemiology, p 5–10. *In* Tärnvik A (ed), WHO guidelines on Tularaemia: epidemic and pandemic alert and response. WHO Press, Geneva, Switzerland.
3. Mörner T, Sandstedt K. 1983. A serological survey of antibodies against Francisella tularensis in some Swedish mammals. Nord Vet Med 35:82–85.
4. Mörner T, Sandström G, Mattsson R. 1988. Comparison of serum and lung extracts for surveys of wild animals for antibodies to Francisella tularensis biovar palaearctica. J Wildl Dis 24:10–14. https://doi.org/10.7589/0090-3558-24.1.10.
5. Schulze C, Heuner K, Myrtennäs K, Karlsson E, Jacob D, Kutzer P, Große K, Forsman M, Grunow R. 2016. High and novel genetic diversity of Francisella tularensis in Germany and indication of environmental persistence. Epidemiol Infect 144:3025–3036. https://doi.org/10.1017/S0950268816001175.
6. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
7. Wick R. 2012. Porechop: adapter trimmer for Oxford Nanopore reads. https://github.com/rrwick/Porechop.
8. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.
9. Larsson P, Oyston PCF, Chain P, Chu MC, Duffield M, Fuxelius H-H, Garcia E, Hälltorp G, Johansson D, Isherwood KE, Karp PD, Larsson E, Liu Y, Michell S, Prior J, Prior R, Malfatti S, Sjöstedt A, Svensson K, Thompson N, Vergez L, Wagg JK, Wren BW, Lindler LE, Andersson SGE, Forsman M, Titball RW. 2005. The complete genome sequence of Francisella tularensis, the causative agent of tularemia. Nat Genet 37:153–159. https://doi.org/10.1038/ng1499.
10. Svensson K, Sjödin A, Byström M, Granberg M, Brittnacher MJ, Rohmer L, Jacobs MA, Sims-Day EH, Levy R, Zhou Y, Hayden HS, Lim R, Chang J, Guenthener D, Kang A, Haugen E, Gillett W, Kaul R, Forsman M, Larsson P, Johansson A. 2012. Genome sequence of Francisella tularensis subspecies holarctica strain FSC200, isolated from a child with tularemia. J Bacteriol 194:6965–6966. https://doi.org/10.1128/JB.01040-12.
11. Li H. 2018. minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
12. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
13. Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics

Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14:178–192. https://doi.org/10.1093/bib/bbs017.

14. Hunt M, De Silva N, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. Genome Biol 16:294. https://doi.org/10.1186/s13059-015-0849-0.

15. Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, Li W, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu F, Marchler GH, Song JS, Thanki N, Yamashita RA, Zheng C, Thibaud-Nissen F, Geer LY, Marchler-Bauer A, Pruitt KD. 2018. RefSeq: an update on prokaryotic genome annotation and curation. Nucleic Acids Res 46:D851–D860. https://doi.org/10.1093/nar/gkx1068.

16. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.

17. Lärkeryd A, Myrtennäs K, Karlsson E, Dwibedi CK, Forsman M, Larsson P, Johansson A, Sjödin A. 2014. CanSNPer: a hierarchical genotype classifier of clonal pathogens. Bioinformatics 30:1762–1764. https://doi.org/10.1093/bioinformatics/btu113.