genomeA
Journals.ASM.org

Complete Genome Sequence of *Francisella guangzhouensis* Strain 08HL01032^T, Isolated from Air-Conditioning Systems in China

Daniel Svensson,[a,b] Caroline Öhrman,[a] Stina Bäckman,[a] Edvin Karlsson,[a] Elin Nilsson,[a] Mona Byström,[a] Adrian Lärkeryd,[a] Kerstin Myrtennäs,[a] Per Stenberg,[a] Ping-hua Qu,[c] Johan Trygg,[b] Holger C. Scholz,[d] Mats Forsman,[a] Andreas Sjödin[a,b]

Division of CBRN Security and Defence, FOI–Swedish Defence Research Agency, Umeå, Sweden[a]; Department of Chemistry, Computational Life Science Cluster (CLiC), Umeå University, Umeå, Sweden[b]; Department of Clinical Laboratory, Guangdong Provincial Hospital of Traditional Chinese Medicine, Guangzhou, China[c]; Department of Bacteriology and Toxinology, German Center for Infection Research (DZIF), Bundeswehr Institute of Microbiology, Munich, Germany[d]

D.S. and C.Ö. contributed equally to this work.

We present the complete genome sequence of *Francisella guangzhouensis* strain 08HL01032^T, which consists of one chromosome (1,658,482 bp) and one plasmid (3,045 bp) with G+C contents of 32.0% and 28.7%, respectively.

Received 3 February 2015 **Accepted** 5 February 2015 **Published** 19 March 2015

Citation Svensson D, Öhrman C, Bäckman S, Karlsson E, Nilsson E, Byström M, Lärkeryd A, Myrtennäs K, Stenberg P, Qu P-H, Trygg J, Scholz HC, Forsman M, Sjödin A. 2015. Complete genome sequence of *Francisella guangzhouensis* strain 08HL01032^T, isolated from air-conditioning systems in China. Genome Announc 3(2):e00024-15. doi:10.1128/genomeA.00024-15.

Address correspondence to Andreas Sjödin, andreas.sjodin@foi.se.

The known diversity of the genus *Francisella* has recently expanded substantially to include several new species. Genome analyses of representative *Francisella* isolates have shown that the genus can be divided into two main genetic clades (1). The first clade contains *F. tularensis*, *F. novicida*, *F. hispaniensis*, and *Francisella*-like endosymbionts (FLE), while the second clade contains *F. noatunensis* and *F. philomiragia*.

F. guangzhouensis type strain 08HL01032^T was isolated from air-conditioning systems in China in 2008 (2, 3). The live strain was obtained from the Public Health England Culture Collection (NCTC 13503) and assigned identifier FSC996 in the *Francisella* Strain Collection. The strain was grown on heart cysteine agar (HCA), and its DNA was extracted using standard techniques (4). Illumina HiSeq instruments generated a total of 17,862,240 paired-end reads (100 bp), with an average insert size of 540 bp, and 20,469,388 mate-pair reads (49 bp), with an average insert size of 4,927 bp. A Pacific Biosciences RSII system (10-kb library, 2-h movie length) generated a total of 90,148 PacBio reads, with an average read length of 3,575 kb, using two single-molecular real-time (SMRT) cells.

The initial draft of the genome was generated by assembling Illumina pair-end reads using the Edena version 3 assembler (5). Scaffolding was performed using Illumina mate-pair reads in SSPACE (6). The SMRT Analysis system version 2.2.0.p3 was used to assemble a second draft genome for PacBio reads. The two draft genomes were compared using progressiveMauve and MUMmer (7), and three regions of discrepancy were manually curated to generate the final assembly.

The final assembly consists of two scaffolds, one for the main chromosome and one for a circular plasmid. The main chromosome contains 1,658,482 bp with a G+C content of 32.0%, and the plasmid contains 3,045 bp with a G+C content of 28.7%. Annotation was carried out using the NCBI annotation service.

F. guangzhouensis strain 08HL01032^T contains 1,423 protein-coding sequences, 75 pseudogenes, 10 rRNAs, 38 tRNAs, and 1 noncoding RNA. The average nucleotide identity (ANI) was calculated by pairwise genome comparisons for publically available genomes within clade I and clade II (1) using JSpecies version 1.2.1 (8). The similarity between *F. guangzhouensis* and clade I genomes was 75.6% to 75.2% and to clade II genomes 80.0% to 74.8%, respectively. Commonly, a threshold of >95% to 96% identity is used to classify genomes as belonging to the same species (9). The plasmid was 89% identical over 1,408 bp to *F. philomiragia* plasmid pF242 (10). The phylogeny shows that *F. guangzhouensis* does not belong to any of the two previously known *Francisella* main clades or the recently published *F. endociliophora* clade (11). This isolate forms a new separate branching clade in the *Francisella* genus. The updated knowledge is essential for improving assays to be used in epidemiological studies of *Francisella* (12).

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession numbers CP010427 (chromosome) and CP010428 (plasmid). The versions described in this paper are the first versions, CP010427.1 and CP010428.1.

ACKNOWLEDGMENTS

This work was supported by the Swedish Ministry of Defence (A4040) and the Swedish Ministry of Foreign Affairs (A4052). We acknowledge the support of the National Genomics Infrastructure (NGI)/Uppsala Genome Center and UPPMAX for providing assistance in massive parallel sequencing and computational infrastructure. The work performed at NGI/Uppsala Genome Center was funded by RFI/VR and Science for Life Laboratory, Sweden.

REFERENCES

1. **Sjödin A, Svensson K, Ohrman C, Ahlinder J, Lindgren P, Duodu S, Johansson A, Colquhoun DJ, Larsson P, Forsman M.** 2012. Genome characterisation of the genus *Francisella* reveals insight into similar evolutionary paths in pathogens of mammals and fish. BMC Genomics **13**:268. http://dx.doi.org/10.1186/1471-2164-13-268.
2. **Qu P-H, Chen S-Y, Scholz HC, Busse H-J, Gu Q, Kämpfer P, Foster JT, Glaeser SP, Chen C, Yang Z-C.** 2013. *Francisella guangzhouensis* sp. nov.,

isolated from air-conditioning systems. Int J Syst Evol Microbiol **63:** 3628–3635. http://dx.doi.org/10.1099/ijs.0.049916-0.

3. **Qu P, Deng X, Zhang J, Chen J, Zhang Q, Xiao Y, Chen S.** 2009. Identification and characterization of the *Francisella* sp. strain 08HL01032 isolated in air condition systems. Wei Sheng Wu Xue Bao **49:**1003–1010.
4. **Larsson P, Oyston PC, 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. http://dx.doi.org/10.1038/ng1499.
5. **Hernandez D, Tewhey R, Veyrieras J-B, Farinelli L, Østerås M, François P, Schrenzel J.** 2013. *De novo* finished 2.8 Mbp *Staphylococcus aureus* genome assembly from 100 bp short and long range paired-end reads. Bioinformatics **30:**40–49. http://dx.doi.org/10.1093/bioinformatics/btt590.
6. **Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W.** 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27:** 578–579. http://dx.doi.org/10.1093/bioinformatics/btq683.
7. **Darling AE, Mau B, Perna NT.** 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One **5:**e11147. http://dx.doi.org/10.1371/journal.pone.0011147.
8. **Richter M, Rosselló-Móra R.** 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci USA **106:** 19126–19131. http://dx.doi.org/10.1073/pnas.0906412106.
9. **Kim M, Oh H-S, Park S-C, Chun J.** 2014. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol **64:**346–351. http://dx.doi.org/10.1099/ijs.0.059774-0.
10. **Le Pihive E, Blaha D, Chenavas S, Thibault F, Vidal D, Valade E.** 2009. Description of two new plasmids isolated from *Francisella philomiragia* strains and construction of shuttle vectors for the study of *Francisella tularensis*. Plasmid **62:**147–157. http://dx.doi.org/10.1016/j.plasmid.2009.07.001.
11. **Sjödin A, Ohrman C, Bäckman S, Lärkeryd A, Granberg M, Lundmark E, Karlsson E, Nilsson E, Vallesi A, Tellgren-Roth C, Stenberg P, Thelaus J.** 2014. Complete genome sequence of *Francisella endociliophora* strain FSC1006, isolated from a laboratory culture of the marine ciliate *Euplotes raikovi*. Genome Announc **2**(6):e01227-14. http://dx.doi.org/10.1128/genomeA.01227-14.
12. **Ahlinder J, Öhrman C, Svensson K, Lindgren P, Johansson A, Forsman M, Larsson P, Sjödin A.** 2012. Increased knowledge of *Francisella* genus diversity highlights the benefits of optimised DNA-based assays. BMC Microbiol **12:**220. http://dx.doi.org/10.1186/1471-2180-12-220.