Francisella tularensis is a highly infectious bacterium with the potential to cause high fatality rates if infections are untreated. To aid in the development of rapid and accurate detection assays, we have sequenced and annotated the genomes of 18 F. tularensis and Francisella philomiragia strains.

Francisella tularensis is the causative agent of tularemia, a zoonotic infection that is mainly transmitted to humans through arthropod bites, direct contact with infected animals, or the inhalation/ingestion of contaminated materials (dust, food, or water). These Gram-negative, nonmotile, facultatively intracellular bacteria are highly infectious (1). The low infectious dose (10 to 50 organisms) and previous weaponization attempts have led to the placement of F. tularensis on the U.S. Centers for Disease Control and Prevention category A bioterrorism agent list (2–4). In 1970, a World Health Organization (WHO) expert committee reported that if sufficient quantities of F. tularensis were dispersed over a metropolitan area with 5 million people, it could result in approximately 250,000 acute febrile nonspecific illnesses 3 to 5 days post-release and 19,000 deaths (5).

Francisella is endemic to the United States and resides within small mammals (mice, voles, rats, squirrels, rabbits, and hares) that serve as natural reservoirs. There are three subspecies of F. tularensis: F. tularensis subsp. tularensis, which can be highly virulent, with a mortality rate of 30 to 60% (without treatment), and the occasional human pathogens F. tularensis subsp. holarctica and F. tularensis subsp. mediaisatica. The close phylogenetic relationship of highly virulent strains to opportunistic pathogens and near neighbors has caused taxonomic and strain differentiation difficulties.

Here, we present the annotated genome assemblies for all but one strain listed in the Standard Method Performance Requirements (SMPR) as derived by the Stakeholder Panel on Agent Detection Assays (SPADA) (FTTN10 is not currently maintained in a publicly available culture collection). The 18 genome sequences presented here were assembled to finished or improved high-quality draft (IHQD) status (6).

The draft genome assemblies included two or more data sets (data types and coverages are listed in the NCBI records): Illumina (short- and/or long-insert paired data), Roche 454 (long-insert paired data), and PacBio long reads. The short- and long-insert paired data were assembled in both Newbler and Velvet (7) and computationally shredded into 1.5-kbp overlapping shreds. If the PacBio coverage was ≥100×, the data were assembled using HGAP (8). All data were additionally assembled in AllPaths (9). The consensus sequences from both HGAP and AllPaths were computationally shredded into 10-kbp overlapping pieces. All shreds were integrated using Phrap. Possible misassemblies were corrected and repeat regions verified using in-house scripts and manual editing in Consed (10–12). When combined with the long-insert (~8 to 10-kb) Illumina data, the HGAP assemblies of recent PacBio data are generally capable of reconstructing the ~30-kb pathogenicity island repeat found in Francisella genomes. These methods resulted in a finished quality genome for 17 of the 18 isolates (Table 1) (6). Each genome assembly was annotated using an Ergatis-based (13) workflow with minor manual curation.

The genome assemblies range from 1.86 to 2.15 Mb (Table 1), with one chromosome and up to two plasmids. As expected for the genus, the G+C% was low, ranging from 32 to 33%, and genomes have between 1,769 and 2,107 coding sequences.

Nucleotide sequence accession numbers. The accession numbers for all 18 genome sequences are listed in Table 1.
TABLE 1  Listing of Francisella isolate genomes released to NCBI as listed in the SMPR by the Association of Analytical Communities

| Organism name | Panel | AOAC name | GenBank accession no. | Quality | Assembly size (bp) | No. of plasmids |
|---------------|-------|-----------|-----------------------|---------|-------------------|-----------------|
| F. tularensis subsp. tularensis | W96   | FT9       | CP010103              | Finished| 2,047,033         | 1               |
|                | Schu_54| FT4       | CP010290              | Finished| 1,892,789         | 0               |
|                | Scherr | FT8       | CP010287              | Finished| 1,890,922         | 0               |
|                | NIH B-38| FT1      | CP010115              | Finished| 1,862,444         | 0               |
| F. tularensis subsp. novicida | U112  | FTNN9     | CP009633              | Finished| 1,910,592         | 0               |
|                | F6168 | FTNN8     | CP009352–CP009353     | Finished| 1,927,240         | 1               |
|                | D9876 | FTNN7     | CP009607              | Finished| 1,870,206         | 0               |
| F. tularensis subsp. holarctica | VT68  | FT6       | CP010288              | Finished| 1,893,733         | 0               |
|                | LVS   | FT2       | CP009694              | Finished| 1,892,177         | 0               |
|                | JAP   | FT5       | IUJU00000000          | IHQD (2) | 1,907,159         | 0               |
|                | FTT_1 | FT3       | CP009693              | Finished| 1,892,758         | 0               |
|                | 425   | FT7       | CP010289              | Finished| 1,894,186         | 0               |
| F. philomiragia | O#319l| FTNN1     | CP010019              | Finished| 2,017,400         | 0               |
|                | O#319-067| FTNN4    | CP009436–CP009437     | Finished| 2,049,711         | 1               |
|                | O#319-036| FTNN3    | CP009442–CP009443     | Finished| 1,924,061         | 1               |
|                | O#319-029| FTNN2    | CP009342–CP009343     | Finished| 2,044,934         | 1               |
|                | GA01-2801| FTNN6   | CP009444–CP009446     | Finished| 2,033,714         | 2               |
|                | GA01-2794| FTNN5    | CP009440–CP009441     | Finished| 2,152,054         | 1               |

a I, inclusivity strain; E, exclusivity strain.
b AOAC, Association of Analytical Communities.
c Quality refers to the quality of the genome assembly, as defined by Chain et al (6). The number of contigs for the IHQD assembly is listed in parentheses.

ACKNOWLEDGMENTS

Funding for this effort was provided by the Defense Threat Reduction Agency’s Joint Science and Technology Office (DTRA J9-CB/JSTO) and the Department of Homeland Security Science and Technology Directorate, award HSHQDC-08-X-00790.

This paper is approved by the LANL for unlimited release (LA-UR-14-29667). The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, or the U.S. Government.

The bacterial strains were obtained from the Department of Defense’s Unified Culture Collection (http://www.usamriid.army.mil/ucc/) or BEI Resources (http://beiresources.org/).

REFERENCES

1. Sjostedt A. 2005. Family XVII. Francisellaceae, genus I. Francisella. In Brenner D, Krieg N, Staley J, Garrity G, Boone D (ed), Bergey’s manual of systematic bacteriology. Springer, New York, NY.
2. Forsman M, Sandstrom G, Sjostedt A. 1994. Analysis of 16S ribosomal DNA sequences of Francisella strains and utilization for determination of the phylogeny of the genus and for identification of strains by PCR. Int J Syst Bacteriol 44:38–46. http://dx.doi.org/10.1099/00207713-44-1-38.
3. Jones RM, Nicas M, Hubbard A, Sylvester MD, Reingold A. 2005. The infectious dose of Francisella tularensis (tularemia). Appl Biosafety 10:227.
4. Rozt LD, Khan AS, Lillibridge SR, Ostroff SM, Hughes JM. 2002. Public health assessment of potential biological terrorism agents. Emerg Infect Dis 8:225–230. http://dx.doi.org/10.3201/eid0802.010164.
5. World Health Organization. 1970. Health aspects of chemical and biological weapons: report of a WHO group of consultants. World Health Organization, Geneva, Switzerland. https://extranet.who.int/iris/restricted/handle/10665/39444.
6. Chain PS, Grafham DV, Fulton RS, FitzGerald MG, Hostetler J, Muzny D, Ali J, Birren B, Bruce DC, Buahy C, Cole JR, Ding Y, Dugan S, Fied D, Garrity GM, Gibbs R, Graves T, Han CS, Harrison SH, Highland S, Hugenholtz P, Khouri HM, Kodira CD, Kolker E, Kyrpides NC, Lang D, Lapidus A, Maffartia SA, Markowitz V, Mertha T, Nelson KE, Parkhill J, Pitluck S, Qin X, Read TD, Schmutz J, Sozhamannan S, Sterk P, Strausberg RL, Sutton G, Thomson NR, Tiedje JM, Weinstock G, Wollam A, Genomic Standards Consortium Human Microbiome Project Jumpstart Consortium, Detter JC. 2009. Genomics. Genome project standards in a new era of sequencing. Science 326:236–237. http://dx.doi.org/10.1126/science.1180614.
7. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Brujin graphs. Genome Res 18:821–829. http://dx.doi.org/10.1101/gr.074492.107.
8. Chin C-S, Alexander DH, Marks P, Klammer AA, Revanna KV, Dong Q, Gordon D, Abajian C, Green P. 2007. Consed: a graphical tool for sequence finishing. Genome Res 17:175–185. http://dx.doi.org/10.1101/gr.7337908.
9. Butler J, MacCallum I, Kleber M, Shlyakhter IA, Belmonte MK, Lander ES, Nusbaum C, Jaffe DB. 2008. ALLPATHS: de novo assembly of whole-genome shotgun microreads. Genome Res 18:810–820. http://dx.doi.org/10.1101/gr.077390.108.
10. Ewing B, Green P. 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. Genome Res 8:161–170. http://dx.doi.org/10.1101/gr.1700703.
11. Ewing B, Hillier L, Wendell MC, Green P. 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res 8:161–170. http://dx.doi.org/10.1101/gr.1700703.
12. Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. Genome Res 8:195–202. http://dx.doi.org/10.1101/gr.8.3.195.
13. Hemmerich C, Bucheclin A, Podicheti R, Revanna KV, Dong Q. 2010. An Ergatis-based prokaryotic genome annotation Web server. Bioinformatics 26:1122–1124. http://dx.doi.org/10.1093/bioinformatics/btq090.