Complete Genome Sequences of *Francisella marina* Strains E95-16 and E103-15, Isolated from Maricultured Spotted Rose Snapper (*Lutjanus guttatus*) on the Pacific Coast of Central America

Hasan C. Tekedar,ª Matt J. Griffin,ª Larry A. Hanson,ª Geoffrey C. Waldbieser,ᵇ Esteban Sotoᶜ

ªCollege of Veterinary Medicine, Mississippi State University, Mississippi State, Mississippi, USA
ᵇWarmwater Aquaculture Research Unit, Agriculture Research Service, U.S. Department of Agriculture, Stoneville, Mississippi, USA
ᶜDepartment of Medicine and Epidemiology, School of Veterinary Medicine, University of California, Davis, Davis, California, USA

ABSTRACT In 2015 and 2016, a previously unrecognized *Francisella* sp. was isolated from disease outbreaks in maricultured spotted rose snapper (*Lutjanus guttatus*) on the Pacific coast of Central America. Polyphasic analysis demonstrated these bacteria differed from any known *Francisella* spp. Here, the complete genomes from the recently described *Francisella marina* strains are released.

*F*rancisella spp. are pleomorphic Gram-negative, nonmotile organisms in the family *Francisellaceae* within the class *Gammaproteobacteria*. Two subspecies, *Francisella noatunensis* subsp. *noatunensis* and *Francisella noatunensis* subsp. *orientalis*, are associated with disease outbreaks in a range of cultured and wild fish species from the Americas, Asia, and Europe (1). Research indicates that many undescribed *Francisella* spp. exist in aquatic environments. Recently, *Francisella marina* sp. nov. was described from mortality events in cultured spotted rose snapper (*Lutjanus guttatus*) in Central America (2). Here, the complete circular genomes of *F. marina* isolates E95-16 and E103-15 are reported.

Bacterial strains were cultured in 9 ml porcine brain heart infusion broth (Becton, Dickinson, Franklin Lakes, NJ, USA) in static, overnight cultures at 28°C. Aliquots (3 ml) of expanded cultures were pelleted by centrifugation (20,000 × *g*). High-molecular-weight genomic DNA was isolated from the concentrated pellets using the Puregene DNA isolation kit (Qiagen, Germantown, MD). Long sequencing reads were produced on a GridION platform (Oxford Nanopore Technologies [ONT], Oxford, UK) using the ligation sequencing kit (LSK109) and v9.4.1 flow cells. Nanopore reads were filtered to a minimal quality score of 13 and a minimal length of 1,000 bp using NanoFilt v2.2.0 (3), producing 461 Mb with an average length of 4.4 kb for E95-16 and 6.9 Mb with an average length of 5.4 kb for E103-15. Long reads from each isolate were assembled with Canu v1.8 (4) using default parameters and declaring an estimated genome size of 2 Mb. An Illumina Nextera XT library was produced from genomic DNA, and paired-end sequences were obtained on a NextSeq platform (Illumina, Inc., San Diego, CA). Paired Illumina reads were trimmed using Trimmomatic v0.38 (LEADING:30 TRAILING:30 SLIDINGWINDOW:4:30 MINLEN:50), and trimmed paired short reads were mapped to each Nanopore contig using minimap2 v2.12 (5). Insertions, deletions, and single-nucleotide variations were corrected using three iterations of Pilon v1.23 (6). Each assembly produced two contigs of approximately 2 Mb and 4 kb, with 101× coverage for E95-16 and 53× coverage for E103-15. Overlapping sequences were identified through a text search and manually trimmed to remove the overlap. The trimmed contig was then cut at position 1000000 and recircularized, and long reads were mapped to the contig to verify continuous coverage across the cut site.

TABLE 1 ANI and GGDC values for newly proposed novel *Francisella marina* genomes and other *Francisella* genus members[a]

	Francisella marina E95-16	*Francisella marina* E103-15	*Francisella* sp. TX077308	*Francisella noatunensis* subsp. *orientalis* LADL-07-285A	*Francisella philomiragia* subsp. *philomiragia* ATCC 25015	*Francisella* sp. FSC1006	*Francisella* sp. TX077310	*Francisella tularensis* subsp. *holarctica* F92	*Francisella tularensis* subsp. *novicida* PA10-7858	*Francisella tularensis* subsp. *tularensis* SCHU S4	
Francisella marina E95-16		99.70%	80.30%	36.00%	39.90%	22.20%	22.00%	24.00%	24.10%	24.10%	GGDC calculation
Francisella marina E103-15	99.95		80.70%	36.00%	40.00%	22.10%	22.00%	24.00%	24.00%	24.00%	
Francisella sp. TX077308	97.9	97.9		35.60%	39.80%	22.00%	22.00%	24.10%	24.40%	23.90%	
Francisella noatunensis subsp. *orientalis* LADL-07-285A	88.62	89	88.78		50.70%	21.80%	21.90%	24.10%	23.90%	23.90%	
Francisella philomiragia subsp. *philomiragia* ATCC 25015	90.47	90.48	90.16	93.13		22.20%	22.00%	25.00%	25.00%	24.50%	
Francisella sp. FSC1006	78.39	78.37	78.47	77.95	78.37		23.00%	22.20%	22.00%	21.80%	
Francisella sp. TX077310	78.08	77.94	78.22	77.78	77.91	79.19		22.80%	22.80%	22.40%	
Francisella tularensis subsp. *holarctica* F92	80.62	80.65	80.5	80.81	81.58	78.19	78.35		81.60%	94.60%	
Francisella tularensis subsp. *novicida* PA10-7858	81.05	80.79	81.25	81.34	81.82	78.84	78.65	97.99		82.80%	
Francisella tularensis subsp. *tularensis* SCHU S4	80.69	80.55	80.67	80.77	81.35	78.51	78.58	99.29	98.11		
	ANI calculation										

[a]Scores of >95% (ANI) or 70% (GGDC) indicate conspecificity (15, 16).

The *F. marina* genomes and plasmids were submitted to the National Center for Biotechnology Information (NCBI) Prokaryotic Genome Annotation Pipeline (PGAP) (7) and Rapid Annotation of microbial genomes using Subsystems Technology (RAST) (8). The Comprehensive Antibiotic Resistance Database (CARD) was used to screen for antibiotic resistance elements (9). Secretion systems were identified with MacSyFinder (10, 11). The Genome-to-Genome Distance Calculator (GGDC) (12) and average nucleotide identity (ANI) analysis (13) were used to identify the relatedness of the *F. marina* strains to other *Francisella* spp. (Table 1). Default parameters were used for all of the tools used in this study. Genome-level similarity estimations for the two *F. marina* strains against other closely related *Francisella* spp. reaffirm previous work establishing *F. marina* as a valid taxon (2), most similar to an undescribed *Francisella* sp. isolate, TX077308 (GenBank accession no. CP002872), which was cultured from seawater from the Gulf of Mexico (14). The genome of *F. marina* strain E95-16 comprises 2,058,175 bp with a GC content of 32.9%, with a 4,106-bp plasmid. The chromosome contains 1,945 proteins, 10 rRNAs, 39 tRNAs, and 4 other RNAs. Comparably, the genome of *F. marina* E103-15 comprises 2,041,198 bp with a GC content of 32.9% and an identical 4,106-bp plasmid. Identified putative genes on the E103-15 chromosome encode 1,936 proteins, 10 rRNAs, 39 tRNAs, and 4 other RNAs.

Based on the functional annotation results, the complete *Francisella marina* E95-16 and E103-15 genomes and plasmids share the same functional elements with one exception; the strain E95-16 genome encodes more type I restriction modification system elements. These genomes carry two different type VI secretion systems (T6SSs) (T6SSi and T6SSii) and one type IV pilus (T4P). The only antibiotic resistance element identified in either genome was FPH-1, a carbapenem-hydrolyzing class A β-lactamase gene.

Data availability. The complete genome sequences of *F. marina* strains E95-16 and E103-15 were deposited in GenBank (BioProject no. PRJNA563512 and PRJNA563510, respectively) under the accession no. CP043552 and CP043553 (E95-16 chromosome and plasmid) and CP043550 and CP043551 (E103-15 chromosome and plasmid). Raw sequence reads were

deposited in the Sequence Read Archive (SRA) under accession no. SRR12615805 (E95-16, Illumina reads), SRR12615804 (E95-16, ONT reads), SRR12616186 (E103-15, Illumina reads), and SRR12616185 (E103-15, ONT reads).

REFERENCES

1. Colquhoun DJ, Duodu S. 2011. *Francisella* infections in farmed and wild aquatic organisms. Vet Res 42:47. https://doi.org/10.1186/1297-9716-42-47.
2. Soto E, Griffin MJ, Morales JA, Calvo EB, de Alexandre Sebastiao F, Porras AL, Viquez-Rodriguez X, Reichley SR, Rosser TG, Ware C, Byrne BA, Garcia JC, LaFrentz BR, Camus AC. 2018. *Francisella marina* sp. nov., etiologic agent of systemic disease in cultured spotted rose snapper (*Lutjanus guttatus*) in Central America. Appl Environ Microbiol 84:e00144-18. https://doi.org/10.1128/AEM.00144-18.
3. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 34:2666–2669. https://doi.org/10.1093/bioinformatics/bty149.
4. Koren S, Walenz BP, Berlin K, Miller JR, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.
5. Li H. 2018. minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
6. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
7. 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.
8. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 42:D206–D214. https://doi.org/10.1093/nar/gkt1226.
9. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S, Sharma AN, Doshi S, Courtot M, Lo R, Williams LE, Frye JG, Elsayegh T, Sardar D, Westman EL, Pawlowski AC, Johnson TA, Brinkman FS, Wright GD, McArthur AG. 2017. CARD 2017: expansion and model-centric curation of the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res 45:D566–D573. https://doi.org/10.1093/nar/gkw1004.
10. Abby SS, Neron B, Menager H, Touchon M, Rocha EP. 2014. MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems. PLoS One 9:e110726. https://doi.org/10.1371/journal.pone.0110726.
11. Abby SS, Cury J, Guglielmini J, Neron B, Touchon M, Rocha EP. 2016. Identification of protein secretion systems in bacterial genomes. Sci Rep 6:23080. https://doi.org/10.1038/srep23080.
12. Meier-Kolthoff JP, Auch AF, Klenk HP, Goker M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14:60. https://doi.org/10.1186/1471-2105-14-60.
13. Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H, Chun J. 2017. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol 67:1613–1617. https://doi.org/10.1099/ijsem.0.001755.
14. Petersen JM, Carlson J, Yockey B, Pillai S, Kuske C, Garbalena G, Pottumarthy S, Chalcraft L. 2009. Direct isolation of *Francisella* spp. from environmental samples. Lett Appl Microbiol 48:663–667. https://doi.org/10.1111/j.1472-765X.2009.02589.x.
15. Richter M, Rossello-Mora R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A 106:19126–19131. https://doi.org/10.1073/pnas.0906412106.
16. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 57:81–91. https://doi.org/10.1099/ijs.0.64483-0.