**Genome Sequence of the Fish Pathogen *Flavobacterium columnare* ATCC 49512**

Hasan C. Tekedar, Attila Karsi, Allison F. Gillaspy, David W. Dyer, Nicole R. Benton, Jeremy Zaitshik, Stefanie Vamenta, Michelle M. Banes, Nagihan Gülsoy, Mary Aboko-Cole, Geoffrey C. Waldbieser, and Mark L. Lawrence

College of Veterinary Medicine, Mississippi State University, Mississippi State, Mississippi, USA; Department of Microbiology and Immunology, The University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, USA; Department of Biology, Faculty of Art and Sciences, Marmara University, Göztepe, Istanbul, Turkey; School of Veterinary Medicine, Tuskegee University, Tuskegee, Alabama, USA; and United States Department of Agriculture, Agricultural Research Service, Stoneville, Mississippi, USA

**Genome Summary**

*Flavobacterium columnare* is a yellow-pigmented, motile, Gram-negative rod in the family *Flavobacteriaceae*, one of the main phylactic lines within the *Bacteroidetes* group (5). *F. columnare* shows wide genetic heterogeneity and colony morphology as well as significant differences in virulence within the species. *F. columnare* strains are divided into three genovars that demonstrate various levels of virulence for different fish species (1, 6, 12–14). *F. columnare* strain ATCC 49512 (CIP 103533 [TG 44/87]), isolated in 1987 from a skin lesion of a brown trout fry in France (3), belongs to genovar I (10) and is avirulent in channel catfish (13). The average reported genome size of bacterial species in the genus *Flavobacterium* and estimated average G+C content of the *F. columnare* genome have been reported as 4.1 ± 1 Mb and 32.5%, respectively (4, 8).

To obtain a complete sequence of the *F. columnare* genome, small and medium insert libraries (average insert sizes, 3 to 6 kb, 6 to 10 kb, and 10 to 12 kb) were constructed at Mississippi State University (9). Shotgun paired-end sequencing was conducted to 8-fold coverage at The Laboratory for Genomics and Bioinformatics at the University of Oklahoma Health Sciences Center. Sequences were assembled using Phred/Phrap. In addition to Sanger sequencing, a sequencing run was conducted using a GS FLX system from 454 Life Sciences, in which *F. columnare* genomic DNA was mixed with amplified gaps from the Sanger assembly and used as the template (269,256 reads with 269.4-bp average read length; Roche). Both Sanger and 454 sequencing generated an assembly with approximately 30-fold coverage. Quality-filtered 454 sequences were assembled using the Newbler assembler, and 454 contigs were added to the Sanger assembly. This hybrid assembly resulted in about 6 large contigs (>2 kb), and gaps were closed by primer walking using small or large insert plasmids or PCR amplicons as templates. The DNA sequence was submitted to the NCBI’s Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) for annotation, and then the annotated genome was submitted to GenBank.

The complete genome of *F. columnare* consists of a 3,162,432-bp chromosome with an average G+C content of 31.5%. The genome is 85.2% coding and has 2,731 predicted genes consisting of 2,642 protein coding, 15 rRNA, and 74 tRNA genes. For the protein coding genes, the average length is 1,021 bp and 1,625 genes (61.5%) have assigned functions. The *F. columnare* genome contains 5 rRNA operons, two of which are tandemly arranged. Analysis of the *F. columnare* genome sequence reveals many features common to other sequenced *Flavobacterium* species. We identified chondroitin AC lyase, proteases, and collagenases as well as genes involved in biofilm formation, secretion systems, and iron acquisition. Thirteen gliding motility genes were also identified.

**Nucleotide sequence accession number.** The genome sequence was deposited in GenBank under accession number CP003222, version CP003222.2, GI:372863588.

**ACKNOWLEDGMENTS**

We thank NCBI for providing the PGAAP Annotation Service.

This project was supported by the United States Department of Agriculture, National Institute of Food and Agriculture (2006-35600-16571). M.A.-C. was supported by the National Institutes of Health (NIH grant number 5T35RR007071).

**REFERENCES**


