Fully Assembled Genome Sequence for *Salmonella enterica* subsp. *enterica* Serovar Javiana CFSAN001992

Marc W. Allard,[a] Tim Muruvanda,[a] Errol Strain,[b] Ruth Timme,[a] Yan Luo,[b] Charles Wang,[a] Christine E. Keys,[a] Justin Payne,[a] Tony Cooper,[c] Khai Luong,[a] Yi Song,[a] Chen-Shan Chin,[c] Jonas Korlach,[c] Rich J. Roberts,[d] Peter Evans,[a] Steven M. Musser,[a] Eric W. Brown[a]

Office of Regulatory Science, Center for Food Safety & Applied Nutrition, U.S. Food & Drug Administration, College Park, Maryland, USA[a]; Office of Food Defense, Communications and Emergency Response, Center for Food Safety & Applied Nutrition, U.S. Food & Drug Administration, College Park, Maryland, USA[b]; Pacific Biosciences, Menlo Park, California, USA[c]; New England Biolabs, Inc., Ipswich, Massachusetts, USA[d]

We report a closed genome of *Salmonella enterica* subsp. *enterica* serovar Javiana (*S.* Javiana). This serotype is a common foodborne pathogen and is often associated with fresh-cut produce. Complete (finished) genome assemblies will support pilot studies testing the utility of next-generation sequencing (NGS) technologies in public health laboratories.

Address correspondence to Marc W. Allard, Marc.Allard@fda.hhs.gov.

*S*almonella enterica subsp. *enterica* serovar Javiana (*S.* Javiana) is one of the top five most common serotypes of *Salmonella* associated with fresh-cut produce, with an average of 11 clusters per year (2008 to 2012) detected by the PulseNet Network. This serotype has been involved in several produce-related outbreak investigations over the past decade, including for commodities such as cantaloupes, green onions, and tomatoes. It is also frequently recovered from the growing environments of these products. In October 2012, an outbreak of *S.* Javiana (PulseNet cluster 1210MLJGG01, pulsed-field gel electrophoresis [PFGE] pattern JGGX01.0500) was determined likely to be produce-related, but a specific commodity responsible for the illnesses was not determined. One clinical isolate (AZ_PI12305015) of this cluster was chosen for complete sequencing and assembly. We believe that the availability of whole-genome sequences and a large reference database will provide the discriminatory power needed to facilitate outbreak cluster detection and source tracking (1, 2). Currently, there is only one *S.* Javiana genome available in GenBank (accession no. ABEH00000000).

DNA was isolated from pure culture using Qiagen DNeasy blood & tissue kit (Qiagen Inc., Valencia, CA). Genome sequencing was performed using Pacific Biosciences RS sequencing technology (Pacific Biosciences, Menlo Park, CA), achieving >20× average genome coverage. The sample was prepared as a 10-kb insert library and was sequenced on 8 single-molecule real-time (SMRT) cells. *De novo* assemblies were created for the genome using the DevNet hierarchical genome assembly process (HGAP)/Quiver software package >4.5 kb 0.8QV (http://www.smrtcommunity.com/Share/Code?id=a1q70000000H2qRAAS), followed by Minimus2, and polished with Quiver to yield a single chromosomal contig and two mobile elements. The complete process from isolate to finished genomic sequence took less than 1 week. The mobile elements are single contigs but do not have overlapping sequences at their ends, suggesting one or more gaps. The genomes were annotated with the NCBI (National Center for Biotechnology Information) Prokaryotic Genomes Automatic Annotation Pipeline (http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html). The methylome was determined using the kinetic score distributions to detect the presence of ^{m6}A methyltransferases (MTases) (3, 4).

Comparative genomic analyses of these data will be included in future publications. This will include descriptions of the 4 active ^{m6}A methyltransferases observed in *S.* Javiana. The methylome of this strain was distinct from other strains analyzed thus far in Rebase (http://rebase.neb.com/cgi-bin/pacbiolist) (5).

This data release is a contribution toward the efforts of the 100K Pathogen Genome Project consortium. The FDA, along with Agilent Technologies, University of California at Davis, and many other federal and private partners, will sequence 100,000 pathogen genomes over the next 5 years (http://100kgenome.vetmed.ucdavis.edu) and will include genome closure of many of these isolates and their mobile elements. The product of this enormous effort will be a public molecular epidemiology reference database useful for designing pathogen detection assays, providing the evolutionary context for outbreaks, and many other applications yet to be realized (6, 7). The NGS (next-generation sequencing) federal and state public health laboratory aspect of the public database will be housed at the NCBI, in Bethesda, MD, under bioproject no. 183844 (http://www.ncbi.nlm.nih.gov/bioproject/183844) and known as the GenomeTrakr database.

Nucleotide sequence accession numbers. Genome sequences of *S.* Javiana and its mobile elements are available in DDBJ/EMBL/GenBank under bioproject no. 184141 and the GenBank accession no. CP004026, CP004027, and CP004028 at NCBI. Kinetic information is deposited in Gene Expression Omnibus (GEO) (GSE45178).

REFERENCES

1. **Lienau EK, Strain E, Wang C, Zheng J, Ottesen AR, Keys CE, Hammack TS, Musser SM, Brown EW, Allard MW, Cao G, Meng J, Stones R.** 2011. Identification of a salmonellosis outbreak by means of molecular sequencing. N Engl J. Med. **364**:981–982.
2. **Allard MW, Luo Y, Strain E, Pettengill J, Timme R, Wang C, Li C, Keys CE, Zheng J, Stones R, Wilson MR, Musser SM, Brown EW.** 2013. On the evolutionary history, population genetics and diversity among isolates of *Salmonella* Enteritidis PFGE pattern JEGX01.0004. PLoS One **8**:e55254.
3. **Murray IA, Clark TA, Morgan RD, Boitano M, Anton BP, Luong K, Fomenkov A, Turner SW, Korlach J, Roberts RJ.** 2012. The methylomes of six bacteria. Nucleic Acids Res. **40**:11450–11462.
4. **Fang G, Munera D, Friedman DI, Mandlik A, Chao MC, Banerjee O, Feng Z, Losic B, Mahajan MC, Jabado OJ, Deikus G, Clark TA, Luong K, Murray IA, Davis BM, Keren-Paz A, Chess A, Roberts RJ, Korlach J, Turner SW, Kumar V, Waldor MK, Schadt EE.** 2012. Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. Nat. Biotechnol. **30**:1232–1239.
5. **Roberts RJ, Vincze T, Posfai J, Macelis D.** 2010. Rebase—a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res. **38**:D234–D236.
6. **Underwood AP, Dallman T, Thomson NR, Williams M, Harker K, Perry N, Adak B, Willshaw G, Cheasty T, Green J, Dougan G, Parkhill J, Wain J.** 2013. Public health value of next-generation DNA sequencing of enterohemorrhagic *Escherichia coli* isolates from an outbreak. J. Clin. Microbiol. **51**: 232–237.
7. **Aarestrup FM, Brown EW, Detter C, Gerner-Smidt P, Gilmour MW, Harmsen D, Hendriksen RS, Hewson R, Heymann DL, Johansson K, Ijaz K, Keim PS, Koopmans M, Kroneman A, Lo Fo Wong D, Lund O, Palm D, Sawanpanyalert P, Sobel J, Schlundt J.** 2012. Integrating genome-based informatics to modernize global disease monitoring, information sharing, and response. Emerg. Infect. Dis. **18**:e1.