


Whole-Genome *Yersinia* sp. Assemblies from 10 Diverse Strains

H. E. Daligault,^a K. W. Davenport,^a T. D. Minogue,^b K. A. Bishop-Lilly,^{c,d} S. M. Broomall,^e D. C. Bruce,^a P. S. Chain,^a S. R. Coyne,^b K. G. Frey,^{c,d} H. S. Gibbons,^e J. Jaissle,^b G. I. Koroleva,^f J. T. Ladner,^f C.-C. Lo,^a C. Munk,^a G. F. Palacios,^f C. L. Redden,^{c,d} C. N. Rosenzweig,^f M. B. Scholz,^{a*}  S. L. Johnson^a

Los Alamos National Laboratory (LANL), Los Alamos, New Mexico, USA^a; Diagnostic Systems Division (DSD), United States Army Medical Research Institute of Infectious Diseases (USAMRIID), Fort Detrick, Maryland, USA^b; Naval Medical Research Center (NMRC)-Frederick, Fort Detrick, Maryland, USA^c; Henry M. Jackson Foundation, Bethesda, Maryland, USA^d; United States Army Edgewood Chemical Biological Center (ECBC), Aberdeen Proving Ground, Aberdeen, Maryland, USA^e; Center for Genome Sciences (CGS), USAMRIID, Fort Detrick, Maryland, USA^f

* Present address: M. B. Scholz, Michigan State University, East Lansing, Michigan, USA.

***Yersinia* spp. are animal pathogens, some of which cause human disease. We sequenced 10 *Yersinia* isolates (from six species: *Yersinia enterocolitica*, *Y. fredericksonii*, *Y. kristensenii*, *Y. pestis*, *Y. pseudotuberculosis*, and *Y. ruckeri*) to high-quality draft or complete status. The genomes range in size from 3.77 to 4.94 Mbp.**

Received 16 September 2014 Accepted 17 September 2014 Published 23 October 2014

Citation Daligault HE, Davenport KW, Minogue TD, Bishop-Lilly KA, Broomall SM, Bruce DC, Chain PS, Coyne SR, Frey KG, Gibbons HS, Jaissle J, Koroleva GI, Ladner JT, Lo C-C, Munk C, Palacios GF, Redden CL, Rosenzweig CN, Scholz MB, Johnson SL. 2014. Whole-genome *Yersinia* sp. assemblies from 10 diverse strains. *Genome Announc.* 2(5):e01055-14. doi:10.1128/genomeA.01055-14.

Copyright © 2014 Daligault et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to S. L. Johnson, shannonj@lanl.gov.

Yersinia is a genus of Gram-negative facultative anaerobes belonging to the *Enterobacteriaceae* family, best known for its three main pathogens, *Yersinia enterocolitica*, *Y. pestis*, and *Y. pseudotuberculosis*. The genus was first described in 1894 by Alexandre Yersin (1), who isolated *Y. pestis* during the third plague pandemic. Generally, *Yersinia* spp. cause animal infections and humans are only incidental hosts (2). *Y. enterocolitica* and *Y. pseudotuberculosis* are both enteric pathogens while *Y. pestis* generally results in lymphadenitis (bubonic plague) and is derived from *Y. pseudotuberculosis* (2, 3). In this study, we sequenced and assembled 10 *Yersinia* strains, including 7 isolates of these 3 pathogenic species and 3 additional congeners.

High-quality genomic DNA was extracted from purified isolates of each strain using a Qiagen Genomic-tip 500 at the USAMRIID Diagnostic Systems Division (DSD). Specifically, 100-mL bacterial cultures were grown to stationary phase and nucleic acid was extracted per the manufacturer's recommendations with one minor variation. For BSL3 *Yersinia pestis*, all cultures were lysed overnight to ensure sterility of the resulting extracted material. If sterility was not achieved, the nucleic acid was passed through a 0.45- μ M filter and rechecked for viable organisms before removal from the BSL3 suite.

Sequence data for each draft genome were generated using a combination of Illumina and 454 technologies (4, 5). For each genome, we constructed and sequenced an Illumina library of 100-bp reads at high coverage (ranging from 119 to 733 bp) and a separate 454 library of long-insert paired-end reads (insert sizes ranging from 7.10 to 10.3 kb with 8- to 57-fold genome

coverage). The two data sets were assembled together in Newbler (Roche) and the consensus sequences computationally shredded into 2-kbp overlapping fake reads (shreds). The raw reads were also assembled in Velvet and those consensus sequences computationally shredded into 1.5-kbp overlapping shreds (6). Draft data from all platforms were then assembled together with Allpaths, and the consensus sequences were computationally shredded into 10-kbp overlapping shreds (7). We then integrated the Newbler consensus shreds, Velvet consensus shreds, Allpaths consensus shreds, and a subset of the long-insert read pairs using parallel Phrap (High Performance Software, LLC). Possible misassemblies were corrected and some gap closure accomplished with manual editing in Consed (8–10).

Automatic annotation for each genome utilized an Ergatis-based workflow at LANL with minor manual curation. Each genome is available in NCBI (accession numbers are listed in Table 1) and raw data can be provided upon request. In-depth comparative analyses of these and other genomes are under way and will be published in subsequent reports.

Nucleotide sequence accession numbers. Genome accession numbers to public databases are listed in Table 1.

ACKNOWLEDGMENTS

Funding for this effort was provided by the Defense Threat Reduction Agency's Joint Science and Technology Office (DTRA J9-CB/JSTO). This article is approved by LANL for unlimited release (LA-UR-14-26046).

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 US Government.

TABLE 1 Strain-identifying information and basic statistics on assemblies and annotations

Species and strain	Alternate strain name	Accession no. (structure)	Size (bp) (% G+C content)	No. of CDS ^b	No. of rRNA genes	No. of tRNA genes	Plasmid ^a		
							pMT (pFra)	pPCP (pPst)	Pgm
<i>Yersinia enterocolitica</i> ATCC 9610	NCTC_12982	JPDV00000000 (1 scaffold, 7 contigs)	4,537,953 (47.3)	4,084	22	81	–	–	–
DATR	YE1013	JPDU00000000 (2 scaffolds, 3 contigs)	4,645,698 (47.3)	4,217	19	79	–	–	+
E265	YE1012	JPDW00000000 (3 scaffolds, 57 contigs)	4,694,189 (46.9)	4,268	18	78	–	–	–
YEA	NA ^c	JPDX00000000 (2 scaffolds; 82 contigs)	4,525,312 (47.0)	4,077	14	73	–	–	–
<i>Yersinia fredericksonii</i> ATCC 33641	CDC1461 to CDC1481	JPPS00000000 (2 scaffolds; 10 contigs)	4,941,072 (47.0)	4,363	22	80	–	–	–
<i>Yersinia kristensenii</i> ATCC 33639	CDC1459 to CDC1481	CP008955 (single closed chromosome)	4,442,328 (47.4)	3,946	22	82	–	–	+
<i>Yersinia pestis</i> CO92	YE0020CO92TA	JPMB00000000 (7 scaffolds; 59 contigs)	4,714,480 (47.6)	4,268	13	69	+	+	+
<i>Yersinia pseudotuberculosis</i> ATCC 4284	447	JPIY00000000 (4 scaffolds; 38 contigs)	4,768,560 (47.6)	4,190	15	78	–	–	+
ATCC 6904	NCTC 2476	CP008943 (single closed chromosome)	4,806,594 (47.6)	4,178	22	81	–	–	+
<i>Yersinia ruckeri</i> ATCC 29473	CDC 2396-61	JPPT00000000 (2 scaffolds; 15 contigs)	3,771,509 (47.4)	3,377	8	72	–	–	–

^a –, not present; +, present.

^b CDS, coding sequences.

^c NA, not applicable.

REFERENCES

- Perry RD, Fetherston JD. 1997. *Yersinia pestis*—etiologic agent of plague. Clin. Microbiol. Rev. 10:35–66.
- Prentice MB, Rahalison L. 2007. Plague. Lancet 369:1196–1207. [http://dx.doi.org/10.1016/S0140-6736\(07\)60566-2](http://dx.doi.org/10.1016/S0140-6736(07)60566-2).
- Achtman M, Zurth K, Morelli G, Torrea G, Guiyoule A, Carniel E. 1999. *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. Proc. Natl. Acad. Sci. USA 96:14043–14048. <http://dx.doi.org/10.1073/pnas.96.24.14043>.
- Bennett S. 2004. Solexa Ltd. Pharmacogenomics 5:433–438. <http://dx.doi.org/10.1517/14622416.5.4.433>.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bembem LA, Berka J, Braverman MS, Chen Y-J, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J-B, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380. <http://dx.doi.org/10.1038/nature03959>.
- Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. 18:821–829. <http://dx.doi.org/10.1101/gr.074492.107>.
- 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.7337908>.
- Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8:175–185. <http://dx.doi.org/10.1101/gr.8.3.175>.
- Ewing B, Green P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8:186–194.
- 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>.