

Thirty-Two Complete Genome Assemblies of Nine *Yersinia* Species, Including *Y. pestis*, *Y. pseudotuberculosis*, and *Y. enterocolitica*

Shannon L. Johnson,^a Hajnalka E. Daligault,^a Karen W. Davenport,^a James Jaissle,^b Kenneth G. Frey,^{c,d} Jason T. Ladner,^e Stacey M. Broomall,^f Kimberly A. Bishop-Lilly,^{c,d} David C. Bruce,^a Susan R. Coyne,^b Henry S. Gibbons,^f Chien-Chi Lo,^a A. Christine Munk,^a C. Nicole Rosenzweig,^f Galina I. Koroleva,^e Gustavo F. Palacios,^e Cassie L. Redden,^{c,d} Yan Xu,^a Timothy D. Minogue,^b Patrick S. Chain^a

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

The genus *Yersinia* includes three human pathogens, of which *Yersinia pestis* is responsible for >2,000 illnesses each year. To aid in the development of detection assays and aid further phylogenetic elucidation, we sequenced and assembled the complete genomes of 32 strains (across 9 *Yersinia* species).

Received 5 February 2015 Accepted 6 March 2015 Published 30 April 2015

Citation Johnson SL, Daligault HE, Davenport KW, Jaissle J, Frey KG, Ladner JT, Broomall SM, Bishop-Lilly KA, Bruce DC, Coyne SR, Gibbons HS, Lo C-C, Munk AC, Rosenzweig CN, Koroleva GI, Palacios GF, Redden CL, Xu Y, Minogue TD, Chain PS. 2015. Thirty-two complete genome assemblies of nine *Yersinia* species, including *Y. pestis*, *Y. pseudotuberculosis*, and *Y. enterocolitica*. *Genome Announc* 3(2):e00148-15. doi:10.1128/genomeA.00148-15.

Copyright © 2015 Johnson 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 Shannon L. Johnson, shannonj@lanl.gov.

The genus *Yersinia* contains 11 species, with three human pathogens, *Y. pestis*, *Y. pseudotuberculosis*, and *Y. enterocolitica*. Of these, *Y. pestis* is the most virulent, causing >2,000 global cases of plague annually, along with three global pandemics (1, 2). *Y. pestis* is a category A pathogen and potential biowarfare agent (3, 4), while *Y. pseudotuberculosis* and *Y. enterocolitica* cause food-borne self-limiting enteric diseases with low mortality rates (5). Recently, the list of strains for consideration in diagnostic assay development was released by the Association of Analytical Communities (AOAC) International, including strains that should be recognized (inclusivity) and ignored (exclusivity) by the assays (6). Here, we present the completed genome assemblies for 32 (see Table 1) of the 33 listed *Yersinia* strains (YPNN7 *Y. pseudotuberculosis* IB was not included due to technical issues).

Each genome was assembled using at least two data sets (specific data types and coverages are listed in the NCBI records), from Illumina (short- and/or long-insert paired data), Roche 454 (long-insert paired data), and/or PacBio long reads. The short- and long-insert paired data were assembled together in both Newbler and Velvet and computationally shredded into 1.5-kbp overlapping shreds. If the PacBio coverage was $\geq 100\times$, the data were assembled using the PacBio Hierarchical Genome Assembly Process (HGAP) (7). All data were additionally assembled in AllPaths (8). 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 (9–11). All genomes were assembled to finished-quality completion (12), and each assembly was annotated using an Ergatis-based (13) workflow, with minor manual curation.

The genome sizes averaged 4.68 ± 0.04 Mb (Table 1; the smallest is *Yersinia ruckeri* YRB, at 3.6 Mb, and the largest is *Y. pestis*

Antiqua, at 4.9 Mb), with up to 4 plasmids (average, 1.6 ± 0.2). Each genome contains 3,161 to 4,419 coding sequences (average, $4,155 \pm 39.9$) and a G+C content of 47 to 48%. As many of the virulence genes are located on plasmids, it is interesting to note that of the 16 *Y. pestis* strains, only 9 had all three “traditional” plasmids (pYV/pCD1 [virulence/calcium dependence], pPCP [plasminogen activator], and pMT [murine toxin]), with one strain (*Y. pestis* Nairobi) containing the pPCP plasmid only.

Nucleotide sequence accession numbers. The GenBank accession numbers for all 32 genomes 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) and the Department of Homeland Security Science and Technology Directorate award HSHQDC-08-X-00790.

This paper is approved by LANL for unlimited release (LA-UR-14-29606).

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 United States Government.

The bacterial strains were obtained from the Department of Defense’s Unified Culture Collection (<http://www.usamriid.army.mil/ucc/>).

REFERENCES

- Perry RD, Fetherston JD. 1997. *Yersinia pestis*—etiologic agent of plague. *Clin Microbiol Rev* 10:35–66.
- World Health Organization (WHO). 2014. Plague. Fact sheet no. 267. World Health Organization, Geneva, Switzerland. <http://www.who.int/mediacentre/factsheets/fs267/en/index.htm>.
- Inglesby TV, Dennis DT, Henderson DA, Bartlett JG, Ascher MS, Eitzen E, Fine AD, Friedlander AM, Hauer J, Koerner JF, Layton M, McDade J, Osterholm MT, O’Toole T, Parker G, Perl TM, Russell PK, Schoch-Spana M, Tonat K, Working Group on Civilian Biodefense. 2000. Plague as a

- biological weapon: medical and public health management. *JAMA* 283: 2281–2290. <http://dx.doi.org/10.1001/jama.283.17.2281>.
4. Riedel S. 2005. Plague: from natural disease to bioterrorism. *Proc (Bayl Univ Med Cent)* 18:116–124.
 5. Huang X-Z, Nikolich MP, Lindler LE. 2006. Current trends in plague research: from genomics to virulence. *Clin Med Res* 4:189–199. <http://dx.doi.org/10.3121/cm.4.3.189>.
 6. AOAC International. 2011. AOAC SMPR 2010.002: standard method performance requirements for polymerase chain reaction (PCR) methods for detection of *Yersinia pestis* in aerosol collection filters and/or liquids. *J AOAC Int* 94:1342–1346.
 7. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Non-hybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <http://dx.doi.org/10.1038/nmeth.2474>.
 8. 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>.
 9. Ewing B, Green P. 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Res* 8:186–194.
 10. Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces UsingPhred. I. Accuracy assessment. *Genome Res* 8:175–185. <http://dx.doi.org/10.1101/gr.8.3.175>.
 11. 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>.
 12. Chain PS, Grafham DV, Fulton RS, FitzGerald MG, Hostetler J, Muzny D, Ali J, Birren B, Bruce DC, Buhay C, Cole JR, Ding Y, Dugan S, Field D, Garrity GM, Gibbs R, Graves T, Han CS, Harrison SH, Highlander S, Hugenholtz P, Khouri HM, Kodira CD, Kolker E, Kyrpides NC, Lang D, Lapidus A, Malfatti SA, Markowitz V, Metha 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>.
 13. Hemmerich C, Buechlein 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>.

TABLE 1 List of strains included in the data set, their accession numbers, and plasmids

| Strain name | AOAC | Panel ^a | Accession no. | Size (Mb) | No. of indicated plasmid | | | |
|---|--------|--------------------|--------------------|-----------|--------------------------|---------|-----|-------|
| | | | | | pPCP | pCD/pYV | pMT | Other |
| <i>Y. aldovae</i> 670-83 | YPNN17 | E | CP009781 | 4.47 | | | | |
| <i>Y. enterocolitica</i> 2516-87 8081 WA | YPNN13 | E | CP009837, CP009838 | 4.60 | | 1 | | |
| | YPNN12 | E | CP009845, CP009846 | 4.68 | | 1 | | |
| | YPNN11 | E | CP009366, CP009367 | 4.61 | | 1 | | |
| <i>Y. frederiksenii</i> Y225 ^b | YPNN15 | E | CP009363, CP009364 | 4.55 | | | | 1 |
| <i>Y. intermedia</i> Y228 | YPNN16 | E | CP009801 | 4.85 | | | | |
| <i>Y. kristensenii</i> Y231 | YPNN14 | E | CP009997 | 4.49 | | | | |
| <i>Y. pestis</i> A1122 Angola Antiqua CO92 pgm- Dodson El Dorado Harbin35 Java9 ^c KIM5 Nairobi Nicholisk 41 PBM19 Pestoides B Pestoides F Pestoides G Shasta | YP12 | I | CP009839–CP009841 | 4.67 | 1 | | 1 | |
| | YP7 | I | CP009934–CP009937 | 4.67 | 1 | | 1 | |
| | YP3 | I | CP009903–CP009906 | 4.88 | 1 | 1 | 1 | |
| | YP1 | I | CP009971–CP009973 | 4.72 | | 1 | 1 | |
| | YP15 | I | CP009842–CP009844 | 4.77 | | 1 | 1 | |
| | YP16 | I | CP009782–CP009785 | 4.83 | 1 | 1 | 1 | |
| | YP9 | I | CP009701–CP009704 | 4.70 | 1 | 1 | 1 | |
| | YP11 | I | CP009992–CP009996 | 4.82 | 1 | 1 | 1 | 2 |
| | YP2 | I | CP009833–CP009836 | 4.78 | 1 | 1 | 1 | |
| | YP8 | I | CP010293, CP010294 | 4.47 | 1 | | | |
| | YP13 | I | CP009988–CP009991 | 4.70 | 1 | 1 | 1 | |
| | YP10 | I | CP009489–CP009492 | 4.86 | 1 | 1 | 1 | |
| | YP4 | I | CP010020–CP010023 | 4.79 | 1 | 1 | 1 | |
| | YP5 | I | CP009713–CP009715 | 4.72 | 1 | 1 | 1 | |
| | YP6 | I | CP010246–CP010248 | 4.73 | | 1 | 1 | |
| YP14 | I | CP009721–CP009724 | 4.83 | 1 | 1 | 1 | | |
| <i>Y. pseudotuberculosis</i> 1 EP2/+ IP32953 MD67 Pa3606 PB1/+ YPIII | YPNN10 | E | CP009786 | 4.72 | | | | |
| | YPNN8 | E | CP009758, CP009759 | 4.77 | | 1 | | |
| | YPNN4 | E | CP009710–CP009712 | 4.83 | | 1 | | 1 |
| | YPNN9 | E | CP009757 | 4.72 | | | | |
| | YPNN6 | E | CP010067–CP010069 | 4.83 | | 1 | | 1 |
| | YPNN3 | E | CP009779, CP009780 | 4.76 | | 1 | | |
| | YPNN5 | E | CP009792 | 4.68 | | | | |
| <i>Y. rohdei</i> ATCC 43380 | YPNN2 | E | CP009787 | 4.37 | | | | |
| <i>Y. ruckeri</i> YRB | YPNN1 | E | CP009539 | 3.60 | | | | |

^a Refers to the AOAC listing (6) of either inclusivity (I) or exclusivity (E) strains.

^b The plasmid in *Y. frederiksenii* is cryptic.

^c The two plasmids listed as “other” for *Y. pestis* JAVA9 are pJARS35 and pJARS36.