

# Genome Assemblies for 11 *Yersinia pestis* Strains Isolated in the Caucasus Region

Ekaterine Zhgenti,<sup>a</sup> Shannon L. Johnson,<sup>b</sup> Karen W. Davenport,<sup>b</sup> Gvantsa Chanturia,<sup>a</sup> Hajnalka E. Daligault,<sup>b</sup> Patrick S. Chain,<sup>b</sup> Mikeljon P. Nikolich<sup>c</sup>

National Center for Disease Control and Public Health (NCDC), Tbilisi, Georgia<sup>a</sup>; Los Alamos National Laboratory (LANL), Los Alamos, New Mexico, USA<sup>b</sup>; Walter Reed Army Institute of Research (WRAIR), Silver Spring, Maryland, USA<sup>c</sup>

***Yersinia pestis*, the causative agent of plague, is endemic to the Caucasus region but few reference strain genome sequences from that region are available. Here, we present the improved draft or finished assembled genomes from 11 strains isolated in the nation of Georgia and surrounding countries.**

Received 29 July 2015 Accepted 5 August 2015 Published 17 September 2015

**Citation** Zhgenti E, Johnson SL, Davenport KW, Chanturia G, Daligault HE, Chain PS, Nikolich MP. 2015. Genome assemblies for 11 *Yersinia pestis* strains isolated in the Caucasus region. *Genome Announc* 3(5):e01030-15. doi:10.1128/genomeA.01030-15.

**Copyright** © 2015 Zhgenti 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.

At the crossroads of Asia and Europe, along the Great Silk Road, Georgia and much of the Caucasus region have been exposed to many epidemics, including plague, which was first mentioned in the 11th century (1). Plague, a global zoonotic disease that has had devastating effects on human populations in three major pandemics, is caused by the Gram-negative coccobacillus *Yersinia pestis*. This disease has impacted the history of the Caucasus region, and unique variants that are known to circulate in the region have been characterized; however, few reference genomes are publicly available (2–4). The strains sequenced in this effort were selected from a historic collection of *Y. pestis* isolates from the Caucasus region preserved in the repository of the National Center for Disease Control and Public Health of Georgia.

High-quality genomic DNA was extracted from purified isolates of each strain and draft sequence data generated using a combination of Illumina and 454 technologies (5, 6). For each genome, we constructed and sequenced an Illumina library of 50 to 151 bp reads at high coverage (57- to 995-fold coverage), a 454

Titanium library of unpaired reads (17- to 42-fold coverage), and, excepting *Y. pestis* strains 1670 and 14735, a 454 Titanium long-insert paired-end library (insert size, 3.8 to 9.1 Kb with 2.1- to 21-fold genome coverage). PacBio long reads were generated for *Y. pestis* strain 790, with an average subread length of 3.5 Kb and 108-fold coverage (7). The 454 data sets were assembled together in Newbler (Roche), and the consensus sequences were computationally shredded into 2-Kbp overlapping fake reads (shreds). The raw Illumina reads were assembled in Velvet, and the Velvet consensus sequences were computationally shredded into 1.5-Kbp overlapping shreds (8). For *Y. pestis* strain 790, Illumina and 454 long-insert data were also assembled together with AllPaths, and the PacBio data were assembled with HGAP (9, 10); the consensus sequences for both assemblies were computationally shredded into 10-Kbp overlapping shreds. We then integrated the consensus shreds from all assemblers and a subset of the long-insert read pairs using parallel Phrap (11, 12). Possible misassemblies were corrected and

**TABLE 1** Listing of *Yersinia pestis* strains, their accession numbers, and genome sizes in this data set

Code	Strain	Country of origin	Accession no. <sup>a</sup>	Assembly status	Size (% G+C) <sup>b</sup>
BYA	1412	Georgia	CP006783 (Chr) CP006780 (pCD) CP006779 (pMT)	Finished	4,733,482 (47.6)
CAY	1413	Georgia	CP006762 (Chr) CP006761 (pCD) CP006760 (pMT)	Finished	4,736,923 (47.6)
DYA	1670	Georgia	AYLR00000000 (unscaffolded)	Improved high-quality draft	4.71 Mb (est) (47.6)
FKY	790	Kyrgyzstan	CP006806 (Chr) CP006807 (pMT) CP006808 (pPCP)	Finished	4,785,488 (47.6)
PYA	1522	Armenia	CP006758 (Chr) CP006757 (pCD) CP006756 (pMT)	Finished	4,738,644 (47.6)
RYP	2944	Russian Federation	CP006792 (Chr) CP006791 (pCD) CP006790 (pMT) CP006793 (pPCP)	Finished	4,672,740 (47.5)
VIP	3067	Georgia	CP006754 (Chr) CP006753 (pCD) CP006752 (pMT)	Finished	4,736,090 (47.5)
WYP	8787	Georgia	CP006748 (Chr) CP006747 (pCD) CP006746 (pMT)	Finished	4,675,987 (47.6)
YAE	3770	Georgia	CP006751 (Chr) CP006750 (pCD) CP006749 (pMT)	Finished	4,735,667 (47.6)
YPY	14735	Armenia	AYLS00000000 (unscaffolded)	Improved high-quality draft	4.69 Mb (est) (47.6)
ZPY	1045	Azerbaijan	CP006794 (Chr) CP006795 (pCD) CP006797 (pMT) CP006796 (pPCP)	Finished, noncontiguous (1 gap)	4.68 Mb (est) (47.5)

<sup>a</sup> Chr, chromosome; pCD, pMT, and pCP are plasmids.

<sup>b</sup> Values for the genome sizes are shown in bp, unless otherwise indicated. est, estimated.

some gap closures accomplished with PCR, manual editing in Consed, and in-house scripts (13).

Automatic annotation for each genome utilized an Ergatis-based workflow with minor manual curation. Each genome is available in the NCBI, 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.** The genome accession numbers to public databases are listed in Table 1.

## ACKNOWLEDGMENTS

Funding for this effort was provided by the U.S. Defense Threat Reduction Agency.

This paper is approved by LANL for unlimited release (LA-UR-15-25578).

The views expressed herein are those of the authors and do not reflect official policy or positions of the Walter Reed Army Institute of Research, the U.S. Army, the U.S. Department of Defense, or the United States Government.

## REFERENCES

- Bakanidze L, Imnadze P, Chubinidze S, Tsertsvadze N, Mgeladze G, Shalutashvili I, Tsanava S, Shavishvili M, Manvelyan J, Ninashvili N, Katsitadze G. 2010. Surveillance on plague in natural foci in Georgia. In O'Connell KP, Skowronski EW, Sulakvelidze A, Bakanidze L (ed), *Emerging and Endemic Pathogens*, p 21–28. Springer Science + Business Media B.V., Dordrecht, Netherlands. [http://dx.doi.org/10.1007/978-90-481-9637-1\\_3](http://dx.doi.org/10.1007/978-90-481-9637-1_3).
- Anisimov AP, Lindler LE, Pier GB. 2004. Intraspecific diversity of *Yersinia pestis*. *Clin Microbiol Rev* 17:434–464. <http://dx.doi.org/10.1128/CMR.17.2.434-464.2004>.
- Gage KL, Kosoy MY. 2005. Natural history of plague: perspectives from more than a century of research. *Annu Rev Entomol* 50:505–528. <http://dx.doi.org/10.1146/annurev.ento.50.071803.130337>.
- Revazishvili T, Rajanna C, Bakanidze L, Tsertsvadze N, Imnadze P, O'Connell K, Kreger A, Stine OC, Morris JG, Jr, Sulakvelidze A. 2008. Characterisation of *Yersinia pestis* isolates from natural foci of plague in the Republic of Georgia, and their relationship to *Y. pestis* isolates from other countries. *Clin Microbiol Infect* 14:429–436. <http://dx.doi.org/10.1111/j.1469-0691.2008.01953.x>.
- 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, Bemben 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>.
- Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, deWinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* 323:133–138. <http://dx.doi.org/10.1126/science.1162986>.
- 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>.
- 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. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <http://dx.doi.org/10.1038/nmeth.2474>.
- Ewing B, Green P. 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Res* 8:186–194. <http://dx.doi.org/10.1101/gr.8.3.186>.
- 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>.
- 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>.