



Five Draft Genome Sequences of Historical *Yersinia pestis* Strains of Phylogroups 2.MED4 and 2.MED1 of the Medieval Biovar

 Alina N. Balykova,^a  Lyubov M. Kukleva,^a  Ekaterina A. Naryshkina,^a  Galina A. Eroshenko,^a  Vladimir V. Kutyrev^a

^aRussian Research Anti-Plague Institute "Microbe", Federal Service for Surveillance in the Sphere of Consumer Rights Protection and Human Welfare, Saratov, Russian Federation

ABSTRACT We announce the genome sequences of five historical highly virulent *Yersinia pestis* strains of the phylogroups 2.MED4 and 2.MED1 of the medieval biovar. They were the etiological agents of plague outbreaks with high mortality rates in the Northern Caspian Sea region at the end of the 19th century and beginning of the 20th.

The plague pathogen, the highly virulent Gram-negative bacterium *Yersinia pestis*, is one of the most striking examples of a pathogen with high epidemic potential. The main subspecies, which is virulent for humans, includes the following biovars: antiqua, medieval, oriental, and possibly also intermedium (1). Our molecular investigations showed that strains of the phylogenetic branch 2.MED of the medieval biovar occupy about 93.3% of the territory of the natural foci in the Commonwealth of Independent States (1, 2). At the end of the 19th century and beginning of the 20th, numerous outbreaks of plague were registered in the Northern Caspian Sea region in Russia and Kazakhstan. Our studies prove that since 1912, outbreaks in the Northern Caspian Sea region have been caused by *Y. pestis* strains of the phylogenetic branches 2.MED1 and 2.MED4 of the medieval biovar (3, 4). The study of the genomes of the etiological agents of plague outbreaks in the Northern Caspian Sea region at the beginning of the 20th century is important for elucidating the causes of the high mortality rates among the population, as well as for historical reconstruction of the evolution of the medieval biovar of the plague pathogen.

The five strains of *Y. pestis* used in this study were isolated from natural foci in the Northern Caspian Sea region between 1925 and 1932 (Table 1). The strains were grown in LB medium (pH 7.2) for 24 to 48 h at 28°C; for DNA isolation, daily cultures of the strains were grown in LB liquid for 24 h at 28°C. DNA was extracted using the PureLink genomic DNA minikit (Invitrogen, USA). Genome sequencing was performed using the Ion Torrent PGM platform (Thermo Fisher Scientific, USA), according to the manufacturer's instructions. DNA libraries were prepared using the Ion PGM reagent 400 kit and Ion 318 chip kit (Thermo Fisher Scientific). For each genome, raw short-read sequences were filtered, quality controlled, and *de novo* assembled using the Ion Torrent Suite v5.12.3 software package (<https://github.com/iontorrent/TS>) and Newbler v2.6 (5). Default parameters were used for all software. Finally, we obtained from 203 to 260 contigs for each genome (Table 1). The average genome size was 4.5 Mb, with an average peak depth of the obtained genomes of 53× and GC contents ranging from 47.46% to 47.54%. Plasmid sequences were identified using the software package DNASTAR Lasergene v15.3 (6). All strains had three plasmids (pMT, pCD, and pPCP), except *Y. pestis* 106, which had no pPCP. The final assemblies were annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v5.3 (7). Each genome contains 3,711 to 4,026 coding sequences.

Editor David Rasko, University of Maryland School of Medicine

Copyright © 2022 Balykova et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Alina N. Balykova, alinabalnik@gmail.com.

The authors declare no conflict of interest.

Received 21 January 2022

Accepted 21 March 2022

Published 12 April 2022

TABLE 1 Accession numbers and basic statistics of the genome assemblies of *Y. pestis* strains

| Strain | Isolation source, yr | SRA accession no. | GenBank assembly accession no. | Data production metrics | | | | | | Total no. of genes | No. of coding genes | %GC | |
|--------|-------------------------------------|-------------------|--------------------------------|-------------------------|--------------------------|-------------------|----------------|----------------------|----------------|--------------------|---------------------|-------|---------------|
| | | | | No. of raw reads | Length of raw reads (bp) | Total length (bp) | No. of contigs | No. of contigs >1 kb | Peak depth (×) | | | | N_{50} (bp) |
| 36 | <i>Spermophilus pygmaeus</i> , 1925 | SRR17333920 | JAJTSR0000000000 | 1,169,251 | 283,896,647 | 4,518,558 | 260 | 213 | 58 | 36,008 | 4,170 | 3,747 | 47.54 |
| 70 | Fleas, 1926 | SRR17333916 | JAJTSV0000000000 | 840,127 | 223,784,416 | 4,516,182 | 244 | 196 | 50 | 37,521 | 4,168 | 3,711 | 47.46 |
| 106 | <i>Meriones meridianus</i> , 1928 | SRR17333918 | JAJTST0000000000 | 1,580,426 | 405,381,148 | 4,488,175 | 220 | 174 | 83 | 45,341 | 4,119 | 3,803 | 47.48 |
| 174 | Human, 1932 | SRR17333917 | JAJTSL0000000000 | 1,106,572 | 283,611,903 | 4,680,401 | 251 | 182 | 53 | 48,683 | 4,315 | 4,026 | 47.54 |
| 107 | Human, 1929 | SRR17333919 | JAJTSS0000000000 | 853,040 | 215,132,355 | 4,585,677 | 203 | 170 | 38 | 47,826 | 4,188 | 3,785 | 47.53 |

The obtained nucleotide sequences of the genomes of historical strains of *Y. pestis* are pivotal for understanding the patterns of microevolution and the spatiotemporal distribution of the medieval biovar in the Caspian Sea region.

Data availability. The whole-genome shotgun projects and raw sequencing reads have been deposited at NCBI GenBank under BioProject accession number [PRJNA792459](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA792459) and are available under the accession numbers listed in Table 1.

REFERENCES

1. Kutyrev VV, Eroshenko GA, Motin VL, Nosov NY, Krasnov JM, Kukleva LM, Nikiforov KA, Al'khova ZV, Oglodin EG, Guseva NP. 2018. Phylogeny and classification of *Yersinia pestis* through the lens of strains from the plague foci of Commonwealth of Independent States. *Front Microbiol* 9:1106–1111. <https://doi.org/10.3389/fmicb.2018.01106>.
2. Popov NV, Eroshenko GA, Karnaukhov IG, Kuznetsov AA, Matrosov AN, Ivanova AV, Oglodin EG, Nikiforov KA, Korzun VM, Verzhutsky DB, Chipanin EV, Ayazbaev TZ, Dzhaparova AK, Berdiev SK, Lopatin AA, Dubyansky VM, Shcherbakova SA, Balakhonov SV, Kulichenko AN, Kutyrev VV. 2021. Epidemiological situation on plague in 2020. Forecast of epizootic activity of natural plague foci in the Russian Federation and other CIS countries for 2021. *Problemy Osobo Opasnykh Infektsii* 1:52–62. (In Russian.) <https://doi.org/10.21055/0370-1069-2021-1-52-62>.
3. Eroshenko GA, Popov NV, Al'khova ZV, Balykova AN, Kukleva LM, Kutyrev VV. 2019. Phylogenetic analysis of *Yersinia pestis* strains of medieval biovar, isolated in Precaspian North-Western Steppe plague focus in the XX century. *Problemy Osobo Opasnykh Infektsii* 2:55–61. (In Russian.) <https://doi.org/10.21055/0370-1069-2019-2-55-61>.
4. Eroshenko GA, Popov NV, Al'khova ZV, Kukleva LM, Balykova AN, Chervyakova NS, Naryshkina EA, Kutyrev VV. 2021. Evolution and circulation of *Yersinia pestis* in the Northern Caspian and Northern Aral Sea regions in the 20th–21st centuries. *PLoS One* 16:e0244615. <https://doi.org/10.1371/journal.pone.0244615>.
5. Life Sciences Corp. 2010. Part COS De Novo Assembler, GS Reference Mapper, SFF Tools, 454. *In* 454 sequencing system software manual, v 2.5.3. Life Sciences Corp., Branford, CT.
6. Burland TG. 2000. DNASTAR's Lasergene sequence analysis software, p 71–91. *In* Misener S, Krawetz SA (ed), *Bioinformatics methods and protocols. Methods in molecular biology*, vol 132. Humana Press, Totowa, NJ. <https://doi.org/10.1385/1-59259-192-2:71>.
7. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>.