Whole-Genome Assemblies for Two \textit{Yersinia pestis} Strains Isolated in Mongolia

Muzi Jin,\textsuperscript{a} B. Byambajav,\textsuperscript{b} Hongyuan Zheng,\textsuperscript{c} Yufei Chen,\textsuperscript{a} B. Natsagdorj,\textsuperscript{b} Chao Yang,\textsuperscript{c} \textsuperscript{Y}ajun Song,\textsuperscript{c} Jing Wang\textsuperscript{d}

\textsuperscript{a}Huhhot Customs District, Huhhot, China
\textsuperscript{b}National Center for Zoonotic Diseases, Ulaanbaatar, Mongolia
\textsuperscript{c}State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing, China
\textsuperscript{d}Chinese Academy of Inspection and Quarantine, Beijing, China

Muzi Jin, B. Byambajav, and Hongyuan Zheng contributed equally to this work. Author order was determined on the basis of seniority.

\textbf{ABSTRACT} \ Here, we report the draft genome sequences of two \textit{Yersinia pestis} bv. Antiqua strains, belonging to the 3.ANT phylogroup, that were isolated in Mongolia and were circulating in marmot populations.

\textbf{P}lague is caused by \textit{Yersinia pestis}, a Gram-negative bacterium that belongs to \textit{Enterobacteriaceae}. It has claimed the lives of millions of people in the three major historical pandemics, and the pathogen has greatly influenced human history on a global scale (1). According to calculations based on a neutral molecular clock, \textit{Y. pestis} differentiated from its ancestor \textit{Yersinia pseudotuberculosis} around 6,000 years ago (2). Based on whole-genome-wide variations, \textit{Y. pestis} strains could be attributed to five major branches (3).

Mongolia is a landlocked nation bordering Russia to the north and China to the south. It is located in central Asia, with an area of about 1.56 million km$^2$; the main groups of \textit{Y. pestis} there are populations of \textit{Y. pestis} 3.ANT2 phylogroup strains belonging to \textit{Y. pestis} bv. Antiqua (3). In this study, we isolated two \textit{Y. pestis} strains from the carcasses of two marmots (\textit{Marmota sibirica}). Strain 3243 was isolated in 2016 from the brain of a dead marmot in Kusgul Aymag, Mongolia (43°07′32″N, 111°08′27″E), and strain 3256 was isolated in 2017 from the spleen of a dead marmot in Dzavhan Aymag, Mongolia (43°07′17″N, 111°08′41″E).

These two isolates were cultured on Hottinger’s agar (pH 6.9 to 7.1; prepared at the National Center for Zoonotic Diseases, Mongolia) at 28°C for 2 days and were confirmed with microscopic examinations, a PCR method, and biochemical identification tests. These two isolates were then cultured on Hottinger’s agar for 18 h for DNA extraction, and DNA samples were extracted using the Qiagen DNAeasy blood and tissue kit.

Sequencing libraries were prepared using the MGIEasy universal DNA library preparation set (BGI, Shenzhen, China), and whole-genome sequencing was performed using an Illumina NovaSeq 6000 system according to the manufacturer’s instructions. For the Illumina sequencing library, the insert size was 350 bp, with a paired-end sequencing length of 150 bp. Finally, we obtained 32,856,824 and 34,512,662 paired-end raw reads for strain 3243 and strain 3256, respectively. We then used Trimmomatic v0.38 (4) to remove low-quality sequencing reads (quality values of >20). After filtering the raw data, we obtained 400 Mb of clean reads for strains 3243 and 3256, with genome coverage values of 86× and 88×, respectively. The paired-end reads were \textit{de novo} assembled by SPAdes v3.12.0 (5), and coding sequences (CDSs) were annotated by Prokka v1.13 (6). All software settings used were the default parameters unless otherwise mentioned. Finally, we obtained 193 and 185 contigs for strains 3243 and 3256, respectively.
3256, respectively. The genome characteristics of the two strains are recorded in Table 1.

**Data availability.** The whole-genome shotgun projects for strains 3243 and 3256 have been deposited in DDBJ/ENA/GenBank under the accession numbers JAABOF000000000 and JAABOG000000000, respectively. The versions described in this paper are the first versions, JAABOF000000000.1 and JAABOG000000000.1, respectively. The publicly available genome assemblies have been annotated via NCBI PGAP v4.11 (7, 8). The Sequence Read Archive (SRA) data for strains 3243 and 3256 have been deposited in the NCBI SRA under the accession numbers SRR10912556 and SRR10912555, respectively.

**ACKNOWLEDGMENTS**

This work was supported by grant 2017JK008 (Chinese Academy of Inspection and Quarantine) and National Science and Technology Major Project grant 2018ZX10101-001 (Ministry of Science and Technology of China).

We sincerely thank Yujun Cui, Yarong Wu, and Yue Xiao (Beijing Institute of Microbiology and Epidemiology) for assistance with this project.

**REFERENCES**

1. Cui Y, Li Y, Gorgé O, Platonov ME, Yan Y, Guo Z, Pourcel C, Dentovskaya SV, Balakhonov SV, Wang X, Song Y, Anisimov AP, Vergnaud G, Yang R. 2008. Insight into microevolution of *Yersinia pestis* by clustered regularly interspaced short palindromic repeats. PLoS One 3:e2652. https://doi.org/10.1371/journal.pone.0002652.

2. Rasmussen S, Allentoft ME, Nielsen K, Orlando L, Sikora M, Sjögren K-G, Pedersen AG, Schubert M, Dam AV, Kapel CMO, Nielsen HB, Brunak S, Avetisyan P, Epimakhov A, Khalyapin MV, Gnuni A, Kristiansen K, Lasik I, Metspalu M, Moiseyev V, Gromov A, Pokutta D, Saag L, Varul L, Yevisekoposyan L, Sicheritz-Pontén T, Foley RA, Lahr MM, Nielsen R, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.

3. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

4. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

5. Haft DH, DiCuccio M, Badr etdin A, Chetvernin V, Navrotsky EP, Zaslavsky L, Prickett KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.

6. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prijibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.

7. Tatusova T, DiCuccio M, Badr etdin A, Chetvernin V, Navrotsky EP, Zaslavsky L, Lomsadze A, Prickett KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.

8. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.