





# Retrospective Analysis of the Relationship between Two Anthrax Outbreaks in Kazakhstan Based on Genomic Data

 Alexandr Shevtsov,<sup>a</sup> Gilles Vergnaud,<sup>b</sup>  Asylulan Amirgazin,<sup>a</sup> Larissa Lukhnova,<sup>c</sup> Uinkul Izbanova,<sup>c</sup> Vladislav Shevtsov,<sup>a</sup> Yerlan Ramankulov<sup>a,d</sup>

<sup>a</sup>National Center for Biotechnology, Nur-Sultan, Kazakhstan

<sup>b</sup>Université Paris-Saclay, Institute for Integrative Biology of the Cell (I2BC), Gif-sur-Yvette, France

<sup>c</sup>M. Aikimbayev's National Scientific Center for Especially Dangerous Infections, Almaty, Kazakhstan

<sup>d</sup>School of Science and Technology, Nazarbayev University, Nur-Sultan, Kazakhstan

**ABSTRACT** We present a retrospective analysis of strains from two anthrax outbreaks in western Kazakhstan in 2009. The outbreaks occurred during the same period and in the same area located close to main roads, favoring a single source of infection. However, multilocus variable-number tandem-repeat analysis (MLVA), canonical single-nucleotide polymorphism (CanSNP) analysis, and genome-wide analysis demonstrated that the outbreaks were not connected.

In June 2009, anthrax outbreaks occurred in two districts (Bayterek and Borili) of the West Kazakhstan region. The distance between the outbreaks is 118 km, and both locations are close to the main roads, suggesting the possibility of carryover contamination between the Bayterek and Borili districts (1). In both outbreaks, the sources of human infection were cows slaughtered without a veterinary examination. Eight strains were recovered from the Borili outbreak, from cattle or humans (one strain). Nine strains were recovered from the Bayterek outbreak, from the soil of the slaughter place or from clinical samples (five strains). Strains isolated from humans were anonymized, and their use was approved by the local ethics committee of the National Center for Biotechnology (protocol no. 1 from 1 January 2020). Strains were isolated using the inoculation method, plated on Hottinger agar, and cultivated at 37°C for 24 hours. All strains were identified as *Bacillus anthracis* based on the cultural and morphological characteristics of the colonies, the absence of hemolysis, catalase, lipase, and phosphatase, and protease activity, as well as their susceptibility to specific  $\gamma$ -phages. DNA was isolated from inactivated cultures using the QIAamp DNA minikit (Qiagen, USA) as previously described (2). Work with pathogens was performed in M. Aikimbayev's National Scientific Center for Especially Dangerous Infections (NSCEDI) biosafety level 3 (BSL3) facilities.

Multilocus variable-number tandem-repeat analysis (MLVA) genotyping was performed *in vitro* with capillary electrophoresis of PCR amplicons using 31 variable-number tandem-repeat (VNTR) loci as previously described (2–4). Two genotypes were observed, genotype 4-20-14-57-21-1-9-7-13-30-7-30-45-10-16-11-11-13-14-75-64-9-8-9-8-6-20-5-5-4-4, shared by all strains from Borili, and genotype 4-20-14-53-17-2-7-9-16-26-7-70-24-10-16-11-11-13-14-57-64-11-8-9-8-6-20-4-4-4-5, shared by all strains from Bayterek. According to the MLVA database at <https://microbesgenotyping.i2bc.paris-saclay.fr/databases/public> (3), the Borili genotype belongs to A.Br.008/009, whereas the Bayterek genotype is assigned to the Ames/Sterne lineage. Strains KZ-100 (Borili, cattle muscle tissue), KZ-107 (Bayterek, human carbuncle), and KZ-114 (Bayterek, soil, slaughter place) were selected for sequencing. The VNTRs located on the pXO1 and pXO2 plasmids could be amplified in all the strains. Sequencing libraries were prepared

**Citation** Shevtsov A, Vergnaud G, Amirgazin A, Lukhnova L, Izbanova U, Shevtsov V, Ramankulov Y. 2020. Retrospective analysis of the relationship between two anthrax outbreaks in Kazakhstan based on genomic data. *Microbiol Resour Announc* 9:e01126-20. <https://doi.org/10.1128/MRA.01126-20>.

**Editor** Julie C. Dunning Hotopp, University of Maryland School of Medicine

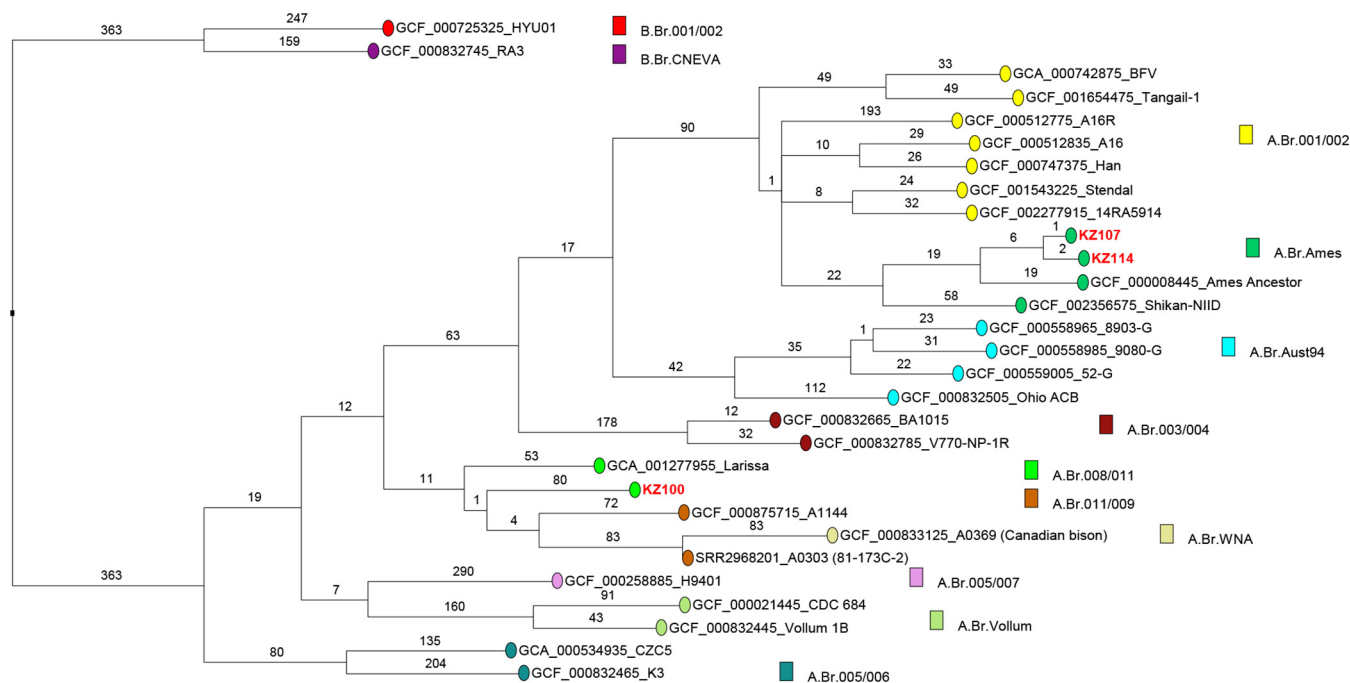
**Copyright** © 2020 Shevtsov et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Alexandr Shevtsov, [ncbshevtsov@gmail.com](mailto:ncbshevtsov@gmail.com).

**Received** 29 September 2020

**Accepted** 18 November 2020

**Published** 10 December 2020



**FIG 1** Whole-genome SNP analysis of the three strains from Kazakhstan, KZ-100, KZ-107, and KZ-114, and comparison with representative fully sequenced genomes. Twenty-six representative full genomes and three strains from Kazakhstan (bold, red type) were mapped on the Ames ancestor reference genome for SNP calling. The 3,784 SNPs identified were used to make a maximum parsimony tree. The tree size is 3,799 SNPs, corresponding to a 0.4% homoplasy level. The tree was rooted using the B branch as the outgroup. Branch lengths are shown, and assembly accession numbers and strain names are indicated. Strains are colored according to their CanSNP assignment.

using the Nextera XT DNA library prep kit (Illumina, San Diego, CA). Sequencing was performed with the MiSeq system using a MiSeq reagent kit v3 (600 cycles). The reads were trimmed using Seqtk v1.3 (5) up to value Q30 and *de novo* assembled using SKESA v2.3.0 (6). All reads were mapped back to the assembly using the Burrows-Wheeler Aligner (BWA) MEM algorithm v0.7.17 (7) to determine the average genome coverage using SAMtools v1.10 with the depth option  $-a$  (8). Single-nucleotide polymorphisms (SNPs) were identified using BioNumerics v7.6.3 (Applied-Maths, Belgium) by mapping reads on the Ames Ancestor reference (GenBank assembly accession no. [GCF\\_000008445](#)) and analyzed as previously described (9–11). The maximum parsimony tree was built using BioNumerics. All software was used with default parameters except when stated otherwise.

The assembly resulted in 42, 85, and 178 contigs, with  $N_{50}$  values of 378,397, 121,323, and 64,674 bp, total genome sizes of 5,449,226, 5,438,930, and 5,429,451 bp, average coverages of 55 $\times$ , 49.5 $\times$ , and 59.9 $\times$ , and GC percentages of 35.1%, 35.09%, and 35.12% for *B. anthracis* strains KZ-100, KZ-107, and KZ-114, respectively. The presence of the pXO1 and pXO2 plasmids was confirmed by aligning the assembled contigs to reference sequences (Ames Ancestor) using BLAST+ v2.6.0 (12). Genome annotation was performed using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v4.13 (13, 14). Totals of 5,364, 5,343, and 5,343 coding DNA sequences (CDSs) were predicted for *B. anthracis* strains KZ-100, KZ-107, and KZ-114, respectively.

Canonical SNP (CanSNP) analysis using CanSNPer v1.0.10 (15) confirmed KZ-100 as A.Br.008/011 and KZ-107/KZ-114 as A.Br.Ames. Strains *B. anthracis* KZ-107 and *B. anthracis* KZ-114 differ by three SNPs, which confirms their common source of origin (Fig. 1). To our knowledge, this is the first report of the Ames lineage in West Kazakhstan (16).

**Data availability.** This whole-genome shotgun project has been deposited in DDBJ/EMBL/GenBank under the accession no. [JACTNR000000000](#), [JACXSN000000000](#), and [JACXSM000000000](#). The versions described in this paper are the first versions,

JACTNR010000000, JACXSN010000000, and JACXSM010000000. The raw data from BioProject PRJNA639508 were submitted to the NCBI SRA under experiment accession no. SRR12560170, SRR12633803, and SRR12633802.

## ACKNOWLEDGMENT

This study was funded by the Ministry of Education and Science of Kazakhstan (grant no. AP05131460).

## REFERENCES

1. Toktassyn Y. 2019. The cadastre of settlements permanently unfavorable for anthrax in the Republic of Kazakhstan (1935–2018). Masgut Aikimbayev's National Scientific Center for Especially Dangerous Infections (NSCEDI), Almaty, Kazakhstan. (In Russian.)
2. Shevtsov AB, Lutsay VB, Kairzhanova AD, Lukhnova LY, Kulatay TZ, Izbanova UA, Karibaev TB, Shustov AV. 2019. Optimization of genotyping protocol for *B. anthracis* using multiple-locus VNTR analysis MLVA-31. *Eurasian J Appl Biotechnol* 2:103–113.
3. Thierry S, Tourterel C, Le Flèche P, Derzelle S, Dekhil N, Mendy C, Colaneri C, Vergnaud G, Madani N. 2014. Genotyping of French *Bacillus anthracis* strains based on 31-loci multi locus VNTR analysis: epidemiology, marker evaluation, and update of the Internet genotype database. *PLoS One* 9: e95131. <https://doi.org/10.1371/journal.pone.0095131>.
4. Beyer W, Bellan S, Eberle G, Ganz HH, Getz WM, Haumacher R, Hilss KA, Killian W, Lazak J, Turner WC, Turnbull PC. 2012. Distribution and molecular evolution of *Bacillus anthracis* genotypes in Namibia. *PLoS Negl Trop Dis* 6:e1534. <https://doi.org/10.1371/journal.pntd.0001534>.
5. Li H. 2012. seqtk toolkit for processing sequences in FASTA/Q formats. <https://github.com/lh3/seqtk>.
6. Souvorov A, Agarwala R, Lipman DJ. 2018. SKESA: strategic k-mer extension for scrupulous assemblies. *Genome Biol* 19:153. <https://doi.org/10.1186/s13059-018-1540-z>.
7. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
8. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
9. Timofeev V, Bahtejeva I, Mironova R, Titareva G, Lev I, Christiany D, Borzilov A, Bogun A, Vergnaud G. 2019. Insights from *Bacillus anthracis* strains isolated from permafrost in the tundra zone of Russia. *PLoS One* 14: e0209140. <https://doi.org/10.1371/journal.pone.0209140>.
10. Blouin Y, Hauck Y, Soler C, Fabre M, Vong R, Dehan C, Cazajous G, Massoure PL, Kraemer P, Jenkins A, Garnotel E, Pourcel C, Vergnaud G. 2012. Significance of the identification in the Horn of Africa of an exceptionally deep branching *Mycobacterium tuberculosis* clade. *PLoS One* 7: e52841. <https://doi.org/10.1371/journal.pone.0052841>.
11. Sahl JW, Pearson T, Okinaka R, Schupp JM, Gillette JD, Heaton H, Birdsell D, Hepp C, Fofanov V, Nosedo R, Fasanella A, Hoffmaster A, Wagner DM, Keim P. 2016. A *Bacillus anthracis* genome sequence from the Sverdlovsk 1979 autopsy specimens. *mBio* 7:e01501-16. <https://doi.org/10.1128/mBio.01501-16>.
12. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>.
13. 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>.
14. Zhao Y, Wu J, Yang J, Sun S, Xiao J, Yu J. 2012. PGAP: pan-genomes analysis pipeline. *Bioinformatics* 28:416–418. <https://doi.org/10.1093/bioinformatics/btr655>.
15. Larkeryd A, Myrtennas K, Karlsson E, Dwibedi CK, Forsman M, Larsson P, Johansson A, Sjodin A. 2014. CanSNPer: a hierarchical genotype classifier of clonal pathogens. *Bioinformatics* 30:1762–1764. <https://doi.org/10.1093/bioinformatics/btu113>.
16. Aikembayev AM, Lukhnova L, Temiraliyeva G, Meka-Mechenko T, Pazylov Y, Zakaryan S, Denissov G, Easterday WR, Van Ert MN, Keim P, Francesconi SC, Blackburn JK, Hugh-Jones M, Hadfield T. 2010. Historical distribution and molecular diversity of *Bacillus anthracis*, Kazakhstan. *Emerg Infect Dis* 16:789–796. <https://doi.org/10.3201/eid1605.091427>.