Allelic Polymorphism of Anthrax Pathogenicity Factor Genes as a Means of Estimating Microbiological Risks Associated with Climate Change

Y. O. Goncharova*, A. G. Bogun*, I. V. Bahtieva*, G. M. Titareva*, R. I. Mironova*, T. B. Kravchenko*, N. A. Ostarkov*, A. V. Brushkov*, V. S. Timofeev*, and S. G. Ignatov**, *

* State Research Center for Applied Microbiology and Biotechnology, Federal Service for Supervision of Consumer Protection and Welfare, Obolensk Settlement, Serukhov, Moscow oblast, 142279 Russia

** e-mail: ignatov@obolensk.org

Received February 11, 2022; revised February 18, 2022; accepted February 28, 2022

Abstract—Climate change brings new risks of emergence of especially dangerous diseases. The paper reports the possibility of assessing the pathogenic potential of bacteria as demonstrated by studying the allelic polymorphism of anthrax bacterium pathogenicity factor genes, which is a prerequisite for assessing the associated microbiological risks. The allelic polymorphism of the capBCADE operon (capB, capC, capA, capD, and capE genes) encoding the capsule biosynthesis proteins of Bacillus anthracis, and the acpA and acpB genes encoding the expression regulators of this operon have been studied for the first time. A number of single nucleotide polymorphisms (SNPs) were described in the strains of the studied sample, including 5 SNPs in the capB gene, 3 in capC, 4 in capA, 14 in capD, 2 in capE, and 15 in acpB, as well as 7 SNPs and one insertion in the acpA gene. As a result, the sample has been divided into sequence types for each gene and 17 genotypes, which are combinations of the identified sequence types. In silico translation of the detected alleles of the studied genes revealed three isoforms of the CapB and CapA proteins, two isoforms of the CapC and CapE proteins, six isoforms of the CapD protein, five isoforms of the AcpA protein, and four isoforms in the AcpB protein. It has been demonstrated that the SNP in the 351A → G position of capC is a marker of A.Br.Aust94 group strains. Based on the results, A.Br.Vollum group strains were divided into two subgroups. The strains in the evolutionary lines B and C differed from the line A strains by the presence of an 853G → A SNP in the acpA gene. In addition, a previously unknown variable number tandem repeat (VNTR), has been found in the acpA gene and the possibility of using it for differentiating and genotyping of B. anthracis strains has been demonstrated.

Keywords: Bacillus anthracis, allelic polymorphism, capBCADE operon, MLVA

INTRODUCTION

Climate change scenarios predict an increase in average annual temperatures worldwide. These climatic changes may have important consequences in terms of new microbiological risks. Specific features of the radiation—heat balance make the Arctic more susceptible to climate changes than other regions. The Arctic ecosystem being a reservoir of microbial genetic diversity represents a virtually unlimited source of microorganisms which may interact with humans. Increased temperatures, melting permafrost and sea ice, and the associated biosphere transformations, in particular the acceleration of microbiological processes, could lead to previously geographically restricted diseases suddenly emerging in economically active Arctic zones. There are risks that microorganisms may be released from permafrost, which people may not be prepared to meet with, including anthrax. Little information is currently available on the occurrence of virulent phenotypes in the Arctic zone. Studying microbial phenotypes allows assessing factors influencing the virulence of bacterial strains in the Arctic environment, as well as the potential risks associated with the changes taking place in polar areas in the face of climate change [1].

Anthrax pathogen, Bacillus anthracis, is a sporulating Gram positive bacterium that primarily infects ungulates but can also be transmitted to other warm-blooded animals including humans [2].
The distinctive feature of *B. anthracis* is the presence of two virulence plasmids, pXO1 and pXO2, which determine host specificity and bear the genetic determinants of the key pathogenicity factors [3]. The pXO1 plasmid contains the *pagA*, *lef*, and *cya* genes, which encode anthrax toxin components, namely, the protective antigen, lethal, and edema factors [4]. The pXO2 plasmid contains the *capBCADE* operon, which encodes the enzymes participating in the synthesis of the *B. anthracis* capsule made up of poly-γ-D-glutamic acid [5]. *Cap* operon genes go in the following order: *capB*, *capC*, *capA*, *capD*, and *capE* and encode the corresponding *B. anthracis* proteins, namely, CapB, CapC, CapA, CapD, and CapE [6]. These proteins are presumably responsible for capsule biosynthesis as well as for the transport and attachment of D-glutamic acid residues to the surface of bacterial cells [7, 8]. The *acpA* and *acpB* genes encoding the AcpA and AcpB proteins with partially overlapping functions are also located on the pXO2 plasmid. AcpA and AcpB independently positively control the transcription of the *cap* operon and many other genes [5, 9]. *B. anthracis* strains containing both plasmids in their genome are virulent for both humans and animals.

The ability to form spores underlies the key features of the *B. anthracis* life cycle. As spores, bacterial cells are capable of remaining viable for tens or hundreds of years [2]. Apart from soil foci formation at the sites of anthrax death, there are also risks of obtaining anthrax-infected animal material, such as wool or hides, and spreading it over long distances through human activity. All this has resulted in the general spread of anthrax across areas with highly developed livestock industry. In addition, there is a hypothesis that anthrax spores spread anthropogenically along the routes of livestock movement during the Mongolian military campaigns [10].

Increased risks of anthrax outbreaks in the Arctic zone of the Russian Federation associated with climate change are of particular importance. Long-term persistence of spores in soil and on household items makes anthrax outbreaks (especially given climatic changes) and in certain cases, major epizootics possible, as it was the case in the summer of 2016 on the Yamal Peninsula. This outbreak resulted in the death of more than 2000 reindeer and 1 person, as well as in the illness of 10 more people [10, 11]. Another interesting development was the discovery of mummies of cave lion cubs in Yakutia, with at least three strains of *B. anthracis* having been found at different depths in the soil beneath their discovery site [10].

Climate change makes it necessary to study the underlying causes of outbreaks and to comprehensively investigate this microorganism in order to prevent the spread of the disease. One aspect of this research is the investigation of the possible origins of the strain that caused the disease outbreak. Several molecular diagnostic methods (phylogenetic studies) are used to determine the origin of a microorganism strain.

In the case of anthrax pathogen, it is important to bear in mind that this microorganism emerged relatively recently (approximately 12000—26000 years ago) [12] and spends most of its life cycle as a spore. As a consequence, *B. anthracis* is characterized by low genetic polymorphism, which makes it impossible to divide it into subspecies and makes phylogenetic studies challenging. In addition, *B. anthracis* along with other members of the *Bacillus* genus is a member of the *Bacillus cereus* group, which includes nine genetically very closely related species [13]. This makes it important to differentiate the anthrax pathogen from other species and to identify the differences between its strains.

Several genotyping methods are now available to draw the genetic profile of a strain. One of the key methods is canSNP genotyping, which involves the identification of nucleotides in 14 positions on the *B. anthracis* chromosome, the so-called canonical SNPs (canSNPs). Using this method the global anthrax pathogen population was divided into three main evolutionary branches A, B, and C including 14 canSNP groups with some of them being associated with certain geographical locations [14, 15]. Multiple Locus VNTR Analysis (MLVA), which is based on identification of the number of tandem repeats in the known VNTR loci, most of which are located on the chromosome, may be named as the second most used method [16]. Several loci located on the plasmids have been suggested for use despite their short repeat length of two to three base pairs (bp), which makes it necessary to use expensive equipment to determine the number of repeats in these loci. Together, these methods make it possible to rapidly obtain information on the phylogenetic status of the strain in question with high precision [14].

Since *B. anthracis* is a pathogenic microorganism, it is only logical that its pathogenicity factors should be studied to the fullest extent. For example, the role of different amino acid residues in pathogenicity factor protein sequences and the polymorphism of genes encoding these factors in the global population of this species is one of the important research directions. This approach would not only contribute to the study of pathogenicity factor protein structures but would also allow the obtained data to be used to develop methods for bacterial strain genotyping using the polymorphisms in the studied genes. There are currently no studies that would describe the polymorphism of genes encoding pathogenicity factors located on the pXO2 plasmid in *B. anthracis* strains.

**Objective**—To describe polymorphism of the *capBCADE* operon genes encoding capsule biosynthesis proteins and the *acpA* and *acpB* genes encoding capsule biosynthesis regulation proteins using whole genome sequencing data of *B. anthracis* and to use the
results for genotyping, differentiation, and phylogenetic study of B. anthracis strains.

MATERIALS AND METHODS

Strains. The study included 77 B. anthracis and 2 B. cereus strains with different geographical origins (Table 1). Among them, 37 strains were isolated in Russia and bordering countries and deposited in the State Collection of Pathogenic Microorganisms and Cell Cultures of the State Research Center for Applied Microbiology and Biotechnology, Federal Service for Supervision of Consumer Protection and Welfare (SRC AMB) (Obolensk, Russia). The other 40 B. anthracis strains, a single B. cereus biovar anthracis strain and a single B. cereus strain, whose genome sequences are deposited in the GenBank database were included as an out group. All the B. anthracis strains under study were assigned to canSNP groups.

Primers. The following primers were used to amplify the VNTRacpA locus: VNTRacpAF = 5'-ATAGG-GAAAACAACATAATATAAA-3' and VNTRacpAR = 5'-TATTTGCTTGCAAAGATTCTC-3'.

Whole genome sequencing. B. anthracis strains were grown on a GRM-agar solid nutritive medium (SRC AMB, Russia) for 12 h at 37°C. Genomic DNA was purified using the Genomic DNA Purification Kit (Thermo Fisher Scientific, United States). Libraries for sequencing were prepared using the Nextera DNA Library Preparation Kit (Illumina, United States). Whole genome sequencing was performed using MiSeq (Illumina, United States) and Ion Torrent PGM (Thermo Fisher Scientific, United States) with the corresponding reagent kits Miseq Reagent Kit v3 (Illumina, United States) and Ion PGM Reagents 400 Kit and Ion 318 Chip Kit (Thermo Fisher Scientific, United States).

Bioinformatics analysis. Whole genome sequencing reads were assembled into the genome nucleotide sequences with the aid of the Lasergene software package (Dnastar, United States) (dnastar.com). The genome of the B. anthracis Ames Ancestor strain (GenBank: GCA_000008445.1) was used as a reference genome.

Searching for the nucleotide sequences from the B. anthracis and B. cereus strains, whose genomes are deposited in GenBank, was performed using BLAST (http://www.ncbi.nlm.nih.gov/blast).

Translation in silico of nucleotide sequences and multiple alignments were carried out with the aid of the MEGA 7.0 software package (http://www.megasoftware.net).

Phylogenetic (cluster) analysis was performed in MEGA 7.0. Phylogenetic trees (dendrograms) were constructed using the unweighted pair group method with arithmetic mean (UPGMA). Joined in silico nucleotide sequences of the studied genes were used for phylogenetic analysis.

MLVA genotyping. MLVA genotyping was performed according to the protocol described in [17] using primers designed in the present study.

PCR to amplify the VNTR locus was carried out using the Taq DNA Polymerase kit (Thermo Fisher Scientific, United States). PCR mixture contained 10× PCR buffer, 2.5 mM MgCl₂, 0.2 mM dNTP, 10 pmol of each primer, 1 U of Taq DNA polymerase, 1 μL of DNA solution in TE buffer for each of the studied strains (5–20 ng), and H₂O to get a final volume of 25 μL. A TE buffer was used as a negative amplification control. PCR was carried out in a T100 Thermal Cycler (Bio-Rad, United States).

The length of the amplified fragments was determined by gel electrophoresis. Results were obtained by counting the number of repeats at the VNTR locus. Electrophoresis was performed in 3% agarose gel (Helicon, Russia) in 0.5× TBE-buffer with subsequent EtBr staining (100 mg/L). Results were registered using the ECX-F15.1 transilluminator (Vilber Lourmat, France), Vzglyad gel documentation system (Helicon, Russia), and IC Measure software (Imaging Source Europe GmbH, Germany) at 254 nm.

Amplified fragment sizes were determined using the EZ Load 20 bp Molecular Ruler ladder (Bio-Rad, United States) with the aid of the PhotoCamptMw 99.04 software program (Vilber Lourmat, France).

RESULTS AND DISCUSSION

Allelic polymorphism. A total of 77 strains of anthrax pathogen were studied, including strains isolated at the site of a major reindeer epizootic on Yamal Peninsula (Russia) in the summer 2016 and the strains isolated from permafrost in Yakutia [10, 11]. B. cereus strains containing pXO2-like plasmids (B. cereus bv. anthracis CI and BC-AK) were also included in the sample. These strains are capable of causing anthrax-like disease in humans and great apes [18–20].

For the strains in the studied sample, the sequences of genes located on the pXO2 virulence plasmid, namely, capBCADE operon genes (capB, capC, capA, capD, and capE) responsible for B. anthracis capsule synthesis and acpA and acpB genes encoding capsule biosynthesis regulation proteins were assembled using the obtained whole-genome sequencing data.

The results allowed us to identify and describe mutations and to split the sample into sequence types (STs). In total, nine STs for the capD gene (D = 0.3546 [0.2178–0.4915]), five STs for the capA4 (D = 0.2691 [0.1436–0.3947]), acpA (D = 0.3636 [02347–0.4925]) and acpB (D = 0.1654 [0.0542–0.2767]) genes, four STs for the capC (D = 0.2293 [0.1073–0.3514]) gene, three STs for the capB (D = 0.0731 [–0.0066–0.1529]) gene and two STs for the capE (D = 0.0491 [–0.0173–0.1155]) gene were identified in the studied sample which included 79 strains. The STs identified, with the indication of their difference from the ST reference genome,
| Strains in the working collection | Strains from GenBank |
|----------------------------------|----------------------|
| **Strain** | **Geographical origin** | **canSNP group** | **Strain** | **Geographical origin** | **canSNP group** | **GenBank reference no.** |
| 1-271 | YaNAD | A.Br.001/002 | 14RA5914 | Germany | A.Br.001/002 | GCA_0022277915.1 |
| 34(738) | Kazakhstan | A.Br.001/002 | A16 | China | A.Br.001/002 | GCA_000512835.2 |
| 52/33 | Chechono-Ingush Republic | A.Br.001/002 | Stendal | Germany | A.Br.001/002 | GCA_001543225.1 |
| 1273 | Volgograd Oblast | A.Br.008/011 | Tangail-1 | Bangladesh | A.Br.001/002 | GCA_001654475.1 |
| 53169 | n/d | A.Br.008/011 | BFV | Jamaica | A.Br.001/002 | GCA_000742875.1 |
| 1(14) Stavropol | Ukraine | A.Br.008/011 | BA1015 | United States | A.Br.003/004 | GCA_000832665.1 |
| 1030/213 | Karachayevo-Cherkesiya | A.Br.008/011 | K3 | RSA | A.Br.005/006 | GCA_000832465.1 |
| 1056/51 | Stavropol Krai | A.Br.008/011 | CZC5 | Zambia | A.Br.005/006 | GCA_000534935.2 |
| 219/6 | Uzbekistan | A.Br.008/011 | A0135 | Tanzania | A.Br.005/006 | SRR2968157 |
| 367/17 | Tula Oblast | A.Br.008/011 | H9401 | Korea | A.Br.005/007 | GCA_000258885.1 |
| 46/27 | Chechono-Ingush Republic | A.Br.008/011 | A2075 | Tanzania | A.Br.005/006 | SRR2968187 |
| 47/28 | Chechono-Ingush Republic | A.Br.008/011 | A2079 | Tanzania | A.Br.005/006 | SRR2968188 |
| 48/29 | Chechono-Ingush Republic | A.Br.008/011 | Larissa | Greece | A.Br.008/011 | GCA_001277955.1 |
| 531/17 | Kalmykiya | A.Br.008/011 | Turkey32 | Turkey | A.Br.008/011 | GCA_000833275.1 |
| 546/714 | Voronezh Oblast | A.Br.008/011 | A1144 | Argentina | A.Br.011/009 | GCA_000875715.1 |
| 555/288 | Orenburg Oblast | A.Br.008/011 | Pollino | Italy | A.Br.011/009 | GCA_000831505.1 |
| 644/268 | Ukraine | A.Br.008/011 | London 499 | Great Britain | A.Br.011/009 | GCA_003227955.1 |
| 68/12 | Azerbaijan | A.Br.008/011 | Shikan-NIID | Japan | A.Br.Ames | GCA_002356575.1 |
| 7(992) | Novgorod Oblast | A.Br.008/011 | Ames Ancestor | United States | A.Br.Ames | GCA_000008445.1 |
| 8(2099) | Tatarstan | A.Br.008/011 | A0248 | United States | A.Br.Ames | GCA_000022865.1 |
| 914/213 | Chechono-Ingush Republic | A.Br.008/011 | A2012 | United States | A.Br.Ames | GCA_000006155.2 |
| Strain * | Geographical origin | canSNP group | Strain | Geographical origin | canSNP group | GenBank reference no.** |
|----------|---------------------|--------------|--------|---------------------|--------------|------------------------|
| LP50/3YA | Yakutia             | A.Br.008/011 | Ohio ACB | United States       | A.Br.Aust94  | GCA_000832505.1        |
| LP51/4YA | Yakutia             | A.Br.008/011 | Kanchipuram | India             | A.Br.Aust94  | GCA_014249775.1        |
| 1183     | Kabardino-Balkaria  | A.Br.008/011 | K1285   | Namibia             | A.Br.008/011 | SRR2071843             |
| 1298     | Volgograd           | A.Br.008/011 | A3716   | Namibia             | A.Br.Aust94  | SRR2968149             |
| 1173     | Stavropol Krai      | A.Br.Aust94  | SK-102  | United States       | A.Br.Vollum  | GCA_000832565.1        |
| 1259     | Stavropol Krai      | A.Br.Aust94  | Vollum 1B | United States   | A.Br.Vollum  | GCA_000832445.1        |
| 1199     | Daghestan           | A.Br.Aust94  | Vollum   | United States       | A.Br.Vollum  | GCA_000742895.1        |
| 331/214  | Azerbaijan          | A.Br.Aust94  | CDC 684  | United States       | A.Br.Vollum  | GCA_000021445.1        |
| 822/7    | Checheno-Ingush Republic | A.Br.Aust94 | Canadian bison | Canada     | A.Br.WNA      | GCA_000833325.1        |
| 11(1940) | Turkmenistan        | A.Br.Vollum  | BA1035   | RSA                 | B.Br.001/002 | GCA_000832725.1        |
| 15(1345) | Tajikistan          | A.Br.Vollum  | HYU01    | Korea               | B.Br.001/002 | GCA_000725325.1        |
| 157(B-1107) | Estonia              | B.Br.001/002 | SVA11    | Sweden              | B.Br.001/002 | GCA_000583105.1        |
| 1-364    | Siberia             | B.Br.001/002 | Tyrol 4675 | Austria         | B.Br.CNEVA   | GCA_001936375.1        |
| LP53/5YA | Yakutia             | B.Br.001/002 | RA3      | France              | B.Br.CNEVA   | GCA_000832745.1        |
| Yamal-2  | Yamal               | B.Br.001/002 | BF1      | Germany             | B.Br.CNEVA   | GCA_000295695.2        |
| 44       | n/d***              | B.Br.CNEVA   | I7OD930  | Switzerland         | B.Br.CNEVA   | GCA_006088855.1        |
|          |                     |              | Pasteur  | n/d                 | A.Br.011/009 | GCA_000832585.1        |
|          |                     |              | 2002013094 | United States    | C.Br.001     | GCA_000832965.1        |
|          |                     |              | 2000003102 | United States     | C.Br.001     | GCA_000742655.1        |
|          | B. cereus BC-AK     | China        | –        |                     |             | GCA_002117465.1        |
|          | B. cereus bv. anthracis CI | Ivory Coast | –        |                     |             | GCA_000143605.1        |

* Listed are *B. anthracis* strains, if not indicated otherwise.
** Reference number for genome sequence or raw whole genome sequencing data archive.
*** n/d—no data.
which consists in the presence of certain mutations, are listed in Tables S1–S6. For each gene, the sequence type to which the Ames Ancestor reference strain belongs is designated ST1. Type numbering was performed in decreasing order of the number of B. anthracis strains included within it, followed by the ST to which the B. cereus strains belonged. All STs differed between themselves by nucleotide substitutions or the insertion, as is the case with the acp4 gene.

To assess the phenotypic expression of the identified nucleotide polymorphism, or in other words, whether the nucleotide substitution at each identified position is synonymous or results in an amino acid substitution in the corresponding protein, or protein inactivation due to a stop codon, in silico translation of the nucleotide sequences into amino acid sequences was carried out (Tables S7–S12). The coordinates of the amino acid substitutions are given for complete translated protein sequences, disregarding their post-translational modifications, which have been described for the CapD protein only and consists in the cleavage of the 28 amino acid N-terminal signal sequence [21]. 3D structures of the proteins encoded by the genes studied here have not been described with the exception of CapD [21, 22], therefore amino acid substitutions are shown without indication in which protein domain they are located.

The distribution of the strains in the studied sample over the translated CapB, CapC, CapA, CapD, AcpA, and AcpB protein types is given in Tables S13–S18. The assignment of strains to protein types was carried out in the same way as it was done in the case of nucleotide STs. In total, six CapD protein isoforms, five AcpA protein isoforms, four AcpB protein isoforms, three isoforms of each of the CapB and CapA proteins, and two isoforms of each of the CapC and CapE proteins were described.

The tables do not contain data for the capE gene since the studied sample appeared to be monomorphic with the only exception for the B. cereus BC-AK strain, which revealed the presence of two SNPs, namely, 49T → G (17L → V) and 138C → T (synonymous). The observed low variability may be partially accounted for by the small length of the capE gene (only 144 bp), as given the same frequency of mutation across the entire cap operon, the probability of mutation in such a small part of it is rather low.

The distribution of strains in the studied sample across genotypes (GTs), each of them including a particular combination of ST genes under study, is shown in Table 2. The genotypes were numbered according to the principle specified above. GT, which included the reference strain was designated GT1. The remaining GTs were numbered in decreasing order of the number of strains in which they were described, the last numbers having been assigned to the GTs characteristic of B. cereus. A total of 17 GTs (D = 0.8908 [0.8604–0.9212]) were described.

A phylogenetic tree showing the relationships between the identified GTs is shown in Fig. 1.

In summary, the present study described the allelic polymorphism of B. anthracis capsule formation genes and cap operon expression regulation genes located on the pXO2 plasmid. Their STs were identified and strains in the study sample were clustered based on their GT, which represented a set of STs for individual genes. In other words, we virtually used multi-virulence-locus sequence typing (MVLST) to study the anthrax pathogen. Earlier, this method was first used by the authors to study this microorganism using the sequences of toxin formation genes located on the pXO1 plasmid. This pipeline was called MVLSTpXO1 genotyping. Accordingly, the genotyping pipeline suggested in this study may be called MVLSTpXO2, while the identified GTs, may be referred to as MVLSTpXO2-GT.

Analyzing the studied sample clustering based on MVLSTpXO2-GTs demonstrated that MVLSTpXO2 profiles correlated well both with the species identity (they were different for B. anthracis and B. cereus) and evolutionary line divergence within the B. anthracis species. However, using MVLSTpXO2 we were not able to accurately divide the sample into the groups corresponding to the canSNP groups that the strains belong to. At the same time [24] reported a strong correlation between MVLSTpXO1-GT of the strain and the canSNP group, which it belongs to, as well as, in some cases, their association with a geographical zone. It should be noted that MVLSTpXO2 genotyping based on gene sequences localized on the pXO2 plasmid had lower resolution (D = 0.6148 [0.488–0.7416]) compared to the similar approach, which used genes located on the pXO1 plasmid (D = 0.8951 [0.8656–0.9245]) [24].

In the A lineage, MVLSTpXO2-GT1 is the most common genotype, which includes strains belonging to almost all canSNP groups. Other GTs in lineage A differed from GT1 by one or two SNPs and contained a single strain or combined small subgroups of strains within the same canSNP group.

The exception was the Korean H9401 strain (canSNP group A.Br.005/007), which differed from MVLSTpXO2GT1 by two SNPs, namely, capD 796G → A (266V → I) and acpB 495A → G (synonymous). However, it is rather doubtful that these two markers are specific to the A.Br.005/007 group, as they are just as likely to be specific to the H9401 strain [25]. The presence of these two mutations in other strains in the A.Br.005/007 group could not be verified because at the time of preparing the manuscript, no whole genome sequencing data were publicly available for the strains in this canSNP group.

Apart from the H9401 strain, two more strains showed the presence of unique markers, namely, the Tangail-1 strain (A.Br.001/002), which contained the capB 230T → C (77V → A) SNP and the Pollino strain
(A.Br.011/009), which revealed the presence of the synonymous SNP capC 147T → C.

As stated above, in several cases, not only individual strains, but also small groups of strains within canSNP groups in lineage A showed unique markers. For example, MLVSTpXO2-GT3 genotype included four strains belonging to the A.Br.Aust94 group, which were isolated in the Caucasus region, namely, in the Stavropol krai, Chechen Republic, and Azerbaijan. These strains form an individual GT based on the presence of the unique synonymous SNP capC 351A → G (capC-ST2). However, this marker is absent in strain 1199 in the same A.Br.Aust94 group isolated in the same region, namely, in Dagestan.

To verify whether this capC 351A → G SNP is specific for the “Caucasian” A.Br.Aust94 strains, its presence was additionally tested in the genomes of 23 strains in this group isolated in different areas located outside the Caucasus region (Table 3). We were not able to detect this SNP in any of the strains included in this additional sample. Thus, it may be suggested that the A.Br.Aust94 capC 351A → G subgroup was formed in the Caucasus region and circulate there together with its “parental” form, and the capC 351A → G SNP may be used as a marker for the presence of which in the strain indicates its region of origin.

The A.Br.Vollum canSNP group was distributed between three GTs in the studied sample. Two strains isolated in Central Asia belong to GT1, a common one for lineage A. Four American strains belonging to GT5 (n = 3) differed from GT1 by the presence of the unique acpB 1381 A → G (461I → V) SNP. GT5 is characterized by the presence of only this substitution. GT12 (n = 1) contains another unique SNP acpB 563C → T (188S → L) in addition to this substitution. In a further search for the presence of these SNPs, we identified nine strains in the additional A.Br.Vollum sample with different geographical origins belonging to GT1 (n = 4) and GT5 (n = 8) (Table 4). Thus, we may at least say that the A.Br.Vollum group has split into two subgroups with different GTs: GT1 (which includes both strains in this group which were isolated in the former Soviet Union) and GT5. GT12 is apparently a strain-specific genotype.

**Table 2.** Distribution of the studied strains across the genotypes based on the combination of the sequence types for capB, capC, capA, capD, capE, acpA, and acpB genes

| MLVSTpXO2-GT | Strain | No. of strains |
|--------------|--------|---------------|
| GT1          | Ames Ancestor, CZC5, London_499, 14RA5914, A16, Stendal, Larissa, Shikan-NIID, A1144, Canadian_bison, BA1015, Ohio ACB, K3, Turkey32, BFV, Kanchipuram, A0248, A2012, Pasteur, K1285, A3716, A0135, A2075, A2079, LP50/3ya, 68/12, LP51/4YA, 367/17, 531/17, 7(992), 1199, 53169, I-271, 34(738), 1030/213, 1183, 219/6, 8(2099), 11(1940), 47/28, 48/29, 914/213, 1298 | 49 |
| GT2          | BF1, 17OD930, RA3, Tyrol 4675 | 4 |
| GT3          | 1259, 331/214, 822/7, 1173 | 4 |
| GT4          | LP53/5YA, I-364, Yamal_2 | 3 |
| GT5          | Vollum 1B, Vollum, CDC 684 | 3 |
| GT6          | 1(14)Stavropol, 555/288, 644/268 | 3 |
| GT7          | BA1035, SVA11 | 2 |
| GT8          | 2000031021, 2002013094 | 2 |
| GT9          | 44 | 1 |
| GT10         | Tangail-1 | 1 |
| GT11         | H9401 | 1 |
| GT12         | SK-102 | 1 |
| GT13         | Pollino | 1 |
| GT14         | HYU01 | 1 |
| GT15         | 157(B-1107) | 1 |
| GT16         | B. cereus bv. anthracis CI | 1 |
| GT17         | B. cereus BC-AK | 1 |
Markers characteristic of the B and C lines are worth mentioning separately. It has been demonstrated that all the studied strains in lineages B and C differ from lineage A strains by the presence of the common SNP \(a_{cpA}^{853G \rightarrow A}\).

The strains in the B.Br.CNEVA canSNP group isolated in Central Europe have a synonymous \(capD^{234T \rightarrow C}\) SNP. However, the only strain in this group originating from the SRCAMB (Obolensk, Russia) did not reveal the presence of this mutation. Unfortunately, there is no data left on where this strain was isolated, so the presence or absence of this \(capD^{234T \rightarrow C}\) SNP cannot be considered as a conclusive marker of geographical origin. At the same time, the presence of this strain in SRCAMB (Obolensk, Russia) still very likely indicates that this strain was isolated in the former Soviet Union. In this case there may be some reason to assume that this SNP does have phylogeographic relevance. However, strain 44 is the only available strain in the B.Br.CNEVA group for which the isolation site is not in Central Europe, which made it impossible to carry out any additional study on the distribution of the \(capD^{234T \rightarrow C}\) SNP among B.Br.CNEVA group strains isolated outside this region.

Evolutionary lineage C (canSNP group C.Br.001) is represented in the studied sample by two strains (2002013094 and 2000031021) isolated in the United States. These strains have the same nucleotide sequences for all the genes studied and therefore fall into the same GT according to the results of phylogenetic analysis of joined sequences. According to the results of the analysis carried out in this study, these strains proved to be phylogenetically closer to \(B.\ cerus\) than the other strains in the studied sample. The following SNPs are common between these strains and nonanthracis (\(B.\ cerus\)) strains: \(capC^{239C \rightarrow T}\) (80T \(\rightarrow\) M), \(capD^{208C \rightarrow T}\) (70H \(\rightarrow\) Y), \(capD^{667A \rightarrow G}\) (223K \(\rightarrow\) E), and \(capD^{1135T \rightarrow A}\) (379F \(\rightarrow\) I).

One of the aims of this study was to examine the anthrax pathogen as an infection, whose outbreaks may potentially be triggered by climatic changes, such as permafrost melting leading to the growth of spores previously trapped in frozen soil layers [10]. Previously, several researchers have suggested the possibil-
ity of outbreaks of several diseases, including anthrax, caused by climatic changes [26–28].

Strains isolated during the Yamal outbreak, as well as one of the strains isolated from the frozen soil layers in Yakutia, were previously assigned to evolutionary lineage B. Several mutations, which are phylogenetically significant markers, have been described in the genomes of these strains here and in the earlier studies [23, 24]. The discovery of anthrax bacilli in the Holocene alluvial deposits in Yakutia is of much interest [10]. The most likely genetic dating, though, indicated that these strains were conserved between

Table 3. The additional A.Br.Aust94 group sample including strains of different geographical origin tested for the presence of the \(\text{capC} \, 351\text{A} \rightarrow \text{G} \) SNP in their genome

| GenBank reference no. | Strain name | Isolation site |
|-----------------------|-------------|----------------|
| GCA_001677295         | K2          | Namibia        |
| GCA_001677305         | K1          | Namibia        |
| GCA_003367985         | 2000031103 (Strain 32) | Great Britain |
| GCA_003368005         | 2007740863 (SK57) | Great Britain |
| DRR000184             | BA_104      | Japan          |
| DRR000186             | BA_104      | Japan          |
| DRR128184             | BA105       | Japan          |
| ERR930300             | 1409        | Denmark        |
| SRR2071849            | K2883       | Denmark        |
| SRR2071866            | K4834       | Australia      |
| SRR2339898            | 2002721539  | RSA            |
| SRR2340304            | 2002734039  | Great Britain  |
| SRR2968133            | A0088       | RSA            |
| SRR2968144            | A0002       | Turkey         |
| SRR2968145            | A0656       | PRC            |
| SRR2968146            | A0659       | PRC            |
| SRR2968155            | A0083       | Germany        |
| SRR2968160            | A0252       | Zimbabwe       |
| SRR2968165            | A0455       | Mozambique     |
| SRR5811187            | 2000031009  | Thailand       |
| SRR1739963            | 2000031027  | United States  |
| SRR2339639            | 2000032893  | United States  |
| SRR2339643            | 2002013170  | United States  |

Table 4. The occurrence of the \(\text{acpB} \, 563\text{C} \rightarrow \text{T} \) and \(\text{acpB} \, 1381\text{A} \rightarrow \text{G} \) SNPs in the additional A.Br.Vollum group sample strains with different geographical origin

| GenBank reference no. | Strain name | Isolation site | SNP \(\text{acpB} \, 563\text{C} \rightarrow \text{T} \) | SNP \(\text{acpB} \, 1381\text{A} \rightarrow \text{G} \) |
|-----------------------|-------------|----------------|-----------------------------------------------|-----------------------------------------------|
| GCA_003045745         | COVASU      | India          | –                                            | +                                            |
| SRR1739961            | 2000031008  | United States  | –                                            | +                                            |
| SRR2339356            | 2002734255  | United States  | –                                            | +                                            |
| SRR2339406            | 2002734264  | United States  | –                                            | +                                            |
| SRR2339549            | 2002734276  | United States  | –                                            | –                                            |
| SRR2339634            | 2002013132  | Pakistan       | –                                            | –                                            |
| SRR2968161            | A0363       | Norway         | –                                            | +                                            |
| SRR2968162            | A0380       | Ireland        | –                                            | +                                            |
| SRR2968173            | A0615       | RSA            | –                                            | –                                            |
| SRR5811120            | 2002734049  | Great Britain  | –                                            | –                                            |
the 13th and 16th centuries. The geological position of the find below the seasonal thawing layer may probably indicate that they are several thousand years old. If the latter is true, then it appears that anthrax bacteria (apparently, as spores) are able to prevent the accumulation of mutations in accordance with the molecular clock hypothesis during their long-term preservation in permafrost leading to the effect of “curiously modern DNA for an ancient bacterium” [29]. The study of such strains is valuable not only in terms of the risks of outbreaks resulting from climate change, but also for better understanding of the evolutionary history of anthrax pathogen and other bacilli.

**MVLST pXO2** genotyping of 40 *B. anthracis* and 2 *B. cereus* strains, whose genomes are deposited in GenBank (Table 1) revealed the presence of a nine bp insertion in the *acpA* gene in several strains belonging to lineages B and C, which at closer look proved to be part of an imperfect tandem repeat ATA**GATA**. It has been demonstrated that the number of tandem repeating units is three in evolutionary lineage A strains, while it is four in lineage B strains belonging to the B.Br.001/002 (*n* = 3) and B.Br.CNEVA (*n* = 4) groups and in lineage C. In such a way, we identified a previously unknown VNTR locus on the pXO2 plasmid of the anthrax pathogen genome, which we named VNTRacpA.

The structure of tandem repeats itself determines the specific nature of the most likely mutations at these loci, which are insertions and deletions of the certain repetitive sequence. Furthermore, the frequency of these events is significantly higher than the average frequency of other mutations in the genome [30, 31]. Therefore, the presence of only two alleles of this tandem repeat in the *acpA* gene in the sample of 42 phylogenetically different strains suggests its low variability. It may be assumed that four repeats of this motif were characteristic of the *B. anthracis* ‘ancestral genome’ prior to its segregation into geographically and genetically distinct groups. It is possible that reducing the repeat fold to three resulted in an AcpA protein variant, which provided the corresponding strain with some selective advantage, for example, more efficient regulation of pathogenicity factor expression, with the archaic four-repeat variant being preserved only in the B and C strains (Table S5).

Previously, several VNTR loci, including those located on the pXO2 plasmid, have been suggested for MLVA genotyping of the anthrax pathogen. However, the loci located on this plasmid represent clusters of 2–3 bp long repeats, with expensive equipment thus being needed to determine their length [32]. The locus described here represents a region containing three or four 9 bp long tandem repeats, which makes it possible to separate the amplified fragments in the agarose gel.

The results showed that the identified polymorphism allows using the number of repeats at this locus as a diagnostic marker to differentiate between the evolutionary lineages. Therefore, we assessed the possibility of using VNTRacpA for MLVA genotyping in this study.
Unfortunately, the method of nucleotide sequence assembly using whole genome sequencing data employed in this study did not allow the number of repeats in the VNTRacpA region to be unambiguously determined in the strains from SRCAMB (Obolensk, Russia) including those belonging to lineage B. For this reason we designed PCR primers flanking the described repeat region and tested them on the DNA isolated from the SRCAMB (Obolensk, Russia) B. anthracis strains included in the studied sample. We used the MLVA genotyping approach, which involved the amplification of the VNTRacpA locus and the amplified fragments separation in the agarose gel.

Figure 2 shows an example of using the VNTRacpA locus for MLVA genotyping of B. anthracis strains and differentiating evolutionary lineages A and B. The results allowed us to differentiate between the strains in evolutionary lines A and B with high confidence and confirmed the hypothesis of the presence of two tandem repeat alleles, specific to the two evolutionary lines of the anthrax pathogen. However, the analysis revealed that the theoretical length of the amplified fragments differed from the observed one. In particular, the theoretical length of the fragment containing three tandem repeats, was 245 bp, and of the fragment containing four repeats, 254 bp (Fig. 2). The observed lengths were 300 and 309 bp.

CONCLUSIONS

To summarize, our findings demonstrate that the segregation of the studied sample into the MVLSTpXO2-GTs corresponds to its segregation into the main B. anthracis evolutionary lineages. The capC 351A → G SNP is the marker of the A.Br.Aust94 group strains circulating in the Caucasus region. MVLSTpXO2 genotyping subdivided the A.Br.Vollum group strains into two subgroups one of which was isolated in the former Soviet Union. The absence of the capD 234T → C SNP in the B.Br.CNEVA group strains may be a possible marker indicating that the strain originated from the former Soviet Union. The possibility of using the VNTRacpA locus to differentiate between anthrax pathogen strains belonging to evolutionary lines A and B by MLVA genotyping was demonstrated. This approach enables the results to be obtained in a short time and allows avoiding errors associated with the assembly of sequencing results. The described genotypes allow evaluating factors associated with microorganism virulence under certain environmental conditions and identifying potential risks caused by climate change.

FUNDING

The study was supported by the Sectoral Program of the Federal Service for Supervision of Consumer Protection and Welfare.

COMPLIANCE WITH ETHICAL STANDARDS

The authors declare that they do not have conflicts of interest. No experimentation involving animals or humans was performed by any of the authors.

SUPPLEMENTARY INFORMATION

The online version contains supplementary material available at https://doi.org/10.1134/S0003683822040056.

OPEN ACCESS

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

REFERENCES

1. Mogrovejo-Arias, D., Brill, F.H.H., and Wagner, D., Environ. Earth Sci., 2020, vol. 79, p. 109.
2. Marinin, L.I., Onishchenko, G.G., Kravchenko, T.B., Dyatlov, I.A., Tyurin, E.A., Stepanov, A.V., and Nikiforov, V.V., Sibirskaya yaza cheloveka: epidemiologiya, profilaktika, diagnostika, lechenie (Human Anthrax: Epidemiology, Prevention, Diagnosis, and Treatment), Moscow: ZAO MP Giggiena, 2008.
3. Bourgogne, A., Drysdale, M., Hilsenbeck, S.G., Peterson, S.N., and Koehler, T.M., Infect. Immun., 2003, vol. 71, no. 5, pp. 2736–2743.
4. Wu, G., Feng, C., Cao, S., Guo, A., and Liu, Z., Appl. Biochem. Biotechnol., 2012, vol. 168, no. 5, pp. 1302–1310.
5. Raynor, M.J., Roh, J.H., Widen, S.G., Wood, T.G., and Koehler, T.M., Mol. Microbiol., 2018. https://doi.org/10.1111/mmi.13961
6. Candela, T. and Fouet, A., Mol. Microbiol., 2006, vol. 60, no. 5, pp. 1091–1098.
7. Makino, S., Watarai, M., Cheun, H.I., Shirahata, T., and Uchida, I., J. Infect. Dis., 2002, vol. 186, pp. 227–233.
8. Makino, S., Uchida, I., Terakado, N., Sasakawa, C., and Yoshikawa, M., J. Bacteriol., 1989, vol. 171, no. 2, pp. 722–730
9. Drysdale, M., Bourgogne, A., Hilsenbeck, S.G., and Koehler, T.M., J. Bacteriol., 2004, vol. 186, no. 2, pp. 307–315.
10. Timofeev, V., Bahtejeva, I., Mironova, R., Titareva, G., Lev, I., Christiany, D., et al., PLoS One, 2019, vol. 14.
11. Simonova, E.G., Kartavaya, S.A., Titkov, A.V., Loktionova, M.N., Raichich, S.R., Tolpin, V.A., et al., Probl. Osobo Opasnykh Infekts., 2017, no. 1, pp. 89–93.

12. Kolsto, A.B., Tourasse, N.J., and Økstad, O.A., Annu. Rev. Microbiol., 2009, vol. 63, pp. 451–476.

13. Carroll, L.M., Kovac, J., Miller, R.A., and Wiedmann, M., Appl. Environ. Microbiol., 2017, vol. 83, no. 17, article ID e01096–17. https://doi.org/10.1128/AEM.01096-17

14. Keim, P., Van Ert, M.N., Pearson, T., Vogler, A.J., Huyhn, L.Y., and Wagner, D.M., Infect. Genet. Evol., 2004, vol. 4, no. 3, p. 205.

15. Van Ert, M.N., Easterday, W.R., Huyhn, L.Y., Okinaka, R.T., Hugh-Jones, M.E., Ravel, J., et al., PLoS One, 2007, vol. 2, article ID e461. https://doi.org/10.1371/journal.pone.0000461

16. Keim, P., Price, L.B., Klevytska, A.M., Smith, K.L., Schupp, J.M., Okinaka, R., et al., J. Bacteriol., 2000, vol. 182, pp. 2928–2936.

17. Le Flèche, P., Hauck, Y., Onteniente, L., Prieur, A., Denœud, F., Ramisse, V., et al., BMC Microbiol., 2001, vol. 1, p. 2. https://doi.org/10.1186/1471-2180-1-2

18. Klee, S.R., Ozel, M., Appel, B., Boesch, C., Ellerbrok, H., Jacob, D., et al., J. Bacteriol., 2006, vol. 188, no. 15, pp. 5333–5344.

19. Leendertz, F.H., Yumlu, S., Pauli, G., Boesch, C., Couacy-Hymann, E., Vigilant, L., et al., PLoS Pathog., 2006, vol. 2, no. 1, article ID e8.

20. Hoffmann, C., Zimmermann, F., Biek, R., Kuehl, H., Nowak, K., Mundry, R., et al., Nature, 2017, vol. 548, no. 7665, pp. 82–86.

21. Wu, R., Richter, S., Zhang, R.G., Anderson, V.J., Misiakas, D., and Joachimiak, A., J. Biol. Chem., 2009, vol. 284, no. 3, pp. 24406–24414.

22. Khavrutskii, I.V., Legler, P.M., Friedlander, A.M., and Wallqvist, A., Biochemistry, 2014, vol. 53, no. 44, pp. 6954–6967.

23. Goncharova, Yu.O., Bakhteeva, I.V., Titareva, G.M., Mironova, R.I., Kislichkina, A.A., Maiskaya, N.V., et al., Bakteriologiya, 2019, vol. 4, no. 2, pp. 7–12.

24. Goncharova, Y., Bahtjeva, I., Titareva, G., Kravchenko, T., Lev, A., Dyatlov, I., and Timofeev, V., Pathogens, 2021, vol. 10, no. 12, p. 1556. https://doi.org/10.3390/pathogens10121556

25. Derzelle, S., Aguilar-Bultet, L., and Frey, J., Infect. Genet. Evol., 2016, vol. 46, pp. 50–58. https://doi.org/10.1016/j.meegid.2016.10.019

26. Hellberg, R.S. and Chu, E., Crit. Rev. Microbiol., 2016, vol. 42, no. 4, pp. 548–572.

27. Rossati, A., Int. J. Occup. Environ. Med., 2017, vol. 8, no. 1, pp. 7–20.

28. Maksimovic, Z., Cornwell, M.S., Semren, O., and Rifaatbegovic, M., Rev. Sci. Tech., vol. 36, no. 3, pp. 959–963.

29. Nickle, D.C., Learn, G.H., Rain, M.W., Mullins, J.I., and Mittler, J.E., J. Mol. Evol., 2002, vol. 4, no. 1, pp. 134–137.

30. Bichara, M., Wagner, J., and Lambert, I.B., Mutat. Res., 2006, vol. 598, nos. 1–2, pp. 144–163.

31. Fan, H. and Chu, J.Y., Genomics Proteomics Bioinformatics, 2007, vol. 5, no. 1, pp. 7–14.

32. Thierry, S., Tourterel, C., Le Flèche, P., Derzelle, S., Dekhil, N., Mendy, C., et al., PLoS One, 2014, vol. 9, no. 6, article ID e95131.

Translated by E. Martynova