Original Study Article
https://doi.org/10.36233/0372-9311-385

In silico analysis of genomes of Bacillus anthracis strains belonging to major genetic lineages

Evgeny I. Eremenko✉, Grigorii A. Pechkovskii, Alla G. Ryazanova, Sergey V. Pisarenko, Dmitry A. Kovalev, Lyudmila Yu. Aksenova, Ol'ga V. Semenova, Alexander N. Kulichenko

Stavropol Plague Control Research Institute, Stavropol, Russia

Abstract

Introduction. The global phylogenetic population structure of *Bacillus anthracis* is represented by major genetic lineages (A, B and C) with nonuniform distribution of isolates, which still cannot be explained. Identification of characteristics of genomes of strains from three lineages, which can affect their spread, is of high importance.

The **aim** of the study is to explore genomic characteristics of different genetic lineages, which may have an effect on their distribution, by using the *in silico* analysis of a representative subset of *B. anthracis* strains.

Materials and methods. The whole-genome sequences of 49 *B. anthracis* strains and *Bacillus cereus biovar anthracis* CI strain were studied. The *in silico* analysis was performed to identify polymorphisms using BLASTn, MEGA X, Tandem Repeat Finder, Parsnp the Harvest Suite software.

Results. The genome variability depended on single nucleotide polymorphisms, single-nucleotide repeats, number of tandem repeats, substitutions and indels. In strains from lineages B and C, they outnumbered 1.6–13.4 times and in the *B. cereus biovar anthracis* strain — 5–150 times those in *B. anthracis* strains from lineage A. Significant substitutions in housekeeping genes and pathogenicity factor genes caused changes in amino acid sequences in proteins significantly more frequently in *B. anthracis* strains from major lineages B and C.

Based on the molecular typing and a multi-virulence-locus sequence typing analysis (MVLST) with a discrimination index of 0.9633, strains were classified into three major genetic lineages including groups different from the canonical group.

Conclusion. The distinctive feature of *B. anthracis* genomes is that they have a larger number of significant nucleotide substitutions in pathogenicity factor genes and housekeeping genes of strains belonging to major lineages B and C compared to lineage A. Changes in proteins encoded by them can cause differences in ecological adaptation and in prevalence, which are higher in strains of lineage A. MVLST having a high discriminating capacity can be used as an additional method to *B. anthracis* molecular typing.

Keywords: *Bacillus anthracis, genetic lineages, pathogenicity factors, molecular typing*

Funding source. This study was not supported by any external sources of funding.

Conflict of interest. The authors declare no apparent or potential conflicts of interest related to the publication of this article.

For citation: Eremenko E.I., Pechkovskii G.A., Ryazanova A.G., Pisarenko S.V., Kovalev D.A., Aksenova L.Yu., Semenova O.V., Kulichenko A.N. *In silico* analysis of genomes of Bacillus anthracis strains belonging to major genetic lineages. *Journal of microbiology, epidemiology and immunobiology = Zhurnal mikrobiologii, èpidemiologii i immunobiologii.* 2023;100(3):155–165.
DOI: https://doi.org/10.36233/0372-9311-385
EDN: https://www.elibrary.ru/ocpnyx

Оригинальное исследование
https://doi.org/10.36233/0372-9311-385

Анализ *in silico* геномов штаммов *Bacillus anthracis* главных генетических линий

Еременко Е.И.✉, Печковский Г.А., Рязанова А.Г., Писаренко С.В., Ковалев Д.А., Аксенова Л.Ю., Семенова О.В., Куличенко А.Н.

Ставропольский противочумный институт Роспотребнадзора, Ставрополь, Россия

Аннотация

Введение. Филогенетическая структура глобальной популяции *Bacillus anthracis* представлена главными генетическими линиями A, B и C с неравной распространённостью изолятов, причина которой неизвестна. Определение особенностей геномов штаммов трех линий, которые могут влиять на распространённость, является актуальным.

Цель — характеристика особенностей геномов разных генетических линий, потенциально влияющих на их распространённость, с использованием анализа *in silico* представительной выборки штаммов *B. anthracis*

Материалы и методы. Изучены полногеномные последовательности 49 штаммов *B. anthracis* и штамма CI *B. cereus biovar anthracis*. Анализ *in silico* проводили с идентификацией полиморфизмов в программах «BLASTn», «MEGA X», «Tandem Repeat Finder», «Parsnp» из пакета «Harvest Suite».

Результаты. Вариабельность геномов определялась однонуклеотидными полиморфизмами, однонуклеотидными повторами, числом тандемных повторов, заменами и инделами. Их количество у штаммов линий B и C было в 1,6–13,4 раза больше, а у штамма *B. cereus biovar anthracis* — в 5–150 раз больше, чем у штаммов *B. anthracis* линии A. Значимые замены в генах домашнего хозяйства и факторов патогенности приводили к изменению аминокислотной последовательности белков также значительно чаще у штаммов *B. anthracis* главных линий B, C.

Молекулярное типирование на основе анализа однонуклеотидных полиморфизмов генов факторов патогенности (MVLST) с индексом дискриминации 0,9633 разделяло штаммы на три главные генетические линии с группами, отличающимися от канонических.

Заключение. Главное отличие геномов *B. anthracis* состоит в большом количестве значимых нуклеотидных замен в генах факторов патогенности и «домашнего хозяйства» штаммов главных линий B и C по сравнению с линией A. Изменения в кодируемых ими белках могут определять разную экологическую адаптацию и распространённость, более высокие у линии A. MVLST с высокой дискриминирующей способностью может быть дополнительным методом молекулярного типирования *B. anthracis*.

Ключевые слова: *Bacillus anthracis, генетические линии, факторы патогенности, молекулярное типирование*

Источник финансирования. Авторы заявляют об отсутствии внешнего финансирования при проведении исследования.

Конфликт интересов. Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Для цитирования: Еременко Е.И., Печковский Г.А., Рязанова А.Г., Писаренко С.В., Ковалев Д.А., Аксенова Л.Ю., Семенова О.В., Куличенко А.Н. Анализ *in silico* геномов штаммов *Bacillus anthracis* главных генетических линий. *Журнал микробиологии, эпидемиологии и иммунобиологии.* 2023;100(3):155–165.
DOI: https://doi.org/10.36233/0372-9311-385
EDN: https://www.elibrary.ru/ocpnyx

Introduction

The global population of *Bacillus anthracis* is represented by three major genetic lineages (A, B, and C) including 14 canonical (canSNP) groups. Pathogen strains are distributed among them nonuniformly; representatives of lineage A prevail (about 90%), lineage B accounts for approximately 10%, while lineage C includes only 3 strains (less than 1%) [1, 2]. The distinct clade in the *Bacillus cereus sensu lato* group including *B. anthracis* is represented by *Bacillus cereus biovar anthracis* strains that can cause anthrax-like infection [3, 4]. The factors causing the nonuniform distribution of *B. anthracis* strains in different genetic lineages are still unknown. The question why strains of lineage A have become widely dispersed throughout the world, while strains of lineage B are limited in number and do not have a wide global distribution remains unanswered.

This nonuniformity can be explained by adaptive genetic differences, which have an impact on survival and reproduction in the environment or in the host body.

The reproduction efficiency in the host body depends on adaptive genetic differences of pathogenicity factors. The pathogenicity of *B. anthracis* depends on the main factors: Two binary exotoxins, lethal and edema factors, and the poly-γ-glutamic acid capsule [2]. Exotoxin components, lethal and edema factors, and the protective antigen (PA) are encoded by *lef*, *cya* and *pagA* genes located on the pXO1 plasmid [5]. The capBCDAE operon for capsule production is encoded by the pXO2 plasmid [2]. The loss of either of the two plasmids results in avirulence of strains. At the same time, strains carrying both plasmids significantly differ by their virulence as demonstrated by laboratory animals. The Carbosap vaccine strain attenuation not associated with the loss of pXO1 and pXO2 plasmids results from chromosomal deletions containing more than 50 genes, which have a known or proposed function that could be linked to virulence [6]. The detected differences in the virulence of strains producing a specific toxin and a typical capsule suggest the existence of additional pathogenicity factors. The role of such factors is claimed by many products of the anthrax microbe:

- the GerXC protein encoded by the *gerXC* plasmid gene and required for spore germination *in vivo* [7];

- phospholipase C (*PlC* gene) [8];
- nitric oxide synthase (*nos* gene) [9];
- bifunctional lysyl-phosphatidylglycerol flippase/synthase (*mprF* gene) [10];
- metalloprotease of the enhancin family (*GBAA_RS16775* locus), homolog of which was first described in baculoviruses [11, 12];
- immune inhibitor A metalloprotease (*inhA* gene) [12];
- quorum-sensing signal molecule autoinducer LuxS (*luxS* gene) [13];
- anthorolysin O (ALO) (cholesterol-dependent cytolysin) (*alo* gene) [14];
- enterotoxin FM (EntFM) (*entFM* gene) [13].

Some *B. anthracis* strains have a set of phenotypic characteristics, which distinguish them from typical virulent strains; among them, the inability to grow in a minimal synthetic medium without tryptophan, and reduced virulence in rabbits. The genetic factors underlying these changes have not been identified [16].

There are publications reporting the effect of amino acid substitutions in the lethal factor protein on its catalytic activity and binding to PA [17–19]. The data were obtained during tests where mutations were introduced into the described genes to assess their impact on the virulence of mutant strains compared to wild-type strains. Allelic polymorphism for the PA gene that is represented by 6 allelic types in natural wild-type strains [20] was also described for other genes of pathogenicity factors [21–23]. However, the variability of chromosomal genes encoding products, for which the impact on the virulence of *B. anthracis* has been identified or assumed, remains unexplored. Quantitative and qualitative distinctive characteristics of polymorphisms inherent in genomes of strains belonging to certain genetic lineages have not been identified.

The molecular typing of *B. anthracis* using multi-virulence-locus sequence typing (MVLST) did not include chromosomal genes.

The urgent need for this study was dictated by the absence of the data that would shed a new light on the pathogenesis of anthrax infection, help identify potential targets for development of new products for treatment and prevention of this infection, contribute to the knowledge of the evolution of the anthrax-causing pathogen and methods of its molecular typing.

The **aim** of the study was to explore characteristics of genomes of different genetic lineages, which have a potential impact on the distribution of strains, by using the *in silico* analysis of a representative subset of *B. anthracis* strains.

Materials and methods

The study was performed using complete genomes of 49 diplasmid strains, including 19 *B. anthracis* isolates from the collection of pathogenic microorganisms from the Stavropol Research Anti-Plague Institute of Rospotrebnadzor and 30 isolates from the GenBank database, which belong to major genetic lineages A, B, C and 14 canSNP groups, as well as the genome of *Bacillus cereus biovar anthracis* strain CI.

GenBank[1] identifiers for genomes:
GCF_000008445.1, GCF_009831565.1, GCF_000167335.1, GCF_003063965.1, GCF_003064045.1, GCF_003860145.1, GCF_000793525.1, GCF_000832965.1, GCF_000310045.1, GCF_000167235.1, GCF_000534935.2, GCF_000258885.1, GCF_000278385.1, GCF_000832465.1, GCF_001273005.1, GCF_001273085.1, GCF_000167295.1, GCF_002896575.1, GCF_014249775.1, GCF_003227955.1, GCF_000831505.1, GCF_000832745.1, GCF_003064005.1, GCF_000008165.1, GCF_000583105.1, GCF_000833275.1, GCF_022221345.1, GCF_000743805.1, GCF_900014355.1, GCF_002356575.1, GCF_000143605.1.

The *in silico* analysis was performed using the genome of the *B. anthracis* Ames Ancestor strain (GenBank: NC_007530.2; NC_007322.2; NC_007323.3) as a reference genome. Polymorphisms were identified using BLAST, BLASTp, MEGA X, MAUVE, Tandem Repeat Finder programs. The alignment of pooled sequences of pathogenicity factor genes and the translation of nucleotide sequences of genes were performed using the MEGA X program. The whole-genome analysis of single nucleotide polymorphisms (SNPs) was performed using the Parsnp program from the Harvest Suite for multiple alignment of genomic sequences. The input data included genomes of 50 strains described above; they were aligned with the chromosomal nucleotide sequence of the reference genome of *B. anthracis* Ames Ancestor (GenBank: NC_007530.2) using Parsnp (parameters c -e -u -C 1000). The detected SNPs were extracted into a VCF file using HarvestTools v. 1.0. The edited file was used as an input file in HarvestTools to compile a FASTA file.

The phylogenetic tree was reconstructed by the neighbor-joining method in MEGA X in accordance with Tamura-Nei model; the bootstrapped confidence interval was based on 1,000 replicates; the goeBURST Full MST method and PHYLOViZ 2.0 program were used for identification of clonal complexes. The FigTree program was used for dendrogram visualization. The Hanter–Gaston discriminatory index was calculated in accordance with [24].

Results

Analysis of polymorphisms of the chromosomal region of the B. anthracis and Bacillus cereus biovar anthracis

[1] URL: ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/833/275/

DOI: https://doi.org/10.36233/0372-9311-385

genome

The variability of the chromosomal region in *B. anthracis* and *B. cereus biovar anthracis* depended on SNPs, single-nucleotide repeats (SNR), tandem repeats, substitutions and indels (insertions/deletions). The main polymorphisms in chromosomal genomes were represented by SNPs (**Table 1**). Strains belonging to different genetic lineages had noticeable differences in the number of polymorphisms.

Chromosomal marker (specific) SNPs were identified for *B. anthracis* major genetic lineages (**Table 2**).

Strains of lineages B and C demonstrated significant differences in the relative amounts of marker SNPs compared to lineage A.

Most of the marker SNPs from lineages A and B were primarily located in housekeeping genes, out of which 4 genes were associated with sporulation and spore germination.

Analysis of SNPs of pathogenicity factor genes

The analysis included 19 variable genes encoding products associated with pathogenicity (**Table 3**).

The gene variability depended on the presence of SNPs, VNTRs and INDELs. The largest number of all SNPs and nonsynonymous SNPs per one genome was found in the *B. cereus biovar anthracis* strain followed by *B. anthracis* strains of lineage C, lineage B, lineage A. The deletion in the *trpA* gene resulted in the formation of pseudogenes and the absence of functional proteins in some strains of lineage B. In the *B. cereus biovar anthracis* strain, the frameshift mutation caused the formation of the pseudogene and absence of the Mobl relaxase (capsule-production plasmid plC-XO2). There are noticeable differences in the number of SNPs in different genes and in strains of different genetic lineages.

Analysis of the polymorphism of pathogenicity factor proteins in strains of different genetic lineages

The lethal factor. In the lethal factor protein, substitutions *E709G* and *E681K* are located within domain 4, which contains the catalytic center, at the distance of 10 and 37 amino acids, respectively, from the zinc binding site; substitutions *A299T*, *L298M* and *R543Q* are located in domain 2, substitutions *E66K* and *V246I* — in domain 1.

The edema factor. All strains of lineage B had substitutions *D180G* and *318T*. The strain of lineage C and *B. cereus biovar anthracis* strain CI have substitutions *K278E*, *I318T* and *N789K*. The *B. cereus var. anthracis* strain CI has substitution *V694A*. Substitutions *D180G* and *K278E* are located within the PAB domain, *I318T* — in the CA segment of the ACD core domain, *V694A* and *N789K* — in the helical region of the edema factor.

The protective antigen. Strains of lineage A were found to have substitutions *A600V* and *P565S*; strains of lineage B had substitutions *I433V* and *A600V*; the strain of lineage C and *B. cereus biovar anthracis* strain CI had substitutions *S66P* and *A600V*. *B. cereus biovar anthracis* strain CI had substitution *S290I*. Substitution *S66P* was located in domain 1, substitution *I433V* — in domain 2, substitution *P565S* – in domain 3; substitution *A600V* — in domain 4 in the receptor-binding region (L595–T735).

The comparative analysis was performed using PA genes and proteins in all live spore vaccines based on strains Carbosap, 34F2_Sterne, A16R, Tsiankovskii-1, STI-1, 55VNIIViM, 228/8 and Brazilian vaccinal as well as in chemical vaccines based on PA of the avirulent V770-NP-1R strain (U.S. vaccines AVA (or BioThrax) and AV7909) and based on PA of vaccine strain 34F2 Sterne (vaccine AVP; United Kingdom). All

Table 1. Polymorphisms of chromosome genomes of *B. anthracis* and *B. cereus biovar anthracis* strains

Major lineage	Strain	Quantity of polymorphisms comparing with *B. anthracis* strain Ames Ancestor					
		SNP	SNR	tandem repeats	substitutions	indels	total
A	Australia 94	411	142	23	8	64	648
A	Vollum	609	233	32	12	68	954
B	SVA11	1693	418	73	16	109	2309
C	2002013094	2381	576	99	134	414	3604
B. cereus biovar anthracis	CI	76 714	1075	188	7857	1783	87 617

Table 2. Marker SNPs for chromosome genomes of *B. anthracis*

Major lineage	Quantity of marker SNP	Ratio marker SNPs/genome	SNP localization	
			gene	intergene space
A	180	4,73	152	28
B	183	18,3	141	42
C	594	594	Not tested	Not tested

Table 3. SNPs in pathogenicity factors genes of *B. anthracis* strains of different lineages

Gene	Ratio quantity of SNPs/genome in strains of lineages							
	A (*n* = 38)		B (*n* = 10)		C (*n* = 1)		*B. cereus biovar anthracis* (*n* = 1)	
	total	nonsynonimic	total	nonsynonimic	total	nonsynonimic	total	nonsynonimic
cya	7/0,18	4/0,1	3/0,3	2/0,2	6	4	6	4
lef	2/0,05	2/0,05	2/0,2	2/0,2	4	3	10	8
pagA	4/0,1	2/0,05	3/0,3	2/0,2	3	2	4	3
atxA	1/0,02	0	0	0	0	0	1	1
capA	3/0,07	2/0,05	1/0,1	1/0,1	1	1	1	0
capC	2/0,05	0	0	0	1	1	1	0
capD	3/0,07	3/0,07	2/0,2	0	3	3	4	4
acpA	1/0,02	0	1/0,1	1/0,1	1	1	3	2
ger XC	1/0,02	1/0,02	2/0,2	2/0,2	1	1	1	1
mprF	2/0,05	1/0,02	2/0,2	2/0,2	2	1	23	1
entFM	2/0,05	2/0,05	1/0,1	1/0,1	0	0	17	5
GBAA_RS16775	1/0,02	1/0,02	2/0,2	2/0,2	0	0	132	31
plC	0	0	1/0,1	1/0,1	1/0,1	1/0,1	9	3
alo	3/0,07	3/0,073	0	0	0	0	57	23
nos	0	0	2/0,2	2/0,2	0	0	32	9
luxS	0	0	0	0	1	1	8	0
trpA	0	0	3/0,3	1/0,1	0	0	18	5
trpD	0	0	1/0,1	1/0,1	1	0	15	4
GBAA_RS06415 (trpG)	0	0	1/0,1	1/0,1	1	1	14	5

the strains belonged to lineage A. Strains V770-NP-1R, Carbosap and all the strains of the Russian origin (Tsiankovskii-1, STI-1, 55VNIIViM, 228/8) had substitutions in gene *C→T* at the position 195 and 1799; in addition, the Tsiankovskii-1 strain had substitution 981 *A→T*. The same strains had substitution *A600V* in PA domain 4 in the receptor-binding region. No substitutions were detected in PA gene and protein in strains 34F2 Sterne, Brazilian vaccinal and A16R.

Capsular polyglutamate synthetase CapA: strains of lineage A had substitution *Q399K*; strains of lineage B had substitution *T345A*; lineage C strains had substitution *V156L*.

Capsule biosynthesis protein CapC: substitution *T80M* in strains of lineage C and *B. cereus* strains.

CapD gamma-glutamyl transferase: strains of lineage A had substitutions *I4M*, *V266I* and *S381E*; strains of lineage B did not have any substitutions; the strain of lineage C and *B. cereus biovar anthracis* strain CI had substitutions *H70Y*, *K223E* and *F379I*; *B. cereus biovar anthracis* strain CI had substitution *G499D*. Substitution *H70Y* was located in the L strand, substitution *F379I* — in the S strand.

Capsule synthesis transcriptional regulator AcpA: strains of lineages B and C and *B. cereus biovar anthracis* strain CI had substitution *E285K*; *B. cereus* also had substitution *Y354H*.

Anthrax toxin trans-activating transcriptional regulator AtxA: substitution *I188N* in *B. cereus biovar anthracis* strain CI.

Anthrolysin O (ALO): strains of the A.Br.Vollum group had substitution *S422F*; strains of the A.Br. Aust94 group, which were isolated in South Africa, had substitution *N221T*; ; strains of the A.Br.005/006 group had substitution *V416G*.

Metalloprotease of the enhancin family: strains of lineage A had substitution *P631S*; strains of lineage B — substitution *L139F*; strains of lineage C — substitution *D444E*, the protein of *B. cereus biovar anthracis* strain CI had 29 substitutions.

GerXC spore germination protein: lineages A and B had substitution *H29R*; lineage B — substitution *T35I*, lineage C and *B. cereus biovar anthracis* — substitution *E219G*.

Protein production autoinducer-2, LuxS: lineage C had substitution *D111G*.

Bifunctional lysyl-phosphatidylglycerol flippase/synthetase MprF: lineage A had substitution *H631R*; lineage B — substitution *L289F*; lineages B and C as well as *B. cereus biovar anthracis* — substitution *V424I*.

Nitric oxide synthase NOS: lineage B — substitutions *Q288H* and *I348F*; *B. cereus biovar anthracis* — 9 substitutions.

DOI: https://doi.org/10.36233/0372-9311-385

Phospholipase C: strains of lineages B and C as well as *B. cereus biovar anthracis* had substitution *H194Y*; *B. cereus biovar anthracis* — had substitutions *N20S* and *A59V*.

Tryptophan synthase alpha subunit: lineage B — substitutions *T222K* and a non-functional protein in some strains of this lineage due to the deletion and formation of a pseudogene; *B. cereus biovar anthracis* — 5 substitutions.

Anthranilate phosphoribosyl transferase TrpD: lineage B — substitution *N300S*; *B. cereus biovar anthracis* — 4 substitutions.

Aminodeoxychorismate/anthranilate synthase component II, TrpG: Lineages B and C — substitutions*N300S*; *B. cereus biovar anthracis* — 5 substitutions. *H70Y* is located in the L strand, *F379I* — in the S strand.

The obtained data on substitutions in plasmid genes and proteins of pathogenicity factors correlate with the data reported in earlier [22, 23]. Substitutions in genes of additional pathogenicity factors having chromosomal localization have been described by us for the first time.

The largest number of significant substitutions was identified in genes of enhancin and anthorolysin O in *B. cereus biovar anthracis* strain CI. In total, 19 pathogenicity factor genes had 15 significant substitutions in 38 strains of lineage A, 20 substitutions in 10 strains of lineage B, 20 substitutions in 1 strain of lineage C, 102 substitutions in *B. cereus biovar anthracis* strain CI. The consistency of pattern is observed: Strains from major lineages B and C, let alone, the *B. cereus biovar anthracis* strain, had significantly larger numbers of amino acid substitutions in pathogenicity factor proteins, which can change their functional activity, than strains from lineage A.

Sequence typing based on SNPs in pathogenicity factors genes

A total of 409 phylogenetically significant SNPs and 33 genotypes of pathogenicity factors were identified in 19 pathogenicity factor gene of 49 *B. anthracis* strains and 1 *B. cereus biovar anthracis* strain identified.

For the multi-virulence-locus sequence typing (MVLST), the Hunter–Gaston discriminatory index was 0.9633, being higher than the respective index for canSNP-typing (0.9056), coming close to the index for WGS-SNP typing (0.9869). Similar results were demonstrated by the comparative analysis of the MVLST and whole-genome SNP typing effectiveness [25]. Compared to canSNP and coreWGS-SNP typing, the MVLST typing based on SNPs of pathogenicity factor genes can be used for chromosomal and plasmid genes.

The dendrogram based on SNPs of pathogenicity factor genes also shows three universally recognized major genetic lineages: A, B, C of *B. anthracis* and the clade of the *B. cereus biovar anthracis* strain; the latter is basic for all the three major genetic *B. anthracis* lineages, while the clade of lineage C is basic for lineages B and C (**Fig. 1**). These findings support the idea of the evolution of *B. anthracis* from its predecessor *B. cereus* to lineage C and then to lineages B and C; they can also suggest that variability of pathogenicity factor genes can be seen as a driving force of the evolution.

The phylogenetic relationships of strains, which were reconstructed using multilocus sequence typing, differed from those demonstrated by whole-genome SNP typing.

The canSNP group B.Br.001/002 and B.Br.Kruger clustering conformity was observed using multilocus sequence typing and whole-genome SNP typing.

The comparison of the dendrograms constructed with multilocus sequence typing and whole-genome SNP typing of the same strains (Fig. 1) showed that some strains belonging to other canSNP groups are clustered with strains from groups that do not conform to their canSNP-group affiliation.

Identification of clonal complexes using multilocus sequence typing

The identification of clonal complexes (CC) using multilocus sequence typing resulted in identification of 33 genotypes (GTs) comprising 5 clonal complexes for 48 *B. anthracis* strains of major lineages A and B as well as two distinct GTs for *B. anthracis* strains of lineage C and *B. cereus bv anthracis* (**Fig. 2**). CC1 is the most numerous complex consisting of 12 GTs of 20 strains.

CC1 includes GTs of 18 strains of 7 groups out of 14 canSNP groups from major genetic *B. anthracis* lineage A. CC2 consists of 2 GTs of the A.Br.008/011 group and serves as an intermediate link between CC1, CC4 having strains of lineage A and CC5having strains of lineage B. CC3 includes 7 GTs from A.Br.Aust94, A.Br.001/002 and A.Br.Ames groups. CC4 consists of 5 GTs of the A.Br.008/011 group of major lineage A. CC5 includes 7 GTs of 10 strains from all canSNP groups of major *B. anthracis* lineage B.

In the clonal complexes, the genetic distances between GTs range from 1 to 6 units. All *B. anthracis* strains of lineage A are separated from strains of lineage B by the distance of 17 units (from GT17 to GT43), from the strain of lineage C — by the distance of 24 units (between GT7 and GT49), from the *B. cereus bv anthracis* strain (GT50) — by the distance of 340 units from the *B. anthracis* strain of lineage C (GT49).

The analysis of the clonal complexes confirms that *B. anthracis* strains are classified into three major genetic lineages and are distributed by genotypes of pathogenicity factor genes, which do not demonstrate complete conformity to canonical SNP groups. The analysis confirms the evolution of pathogenicity factors from *B. cereus biovar anthracis* to *B. anthracis* of lineage C and then to lineages A and B, thus making it possible to identify genetic groups different from canonical SNP groups.

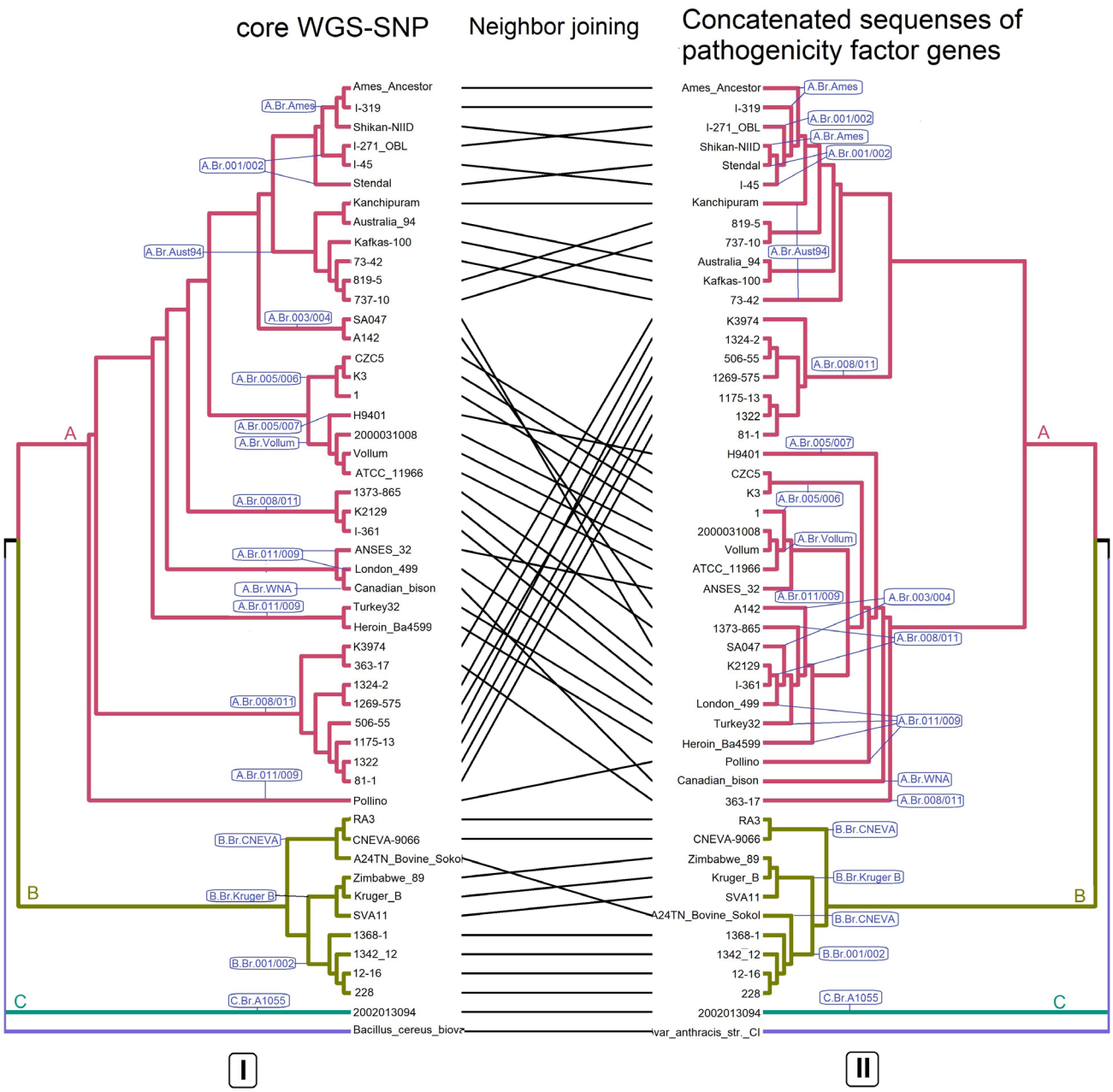

Fig. 1. Comparison of the phylogenetic reconstruction of the results received using the multilocus sequence typing and coreWGS-SNP.

Discussion

The number of all types of polymorphisms in the chromosomal region of genomes in *B. anthracis* strains of lineages B and C was 1.6-13.4 times and in the *B. cereus var. anthracis* strain 5-150 times as large as their number in *B. anthracis* strains of lineage A. Especially significant differences were observed in the numbers of substitutions, SNPs, and indels in the genome of the *B. cereus var. anthracis* strain, which were 785.7, 150 and 27 times as large as their numbers in *B. anthracis* strains of lineage A. These differences can be explained by the conformity between the chromosomal region of the genome of *B. cereus var. anthracis* strains and the genome of representatives of the *B. cereus sensu lato* group, except for *B. anthracis*, while pCI-XO1 and pCI-XO2 plasmids do not differ significantly from pXO1 and pXO2 plasmids of *B. anthracis* [4].

The number of chromosomal specific marker SNPs per genome had an inverse relationship with the number of lineage strains; for the only strain of major lineage C, this number was 24 times as large as the number for 10 strains of lineage B and exceeded 170 times the number for 38 strains of lineage A. This phenomenon can result from a longer evolution history involving accumulation of mutations in genetic lineages B and C compared to lineage A. As SNPs were located primarily in housekeeping genes, there is a high probability that these mutations were significant and could

GT	Strains	canSNP group	GT	Strains	canSNP group
GT1	Ames_Ancestor	A.Br.Ames	GT20	1175-13	A.Br.008/011
	Shikan-NIID	A.Br.Ames		1322	A.Br.008/011
	I-319	A.Br.Ames	GT21	1324	A.Br.008/011
	I-45	A.Br.001/002	GT22	363-17	A.Br.008/011
	I-271_OBL_	A.Br.001/002	GT19	506-55	A.Br.008/011
GT6	Stendal	A.Br.001/002	GT11	CZC5	A.Br.005/006
GT30	Kafkas-100	A.Br.AuGT94	GT34	Vollum	A.Br.Vollum
	Australia_94	A.Br.AuGT94		ATCC_11966	A.Br.Vollum
GT27	73-42	A.Br.AuGT94	GT38	Canadian_bison	A.Br.WNA
GT28	737-10	A.Br.AuGT94	GT37	2000031008	A.Br.Vollum
GT29	819-5	A.Br.AuGT94	GT10	1	A.Br.005/006
GT7	SA047	A.Br.003/004	GT36	H9401	A.Br.005/007
	A142	A.Br.003/004	GT9	K3	A.Br.005/006
	London_499	A.Br.011/009	GT43	RA3	B.Br.CNEVA
	K2129	A.Br.008/011		CNEVA-9066	B.Br.CNEVA
	1373-865	A.Br.008/011	GT39	228	B.Br.001/002
	I-361	A.Br.008/011		12-16	B.Br.001/002
	Turkey32	A.Br.011/009		1342_12	B.Br.001/002
	Heroin_Ba4599	A.Br.011/009	GT42	1368-1	B.Br.001/002
GT12	ANSES_32	A.Br.011/009	GT45	A24TN_Bovine_Sokol	B.Br.CNEVA
GT33	Polino	A.Br.011/009	GT46	Kruger_B	B.Br.Kruger B
GT17	K3974	A.Br.008/011	GT48	SVA11	B.Br.001/002
GT18	81-1	A.Br.008/011	GT47	Zimbabwe_89	B.Br.001/002
GT23	1269	A.Br.008/011	GT49	2002013094	C.Br.A1055
GT32	Kanchipuram	A.Br.Aust94	GT50	Bacillus_cereus_biovar anthracis_str._CI	B. c. bv anthr

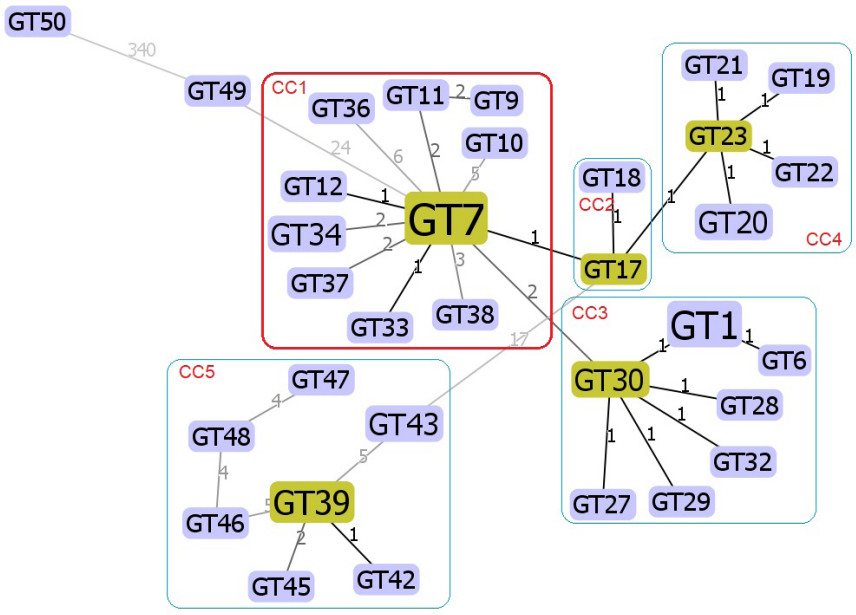

Fig. 2. Clonal complexes of *B. anthracis* strains. The dendrogram was constructed using multilocus sequence typing and the goeBURST Full MST algorithm in the PHYLOViZ 2.0 program; numbers on the clades correspond to genetic distances.

affect the ecological adaptation of lineages B and C, consequently, their limited spread.

The variability of 19 pathogenicity factor genes, out of which 9 were encoded by pXO1 and pXO2 plasmids of *B. anthracis* or pCI-XO1 and pCI-XO2 of *B. cereus biovar anthracis*, and another 10 — by a chromosome, was manifested by the presence of SNPs and indels. It should be noted that no indels have been detected in *B. anthracis* strains from lineage A, except for the strain 2000031008, as opposed to other lineages. In strains of lineages B and C, insertion ATATAGATA in the *acpA* gene caused the insertion of 3 amino acids NID (asparagine-isoleucine-aspartic acid) in the capsule synthesis transcriptional regulator AcpA. This insertion, a tandem repeat unit, was described as a new VNTR-locus first in our publication [26] and later by other authors [23]. The deletion in the 107–124 bp region in the *trpA* gene turned it into a pseudogene, deprived the TrpA tryptophan synthase alpha subunit of 35–40 N-terminal amino acids and made the enzyme nonfunctional in most of the strains of lineage B, which we have studied. These effects could explain the dependence of these strains on tryptophan [16].

PA is the main immunogenic component of anthrax vaccines. All vaccine strains of live vaccines as well as strains, PA of which is used in chemical vaccines, belonged to major genetic lineage A. Vaccine strains of Russian origin, strains V770-NP-1R and Carbosap have the alanine-to-valine substitution in the receptor-binding region of PA domain 4, which is absent in vaccine strains from China, Brazil and in the 34F2 Sterne strain used for vaccination of livestock in Western countries. These data can be useful in development of new anthrax vaccines.

The global dominance of genotypes of lineage A serves as a proof of a bid reproductive success (and, consequently, adaptation) and considerable dispersion over large distances [27]. Smith *et al.* assume that strains of lineage A, though not lineage B, have a hypothetical ability to cause latent infection in animals, which is associated with their global distribution and the restricted distribution of lineage B. The comparison of isolates from lineages A and B from South Africa showed that strains A were adapted to more diverse environments than strains B, which were restricted by narrower environments [28]. The limited number and restricted geographical distribution of rarer lineages can be caused by higher adaptation costs, which are associated with niche specialization [29].

The genotypes from lineage C and, to a lesser extent, from lineage B, apparently have a very poor adaptation compared to the genotypes from lineage A. Indeed, the clade of lineage C is characterized by significantly slower rates of evolution than the clade of lineage A, implying the smaller number of infectious cycles in nature [1].

The differences between evolutionary lineages in terms of susceptible hosts can also contribute to differences in their distribution. Strains of the B.Br.CNEVA group of lineage B were reported only in France, Southern Germany, Switzerland, Northern Italy, Bosnia and Herzegovina, Croatia, Slovenia, Slovakia and Poland. They constitute a transalpine axis represented by valley grasslands with rich meadows, traditionally focusing on specific breeds of cattle, which were kept isolated and were not exchanged for centuries. Such geographic isolation might have provided favorable environment for spore survival and replication of *B. anthracis* of the B.Br.CNEVA group [2, 30].

Conclusion

There are significant differences in the numbers of polymorphisms in genomes of representatives of major genetic *B. anthracis* lineages A, B and C. Strains of the most geographically restricted lineage C had 4.5 times and strains of lineage B, which is also quite restricted, 3 times as many types of polymorphisms compared to strains of numerous lineage A. The dominating location of nucleotide substitutions, including significant substitutions, within housekeeping genes and pathogenicity factors could contribute to changes in functions of respective proteins. The expansion of lineage A can be explained by its advantages over lineages B and C, which were established during the evolution. Lineage C, which is evolutionary older and is primary in relation to lineages B and A as well as least adapted, is restricted in distribution by negative selection. The evolution of *B. anthracis*, which is associated with the variability of pathogenicity factors, helps identify genetic groups that are different from canonical SNP groups. MVLST having a good discriminatory power can be used as an additional method of molecular typing of the anthrax pathogen, making it possible to differentiate strains using pathogenicity determinant.

In our study, we, for the first time, found that strains of lineages B and C have significantly higher numbers of polymorphisms in genomes, including pathogenicity factor genes, compared to lineage A. We identified significant substitutions in chromosomal and plasmid genes, which have a potential effect on virulence, and demonstrated a high discriminatory power of MVLST) using the analysis of SNPs of 19 pathogenicity factor genes. For the first time, the mechanism underlying the tryptophan dependance in some *B. anthracis* strains of lineage B was identified, demonstrating its association with mutations in the gene of the tryptophan synthase alpha subunit.

The variability of genes associated with spore germination and spore formation should be studied further, as they can also have an impact on adaptation and distribution of the genetic lineages of *B. anthracis*.

Thus, the dominance of major genetic lineage A of *B. anthracis* can be explained, among other versions, by

a significantly smaller numbers of genomic mutations compared to B and, especially, C lineages as well as by its better adaptation to external and host environments.

СПИСОК ИСТОЧНИКОВ | REFERENCES

1. Pearson T., Busch J.D., Ravel J., et al. Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing. *Proc. Natl Acad. Sci. USA*. 2004;101(37):13536–41. DOI: https://doi.org/10.1073/pnas.0403844101
2. Pilo P., Frey J. Pathogenicity, population genetics and dissemination of *Bacillus anthracis*. *Infect. Genet. Evol.* 2018;64:115–25. DOI: https://doi.org/10.1016/j.meegid.2018.06.024
3. Leendertz F.H., Ellerbrok H., Boesch C., et al. Anthrax kills wild chimpanzees in a tropical rainforest. *Nature.* 2004; 430(6998):451–2. DOI: https://doi.org/10.1038/nature02722.
4. Klee S.R., Brzuszkiewicz E.B., Nattermann H., et al. The genome of a *Bacillus* isolate causing anthrax in chimpanzees combines chromosomal properties of *B. cereus* with *B. anthracis* virulence plasmids. *PLoS One*. 2010;5(7):e10986. DOI: https://doi.org/10.1371/journal.pone.0010986
5. Okinaka R.T., Cloud K., Hampton O., et al. Sequence and organization of pXO1, the large *Bacillus anthracis* plasmid harboring the anthrax toxin genes. *J. Bacteriol*. 1999;181(20): 6509–15. DOI: https://doi.org/10.1128/JB.181.20.6509-6515.1999
6. Harrington R., Ondov B.D., Radune D., et al. Genome sequence of the attenuated Carbosap vaccine strain of *Bacillus anthracis*. *Genome Announc.* 2013;1(1):e00067-12. DOI: https://doi.org/10.1128/genomea.00067-12
7. Sirard J.C., Guidi–Rontani C., Fouet A., Mock M. Characterization of a plasmid region involved in *Bacillus anthracis* toxin production and pathogenesis. *Int. J. Med. Microbiol*. 2000;290(4-5):313–6. DOI: https://doi.org/10.1016/S1438-4221(00)80030-2
8. Heffernan B.J., Thomason B., Herring-Palmer A., et al. *Bacillus anthracis* phospholipases C facilitate macrophage-associated growth and contribute to virulence in a murine model of inhalation anthrax. *Infect. Immun*. 2006;74(7):3756–64. DOI: https://doi.org/10.1128/IAI.00307-06
9. Shatalin K., Gusarov I., Avetissova E., et al. *Bacillus anthracis*-derived nitric oxide is essential for pathogen virulence and survival in macrophages. *Proc. Natl Acad. Sci. USA.* 2008;105(3):1009–13. DOI: https://doi.org/10.1073/pnas.0710950105
10. Samant S., Hsu F.F., Neyfakh A.A., Lee H. The *Bacillus anthracis* protein MprF is required for synthesis of lysylphosphatidylglycerols and for resistance to cationic antimicrobial peptides. *J. Bacteriol*. 2009;191(4):1311–9. DOI: https://doi.org/10.1128/JB.01345-08
11. Lepore L.S., Roelvink P.R., Granados R.R. Enhancin, the granulosis virus protein that facilitates nucleopolyhedrovirus (NPV) infections, is a metalloprotease. *J. Invertebr. Pathol*. 1996;68(2):131–40. DOI: https://doi.org/10.1006/jipa.1996.0070
12. Read T.D., Peterson S.N., Tourasse N., et al. The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature.* 2003;423(6935):81–6. DOI: https://doi.org/10.1038/nature01586
13. Jones M.B., Blaser M. Detection of a luxS-signaling molecule in *Bacillus anthracis*. *Infect. Immun*. 2003;71(7):3914–9. DOI: https://doi.org/10.1128/IAI.71.7.3914-3919.2003
14. Shannon J.G., Ross C.L., Koehler T.M., Rest R. Characterization of anthrolysin O, the *Bacillus anthracis* cholesterol-dependent cytolysin. *Infect. Immun*. 2003;71(6):3183–9. DOI: https://doi.org/10.1128/iai.71.6.3183-3189.2003
15. Tran S.L., Guillemet E., Gohar M., et al. CwpFM (EntFM) is a *Bacillus cereus* potential cell wall peptidase implicated in adhesion, biofilm formation, and virulence. *J. Bacteriol*. 2010; 192(10):2638–42. DOI: https://doi.org/10.1128/jb.01315-09
16. Еременко Е.И., Рязанова А.Г., Цыганкова Е.А. и др. Генотипические особенности штаммов *Bacillus anthracis* с разным проявлением признаков, ассоциированных с патогенностью. *Проблемы особо опасных инфекций*. 2010; (2):53–6. Eremenko E.I., Ryazanova A.G., Tsygankova E.A., et al. Genotypic peculiarities of *Bacillus anthracis* strains with different manifestation of pathogenicity-associated features. *Problems of Particularly Dangerous Infections.* 2010;(2):53–6. EDN: https://elibrary.ru/mstsxj
17. Cao S., Guo A., Wu G., et al. Residue histidine 669 is essential for the catalytic activity of *Bacillus anthracis* lethal factor. *J. Bacteriol*. 2010;192(21):5799–805. DOI: https://doi.org/10.1128/JB.00485-10
18. Klimpel K.R., Arora N., Leppla S.H. Anthrax toxin lethal factor contains a zinc metalloprotease consensus sequence which is required for lethal toxin activity. *Mol. Microbiol*. 1994;13(6):1093–100. DOI: https://doi.org/10.1111/j.1365-2958.1994.tb00500.x
19. Tonello F., Naletto L., Romanello V., et al. Tyrosine-728 and glutamic acid-735 are essential for the metalloproteolytic activity of the lethal factor of *Bacillus anthracis. Biochem. Biophys. Res. Commun*. 2004;313(3):496–502. DOI: https://doi.org/10.1016/j.bbrc.2003.11.134
20. Price L.B., Hugh-Jones M., Jackson P.J., Keim P. Genetic diversity in the protective antigen gene of *Bacillus anthracis. J. Bacteriol*. 1999;181(8):2358–62. DOI: https://doi.org/10.1128/JB.181.8.2358-2362.1999
21. Куличенко А.Н., Еременко Е.И., Рязанова А.Г. и др. Биологические свойства и молекулярно-генетическая характеристика штаммов *Bacillus anthracis*, выделенных во время вспышки сибирской язвы в Ямало-Ненецком автономном округе в 2016 г. *Проблемы особо опасных инфекций*. 2017;(1):94–9. Kulichenko A.N., Eremenko E.I., Ryazanova A.G., et al. Biological properties and molecular-genetic characteristics of *Bacillus anthracis* strains, isolated during the outbreak of anthrax in the Yamalo-Nenets autonomous district in 2016. *Problems of Particularly Dangerous Infections.* 2017;(1):94–9. DOI: https://doi.org/10.21055/0370-1069-2017-1-94-99 EDN: https://elibrary.ru/yixyqj
22. Goncharova Y., Bahtejeva I., Titareva G., et al. Sequence variability of pXO1-located pathogenicity genes of *Bacillus anthracis* natural strains of different geographic origin. *Pathogens*. 2021;10(12):1556. DOI: https://doi.org/10.3390/pathogens10121556
23. Goncharova Y.O., Bogun A.G., Bahtejeva I.V., et al. Allelic polymorphism of anthrax pathogenicity factor genes as a means of estimating microbiological risks associated with climate change. *Appl. Biochem. Microbiol*. 2022;58(4):382–93. DOI: https://doi.org/10.1134/S0003683822040056
24. Hunter P.R., Gaston M.A. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J. Clin. Microbiol.* 1988;26(11):2465–6. DOI: https://doi.org/10.1128/jcm.26.11.2465-2466.1988
25. Еременко Е.И., Рязанова А.Г., Писаренко С.В. и др. Сравнительный анализ методов генетического типирования *Bacillus anthracis. Генетика.* 2019;55(1):40–51. Eremenko E.I., Ryazanova A.G., Pisarenko S.V., et al. Comparative analysis of genotyping methods for *Bacillus anthracis. Russian Journal of Genetics*. 2019;55(1):35–44. DOI: https://doi.org/10.1134/S102279541901006X EDN: https://elibrary.ru/jshcda

26. Еременко Е.И., Рязанова А.Г., Писаренко С.В. и др. Новые генетические маркеры для молекулярного типирования штаммов *Bacillus anthracis. Проблемы особо опасных инфекций*. 2019;(3):43–50. Eremenko E.I., Ryazanova A.G., Pisarenko S.V., et al. New genetic markers for molecular typing of *Bacillus anthracis* strains. *Problems of Particularly Dangerous Infections.* 2019;(3):43–50.
DOI: https://doi.org/10.21055/0370-1069-2019-3-43-50
EDN: https://elibrary.ru/pgefkd
27. Van Ert M.N., Easterday W.R., Huynh L.Y., et al. Global genetic population structure of *Bacillus anthracis. PLoS One*. 2007;2(5):e461.
DOI: https://doi.org/10.1371/journal.pone.0000461
28. Smith K.L., DeVos V., Bryden H., et al. *Bacillus anthracis* diversity in Kruger National Park. *J. Clin. Microbiol*. 2000; 38(10):3780–4.
DOI: https://doi.org/10.1128/JCM.38.10.3780-3784.2000
29. Kassen R., Llewellyn M., Rainey P.B. Ecological constraints on diversification in a model adaptive radiation. *Nature*. 2004;431(7011):984–8.
DOI: https://doi.org/10.1038/nature02923
30. Derzelle S., Aguilar-Bulteta L., Frey J. Whole genome SNP analysis of bovine *B. anthracis* strains from Switzerland reflects strict regional separation of Simmental and Swiss Brown breeds in the past. *Vet. Microbiol*. 2016;196:1–8.
DOI: https://doi.org/10.1016/j.vetmic.2016.10.014

Information about the authors

Evgeny I. Eremenko✉ — D. Sci. (Med.), Prof, Chief Scientist, Anthrax laboratory, Stavropol Plague Control Research Institute, Stavropol, Russia, ejer@mail.ru, https://orcid.org/0000-0002-1117-1185.

Grigorii A. Pechkovskii — junior researcher, Anthrax laboratory, Stavropol Plague Control Research Institute, Stavropol, Russia, https://orcid.org/0000-0001-7033-9972

Alla G. Ryazanova — Cand. Sci. (Med.), Head, Anthrax laboratory, Stavropol Plague Control Research Institute, Stavropol, Russia, https://orcid.org/0000-0002-5196-784X

Sergey V. Pisarenko — Cand. Sci. (Chem.), leading researcher, Laboratory of biochemistry, Stavropol Plague Control Research Institute, Stavropol, Russia, https://orcid.org/0000-0001-6458-6790

Dmitry A. Kovalev — Cand. Sci. (Chem.), Head, Laboratory of biochemistry, Stavropol Plague Control Research Institute, Stavropol, Russia, https://orcid.org/0000-0002-9366-5647

Lyudmila Yu. Aksenova — Cand. Sci. (Med.), senior researcher, Anthrax laboratory, Stavropol Plague Control Research Institute, Stavropol, Russia, https://orcid.org/0000-0002-5196-784X

Ol'ga V. Semenova — Cand. Sci. (Biol.), researcher, Anthrax laboratory, Stavropol Plague Control Research Institute, Stavropol, Russia, https://orcid.org/0000-0003-0274-898X

Alexander N. Kulichenko — D. Sci. (Med.), Prof., Academician of RAS, Director, Stavropol Plague Control Research Institute, Stavropol, Russia, https://orcid.org/0000-0002-9362-3949

Author contribution. All authors made a substantial contribution to the conception of the work, acquisition, analysis, interpretation of data for the work, drafting and revising the work, final approval of the version to be published.

The article was submitted 21.03.2023;
accepted for publication 29.05.2023;
published 28.06.2023

Информация об авторах

Еременко Евгений Иванович✉ — д.м.н., профессор, г.н.с. лаб. сибирской язвы Ставропольского противочумного института, Ставрополь, Россия, ejer@mail.ru,
https://orcid.org/0000-0002-1117-1185

Печковский Григорий Александрович — м.н.с.. лаб. сибирской язвы Ставропольского противочумного института, Ставрополь, Россия, https://orcid.org/0000-0001-7033-9972

Рязанова Алла Геннадиевна — к.м.н., зав. лаб. сибирской язвы Ставропольского противочумного института, Ставрополь, Россия, https://orcid.org/0000-0002-5196-784X

Писаренко Сергей Владимирович — к.х.н., в.н.с. лаб. биохимии Ставропольского противочумного института, Ставрополь, Россия, https://orcid.org/0000-0001-6458-6790

Ковалев Дмитрий Анатольевич — к.х.н., зав. лаб. биохимии Ставропольского противочумного института, Ставрополь, Россия, https://orcid.org/0000-0002-9366-5647

Аксенова Людмила Юрьевна — к.м.н., с.н.с. лаб. сибирской язвы Ставропольского противочумного института, Ставрополь, Россия, https://orcid.org/0000-0002-7744-3112

Семенова Ольга Викторовна — к.б.н., н.с. лаб. сибирской язвы Ставропольского противочумного института, Ставрополь, Россия, https://orcid.org/0000-0003-0274-898X

Куличенко Александр Николаевич — академик РАН, д.м.н., профессор, директор Ставропольского противочумного института, Ставрополь, Россия, https://orcid.org/0000-0002-9362-3949

Участие авторов. Все авторы внесли существенный вклад в проведение поисково-аналитической работы и подготовку статьи, прочли и одобрили финальную версию до публикации.

Статья поступила в редакцию 21.03.2023;
принята к публикации 29.05.2023;
опубликована 28.06.2023