



## ORIGINAL STUDY ARTICLE

DOI: <https://doi.org/10.36233/0507-4088-357>

© SKOROBAGATKO D.A., TOLKOVA E.S., SHEPELEVA O.A., SHAPOVALOV S.O., 2026



# African swine fever virus (*Asfarviridae*, *Asfivirus*) strains from the central regions of Russia, carrying variant 5 of the central variable region (CVR), are characterized by tandem duplication in the intergenic region *MGF 360-13L – MGF 360-14L*

Daria A. Skorobagatko<sup>✉</sup>, Ekaterina S. Tolkova, Olga A. Shepeleva, Sergey O. Shapovalov

Cherkizovo Research and Testing Center LLC, 108805, Moscow, Russia

**Abstract**

**Introduction.** The high economic losses and the lack of effective and safe vaccines against African swine fever (ASF) indicate the need for further in-depth studies of the virus genome, its changes and the circulation of genetic lineages. Whole genome sequencing of virus isolates is best suited for this purpose.

**The aim of the study.** Whole-genome analysis of African swine fever virus (ASFV) isolates obtained in the Lipetsk, Penza and Tambov regions in 2016–2021 and identification of additional diversity markers within the genetic lineage.

**Materials and methods.** Domestic pig tissue samples were analyzed using whole-genome sequencing and Sanger sequencing. The following programs were used for sequence assembly: CLC Genomics Workbench 22, Trimmomatic v. 0.39, SPAdes v. 4.2.0, BWA-MEM v. 0.7.17-r1188 and bcftools v.1.22. The phylogenetic tree was constructed in MEGA11 based on the alignment in MAFFT v. 7.526 with 67 genomes from GenBank.

**Results.** Based on the presence of MGF 360-10L III polymorphism, the analyzed isolates belong to the CVR-V variant of the Russia genetic lineage of the Georgia 2007 clade. Based on the order of formation of MGF 360-10L III and CVR-V, any sequences carrying CVR-V belong to the same genetic lineage. A 12-nucleotide insertion CAGTCTATAAGA was detected, forming a tandem duplication in *IGR MGF 360-13L – MGF 360-14L*, and polymorphisms in *IGR C62L – C962R* and in genes *D1133L* and *Q706L* were proposed as having phylogenetic potential for differentiation ASFV strains in the central regions of Russia.

**Conclusion.** The proposed new potential diversity markers have a resolving power for ASFV strains from Central Russia.

**Keywords:** African swine fever virus; whole genome sequencing; central variable region; Central Russia

**For citation:** Skorobagatko D.A., Tolkova E.S., Shepeleva O.A., Shapovalov S.O. African swine fever virus (*Asfarviridae*, *Asfivirus*) strains from the central regions of Russia, carrying variant 5 of the central variable region (CVR), are characterized by tandem duplication in the intergenic region *MGF 360-13L – MGF 360-14L*. *Problems of Virology (Voprosy Virusologii)*. 2026; 71 (1): 42–52. DOI: <https://doi.org/10.36233/0507-4088-357>  
EDN: <https://elibrary.ru/efobwo>

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

**Acknowledgments.** The authors are grateful to the scientists working at the Russian State Agrarian University – Moscow Timiryazev Agricultural Academy for permission to use the DNBSEQ-G50 platform for this study.

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

**Ethics approval.** Authors confirm compliance with institutional and national standards for the use of laboratory animals in accordance with “Consensus author guidelines for animal use” (IAVES 23 July 2010). The research protocol was approved by the Ethics Committee of the institution Cherkizovo Research and Testing Center LLC (Protocol No. 1 dated 04 March 2025).

## Штаммы вируса африканской чумы свиней (*Asfarviridae*, *Asfivirus*, *African swine fever virus*) центральных регионов России, несущие вариант 5 центральной вариабельной области (CVR), характеризуются тандемной дупликацией в межгенном регионе *MGF 360-13L – MGF 360-14L*

Скоробагатько Д.А.✉, Толькова Е.С., Шепелева О.А., Шаповалов С.О.

ООО Научно-испытательный центр «Черкизово», 108805, г. Москва, Россия

### Резюме

**Введение.** Большой экономический ущерб и отсутствие действенных и безопасных вакцин против африканской чумы (АЧС) указывают на необходимость продолжения углубленного изучения генома вируса АЧС, его изменений и циркуляции генетических линий. Для этого наилучшим образом подходит полногеномное секвенирование изолятов вируса.

**Цель работы.** Полногеномный анализ изолятов вируса АЧС, полученных на территории Липецкой, Пензенской и Тамбовской областей в 2016–2021 гг., и выявление дополнительных маркеров разнообразия в пределах генетической линии.

**Материалы и методы.** Образцы тканей домашних свиней исследовали с помощью полногеномного секвенирования и секвенирования по Сэнгеру. Для сборки последовательностей использовали программы CLC Genomics Workbench 22, Trimmomatic v. 0.39, SPAdes v. 4.2.0, BWA-MEM v. 0.7.17-r1188 и bcftools v. 1.22. Филогенетическое древо строили в программе MEGA11 на основе выравнивания в MAFFT v. 7.526 с 67 геномами из GenBank.

**Результаты.** На основании присутствия полиморфизма *MGF 360-10L III* проанализированные изоляты относились к варианту CVR-V генетической линии «Россия» клады Georgia 2007. Исходя из очередности образования *MGF 360-10L III* и CVR-V, любые последовательности, несущие CVR-V, принадлежат к этой же генетической линии. Была обнаружена инсерция в 12 нуклеотидов CAGTCTATAAGA, образующая тандемную дупликацию в *IGR MGF 360-13L – MGF 360-14L*, и предложены полиморфизмы в *IGR C62L – C962R* и в генах *D1133L* и *Q706L*, имеющие филогенетический потенциал для дифференцировки штаммов вируса АЧС в центральных регионах России.

**Заключение.** Предложенные к использованию новые вероятные маркеры разнообразия обладают разрешающей способностью в отношении штаммов вируса АЧС из Центральной России.

**Ключевые слова:** вирус африканской чумы свиней; полногеномное секвенирование; центральный вариабельный регион; Центральная Россия

**Для цитирования:** Скоробагатько Д.А., Толькова Е.С., Шепелева О.А., Шаповалов С.О. Штаммы вируса африканской чумы свиней (*Asfarviridae*, *Asfivirus*, *African swine fever virus*) центральных регионов России, несущие вариант 5 центральной вариабельной области (CVR), характеризуются тандемной дупликацией в межгенном регионе *MGF 360-13L – MGF 360-14L*. *Вопросы вирусологии*. 2026; 71 (1): 42–52. DOI: <https://doi.org/10.36233/0507-4088-357> EDN: <https://elibrary.ru/efobwo>

**Финансирование.** Авторы заявляют об отсутствии внешнего финансирования исследования.

**Благодарности.** Авторы благодарны ученым, работающим в Российском государственном аграрном университете – Московской сельскохозяйственной академии имени К.А. Тимирязева, за разрешение использовать платформу DNBSEQ-G50 для данного исследования.

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

**Этическое утверждение.** Авторы подтверждают соблюдение институциональных и национальных стандартов по использованию лабораторных животных в соответствии с “Consensus author guidelines for animal use” (IAVES 23 July 2010). Протокол исследования одобрен Этическим комитетом организации Научно-испытательный Центр “Черкизово” (Протокол № 1 от 04.03.2025).

## Introduction

African swine fever (ASF) is a contagious viral disease of pigs and wild boars that has spread widely throughout the world since 2007, causing significant economic damage. The lack of a reliable vaccine, as well as the identified trend towards a decrease in virulence and an increase in the infectivity of the ASF virus (*Asfarviridae*, *Asfivirus*, *African swine fever virus*) [1] make it difficult to control outbreaks of the disease. This points to the necessity for further in-depth study of the spatial and temporal circulation of the virus, changes in its genome, and the characteristics of genetic lines. In this regard, it seems important not only to classify the samples obtained according to known schemes, but also to perform their whole-genome analysis for the possible detection of previously undescribed regions that have potential phylogenetic significance.

To classify ASF virus strains of genotype II circulating in Eurasia, subgenotyping schemes based on a set of marker sites are currently used [2–4]. The central variable region (CVR) in the *B602L* gene is one such phylogenetic marker with regional resolution [2–5]. Since 2013, three CVR variants have been registered in the central regions of Russia: CVR-I, CVR-V, and CVR-VI [4, 6], but there are discrepancies between the frequency of occurrence of different CVR variants and the proportion of described whole-genome sequences carrying these variants. Since 2016, the frequency of detection of the CVR-V single nucleotide polymorphism (SNP) has increased [4, 6], but so far only one complete genome containing this SNP has been described [7]. It is of interest to analyze a larger number of whole-genome sequences carrying the CVR-V polymorphism to better understand the spread of the virus, its variability, and to identify additional markers of diversity.

**The aim** of the study is to perform whole-genome sequencing, subgenotyping, and bioinformatics analysis of previously unstudied ASF virus isolates obtained in the Lipetsk, Penza, and Tambov regions in 2016–2021, and to identify additional markers of diversity within the genetic lineage.

## Materials and methods

The study analyzed ASF virus sequences from samples obtained during three ASF outbreaks in the Russian Federation in the Lipetsk (2016), Penza (2019), and Tambov (2021) regions. The obtained genomes were published in GenBank under numbers PX488473–PX488476.

**DNA isolation.** DNA was isolated from tissue samples of domestic pigs from pig farms. The presence of ASF virus in the samples had previously been confirmed by polymerase chain reaction (PCR) (ASF test system form 4, AmpliSens, Russia).

Viral DNA enrichment was not performed. The tissue samples were homogenized, then a 10% suspension was prepared with 800  $\mu$ L of sterile saline. The samples were centrifuged at 200g for 1 min and the supernatant was used for extraction using the HiPure Blood DNA mini kit (Magen, China). DNA quality was confirmed by measurement with a nanophotometer (Implen); the A260/

A280 ratio was in the range of 1.65–1.90. The concentration of the isolated DNA was measured using a Qubit fluorometer (Applied Biosystems, Thermo Fisher Scientific, USA) and the HS QuDye dsDNA assay kit (Lumiprobe, Russia). It was 8–10 ng/ $\mu$ L.

**Sanger sequencing.** Prior to whole-genome sequencing, several regions (the C-terminal region of the p72 protein, the central variable region in the *B602L* (CVR) gene, and the region located between the *I73R* and *I329L* genes) were sequenced using Sanger sequencing with an AB3500 genetic analyzer. For this purpose, total genomic DNA was extracted using the MagnoSorb kit (AmpliSens, Russia) on the KingFisher Flex system (Thermo Fisher Scientific, USA). DreamTaq PCR Master Mix (2X) (Thermo Fisher Scientific, USA) was used to amplify each primer pair. The ExoSAP-IT kit (Thermo Fisher Scientific, USA) was used to purify PCR reaction products, the BigDye Terminator v. 1.1 kit (Thermo Fisher Scientific, USA) was used for PCR analysis, and the BigDye XTerminator kit (Thermo Fisher Scientific, USA) was used for post-PCR purification.

**DNA library preparation and high-throughput sequencing.** For sample preparation, 200 ng of DNA was used with the MGIEasy FS DNA library preparation kit (MGI Tech, China) according to the manufacturer's instructions. The resulting DNA library was sequenced using the DNBSEQ-G50RS high-throughput sequencing kit (FCL PE100, China) on a DNBSEQ-G50 instrument according to the manufacturer's instructions.

**Data analysis.** Sanger sequencing data were analyzed using CLC Main Workbench 20 (Qiagen, Netherlands).

Whole-genome sequencing data were processed to remove adapters and low-quality reads. The processed pairs were then aligned to the *Sus scrofa* reference genome (GenBank accession number GCA\_000003025.6) to remove host data using CLC Genomics Workbench 22 (Qiagen, Netherlands).

Unmatched reads were collected and filtered for quality using Trimmomatic v. 0.39 [8] to remove or trim the ends of low-quality reads. After that, read errors were corrected in SPAdes v. 4.2.0 [9]. Based on the corrected reads, genomes were assembled by aligning short reads to the Georgia 2007/1 reference sequence using the BWA-MEM v. 0.7.17-r1188 algorithm on the UGENE v. 52.1 platform [10]. To search for variations and create a consensus using the BWA-MEM method, the assembly was analyzed using the bcftools v. 1.22 package [11], after which the resulting consensus sequence was manually verified. To determine open reading frames (ORFs) using GATU software [12], the Georgia 2007/1 reference strain was used.

Multiple alignment of the obtained virus sequences with 67 genomes downloaded from GenBank (See **Appendix 1** in the supplementary files of the article on the journal site: <https://doi.org/10.36233/0507-4088-357-1>) was performed using MAFFT v. 7.526 [13]. Phylogenetic analysis of the isolates was performed using Mega 11 [14] according to the recommended Tamura (T92) model with optional gamma distribution (G) parameters and consideration of invariant sites (I) selected based on the calculation of the Bayesian coefficient (BIC = 524,795.816) and

the adjusted Akaike coefficient (AICc = 523,469.554). Phylogenetic trees were constructed using the Maximum Likelihood (ML) method with Bootstrap support of 1000 repetitions.

Molecular-epizootological maps were created in Evergis Online Map.

Authors confirm compliance with institutional and national standards for the use of laboratory animals in accordance with “Consensus author guidelines for animal use” (IAVES 23 July 2010). The research protocol was approved by the Ethics Committee of the institution Cherkizovo Research and Testing Center LLC (Protocol No. 1 dated 04 March 2025).

## Results

As expected, preliminary Sanger sequencing of all isolates identified genotype II according to the generally accepted classification.

All isolates had a glutamic acid substitution for lysine at position 201 of the B602L protein and were therefore classified as CVR-V [6]. Also, all *IGR 173R/1329L* sequences obtained by Sanger method included a tandem repeat corresponding to variant II among the known IGR variants [15].

The lengths of the obtained genomes were 190,596 bp for ASFV\_Lipezk-2016, 190,589 for ASFV\_Penza-2019a, 190,596 for ASFV\_Penza-2019b, and 190,592 for ASFV\_Tambov-2021.

Cluster analysis was performed according to the supplemented scheme for subgenotyping ASF virus genotype II (modification of the subgenotype classification by C. Gallardo et al., 2023)<sup>1</sup> [2]. The presence of a specific SNP at position 27009 (MGF 360-10L III polymorphism) allowed the sequences to be assigned to the Georgia-2007 clade of the “Russia” genetic lineage according to the above scheme. To date, this combination of MGF 360-10L III and CVR-V has been found only once, in the ASFV/Ulyanovsk/19WB-5699 isolate [7]. The other known cases of CVR-V variant detection were not based on whole-genome sequencing, but on Sanger sequencing of CVR regions [4, 6].

Based on 67 complete ASF virus sequences downloaded from GenBank and the four genomes described above, a phylogenetic analysis was performed, the results of which are shown in **Fig. 1**. The MAD/01/1998 genotype strain isolated in Madagascar in 1998 was used as an outgroup. The results classify the isolates under consideration as belonging to the Georgia 2007 clade, which is characterized by close relatedness to the Georgia 2007/1 strain and relatively early isolation [16]. The highest homology was observed with the ASFV/Ulyanovsk/19WB-5699 isolate. The data obtained confirm the results of subgenotyping.

Three insertions, one reading frame shift, and 22 SNPs relative to the reference sequence were found in the genomes described, of which five SNPs and two insertions are unique (**Table**). The polyC, polyG, and variation re-

gions at the ends of the genome are considered unreliable and were therefore not taken into account.

The greatest similarity of all isolates studied is observed with ASFV/Ulyanovsk/19WB-5699. These are 7 synonymous substitutions at positions 27009, 39215, 62387, 63729, 69509, 120769, and 157890, a deletion at position 103309 in the intergenic region *B602L – B385R*, as well as 4 non-synonymous SNPs leading to the replacement of histidine with tyrosine at position 465 of *MGF 505-9R*, glutamic acid to lysine at position 201 of the *B602L* gene, arginine to histidine at position 186 of the *H240R* gene, and lysine to glutamic acid in the *I9R* gene.

The close relationship between ASFV/Ulyanovsk/19WB-5699 isolates from Penza and Tambov is also confirmed by transitions at position 90001 in *IGR C62L – C962R*, proline to serine substitution at position 457 of the *D113L* gene, and alanine to valine substitution at position 607 of the *Q706L* gene.

In this study, we discovered for the first time an insertion of 12 nucleotides CAGTCTATAAGA in *IGR MGF 360-13L – MGF 360-14L* in all sequenced samples, resulting in a tandem repeat (**Fig. 2**). Sanger sequencing of this region confirmed the presence of the insertion. Furthermore, based on multiple alignment data and the fact that tandem repeats are poorly mapped when assembled using standard methods, a similar repeat may be present instead of an adenine insertion at the same position in the ASFV/Ulyanovsk/19WB-5699 isolate.

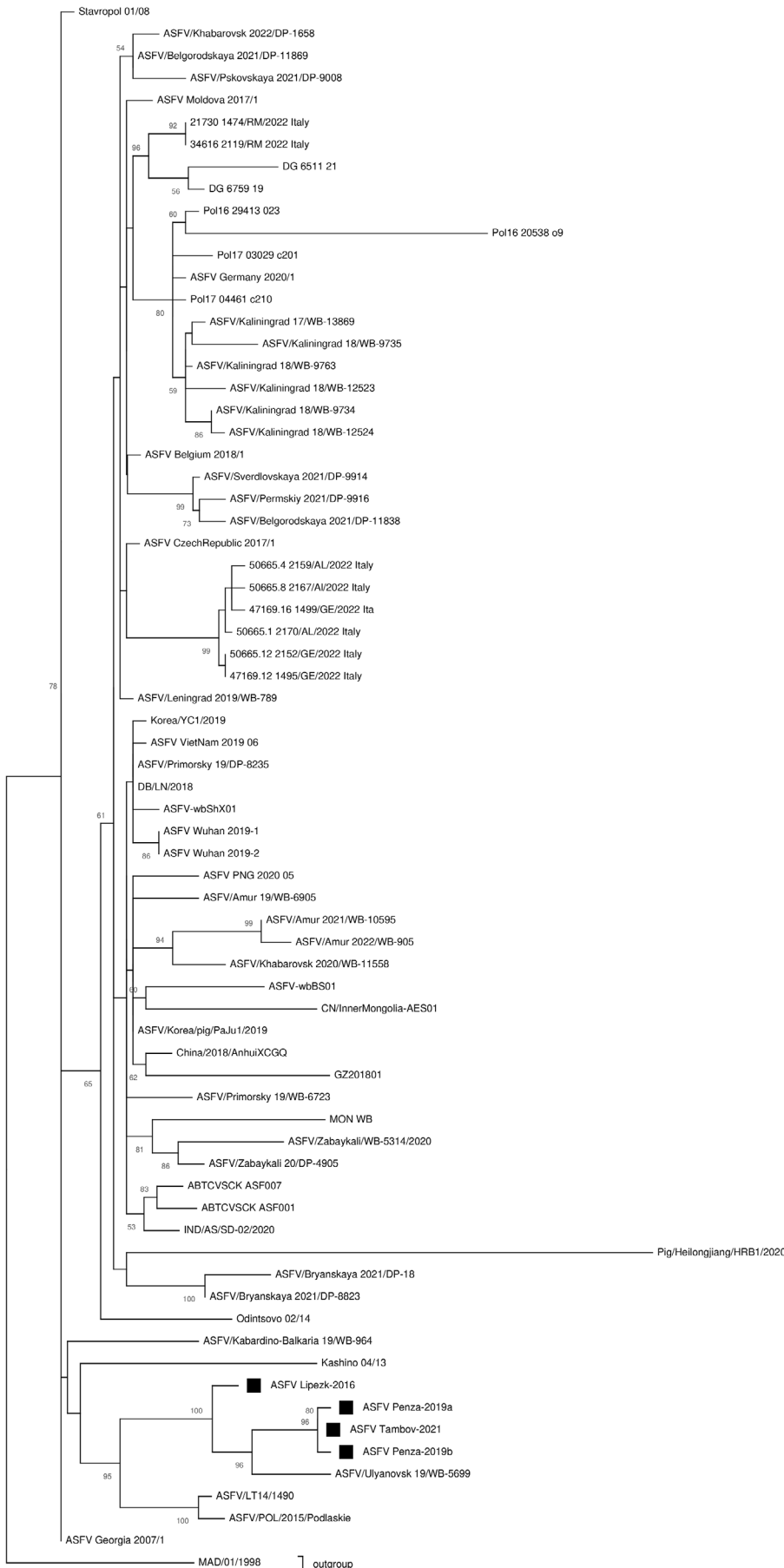
Unique polymorphisms at positions 2153 and 2161 in the ASFV\_Penza-2019a, b and ASFV\_Tambov-2021 samples lead to the replacement of cysteine with glycine at position 165 and methionine with threonine and serine at position 162 of the predicted MGF 360-1La protein, respectively. Synonymous SNPs at positions 15320 and 41616, as well as the replacement of glutamine with histidine at position 22 of MGF 360-11L, are also described for the first time for the outbreaks in Penza and Tambov.

## Discussion

According to the methodological recommendations developed by the Federal Center for Animal Health (ARRIAH) (Russia)<sup>1</sup>, the Georgia 2007 clade of the “Russia” genetic lineage includes those variants of the virus in which SNP A>G is present at position 27009 (the *MGF 360-10L* gene was proposed as a possible marker in [7], the Georgia 2007 clade was defined in [16]). These are isolates Kashino 04/13, ASFV/POL/2015/Podlaskie, ASFV/LT14/1490, ASFV/Ulyanovsk/19WB-5699, and those described by us, ASFV\_Lipezk-2016, ASFV\_Penza-2019a, ASFV\_Penza-2019b, and ASFV\_Tambov-2021 (**Fig. 3**).

Within this lineage, several subgroups can be distinguished based on the CVR sequence. Since 2013, three CVR variants have been described in the central regions: CVR-I, CVR-V, and CVR-VI [4, 6]. CVR-I (identical to the reference sequence) is the most common. CVR-VI

<sup>1</sup>Chernyshev R.S., Sprygin A.S., Lavrentyev I.A., Igolkin A.S. Methodological recommendations for subgenotyping of ASF virus isolates of genotype II into topotypes and genetic lines. Vladimir; 2025, 14 p. (In Russ.).



**Fig. 1.** Phylogenetic tree constructed based on 71 complete genome sequences of ASFV, genotype II.

The isolates covered by this research are labeled with ■.

**Рис. 1.** Филогенетическое древо, построенное на основании 71 полногеномной последовательности вируса АЧС, II генотип.

Изоляты, полученные в данном исследовании, отмечены ■.

**Table.** Variations in the complete genome sequences of the ASF virus of the studied isolates

**Таблица.** Вариации полногеномных последовательностей вируса АЧС исследованных изолятов

Position in Georgia-2007/1 Позиция в Georgia-2007/1	Gene / intergenic region Ген / межгенная область	Georgia 2007/1	Lipezk-2016	Penza-2019a	Penza-2019b	Tambov-2021	Changes in amino acids Изменения в аминокислотах
2153*	<i>MGF 360-1La</i>	A	A	C	C	C	C > G
2161*	<i>MGF 360-1La</i>	A	A	GTAC	GTAC	GTAC	M > ST
12568	<i>G ACD 00190</i>	A	–	–	–	–	Reading frame shift Сдвиг рамки считывания
15320*	<i>MGF 110-13La</i>	A	A	G	G	G	–
27009	<i>MGF 360-10L</i>	A	G	G	G	G	–
28434*	<i>MGF 360-11L</i>	T	T	A	A	A	Q > H
32767–32768*	<i>IGR MGF 360-13L – MGF 360-14L</i>	–	CAGTCTATAAGA	CAGTCTATA-AGA	CAGTCTATA-AGA	CAGTCTATA-AGA	–
39215	<i>MGF 505-5R</i>	G	A	A	A	A	–
41616*	<i>IGR MGF 505-6R – MGF 505-7R</i>	T	T	A	A	A	–
45002	<i>MGF 505-9R</i>	C	T	T	T	T	H > Y
62387	<i>F1055L</i>	G	A	A	A	A	–
63081*	<i>F1055L</i>	G	G	A	G	G	–
63729	<i>F1055L</i>	A	G	G	G	G	–
69509	<i>EPI242L</i>	G	A	A	A	A	–
90001	<i>IGR C62L – C962R</i>	T	T	C	C	C	–
98321*	<i>B438L</i>	T	C	T	T	T	–
102668	<i>B602L</i>	C	T	T	T	T	E > K
103309	<i>IGR B602L – B385R</i>	G	–	–	–	–	–
120769	<i>CP2475L</i>	C	T	T	T	T	–
143368	<i>D1133L</i>	G	G	A	A	A	P > S
156891	<i>H240R</i>	G	A	A	A	A	R > H
157890	<i>R298L</i>	C	T	T	T	T	–
158225	<i>Q706L</i>	G	G	A	A	A	A > V
167129	<i>E199L</i>	T	G	T	T	T	Q > H
167188	<i>E199L</i>	C	C	C	G	C	A > P
173408–173409	<i>IGR I73R – I329L</i>	–	GAATATATAG	GAATATATAG	GAATATATAG	GAATATATAG	–
182971	<i>I9R</i>	A	G	G	G	G	K > E

**Note.** \* – the positions of previously undescribed polymorphisms are marked.

**Примечание.** \* – позиции ранее не описанных полиморфизмов.

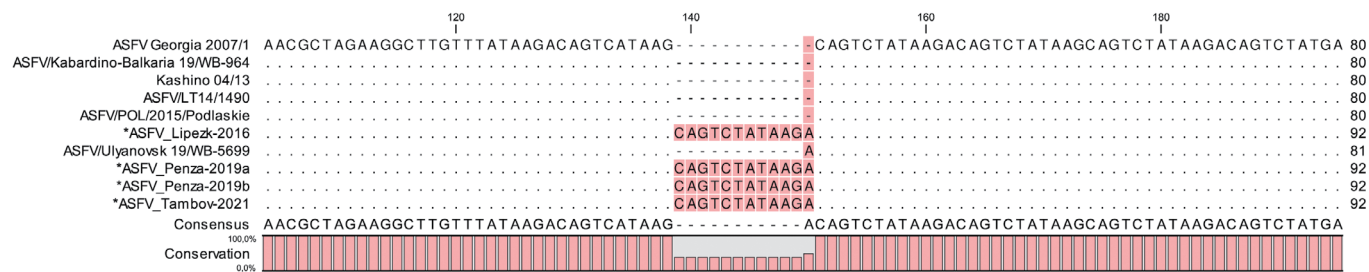
polymorphism (A>T substitution at position 459 of the *B602L* gene) was described for isolates from the Tver, Moscow, and Vladimir regions from 2013 to 2016 and is currently not detected. Although the prevalence of the CVR-V variant (C>T at position 601 of the *B602L* gene), described only for the central regions of Russia, has increased significantly over time [4], only one complete genome containing CVR-V and CVR-VI has been described so far: ASFV/Ulyanovsk/19WB-5699 and Kashino 04/13, respectively.

The ASFV/POL/2015/Подлaskie and ASFV/LT14/1490 strains, which are closely related and were isolated in Eastern Europe, have the CVR-I variant. However, they do not cluster with other strains with CVR-I circulating in Russia since 2013 and only carry the original CVR variant.

The other two subgroups identified in central Russia differ in terms of CVR-V and CVR-VI variants, which are not

found among representatives of other genetic lines. It should be noted that the authors who described these polymorphisms and analyzed their distribution in Russia worked mainly with partial sequences of the *B602L* gene rather than complete genomes, describing 6 isolates containing CVR-VI and 26 isolates containing CVR-V [4, 6].

CVR-V, circulating in central Russia, the Volga region, European South, and part of the Urals between 2016 and 2022 [4, 6], has a non-synonymous substitution of glutamic acid for lysine at position 201 of the *B602L* protein. This variant was present in the ASFV/Ulyanovsk/19WB-5699 strain [7] and in all the sequences we isolated. The appearance of CVR-V and CVR-VI variants in isolates of the “Russia” lineage after the formation of MGF 360-10L-III, which characterizes this lineage, suggests that these variants were formed within this lineage and that all sequences with these polymorphisms belong to it.

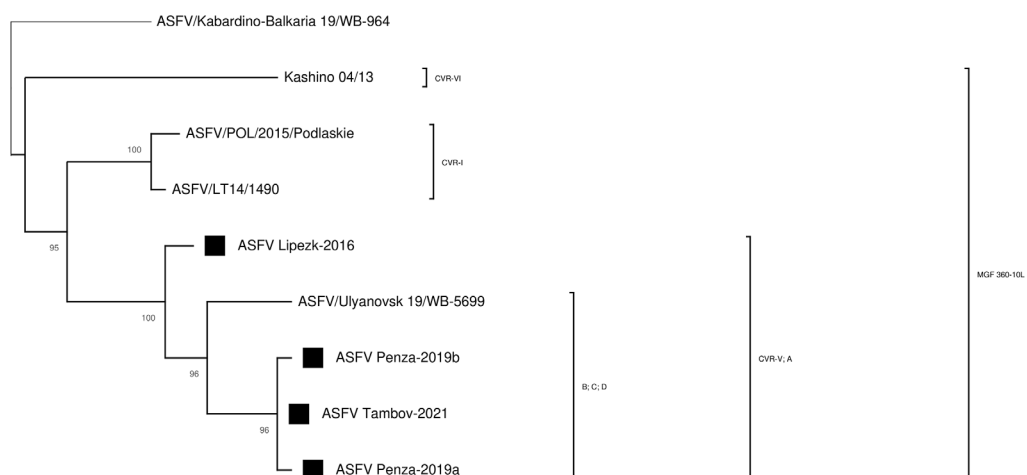


**Fig. 2.** Insertion in *IGR MGF 360-13L – MGF 360-14L*. Clade Georgia 2007.

The isolates covered by this research are labeled with \*.

**Рис. 2.** Инсерция в *IGR MGF 360-13L – MGF 360-14L*. Клада Georgia 2007.

Изоляты, полученные в данном исследовании, отмечены \*.



**Fig. 3.** Markers that allow grouping of isolates within the Russia lineage of the Georgia 2007 clade.

Sequences united by the presence of SNPs MGF 360-10L III; CVR-I, CVR-V and CVR-VI are shown; A – CAGTCTATAAGA insertion in *IGR MGF 360-13L – MGF 360-14L*; B – T>C substitution at position 90001 (*IGR C62L – C962R*); C – G>A transition at position 143368 (*D1133L*); D – G>A transition at position 158225 (*Q706L*). The isolates covered by this research are labeled with ■.

**Рис. 3.** Маркеры, группирующие изоляты внутри линии «Россия» клды Georgia 2007.

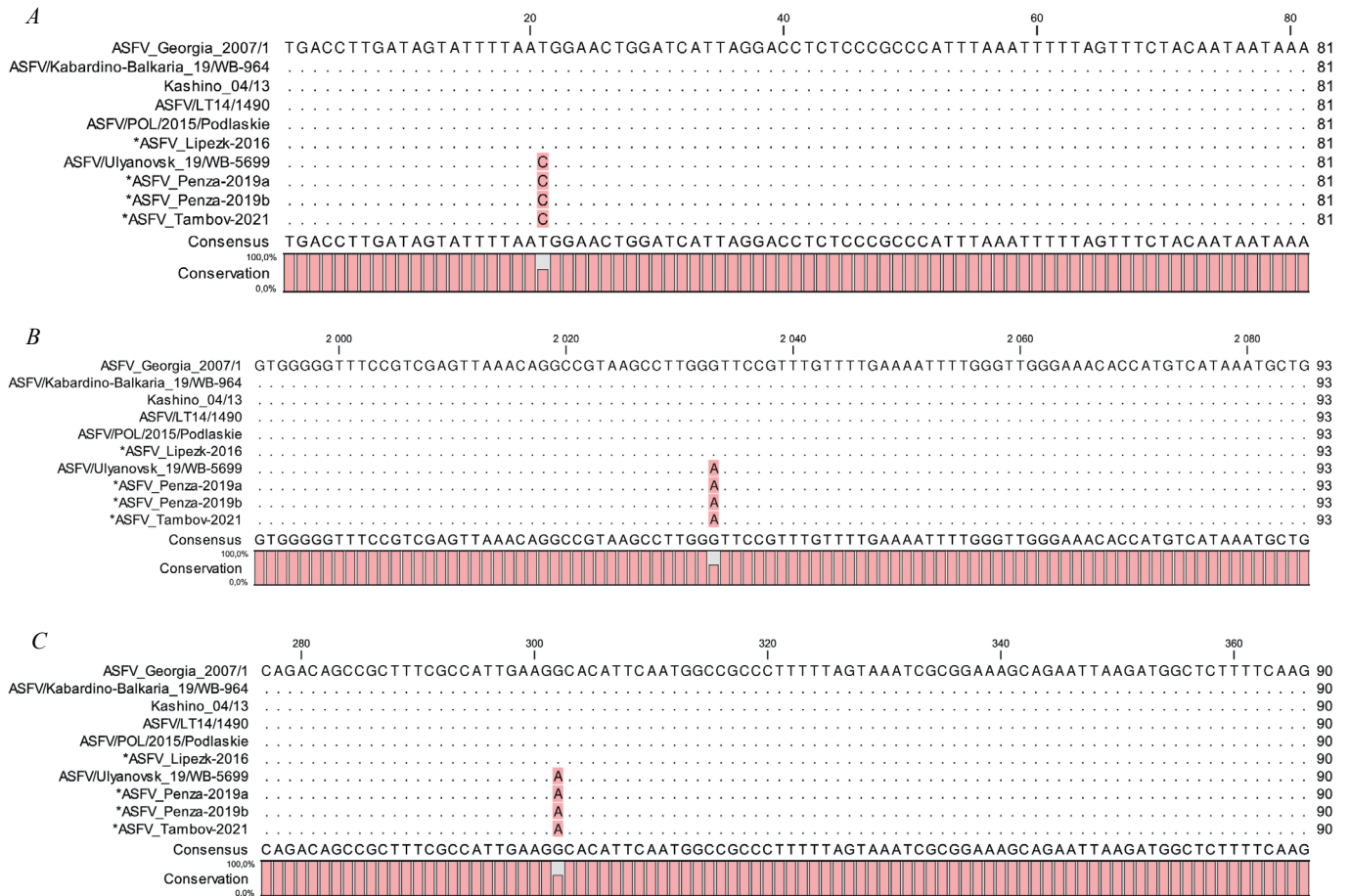
Показаны последовательности, объединенные наличием ОНП MGF 360-10L III; CVR-I, CVR-V и CVR-VI; A – инсерция CAGTCTATAAGA в *IGR MGF 360-13L – MGF 360-14L*; B – замена T>C в позиции 90001 (*IGR C62L – C962R*); C – транзиция G>A в позиции 143368 (*D1133L*); D – транзиция G>A в позиции 158225 (*Q706L*). Изоляты, полученные в данном исследовании, отмечены ■.

To differentiate genomes with the CVR-V variant circulating in central regions of Russia, four sites with phylogenetic potential are proposed. The grouping of isolates using the proposed markers is shown in Fig. 3. The first site is an additional copy of the CAGTCTATAAGA sequence in *IGR MGF 360-13L – MGF 360-14L*, which we found and which leads to the formation of a tandem duplication. In the reference genome, there are four more similar sequences 9–12 nucleotides long on both sides of the duplicated region, one of which is identical to the insertion (Fig. 2). These regions probably arose as a result of duplication events and then changed independently. The ASF virus is characterized by variation in the number of tandem repeats both within genes and in non-coding regions [17]. In [18], a mechanism of chain slippage during replication, known for bacterial genomes [19], was proposed for the formation of indels in poxviruses. In this case, even if the repeat is not completely identical, the template and emerging strands shift, and duplication occurs. As a result, there are

at least three copies on the new strand. On the other hand, there is an opinion that indels are significantly determined by homologous recombination events [20].

The CAGTCTATAAGA duplication is present (ASFV\_Lipezk-2016, ASFV\_Penza-2019a, ASFV\_Penza-2019b, and ASFV\_Tambov-2021), or is presumed to be present (ASFV/Ulyanovsk/19/WB-5699) in the analyzed strains of the Russia lineage of the CVR-V subgroup. In our opinion, this insertion may serve as a potential marker of diversity for a more detailed classification of ASF virus within the Russia genetic lineage. Sequencing of additional regions or whole-genome sequencing of previously selected and new ASF isolates could provide greater clarity on the extent of this insertion.

The following three sections are common transitions for four isolates: ASFV/Ulyanovsk/19/WB-5699, ASFV\_Penza-2019a, ASFV\_Penza-2019b, and ASFV\_Tambov-2021, which are linked by the geographical proximity of their isolation sites in central Russia and a



**Fig. 4.** Polymorphisms that allow grouping of isolates in the Georgia 2007 clade sequences.

A – *IGR C62L – C962R*; B – partial sequence of the *D1133L* gene; C – partial sequence of the *Q706L* gene. The isolates covered by this research are labeled with –\*.

**Рис. 4.** Полиморфизмы, группирующие изоляты в последовательностях клады Georgia 2007.

A – *IGR C62L – C962R*; B – частичная последовательность гена *D1133L*; C – частичная последовательность гена *Q706L*. Изоляты, полученные в данном исследовании, отмечены\*.

short time interval. These SNPs may be of phylogenetic significance and allow differentiation of sequences obtained since 2019 in central regions from earlier isolates of this lineage (Fig. 4).

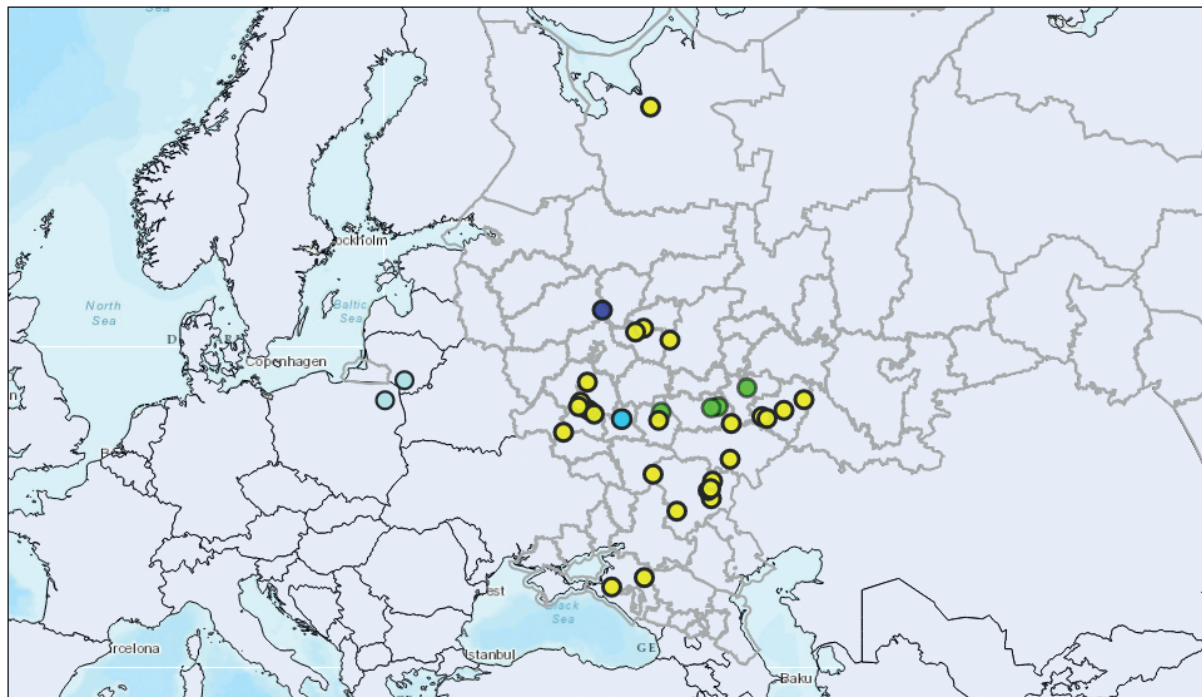
The first of these is the T>C transition in *IGR C62L – C962R* (Fig. 4 A). The substitution is expected to be fairly stable and suitable as a marker site, as intergenic regions do not experience significant selection pressure.

The second is the replacement of proline with serine at position 457 of the D1133L protein, caused by a G>A transition (Fig. 4 B). It is not strictly conservative, since although the charge of the polar groups of serine in a neutral environment is compensated and equal to that of neutral proline, these amino acids have different hydrophobicity, which may slightly affect the protein structure. However, the detection of this mutation in several isolates suggests that it does not critically affect the normal function of D1133L. D1133L is an intermediate-late protein with nuclear and cytoplasmic localization during infection [21]. It has helicase activity and may also be involved in RNA metabolism and ribosomal function [22]. When D1133L is overexpressed, ASF virions replicate more efficiently in cells [23]. *In vitro*, it interacts with several

host proteins—VIM and TRIM21, which enhance virus replication, and the translation elongation factor TUFM, which inhibits ASF virus replication in cell culture [22]. The polymorphism found may be a probable marker, but due to the possible influence of selection, it is not as reliable as a conservative replacement.

The third substitution, G>A transition, results in a fairly conservative substitution of alanine for valine at position 607 of the Q706L protein (Fig. 4 C), which should not lead to functional changes and should not be subject to increased selection pressure. This protein is an RNA helicase and is quite conservative, although it has a small variability between ASF virus genotypes from different regions [24]. Another SNP in this gene, characterizing isolates from the Bryansk region, is reported in [16]. The SNP we propose can also be considered a marker for the central regions of Russia since 2019.

The marker sites proposed for use allow differentiation of ASF virus strains from Central Russia carrying the CVR-V polymorphism, as well as strains that appeared after 2019, from earlier variants. The question of the prevalence of the new insertion in *IGR MGF 360-13L – MGF 360-14L* in relation to the occurrence of CVR-V



**Fig. 5.** Distribution of isolates of the Russia lineage.

● – variant CVR-I; ● – variant CVR-VI; ● – variant CVR-V, contains the CAGTCTATAAGA insertion in *IGR MGF 360-13L – MGF 360-14L*; ● – variant CVR-V, contains an insertion in *IGR MGF 360-13L – MGF 360-14L*, as well as transitions in *IGR C62L – C962R* and in genes *D1133L* and *Q706L*; ● – isolates carrying the CVR-V variant, presumably belonging to the Russia lineage.

**Рис. 5.** Распространение изолятов линии «Россия».

● – вариант CVR-I; ● – вариант CVR-VI; ● – вариант CVR-V, содержит инсерцию CAGTCTATAAGA в *IGR MGF 360-13L – MGF 360-14L*; ● – вариант CVR-V, содержит инсерцию в *IGR MGF 360-13L – MGF 360-14L*, а также транзиции в *IGR C62L – C962R* и в генах *D1133L* и *Q706L*; ● – изоляты, несущие вариант CVR-V, предположительно относящиеся к линии «Россия».

can be clarified by whole-genome sequencing of a larger number of ASF virus isolates. A molecular-epizootological map showing the distribution of both confirmed and presumed Russia lineage isolates, with their marker sites indicated, is shown in **Fig. 5**. **Fig. 6** shows the prevalence of ASF genotype II genetic lineages in Europe according to the above-mentioned methodological recommendations.

The isolates we studied also contain previously undescribed variations. A structure similar to polymorphisms 2153 and 2161 is observed in the paralog of the *MGF 360-1L – MGF 360-21R* gene, but no other substitutions characteristic of this region were found. This combination was also found in some incorrectly assembled sequences from GenBank, which we did not use for analysis. These two paralogs are among the most closely related in the ASF virus genome (69% identity) [17], which increases the likelihood of errors during genome assembly. The *MGF-360-1L* gene does not play a significant role in ASF virulence [25], therefore, these mutations should not lead to changes in the properties of the virus.

The Q>H substitution at position 22 observed in the *MGF 360-11L* protein of ASFV\_Penza-2019a, b and ASFV\_Tambov-2021 is non-conservative, which should affect the conformation of the protein molecule. The *MGF 360-11L* protein has an inhibitory effect on the pro-

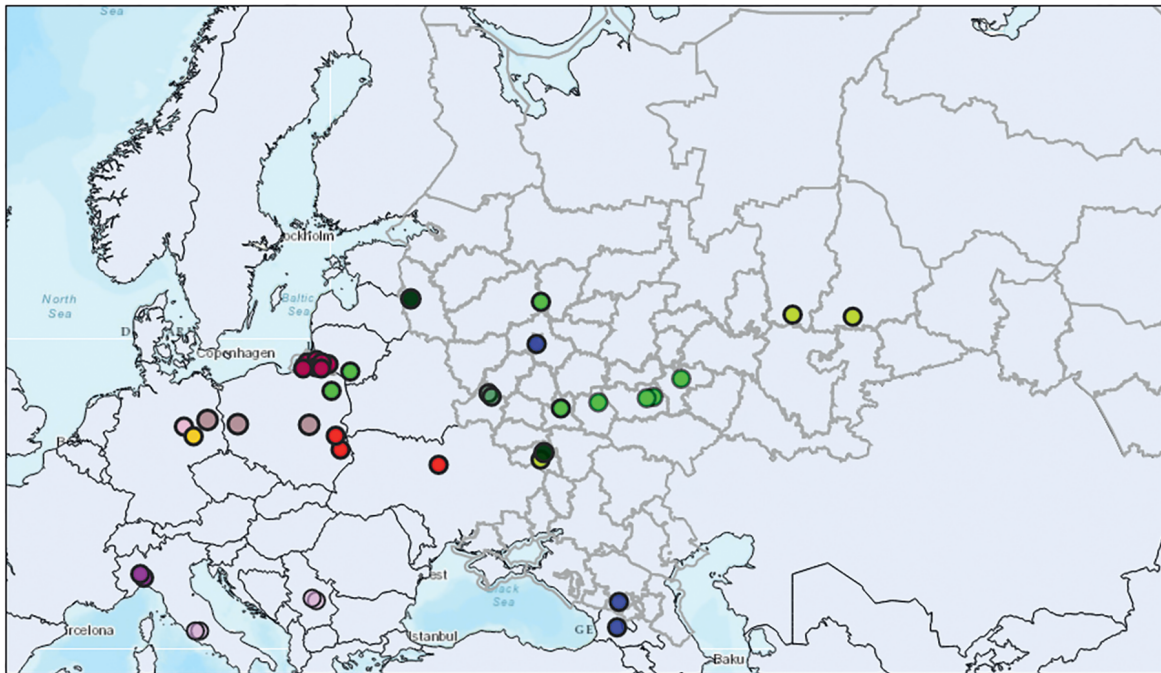
duction of type I interferon (IFN) and interleukins IL-1 $\beta$  and IL-6, inhibiting the activity of the IFN- $\beta$  promoter and ISRE sequence, interfering with the normal functioning of the cellular immune response and limiting the synthesis of inflammatory factors [26].

The substitution of alanine for proline at position 85 of the E199L protein found in ASFV\_Penza-2019b was also found in other isolates in Russia and other countries. However, as shown in [16], the polymorphism of the E199L gene is not significant as a marker of diversity, since its changes are associated not with geographical distribution, but with the adaptation of the virus to the host. This protein is necessary for the fusion event that leads to the penetration of the virus into the host cell [27] and also induces autophagy [28]. The same applies to the replacement of glutamine with histidine at position 104 of E199L in ASFV\_Lipezk-2016.

The remaining unique polymorphisms detected do not lead to changes in the amino acid sequence, although they may also have potential phylogenetic significance in the future when describing new complete virus sequences isolated in central regions of Russia.

### Conclusion

Based on the analysis, it was established that ASF virus



**Fig. 6.** Distribution of ASFV genotype II genetic lineages in Europe. Isolates presumably belonging to the Russia lineage are not listed.

● – Georgia 2007/Russia; ● – Georgia 2007/Unclassified lineages; ● – Eastern Europe/Poland; ● – Eastern Europe/Kyiv 2016; ● – Eastern Europe/Kaliningrad; ● – Europe/Italy; ● – Europe/Serbia; ● – Eastern Europe/Germany 2020; ● – Eastern Europe/Germany 2021; ● – Europe/Pskov 2021; ● – Europe/Central Russia 2021; ● – Europe/Bryansk 2021.

**Рис. 6.** Встречаемость генетических линий АЧС II генотипа в Европе. Изоляты, предположительно относящиеся к линии «Россия», не указаны.

● – Georgia 2007/Россия; ● – Georgia 2007/Неклассифицированные линии; ● – Восточная Европа/Польша; ● – Восточная Европа/Киев 2016; ● – Восточная Европа/Калининград; ● – Европа/Италия; ● – Европа/Сербия; ● – Восточная Европа/Германия 2020; ● – Восточная Европа/Германия 2021; ● – Европа/Псков 2021; ● – Европа/Центр России 2021; ● – Европа/Брянск 2021.

isolates detected in the Lipetsk, Penza, and Tambov regions between 2016 and 2021 belong to the CVR-V variant of the Georgia 2007 clade of the Russia genetic lineage. A new probable marker of diversity was found in *IGR MGF 360-13L – MGF 360-14L*, and polymorphisms were proposed in *IGR C62L – C962R* and in the *D1133L* and *Q706L* genes, which have phylogenetic potential for differentiating ASF virus strains in the central regions of Russia.

**REFERENCES / ЛИТЕРАТУРА**

- Chen S., Wang T., Luo R., Lu Z., Lan J., Sun Y., et al. Genetic variations of African swine fever virus: major challenges and prospects. *Viruses*. 2024; 16(6): 913. <https://doi.org/10.3390/v16060913>
- Gallardo C., Casado N., Soler A., Djadjovski I., Krivko L., Madueño E., et al. A multi gene-approach genotyping method identifies 24 genetic clusters within the genotype II-European African swine fever viruses circulating from 2007 to 2022. *Front. Vet. Sci.* 2023; (10): 1112850. <https://doi.org/10.3389/fvets.2023.1112850>
- Mazloum A., van Schalkwyk A., Chernyshev R., Igolkin A., Heath L., Sprygin A.A. Guide to molecular characterization of genotype II African swine fever virus: essential and alternative genome markers. *Microorganisms*. 2023; 11(3): 642. <https://doi.org/10.3390/microorganisms11030642>
- Chernyshev R., Igolkin A., van Schalkwyk A., Zinyakov N., Kolbin I., Shotin A., et al. A proposed update of African swine fever virus (genotype II) subgenotyping based on the central variable region (CVR) of Russian isolates. *Arch. Virol.* 2024; 169(7): 147. <https://doi.org/10.1007/s00705-024-06064-w>
- Nix R.J., Gallardo C., Hutchings G., Blanco E., Dixon L.K. Mo-

- lecular epidemiology of African swine fever virus studied by analysis of four variable genome regions. *Arch. Virol.* 2006; 151(12): 2475–94. <https://doi.org/10.1007/s00705-006-0794-z>
- Mazloum A., van Schalkwyk A., Chernyshev R., Shotin A., Korennoy F.I., Igolkin A., et al. Genetic characterization of the central variable region in African swine fever virus isolates in the Russian Federation from 2013 to 2017. *Pathogens*. 2022; 11(8): 919. <https://doi.org/10.3390/pathogens11080919>
- Mazloum A., van Schalkwyk A., Shotin A., Igolkin A., Shevchenko I., Gruzdev K.N., et al. Comparative analysis of full genome sequences of African swine fever virus isolates taken from wild boars in Russia in 2019. *Pathogens*. 2021; 10(5): 521. <https://doi.org/10.3390/pathogens10050521>
- Bolger A.M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*. 2014; 30(15): 2114–20. <https://doi.org/10.1093/bioinformatics/btu170>
- Prijbelski A., Antipov D., Meleshko D., Lapidus A., Korobeynikov A. Using SPAdes de novo assembler. *Curr. Protoc. Bioinformatics*. 2020; 70(1): e102. <https://doi.org/10.1002/cpbi.102>
- Okonechnikov K., Golosova O., Fursov M.; UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*. 2012; 28(8): 1166–7. <https://doi.org/10.1093/bioinformatics/bts091>
- Danecek P., Bonfield J.K., Liddle J., Marshall J., Ohan V., Pollard M.O., et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021; 10(2): giab008. <https://doi.org/10.1093/gigascience/giab008>
- Tcherepanov V., Ehlers A., Upton C. Genome Annotation Transfer Utility (GATU): rapid annotation of viral genomes using a closely related reference genome. *BMC Genomics*. 2006; 7: 150. <https://doi.org/10.1186/1471-2164-7-150>
- Katoh K., Standley D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 2013; 30(4): 772–80. <https://doi.org/10.1093/molbev/mst010>
- Tamura K., Stecher G., Kumar S. MEGA11: Molecular Evolution-

- ary Genetics Analysis Version 11. *Mol. Biol. Evol.* 2021; 38(7): 3022–7. <https://doi.org/10.1093/molbev/msab120>
15. Mazur-Panasiuk N., Walczak M., Juszkiewicz M., Woźniakowski G. The spillover of African swine fever in western Poland revealed its estimated origin on the basis of O174L, K145R, MGF 505-5R and IGR I73R/I329L genomic sequences. *Viruses*. 2020; 12(10): 1094. <https://doi.org/10.3390/v12101094>
  16. Chernyshev R., Mazloum A., Zinyakov N., Kolbin I., Shotin A., Korennoy F.I., et al. Unique nucleotide polymorphism of African swine fever virus circulating in East Asia and Central Russia. *Viruses*. 2024; 16(12): 1907. <https://doi.org/10.3390/v16121907>
  17. Dixon L.K., Chapman D.A., Netherton C.L., Upton C. African swine fever virus replication and genomics. *Virus Res.* 2013; 173(1): 3–14. <https://doi.org/10.1016/j.virusres.2012.10.020>
  18. Coulson D., Upton C. Characterization of indels in poxvirus genomes. *Virus Genes*. 2011; 42(2): 171–7. <https://doi.org/10.1007/s11262-010-0560-x>
  19. Lovett S.T. Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences. *Mol. Microbiol.* 2004; 52(5): 1243–53. <https://doi.org/10.1111/j.1365-2958.2004.04076.x>
  20. Zhu Z., Xiao C.T., Fan Y., Cai Z., Lu C., Zhang G., et al. Homologous recombination shapes the genetic diversity of African swine fever viruses. *Vet. Microbiol.* 2019; 236: 108380. <https://doi.org/10.1016/j.vetmic.2019.08.003>
  21. Hou J., Shen C., Dajun Z., Yang B., Shi X., Zhang T., et al. Gene sequence analysis, protein structure prediction and subcellular localization of African swine fever virus helicase D1133L. *Acta Veterinaria et Zootechnica Sinica*. 2021; 52(7): 1953–62. <https://doi.org/10.11843/j.issn.0366-6964.2021.07.017>
  22. Hao Y., Yang J., Yang B., Zhang T., Shi X., Yang X., et al. Identification and analysis of the interaction network of African swine fever virus D1133L with host proteins. *Front. Microbiol.* 2022; 13: 1037346. <https://doi.org/10.3389/fmicb.2022.1037346>
  23. Zhang T., Yang B., Cui H., Yuan X., Zhao D., Yang J., et al. Establishment and preliminary application of MA-104 cell line over-expressing African swine fever virus D1133L protein. *Acta Veterinaria et Zootechnica Sinica*. 2022; 53(6): 1877–85. <https://doi.org/10.11843/j.issn.0366-6964.2022.06.021>
  24. Freitas F.B., Frouco G., Martins C., Ferreira F. The QP509L and Q706L superfamily II RNA helicases of African swine fever virus are required for viral replication, having non-redundant activities. *Emerg. Microbes Infect.* 2019; 8(1): 291–302. <https://doi.org/10.1080/22221751.2019.1578624>
  25. Ramirez-Medina E., Vuono E.A., Rai A., Pruitt S., Silva E., Velazquez-Salinas L., et al. Evaluation in swine of a recombinant African swine fever virus lacking the MGF-360-1L gene. *Viruses*. 2020; 12(10): 1193. <https://doi.org/10.3390/v12101193>
  26. Yang K., Xue Y., Niu H., Shi C., Cheng M., Wang J., et al. African swine fever virus MGF360-11L negatively regulates cGAS-STING-mediated inhibition of type I interferon production. *Vet. Res.* 2022; 53(1): 7. <https://doi.org/10.1186/s13567-022-01025-0>
  27. Matamoros T., Alejo A., Rodríguez J.M., Hernández B., Guerra M., Fraile-Ramos A., et al. African swine fever virus protein pE199L mediates virus entry by enabling membrane fusion and core penetration. *mBio*. 2020; 11(4): e00789-20. <https://doi.org/10.1128/mbio.00789-20>
  28. Chen S., Zhang X., Nie Y., Li H., Chen W., Lin W., et al. African swine fever virus protein E199L promotes cell autophagy through the Interaction of PYCR2. *Virol. Sin.* 2021; 36(2): 196–206. <https://doi.org/10.1007/s12250-021-00375-x>

**Information about the authors:**

**Daria A. Skorobagatko** ✉ – Candidate of Biological Sciences (Ph. D.), Specialist in Department of Genomics and Molecular Biology, Cherkizovo Research and Testing Center LLC, Moscow, Russia. E-mail: [d.skorobagatko86@gmail.com](mailto:d.skorobagatko86@gmail.com); <https://orcid.org/0000-0003-3239-9004>

**Ekaterina S. Tolkova** – Main Specialist in Department of Genomics and Molecular Biology, Cherkizovo Research and Testing Center LLC, Moscow, Russia. E-mail: [katerina.tolkova@gmail.com](mailto:katerina.tolkova@gmail.com); <https://orcid.org/0000-0002-4770-3796>

**Olga A. Shepeleva** – Head of the Department of Genomics and Molecular Biology, Cherkizovo Research and Testing Center LLC, Moscow, Russia. E-mail: [o.shepeleva@cherkizovo.com](mailto:o.shepeleva@cherkizovo.com); <https://orcid.org/0000-0002-3366-8184>

**Sergey O. Shapovalov** – Doctor of Sciences in Biology, Head of the Cherkizovo Research and Testing Center LLC, 108805, Moscow, Russia. E-mail: [s.shapovalov@cherkizovo.com](mailto:s.shapovalov@cherkizovo.com); <https://orcid.org/0000-0002-5630-5247>

**Contribution:** Skorobagatko D.A. – conducting bioinformatic analysis, data interpretation, critical processing and text preparation; Tolkova E.S. – design and concept of research, conducting experiments, data collection, bioinformatic analysis, text preparation; Shepeleva O.A. – conducting experiments, text preparation; Shapovalov S.O. – final approval of the article for publication.

Received 10 November 2025

Accepted 16 January 2026

Published 28 February 2026

**Информация об авторах:**

**Скоробагатько Дарья Александровна** ✉ – канд. биол. наук, специалист направления геномики и молекулярной биологии ООО НИЦ «Черкизово», Москва, Россия. E-mail: [d.skorobagatko86@gmail.com](mailto:d.skorobagatko86@gmail.com); <https://orcid.org/0000-0003-3239-9004>

**Толькова Екатерина Сергеевна** – главный специалист направления геномики и молекулярной биологии ООО НИЦ «Черкизово», Москва, Россия. E-mail: [katerina.tolkova@gmail.com](mailto:katerina.tolkova@gmail.com); <https://orcid.org/0000-0002-4770-3796>

**Шепелева Ольга Александровна** – руководитель направления геномики и молекулярной биологии ООО НИЦ «Черкизово», Москва, Россия. E-mail: [o.shepeleva@cherkizovo.com](mailto:o.shepeleva@cherkizovo.com); <https://orcid.org/0000-0002-3366-8184>

**Шаповалов Сергей Олегович** – д-р биол. наук, директор ООО НИЦ «Черкизово», Москва, Россия. E-mail: [s.shapovalov@cherkizovo.com](mailto:s.shapovalov@cherkizovo.com); <https://orcid.org/0000-0002-5630-5247>

**Участие авторов:** Скоробагатько Д.А. – проведение биоинформатического анализа, интерпретация данных, критическая обработка и подготовка текста; Толькова Е.С. – разработка и концепция исследования, проведение экспериментов, сбор данных, биоинформатический анализ, подготовка текста; Шепелева О.А. – проведение экспериментов, подготовка текста; Шаповалов С.О. – окончательное утверждение статьи для публикации.

Поступила 10.11.2025

Принята в печать 16.01.2026

Опубликована 28.02.2026