

ORIGINAL RESEARCHES



Original Study Article

<https://doi.org/10.36233/0372-9311-554>



Genomic surveillance of SARS-CoV-2 in Russia: insights from the VGARus platform

Ivan A. Kotov^{1,2}, Matvey R. Agletdinov^{1,2}, German V. Roev^{1,2}, Ekaterina V. Pimkina¹, Maksim I. Nadtoka¹, Arina V. Peresadina¹, Anna Yu. Bukharina¹, Dmitry V. Svetlichny¹, Sergey E. Goncharov¹, Anastasiia V. Vykhodtseva¹, Nadezhda I. Borisova¹, Vladislav G. Lysenkov¹, Mikhail D. Chanyshv¹, David N. Agabalaev¹, Valeriia V. Saenko¹, Anna S. Cherkashina¹, Tatiana A. Semenenko³, Dmitry V. Dubodelov¹, Kamil F. Khafizov^{1✉}, Vasily G. Akimkin¹

¹Central Research Institute for Epidemiology, Moscow, Russia;

²Moscow Institute of Physics and Technology (National Research University), Dolgoprudny, Russia;

³National Research Centre of Epidemiology and Microbiology named after the Honorary Academician N.F. Gamaleya, Moscow, Russia

Abstract

Introduction. In response to the COVID-19 pandemic in the Russian Federation, comprehensive response measures were taken. One of these measures was the development of a viral genome aggregation platform (VGARus) to monitor virus variability.

The **aim** of this paper is to describe the role of the VGARus platform in tracking genetic variation in SARS-CoV-2.

Materials and methods. VGARus utilizes sequencing data and bioinformatics tools to monitor genetic variations in SARS-CoV-2. The viral genomes were aligned using NextClade, which also translated them into amino acids and identified mutations. The viral variability over time was analyzed by counting the number of amino acid changes compared to the reference sequence.

Results. The analysis of data within VGARus enabled the identification of new virus variants, contributing to improved diagnostic tests and vaccine development. The platform allowed for the prediction of epidemiologic trends, facilitating a rapid response to changes in the epidemiologic situation. For example, using VGARus, an increase in COVID-19 incidence was accurately predicted in the summer of 2022 and early 2023, which were associated with the emergence of Omicron subvariants BA.5 and XBB. Data from the platform helps validate the effectiveness of primers and DNA probes to ensure high diagnostic accuracy and reduce the risk of false negatives.

Conclusion. VGARus demonstrates the growing role of genomic surveillance in combating COVID-19 and improving preparedness for future infectious disease outbreaks. The platform is a powerful tool for generating evidence-based solutions to combat a pandemic and mitigate its health, economic and societal impacts. It provides the ability to promptly obtain information on the epidemiologic situation in a particular region of the Russian Federation, use genomic data for phylogenetic analysis, compare the mutational spectrum of SARS-CoV-2 sequences with foreign samples. VGARus data allow for both retrospective analysis and predictive hypotheses. For example, we can clearly see the dynamics of the change of different virus variants: sequences belonging to the Alpha, Beta, Delta, Omicron lineages and many less common ones, clearly form the upsurges of morbidity, the interaction of which is reflected in the epidemiological picture. It is also currently being expanded to monitor other pathogens, increasing its public health relevance.

Keywords: *genomic epidemiology, molecular epidemiology, SARS-CoV-2, next generation sequencing, VGARus platform, genomic surveillance*

Ethics approval. The study was conducted with the informed consent of the patients. The research protocol was approved by the Ethics Committee of the Central Research Institute for Epidemiology (protocol No. 111, December 22, 2020).

Acknowledgement. We express our sincere gratitude to everyone who took part in the creation, maintenance and filling of the VGARus platform with genomic data.

Funding source. The sequencing and data analysis efforts at the Institute were supported by funding from the "Sanitary Shield" federal project, subsidies provided under an order from the Russian Federation Government, and the Institute's internal financial resources.

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

For citation: Kotov I.A., Agletdinov M.R., Roev G.V., Pimkina E.V., Nadtoka M.I., Peresadina A.V., Bukharina A.Yu., Svetlichny D.V., Goncharov S.E., Vykhodtseva A.V., Borisova N.I., Lysenkov V.G., Chanyshv M.D., Agabalaev D.N., Saenko V.V., Cherkashina A.S., Semenenko T.A., Dubodelov D.V., Khafizov K.F., Akimkin V.G. Genomic surveillance of SARS-CoV-2 in Russia: insights from the VGARus platform. *Journal of microbiology, epidemiology and immunobiology*. 2024;101(4):435–447.

DOI: <https://doi.org/10.36233/0372-9311-554>

EDN: <https://www.elibrary.ru/irjxcx>

Оригинальное исследование

<https://doi.org/10.36233/0372-9311-554>

Геномный надзор за SARS-CoV-2 в Российской Федерации: возможности платформы VGARus

Котов И.А.^{1,2}, Аглетдинов М.Р.^{1,2}, Роев Г.В.^{1,2}, Пимкина Е.В.¹, Надтока М.И.¹, Пересади́на А.В.¹, Бухарина А.Ю.¹, Светличный Д.В.¹, Гончаров С.Е.¹, Выходцева А.В.¹, Борисова Н.И.¹, Лысенков В.Г.¹, Чанышев М.Д.¹, Агабалаев Д.Н.¹, Саенко В.В.¹, Черкашина А.С.¹, Семенов Т.А.³, Дубоделов Д.В.¹, Хафизов К.Ф.¹, Акимкин В.Г.¹

¹Центральный научно-исследовательский институт эпидемиологии Роспотребнадзора, Москва, Россия;

²Московский физико-технический институт (национальный исследовательский университет), Долгопрудный, Россия;

³Национальный исследовательский центр эпидемиологии и микробиологии имени почётного академика Н.Ф. Гамалеи, Москва, Россия

Аннотация

Введение. В ответ на пандемию COVID-19 в России были приняты комплексные меры реагирования. Одной из них стала разработка платформы агрегации вирусных геномов (VGARus) для мониторинга изменчивости вируса.

Цель работы — описать роль VGARus в отслеживании генетических изменений SARS-CoV-2.

Материалы и методы. Выравнивание вирусных геномов, последующую трансляцию в аминокислоты и поиск мутаций производили с помощью программы NextClade. С целью анализа геномной изменчивости подсчитывали число аминокислотных изменений относительно референсной последовательности.

Результаты. Анализ данных VGARus позволил идентифицировать новые варианты вируса, что способствовало улучшению диагностических тестов и может помочь в разработке вакцин. Платформа предоставила возможность прогнозировать эпидемиологические тенденции и оперативно реагировать на изменения эпидемиологической ситуации. Например, с использованием VGARus был точно предсказан рост заболеваемости COVID-19 летом 2022 г. и в начале 2023 г., связанный с появлением субвариантов Omicron BA.5 и XBB. Данные платформы помогают проверять эффективность праймеров и ДНК-зондов, что обеспечивает высокую точность диагностики и снижает риск ложноотрицательных результатов.

Заключение. VGARus демонстрирует растущую роль геномного эпиднадзора в борьбе с COVID-19 и повышении готовности к будущим вспышкам инфекционных заболеваний. Платформа является мощным инструментом для формирования научно обоснованных решений по борьбе с пандемией и смягчению её последствий для здоровья населения, экономики и общества. Она предоставляет возможность оперативно получать информацию об эпидемиологической обстановке в конкретном регионе России, использовать геномные данные для проведения филогенетического анализа, сравнивать мутационный спектр последовательностей SARS-CoV-2 с зарубежными образцами. Данные VGARus позволяют проводить ретроспективный анализ и выдвигать гипотезы прогностического характера. Так, явно можно увидеть динамику смены различных вариантов вируса: последовательности, принадлежащие линиям Alpha, Beta, Delta, Omicron и многим менее распространённым, отчётливо формируют подёмы заболеваемости, которые отражаются на эпидемиологической ситуации. В данный момент платформа расширяется для мониторинга изменчивости других патогенов, что увеличивает её значимость для общественного здравоохранения.

Ключевые слова: геномная эпидемиология, молекулярная эпидемиология, SARS-CoV-2, секвенирование следующего поколения, платформа VGARus, геномный надзор

Этическое утверждение. Исследование проводилось при добровольном информированном согласии пациентов. Протокол исследования одобрен Этическим комитетом ЦНИИ Эпидемиологии (протокол № 111 от 22.12.2020).

Благодарность. Мы выражаем искреннюю благодарность всем, кто принимал участие в создании, обслуживании и наполнении геномными данными базы VGARus.

Источник финансирования. Работы по секвенированию и анализу данных в ЦНИИ Эпидемиологии Роспотребнадзора проводились при финансовой поддержке федерального проекта «Санитарный щит», субсидий, выделенных по распоряжению Правительства РФ, а также за счёт внутренних финансовых ресурсов ЦНИИ Эпидемиологии.

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

Для цитирования: Котов И.А., Аглетдинов М.Р., Роев Г.В., Пимкина Е.В., Надтока М.И., Пересадына А.В., Бухарина А.Ю., Светличный Д.В., Гончаров С.Е., Выходцева А.В., Борисова Н.И., Лысенков В.Г., Чанышев М.Д., Агабалаев Д.Н., Саенко В.В., Черкашина А.С., Семенов Т.А., Дубоделов Д.В., Хафизов К.Ф., Акимкин В.Г. Геномный надзор за SARS-CoV-2 в Российской Федерации: возможности платформы VGARus. *Журнал микробиологии, эпидемиологии и иммунобиологии*. 2024;101(4):435–447.

DOI: <https://doi.org/10.36233/0372-9311-554>

EDN: <https://www.elibrary.ru/irjxcx>

Introduction

The advent of high-throughput sequencing technologies, also known as next-generation sequencing (NGS), has led to a significant reduction in the cost of genomic sequencing experiments over the past 15 years. NGS is increasingly utilized across various biological and medical fields, including virology, where its application in studying viral genomes has become widespread [1–5]. Furthermore, contemporary bioinformatics tools have expanded the capabilities for the development and analysis of databases containing the genomes of pathogens responsible for various infectious diseases [6–8]. Genomic epidemiology has emerged as a crucial component in epidemic control. It enables the examination of genetic alterations in the genomes of pathogens, the identification and classification of distinct lineages, and the evaluation of their pathogenic potential and transmissibility [9–15]. These studies are very important for the development of new diagnostic kits, the creation of modern and effective vaccines, the formulation of optimal epidemic response strategies, and the prediction of disease incidence.

A striking example of the application of molecular genetic monitoring is the detailed study of a new coronavirus infection during the COVID-19 pandemic [16]. By analyzing the genomes of SARS-CoV-2, associations were established between different virus variants and the characteristics of the course of the epidemic. This approach allows for accurate monitoring, understanding the relationship between genetic variants and their ability to cause disease, and implementing targeted measures to prevent the spread of infection.

At the onset of the COVID-19 pandemic, Professor Edward Holmes from the University of Sydney, representing a research team led by Yong-Zheng Zhang from Fudan University in Shanghai, published the nucleotide sequence of the SARS-CoV-2 genome. This information was posted on the Virological.org platform¹, which allowed the international scientific community to begin taking immediate action to counter the spread of the pathogen, among which were the

development of new diagnostic tests and subsequent vaccine development [17, 18]. As the pandemic progressed, countries that had typically relied less on their own genomic data began to conduct extensive sequencing experiments. The knowledge gained was used to develop strategic plans to contain the spread of infection [19]. The widespread use of SARS-CoV-2 genome sequencing led to a significant increase in the number of new sequences uploaded to international databases. The well-known database is GISAID (<https://www.gisaid.org>) with more than 16 million sequences from more than 200 countries [8].

The aim of the study is to determine the role of the VGARus platform and its data for analyzing genomic sequences of SARS-CoV-2 virus collected in Russia.

Materials and methods

Before starting this study, informed consent was obtained from patients, and the protocol was approved by the ethical committee of the Central Research Institute of Epidemiology (protocol No. 111 of 22.12.2020). Biological material was obtained by taking nasopharyngeal swabs from patients with COVID-19 symptoms. The samples were collected from different regions of Russia, with most of them coming from Moscow and the Moscow region. The presence of SARS-CoV-2 RNA was confirmed by real-time reverse transcription polymerase chain reaction (RT-PCR). RIBO-prep kit (AmpliSense, Russia) was used for RNA isolation, and REVERTA-L reagent kit (AmpliSense, Russia) was used for reverse transcription.

High-throughput sequencing was performed on the Illumina MiSeq platform (Illumina, San Diego, CA, USA) using MiSeq Reagent Kit v2 (PE 150 + 150 or PE 250 + 250 cycles) or MiSeq Reagent Kit v3 (PE 300 + 300 cycles), Illumina NextSeq 2000 using NextSeq 1000/2000 P2 reagents v3 (300 cycles), MinION using Midnight Kit (Oxford Nanopore Technologies Oxford, UK), DNBSEQ-G50 using ATOplex RNA Library Prep Set (MGI Tech, Shenzhen, China). The Sanger method was used to sequencing fragments of the spike protein gene, but this information was barely utilized in the analysis. In addition, nucleotide sequence data from the GISAID database were used in case of their geographi-

¹ Novel 2019 Coronavirus Genome.

URL: <https://virological.org/t/novel-2019-coronavirus-genome/319>

cal affiliation to Russia. The program “Pangolin” [19], as well as internal tools and scripts, were used to classify different variants of SARS-CoV-2.

In total, more than 82,000 complete SARS-CoV-2 genomes with the date of biomaterial collection from 01.01.2020 to 31.12.2023 were used. Only genomic sequences that met the specified quality criteria were selected. The selected genomes were aligned to the reference sequence NC_045512.2 using the NextClade tool and then translated into amino acid sequences. A specialized script written in Python was used to count the number of amino acid changes compared to the reference.

Results

Development and creation of the Russian viral genome aggregation platform VGARus

In 2021, the VGARus platform (Virus Genome Aggregator of Russia; registration date 06.07.2023, No. 2023622263) was developed and established at the Central Research Institute of Epidemiology (CRIE) of Rospotrebnadzor in accordance with the decree of the Government of the Russian Federation. The key tasks of this platform are collection of data on viral genomes, centralized analysis of genetic diversity and temporal dynamics of identified SARS-CoV-2 variants in Russia. A scientific consortium was established, comprising institutions from Rospotrebnadzor, the Ministry of Health of the Russian Federation, various scientific institutes, and other organizations. Currently, more than 150 organizations are members of the consortium, many of which are actively conducting extensive genomic sequencing of SARS-CoV-2 and uploading the obtained sequences to the VGARus database for further analysis. In addition, the Republic of Armenia and the Republic of Belarus are participating in the project, which makes it possible to track pathogen variability in neighboring countries with active transport connections.

The process of monitoring viral genome variability involves the following steps (**Figure 1**):

- the sequencing laboratory receives biological material from diagnostic laboratories, including those affiliated with hospitals. The quality of these samples is preliminarily assessed, typically through PCR analysis, to determine viral load and assess the sample's suitability for next-generation sequencing (NGS);
- the laboratory that provided the biological material must enter the relevant metadata into the VGARus platform (information on sex, age, vaccination status of the patient, date of collection of the biological material, region of collection, etc.);
- a specialized sequencing laboratory conducts the essential sample preparation followed by the sequencing of viral genomes;

- primary bioinformatics analysis is performed, which includes data quality control, genome assembly (usually by alignment to a reference genome) and sequence validity check (assessment of genome coverage);
- uploaded genomic information is validated and processed automatically using the “Pangolin” program for complete genomes [19, 20] and the “V-TRACE” program (developed by CRIE) for fragment sequencing results. Samples with genomes that fail quality control are marked as invalid on the platform, and a corresponding notification is sent to the originating laboratory.

All SARS-CoV-2 genomic sequences in the country, obtained through routine epidemiologic monitoring, are registered in the VGARus database. The system supports both manual uploading and uploading via specialized APIs, facilitating the addition of large volumes of sequences. Each sample entry in the system includes not only the nucleotide sequence but also associated technical data. Upon registration in the database, the sample is automatically assigned an internal identifier, and the SARS-CoV-2 genome sequence is appended to the sample information field. The technical information encompasses data on the organizations involved in sample collection and laboratory processing, the dates of sample receipt, registration in the system, and sequence upload.

The S-protein SARS-CoV-2 plays a key role in virus variant identification. This is due to its role in vi-

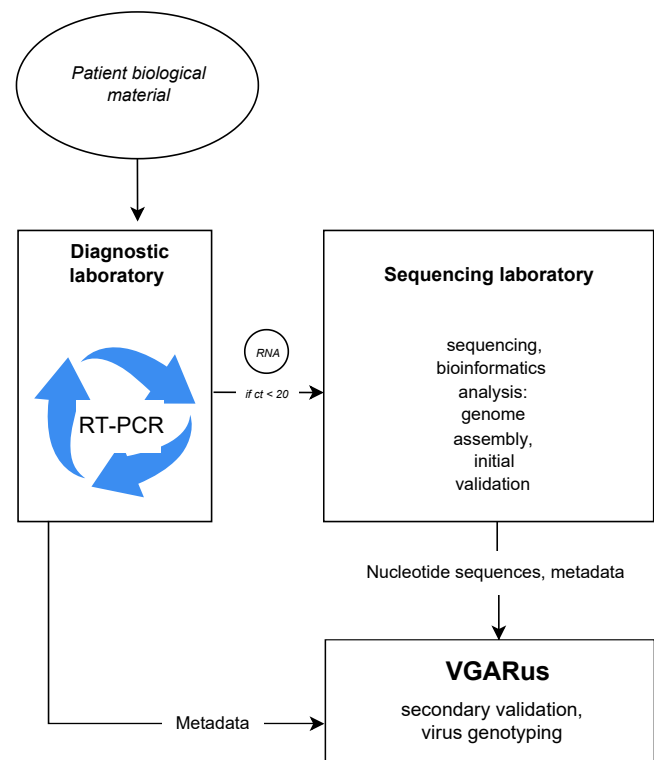


Fig. 1. The stages of the process of monitoring the variability of viral genomes.

tion entry into host cells and high mutation frequency/variability of the sequence [21]. However, attempting to establish a virus variant solely on the basis of mutations in the S-protein gene may lead to incomplete or inconsistent results [22]. Given the complexity of viral evolution, in which individual mutations may affect virus functions differently and interact in ways that may alter the overall effect, bioinformaticians at the Central Research Institute for Virus Evolution have developed the V-TRACE algorithm to address this problem. The algorithm identifies mutations in the S-protein gene of SARS-CoV-2, after which a plausibility measure of whether the sequence under study belongs to different virus lineages is estimated.

Temporal dynamics and evolutionary trajectories of SARS-CoV-2 variants

The VGARus platform is a valuable resource for tracking the dynamics of the COVID-19 pandemic in Russia and studying its peculiarities. In particular, systematically collected sequence information allows us to study the genomic diversity of the virus.

In 2020, a substantial diversity of SARS-CoV-2 lineages was observed [23]. These lineages did not exhibit significant advantages over one another, resulting in none of the variants becoming dominant. In December 2020, about a year after the new coronavirus began to spread worldwide, the UK authorities informed the World Health Organization of the discovery of a new SARS-CoV-2 lineage, named VOC-202012/01. It had numerous mutations in its genome and was originally named “British” but was later renamed Alpha to avoid naming variants by country. Among the mutations found in the S-protein gene, the most important were N501Y, P681H and Δ 69-70 [24, 25]. These mutations affected the ability of the virus to infect cells and evade the host immune response and, as a consequence, allowed it to spread more efficiently. This variant was detected in Russia in late 2020 and persisted into early 2021, coinciding with a sharp increase in the number of cases.

The Beta variant was identified shortly thereafter, but it had a much lower prevalence than Alpha. In the spring of 2021, the Delta variant appeared and quickly became dominant, leading to a significant increase in the incidence and hospitalization rate [26]. After a period of relatively favorable epidemiological conditions, the Omicron variant appeared in December 2021 (**Figure 2**), which led to a marked increase in the number of cases in Russia. However, the incidence of the disease declined just as quickly.

Despite a period of low numbers of COVID-19 cases in the spring of 2022, the emergence of Omicron subvariants BA.4 and BA.5 caused an increase in incidence that lasted until the end of October (**Figure 3**). In late 2022 and early 2023, highly contagious variants such as BQ.1* emerged. Such shifts in dominant lineages well illustrate the ever-changing and complex na-

ture of SARS-CoV-2 evolution. Notably, in early 2023, modified versions of pre-existing lineages returned to the virus population, notably Omicron BA.2, presented as recombinant forms of XBB*. In November 2023, a variant of coronavirus BA.2.86, unofficially named Pirola, began to spread rapidly in several countries, including Russia. It was notable for the large number of accumulated changes in the genome compared to earlier lineages and by the end of 2023 had become the predominant virus lineage, and in early 2024 its JN.1 sublineage was almost completely dominant in most countries of the world.

The pathogen variability described above underscores the importance of ongoing epidemiologic monitoring and sequencing of virus genomes for the timely detection of new variants or changes in viral population structure. Rapid identification of such changes can help in the development of public health strategies and in controlling the spread of these variants.

Comparative analysis of Figure 2 and Figure 3 reveals a trend wherein each new significant virus lineage becomes dominant once it reaches a threshold of 50% of the total population size, typically within 1.5 to 3.0 months. Additionally, the period during which an individual lineage remains dominant ranges from 3 months to 1 year.

We then attempted to explain the dynamics of COVID-19 incidence in Russia and to explore the possibility of predicting the rate of spread of a particular virus sublineage on the basis of SARS-CoV-2 sequence data. Our main hypothesis is that specific mutations in the virus genome significantly affect the incidence rate. However, the reported incidence rate undoubtedly depends on other critical factors such as population immunization rates, PCR testing coverage, and seasonal factors, whose exact contribution is difficult to estimate. Therefore, these factors were not used for the analysis.

The period from May 2020 to December 2023 was divided into 21-day intervals. The time period studied was limited to December 2023, allowing for a detailed examination of trends at that time. However, the subsequent emergence of the BA.2.86 variant has shown that predicting future trends can be extremely difficult. Given the key role that missense mutations play in viral transmission rates, nucleotide sequences were aligned to the reference genome and translated into amino acid sequences using NextClade [27]. The number of amino acid changes compared to the reference sequence was chosen as a metric of viral variability. For example, a rapid increase in the number of frequently occurring changes may indicate an active mutational process or the importation of a new lineage into the study region. The higher the mutational activity of a virus, the more likely it is that some subset of acquired mutations can affect the properties of the virus, such as its transmissibility.

The amino acid changes compared to the reference sequence from May 2020 to December 2023 are

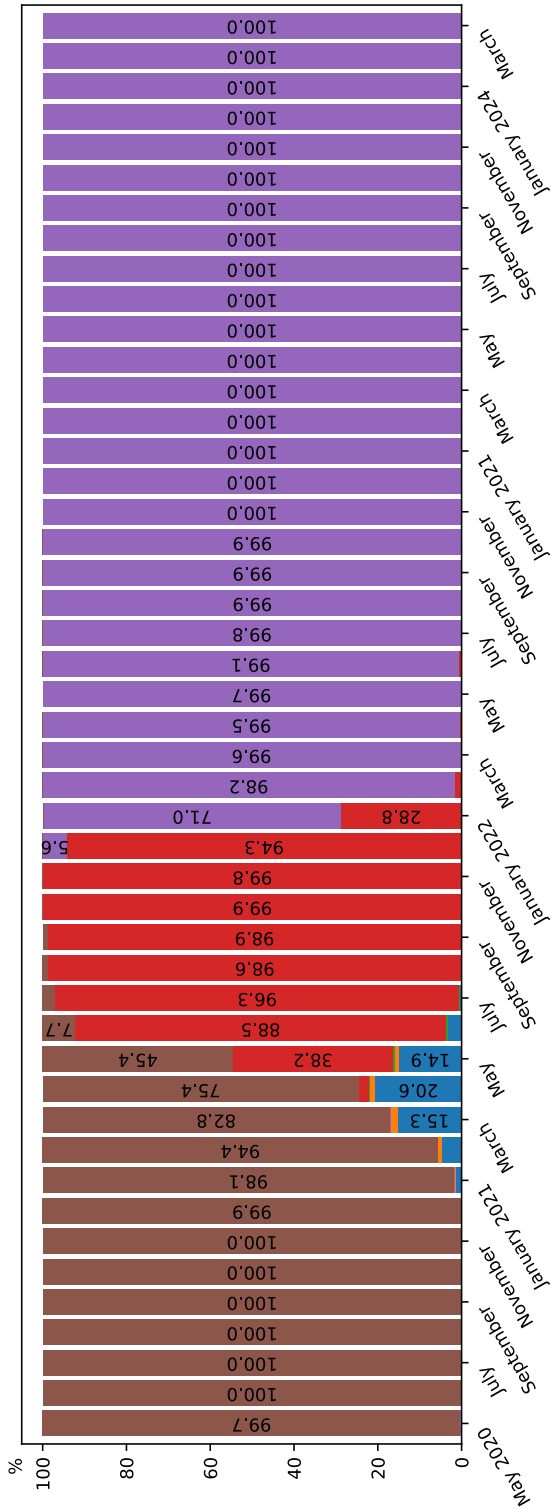


Fig. 2. A diagram illustrating the occurrence of significant SARS-CoV-2 variants of concern in the Russian Federation throughout 2020 until the March of 2024.

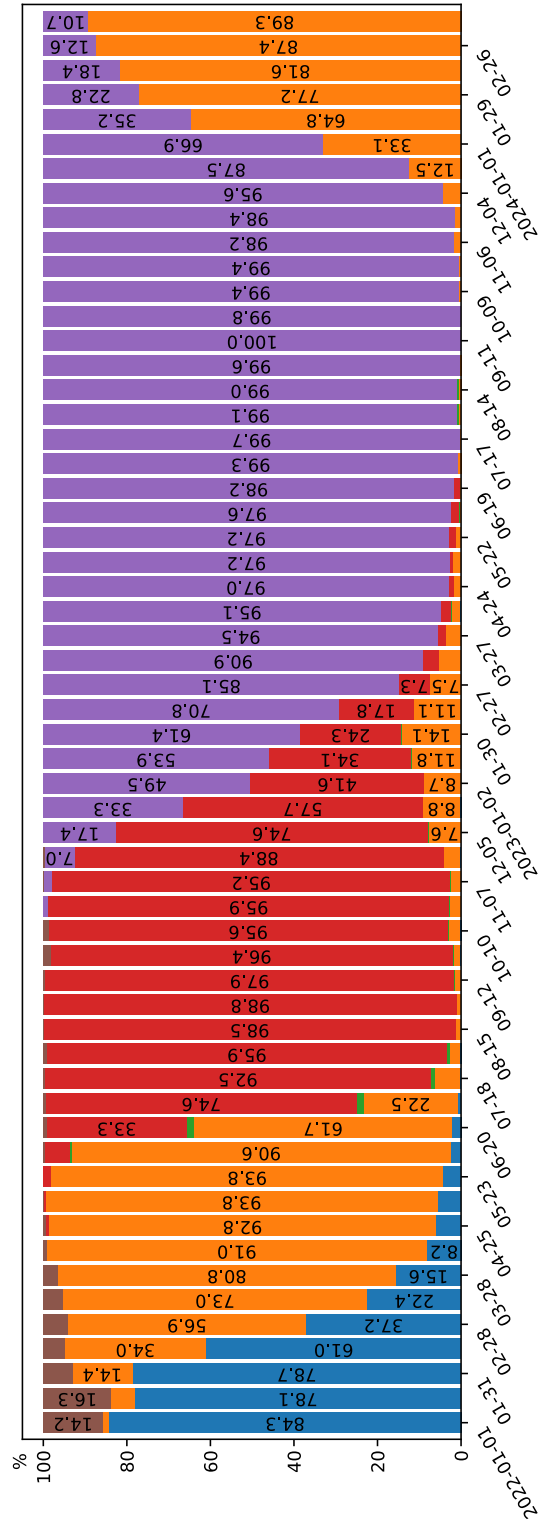


Fig. 3. Chart showing the frequency of occurrence of various Omicron sublineages from early 2022 to March of 2024.

The rise in the prevalence of the BA.2 sublineage by the end of 2023 is attributed to the BA.2.86 variant.

represented by the black line in **Figure 4**. For each interval, only mutations with a frequency of at least 50% are plotted. This approach enables the evaluation of mutations that significantly affect the adaptability of a viral variant or are inherited alongside such mutations. The graph clearly highlights three distinct intervals during which the number of frequent mutations in the SARS-CoV-2 genome increased. These intervals correspond to June 2021, February 2022, and January 2023, aligning with the widespread distribution of the Delta, Omicron, and XBB variants in Russia, respectively.

One potential limitation of the method described above is that it cannot effectively reflect the dynamics of change when new and old dominant lineages have different characteristic mutations, but their absolute number differs only slightly. A simple counting of mutations in such a case will lead to the erroneous conclusion that genetic evolution does not occur. To solve this problem, the qualitative composition of mutations was analyzed.

Sets of amino acid changes with a frequency greater than 50% were considered as separate sets for each time interval. Differences between them for adjacent time intervals were evaluated and used to measure the genetic variability of the virus. In this case, both the appearance and disappearance of a mutation with a frequency of at least 50% compared to the previous period was considered a change (Figures 2, 4). This auxiliary strategy demonstrated its reliability. Thanks to its application, we observed qualitative changes in the set of frequent mutations in the population caused by the transition from lineage BA.1 to BA.2. At the same time, fluctuations in the absolute number of common mutations were minimal. Both data sets described above are consistent with the World Health Organization data on the incidence of SARS-CoV-2 in Russia (Figures 3, 4).

Increases in disease incidence are often preceded by significant changes in the pathogen genome, as was the case in the summer of 2021 with the appearance of the Delta variant, in December 2021 with Omicron (BA.1/BA.2), in July 2022 with Omicron (BA.5), and in early 2023 with XBB (**Figure 5**). However, seasonal factors also play an important role.

We investigated the dynamics of mutation frequencies in the SARS-CoV-2 genome in Russia (**Figure 5**). It was noted that during the spread of the new dominant lineage at the above-mentioned time points, the frequency of mutations characteristic of them demonstrated a rapid S-shaped growth. When the new lineage began to dominate in the country, the frequency of mutations characteristic of the replaced lineage decreased along a similar S-shaped trajectory. In addition to the general picture, local trends can be observed on the mutation frequency distribution graph.

The frequency distribution of mutations also reveals two distinct trends on the right side of the graph. A more precise analysis of the mutations and sequences comprising these trends shows that both groups of lineages include XBB.1.9.1, FL.24, FL.1.5.1, XBB.1.16, XBB.1.16.11, and XBB.1.16.17. The upper trend, situated in the high-frequency range (70–90%) and indicated by the green dashed line, is formed by mutations common to all these lineages, such as S:G252V and ORF1b:S959P. Meanwhile, in the low-frequency range, the trend indicated by the dashed line is formed by mutations found in subgroups of the same lineages, such as S:E180V, which is present only in XBB.1.16, XBB.1.16.11, and XBB.1.16.17, and ORF1a:G1819S, found in the remaining lineages.

Studying the spreading dynamics of these lineages with time can potentially help predict the evolution of SARS-CoV-2. For this reason, the frequency dynamics for several of the discussed sublineages from June

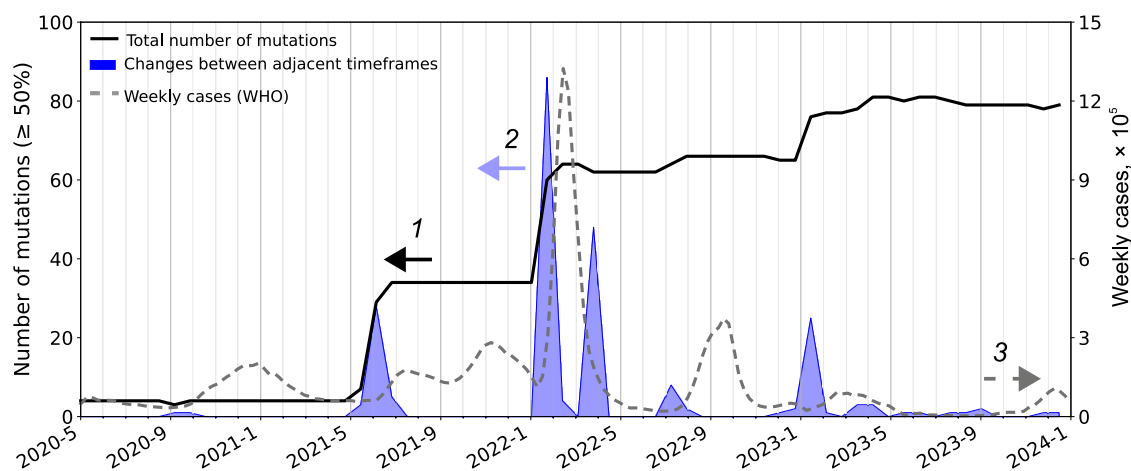


Fig. 4. Changes in amino acid sequences compared to the reference sequence from May 2020 to December 2023.

- 1 — number of amino acid substitutions relative to the reference variant (NC_045512.2) with a frequency of 50% or higher;
- 2 — number of changes in the set of mutations with a frequency of at least 50% compared to the previous period;
- 3 — temporal dynamics of COVID-19 cases in Russia according to WHO data.

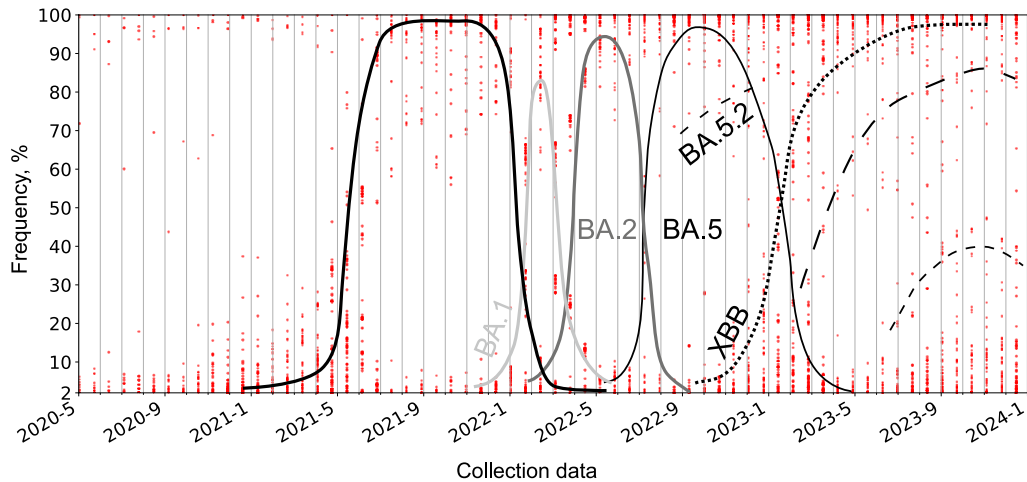


Fig. 5. Dynamics of the frequency distribution of amino acid mutations.

The lines indicate trends associated with changes in the dominant virus lineage. Each point represents a mutation plotted on the timeline, with its position indicating the frequency of occurrence at the time of sample collection. The major lineages leading to changes in mutation frequencies are shown. Local trends are highlighted with dashed lines.

through December 2023 are presented in **Figure 6**. Lineages XBB.1.9.1 and XBB.1.16 were intentionally excluded from this analysis due to the decreasing or static nature of their frequency trend during the period under consideration. The former line significantly decreased its prevalence to almost zero, while the latter line did not change its frequency, remaining between 6–16%. Consequently, they are not considered potentially future dominant lineages.

However, after this period, the BA.2.86* (“Pirrola”) lineage became dominant shortly after XBB, which could not have been predicted from data obtained by mid-December 2023, when this lineage was found in only a few samples. **Figure 6** also shows that most of the lineages previously considered potentially dominant, although becoming more common, reach a frequency of only about 16% in a few months. Meanwhile, earlier observations in this study suggest that a dominant line typically reaches 50% frequency within 1.5–3.0 months of emergence.

These data suggest that a future potentially dominant lineage must possess a certain minimum spreading rate; otherwise, it is likely to be displaced by others. This hypothesis aligns with the periodic nature of lineage changes. Additionally, such events complicate the prediction of the pandemic’s trajectory, particularly in determining the dominant lineage in the near future and its impact on public health. Finally, the rapid emergence and spread of an entirely new variant can render all previous predictions irrelevant.

Data on the distribution of mutations in the genome were also obtained. **Figure 7** highlights mutations that reached a frequency of at least 50% at the considered time points. Many of these mutations were found in the S-protein gene, which aligns with the findings of other research groups [28]. It is notable that each new dominant lineage introduces progressively fewer new mutations in this gene. While it is challenging to interpret this observation unambiguously at present, it may suggest the degree of relatedness between the lineages

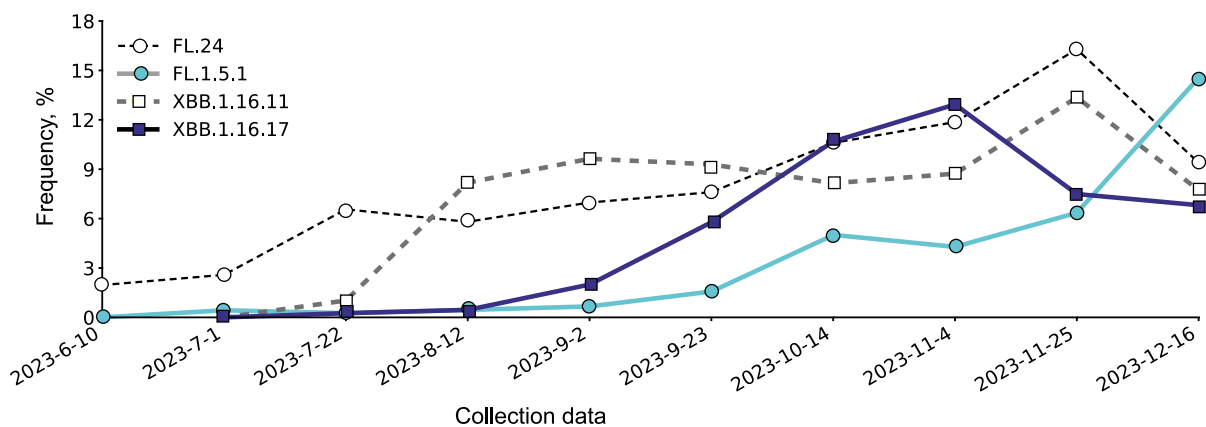


Fig. 6. Prevalence of SARS-CoV-2 from June 2023 through mid-December 2023.

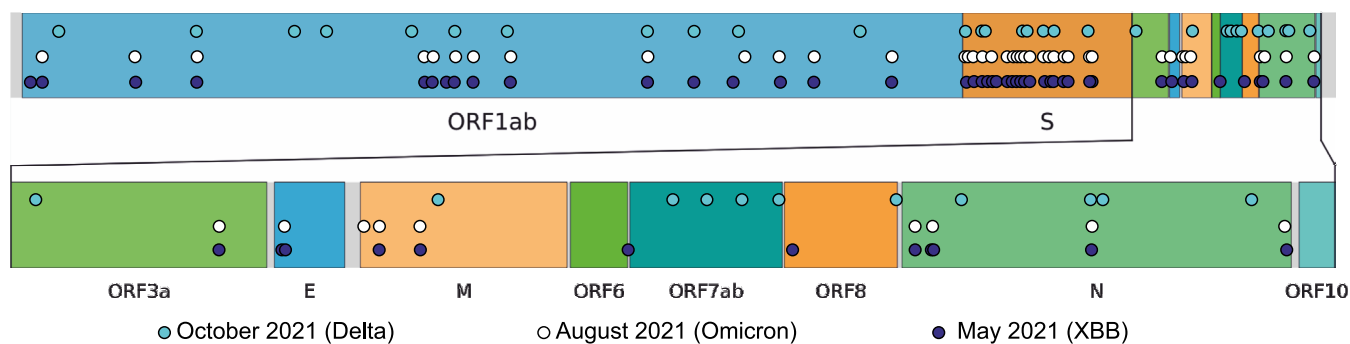


Fig. 7. Distribution of mutations across the SARS-CoV-2 genome presented in at least 50% of the samples for October 2021, August 2022, and May 2023. For each timeframe dominant strain is presented.

or that the gene is approaching its optimal structure for maximizing affinity to the human ACE2 receptor.

Discussion

The development of the Russian viral genome aggregation platform VGARus has become an important aspect in the fight against the COVID-19 pandemic in the country. This database contains over 320,000 SARS-CoV-2 genome sequences, including approximately 200,000 complete genomes. VGARus supports many essential functions, such as the identification of novel virus variants, the creation of effective diagnostic tools, and the formulation of public health policies [12–15].

VGARus has significantly contributed to the understanding of the spatial and temporal dynamics of the COVID-19 pandemic. By providing detailed information on the time and location of each genomic sample, the platform allows visualizing the distribution of specific viral variants in Russia and their evolution over time. This detailed knowledge provides an important advantage in predicting epidemiologic trends in the coming months. For example, using VGARus, we accurately predicted an increase in COVID-19 incidence in the summer of 2022 and early 2023, associated with the emergence of Omicron BA.5 and XBB subvariants, respectively. These predictive capabilities enable public health authorities to respond rapidly to changing epidemiologic situations.

In general, the results presented above on the frequency of occurrence of various SARS-CoV-2 variants in Russia are consistent with the data of other studies in different countries. For example, B. Xiao et al. used data from the GISAID platform [29]. They considered sequences obtained in the period from January 2020 to May 20, 2023 in the USA, UK, India, South Africa, Brazil, and Russia. During this period, the GISAID platform collected about 79,000 genomes from Russia.

The overall picture of the evolution and spread of the virus in different countries is similar, but there are significant regional differences. For example, in the

USA and the UK, the peak proportion of the Alpha variant was about 65% and 100%, respectively. At the same time, in Russia, the estimate of its peak prevalence was about 40%, according to the foreign study, and 20% in the current work. This difference can be explained by a wider geography of our study than the one based only on GISAID data.

The Delta variant arrived in Russia in approximately the same period as in the other countries studied — April–May 2021. Exceptions included India, where this variant appeared in March 2021, and Brazil, where the spread of the strain began only in July 2021. As in most of the other countries studied, the Gamma variant was practically undetectable in Russia, while it was dominant in Brazil and was registered in the USA with a peak frequency of about 10%.

The frequency dynamics of the BA.1 and BA.2 sublineages were nearly identical in Russia and all the countries studied, except for India. In India, BA.1 did not become dominant, remaining below 40%, while BA.2 spread earlier and remained dominant for almost 8 months. A unique characteristic of Russia and India was the almost complete absence of BA.4, which was found in other countries with peak frequencies ranging from 20% to 60%.

The presented comparison demonstrates that individual virus variants may begin their spread in different regions of the planet with a difference of several months, and the time of their maximum prevalence may vary considerably. The data on the frequency dynamics of SARS-CoV-2 variants are consistent with the findings on the timing of their spread and dominance made in this study.

In this study, we also demonstrated the main capabilities of the VGARus platform, presented the dynamics of variant diversity and mutations of SARS-CoV-2 in Russia, and examined in detail the scenario preceding the change of the dominant lineage in December 2023. The results of this part of the study revealed certain limitations in predicting the dominant lineage in the near future. These limitations, caused by the ra-

pid emergence of new lineages subsequently becoming dominant, represent a significant challenge for existing prediction tools [30, 31]. Nevertheless, we remain optimistic that future approaches will effectively address this factor. This work has, however, highlighted some general trends in the course of a pandemic, with the periodic nature of the change in the dominant variant being particularly crucial.

In addition to its role in tracking the course of epidemics, VGARus has practical implications, such as the development and assessment of diagnostic tests. For example, researchers at the Central Research Institute of Epidemiology regularly use VGARus data to test the effectiveness of primers and DNA probes used in test kits. Thus, the platform provides information for the development and processing of oligonucleotides for diagnostic test systems, information about the variants circulating in the country, about the variability of the pathogen sequence in the annealing site of primers and probes [32].

The capabilities of the VGARus platform have now been extended beyond SARS-CoV-2 research. The platform is currently being expanded to include data on additional pathogens, such as the viruses causing hepatitis, influenza, varicella, measles, and others. This multi-pathogen functionality will serve as a valuable tool for virologists and infectious disease specialists, enabling enhanced monitoring of the spread of various diseases and facilitating timely public health interventions.

Overall, VGARus is a significant achievement of the joint efforts of numerous scientific institutes of Rospotrebnadzor and other agencies [12]. Its implementation has expanded our understanding of SARS-CoV-2 and contributed to the study and control of the COVID-19 pandemic. VGARus emphasizes the critical importance of epidemiological monitoring in controlling infectious disease outbreaks and the significance of collaborative efforts in addressing global health crises.

СПИСОК ИСТОЧНИКОВ | REFERENCES

1. Barzon L., Lavezzo E., Militello V., et al. Applications of next-generation sequencing technologies to diagnostic virology. *Int. J. Mol. Sci.* 2011;12(11):7861–84. DOI: <https://doi.org/10.3390/ijms12117861>
2. Quer J., Colomer-Castell S., Campos C., et al. Next-generation sequencing for confronting virus pandemics. *Viruses.* 2022;14(3):600. DOI: <https://doi.org/10.3390/v14030600>
3. Capobianchi M.R., Giombini E., Rozera G. Next-generation sequencing technology in clinical virology. *Clin. Microbiol. Infect.* 2013;19(1):15–22. DOI: <https://doi.org/10.1111/1469-0691.12056>
4. Singh D.D. Next-generation sequencing technologies as emergent tools and their challenges in viral diagnostic. *Biomed. Res.* 2018;29(8):1637–44. DOI: <https://doi.org/10.4066/biomedicalresearch.29-18-362>
5. Mokili J.L., Rohwer F., Dutilh B.E. Metagenomics and future perspectives in virus discovery. *Curr. Opin. Virol.* 2012;2(1):63–77. DOI: <https://doi.org/10.1016/j.coviro.2011.12.004>
6. Akermi S., Jayant S., Ghosh A., et al. Viroinformatics for Viral diseases: tools and databases. In: *Translational Bioinformatics in Healthcare and Medicine*. Elsevier;2021:171–82.
7. Olson R.D., Assaf R., Brettin T., et al. Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. *Nucleic. Acids. Res.* 2023;51(D1):D678–89. DOI: <https://doi.org/10.1093/nar/gkac1003>
8. Shu Y., McCauley J. GISAID: Global initiative on sharing all influenza data — from vision to reality. *Euro. Surveill.* 2017;22(13):30494. DOI: <https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494>
9. Hill V., Ruis C., Bajaj S., et al. Progress and challenges in virus genomic epidemiology. *Trends Parasitol.* 2021;37(12):1038–49. DOI: <https://doi.org/10.1016/j.pt.2021.08.007>
10. Giovanetti M., Slavov S.N., Fonseca V., et al. Genomic epidemiology of the SARS-CoV-2 epidemic in Brazil. *Nat. Microbiol.* 2022;7(9):1490–500. DOI: <https://doi.org/10.1038/s41564-022-01191-z>
11. Klink G.V., Safina K.R., Nabieva E., et al. The rise and spread of the SARS-CoV-2 AY.122 lineage in Russia. *Virus Evol.* 2022;8(1):veac017. DOI: <https://doi.org/10.1093/ve/veac017>
12. Akimkin V., Semenenko T.A., Ugleva S.V., et al. COVID-19 epidemic process and evolution of SARS-CoV-2 genetic variants in the Russian Federation. *Microbiol. Res.* 2024;15(1):213–24.
13. Акимкин В.Г., Попова А.Ю., Плоскирева А.А. и др. COVID-19: эволюция пандемии в России. Сообщение I: проявления эпидемического процесса COVID-19. *Журнал микробиологии, эпидемиологии и иммунобиологии.* 2022;99(3):269–286. Akimkin V.G., Popova A.Yu., Ploskireva A.A., et al. COVID-19: the evolution of the pandemic in Russia. Report I: manifestations of the COVID-19 epidemic process. *Journal of microbiology, epidemiology and immunobiology.* 2022;99(3):269–286. DOI: [10.36233/0372-9311-276](https://doi.org/10.36233/0372-9311-276)
14. Акимкин В.Г., Попова А.Ю., Хафизов К.Ф. и др. COVID-19: эволюция пандемии в России. Сообщение II: динамика циркуляции геновариантов вируса SARS-CoV-2. *Журнал микробиологии, эпидемиологии и иммунобиологии.* 2022;99(4):381–396. Akimkin V.G., Popova A.Yu., Khafizov K.F., et al. COVID-19: evolution of the pandemic in Russia. Report II: dynamics of the circulation of SARS-CoV-2 genetic variants. *Journal of microbiology, epidemiology and immunobiology.* 2022;99(4):381–396. DOI: [10.36233/0372-9311-295](https://doi.org/10.36233/0372-9311-295)
15. Акимкин В.Г., Семенов Т.А., Хафизов К.Ф. и др. Стратегия геномного эпидемиологического надзора. Проблемы и перспективы. *Журнал микробиологии, эпидемиологии и иммунобиологии.* 2024;101(2):163–172. Akimkin V.G., Semenenko T.A., Khafizov K.F., et al. Genomic surveillance strategy. Problems and perspectives. *Journal of microbiology, epidemiology and immunobiology.* 2024;101(2):163–172. DOI: [10.36233/0372-9311-507](https://doi.org/10.36233/0372-9311-507)
16. Hill V., Githinji G., Vogels C.B.F., et al. Toward a global virus genomic surveillance network. *Cell Host Microbe.* 2023;31(6):861–73. DOI: <https://doi.org/10.1016/j.chom.2023.03.003>
17. Lu R., Zhao X., Li J., et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet.* 2020;395(10224):565–74. DOI: [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8)
18. Corman V.M., Landt O., Kaiser M., et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill.* 2020;25(3):2000045. DOI: <https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045>
19. Rambaut A., Holmes E.C., O’Toole Á., et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* 2020;5(11):1403–7. DOI: <https://doi.org/10.1038/s41564-020-0770-5>
20. O’Toole Á., Scher E., Underwood A., et al. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus. Evol.* 2021;7(2):veab064. DOI: <https://doi.org/10.1093/ve/veab064>
21. Lan J., Ge J., Yu J., et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature.* 2020;581(7807):215–20. DOI: <https://doi.org/10.1038/s41586-020-2180-5>
22. O’Toole Á., Pybus O.G., Abram M.E., et al. Pango lineage designation and assignment using SARS-CoV-2 spike gene nucleotide sequences. *BMC Genomics.* 2022;23(1):121. DOI: <https://doi.org/10.1186/s12864-022-08358-2>
23. Komissarov A.B., Safina K.R., Garushyants S.K., et al. Genomic epidemiology of the early stages of the SARS-CoV-2 outbreak in Russia. *Nat. Commun.* 2021;12(1):649. DOI: <https://doi.org/10.1038/s41467-020-20880-z>
24. Liu Y., Liu J., Plante K.S., et al. The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. *Nature.* 2022;602(7896):294–9. DOI: <https://doi.org/10.1038/s41586-021-04245-0>
25. Lista M.J., Winstone H., Wilson H.D., et al. The P681H mutation in the spike glycoprotein of the Alpha variant of SARS-CoV-2 escapes IFITM restriction and is necessary for type I Interferon resistance. *J. Virol.* 2022;96(23):e0125022. DOI: <https://doi.org/10.1128/jvi.01250-22>
26. Борисова Н.И., Котов И.А., Колесников А.А. и др. Мониторинг распространения вариантов SARS-CoV-2 (Coronaviridae: Coronavirinae: Betacoronavirus; Sarbecovirus) на территории московского региона с помощью таргетного высокопроизводительного секвенирования. *Вопросы вирусологии.* 2021;66(4):269–78. Borisova N.I., Kotov I.A., Kolesnikov A.A., et al. Monitoring the spread of the SARS-CoV-2 (Coronaviridae: Coronavirinae: Betacoronavirus; Sarbecovirus) variants in the Moscow region using targeted high-throughput sequencing. *Problems of Virology.* 2021;66(4):269–78. DOI: [10.36233/0507-4088-72](https://doi.org/10.36233/0507-4088-72) EDN: <https://elibrary.ru/qdsujp>
27. Aksamentov I., Roemer C., Hodcroft E., Neher R. Nextclade: clade assignment, mutation calling and quality control for viral genomes. *J. Open Source Softw.* 2021;6(67):3773. DOI: <https://doi.org/10.21105/joss.03773>
28. Kumar R., Srivastava Y., Muthuramalingam P., et al. Understanding mutations in human SARS-CoV-2 spike glycoprotein: a systematic review & meta-analysis. *Viruses.* 2023;15(4):856. DOI: <https://doi.org/10.3390/v15040856>
29. Xiao B., Wu L., Sun Q., et al. Dynamic analysis of SARS-CoV-2 evolution based on different countries. *Gene.* 2024;916:148426. DOI: <https://doi.org/10.1016/j.gene.2024.148426>
30. Fernandes Q., Inchakalody V.P., Merhi M., et al. Emerging COVID-19 variants and their impact on SARS-CoV-2 diagnosis, therapeutics and vaccines. *Ann. Med.* 2022;54(1):524–40. DOI: <https://doi.org/10.1080/07853890.2022.2031274>

31. Raghwani J., du Plessis L., McCrone J.T., et al. Genomic epidemiology of early SARS-CoV-2 transmission dynamics, Gujarat, India. *Emerg. Infect. Dis.* 2022;28(4):751–8. DOI: <https://doi.org/10.3201/eid2804.212053>
32. Kotov I., Saenko V., Borisova N., et al. Effective approaches to study the genetic variability of SARS-CoV-2. *Viruses.* 2022;14(9):1855. DOI: <https://doi.org/10.3390/v14091855>

Information about the authors

Ivan A. Kotov — bioinformatician, Laboratory for genomic research, Central Research Institute for Epidemiology, Moscow, Russia; postgraduate student, Moscow Institute of Physics and Technology (National Research University), Dolgoprudny, Russia, <https://orcid.org/0000-0003-2416-5689>

Matvey R. Agletdinov — bioinformatician, Laboratory for genomic research, Central Research Institute for Epidemiology, Moscow, Russia; postgraduate student, Moscow Institute of Physics and Technology (National Research University), Dolgoprudny, Russia, <https://orcid.org/0000-0003-2249-7196>

German V. Roev — bioinformatician, Laboratory for genomic research, Central Research Institute for Epidemiology, Moscow, Russia; postgraduate student, Moscow Institute of Physics and Technology (National Research University), Dolgoprudny, Russia, <https://orcid.org/0000-0002-2353-5222>

Ekaterina V. Pimkina — junior researcher, Scientific group of genetic technology, Laboratory for genomic research, Central Research Institute for Epidemiology, Moscow, Russia, <https://orcid.org/0000-0002-0591-3525>

Maksim I. Nadtoka — researcher, Scientific group of genetic technology, Laboratory for genomic research, Central Research Institute for Epidemiology, Moscow, Russia, <https://orcid.org/0009-0002-3217-0963>

Arina V. Peresadina — bioinformatician, Laboratory for genomic research, Central Research Institute for Epidemiology, Moscow, Russia, <https://orcid.org/0009-0002-4981-6716>

Anna Yu. Bukharina — junior researcher, Scientific group of genetic technology, Laboratory for genomic research, Central Research Institute for Epidemiology, Moscow, Russia, <https://orcid.org/0000-0002-6892-3595>

Dmitry V. Svetlichny — junior researcher, Scientific group of genetic technology, Laboratory for genomic research, Central Research Institute for Epidemiology, Moscow, Russia, <https://orcid.org/0009-0008-6864-2807>

Sergey E. Goncharov — laboratory assistant-researcher, Scientific group of genetic technology, Laboratory for genomic research, Central Research Institute for Epidemiology, Moscow, Russia, <https://orcid.org/0009-0000-2842-5818>

Anastasiia V. Vykhodtseva — technologist, Scientific group of genetic technology, Laboratory for genomic research, Central Research Institute for Epidemiology, Moscow, Russia, <https://orcid.org/0009-0005-1911-9620>

Nadezhda I. Borisova — junior researcher, Scientific group of genetic technology, Laboratory for genomic research, Central Research Institute for Epidemiology, Moscow, Russia, <https://orcid.org/0000-0002-9672-0648>

Vladislav G. Lysenkov — bioinformatician, Laboratory for genomic research, Central Research Institute for Epidemiology, Moscow, Russia, <https://orcid.org/0000-0002-1468-1631>

Mikhail D. Chanyshev — Cand. Sci. (Biol.), senior researcher, Laboratory for genomic research, Central Research Institute for Epidemiology, Moscow, Russia, <https://orcid.org/0000-0002-6943-2915>

David N. oglu Agabalaev — bioinformatician, Laboratory for genomic research, Central Research Institute for Epidemiology, Moscow, Russia, <https://orcid.org/0000-0001-8125-355X>

Valeria V. Saenko — researcher, Scientific group of genetic technology, Laboratory for genomic research, Central Research Institute for Epidemiology, Moscow, Russia, <https://orcid.org/0000-0003-0952-0830>

Anna S. Cherkashina — Cand. Sci. (Chem.), Head, Scientific group of genetic engineering and biotechnology, Central Research Institute for Epidemiology, Moscow, Russia, <https://orcid.org/0000-0001-7970-7495>

Информация об авторах

Котов Иван Андреевич — м.н.с. лаб. геномных исследований ЦНИИ Эпидемиологии Роспотребнадзора, Москва, Россия; аспирант МФТИ (НИУ), Долгопрудный, Россия, <https://orcid.org/0000-0003-2416-5689>

Аглетдинов Матвей Рашидович — биоинформатик лаб. геномных исследований ЦНИИ Эпидемиологии Роспотребнадзора, Москва, Россия; аспирант МФТИ (НИУ), Долгопрудный, Россия, <https://orcid.org/0000-0003-2249-7196>

Роев Герман Викторович — биоинформатик лаб. геномных исследований ЦНИИ Эпидемиологии Роспотребнадзора, Москва, Россия; аспирант МФТИ (НИУ), Долгопрудный, Россия, <https://orcid.org/0000-0002-2353-5222>

Пимкина Екатерина Владимировна — м.н.с. научной группы геномных технологий лаб. геномных исследований ЦНИИ Эпидемиологии Роспотребнадзора, Москва, Россия, <https://orcid.org/0000-0002-0591-3525>

Надтока Максим Игоревич — н.с. научной группы геномных технологий лаб. геномных исследований ЦНИИ Эпидемиологии Роспотребнадзора, Москва, Россия, <https://orcid.org/0009-0002-3217-0963>

Пересадина Арина Валерьевна — биоинформатик лаб. геномных исследований ЦНИИ Эпидемиологии Роспотребнадзора, Москва, Россия, <https://orcid.org/0009-0002-4981-6716>

Бухарина Анна Юрьевна — м.н.с. научной группы геномных технологий лаб. геномных исследований ЦНИИ Эпидемиологии Роспотребнадзора, Москва, Россия, <https://orcid.org/0000-0002-6892-3595>

Светличный Дмитрий Васильевич — м.н.с. научной группы геномных технологий лаб. геномных исследований ЦНИИ Эпидемиологии Роспотребнадзора, Москва, Россия, <https://orcid.org/0009-0008-6864-2807>

Гончаров Сергей Евгеньевич — лаборант-исследователь научной группы геномных технологий лаб. геномных исследований ЦНИИ Эпидемиологии Роспотребнадзора, Москва, Россия, <https://orcid.org/0009-0000-2842-5818>

Выходцева Анастасия Владимировна — технолог научной группы геномных технологий лаб. геномных исследований ЦНИИ Эпидемиологии Роспотребнадзора, Москва, Россия, <https://orcid.org/0009-0005-1911-9620>

Борисова Надежда Ивановна — м.н.с. научной группы геномных технологий лаб. геномных исследований ЦНИИ Эпидемиологии Роспотребнадзора, Москва, Россия, <https://orcid.org/0000-0002-9672-0648>

Лысенков Владислав Геннадиевич — биоинформатик лаб. геномных исследований ЦНИИ Эпидемиологии Роспотребнадзора, Москва, Россия, <https://orcid.org/0000-0002-1468-1631>

Чанышев Михаил Дамирович — к.б.н., с.н.с. лаб. геномных исследований ЦНИИ Эпидемиологии Роспотребнадзора, Москва, Россия, <https://orcid.org/0000-0002-6943-2915>

Агабалаев Давид Накам оглы — биоинформатик лаб. геномных исследований ЦНИИ Эпидемиологии Роспотребнадзора, Москва, Россия, <https://orcid.org/0000-0001-8125-355X>

Саенко Валерия Владимировна — н.с. научной группы геномных технологий лаб. геномных исследований ЦНИИ Эпидемиологии Роспотребнадзора, Москва, Россия, <https://orcid.org/0000-0003-0952-0830>

Черкашина Анна Сергеевна — к.х.н., рук. научной группы генной инженерии и биотехнологии ЦНИИ Эпидемиологии Роспотребнадзора, Москва, Россия, <https://orcid.org/0000-0001-7970-7495>

Семененко Татьяна Анатольевна — д.м.н., профессор, научный консультант отдела эпидемиологии НИЦ эпидемиологии и микробиологии им. почетного академика Н.Ф. Гамалеи, Москва, Россия, <https://orcid.org/0000-0002-6686-9011>

Tatiana A. Semenenko — D. Sci. (Med.), Professor, Scientific advisor, Epidemiology department, National Research Center of Epidemiology and Microbiology named after Honorary Academician N.F. Gamaleya, Moscow, Russia, <https://orcid.org/0000-0002-6686-9011>

Dmitry V. Dubodelov — senior researcher, Laboratory of viral hepatitis, Central Research Institute for Epidemiology, Moscow, Russia, <https://orcid.org/0000-0003-3093-5731>

Kamil F. Khafizov — Cand. Sci. (Biol.), Head, Laboratory for genomic research, Central Research Institute for Epidemiology, Moscow, Russia, khafizov@cmd.su, <https://orcid.org/0000-0001-5524-0296>

Vasily G. Akimkin — D. Sci. (Med.), Professor, Full Member of RAS, Director, Central Research Institute for Epidemiology, Moscow, Russia, <https://orcid.org/0000-0003-4228-9044>

Author contributions: *Kotov I.A.* — methodology, software, data curation, investigation, formal analysis, visualization, writing of original draft and review, editing; *Akletdinov M.R.*, *Roev G.V.* — methodology, software, data curation, investigation, formal analysis, writing of review, editing; *Pimkina E.V.*, *Nadtoka M.I.* — investigation, validation; *Peresadina A.V.* — software, data curation, investigation, formal analysis, visualization; *Bukharina A.Yu.*, *Svetlichny D.V.*, *Goncharov S.E.*, *Vykhodtseva A.V.*, *Agabalaev D.N.*, *Cherkashina A.S.* — investigation; *Borisova N.I.* — investigation, validation; *Lysenkov V.G.* — software, data curation; *Chanyshev M.D.* — investigation, writing of review, editing; *Saenko V.V.* — investigation, validation, supervision; *Semenenko T.A.* — writing of review, editing; *Dubodelov D.V.* — data curation; *Khafizov K.F.* — conceptualization, supervision, project administration, writing of original draft and review, editing, investigation, methodology, software, data curation; *Akimkin V.G.* — conceptualization, resources, supervision, project administration, writing of review, editing, funding acquisition. All authors confirm that they meet the International Committee of Medical Journal Editors criteria for authorship, made a substantial contribution to the conception of the article, acquisition, analysis, interpretation of data for the article, drafting and revising the article, final approval of the version to be published.

The article was submitted 24.06.2024;
accepted for publication 10.08.2024;

Дубоделов Дмитрий Васильевич — к.м.н., с.н.с. лаб. вирусных гепатитов ЦНИИ Эпидемиологии Роспотребнадзора, Москва, Россия, <https://orcid.org/0000-0003-3093-5731>

Хафизов Камиль Фаридович — к.б.н., зав. лаб. геномных исследований ЦНИИ Эпидемиологии Роспотребнадзора, Москва, Россия, khafizov@cmd.su, <https://orcid.org/0000-0001-5524-0296>

Акимкин Василий Геннадьевич — д.м.н., профессор, академик РАН, директор ЦНИИ Эпидемиологии Роспотребнадзора, Москва, Россия, <https://orcid.org/0000-0003-4228-9044>

Участие авторов: *Котов И.А.* — методология, разработка программного обеспечения, курирование данных, проведение исследований, формальный анализ, визуализация результатов, написание оригинального черновика и обзора, редактирование; *Аглетдинов М.Р.*, *Роев Г.В.* — методология, разработка программного обеспечения, курирование данных, проведение исследований, формальный анализ, написание обзора и редактирование; *Пимкина Е.В.*, *Надтока М.И.* — проведение исследований, валидация результатов; *Пересадина А.В.* — разработка программного обеспечения, курирование данных, проведение исследований, формальный анализ, визуализация результатов; *Бухарина А.Ю.*, *Светличный Д.В.*, *Гончаров С.Е.*, *Выходцева А.В.*, *Агабалаев Д.Н.*, *Черкашина А.С.* — проведение исследований; *Борисова Н.И.* — проведение исследований, валидация результатов; *Лысенков В.Г.* — программное обеспечение, курирование данных; *Чанышев М.Д.* — проведение исследований, написание обзора и редактирование; *Саенко В.В.* — проведение исследований, валидация результатов, руководство; *Семененко Т.А.* — написание обзора и редактирование; *Дубоделов Д.В.* — курирование данных; *Хафизов К.Ф.* — концептуализация, руководство, администрирование проекта, написание оригинального черновика и обзора, редактирование, проведение исследований, методология, программное обеспечение, курирование данных; *Акимкин В.Г.* — концептуализация, ресурсы, руководство, администрирование проекта, написание обзора и редактирование, привлечение финансирования. Все авторы подтверждают соответствие своего авторства критериям Международного комитета редакторов медицинских журналов, внесли существенный вклад в проведение поисково-аналитической работы и подготовку статьи, прочли и одобрили финальную версию до публикации.

Статья поступила в редакцию 24.06.2024;
принята к публикации 10.08.2024;
опубликована 29.08.2024