

Scientific Session of the General Meeting of the RAS Members “The Role of Science in Overcoming Pandemics and Postcrisis Development of Society”

COVID-19 Epidemiology and Diagnosis: Monitoring Evolutionary Changes in the SARS-CoV-2 Virus

V. G. Akimkin[#]

Central Research Institute of Epidemiology, Federal Service for Oversight of Consumer Rights Protection and Human Wellbeing,
Moscow, Russia

e-mail: vgakimkin@yandex.ru

Received December 29, 2021; revised January 10, 2022; accepted February 28, 2022

Abstract—This article presents the results of scientific research on the epidemiological features of the new coronavirus infection (COVID-19) and the molecular genetic characteristics of SARS-CoV-2 virus genovariants, based on the experience of the Central Research Institute of Epidemiology of the Federal Service for Oversight of Consumer Rights Protection and Human Wellbeing (Rosпотребнадзор).

Keywords: new coronavirus infection, COVID-19, epidemiology, molecular diagnostics, SARS-CoV-2

DOI: 10.1134/S1019331622040013

Since December 2019, the pandemic of a new coronavirus infection, COVID-19, has been developing in the world; it is caused by an enveloped single-stranded RNA virus SARS-CoV-2, belonging to the family Sarbecovirus of the genus *Betacoronavirus* [1–4]. Data on this virus are contained in the GISAID (<https://www.gisaid.org>) and GenBank (<https://www.ncbi.nlm.nih.gov>) genome databases. In Russia, the corresponding VGARuS database (<https://genome.crie.ru/app/index>) was created at the Central Research Institute of Epidemiology. Phylogenetic analysis has been carried out using the Nexstrain (<https://nextstrain.org>) and Pangolin (<https://cov-lineages.org/resources/pangolin.html>) classifiers.

The main transmission mechanism of SARS-CoV-2 is aerogenic. Aerosol microparticles with dimensions of 1 to 10 μm can penetrate the deep sections of the respiratory tract, playing the most significant role in the spread of COVID-19. The second most important transmission mechanism is contact [5, 6]. Hence, it became necessary to limit the social contacts of the infected and to improve medical control over their condition.

On March 11, 2020, the World Health Organization recognized that the spread of the new coronavirus infection had become a pandemic. As of December 13, 2021, in terms of the total number of detected cases among the regions of the world, the United States

ranked first (97679255), and the largest number of deaths was also recorded there (2360315). The highest increase in detected cases for the week from December 5 to 12, 2021, was recorded in Europe (2687257) and in the Americas (935062). The maximum increase in deaths for the same week was registered in Europe and the American region (28990 and 12987, respectively).

For the world, according to data available as of December 13, 2021, over 270 million confirmed cases of COVID-19 had been recorded, of which about 5.3 million were fatal. The Russian Federation ranks 5th in the world in terms of the total number of registered cases, although in terms of the overall morbidity per 100000 people, our country is in 77th place; in terms of the absolute number of deaths, it ranks 5th; in terms of lethality, 37th; and in terms of mortality per 100000 people, 36th.

According to statistics, as of December 13, 2021, 10046454 cases had been registered in the Russian Federation, 8769916 people had recovered, 290604 deaths had been recorded, more than 232.1 million tests had been carried out, and 61074945 people were fully vaccinated (135478571 vaccinations).

Analysis of the manifestations of the epidemic process of COVID-19 in the Russian Federation for 2020–2021 made it possible to identify four increases in the incidence rate (Fig. 1). In the age structure of the diseased, people of three ages predominated in total: middle (30–49 years old), 33–34%; older (from 50 to 64 years old), 24.5–26.5%; and the elderly (over 65 years old), 18–21%. This is partly because young

[#] RAS Academician Vasilii Gennad'evich Akimkin is Director of the Central Research Institute of Epidemiology of the Federal Service for Oversight of Consumer Rights Protection and Human Wellbeing (Rosпотребнадзор CRIE).

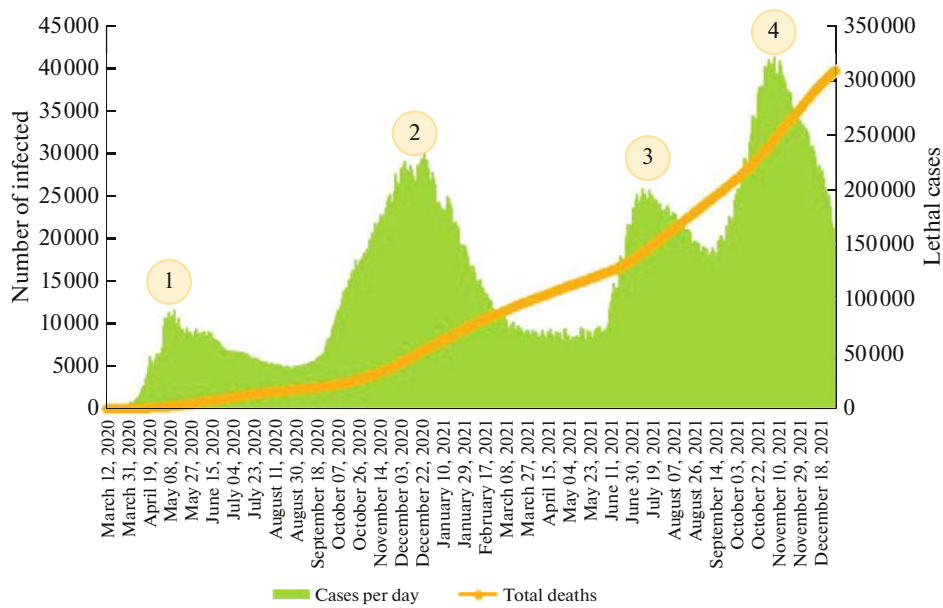


Fig. 1. Dynamics of changes in the absolute number of daily registered cases and the number of deaths in the Russian Federation.

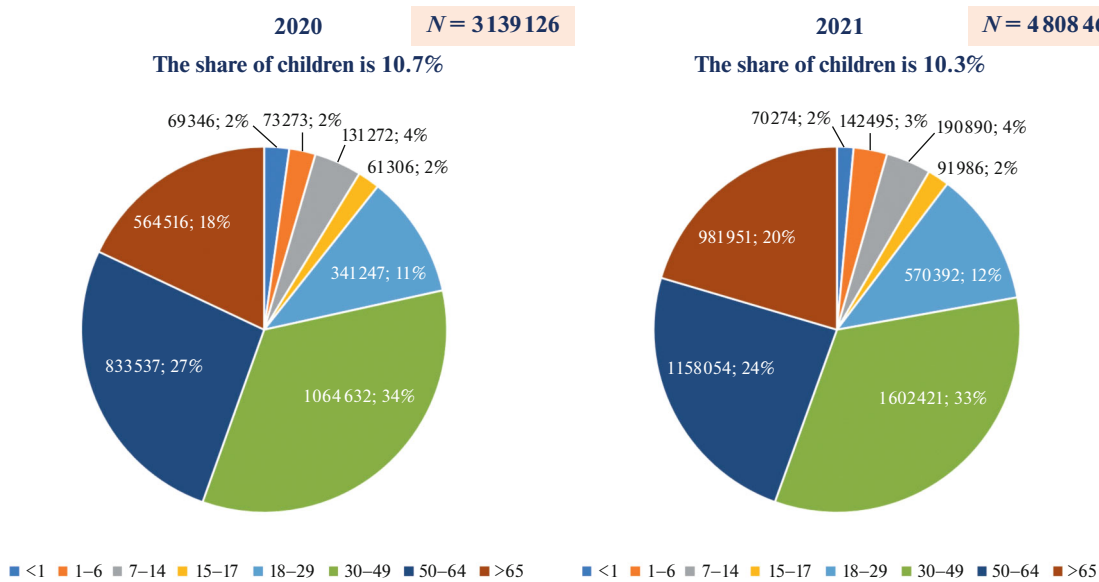


Fig. 2. Age structure of COVID-19 cases in the Russian Federation for 2020 and 2021.

people and children usually do fine with COVID-19; they constitute a higher proportion of asymptomatic patients who do not resort to testing, do not seek medical help, and are not registered in the statistics, but at the same time may infect others. The proportion of patients under the age of 17 was 19–20% (Fig. 2).¹

Laboratory diagnosis of the new coronavirus infection is carried out in accordance with the requirements

¹ Full-color versions of Figs. 2–4 are available in the electronic version of the journal Vestnik RAN on the IKTs Akademkniga website.

of the Sanitary and Epidemiological Rules SP 1-P. The biological material is a swab taken from the nasopharynx. When determining the pathogenicity group, the method of nucleic acid amplification is mainly used, specifically, polymerase chain reaction (PCR) with reverse transcription (conversion of RNA molecules into DNA molecules) in real time (real-time RT-PCR). PCR is the “gold standard” in diagnostics, but recently, along with it, methods of isothermal amplification of nucleic acids have been actively introduced.

The COVID-19 pandemic has created an urgent need for available mass testing, which quantitatively

exceeds the capabilities of the existing medical and laboratory infrastructure using traditional diagnostic methods by orders of magnitude. In detecting the disease at an early stage, the speed of the tests is also very important. Tests based on RT-PCR require a minimum of 3–4.5 h, and the time from the moment the material arrives at the laboratory until the official result is issued can be 8–12 h. In this context, the issue of introducing new methods of rapid diagnostics has become topical.

In terms of testing volume, Russia, as of December 1, 2021, ranks 4th in the world after Britain, India, and the United States, while we are in 5th place per 1000 inhabitants.

In 2020–2021, CRIE staff conducted 3.5 million PCR examinations; compared with previous years, their volume increased by 50 times (!). More than one million enzyme-linked immunosorbent assay (ELISA) tests were performed to determine antibodies to COVID-19. More than 20000 SARS-CoV-2 genome sequencings were conducted.

The Central Research Institute of Epidemiology was one of the first in our country to register the AmpliSens® Cov-Bat-FL PCR-based test system for detecting SARS-CoV-2 RNA, which is highly sensitive (0.1 TCID₅₀/mL,² 1000 RNA copies/mL) and specific. Using it, we can diagnose not only SARS-CoV-2 but also SARS and MERS.

The above-mentioned nucleic acid amplification methods are the most sensitive and specific currently available means to detect SARS-CoV-2 RNA [7–9]. In April 2021, the AmpliSens® COVID-19-FL test system based on the nucleic acid amplification method for the quantitative determination of virus RNA using RT-PCR was registered. In terms of sensitivity, it is one of the best in the world. Its sensitivity is 5×10 GE³ (copies)/mL with a wide measurement range: $5 \times 10 - 3 \times 10$ GE (copies)/mL, and the specificity is practically 100%. This makes it possible to detect the presence of the virus, track its changes, and determine the viral load, which is important for predicting the course of the disease in hospitalized patients. Automatic RNA extraction reduces the testing time by 4–5 times. Thus, when using the CFX96 cycler, it is 93 min; Rotor-Gene, 99 min and DT-96, 101 min. Determining the viral load makes it possible to assess the likelihood of transmission of the pathogen through contact with the infected; the epidemiological risk decreases as the amount of RNA in the sample decreases.

Loop-mediated Isothermal amplification (LAMP) is the most popular and effective among the new

express diagnostic methods. CRIE associates developed the LAMP-based AmpliSens® SARS-CoV-2-GG test system, which has several advantages: it makes it possible to reduce the study time by 3–4 times without reducing the sensitivity and specificity of classical RT-PCR (100%). Amplification takes 20 min, and the total testing time, together with virus RNA isolation, is about one hour, which is much faster than with classical PCR testing. In addition, it is possible to use simpler equipment (thermostat) and not only expensive amplifiers. Note that the kit was developed on the basis of the Central Research Institute of Epidemiology's own enzyme base, which makes it possible not to depend on the supply of reagents from abroad and to keep prices low.

Specialists of Rospotrebnadzor and other departments are conducting large-scale work on sequencing and bioinformatic analysis of the SARS-CoV-2 genome, identifying both known and potentially new genovariants. Whole genome sequencing makes it possible to detect all possible mutations in the genome. Fragment sequencing, on the other hand, usually focuses on key changes in the genome that are responsible for such properties of the pathogen that would be more successful for the virus.

In July 2021, scientists from our institute developed the AmpliSens® SARS-CoV-2-N501Y-IT NGS panel to identify the most significant mutations in the S-protein gene. This is a fast, efficient, and cheap sequencing method: the primers were positioned so that they “stick” on conservative areas, and the read sections cover all significant changes. At that time, the genovariants Alpha, British (B.1.1.7), Beta, South African (B.1.351), and Brazilian (P.1) were known. Our panel also identifies other epidemiologically significant strains of the N501Y virus, including Delta/Kappa, Indian (B.1.617.1/B.1/617.2), which is now ubiquitous. The panel makes it possible to examine a large number of samples and determine the belonging of isolates to specific genovariants.

It is clear that SARS-CoV-2 will continue to mutate, and new changes may not get into readable areas; hence, we have developed a second version of the panel. The S-protein gene, which includes about 4000 sequences, is completely covered, which makes it possible to detect any changes in it. It has become possible to determine all strains, while the cost remains quite low (€3000 per sample). The panel also makes it possible to increase the speed of sequencing compared to other commercially available products. Preparation takes about a day, sequencing itself takes less than six hours, and as a result, reading the S-protein gene takes less than three days.

To reduce the likelihood of false negative results associated with new strains, CRIE specialists have developed a new test system that is less sensitive to changes in SARS-CoV-2 due to the simultaneous use of two targets in the virus genome. Since the test uses

² TCID₅₀, mL, is a quantitative unit of the viral particle titer determined by the final dilution required for killing 50% of infected hosts or producing a cytopathic effect in 50% of inoculated tissue culture cells under laboratory conditions.

³ GE, mL, is the concentration of the genomic equivalent (GE) of a microorganism in 1 mL of a sample of the test material.

nonspecific detection of amplification products, the simultaneous use of two targets slightly affects the cost of the test but at the same time can significantly reduce the number of false negative results. The test itself is based on loop isothermal amplification technology, which makes testing much faster and easier than traditional PCR. In parallel, the Institute regularly analyzes the genomic sequences of the new coronavirus obtained during sequencing by institutes of Rospotrebnadzor and other departments not to miss the moment of such mutations in targeted regions of the genome, the frequency of which will begin to increase in the general viral population. Using two targets in the test will save time: at least one region remains unchanged, and the developers have time to correct the primers in the mutated locus. Several target areas in quick tests have never been used in Russia before, and their use abroad is known only in the form of scientific publications.

During the pandemic period (2020–2021), the Central Research Institute of Epidemiology of Rospotrebnadzor complied with seven orders of the Russian Government, within which the production and supply of diagnostic test systems are fully ensured, including more than 27 million for testing to detect SARS-CoV-2 RNA, more than 28 million for extraction of nucleic acids, and more than 21 million for obtaining complementary DNA on a DNA matrix. In total, more than 76 million studies have been completed. Almost every third test system in Russia for diagnosing SARS-CoV-2 was produced at the Central Research Institute of Epidemiology. Note that within the framework of international cooperation, our institute supplies diagnostic test systems to 40 countries of the world.

Monitoring changes in the properties of the pathogen is the most important area of epidemiological surveillance for COVID-19 (Decree of the Government of the Russian Federation no. 448 of March 23, 2021). Rospotrebnadzor specialists have launched large-scale studies (sequencing and bioinformatics analysis) of the SARS-CoV-2 coronavirus genomes, during which both known and potentially new variants of the pathogen are identified.

On the basis of the Central Research Institute of Epidemiology, the Russian platform VGARuS has been deployed to analyze the genome of the SARS-CoV-2 virus [10]. Epidemiological analysis of the distribution of its genovariants for the territory of the Russian Federation is carried out on the basis of the VGARuS database and information from Rospotrebnadzor organizations that carry out sequencing. A web platform has been created to download and analyze data and display sequencing results, and bioinformatic data analysis tools have been developed to identify mutations and display their belonging to epidemiologically significant strains. The deposition of sequencing data, including metadata, continues. As of December 13, 2021, more than 46000 pathogen sequences had been

uploaded into the database, including 22908 full genome sequences and 23841 genome fragment sequences. In addition, 10872 sequences (10576 full genomes, 296 genome fragments) have been uploaded to GISAID⁴ in Russia. More than 30000 samples belonging to genovariants other than Wuhan have been identified. A scientific consortium has been created, including 28 organizations that conduct sequencing and upload data to the platform.

CRIE specialists have repeatedly modified the NGS panel to detect significant mutations in the S-protein gene [11], which includes all known significant mutations in the S-protein gene, including those inherent in the Omicron strain. A special modification of primers eliminates the need for a ligation step, significantly reducing the cost and duration of sample preparation, and allows adding new genome fragments to the panel. The NGS panel for sequencing the entire S-protein gene covers ~4000 nucleotide sequences and allows one to detect any possible mutations in the gene and determine all SARS-CoV-2 genovariants; it is suitable for mass screening of samples.

At the end of September 2021, at a press conference of the World Health Organization, it was reported that the Delta variant had finally supplanted all other genovariants around the world. The list of genovariants was shortened, and since December 6, 2021, the WHO has been highlighting the following:

- VOC, Variants of Concern (those that contain dangerous mutations, have caused large-scale outbreaks);
- VUI/VOI, Variants of Interest (those that contain dangerous mutations, caused/may cause local outbreaks or multiple clusters in many countries);
- VUM, Variants Under Monitoring (those that contain dangerous or presumably dangerous mutations; their properties and ability to spread are still unclear) [10].

Here is a list of WHO genovariants, current as of December 3, 2021.

The first category (VOC) is variants of concern: Alpha (British, B.1.1.7), Beta (South African B.1.351), Gamma (Brazilian, P.1), Delta (Indian-2, B.1.617.2 + AY.*), and Omicron (B.1.1.529).

The second category (VOI) is variants of interest: Lambda (C.37), Mu (B.1.621).

The third category (VUM) is variants under monitoring: AZ.5, C.1.2, Kappa (B.1.617.1), Iota (B.1.526), Eta (B.1.525), B.1.630, B.1.640.

On November 26, 2021, the WHO attributed the Omicron variant to Category 1 (VOC). As of December 6, 2021, 582 genomes of the Omicron strain (B.1.1.529), more than 45 mutations in the entire

⁴ GISAID (Global Initiative on Sharing All Influenza Data) is an international open platform for collecting information on viral genomes, created in 2008 after the H5N1 avian influenza outbreak.

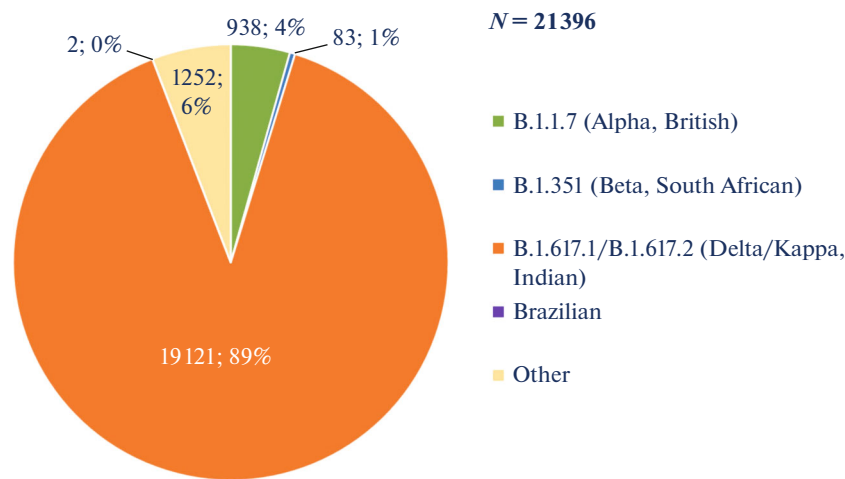


Fig. 3. Structure of SARS-CoV-2 genovariants detected on the territory of the Russian Federation (as of December 3, 2021).

genome, and more than 30 mutations in the S-protein gene were published in the GISAID database.

Let us now turn to the location of sequences for the VOC and VOI categories on the phylogenetic tree and the detectability of genovariants on the territory of the Russian Federation: the Alpha genovariant (British, B.1.1.7) was distributed throughout the country in the winter of 2021; the Delta genovariant (Indian-2, B.1.617.2 + AY.*) spread throughout the country in the second half of April 2021 and dominated as of December 14, 2021; the Beta genovariant (South African, B.1.351) occurred in winter 2021 at the same time as Alpha but was not spread; the genovariant Gamma (Brazilian, P.1) occurred at the beginning of 2021 and did not receive a noticeable distribution; the genovariant Lambda (C.37) and the genovariant Mu (B.1.621) have no epidemic significance on the territory of the Russian Federation today; and the genovariant Omicron (B.1.1.529) was found on the territory of our country in citizens arriving from abroad. The SARS-CoV-2 BA.1 variant was detected almost simultaneously in South Africa, Botswana, and Hong Kong (according to GISAID). The first genomes are dated November 9 and 11, 2021 (from South Africa), November 11 (from Botswana), and November 13 (from Hong Kong). Later, Omicron was detected in Israel (November 20), Canada (November 23), Belgium (November 24), and Australia (November 28) (indicated by the date of biomaterial sampling). As of December 27, 2021, according to GISAID data, a total of 53 737 genome sequences of the Omicron strain were known worldwide.

According to the national genomic database VGARus, among the SARS-CoV-2 sequences classified by the Pangolin program (version 3.1.17) as Omicron (BA.1), 89 sequences were found as of December 27, 2021.

According to the monitoring of the circulation of SARS-CoV-2 genotypes, the genetic lineage Delta (B.1.617.2+AY.*) dominates on the territory of the Russian Federation; since April 2021, it has been

assigned to the VOC group. At the beginning of December 2021, the share of Delta in the structure of identified SARS-CoV-2 genovariants remained at the level of 89% (Fig. 3). This genetic lineage dominated on the territory of all federal districts and subjects of the Russian Federation. According to our updated data, its share in the structure of Delta subvariants fluctuated: 82.7–88.5% in September, 82–93.5% in October, 91.2% in the first week of November, 85.7% in the second week of November. On the territory of the Russian Federation, the Delta variant became dominant in April–May 2021. In the structure of subvariants of the Delta lineage (B.1.617.2 + AY.*) isolated in the Russian Federation, the subvariant AY.122 (81%) prevails (Fig. 4).

As of the beginning of December 2021, in addition to AY.122, such sublineages as AY.82, B.1.617.2 prevailed on the territory of the Russian Federation. Note that more than 30 sublineages of Delta genovariants have been registered in Russia, and we continue to monitor the genovariants on a regular basis.

Within the framework of Resolution of the Russian Government no. 452 of March 27, 2021, and Instruction of the Russian Government no. 635-r of March 16, 2020 (as amended of December 10, 2021), the integration platform “SOLAR” was created on the basis of the Central Research Institute of Epidemiology for the rapid transmission of PCR results to all interested citizens of the Russian Federation; from the moment the information is uploaded into the database, this takes less than 60 min. From November 1, 2020, to December 13, 2021, regions transmitted 56740613 results to the platform. More than 1000 laboratories (including network ones) are connected to the platform; 85 regions transmit data automatically. More than 427000 results are transmitted daily to the Unified Portal of Public Services of the Russian Federation. The “I travel without COVID-19” application has also been created; 3200 biomaterial sampling points are connected to it; 6000 results are uploaded monthly.

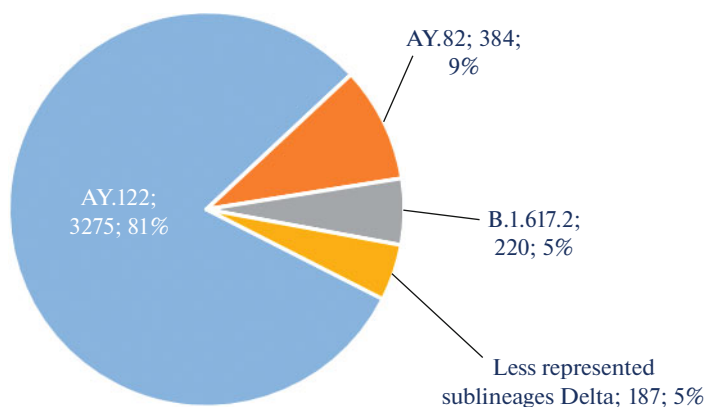


Fig. 4. Structure of subvariants of the Delta lineage (B.1.617.2 + AY.*) detected in the Russian Federation. *Note:* Sampling of bio-material from October 5 to December 6, 2021, according to the VGARus database, $N = 4066$.

* * *

Summing up, let us note the following.

As of mid-December 2021, the active process of evolution of the COVID-19 pathogen was continuing. The genetic variant Delta (B.1.617.2 + AY.*) dominated the territory of the Russian Federation, its share among all identified variants being more than 95%, while its dissociation was noted. During all the months of observation after the start of the Delta genovariant registration, the dominant variant was AY.122, named so by the Pangolin classifier on November 26, 2021. The share of AY.122 in the structure of Delta subvariants in Russia was 85–95%. In addition to AY.122, the most common Delta subvariants were the “parent” B.1.617.2 (its share from October to December 2021 was 5%), AY.82 (4%), and AY.99 (2%). Other subvariants (there are 30 of them) are represented by single samples and totaled about 1% of all Delta sequences.

The dynamic change in the proportion of these subvariants in the overall structure in recent months has not affected the intensity of the manifestations of the epidemic process or the severity of the clinical course of the disease.

Genovariants that require special attention, in particular Lambda (C. 37) and Mu (B.1.621), were practically absent on the territory of the Russian Federation and did not form epidemic foci. In connection with the increase in the number of cases of the disease caused by the Omicron variant detected in different countries of the world, it is advisable to continue monitoring this genovariant among people entering the territory of the Russian Federation from epidemiologically disadvantaged countries.

Molecular genetic monitoring of circulating genovariants should remain the basis of epidemiological surveillance for COVID-19, and its results must be considered when creating vaccine preparations.

CONFLICT OF INTEREST

The author declares that he has no conflicts of interest.

REFERENCES

1. Severe acute respiratory syndrome coronavirus 2. <https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=2697049>
2. A. P. Fehr and S. Perlman, “Coronaviruses: An overview of their replication and pathogenesis,” *Methods Mol. Biol.* **1282**, 1–23 (2015).
3. G. S. Randhawa, M. P. M. Soltysiak, H. E. Roz, et al., “Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study,” *PLOS One* **15** (4), e0232391 (2020).
4. M. Yu. Shchelkanov, A. Yu. Popova, V. G. Dedkov, et al., “History of study and modern classification of coronaviruses (Nidovirales: Coronaviridae),” *Infektsiya Immunitet*, No. 2, 221–246 (2020).
5. V. Stadnytskyi, C. E. Bax, A. Bax, et al., “The airborne lifetime of small speech droplets and their potential importance in SARS-CoV-2 transmission,” *PNAS* **117** (22), 11875–11877 (2020). <https://doi.org/10.1073/pnas.2006874117>
6. S. Asadi, A. S. Wexler, C. D. Cappa, et al., “Effect of voicing and articulation manner on aerosol particle emission during human speech,” *PLOS One* **15** (1), e0227699 (2020).
7. L. Geddes, “Puzzle over viral load,” *New Sci.* **245** (3276), 8 (2020).
8. G. M. Joynt and W. K. Wu, “Understanding COVID-19: What does viral RNA load really mean?,” *Lancet Infect. Dis.* **20**, 635–636 (2020).
9. B. La Scola, M. Le Bideau, J. Andreani, et al., “Viral RNA load as determined by cell culture as a management tool for discharge of SARS-CoV-2 patients from infectious disease wards,” *Eur. J. Clin. Microbiol. Infect. Dis.* **39**, 1059–1061 (2020).
10. <https://www.genome.crie.ru/app/index>. Cited December 13, 2021.
11. N. I. Borisova, I. A. Kotov, A. A. Kolesnikov, et al., “Monitoring the spread of the SARS-CoV-2 (Coronaviridae; Coronavirinae; Betacoronavirus; Sarbecovirus) variants in the Moscow region using targeted high-throughput sequencing,” *Vopr. Virusol.*, No. 4, 269–278 (2021).

Translated by B. Alekseev