ORIGINAL RESEARCHES

© Authors, 2020

Genetic variability of SARS-CoV-2 in biological samples from patients in Moscow

Anna S. Speranskaya✉**, Valeria V. Kaptelova, Andrey E. Samoilov, Anna Yu. Bukharina, Olga Yu. Shipulina, Elena V. Korneenko, Vasily G. Akimkin**

Central Research Institute of Epidemiology, 111123, Moscow, Russia

Currently, a lot of attention is given to SARS-CoV-2 subpopulations and their coexistence with different genomic variants within the same patient. In this study, we performed next-generation whole-genome sequencing and assembly of viruses from samples representing swabs or autopsy specimens obtained from patients diagnosed with COVID-19, which were initially confirmed by the real-time polymerase chain reaction (Ct = 10.4–19.8). Samples were prepared for sequencing by using the SCV-2000bp protocol. The obtained data were checked for presence of more than one SARS-CoV-2 genetic variants in a sample. Variants of nucleotide substitutions, coverage for each variant, and location of the variable position in the reference genome were detected with tools incorporated in the CLC Genomics Workbench program. In our search for variable nucleotide positions, we assumed that the sample had two genetic variants (not more); the threshold value ≥ 90% was set for probability of the identified variant. Variants represented by less than 20% of the reads in the total coverage were not taken into consideration. The obtained results showed that 5 samples had variability, i.e. they had several genetic variants of SARS-CoV-2. In 4 samples, both of the detected genomic variants differed only in one nucleotide position. The fifth sample demonstrated more substantial differences: a total of 3 variable positions and one three-nucleotide deletion. Our study shows that different genetic variants of SARS-CoV-2 can coexist within the same patient.

Keywords: *SARS-CoV-2; next generation sequencing; dual infection; quasispecies.*

Funding. This work was supported by (1) Russian Foundation for Basic Research (RFBR) grant № 20-04-60561 (sequencing) and (2) Ministry of Science and Higher Education of the Russian Federation within the framework of a grant in the form of a subsidy for the creation and development of the «World-class Genomic Research Center for Ensuring Biological Safety and Technological Independence under the Federal Scientific and Technical Program for the Development of Genetic Technologies», agreement No. 075-15-2019-1666 (data analysis).

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

For citation: Speranskaya A.S., Kaptelova V.V., Samoilov A.E., Bukharina A.Yu., Shipulina O.Yu., Korneenko E.V., Akimkin V.G. Genetic variability of SARS-CoV-2 in biological samples from patients in Moscow. *Journal of microbiology, epidemiology and immunobiology = Zhurnal mikrobiologii, èpidemiologii i immunobiologii.* 2020; 97(6): 511-517. (In Russ.).
DOI: https://doi.org/10.36233/0372-9311-2020-97-6-1

Received 30 November 2020
Accepted 1 December 2020

Генетическая вариабельность SARS-CoV-2 в биологических образцах от пациентов г. Москвы

Сперанская А.С.✉**, Каптелова В.В., Самойлов А.Е., Бухарина А.Ю., Шипулина О.Ю., Корнеенко Е.В., Акимкин В.Г.**

ФБУН «Центральный научно-исследовательский институт эпидемиологии» Роспотребнадзора, 111123, Москва, Россия

Сосуществование субпопуляций SARS-CoV-2 с различными вариантами генома внутри организма одного пациента — один из обсуждаемых в настоящее время феноменов. В данной работе мы провели высокопроизводительное секвенирование и сборку полных геномов вирусов из образцов, которые представляли

собой мазки или аутопсийный материал от пациентов с диагнозом COVID-19, предварительно подтвержденным методом полимеразной цепной реакции в реальном времени (Ct = 10,4–19,8). Подготовку образцов к секвенированию проводили с помощью протокола SCV-2000bp. Полученные данные проверяли на присутствие в образце более чем одного генетического варианта SARS-CoV-2. Варианты нуклеотидных замен, покрытие для каждого варианта, а также координаты вариабельной позиции в референсном геноме определяли с помощью инструментов программы «CLC Genomics Workbench». При поиске вариабельных нуклеотидных позиций исходили из предположения, что в образце имеются 2 генетических варианта (не более), для вероятности определяемого варианта использовали пороговое значение ≥ 90%. Также игнорировали варианты, которые были представлены менее чем 20% прочтений от общего покрытия. Полученные результаты показали, что в 5 образцах имеется вариабельность, т.е. содержится несколько генетических вариантов SARS-CoV-2. В 4 образцах оба найденных варианта геномов различались лишь в одной нуклеотидной позиции. В пятом образце были найдены более существенные различия: сразу 3 вариабельных позиции и одна делеция длиной в 3 нуклеотида. Наше исследование показывает возможность сосуществования различных генетических вариантов SARS-CoV-2 в организме пациента.

Ключевые слова: *SARS-CoV-2; высокопроизводительное секвенирование; двойная инфекция; квазивиды.*

Источник финансирования. Данная работа поддержана частично грантом РФФИ № 20-04-60561 (секвенирование), а также частично грантом в форме субсидии на создание и развитие «Центра геномных исследований мирового уровня по обеспечению биологической безопасности и технологической независимости в рамках Федеральной научно-технической программы развития генетических технологий», соглашение № 075-15-2019-1666 (анализ данных).

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

Для цитирования: Сперанская А.С., Каптелова В.В., Самойлов А.Е., Бухарина А.Ю., Шипулина О.Ю., Корнеенко Е.В., Акимкин В.Г. Генетическая вариабельность SARS-CoV-2 в биологических образцах от пациентов г. Москвы. *Журнал микробиологии, эпидемиологии и иммунобиологии.* 2020; 97(6): 511–517. DOI: https://doi.org/10.36233/0372-9311-2020-97-6-1

Поступила 30.11.2020
Принята в печать 1.12.2020

Introduction

The COVID-19 pandemic caused by the SARS-CoV-2 coronavirus began in Wuhan, China, at the end of December in 2019. In Russia, the peak of the first wave was recorded in the middle of May in 2020; the second wave started rising at the end of August. By 11/11/2020, the number of new COVID-19 cases daily registered in Russia had reached approximately 19.9 thousand (more than 1.8 million confirmed cases countrywide.[1] In February 2020, the published data obtained by foreign researchers performing next generation sequencing (NGS) of SARS-CoV-2 genomes showed that in the phylogenetic tree, sequences are grouped into two major clades (lineages/types/genotypes) known as L and S. They are differentiated by single-nucleotide polymorphisms in ORF1ab and ORF8 [1]. By 10/11/2020, genomes belonging to 8 major clades: L, O, S, V, G, GR, GH, and GV had been deposited in the GISAID database.[2] Out of them, 6 clades were identified among Russian isolates at the end of April 2020 [2]; representatives of 7 clades (except for GV) were deposited in GISAID in November 2020 [3].

The SARS-CoV-2 belongs to RNA viruses characterized by high mutation rates, which, in their turn, lead to evolution of quasispecies (sub-populations) within the same host. Currently, the existing quasispecies for SARS-CoV-2 have been identified and recorded [4–8]; there have been studies based on a large amount of the related data. The work [9] includes a bioinformatic analysis of "raw" NGS data from nearly 4 thousand samples obtained at different laboratories and available in the SRA database[3]. In addition, the same researchers have performed a bioinformatic analysis of NGS data for RNA isolated from swabs of patients from Switzerland and have found that different variants of the SARS-CoV-2 genome coexist within the same patient. U. Fahnøe et al. Explain the phenomenon by natural genetic diversity caused by rapid evolution of viruses (assuming that some of the genomic variants can be artifacts that evolved during preparation of libraries and sequencing). Indeed, NGS is a recognized technique for assessment of genetic variability of viral populations [10, 11] and is used for confirming the existence of viral quasispecies in a patient's body [12]. However, it is not easy to differentiate truly single-nucleotide variants (SNV) from errors in sequencing and artifacts of sample preparation, especially if it refers to detection of rarely occurring sub-populations.

The phenomenon when the same patient has concurrently two or more variants of the same virus is known as dual infection or coinfection, if it occurs simultaneously with the first infection or sometime later [13]. This phenomenon has been quite extensively studied among viruses of different families and genera [13–

[1] statista.com [Electronic resource]. URL: www.statista.com/statistics/1102303/coronavirus-new-cases-development-russia

[2] gisaid.org [Electronic resource]. URL: www.gisaid.org

[3] URL: www.ncbi.nlm.nih.gov/sra

17]. The possibility of coinfection for SARS-CoV-2 is still questionable, though the number of arguments supporting this assumption is gradually increasing. Some authors interpret the heterogeneity found in SARS-CoV-2 genome sequences as dual infection. For example, this conclusion was offered in the research work [18] describing the results of sequencing of the fragment (795 bp long) of the gene encoding the viral spike protein, when sample were obtained in Iraq, from 19 patients having obvious symptoms of COVID-19. By using Sanger sequencing, double peaks in chromatograms were detected in each of 19 samples. The authors explain the heterogeneity found in the sequencing results as coinfection, while the high percentage (19/19) is explained by the specifics of the national approach to compliance with sanitary regulation (accidental contamination was not considered). In addition to the above preprint, at the end of September 2020, a case of dual infection with SARS-CoV-2 was reported by another group of authors [19] who referred to the data published by S. Ilmjärv et al. [20]. Our preprint article also provides evidence supporting the possibility of dual SARS-CoV-2 infection: We have described the case when genomes belonging to GR and GH clades were detected in a female 90-year-old patient and when the dominant strain changed during the disease course [21].

Mutations occurring in the population and resulting from natural evolution of the virus, similar to infection with another strain, can affect the immune response of the host and change the clinical course of the disease. Some researchers assume that individual mutations of SARS-CoV-2 or their combinations can affect the viral replication rate and speed of disease transmission, cause problems during treatment due to developing antiviral drug resistance mutations [22, 23].

The purpose of this study is to demonstrate viral quasispecies at least in part of biological samples from Moscow patients with confirmed SARS-CoV-2 infection.

Materials and methods

The study was conducted on swabs (19 samples) and autopsy material (2 samples) in transport media from patients with ARVI symptoms or with suspected COVID-19, which had been initially delivered to the Department of Molecular Diagnostic Methods at the Central Research Institute of Epidemiology and had been identified as SARS-CoV-2 positive. The identification involved the real-time polymerase chain reaction that was performed with an AmpliSens® Cov-Bat-FL reagent kit (AmpliSens, Russia) in accordance with the user manual. All the samples used in the study contained viral RNA at high concentration (Ct = 10.4–19.8).

The reverse transcription reaction was conducted by using 10 µl of RNA samples and Reverta-L kit (AmpliSens, Russia) in accordance with the user manual. The obtained cDNA was used as a template for amplification of genomic fragments. The SARS-CoV-2 genome fragments were amplified by using the SCV-200bp primer panel of our design [25].

The samples were prepared for sequencing in accordance with the protocol [26]. The Q5 High-Fidelity DNA Polymerase (New England BioLabs) was used for amplification of genomic fragments, 35 cycles of amplification. The same polymerase was used for preparation of libraries, 8 cycles of amplification. NGS was completed by using the Illumina HiSeq 1500 platform as well as HiSeq PE Rapid Cluster Kit v2 and HiSeq Rapid SBS Kit v2 reagent kits (500 cycles) or the Illumina MiSeq platform and a MiSeq Reagent Kit v2 reagent kit (500 cycles).

The obtained reads were filtered by quality with the help of Trimmomatic [27]; the sequences of primers were removed by using cutadapt software [28]; the obtained reads were mapped onto the reference sequence EPI_ISL_402124 by using the bowtie2 tool [29]; the SAMtools [30] was used to remove chimeric reads and to receive bam-files. Consensus sequences were obtained with the help of BEDtools [31]. The built-in tools of the CLC Genomics Workbench 8.5 program were used for identification and estimation of the SNV coverage [32]. During the analysis, we proceeded from the assumption that the sample can concurrently have not more than 2 genomic variants (we used the fixed ploidy variant = 2). We set the coverage at ≥ 20% as a threshold value for SNV detection (substitutions with low coverage were disregarded).

Results

The consensus sequences of genomes from the studied samples were deposited in the GISAID database, including SNV in the degenerate nucleotide code format (accession numbers are given in **Table 1**), except for one sample - d186dl477 (see below for details).

After quality filtering and PCR primer trimming, the amount of information per each sample ranged from 0.557 million to 11.965 million reads (the median - 6 million). Then SNVs were analyzed in the viral population for each sample. The mapping of reads onto the reference genome hCoV-19/Wuhan/WIV04/2019 (the GISAID accession number EPI_ISL_402124) showed that some samples had variable/degenerate SNV positions. The proportion of minor subpopulations (estimated by the minor SNV coverage) ranged from 24% to 46%. The presence of variability did not depend on the number of reads or on the viral load in the sample (see samples d186s56, Ct = 10.4 or d186s144, Ct = 11.2, in which no variable positions were found).

SNVs in gene encoding S-protein (Spike Glycoprotein) were detected in 4 samples: d186s128, d186s137, d186dl290, and d186dl477. In the meantime, in one of the samples, the same *spike glycoprotein* gene had a heterogeneous section represented by the TTA/--- deletion in one of the two genomic variants.

JOURNAL OF MICROBIOLOGY, EPIDEMIOLOGY AND IMMUNOBIOLOGY. 2020; 97(6)
DOI: https://doi.org/10.36233/0372-9311-2020-97-6-1

Table 1. Description of degenerate positions in analyzed samples

Sample	Ct	Reads after filtration	Nucleotide position	Total coverage	Nucleotide variants	Variant coverage	Percent of reads	Protein mutation	GISAID accession number
d186s128	16	7 575 603	23613	450	C	207	46%	Spike A684V	EPI_ISL_660437
					T	243	54%		
d186l325	13,6	6 337 032			No degenerate positions				EPI_ISL_610232
d186s56	10,4	6 481 901			No degenerate positions				EPI_ISL_610240
d186s51	13,7	12790208			No degenerate positions				EPI_ISL_610239
d186s137	16,3	7 992 766	23916	15388	T	9314	60,53%	Spike V785A	EPI_ISL_660438
					C	6065	39,41%		
d186s144	11,2	6 928 932			No degenerate positions				EPI_ISL_610241
d186s149	13,5	7 026 576			No degenerate positions				EPI_ISL_610242
d186s155	12,6	6 815 104			No degenerate positions				EPI_ISL_610243
d186s434	14	7 722 664			No degenerate positions				EPI_ISL_610244
d186dl199	15,1	3 163 502			No degenerate positions				EPI_ISL_610227
d186dl222	18,1	2 885 286			No degenerate positions				EPI_ISL_610228
d186dl240	19,8	1 467 205	8139	24	C	8	33,33%	NSP3S1807F	EPI_ISL_660433
					T	16	66,67%		
d186dl282	13,4	4 043 559			No degenerate positions				EPI_ISL_610229
d186dl290	13	557 624	24794	4424	G	1596	36,08%	Spike A1078S	EPI_ISL_660434
					T	2825	63,86%		
d186dl294	12,9	6 481 228			No degenerate positions				EPI_ISL_610230
d186dl302	12,4	3 742 501			No degenerate positions				EPI_ISL_610231
d186dl365	14,3	5625499			No degenerate positions				EPI_ISL_610233
d186dl381	13,4	3 066 646			No degenerate positions				EPI_ISL_610234
d186dl389	13	3 424 997			No degenerate positions				EPI_ISL_610235
d186dl441	14,3	4 940 467			No degenerate positions				EPI_ISL_610236
d186dl477	12	2 500 575	21588	17058	T	8950	52,47%	Spike P9L	Nucleotide variants with major coverage — EPI_ISL_660435, with minor coverage — EPI_ISL_660436
					C	8107	47,53%		
			21991–21993	14810	TTA	11250	75,96%	Spike Y145del	
					---(del)	3544	23,93%		
			27247	7923	C	4870	61,47%	ORF6 synonymous	
					T	3043	38,41%		
			29250	1245	T	873	70,12%	N P326L	
					C	370	29,72%		

In addition, SNVs were found at the site of the *ORF1ab* gene encoding the nsp3 protein (see sample d186dl240) as well as in genes *N* and *ORF6* (encoding same-name proteins; see sample d186dl477).

Sample d186dl477 is especially noteworthy, as 3 SNVs and a three-nucleotide deletion were detected in it. One of the SNVs leads to synonymous substitution in ORF6; the other cause mutations in the spike protein (P9L and Y145del) as well as in the N protein (P326L). All the detected mutations belong to the category of rare or new mutations (**Table 2**). The coverage of minor mutation variants in this sample ranged from 29.7% to 47%. We deposited two sequences in the GISAID database: Sequence [EPI_ISL_660435] includes major mutation variants, while sequence [EPI_ISL_660436] includes minor variants.

Discussion

In our study, we intentionally narrowed down our search to strongly represented variants, as our main objective was to show wide occurrence of heterogeneous populations of SARS-CoV-2. We analyzed the NGS data for SARS-CoV-2 genomes and evaluated the representation of viral subpopulations by using variability search algorithms implemented as built-in tools of the CLC Genomics Workbench program. We did not try to find optimum criteria for SNV identification, being satisfied by using rigid criteria (min 20% of the total coverage of the analyzed position at detection reliability of at least 90%). We identified SNVs in 5 samples out of 21. In 4 samples, quasispecies differed in occasional SNVs.

RNA viruses are characterized by high mutation rates, which frequently cause developing of quasispecies within the same host. Numerous studies show that several viruses with different SNV in genomes have been found concurrently existing in samples from COVID-19 patients [8–11, 21]. We assume that occasional SNVs found in 4 samples can be also explained by natural evolution of viral genomes in the host.

The fifth sample (d186dl477) differed from other heterogeneous samples by 3 SNPs and 1 heterogeneous three-nucleotide section that were detected in it. However, the values of relative coverage in these positions did not show any substantial difference, thus leading to the conclusion that the sample has a combination of strains. We assume that this phenomenon can be explained in two ways: unprecedentedly fast evolution of the virus in the body of this patient or infection with different strains of SARS-CoV-2.

Recent publications show that the mutation rate for SARS-CoV-2 is almost identical to the mutation rate of the SARS-CoV genome ($0.80–2.38 \times 10^{-3}$ nucleotide substitutions per site per year) [19, 30, 31]. Having statistically analyzed a large amount of "raw" sequencing data from different laboratories, J. Kuipers *et al.* [11] demonstrated that heterogeneity of the viral population in a sample may correlate with the age of the patient. Sample d186dl477 was obtained from an 84-year-old female patient. If in our theoretical calculations, we assume that the mutation rate equals the highest possible rate (2.38×10^{-3} nucleotide substitutions per site per year), then up to 10 mutations can

Table 2. Characterization of mutations found in sample d186dl477

<table>
<tr><th rowspan="2">Protein</th><th rowspan="2">EPI_ISL_660435</th><th rowspan="2">EPI_ISL_660436</th><th colspan="3">Frequency of found mutations among known SARS-CoV-2 sequences according to GISAID data, as of November 27, 2020</th></tr>
<tr><th>%</th><th>number of genoms</th><th>number of countries</th></tr>
<tr><td>NSP2</td><td colspan="2">**T547I**</td><td>**0,07**</td><td>**55**</td><td>**10**</td></tr>
<tr><td>NSP3</td><td colspan="2">**V1847F**</td><td>**>0,01**</td><td>**7**</td><td>**2**</td></tr>
<tr><td>NSP12</td><td colspan="2">P323L</td><td>88,88</td><td>184 070</td><td>125</td></tr>
<tr><td>Spike</td><td>**P9L**</td><td>**–**</td><td>**0,01**</td><td>**27**</td><td>**7**</td></tr>
<tr><td></td><td>**–**</td><td>**Y145del**</td><td colspan="3">**Found for the first time**</td></tr>
<tr><td></td><td colspan="2">D614G</td><td>89,28</td><td>18 5187</td><td>26</td></tr>
<tr><td>N</td><td colspan="2">R203K</td><td>36,46</td><td>75 504</td><td>104</td></tr>
<tr><td></td><td colspan="2">G204R</td><td>36,15</td><td>74 866</td><td>103</td></tr>
<tr><td></td><td colspan="2">**M210I**</td><td>**0,24**</td><td>**488**</td><td>**18**</td></tr>
<tr><td></td><td colspan="2">**A211V**</td><td>**0,08**</td><td>**175**</td><td>**22**</td></tr>
<tr><td></td><td>**P326L**</td><td>**–**</td><td>**0,02**</td><td>**41**</td><td>**8**</td></tr>
</table>

Note. Mutations for which the frequency of occurrence in the world is ≤ 0.25 are in bold.

evolve in the SARS-CoV-2 genome during 5 days of disease progression.

The idea of evolution is also supported by the fact that four mutations out of the mutations common to both strains are characterized by low occurrence worldwide — >0.01–0.24%. The SNV occurrence resulting from the consequential evolution should have affected the mutation rate. However, the observed difference in the coverage values for minor SNVs is not significant. In the absence of clinical data and information about the duration of the disease in the patient we obtained sample d186dl477 from, we cannot decidedly assert that heterogeneity is a consequence of the natural evolution of the virus. A lack of data gives no support for the alternative assumption (i.e. coinfection resulting from the infection with the second strain of SARS-CoV-2).

Being in agreement with the authors [9], we think that it is important to find out (in the future) if heterogeneity of SARS-CoV-2 populations depends on the disease progression, if the probability of detection of heterogeneous samples increases with the patient's age. Special attention should be given to developing criteria for differentiation between repeat infection and heterogeneity resulting from the natural evolution of the virus.

REFERENCES

1. Tang X., Wu C., Li X., Song Y., Yao X., Wu X., et al. On the origin and continuing evolution of SARS-CoV-2. *Natl. Sci. Rev.* 2020; 7(6): 1012–23. https://doi.org/10.1093/nsr/nwaa036
2. Komissarov A.B., Safina K.R., Garushyants S.K., Fadeev A.V., Sergeeva M.V., Ivanova A.A., et al. Genomic epidemiology of the early stages of SARS-CoV-2 outbreak in Russia. *medRxiv.* Preprint. 2020. https://doi.org/10.1101/2020.07.14.20150979
3. Sýkorová E., Fajkus J., Mezníková M., Lim K.Y., Neplechová K., Blattner F.R., et al. Minisatellite telomeres occur in the family Alliaceae but are lost in Allium. *Am. J. Bot.* 2006; 93(6): 814–23. https://doi.org/10.3732/ajb.93.6.814
4. Nyayanit D., Yadav P.D., Kharde R., Shete-Aich A. Quasispecies analysis of the SARS-CoV-2 from representative clinical samples: A preliminary analysis. *Indian J. Med. Res.* 2020; 152(1): 105. https://doi.org/10.4103/ijmr.ijmr_2251_20
5. Jary A., Leducq V., Malet I., Marot S., Klement-Frutos E., Teyssou E., et al. Evolution of viral quasispecies during SARS-CoV-2 infection. *Clin. Microbiol. Infect.* 2020; 26(11): 1560.e1-1560.e4. https://doi.org/10.1016/j.cmi.2020.07.032
6. Chaudhry M.Z., Eschke K., Grashoff M., Abassi L., Kim Y., Brunotte L., et al. SARS-CoV-2 quasispecies mediate rapid virus evolution and adaptation. *bioRxiv.* Preprint. 2020. https://doi.org/10.1101/2020.08.10.241414
7. Xu D., Zhang Z., Wang F.S. SARS-associated coronavirus quasispecies in individual patients. *N. Engl. J. Med.* 2004; 350(13): 1366–7. https://doi.org/10.1056/nejmc032421
8. Park D., Huh H.J., Kim Y.J., Son D.S., Jeon H.J., Im E.H., et al. Analysis of intrapatient heterogeneity uncovers the microevolution of Middle East respiratory syndrome coronavirus. *Mol. Case Stud.* 2016; 2(6): a001214. https://doi.org/10.1101/mcs.a001214
9. Kuipers J., Batavia A.A., Jablonski K.P., Bayer F., Borgsmüller N., Dondi A., et al. Within-patient genetic diversity of SARS-CoV-2. *bioRxiv.* Preprint. 2020. https://doi.org/10.1101/2020.10.12.335919
10. McElroy K., Zagordi O., Bull R., Luciani F., Beerenwinkel N. Accurate single nucleotide variant detection in viral populations by combining probabilistic clustering with a statistical test of strand bias. *BMC Genomics.* 2013; 14(1): 501. https://doi.org/10.1186/1471-2164-14-501
11. Fahnøe U., Pedersen A.G., Dräger C., Orton R.J., Blome S., Höper D., et al. Creation of functional viruses from non-functional cDNA clones obtained from an RNA virus population by the use of ancestral reconstruction. *PLoS One.* 2015; 10(10): e0140912. https://doi.org/10.1371/journal.pone.0140912
12. Kireev D.E., Lopatukhin A.E., Murzakova A.V., Pimkina E.V., Speranskaya A.S., Neverov A.D., et al. Evaluating the accuracy and sensitivity of detecting minority HIV-1 populations by Illumina next-generation sequencing. *J. Virol. Methods.* 2018; 261: 40–5. https://doi.org/10.1016/j.jviromet.2018.08.001
13. Лаповок И.А., Лопатухин А.Э., Киреев Д.Е. Двойная ВИЧ-инфекция: эпидемиология, клиническая значимость и диагностика. *Инфекционные болезни.* 2019; 17(2): 81–7. [Lapovok I.A., Lopatukhin A.E., Kireev D.E. Dual HIV infection: epidemiology, clinical significance, and diagnosis. *Infektsionnye bolezni.* 2019; 17(2): 81–7. (in Russ.)] https://doi.org/10.20953/1729-9225-2019-2-81-87
14. Falchi A., Arena C., Andreoletti L., Jacques J., Leveque N., Blanchon T., et al. Dual infections by influenza A/H3N2 and B viruses and by influenza A/H3N2 and A/H1N1 viruses during winter 2007, Corsica Island, France. *J. Clin. Virol.* 2008; 41(2): 148–51. https://doi.org/10.1016/j.jcv.2007.11.003
15. Semple M.G., Cowell A., Dove W., Greensill J., McNamara P.S., Halfhide C., et al. Dual infection of infants by human metapneumovirus and human respiratory syncytial virus is strongly associated with severe bronchiolitis. *J. Infect. Dis.* 2005; 191(3): 382–6. https://doi.org/10.1086/426457
16. van der Kuyl A.C., Cornelissen M. Identifying HIV-1 dual infections. *Retrovirology*. 2007; 4: 67. https://doi.org/10.1186/1742-4690-4-67
17. Weinberg A., Bloch K.C., Li S., Tang Y.W., Palmer M., Tyler K.L. Dual infections of the central nervous system with Epstein-Barr virus. *J. Infect. Dis.* 2005; 191(2): 234–7. https://doi.org/10.1086/426402
18. Hashim H.O., Mohammed M.K., Mousa M.J., Abdulameer H.H., Alhassnawi A.T., Hassan S.A., et al. Unexpected co-infection with different strains of SARS-CoV-2 in patients with COVID-19. *Preprints.org.* 2020. Preprint. https://doi.org/10.20944/preprints202009.0375.v1
19. Liu S., Shen J., Fang S., Li K., Liu J., Yang L., et al. Genetic spectrum and distinct evolution patterns of SARS-CoV-2. *Front. Microbiol.* 2020; 11: 593548. https://doi.org/10.3389/fmicb.2020.593548
20. Gudbjartsson D.F., Helgason A., Jonsson H., Magnusson O.T., Melsted P., Norddahl G.L., et al. Spread of SARS-CoV-2 in the Icelandic population. *N. Engl. J. Med.* 2020; 382(24): 2302–15. https://doi.org/10.1056/nejmoa2006100
21. Samoilov A., Kaptelova V.V., Bukharina A.Y., Shipulina O.Y., Korneenko E.V., Lukyanov A.V. et al. Change of dominant strain during dual SARS-CoV-2 infection: preprint. *medRxiv*. Preprint. 2020. https://doi.org/10.1101/2020.11.29.20238402
22. Ilmjärv S., Abdul F., Acosta-Gutiérrez S., Estarellas C., Galdadas I., Casimir M., et al. Epidemiologically most successful SARS-CoV-2 variant: concurrent mutations in RNA-dependent RNA polymerase and spike protein. *medRxiv.* Preprint. 2020. https://doi.org/10.1101/2020.08.23.20180281
23. Speranskaya A., Kaptelova V., Valdokhina A., Bulanenko V., Samoilov A., Korneenko E., et al. SCV-2000bp: a primer panel for SARS-CoV-2 full-genome sequencing. *bioRxiv.* Preprint. 2020. https://doi.org/10.1101/2020.08.04.234880
24. Kaptelova V.V., Speranskaya A.S. Protocol for SCV-2000bp: a primer panel for SARS-CoV-2 full-genome sequencing v1. https://dx.doi.org/10.17504/protocols.io.bn77mhrn
25. 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

26. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal.* 2011; 17(1): 10. https://doi.org/10.14806/ej.17.1.200
27. Langmead B., Salzberg S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods.* 2012; 9(4): 357–9. https://doi.org/10.1038/nmeth.1923
28. Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25(16): 2078–9. https://doi.org/10.1093/bioinformatics/btp352
29. Quinlan A.R., Hall I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010; 26(6): 841–2. https://doi.org/10.1093/bioinformatics/btq033
30. Romano M., Ruggiero A., Squeglia F., Maga G., Berisio R. A structural view of SARS-CoV-2 RNA replication machinery: RNA synthesis, proofreading and final capping. *Cells*. 2020; 9(5): 1267. https://doi.org/10.3390/cells9051267
31. Zhao Z., Li H., Wu X., Zhong Y., Zhang K., Zhang Y.P., et al. Moderate mutation rate in the SARS coronavirus genome and its implications. *BMC Evol. Biol.* 2004; 4: 21. https://doi.org/10.1186/1471-2148-4-21

Information about the authors

Anna S. Speranskaya✉ — PhD, Head, Group for genomics and post-genome technology, Central Research Institute of Epidemiology, 111123, Moscow, Russia.
ORCID ID: https://orcid.org/0000-0001-6326-1249.
E-mail: hanna.s.939@gmail.com

Valeria V. Kaptelova — junior researcher, Group for genomics and post-genome technology, Central Research Institute of Epidemiology, 111123, Moscow, Russia.
ORCID ID: https://orcid.org/0000-0003-0952-0830

Andrei E. Samoilov — researcher, Group for genomics and post-genome technology, Central Research Institute of Epidemiology, 111123, Moscow, Russia.
ORCID ID: https://orcid.org/0000-0001-8284-3164

Anna Yu. Bukharina — laboratory researcher assistant (Molecular diagnostic methods department), Central Research Institute of Epidemiology, 111123, Moscow, Russia.
ORCID ID: https://orcid.org/0000-0002-6892-3595

Olga Yu. Shipulina — PhD, Head, Molecular diagnostic methods department, Central Research Institute of Epidemiology, 111123, Moscow, Russia.
ORCID ID: https://orcid.org/0000-0003-4679-6772

Elena V. Korneenko — junior researcher, Group for genomics and post-genome technology, Central Research Institute of Epidemiology, 111123, Moscow, Russia.
ORCID ID: https://orcid.org/0000-0003-4679-6772

Vasily G. Akimkin — D. Sci. (Med.), Prof., Academician of RAS, Director, Central Research Institute of Epidemiology, 111123, Moscow, Russia.
ORCID ID: https://orcid.org/0000-0003-4228-9044

Сведения об авторах

Сперанская Анна Сергеевна✉ — к.б.н., рук. научной группы геномики и постгеномных технологий ЦНИИ Эпидемиологии, 111123, Москва, Россия.
ORCID ID: https://orcid.org/0000-0001-6326-1249.
E-mail: hanna.s.939@gmail.com

Каптелова Валерия Владимировна — м.н.с. группы геномики и постгеномных технологий ЦНИИ Эпидемиологии, 111123, Москва, Россия.
ORCID ID: https://orcid.org/0000-0003-0952-0830.

Самойлов Андрей Евгеньевич — н.с. группы геномики и постгеномных технологий ЦНИИ Эпидемиологии, 111123, Москва, Россия.
ORCID ID: https://orcid.org/0000-0001-8284-3164.

Бухарина Анна Юрьевна — лаборант-исследователь подразделения молекулярных методов диагностики ЦНИИ Эпидемиологии.
ORCID ID: https://orcid.org/0000-0002-6892-3595.

Шипулина Ольга Юрьевна — к.м.н., рук. подразделения лабораторной медицины и продвижения лабораторных услуг отдела молекулярной диагностики и эпидемиологии ЦНИИ Эпидемиологии, 111123, Москва, Россия.
ORCID ID: https://orcid.org/0000-0003-4679-6772.

Корнеенко Елена Васильевна — м.н.с. группы геномики и постгеномных технологий ЦНИИ Эпидемиологии, 111123, Москва, Россия.
ORCID ID: https://orcid.org/0000-0003-4679-6772.

Акимкин Василий Геннадьевич — д.м.н., проф., акад. РАН, директор ФБУН ЦНИИ Эпидемиологии, 111123, Москва, Россия.
ORCID ID: https://orcid.org/0000-0003-4228-9044.

Участие авторов: все авторы сделали эквивалентный вклад в подготовку публикации.