ОРИГИНАЛЬНЫЕ СТАТЬИ

© Коллектив авторов, 2020

Check for updates

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 01.12.2020

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 as-

sembly 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.), (In Engl.).
DOI: https://doi.org/10.36233/0372-9311-2020-97-6-1

Received 30 November 2020
Accepted 1 December 2020

Введение

Пандемия COVID-19, вызванная коронавирусом SARS-CoV-2, началась в конце декабря 2019 г. в Ухане (КНР). В России пик первой волны пришелся на середину мая 2020 г., вторая волна начала подниматься в конце августа. К 11.11.2020 г. число ежедневно регистрируемых в России новых случаев COVID-19 достигло примерно 19,9 тыс. (более 1,8 млн подтвержденных случаев в стране)[1]. Уже в феврале 2020 г. опубликованы зарубежные данные высокопроизводительного секвенирования (next generation sequencing, NGS) геномов SARS-CoV-2, из которых следовало, что на филогенетическом дереве последовательности делятся на две основные клады (линии/типы/генотипы), которые были названы L и S. Они различаются однонуклеотидными полиморфизмами в ORF1ab и ORF8 [1]. К 10.11.2020 г. в базе данных GISAID[2] зарегистрированы геномы, относящиеся к 8 основным кладам: L, O, S, V, G, GR, GH, GV. Из них 6 клад были найдены среди российских изолятов уже в конце апреля 2020 г. [2], представители 7 клад (кроме GV) были зарегистрированы в GISAID в ноябре 2020 г. [3].

SARS-CoV-2 относится к РНК-вирусам, для которых характерна высокая частота мутаций. Следствием этого служит формирование квазивидов (субпопуляций) в организме одного хозяина. В настоящее время уже зафиксирован факт существования квазивидов для SARS-CoV-2 [4–8], в том числе есть исследования, проведенные на большой выборке данных. В работе [9] осуществлен биоинформатический анализ результатов «сырых» данных NGS почти 4 тыс. образцов, полученных ранее различными лабораториями и выложенных в открытую базу данных SRA[3]. Кроме того, те же авторы выполнили биоинформатический анализ данных NGS РНК, выделенной из мазков от пациентов из Швейцарии, и обнаружили сосуществование различных вариантов генома SARS-CoV-2 внутри организма пациента. U. Fahnøe и соавт. объясняют это естественным генетическим разнообразием из-за быстрой вирусной эволюции (допуская, что часть вариантов генома могут быть артефактами, образовавшимися в процессе подготовки библиотек и секвенирования). Действительно, NGS — признанный инструмент для оценки генетической вариабельности вирусных популяций [10, 11], для подтверждения существования квазивидов вируса в организме пациента [12]. Однако отличить истинные однонуклеотидные варианты (SNV) от ошибок секвенирования, артефактов пробоподготовки — сложная задача, особенно если речь идет об обнаружении редко встречающихся субпопуляций.

[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

Явление, при котором в организме пациента присутствуют одновременно два или более вариантов одного вируса, также называют двойной инфекцией или ко-инфекцией, если она происходит одновременно с первым заражением или спустя некоторое время [13]. Это явление довольно хорошо исследовано среди вирусов различных семейств и родов [13–17]. Возможность ко-инфекции для SARS-CoV-2 пока что остается под вопросом, хотя доводы в пользу этого предположения постепенно накапливаются. Некоторые авторы интерпретируют как двойную инфекцию наблюдаемую гетерогенность в результатах секвенирования геномов SARS-CoV-2. Например, такие выводы были сделаны в работе [18], описывающей результаты секвенирования фрагмента (длиной 795 п.н.) гена, кодирующего Spike-белок вируса, из образцов от 19 пациентов в Ираке с явными симптомами COVID-19. Секвенирование проводилось методом Сэнгера, двойные пики на хроматограммах были обнаружены во всех 19 образцах.

Наличие гетерогенности в результатах секвенирования авторы объясняют ко-инфекцией, а столь высокий процент (19/19) — особенностями национального подхода к соблюдению санитарных норм (возможность контаминации не рассматривалась). Помимо вышеупомянутого препринта, в конце сентября 2020 г. о случае двойной инфекции SARS-CoV-2 у одного пациента сообщили авторы работы [19] со ссылкой на данные S. Ilmjärv и соавт. [20]. В нашей работе, опубликованной в виде препринта, также приводятся доказательства в пользу возможности двойной инфекции SARS-CoV-2: описан случай обнаружения геномов, относящихся к кладам GR и GH, у пациентки 90 лет, а также смены доминирующего штамма в процессе заболевания [21].

Мутации, возникающие в популяции вследствие естественной эволюции вируса, как и инфицирование другим штаммом, могут влиять на иммунный ответ организма хозяина, изменять клиническое течение заболевания. Высказывались предположения, что отдельные мутации SARS-CoV-2 или их комбинации могут влиять на скорость репликации вируса и передачи заболевания, приводить к проблемам при лечении из-за формирующихся мутаций устойчивости к противовирусным препаратам [22, 23].

Цель этой работы — показать наличие квазивидов вируса, по крайней мере, в некоторых биологических образцах от пациентов из Москвы с подтвержденной инфекцией SARS-CoV-2 .

Материалы и методы

Для проведения исследования были использованы мазки (19 образцов) и аутопсийный материал (2 образца) в транспортной среде от пациентов с симптомами ОРВИ или подозрениями на COVID-19, которые предварительно поступили в подразделение молекулярных методов диагностики ФБУН ЦНИИ Эпидемиологии и там же были идентифицированы как положительные на наличие SARS-CoV-2. Для идентификации использовали метод полимеразной цепной реакции в реальном времени, который проводили с помощью набора реагентов «AmpliSens® Cov-Bat-FL» («AmpliSens», Россия) в соответствии с инструкцией производителя. Все использованные в работе образцы содержали вирусную РНК в высокой концентрации (Ct = 10,4–19,8).

Реакцию обратной транскрипции проводили с использованием 10 мкл образцов РНК и набора «Reverta-L kit» («AmpliSens», Россия) согласно инструкции производителя. Полученную кДНК применяли в качестве матрицы для амплификации фрагментов генома. Фрагменты генома SARS-CoV-2 амплифицировали с помощью разработанной нами праймерной панели SCV-200bp [23].

Образцы готовили к секвенированию согласно протоколу [24]. Для амплификации фрагментов генома использовали Q5 High-Fidelity ДНК-полимеразу («New England BioLabs»), 35 циклов амплификации. Для подготовки библиотек использовали ту же полимеразу, 8 циклов амплификации. NGS выполнено на платформе «Illumina HiSeq 1500» с использованием наборов реагентов «HiSeq PE Rapid Cluster Kit v2» и «HiSeq Rapid SBS Kit v2» (500 циклов) или на платформе «Illumina MiSeq» с набором «MiSeq Reagent Kit v2» (500 циклов).

Полученные прочтения были отфильтрованы по качеству с помощью «Trimmomatic» [25]; последовательности праймеров удалены с помощью cutadapt [26]; полученные прочтения картированы на референсную последовательность EPI_ISL_402124 с использованием bowtie2 [27]; с помощью SAMtools [28] удалены химерные прочтения и получены bam-файлы. Консенсусные последовательности получали с помощью BEDtools [29]. Для идентификации и расчета покрытия SNV был использован встроенный инструмент программы «CLC Genomics Workbench 8.5»[4]. В процессе анализа мы исходили из предположения, что в образце может находиться одновременно не более 2 вариантов геномов (использовали алгоритм fixed ploidy variant = 2). В качестве порогового значения детекции SNV использовали покрытие ≥ 20% (варианты замен с низким покрытием не учитывали).

Результаты

Консенсусные последовательности геномов исследованных образцов с указанием SNV в формате вырожденного нуклеотидного кода были депонированы в базу данных GISAID (номера доступа

[4] CLC Genomics Workbench. [Electronic resource]. URL: www.qiagenbioinformatics.com/products/clc-genomics-workbench

Таблица 1. Описание гетерогенных позиций в проанализированных образцах
Table 1. Description of degenerate positions in analyzed samples

Образец Sample	Ct	Количество прочтений после фильтрации Reads after filtration	Позиция нуклеотида Nucleotide position	Покрытие Total coverage	Варианты нуклеотида Nucleotide variants	Вариант покрытия Variant coverage	Доля прочтений Percent of reads	Мутация белка Protein mutation	Номер доступа в GISAID 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	Нет вырожденных позиций / No degenerate positions						EPI_ISL_610232
d186s56	10,4	6 481 901	Нет вырожденных позиций / No degenerate positions						EPI_ISL_610240
d186s51	13,7	12790208	Нет вырожденных позиций / 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	Нет вырожденных позиций / No degenerate positions						EPI_ISL_610241
d186s149	13,5	7 026 576	Нет вырожденных позиций / No degenerate positions						EPI_ISL_610242
d186s155	12,6	6 815 104	Нет вырожденных позиций / No degenerate positions						EPI_ISL_610243
d186s434	14	7 722 664	Нет вырожденных позиций / No degenerate positions						EPI_ISL_610244
d186dl199	15,1	3 163 502	Нет вырожденных позиций / No degenerate positions						EPI_ISL_610227
d186dl222	18,1	2 885 286	Нет вырожденных позиций / 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	Нет вырожденных позиций / 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	Нет вырожденных позиций / No degenerate positions						EPI_ISL_610230
d186dl302	12,4	3 742 501	Нет вырожденных позиций / No degenerate positions						EPI_ISL_610231
d186dl365	14,3	5625499	Нет вырожденных позиций / No degenerate positions						EPI_ISL_610233
d186dl381	13,4	3 066 646	Нет вырожденных позиций / No degenerate positions						EPI_ISL_610234
d186dl389	13	3 424 997	Нет вырожденных позиций / No degenerate positions						EPI_ISL_610235
d186dl441	14,3	4 940 467	Нет вырожденных позиций / No degenerate positions						EPI_ISL_610236
d186dl477	12	2 500 575	21588	17058	T	8950	52,47%	Spike P9L	Варианты нуклеотидов с мажорным покрытием — EPI_ISL_660435, с минорным покрытием — EPI_ISL_660436 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 синонимичная замена ORF6 synonymous	
					T	3043	38,41%		
			29250	1245	T	873	70,12%	N P326L	
					C	370	29,72%		

приведены в **табл. 1**), за исключением одного образца — d186dl477 (детали см. ниже).

После фильтрации по качеству, тримминга праймеров полимеразной цепной реакции количество информации на каждый образец составило от 0,557 млн до 11,965 млн прочтений (медиана — 6 млн). Затем был проведен анализ SNV в вирусной популяции для каждого образца. Картирование прочтений на референсный геном hCoV-19/Wuhan/WIV04/2019 (номер доступа в базе GISAID EPI_ISL_402124) показало, что в некоторых образцах имеются вариабельные/вырожденные SNV-позиции. Доля минорных субпопуляций (оценивалась по покрытию минорного варианта SNV) варьировала от 24% до 46%. Наличие вариабельности не зависело ни от количества прочтений, ни от вирусной нагрузки в образце (см., например, образцы d186s56, Ct = 10,4 или d186s144, Ct = 11,2, в которых вариабельные позиции не найдены).

SNV в гене, кодирующем белок S (Spike Glycoprotein), были обнаружены в 4 образцах: d186s128, d186s137, d186dl290 и d186dl477. При этом в одном из образцов в том же гене *spike glycoprotein* был найден гетерогенный участок, представленный в одном из двух вариантов геномов делецией TTA/---. Кроме того, SNV были найдены на участке *ORF1ab*, кодирующем белок nsp3 (см. образец d186dl240), а также в генах *N* и *ORF6* (кодируют одноименные белки, см. образец d186dl477).

Образец d186dl477 заслуживает особого внимания, поскольку в нем были найдены сразу 3 SNV, а также делеция в три нуклеотида. Один из SNV приводит к синонимичной замене в ORF6, другие — к мутациям в белке Spike (P9L и Y145del), а также в белке N (P326L). Все найденные мутации относятся к категории редко встречающихся или новых (**табл. 2**). Покрытие минорных вариантов мутаций варьировало в этом образце от 29,7% до 47%. Мы депонировали в базу данных GISAID две последовательности: последовательность [EPI_ISL_660435] включает в себя мажорные варианты мутаций, тогда как [EPI_ISL_660436] — минорные варианты.

Обсуждение

В настоящей работе мы намеренно ограничились поиском только лишь высокопредставленных вариантов, поскольку основной целью было показать широкую распространенность гетерогенных популяций SARS-CoV-2. Мы проанализировали данные NGS геномов SARS-CoV-2 и оценили представленность вирусных субпопуляций, используя алгоритмы поиска вариабельности, реализованные в виде встроенных инструментов программы «CLC Genomics Workbench». Мы не определяли оптимальные критерии для идентификации SNV, ограничившись применением жестких критериев (не менее 20% от общего покрытия анализируемой позиции при достоверности определения 90% и более). Мы

Таблица 2. Характеристика мутаций, найденных в образце d186dl477
Table 2. Characterization of mutations found in sample d186dl477

Белок Protein	EPI_ISL_660435	EPI_ISL_660436	Частота встречаемости мутаций среди известных последовательностей геномов SARS-CoV-2 по данным GISAID, депонированных в базу данных на 27.11.2020 г. Frequency of found mutations among known SARS-CoV-2 sequences according to GISAID data, as of November 27, 2020		
			%	количество геномов number of genoms	количество стран number of countries
NSP2		**T547I**	**0,07**	**55**	**10**
NSP3		**V1847F**	**>0,01**	**7**	**2**
NSP12		P323L	88,88	184 070	125
Spike	**P9L**	**–**	**0,01**	**27**	**7**
	–	**Y145del**		**Неизвестная новая мутация Found for the first time**	
		D614G	89,28	18 5187	26
N		R203K	36,46	75 504	104
		G204R	36,15	74 866	103
		M210I	**0,24**	**488**	**18**
		A211V	**0,08**	**175**	**22**
	P326L	**–**	**0,02**	**41**	**8**

Примечание. Выделены мутации, для которых частота встречаемости в мире ≤ 0,25.
Note. Mutations for which the frequency of occurrence in the world is ≤ 0.25 are in bold.

идентифицировали SNV в 5 образцах из 21. В 4 образцах квазивиды различались единичными SNV.

Для РНК-вирусов характерна высокая скорость мутаций, что часто приводит к формированию квазивидов в организме одного хозяина. Во многих работах в образцах от больных COVID-19 идентифицированы одновременно несколько вирусов с разными SNV в геномах [6–9, 19]. Мы полагаем, что найденные в 4 образцах единичные SNV также можно объяснить естественной эволюцией вирусных геномов в организме хозяина.

Пятый образец (d186dl477) отличался от прочих гетерогенных образцов тем, что в нем были обнаружены сразу 3 SNP и 1 гетерогенный участок длиной в 3 нуклеотида. При этом значения относительного покрытия в этих позициях, хотя и различались, но не принципиально, что позволяет утверждать, что в этом образце имеется смесь штаммов. Мы полагаем, что для объяснения этого явления можно выдвинуть два предположения: беспрецедентно быстрая эволюция вируса в организме именно этого пациента либо инфекция разными штаммами SARS-CoV-2.

Недавно опубликованные работы показывают, что частота мутаций SARS-CoV-2 — примерно такая же, как и в геноме SARS-CoV ($0{,}80–2{,}38 \times 10^{-3}$ нуклеотидных замен на сайт в год) [19, 30, 31]. J. Kuipers и соавт. [9], статистически проанализировав большое количество «сырых» данных секвенирования из разных лабораторий, показали, что гетерогенность вирусной популяции в образце положительно коррелирует с увеличением возраста пациентов. Образец d186dl477 был получен от 84-летней пациентки. Если мы примем для теоретических расчетов скорость мутаций, равную максимально возможной ($2{,}38 \times 10^{-3}$ нуклеотидных замен на сайт в год), то при продолжительности развития заболевания 5 дней в геноме SARS-CoV-2 могло сформироваться до 10 мутаций.

В пользу предположения об эволюции также свидетельствует тот факт, что четыре из общих для обоих штаммов мутаций имеют низкую частоту встречаемости в мире — >0,01–0,24%. Образование SNV в результате последовательной эволюции должно было бы отразиться на частоте мутаций. Однако наблюдаемая разница в значениях покрытий минорного SNV невелика. В отсутствие клинических данных и информации о продолжительности заболевания у пациента, от которого был получен образец d186dl477, мы не можем однозначно утверждать, что гетерогенность является следствием естественной эволюции вируса. Данных для обоснования альтернативной гипотезы (о ко-инфекции, возникшей вследствие заражения вторым штаммом SARS-CoV-2) также недостаточно.

Вслед за авторами работы [9] мы полагаем, что важно установить (в перспективе), связана ли гетерогенность популяций SARS-CoV-2 с прогрессированием заболевания, увеличивается ли вероятность обнаружения гетерогенных образцов вместе с возрастом пациента. Также необходимо определить критерии, позволяющие отличить случаи повторного заражения от гетерогенности, возникающей вследствие естественной эволюции вируса.

ЛИТЕРАТУРА/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

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.

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