

РАСПРОСТРАНЕНИЕ ВАРИАНТОВ С ЧАСТЫМИ МУТАЦИЯМИ В ГЕНЕ КАПСИДНОГО БЕЛКА N В РОССИЙСКИХ ИЗОЛЯТАХ SARS-COV-2

С. А. Кирьянов [✉], Т. А. Левина, М. Ю. Кириллов

ООО «ДНК-Технология», Москва, Россия

SARS-CoV-2 представляет собой РНК-вирус семейства β -коронавирусов, вызвавший пандемию тяжелого острого респираторного синдрома (COVID-19). Предполагают, что эволюционные изменения генома ныне преобладающих исходных клонов могут быть обусловлены географическими особенностями, поэтому для штаммов разного происхождения могут быть характерны различные паттерны мутаций. Целью исследования было провести мутационный и филогенетический анализы геномов SARS-CoV-2 российского происхождения в разные периоды и в разных регионах, а также охарактеризовать мутационные профили изолятов, используя подходы биоинформатики. Проведено сравнение накопления мутаций в 86 полногеномных последовательностях SARS-CoV-2 из России и 220 из Европы и Северной Америки для выявления характерных генных вариаций, возможных паттернов селективного отбора. Помимо известных мутаций в генах структурных белков, типичных для изолятов европейского происхождения, выявлены дополнительные синонимичные мутации российского происхождения, в том числе в гене *M* (C26750T). Двойная мутация R203K и G204R в гене нуклеокапсида, ранее появившаяся в Европе, распространилась и быстро (в течение месяца) стала доминирующим вариантом в России. Полученные результаты свидетельствуют о том, что геном значительной части изолятов SARS-CoV-2 из России эволюционирует с накоплением новых мутаций, способствующих повышению трансмиссивности вируса. Данные о специфических мутационных паттернах генома SARS-CoV-2 могут быть использованы для выявления вируса, отслеживания и контроля его распространенности.

Ключевые слова: SARS-CoV-2, COVID-19, геном, коронавирус, мутации, российские изоляты

Вклад авторов: авторы внесли равный вклад в написание статьи.

✉ **Для корреспонденции:** Сергей Альбертович Кирьянов
а/я 181, г. Москва, 117587; kiryanov@dna-technology.ru

Статья получена: 26.06.2020 **Статья принята к печати:** 10.07.2020 **Опубликована онлайн:** 27.07.2020

DOI: 10.24075/vrgmu.2020.045

SPREAD OF VARIANTS WITH GENE N HOT SPOT MUTATIONS IN RUSSIAN SARS-COV-2 ISOLATES

Kiryanov SA [✉], Levina TA, Kirillov MYu

"DNA-Technology", Moscow, Russia

SARS-CoV-2 is a RNA coronavirus of the β -CoV family responsible for the pandemic of the severe acute respiratory syndrome (COVID-19). It is hypothesized that the viral genome of the contemporary predominant founder clones is likely to be evolving in geographic-dependent manner. Thus, strains of different origin may be characterized by different mutation patterns. The study was aimed to perform the mutational and phylogenetic analysis of the Russian SARS-CoV-2 genomes at different time periods and in various regions, as well as to characterize the mutational profiles of isolates using the bioinformatics approaches. The mutation accumulation was compared in 86 SARS-CoV-2 whole-genome sequences from Russia and 220 from Europe and North America in order to reveal the characteristic gene variations, the possible positive selection patterns. Along with the known mutation variants in the structural proteins genes, typical for isolates of European origin, several additional mutations including the synonymous mutation in gene *M* (C26750T) characteristic for Russian isolates were revealed. Double mutation R203K and G204R in the nucleocapsid gene which previously emerged in Europe began spreading and rapidly (within a month) became the dominant form in Russia. The results obtained indicate that the viral genome of most Russian isolates evolves with accumulation of new mutations associated with increased viral transmission. Data on the SARS-CoV-2 genome specific mutation patterns might be used for the detection of the virus, as well as for tracking and controlling of its spread.

Keywords: SARS-CoV-2, COVID-19, genome, coronavirus, mutations, Russian isolates

Author contribution: the authors contributed to the manuscript equally.

✉ **Correspondence should be addressed:** Sergei A. Kiryanov
PO Box 181, Moscow, 117587; kiryanov@dna-technology.ru

Received: 26.06.2020 **Accepted:** 10.07.2020 **Published online:** 27.07.2020

DOI: 10.24075/brsmu.2020.045

Характерной особенностью пандемии COVID-19 является быстрое распространение вируса во многих (более 187) странах мира [1]. Выявлены значительные различия в распространенности, уровне смертности и тяжести заболевания между географическими регионами, странами, а также среди возрастных групп инфицированных [2–4].

Обнаруженные различия можно объяснить тем, что геном SARS-CoV-2 эволюционирует и дифференцируется по географическому признаку (благодаря введению режима изоляции и физического дистанцирования) с образованием нескольких типов от исходного типа из г. Ухань (провинция Хубэй, КНР) [5, 6]. Известно, что РНК-вирусы легко накапливают мутации. В случае возникновения неадаптивных мутаций вирусы элиминируются. При наличии адаптивных мутаций можно ожидать, что вирус получит

селективное преимущество, а это, как правило, проявляется в высокой частоте указанных мутаций и, соответственно, в более высокой трансмиссивности вируса.

Быстрое распространение SARS-CoV-2 в мире вызывает вопрос о том, обусловлена ли его эволюция драйверными адаптивными мутациями и если да, то в каких генах.

Размер генома коронавируса составляет около 29 900 нуклеотидов. Геном SARS-CoV-2 кодирует протяженный полипротеин репликативного комплекса ORF1ab и четыре структурных белка: мембранный (M), поверхностный гликопротеин (S), оболочечный (E) и нуклеокапсидный фосфопротеин (N), как и у других β -коронавирусов [7].

Геном SARS-CoV-2 интенсивно исследуют, как в целях диагностики и анализа патогенности этого вируса, так и в целях отслеживания его эволюции. В настоящее

время в базе данных GISAID находится более 25 000 последовательностей вируса, выделенных в нескольких десятках стран. Ранее эволюционные изменения генома вируса SARS-CoV-2 прослеживали по паттернам мутаций в генах *ORF1ab*, гликопротеина S и неструктурных протеинов *nsp6* и *nsp8* [8–10]. Так, высокая частота мутаций C241T, C3037T и C14408T в гене *ORF1ab*, кодирующем белки репликативного комплекса, и мутации A23403G в гене S, продукты которого взаимодействуют с рецептором ACE2, были обнаружены в изолятах вируса SARS-CoV-2 из стран Западной Европы, где заболевание COVID-19 протекает более тяжело, чем в других географических регионах [11]. В совокупности указанные мутации, определяющие клад 20A (ранее известный как клад G), по-видимому, обуславливают повышенную трансмиссивность вируса и его доминирование в Европе.

Интерпретация данных по эволюции генома SARS-CoV-2 остается неполной, поскольку существующие публикации по-прежнему охватывают преимущественно изоляты, представленные в США, странах Европы, КНР, и некоторых других странах. В частности, недостаточно информации о мутационных профилях геномов изолятов SARS-CoV-2 российского происхождения, изучение которых особенно актуально в связи с контрастирующими друг с другом высоким ростом инфицирования и невысокой смертностью.

Целью исследования было провести мутационный и филогенетический анализы геномов SARS-CoV-2 из России в разные периоды и в разных регионах, а также охарактеризовать мутационные профили изолятов, используя подходы биоинформатики.

МАТЕРИАЛЫ И МЕТОДЫ

В период с 1 марта по 29 апреля 2020 г. с использованием баз данных NCBI и GISAID было отобрано 86 нуклеотидных последовательностей SARS-CoV-2 из России и 220 из Европы и США. Критерии включения: длина полноразмерной последовательности 26 000–30 000 п.н., последовательности аннотированы как SARS-CoV-2. Критерии исключения: повторно представленные последовательности, последовательности со слишком большим числом неопределенных нуклеотидов. Для выравнивания множественных последовательностей применяли программы Clustal Omega (EMBL-EBI; Великобритания) и Blast (NCBI; США). В качестве последовательности сравнения для анализа последовательностей, выделенных у пациентов из России, выбрали MT233519, SARS-CoV-2/human/ESP/Valencia5/2020.

Для филогенетического анализа последовательностей SARS-CoV-2, временного датирования предковых узлов, реконструкции дискретных признаков, частоты и датирования появления анкорных специфических мутаций на филогенетических ответвлениях использовали программу <https://nextstrain.org/> [12].

Предсказание В-клеточных эпитопов проводили по результатам анализа и в соответствии с ранее предложенным группой исследователей для SARS-CoV алгоритмом [13]. Для предсказания потенциальных В-клеточных эпитопов в первичной аминокислотной последовательности фосфопротеина N применяли следующие инструменты прогнозирования: для линейных В-клеточных эпитопов программу VeriPred-2.0 (DTU; Дания) [14], для конформационных эпитопов — программу

DiscoTope 2.0 (DTU; Дания) [15], предоставляемые сервером с базой данных эпитопов IEDB Immunobrowser (NIAID; США).

Потенциальные линейные В-клеточные эпитопы были предсказаны с использованием программы VeriPred-2.0 (DTU; Дания) [14] при максимальном пороговом значении 0,75, что соответствовало специфичности выше 0,85 и чувствительности ниже 0,40 (с учетом последовательностей, содержащих более семи аминокислотных остатков). Прогнозирование конформационных эпитопов проводили с использованием программы DiscoTope 2.0 (DTU; Дания) [15], применив ограничение позитивности выше –3,7, что соответствовало специфичности больше или равной 0,75 и чувствительности ниже 0,40.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

В рамках исследования было проанализировано 86 полногеномных последовательностей SARS-CoV-2, выделенных у российских пациентов в марте-апреле 2020 г. Из них 38 изолятов (44%) получили в марте и 56 (56%) — до конца апреля 2020 г. Нуклеотидные последовательности генома вируса подвергли выравниванию и сравнили с последовательностями SARS-CoV-2 от 220 европейских и американских пациентов, случайным образом отобранных из базы данных GISAID. Филогенетический анализ выбранных нуклеотидных последовательностей российского и европейского происхождения показал, что все российские изоляты, за исключением одного, принадлежали к кладу 20A, который ранее классифицировали как G (рис. 1). Все остальные последовательности, кроме одного выделенного в марте образца из Кабардино-Балкарии, содержали мутацию A23403G с заменой D614G в гене гликопротеина S, мутацию C14408T с заменой P314L в гене белка ORF1b, а также синонимичные мутации C241T и C3037T. В Европе указанные мутации ранее обнаружили в изоляте из Германии (Germany/BavPat1/2020), а затем в изолятах из Италии, выделенных в феврале. По-видимому, все перечисленные мутации образуют эволюционно стабильный гаплотип, который сейчас доминирует в европейских изолятах, изолятах с восточного побережья США, а также изолятах российского происхождения.

По наличию или отсутствию тройной мутации G28881A, G28882A и G28883C в гене N, вызывающей двойную несинонимичную мутацию R203K и G204R, последовательности геномов SARS-CoV-2 изолятов из России можно разделить на две неравные группы, из 59 и 26 последовательностей соответственно. Филогенетический анализ изолятов из России, Европы и США показал, что двойная мутация R203K и G204R, ранее обнаруженная в изоляте из Валенсии, Испания (MT233522, 2 марта 2020 г.), также образует отдельный субклад 20B (рис. 1). Следует отметить, что в отличие от российских изолятов большинство европейских и американских изолятов образует субклады без тройной мутации G28881A, G28882A и G28883C.

Субклад российских изолятов с двойной характеристической мутацией R203K и G204R делится на три неравные группы. Наиболее многочисленной является группа AP1, содержащая более 40 изолятов, выделенных преимущественно в Санкт-Петербурге и, по-видимому, имеющих итальянское происхождение. Указанная группа дивергировала от предшественника, ее определяет синонимичная мутация C26750T в гене M, специфичная только для рассмотренных российских

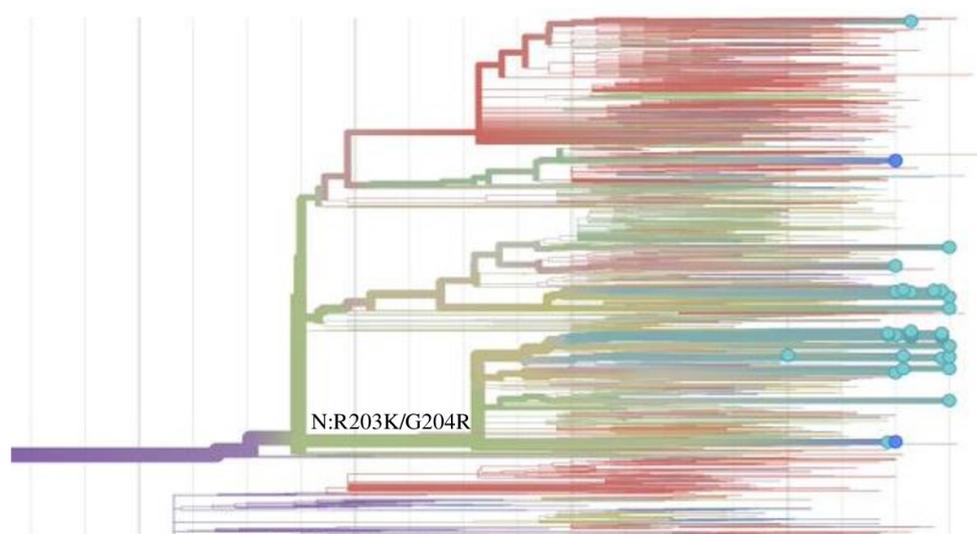


Рис. 1. Филогенетическое дерево геномов российских, европейских и американских изолятов SARS-CoV-2, дифференцированных по наличию или отсутствию двойной мутации R203K и G204R в гене *N*. Голубым цветом выделены российские изоляты, зеленым и темно-зеленым — европейские изоляты, темно-красным — американские изоляты

изолятов. Время появления данной мутации — не позднее начала марта. Для этой группы также характерен эффект микроклональности, определяемый накоплением преимущественно синонимичных мутаций в 5'-области гена *ORF1ab*, которые делят варианты геномов на дополнительные субпопуляции. Сведения о выделенных подгруппах и мутациях представлены в табл. 1.

Группу AP2 (шесть изолятов из Москвы и два из Якутии) определили мутации в гене *ORF1a* (G3278S, T1246I, L3606F) и синонимичная мутация C23731T в гене *S*. Подгруппа из четырех изолятов из Москвы и Якутии впоследствии дивергировала с накоплением мутаций A364S и дополнительной мутации M1499I в гене *ORF1ab*. Последняя мутация специфична только для изолятов российского происхождения, она появилась не позднее середины марта.

В еще одной группе изолятов AP3 (шесть изолятов из Москвы, Липецка и Краснодара) итальянского происхождения есть характерная мутация T175M в гене *M*. Изоляты различаются также по дальнейшему мутированию гена *ORF1ab* (мутации P892S, I1887V). В наименее представленной группе из трех изолятов (Москва) присутствует дополнительная мутация A152S в гене *N*, выявленная не позднее середины марта.

В группе из 26 изолятов без двойной мутации R203K и G204R в гене *N* накопление мутаций также происходит в основном в гене *ORF1ab*. Наиболее часто встречающейся дифференцирующей мутацией является синонимичная мутация в позиции A20268G испанского происхождения (обнаружена в 16 изолятах, преимущественно из Санкт-Петербурга). Наличие несинонимичных мутаций в гене *ORF1ab* (T265I, P3395L и др.), а также в генах *ORF3a* (Q57H) и *M* (D3G) позволило выделить несколько подгрупп с незначительным числом изолятов (4–6). Происхождение указанных изолятов и дополнительные мутации представлены в табл. 1. Только в трех изолятах из 26 обнаружили дополнительные несинонимичные мутации в гене *N*: двойную мутацию N140K и T205I, а также N140T и A397V.

Таким образом, независимо от происхождения варианты SARS-CoV-2 с двойной мутацией R203K и G204R в гене *N* доминируют в разных регионах России.

Для определения времени появления и распространения двойной мутации R203K и G204R был проведен анализ

наиболее представленных геномов вируса, полученных у пациентов из Москвы и Санкт-Петербурга в марте-апреле 2020 г., которые классифицировали в соответствии с датами выделения образцов (указанными в базе данных GISAID). Четыре временные подгруппы выделили следующим образом: 10–12 марта 2020 г. (геномы от восьми пациентов), 19–21 марта 2020 г. (геномы от девяти пациентов), 1–3 апреля 2020 г. (геномы от 16 пациентов), 10–12 апреля 2020 г. (геномы от 29 пациентов). Число других накапливаемых мутаций (в основном в гене *ORF1ab*) менялось в течение каждого временного периода: 2, 4 и 4 в группе с двойной мутацией R203K и G204R, и 2, 3 и 3 в группе геномов без такой мутации. При этом в последних присутствовали мутации в гене *N*: двойная мутация N140K и T205I, а также N140T. Дивергенция других генов не влияла на распространение вариантов с двойной мутацией R203K и G204R в гене *N*. В конце марта — начале апреля доля изолятов с двойной мутацией R203K и G204R выросла более чем в два раза и к середине апреля составила более 69,5% (рис. 2).

Были проверены общее распределение и представленность мутаций в гене *N* в нуклеотидных последовательностях изолятов из Европы и США, размещенных в базах данных GISAID и NCBI. Интересно, что в европейских популяциях представленность геномов субклада с двойной мутацией R203K и G204R в гене *N* оказалась существенно ниже, чем в России, и составила 32,6% (1068 геномов против 3241 без этой мутации). В США доля геномов того же субклада еще ниже, она составляет 13,3% (464 генома против 3479). Распределение несинонимичных мутаций в гене *N* оказалось неравномерным: 58,7% мутаций были локализованы на участке гена $N_{179-217}$. С помощью алгоритмов предсказания для линейных В-клеточных эпитопов в протеине *N* предсказали два варианта возможных линейных В-клеточных эпитопов с максимальным пороговым значением $> 0,758$, в позициях 23–36 и 178–207 соответственно (табл. 2). Конформационные В-клеточные эпитопы в гене *N* с пороговым значением $> -0,37$ и специфичностью 0,75 были предсказаны с использованием соответствующего алгоритма примерно в тех же позициях (26–36 и 193–207), включая фланкирующие R203, G204 и T205 (данные не представлены). Однако

Таблица 1. Субкластеры и характеристические мутации, выявленные в российских изолятах SARS-CoV-2

Субкластеры — мутации в гене <i>N</i>	*Мутации в <i>ORF1ab</i>	**Мутации в других генах	***Дополнительные мутации	Вероятное происхождение	Время появления	Регион России и число изолятов
AP1 - R203K G204R		M: c26750t	ORF1a: P959S - 7 ORF1a: P309L 8	Италия Италия	02.03.2020 16.04.2020	г. Санкт-Петербург (40) респ. Саха (Якутия) респ. Бурятия (5)
AP2 - R203K G204R	ORF1a: G3278S (g10097a) S: c23731t	ORF1a: A364S	ORF1b: M1499I	Англия	09–16.03.2020	г. Москва (2) респ. Саха (Якутия) (2)
		ORF1a: T1246I L3606F	ORF7a: L5F	Англия	02.03.2020	г. Москва (4)
AP3 - R203K G204R		M: T175M	ORF1a: P892S, ORF1b: I1887V N: A152S	Италия	01–10.03.2020 14.03.2020	г. Липецк (2) г. Краснодар (1) г. Москва (3)
R203 G204	a20268g		N: N140T 8 N: N140K, T205I; ORF8: R101L	Испания	04–08.03.2020	г. Санкт-Петербург (14) г. Псков (1) г. Омск (1)
R203 G204	ORF1a: T265I (c1059t)	ORF3a: Q57H (g25563t)	ORF1a: P3395L S: G261V, ORF1a: L3606F c29200t	Франция Австрия	04–16.03.2020 08–15.03.2020	г. Оренбург (2) о. Сахалин (1) г. Челябинск (1) г. Москва (2)
R203 G204		M: D3G (a26530g)	G4255T ORF1a: L890F; ORF1b: Q348H, M1156I; ORF1b: L1701F, V1615I	Англия	04–05.03.2020 11.03.2020	г. Москва (2) г. Москва (2)

Примечание: * — дифференцирующие мутации в гене *ORF1ab*, определяющие принадлежность изолята к подгруппе; ** — дифференцирующие мутации в других генах, определяющие принадлежность изолята к подгруппе; *** — другие частые мутации, выявленные в изолятах.

с учетом критерия специфичности (0,75) около 25% аминокислотных остатков могут быть предсказаны как часть В-клеточного эпитопа неверно.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

По имеющимся данным, геном SARS-CoV-2 эволюционирует с образованием нескольких типов, кластеризующихся в отдельные географически обусловленные группы [16]. Мутационный анализ географически обусловленных изолятов позволяет получить представление о наиболее частых адаптивных мутациях, обеспечивающих высокую трансмиссивность вируса. Ранее было обнаружено, что не менее пяти основных мутаций (C241T, C3037T, T28144C, C14408T, A23403G) оказались наиболее представленными в изолятах вируса SARS-CoV-2 из Западной Европы [11]. Указанные комутации, определяющие клад 20A, по-видимому обеспечивают повышенную трансмиссивность вируса и его доминирование в Европе. По результатам

мутационного и филогенетического анализов геномов SARS-CoV-2, обнаруженных в России в марте-апреле 2020 г., наиболее распространенным также оказался клад 20A, что указывает на европейское происхождение российских изолятов. Однако в России, в отличие от Западной Европы, распространился и доминирует субклад 20B с тройной мутацией G28881A, G28882A и G28883C, вызывающей двойную замену R203K и G204R в гене протеина N. Так, в России в конце апреля представленность геномов с двойной мутацией R203K и G204R составила более 69,5%, тогда как в Европе — 32,6%. В США число геномов того же субклада с мутациями R203K и G204R было еще ниже, их доля достигла 13,3%. Предположительно, указанный вариант появился в России в начале-середине марта 2020 г. Его дальнейшее распространение сопровождалось формированием новых подтипов с накоплением характерных мутаций в гене *M* (C26750T) или *ORF1b* (M1499I или G17964T) и последующей дивергенцией за счет новых единичных (в основном синонимичных)

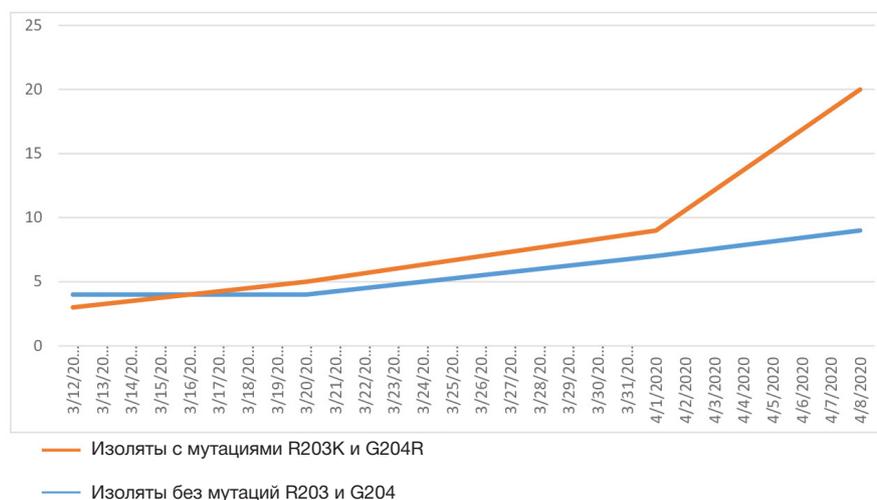


Рис. 2. Динамика представленности изолятов SARS-CoV-2 с мутациями R203K и G204R в гене *N* в России в марте-апреле 2020 г.

Таблица 2. Варианты предсказанных В-клеточных эпитопов в нуклеокапсидном белке N SARS-CoV2

Тип эпитопа	Позиция в белке N	Позиции мутаций	Аминокислотная последовательность эпитопа
В-клеточный	178–207	R203K G204R T205I	GGSQASSRSSSRNSSRNSTPGSSRGTS
В-клеточный	23–36		STGSNQNGERSGAR

мутаций в гене *ORF1ab*. Быстрое мутирование варианта с двойной мутацией R203K и G204R в гене *N*, по-видимому, отражает его способность к адаптации и повышению трансмиссивности, а не модуляции вирулентности.

Пока возможный функциональный эффект внедрения мутантного мотива AAACGA в ген нуклеокапсида неизвестен. Белок N отвечает за образование спирального нуклеокапсида в ходе сборки вириона и играет ключевую роль в его репликации и транскрипции. Он может вызывать иммунный ответ и быть потенциальной мишенью при разработке вакцин [17]. Ранее сообщали о локализации потенциальных В-клеточных и Т-клеточных эпитопов в гликопротеине S, мембранном M и капсидном фосфопротеине N, предсказанных по гомологичным участкам в геноме коронавируса SARS-CoV [18]. Предпринятая нами попытка картирования предсказанного пептида В-клеточного эпитопа N_{179–207} аминокислотной последовательности гена *N* допускает возможность того, что позиции R203 и G204 располагаются в пределах эпитопа. В результате мутаций R203K и G204R возникают два сильных положительно заряженных аминокислотных остатка в близких положениях, в отличие от только одного положительно заряженного остатка в немутантном генотипе, что, возможно, способствует снижению конформационной энтропии по сравнению с исходным генотипом. В настоящий момент биоинформационные подходы в отсутствие экспериментальных данных не позволяют оценить биологическую значимость этих мутаций.

Кроме того, нет оснований связывать распространенность данных мутаций вируса SARS-CoV-2 в России и его патогенность. Дальнейшее изучение эволюции генома вируса SARS-CoV-2 позволит не только реконструировать текущие эпидемиологические процессы, но и оптимизировать существующие ПЦР-РВ-диагностические тест-системы и осуществить поиск новых мишеней для разработки вакцин.

ВЫВОДЫ

Полученные данные свидетельствуют о том, что подавляющее большинство российских изолятов SARS-CoV-2 имеет европейское происхождение. Вирусный геном значительной части российских изолятов SARS-CoV-2 эволюционирует с накоплением новых мутаций, способствующих повышению трансмиссивности вируса. Накопление двойной мутации R203K и G204R в гене *N* в изолятах российского происхождения привело к быстрой популяционной экспансии данного варианта и его доминированию в России.

Выявление характерных вариантов геномов вируса SARS-CoV-2 в России дает представление о его дальнейшей адаптивной эволюции. Данные о специфичных мутационных паттернах генома SARS-CoV-2, в том числе в генах структурных белков N и M, могут быть использованы для выявления SARS-CoV-2, отслеживания и контроля его распространенности.

Литература

- Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020; 579 (7798): 270–3. PubMed PMID: 32015507.
- Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect Dis*. 2020; 20 (6): 669–7. PubMed PMID: 32240634.
- Spychalski P, Błazyńska-Spychalska A, Kobiela J. Estimating case fatality rates of COVID-19. *Lancet Infect Dis*. 2020; 20 (7): 774–5. PubMed PMID: 32243815.
- Khafaie MA, Rahim F. Cross-country comparison of case fatality rates of Covid-19/SARS-CoV-2. *Osong Public Heal Res Perspect*. 2020; 11 (2): 74–80. PubMed PMID: 32257772.
- Phan T. Novel coronavirus: from discovery to clinical diagnostics. *Infect Genet Evol*. 2020; 79: 104211. PubMed PMID: 32007627.
- Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of SARS-CoV-2 genomes. *Proc Natl Acad Sci*. 2020; 117 (17): 9241–3. PubMed PMID: 32269081.
- Boopathi S, Poma A, Kolandaivel P. Novel 2019 coronavirus structure, mechanism of action, antiviral drug promises and rule out against its treatment. *J Biomol Struct Dyn*. 2020; 1–10. PubMed PMID: 32306836.
- Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J Transl Med*. 2020; 18: 179. Available from: <https://doi.org/10.1186/s12967-020-02344-6>.
- Benvenuto D, Angeletti S, Giovanetti M, Bianchi M, Pascarella S, Cauda R, et al. Evolutionary analysis of SARS-CoV-2: how mutation of Non-Structural Protein 6 (NSP6) could affect viral autophagy. *J Infect*. 2020; 81 (1): 24–7. PubMed PMID: 32283146.
- Stefanelli P, Faggioni G, Lo Presti A, Fiore S, Marchi A, Benedetti E, et al. Whole genome and phylogenetic analysis of two SARS-CoV-2 strains isolated in Italy in January and February 2020: additional clues on multiple introductions and further circulation in Europe. *Euro Surveill*. 2020; 25 (13): 1–5. PubMed PMID: 2000305.
- Yin C. 2020 Genotyping coronavirus SARS-CoV-2: methods and implications. *Genomics*. 2020; 112 (5): 3588–6. PubMed PMID: 32353474.
- Hadfield J, Megill K, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*. 2018; 34 (23): 4121–3. PubMed PMID: 29790939.
- Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. *Cell Host Microbe*. 2020; 27 (4): 671–80. PubMed PMID: 32183941.
- Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res*. 2017; 45 (1): 24–9. PubMed PMID: 28472356.
- Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *PLoS Comput Biol*. 2012; 8 (12):

- e1002829: 1–10. PubMed PMID: 23300419.
16. Hahn G, Lee S, Weiss S. and Lange C. Unsupervised cluster analysis of SARS-CoV-2 genomes reflects its geographic progression and identifies distinct genetic subgroups of SARS-CoV-2 virus. *bioRxiv*. Preprint. 2020 Jun 30. Available from: <https://www.biorxiv.org/content/10.1101/2020.05.05.079061v2>
 17. Zhao P, Cao J, Zhao L-J, Qin Z-L, Ke J-S, Pan W, Ren H, et al. Immune responses against SARS-coronavirus nucleocapsid protein induced by DNA vaccine. *Virology*. 2005; 331 (1): 128–5. PubMed PMID: 15582659.
 18. Koyama T, Weeraratne D, Snowdon J, and Parida L. Emergence of drift variants that may affect COVID-19 vaccine development and antibody treatment. *Pathogens*. 2020; 9 (5): 324. PubMed PMID: 32357545.

References

1. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020; 579 (7798): 270–3. PubMed PMID: 32015507.
2. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect Dis*. 2020; 20 (6): 669–7. PubMed PMID: 32240634.
3. Spychalski P, Błażyńska-Spychalska A, Kobiela J. Estimating case fatality rates of COVID-19. *Lancet Infect Dis*. 2020; 20 (7): 774–5. PubMed PMID: 32243815.
4. Khafaie MA, Rahim F. Cross-country comparison of case fatality rates of Covid-19/SARS-CoV-2. *Osong Public Heal Res Perspect*. 2020; 11 (2): 74–80. PubMed PMID: 32257772.
5. Phan T. Novel coronavirus: from discovery to clinical diagnostics. *Infect Genet Evol*. 2020; 79: 104211. PubMed PMID: 32007627.
6. Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of SARS-CoV-2 genomes. *Proc Natl Acad Sci*. 2020; 117 (17): 9241–3. PubMed PMID: 32269081.
7. Boopathi S, Poma A, Koldaivel P. Novel 2019 coronavirus structure, mechanism of action, antiviral drug promises and rule out against its treatment. *J Biomol Struct Dyn*. 2020: 1–10. PubMed PMID: 32306836.
8. Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J Transl Med*. 2020; 18: 179. Available from: <https://doi.org/10.1186/s12967-020-02344-6>.
9. Benvenuto D, Angeletti S, Giovanetti M, Bianchi M, Pascarella S, Cauda R, et al. Evolutionary analysis of SARS-CoV-2: how mutation of Non-Structural Protein 6 (NSP6) could affect viral autophagy. *J Infect*. 2020; 81 (1): 24–7. PubMed PMID: 32283146.
10. Stefanelli P, Faggioni G, Lo Presti A, Fiore S, Marchi A, Benedetti E, et al. Whole genome and phylogenetic analysis of two SARSCoV-2 strains isolated in Italy in January and February 2020: additional clues on multiple introductions and further circulation in Europe. *Euro Surveill*. 2020; 25 (13): 1–5. PubMed PMID: 2000305.
11. Yin C. 2020 Genotyping coronavirus SARS-CoV-2: methods and implications. *Genomics*. 2020; 112 (5): 3588-6. PubMed PMID: 32353474.
12. Hadfield J, Megill K, Bell SM. Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*. 2018; 34 (23): 4121–3. PubMed PMID: 29790939.
13. Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. *Cell Host Microbe*. 2020; 27 (4): 671–80. PubMed PMID: 32183941.
14. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res*. 2017; 45 (1): 24–9. PubMed PMID: 28472356.
15. Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *PLoS Comput Biol*. 2012; 8 (12): e1002829: 1–10. PubMed PMID: 23300419.
16. Hahn G, Lee S, Weiss S. and Lange C. Unsupervised cluster analysis of SARS-CoV-2 genomes reflects its geographic progression and identifies distinct genetic subgroups of SARS-CoV-2 virus. *bioRxiv*. Preprint. 2020 Jun 30. Available from: <https://www.biorxiv.org/content/10.1101/2020.05.05.079061v2>
17. Zhao P, Cao J, Zhao L-J, Qin Z-L, Ke J-S, Pan W, Ren H, et al. Immune responses against SARS-coronavirus nucleocapsid protein induced by DNA vaccine. *Virology*. 2005; 331 (1): 128–5. PubMed PMID: 15582659.
18. Koyama T, Weeraratne D, Snowdon J, and Parida L. Emergence of drift variants that may affect COVID-19 vaccine development and antibody treatment. *Pathogens*. 2020; 9 (5): 324. PubMed PMID: 32357545.