

Анализ геномного разнообразия SARS-CoV-2 и эпидемиологических признаков адаптации возбудителя COVID-19 к человеческой популяции (Сообщение 1)

Я.М. Краснов¹, А.Ю. Попова², В.А. Сафронов¹, А.В. Федоров¹, Д.В. Баданин¹, С.А. Щербакова¹, В.В. Кутырев¹

1. ФКУЗ «Российский научно-исследовательский противочумный институт «Микроб»
Роспотребнадзора, Саратов, Российская Федерация,
2. Российская медицинская академия непрерывного профессионального образования,
Москва, Российская Федерация

Резюме

Вспышка COVID-19 началась в середине декабря 2019 года в городе Ухань (Китай) и быстро переросла в пандемию, которая продолжается и в настоящее время. В этом исследовании мы проанализировали информацию о 7947 полных геномах вируса SARS-CoV-2 опубликованную в базе GISAID, для изучения эволюции генома вируса при адаптации к организму человека за последние шесть месяцев. Проведен филогенетический анализ и оценка изменений в исследуемых вирусных геномах, которые могут быть разделены на четыре основных кластера. Ключевое направление эволюции генома вируса SARS-CoV-2 сопряжено с появлением единичной мутации в гене S (D614G). Установлено, что распространение штаммов с G мутацией ассоциировано не только с ростом заболеваемости, но и со снижением летальности в мире. Наблюдаемое увеличение потенциала распространения на фоне признаков снижения вирулентности, вероятно, является основной формой адаптации нового коронавируса к человеческой популяции и, по-видимому, будет продолжаться в дальнейшем в виде интеграции SARS-CoV-2 в структуру сезонных возбудителей ОРВИ.

Ключевые слова: COVID-19, SARS-CoV-2, пандемия, передача от человека человеку, вирулентность, эволюция возбудителя.

Введение

Коронавирусы вызывали эпидемические осложнения задолго до развивающейся в настоящее время пандемии COVID-19. В частности, бетакоронавирусы SARS-CoV и MERS-CoV были этиологическими агентами двух крупных вспышек SARS и MERS (Cui et al., 2019). Новый бетакоронавирус SARS-CoV-2, вызывающий пневмонию у человека – COVID-19, впервые зарегистрирован в декабре 2019 года в городе Ухань (провинция Хубэй, центральный Китай) (Wu F. et al., 2020; Zhou P. et al., 2020). Предполагается, что основными носителями в дикой природе вируса SARS-CoV-2 являются летучие мыши рода подковоносов (Lu R.J. et al., 2020; Zhou P. et al., 2020).

С момента первого обнаружения SARS-Cov-2 в декабре 2019 года в Китае накопился значительный массив сведений об изменчивости генома и динамике распространения COVID-19 в странах мира. За 6 с лишним месяцев интенсивной передачи от человека к человеку геном SARS-CoV-2, приспосабливаясь к новому хозяину, приобрел определенное количество нуклеотидных замен. В отличие от упомянутых вспышек SARS и MERS, распространение COVID-19 не удастся взять под контроль за счет использования профилактических и противоэпидемических мероприятий, что делает наиболее вероятным интеграцию этого возбудителя в этиологическую структуру возбудителей ОРВИ в среднесрочной перспективе. В свою очередь внедрение SARS-Cov-2 в ряд возбудителей ОРВИ может сопровождаться снижением его вирулентных свойств по причине селективных преимуществ тех геновариантов, которые легче передаются от человека к человеку. В противовес этому, отрицательной селекции подвергаются штаммы возбудителя, способствующие развитию более тяжелого инфекционного процесса, так как пациенты с ярко выраженной клиникой чаще обращают на себя внимание, госпитализируются в медицинские учреждения и, вследствие чего, снижается возможность дальнейшей передачи вируса.

Наша работа направлена на характеристику вариаций вируса SARS-CoV-2 и выявление эволюционных изменений его генома с учетом тенденций приобретения новых эпидемиологических свойств.

Материалы и методы

Для проведения филогенетического анализа использовались данные, представленные на сайте организации GISAID (<https://www.gisaid.org>), по нуклеотидным последовательностям геномов вируса SARS-CoV-2, полученных из образцов клинического материала. Кроме того, была использована информация на сайте GISAID о количестве и качестве мутаций, закрепившихся в геноме вируса SARS-CoV-2 в период пандемии. По состоянию на 06 июля 2020 года в этой базе содержалось 37952 полных последовательностей геномов вируса SARS-CoV-2 с высокой степенью покрытия, полученных из всех регионов мира. В анализ мы взяли только нуклеотидные последовательности геномов вируса с высоким качеством прочтения. Для отбора, последовательности геномов выравнивали (использовался сервис <https://mafft.cbrc.jp/alignment/server/>) по референсу - Wuhan-Hu-1 (NC_045512.2). Затем последовательности выровненных геномов обрезами на 106 нуклеотидов от начала генома и на 390 нуклеотидов от его правого края по геному референсной последовательности. В случае наличия в оставшейся части последовательности генома хотя бы одного не распознанного нуклеотида (N) этот геном удалялся из выборки для анализа. По представленным критериям отобрано 7947 нуклеотидных последовательностей геномов вируса SARS-CoV-2. Построение филогенетических деревьев проводили с помощью программного обеспечения BioNumerics 7.6 (<https://www.applied-maths.com>) с использованием алгоритма максимальной экономии (maximum parsimony).

Для оценки эпидемиологических тенденций использованы данные о ежедневном числе регистрируемых случаев COVID-19 и летальных исходов в странах мира по данным университета Джонса Хопкинса на 13 июля 2020 года - 32524 значения по 188 странам за 173 дня наблюдений (<https://github.com/CSSEGISandData/COVID-19>).

Результаты

Нами изучена динамика эволюционной изменчивости генома вируса SARS-CoV-2 за период с конца декабря 2019 по июнь 2020 года (рисунки 1–5). Подтверждено наличие ключевых мутаций в геноме вируса SARS-CoV-2 в процессе его адаптации к организму человека при многократной передаче внутри человеческой популяции. Общее число мутаций, которые в совокупности дают все секвенированные геномы вируса SARS-CoV-2, составляет несколько тысяч, однако редкие из них зафиксировались и стабильно наследуются. На примере сделанной нами выборки, включающей 7947 геномов вируса SARS-CoV-2, секвенированных с конца декабря 2019 по июнь 2020 года, мы установили, что в совокупности эти геномы насчитывают 4887 единичных мутаций, отличающих их от исходного (референсного) штамма Wuhan-Hu-1 (NC_045512.2). Мы определили только 22 единичные мутации, сохраняющиеся в $\geq 3\%$ от числа исследуемых геномов вируса SARSCoV-2 (рисунок 8).

Результаты проведенного нами анализа показывают какие из мутаций в геноме вируса дали преимущества вариантам штаммов вируса SARS-CoV-2 и позволили получить широкое распространение. Так, на ранних этапах развития пандемии COVID-19 (до середины января 2020 года) в геноме вируса SARS-CoV-2 стабильно зафиксировались три основные группы мутаций, сформировавшие, соответственно, три ветви эволюционного развития (кластер 1–3, на рисунках 1–5). Ключевые мутации, сформировавшие эти кластеры, следующие: кластер - 1 (C8782T, T28144C), также известный как кластер «S»; кластер - 2 (G11083T, G26144T), также известный как кластер «V»; кластер - 3 (G1397A/C6312A, G11083T, C28311T/T28688C). Как видно, основное число этих мутаций относится к гену ORF1ab, однако всего одна из них приводит к замене аминокислоты (подчеркнутая позиция G11083T присутствует в кластере 2 и 3), остальные единичные мутации являются синонимичными. Три другие, не синонимичные мутации, относятся к генам: ORF8 (T28144C, кластер 1), ORF3a (G26144T, кластер 2) и N (C28311T, кластер 3). Развитие этих эволюционных направлений (трех кластеров) замедляется к концу марта 2020 года, когда в основном начинают обнаруживаться варианты штаммов вируса с G мутацией в S гене (позиция D614G).

Появление первых вариантов вируса с этой мутацией произошло, по-видимому, в первой половине января 2020 года и в дальнейшем привело к бурному распространению

штаммов вируса SARS-CoV-2, содержащих данную мутацию. Геномы штаммов вируса SARS-CoV-2, не содержащих мутацию D614G, уже к апрелю 2020 года встречаются только в 25 % случаев (таблица). На территории России также наблюдается наличие вариантов штаммов вируса с G мутацией в большинстве представленных геномов. Из 223 геномов вируса SARS-CoV-2, секвенированных в России, 220 геномов (98,6 %) содержат G мутацию (рисунок 6). Детальное сравнение геномов вируса SARS-CoV-2 на территории России и зарубежных стран показывает, что именно варианты штаммов вируса, имеющие G мутацию и циркулирующие на территории Западной Европы в марте–апреле 2020 года были массово завезены в Россию.

На основе обобщения эпидемиологических данных нами построены диаграммы, отражающие динамическое соотношение встречаемости G-геновариантов и показателей ежедневного выявления инфицированных и умерших от COVID-19 (рисунок 7). На графике показано, что уже с марта 2020 года наблюдается устойчивая тенденция к росту ежедневного числа регистрируемых больных на фоне распространения штаммов относящихся к G-геноварианту с 16 % к началу марта до 97 % в конце июня (таблица), что косвенно подтверждает наличие более высокого потенциала к передаче вируса от человека к человеку. При этом с середины апреля 2020 года наблюдается динамическое снижение коэффициента суточного числа летальных исходов к суточному числу зарегистрированных больных с 10 до 2 %.

Представленные данные свидетельствуют в пользу того, что распространение штаммов вируса с G мутацией может быть связано с его биологическими свойствами, облегчающими передачу от человека к человеку, при этом ассоциированное снижение показателя летальности может отражать ослабление его вирулентных свойств.

Обсуждение

Скорость мутаций в геноме вируса SARS-CoV-2 на ранних этапах эпидемии (середина января 2020 года) оценивалась в интервале $5,2-8,1 \cdot 10^{-3}$ на нуклеотид в год (Benvenuto D. et al., 2020). Позже на сайте GISAID на выборке из тысяч геномов вируса SARS-CoV-2, представленных за период в несколько месяцев эпидемии, скорость его мутации определена на порядок ниже, – около $8 \cdot 10^{-4}$ на нуклеотид в год (<https://www.gisaid.org/epiflu-applications/next-hcov-19-app/>). Для вируса SARS-CoV, послужившего причиной вспышки атипичной пневмонии в Китае в 2003 году, скорость мутации оценивалась $0,8-2,38 \cdot 10^{-3}$ на нуклеотид в год (Zhao Z. et al., 2004). По-видимому, новый вариант вируса SARS-CoV-2 значительно уступает в скорости мутации близкородственному вирусу SARS-CoV.

Проведенный нами сравнительный анализ частоты мутаций показывает, что большинство вновь секвенированных геномов вируса SARS-CoV-2, включая варианты конца июня 2020 года, до сих пор не отличаются от генома штамма референса (Wuhan-Hu-1, NC_045512.2) более чем на 10–12 единичных мутаций и за редким исключением – более 15. Исходя из размера генома вируса SARS-CoV-2 в 29903 нуклеотида, однородность его генома в течение 6 месяцев развития пандемии сохраняется на уровне $\geq 99,93\%$. На примере сделанной нами выборки установлено, что в совокупности эти геномы содержат 4887 единичных мутаций, отличающих их от исходного (референсного) штамма Wuhan-Hu-1 (NC_045512.2). Однако мы определили только 22 единичные мутации, которые сохраняются в значительном количестве геномов вируса SARS-CoV-2 в исследуемой выборке (число геномов $\geq 3\%$). Остальные единичные мутации – уникальные, или редко встречающиеся (рисунок 8). Данные 22 единичные мутации располагаются в следующих генах: ORF1ab (13 ед. мут.), S (1 ед. мут.), ORF3a (2 ед. мут.), ORF8 (2 ед. мут.) и N (1 ед. мут.); мутация в позиции C241T относится к межгенной области. Остальные гены остались без изменений (ген M, E, ORF6, ORF7ab и ORF10). Функциональные характеристики указанных выше генов, несущих закрепившиеся мутации, можно разделить на две основные группы: гены, ответственные за синтез не структурных или вспомогательных белков вируса – ORF1ab, ORF3a и ORF8, и гены, синтезирующие соответствующие структурные белки вириона, – S и N. Белок S образует гомотримерные шипы на поверхности вирусной частицы и опосредует вход вируса в клетку хозяина. Белок N связывает вирусную РНК и требуется для упаковки вирусной РНК в вирусную

частицу во время ее сборки. В то же время белки М и Е не имеют мутаций в значительном числе геномов вируса, хотя являются важными белками в структуре вириона – белок М придает ему форму и имеет решающее значение вместе с белком Е в организации сборки вируса и формировании зрелых вирусных частиц (Mousavizadeh L., 2020; Ashour H.M., 2020). Таким образом, наиболее важные гены вируса SARS-CoV-2, отвечающие за его взаимодействие с клеткой хозяина, имеют закрепившиеся мутации в процессе эволюционной адаптации к организму человека. Следует отметить, что все эти 22 основные закрепившиеся мутации появились в геноме вируса в период с конца декабря 2019 года по середину февраля 2020 года. Далее за последующие 4,5 месяца в эволюции генома вируса не наблюдается появление новых закрепившихся мутаций у значительного числа штаммов вируса.

Исходя из филогенетических связей, представленных на рисунках 1–5, видно, что геном вируса SARS-CoV-2 с января 2020 года практически одновременно приобретает ряд ключевых мутаций, которые в дальнейшем сформируют четыре основных кластера его геновариантов, в трех из которых отсутствует мутация D614G в S гене, а четвертый кластер (G кластер) содержит G мутацию (S ген, D614G), и является основным эволюционным вектором развития. Появление и рост встречаемости данной мутации уже отмечено в ряде публикаций (Pachetti M. *et al.*, 2020; Korber B. *et al.*, 2020), а также представлено на сайте GISAID.

Ключевые мутации, сформировавшие три первые кластера, следующие: кластер - 1 (C8782T, T28144C), так же известный как кластер «S»; кластер - 2 (G11083T, G26144T), так же известный как кластер «V»; кластер - 3 (G1397A/C6312A, G11083T, C28311T/T28688C). При этом по одной единичной мутации, приводящей к замене аминокислоты (подчеркнутая позиция) присутствует в четырех генах: ORF1ab (G11083T, кластер 2 и 3), ORF8 (T28144C, кластер 1), ORF3a (G26144T, кластер 2) и N (C28311T, кластер 3). Все эти гены, за исключением гена N (нуклеокапсид), ответственны за синтез неструктурных или вспомогательных белков вируса (Mousavizadeh L., 2020; Ashour H.M., 2020). Проведя анализ по датам забора клинического материала и странам, на территории которых зафиксированы геноварианты из трех вышеописанных кластеров становится очевидным, что к апрелю 2020 года варианты штаммов вируса SARS-CoV-2 из кластеров 1–3 распространились по большинству стран и регионов, расположенных на всех континентах. Мы построили филогенетические связи между геномами вируса, обозначив

цветом каждый месяц, добавляя данные о геномах вируса SARS-CoV-2 в каждом новом месяце к предыдущим (рисунки 1–5). На рисунке 1 отображена ситуация филогенетических связей с конца декабря 2019 по февраль 2020 года и показано начало формирования G кластера на фоне уже сильно распространившихся геновариантов из кластеров 1–3 вариантов (коричневый – декабрь 2019 года, красный – январь и желтый – февраль 2020 года). Далее на рисунке 2 показано, как в течение марта 2020 года число геномов из G кластера превысило число геновариантов, приходящихся на кластеры 1–3 (геномы за март выделены зеленым цветом). Сам G кластер разделился на несколько подкластеров, что говорит о бурном эволюционном процессе и распространении геновариантов из G кластера в человеческой популяции. В апреле (выделен синим), мае (фиолетовый) и июне (розовый) наблюдается дальнейший рост количества геновариантов в G кластере на фоне постепенного уменьшения числа геновариантов вируса SARS-CoV-2 в кластерах 1–3 (рисунки 3–5). В численном соотношении этот процесс представлен в таблице.

Таким образом, вероятно, что на текущий момент геноварианты вируса SARS-CoV-2 из кластеров 1-3 оказывают слабое влияние на эпидемиологическую ситуацию, так как их встречаемость не превышает 3%. Причиной этому послужило появление и широкое распространение геновариантов из G кластера, более адаптированных к организму человека, которые конкурентно вытеснили более ранние геноварианты, циркулирующие на территории всех континентов.

Первые штаммы вируса SARS-CoV-2 с мутацией D614G (S-белок) появляются в январе 2020 года. В нуклеотидной последовательности генома вируса эта мутация находится в позиции 23403 (S ген, замена A на G), что в аминокислотной последовательности S-белка привело к замене в позиции 614 аспарагиновой кислоты (D) на глицин (G). Кодированный S геном белок – гликопротеин (spike glycoprotein), отвечает за функцию узнавания клетки хозяина, связывания с ней и проникновения в клетку (Chen W. et al., 2020). Помимо основной мутации D614G в S гене, в феврале и марте 2020 года происходит формирование ряда дополнительных единичных замен в геноме вируса SARS-CoV-2, которые разделяют сформированный ими кластер G на подкластеры (рисунки 1–2). Профиль основных генетических мутаций в кластере G представлен следующими единичными мутациями: исходный кластер G (C241T, C3037T, C14408T, A23403G); подкластер G-1 (C241T, C1059T, C3037T, C14408T, A23403G, G25563T), известный как GH кластер; подкластер G-

2 (C241T, C3037T, C14408T, A23403G, комплексная мутация трех нуклеотидов подряд – GGG28881-28883AAC), известный как GR кластер; подкластер G-3 (C241T, C3037T, C14408T, A20268G, A23403G); подкластер G-4 (C241T, C2416T/C3037T, C14408T, A23403G, G25563T) и подкластер G-5 (C241T, C3037T, C14408T, C15324T, A23403G). В связи с этим следует отметить, что первые секвенированные геномы геновариантов вируса, содержащие мутацию D614G (24.01.2020, Китай, EPI_ISL_451345; 27.01.2020, Великобритания, EPI_ISL_474597; 28.01.2020, Германия, EPI_ISL_450198), не получили массового распространения в человеческой популяции пока в их геноме, ориентировочно в середине февраля, не появилась синонимичная единичная мутация в позиции 14408 (замена С на Т, ген ORF1ab), что соответствует полному профилю нуклеотидных замен у исходных вариантов генома вируса в G кластере (C241T, C3037T, C14408T, A23403G). Далее, в феврале происходит появление ряда единичных мутаций формирующих пять подкластеров G. Подкластер G-1 сформировался благодаря дополнительным мутациям в ORF1ab гене (C1059T) и гене ORF3a (G25563T); геномы вируса из подкластера G-2, помимо исходного профиля мутаций G кластера, имеют комплексную мутацию трех нуклеотидов подряд в последовательности гена N – GGG28881-28883AAC. Подкластеры G-3 (A20268G, ген ORF1ab), G-4 (C2416T, ген ORF1ab и G25563T, ген ORF3a) и G-5 (C15324T, ген ORF1ab) формируются за счет дополнительных синонимичных и не синонимичных мутаций, указанных в скобках. Как видно, кроме ключевой мутации A23403G (или D614G в аминокислотном варианте), геномы вируса из кластера G имеют ряд закрепившихся мутаций, отличных от геномов вируса из кластеров 1–3. В итоге в конце февраля и первых числах марта от основания кластера G появляются пять быстро развивающихся ответвлений (подкластеров) G-1, G-2, G-3, G-4 и G-5, наиболее представленными из которых будут подкластеры G-1 и G-2 (рисунки 1–5).

Частота встречаемости вариантов вируса SARS-CoV-2 с геномами, образующими G кластер в образцах клинического материала положительного на COVID-19, составляет на сегодняшний день около 97 % (рисунок 5, таблица).

На территорию России вирус SARS-CoV-2 завезен массово в марте 2020 года, прибывшими из-за рубежа. На рисунке 6 представлено филогенетическое дерево из 223 геномов вируса SARS-CoV-2, полученных на территории России. Только в 3-х геномах (1,4 %) вируса SARS-CoV-2 из данной выборки не было G мутации. Филогенетический анализ показал, что основное количество геномов (87%) вируса SARS-CoV-2,

циркулирующего в России относится к подкластерам G-2 и G-3, а также к исходным вариантам G кластера (146, 24 и 24 геномам соответственно). Варианты данных геномов вируса SARS-CoV-2 типичны для территории Западной Европы в период конца февраля - апреля 2020 года. Близкие им варианты генома только в небольших количествах секвенированы из клинического материала от пациентов из Северной и Южной Америки, Саудовской Аравии и Индии. К подкластеру G-1 относились всего 19 геномов вируса, полученных из клинического материала от пациентов из России. Варианты геномов вируса SARS-CoV-2, формирующих подкластер G-1, типичны для территории Северной Америки. Таким образом, почти все генетическое разнообразие вариантов генома вируса SARS-CoV-2, распространенных на территории России, относится к кластеру G и в пределах этого кластера разделяется на 5 групп: G, G-1, G-2, G-3 и G-4.

В результате, по полученным нами данным, наиболее вероятным регионом, из которого осуществлен завоз штаммов вируса SARS-CoV-2 на территорию России, является Западная Европа. В это время в Западной Европе в большинстве секвенированных образцов уже содержались геноварианты вируса с мутацией D614G (Pachetti M. *et al.*, 2020).

В литературе описано две гипотезы, объясняющие как мутация D614G связана с повышенной передачей вируса SARS-CoV-2. Первый гипотеза основан на структурном положении мутации D614G (Wrapp D. *et al.*, 2020; Korber B. *et al.*, 2020) и связана с уменьшением взаимодействия между субъединицами S1 и S2 в S-белке, облегчая их расщепление и, соответственно, дальнейшее слияние вируса со стенкой клетки хозяина и последующим проникновением в клетку. Вторая гипотеза состоит в том, что эта мутация может косвенно усилить взаимодействие рецептор-связывающего домена на поверхности субъединицы S1 в S-белке вируса с ангиотензин-превращающим ферментом 2 (связь RBD-ACE2), что также способствует ускоренному взаимодействию частицы вируса с клеткой (Walls A.C. *et al.*, 2020; Wrapp D. *et al.*, 2020).

Таким образом, мы наблюдаем ярко выраженный вектор эволюционного развития и повсеместного распространения штаммов вируса SARS-CoV-2, имеющих G мутацию, что, видимо, обусловлено ускоренным процессом связывания вируса с клеткой хозяина и проникновением в нее, по отношению к штаммам вируса SARS-CoV-2, не имеющим данной мутации. Характерно, что указанный тренд наблюдается как в мире в целом, так и

при независимом анализе в разрезе отдельных регионов, что подтверждает наличие эволюционного преимущества G-геновариантов (рисунок 7). Обращает внимание тот факт, что распространение штаммов G кластера ассоциировано не только с ростом заболеваемости (до 200 и более тысяч ежедневно выявляемых инфицированных), но и со снижением летальности, как показано на рисунке. Уже с марта 2020 года наблюдается устойчивая тенденция к росту ежедневного числа регистрируемых больных на фоне распространения штаммов относящихся к G кластеру с 16% к началу марта до 97% в конце июня, что косвенно подтверждает наличие более высокого потенциала передачи вируса от человека к человеку. При этом с середины апреля 2020 года наблюдается динамическое снижение летальности с 10% до 2% (оцененной в данном исследовании как отношение суточного числа летальных исходов к суточному числу зарегистрированных больных в процентном выражении). Представленные данные дополнительно свидетельствуют в пользу того, что распространение штаммов вируса относящихся к G кластеру может быть связано с его биологическими свойствами, облегчающими передачу от человека к человеку, при этом ассоциированное снижение летальности может отражать ослабление вирулентных свойств.

Одним из актуальных вопросов оценки глобального риска, связанного с пандемией COVID-19, остается интерпретация наблюдаемых изменений в геноме, что имеет значение для долгосрочного прогноза развития эпидемического процесса. До тех пор, пока отсутствует общедоступная и эффективная вакцина, развитие эпидемического процесса обуславливается, с одной стороны, биологическими свойствами возбудителя, а с другой – комплексом неспецифических профилактических и противоэпидемических мероприятий. При сохранении текущих темпов распространения COVID-19 формирование иммунной прослойки будет в основном зависеть от напряженности и длительности формируемого постинфекционного иммунитета. В описанных условиях SARS-CoV-2 будет встраиваться в этиологическую структуру возбудителей сезонных острых респираторных инфекционных болезней.

Следует подчеркнуть, что эволюция SARS-CoV-2 продолжается и требует непрерывной научной оценки с использованием передовых методов анализа генетических последовательностей. При этом представленные данные свидетельствуют в пользу того, что распространение штаммов вируса из G кластера может быть связано с их

биологическими свойствами, облегчающими передачу от человека к человеку, в то время как снижение летальности может отражать ослабление вирулентных свойств.

Дальнейшее развитие представленной работы будет вестись в направлении детального анализа эволюции генома SARS-CoV-2 в регионах мира с акцентом на Российскую Федерацию. Отдельного внимания заслуживает изучение биологических свойств геновариантов SARS-CoV-2.

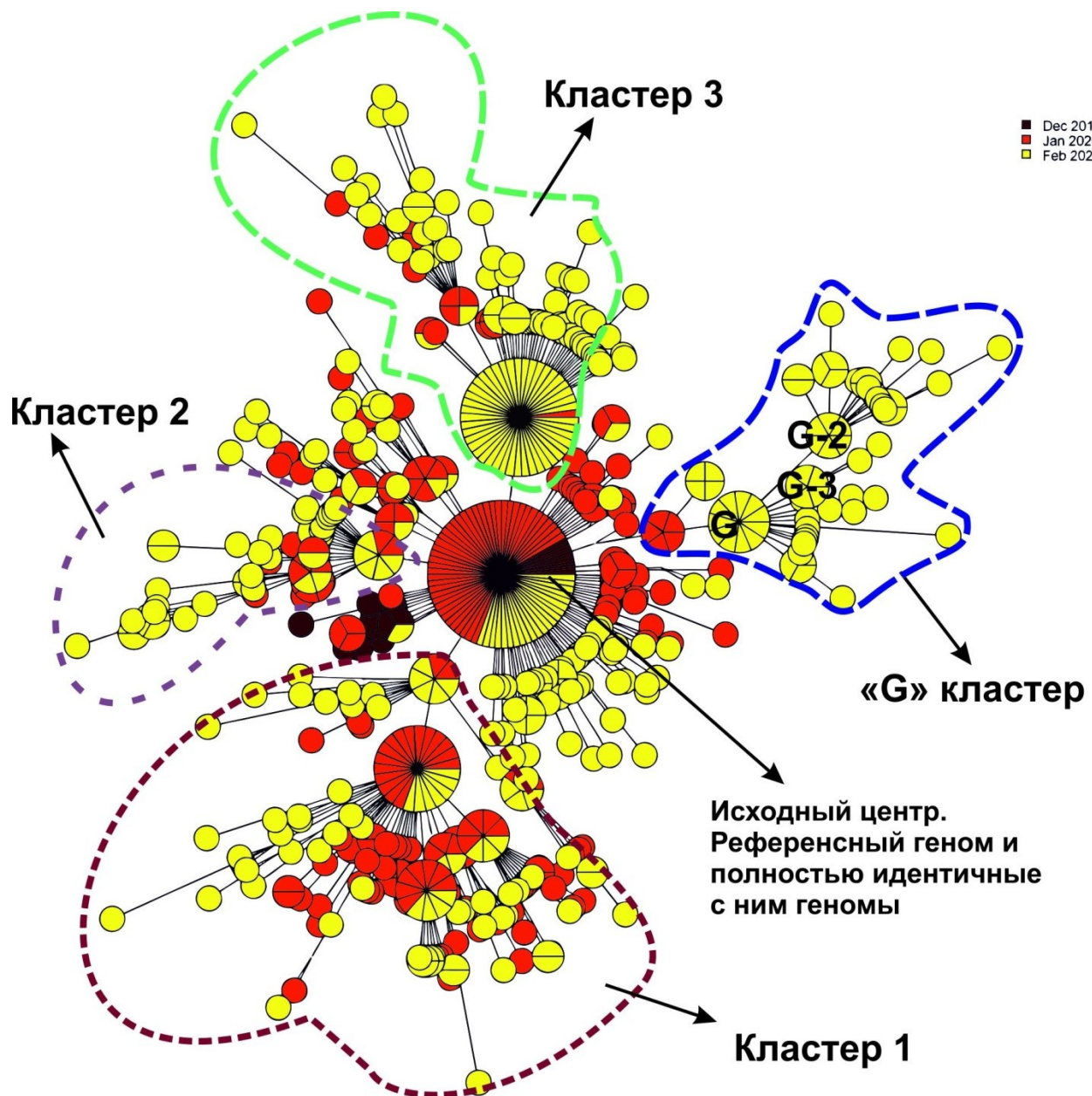


Рисунок 1. Филогенетические связи геномов вируса SARS-CoV-2, полученных из клинического материала от пациентов из разных регионов мира в период с декабря 2019 г. по февраль 2020 г. (коричневый – декабрь, красный – январь, желтый - февраль). Первые геномы штаммов вируса SARS-CoV-2, имеющие мутацию D614G в S гене, обнаружены в конце января 2020 г. в Китае, затем в Великобритании и Германии. Новые штаммы SARS-CoV-2 сформировали в феврале G кластер, с последующим разветвлением на подкластеры G-2 и G-3.

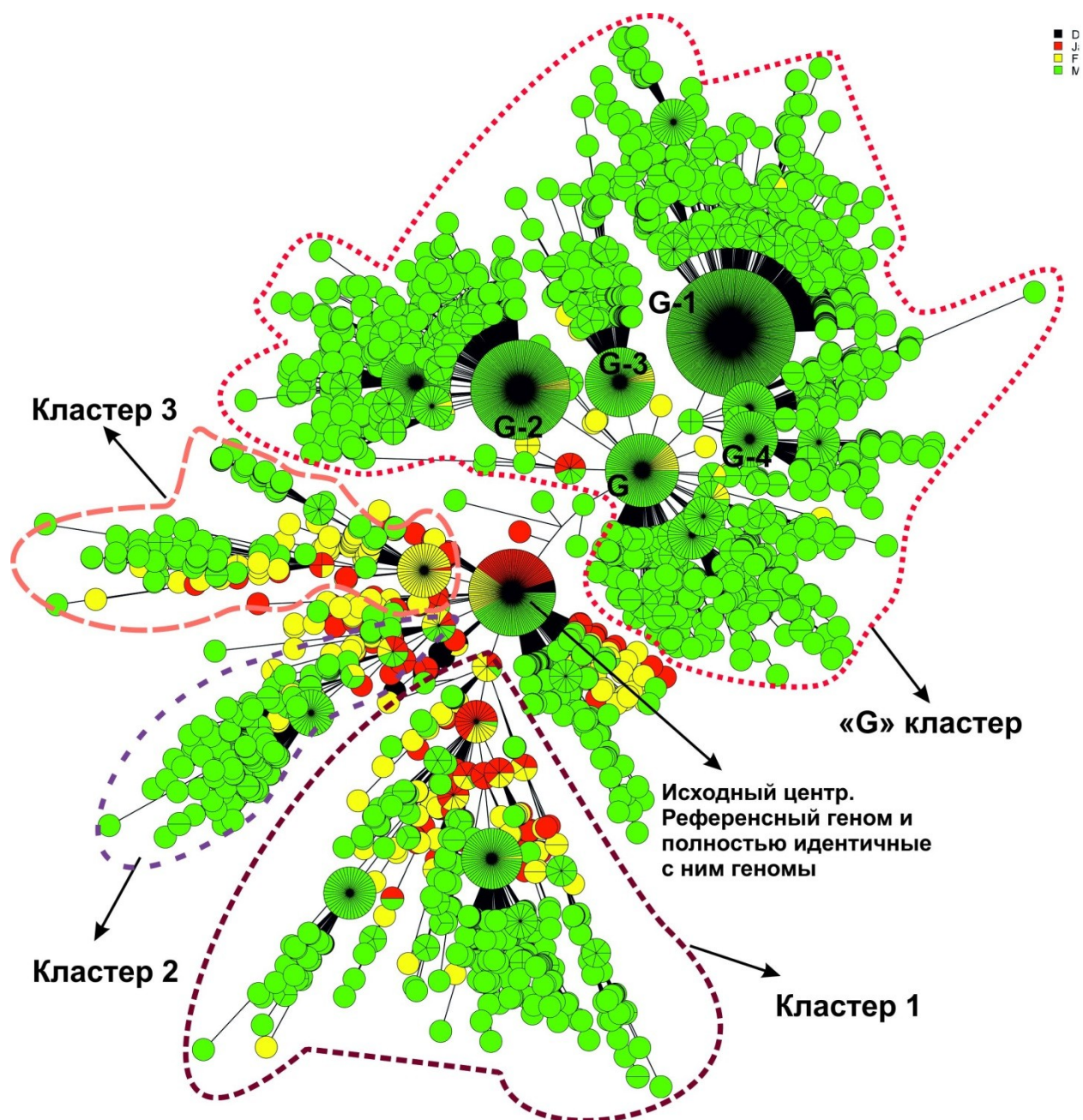


Рисунок 2. Филогенетические связи геномов вируса SARS-CoV-2, полученных из клинического материала от пациентов из разных регионов мира в период с декабря 2019 г. по март 2020 г. В течение марта (выделен зеленым) наблюдается бурный рост числа образцов клинического материала, положительного на COVID-19, в которых геном вируса SARS-CoV-2 имеет мутацию D614G в S гене.

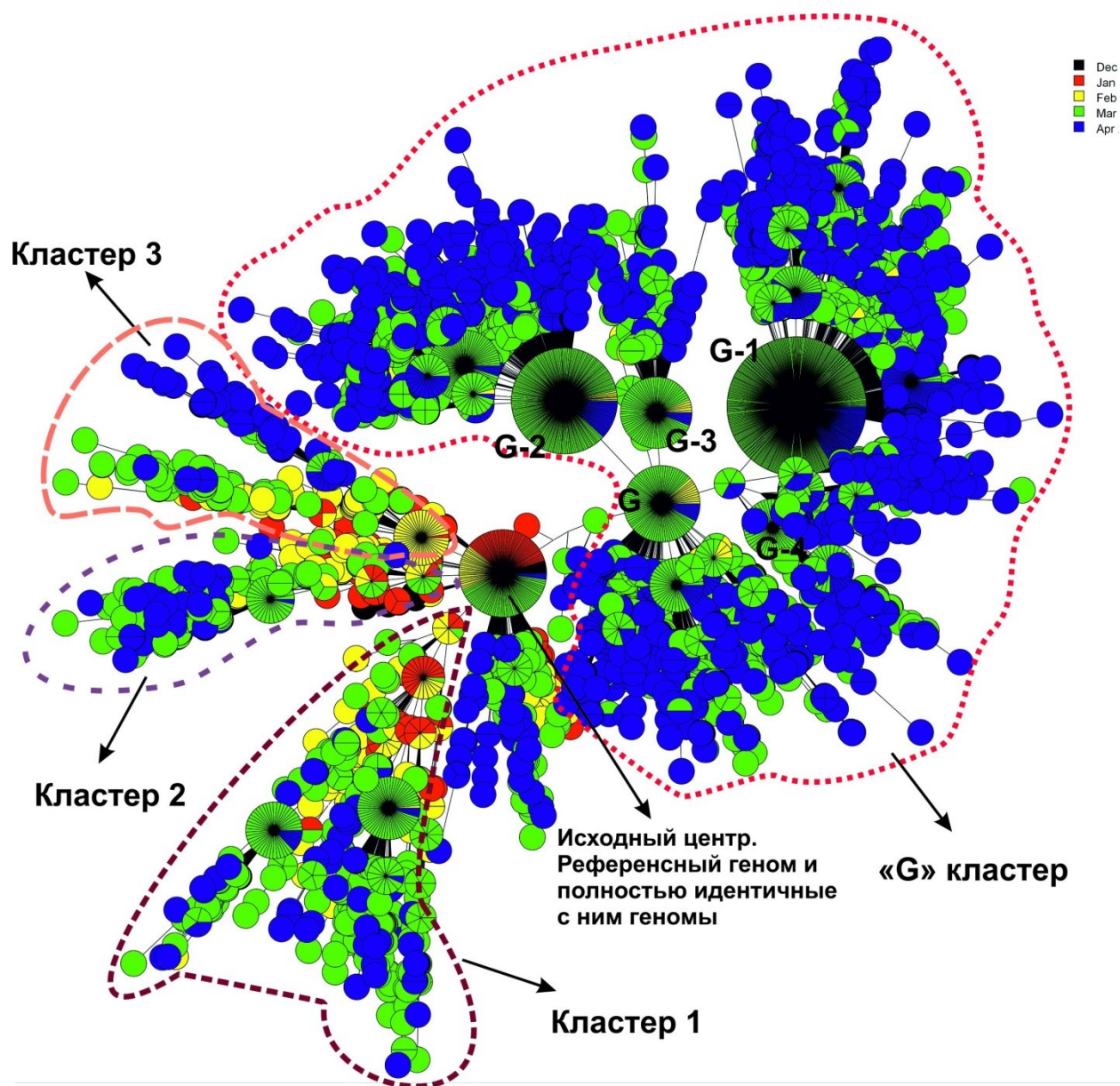


Рисунок 3. Филогенетические связи геномов вируса SARS-CoV-2, полученных из клинического материала от пациентов из разных регионов мира в период с декабря 2019 г. по апрель 2020 г. В апреле (выделен синим) наблюдается дальнейший бурный рост числа образцов клинического материала, положительного на COVID-19, в которых геном вируса SARS-CoV-2 имеет мутацию D614G в S гене. Число новых секвенированных геномов вируса SARS-CoV-2, не имеющих мутацию D614G, к маю 2020 г. не превышает 14 %

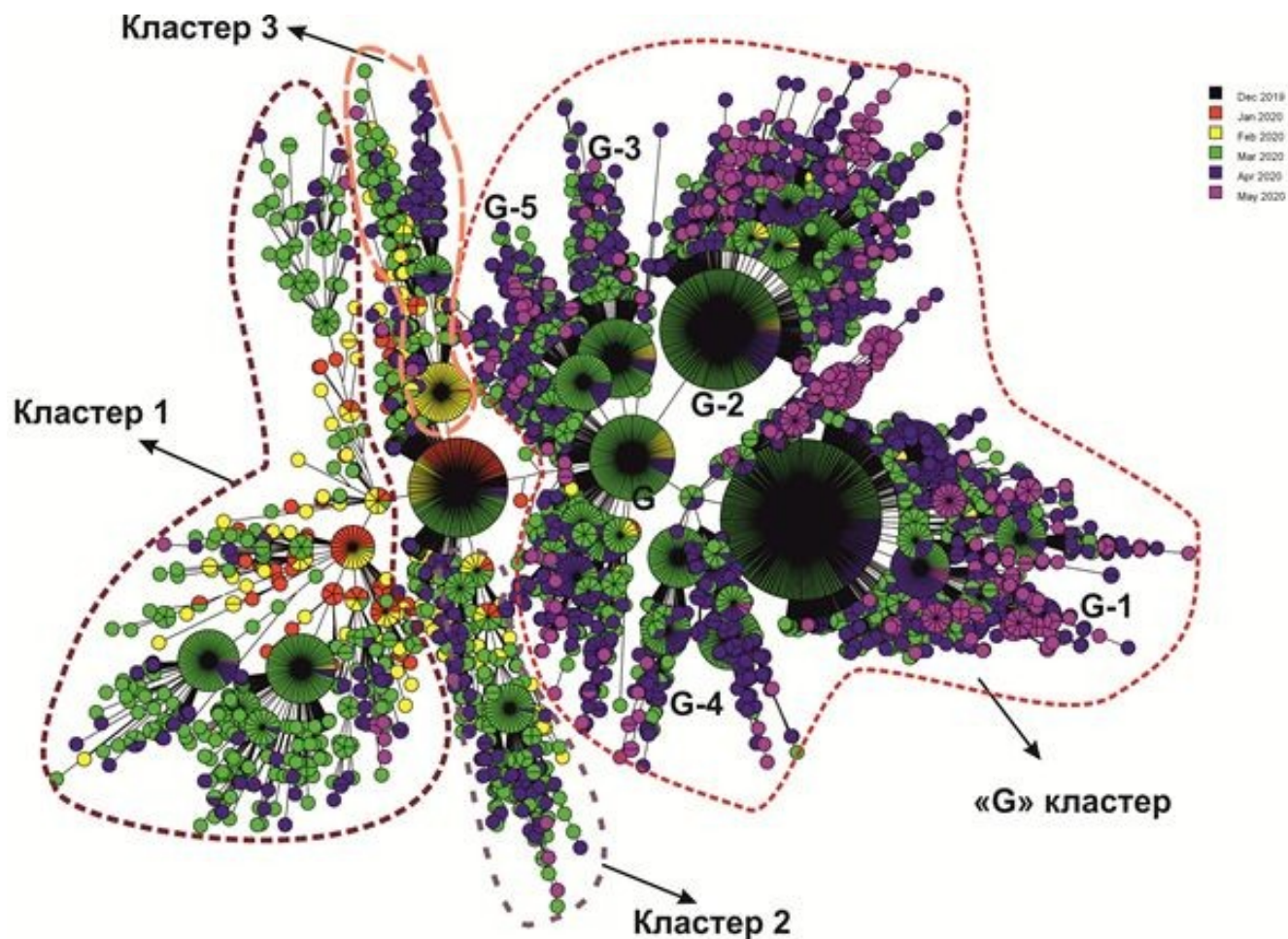


Рисунок 4. Филогенетические связи геномов вируса SARS-CoV-2, полученных из клинического материала от пациентов из разных регионов мира в период с декабря 2019 г. по май 2020 г. В мае (выделен фиолетовы) сохраняется рост числа образцов клинического материала, положительного на COVID-19, в которых геном вируса SARS-CoV-2 имеет мутацию D614G в S гене. Число новых секвенированных геномов вируса SARS-CoV-2, не имеющих мутацию D614G, к июню 2020 г. около 4 %

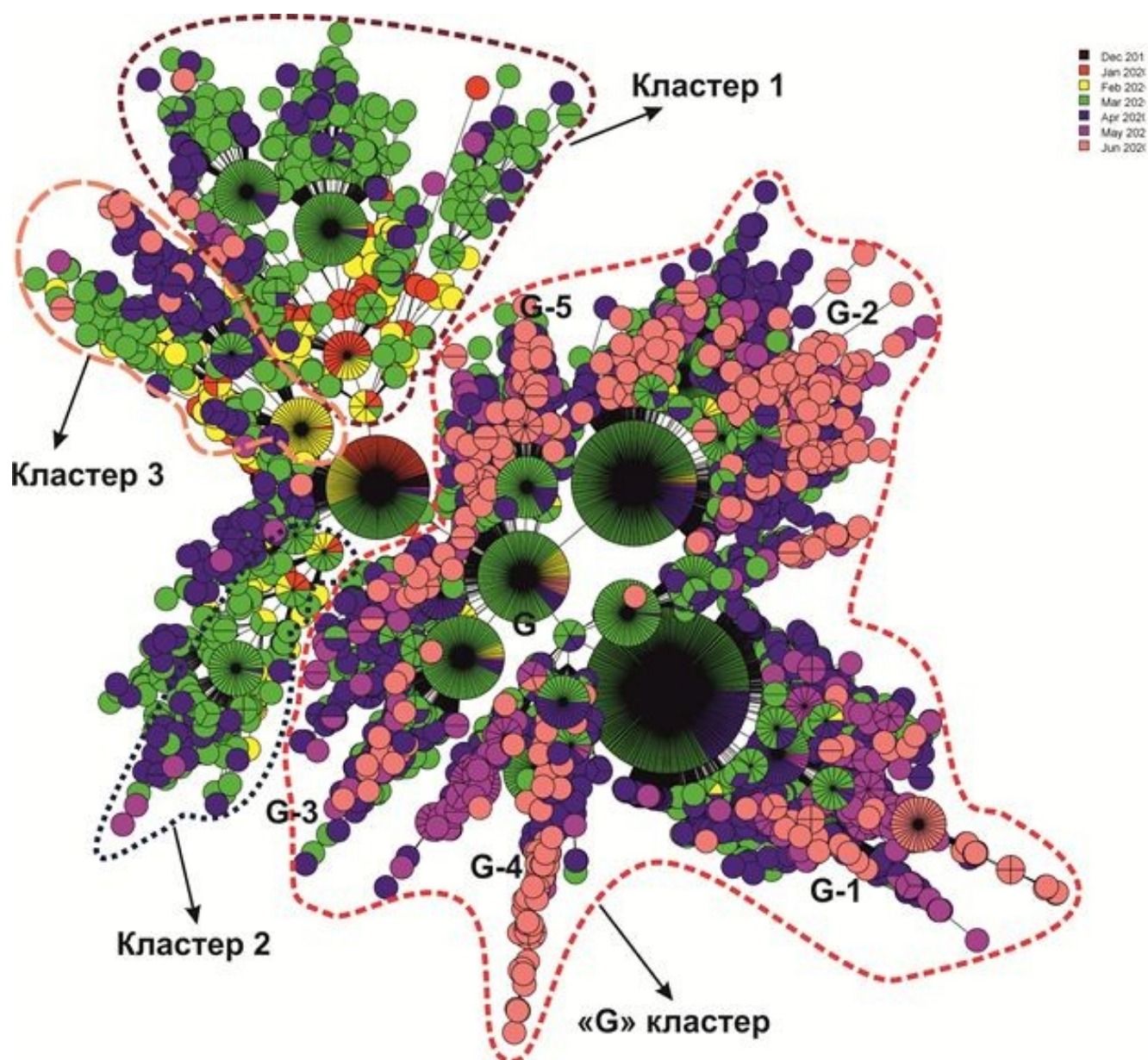


Рисунок 5. Филогенетические связи геномов вируса SARS-CoV-2, полученных из клинического материала от пациентов из разных регионов мира в период с декабря 2019 г. по июнь 2020 г. В июне (выделен розовым) продолжается рост числа образцов клинического материала, положительного на COVID-19, в которых геном вируса SARS-CoV-2 имеет мутацию D614G в S гене. Число новых секвенированных геномов вируса SARS-CoV-2, не имеющих мутацию D614G, к июлю 2020 г. около 3 %

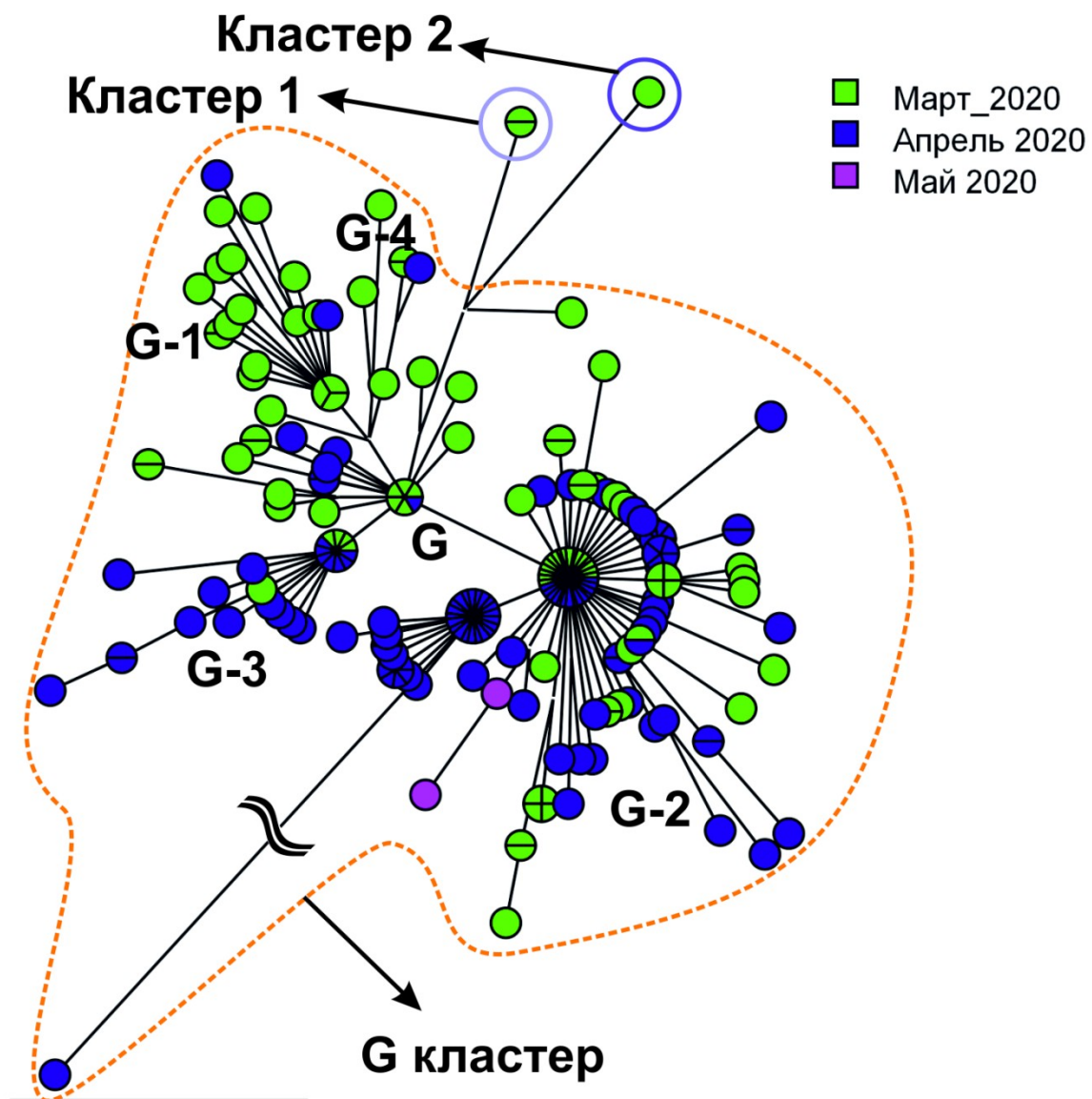


Рисунок 6. Филогенетические связи геномов вируса SARS-CoV-2, полученных из клинического материала от пациентов на территории России в марте–мае 2020 г. Наиболее представлен подкластер G-2 (146 геномов), затем подкластер G-3 (24 генома), а также исходные варианты G кластера (24 генома). Кластеры 1 и 2 представлены двумя и одним геномом вируса SARS-CoV-2, соответственно.

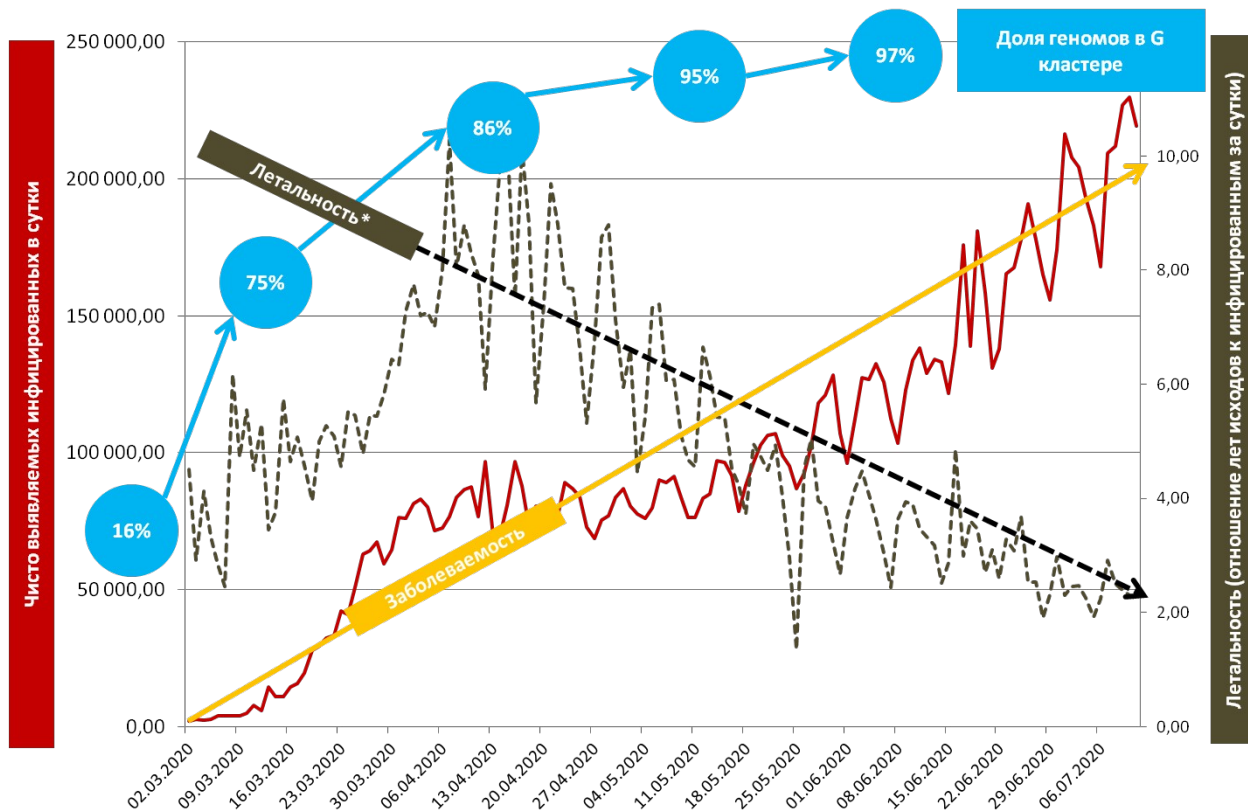


Рисунок 7. Сводный график заболеваемости (красная линия), соотношение суточного числа летальных исходов и числа инфицированных (черная пунктирная линия, обозначенная как летальность*) и распространенности геномов в G кластере (светло-синяя линия).

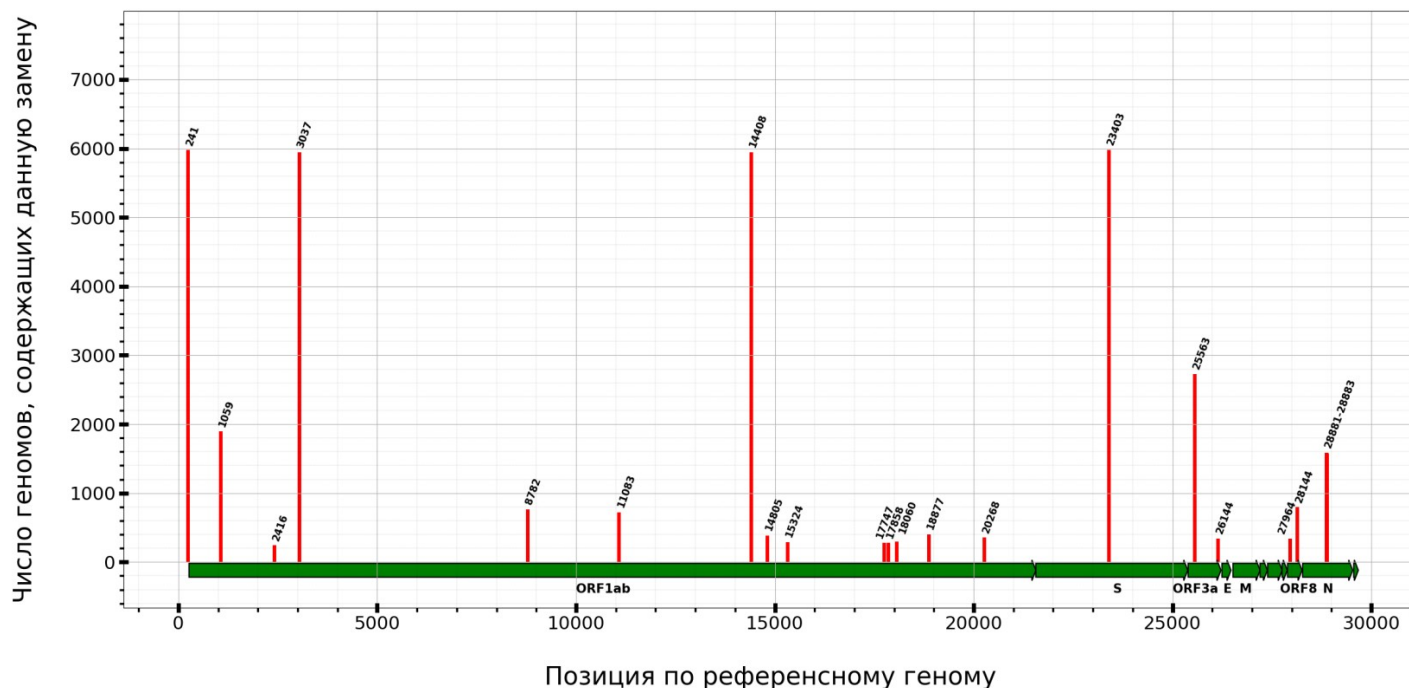


Рисунок 8. Распределение единичных мутаций, стабильно сохраняющихся во многих геномах вирусов SARS-CoV-2 в последующих инфекционных циклах. Расчет сделан по выборке, включающей 7947 геномов вируса SARS-CoV-2 из базы данных на сайте GISAID (<https://www.gisaid.org>) с конца декабря 2019 г. по июнь 2020 г. Определены 22 единичные мутации, которые сохраняются более чем у 3 % от числа исследуемых геномов вируса SARS-CoV-2.

Таблица

Рост числа проб клинического материала, содержащих вирус SARS-CoV-2, в геноме которого обнаружена G мутация (D614G) в S гене

Месяц	Общее число геномов	Число геномов имеющих G мутацию	Процент от общего по количеству вариантов с G мутацией
январь	266	5	1,87
февраль	405	65	16,05
март	4380	3312	75,62
апрель	1916	1662	86,74
май	470	451	95,95
июнь	493	477	96,75

Литература

Cui J., Li F., Shi Z.L. Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology*. 2019; 17(3):181–192. DOI: 10.1038/s41579-018-0118-9.

Wu F., Zhao S., Yu B., Chen Y.M., Wang W., Song Z.G., Hu Y., Tao Z.W., Tian J.H., Pei Y.Y., Yuan M.L., Zhang Y.L., Dai F.H., Liu Y., Wang Q.M., Zheng J.J., Xu L., Holmes E.C., Zhang Y.Z. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020a; 579(7798):265–269. DOI: 10.1038/s41586-020-2008-3.

Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020; 579(7798):270–273. DOI: 10.1038/s41586-020-2012-7.

Lu R.J., Zhao X., Li J., Niu P.H., Yang B., Wu H.L., Wang W.L., Song H., Huang B.Y., Zhu N., Bi Y.H., Ma X.J., Zhan F.X., Wang L., Hu T., Zhou H., Hu Z.H., Zhou W.M., Zhao L., Chen J., Meng Y., Wang J., Lin Y., Yuan J.Y., Xie Z.H., Ma J.M., Liu W.J., Wang D.Y., Xu W.B., Holmes E.C., Gao G.F., Wu G.Z., Chen W.J., Shi W.F., Tan W.J. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet*. 2020b; 395(10224):565–574. DOI: 10.1016/S0140-6736(20)30251-8.

Wrapp D., Wang N., Corbett K.S., Goldsmith J.A., Hsieh C.-L., Abiona O., Graham, B.S., McLellan J.S. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*. 2020; 367:1260.

Chen W.H., Hotez P.J., and Bottazzi M.E. Potential for developing a SARS-CoV receptor-binding domain (RBD) recombinant protein as a heterologous human vaccine against coronavirus infectious disease (COVID)-19. *Hum Vaccin Immunother*. 2020; 1–4.

Walls A.C., Park Y.-J., Tortorici M.A., Wall A., McGuire A.T., and Veelsler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*. 2020; 181:281–292.e286.

Pachetti M., Marini B., Benedetti F., Giudici F., Mauro E., Storici P., Masciovecchio C., Angeletti S., Ciccozzi M., Gallo R.C. *et al.* Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *Journal of Translational Medicine*. 2020; 18: 179.

Korber B., Fischer W.M., Gnanakaran S., Yoon H., Theiler J., Abfalterer W., at all. Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. *bioRxiv*. DOI: 10.1101/2020.04.29.069054.

Ashour H.M., Elkhatab W.F., Rahman Md.M. and Elshabrawy H.A. Insights into the Recent 2019 Novel Coronavirus (SARS-CoV-2) in Light of Past Human Coronavirus Outbreaks. *Pathogens*. 2020; 9:186. DOI: 10.3390/pathogens9030186.

Mousavizadeh L., Ghasemi S. Genotype and phenotype of COVID-19: Their roles in pathogenesis. *Journal of Microbiology, Immunology and Infection*. DOI: 10.1016/j.jmii.2020.03.022.

Benvenuto D., Giovanetti M., Salemi M., Prosperi M., De Flora C., Alcantara L.C.J., Angeletti S., Ciccozzi M. The global spread of 2019-nCoV: a molecular evolutionary analysis, *Pathogens and Global Health*. 114(2):64–67. DOI: 10.1080/20477724.2020.1725339.

Zhao Z., Li H., Wu X. et al. Moderate mutation rate in the SARS coronavirus genome and its implications. *BMC Evol Biol*. 2004; 4:21 . DOI: 10.1186/1471-2148-4-21.