

DOI: 10.21055/0370-1069-2023-1-111-119

УДК 616.98:578.834.1

Л.В. Миронова<sup>1</sup>, А.Н. Бондарюк<sup>1,2</sup>, Е.А. Сидорова<sup>1</sup>, Н.О. Бочалгин<sup>1</sup>, И.С. Федотова<sup>1</sup>,  
Ю.С. Букин<sup>2,3</sup>, А.С. Пономарева<sup>1</sup>, Е.И. Андаев<sup>1</sup>, С.В. Балахонов<sup>1</sup>

## Популяционная геномика SARS-CoV-2 в субъектах Сибирского и Дальневосточного федеральных округов

<sup>1</sup>ФКУЗ «Иркутский научно-исследовательский противочумный институт Сибири и Дальнего Востока», Иркутск, Российская Федерация; <sup>2</sup>ФГБУН «Лимнологический институт» Сибирского отделения РАН, Иркутск, Российская Федерация; <sup>3</sup>ФГБОУ ВО «Иркутский государственный университет», Иркутск, Российская Федерация

**Цель** работы – анализ генетической структуры популяций и изучение закономерностей эволюционной изменчивости возбудителя новой коронавирусной инфекции в Сибирском и Дальневосточном федеральных округах. **Материалы и методы.** Исследовано 1033 генома SARS-CoV-2 из проб от лиц с диагнозом COVID-19 из восьми субъектов Сибири и Дальнего Востока с декабря 2020 г. по ноябрь 2021 г. Секвенирование проводилось на платформе MinION Oxford Nanopore по протоколу ARTIC v.3. Степень генетической изоляции SARS-CoV-2 оценивалась по критерию  $F_{st}$ . Филогенетический анализ осуществлялся с использованием метода максимального правдоподобия и байесовского филогенетического вывода. При реконструкции популяционной динамики применена непараметрическая модель Bayesian Skyline Plot (BSP). **Результаты и обсуждение.** На начальных этапах в 100 % идентифицирован исходный вариант SARS-CoV-2 (B.1), геновариант Alpha обнаруживался в марте – июне 2021 г., Beta – в единичных образцах в марте – мае 2021 г., Delta впервые выявлен в апреле 2021 г. Максимальная степень генетической изоляции SARS-CoV-2 ( $F_{st}=0,18$ ) установлена для наиболее удаленных территорий (Алтайский край ↔ Республика Бурятия и Алтайский край ↔ Иркутская область). Между Иркутской областью, Республикой Бурятия и Красноярским краем выявлена относительно свободная циркуляция вируса. По результатам популяционно-генетических тестов показано, что определяющим механизмом формирования генетического разнообразия SARS-CoV-2 стал резкий рост эффективного размера популяции вируса. Реконструкция популяционной динамики в BEAST (модель BSP) выявила согласованность трендов генетического разнообразия вируса и количества активных случаев заболевания. В кластере Delta выявлено два субкластера, состоящих преимущественно из образцов, выделенных в Иркутской области и Красноярском крае. Установлена смена доминирующего варианта SARS-CoV-2 в динамике. Молекулярно-эпидемиологические данные свидетельствуют о существовании множественных путей проникновения различных генотипов SARS-CoV-2 в субъекты с формированием отдельных монофилетических кластеров и дальнейшим внутри- и экстерриториальным распространением входящих в кластеры вариантов.

**Ключевые слова:** SARS-CoV-2, геновариант, филогенетический анализ, генетическое разнообразие, популяционная динамика, Delta, ВЭП.

Корреспондирующий автор: Миронова Лилия Валерьевна, e-mail: mironova-lv@yandex.ru.

Для цитирования: Миронова Л.В., Бондарюк А.Н., Сидорова Е.А., Бочалгин Н.О., Федотова И.С., Букин Ю.С., Пономарева А.С., Андаев Е.И., Балахонов С.В. Популяционная геномика SARS-CoV-2 в субъектах Сибирского и Дальневосточного федеральных округов. *Проблемы особо опасных инфекций*. 2023; 1:111–119. DOI: 10.21055/0370-1069-2023-1-111-119

Поступила 05.08.2022. Принята к публ. 10.08.2022.

L.V. Mironova<sup>1</sup>, A.N. Bondaryuk<sup>1,2</sup>, E.A. Sidorova<sup>1</sup>, N.O. Bochalgina<sup>1</sup>, I.S. Fedotova<sup>1</sup>, Yu.S. Bukin<sup>2,3</sup>,  
A.S. Ponomareva<sup>1</sup>, E.I. Andaev<sup>1</sup>, S.V. Balakhonov<sup>1</sup>

## Population Genomics of SARS-CoV-2 in the Constituent Entities of Siberian and Far Eastern Federal Districts

<sup>1</sup>Irkutsk Research Anti-Plague Institute of Siberia and Far East, Irkutsk, Russian Federation;

<sup>2</sup>Limnological Institute of the Siberian Branch of the Russian Academy Sciences, Irkutsk, Russian Federation;

<sup>3</sup>Irkutsk State University, Irkutsk, Russian Federation

**Abstract.** The aim of the study was to analyze the genetic structure of populations and the patterns of evolutionary variability of the novel coronavirus infection in the Siberian and Far Eastern Federal Districts. **Materials and methods.** 1033 SARS-CoV-2 genomes from samples from individuals diagnosed with COVID-19 from eight entities of the Siberia and Far East between December 2020 and November 2021 were assessed. Sequencing was performed on the MinION Oxford Nanopore platform using the ARTIC v.3 protocol. The degree of SARS-CoV-2 genetic isolation was estimated applying the  $F_{st}$  criterion. Phylogenetic analysis was carried out using maximum likelihood method and Bayesian phylogenetic inference. A nonparametric Bayesian Skyline Plot (BSP) model was used to reconstruct population dynamics. **Results and discussion.** The original SARS-CoV-2 variant (B.1) was identified in 100 % of the cases at the initial stages. The Alpha variant was detected in March–June, 2021; Beta – in single samples in March–May, 2021. Delta was first identified in April, 2021. The maximum degree of SARS-CoV-2 genetic isolation ( $F_{st}=0.18$ ) was established for the most remote territories (Altai Territory ↔ Republic of Buryatia and Altai Territory ↔ Irkutsk Region). A relatively free circulation of the virus was detected between Irkutsk Region, Republic of Buryatia and Krasnoyarsk Territory. According to the results of population genetic tests, a sharp increase in the effective virus population size was the determining mechanism of SARS-CoV-2 genetic diversity formation. Reconstruction of population dynamics in BEAST (BSP model) has re-

vealed the consistency of trends in the genetic diversity of the virus and the number of active cases. Two subclusters have been identified in the Delta cluster, consisting predominantly of samples isolated in the Irkutsk Region and Krasnoyarsk Territory. Change in the dominant variant of SARS-CoV-2 has been traced in dynamics. Molecular-epidemiological data point to the multiple pathways of spatial expansion of different SARS-CoV-2 genotypes into the constituent entities with generation of individual monophyletic clusters and further intra- and extraterritorial spread of the decedents.

**Key words:** SARS-CoV-2, genovariant, phylogenetic analysis, genetic diversity, population dynamics, Delta, VOC.

**Conflict of interest:** The authors declare no conflict of interest.

**Corresponding author:** Liliya V. Mironova, e-mail: mironova-lv@yandex.ru.

**Citation:** Mironova L.V., Bondaryuk A.N., Sidorova E.A., Bochalgin N.O., Fedotova I.S., Bukin Yu.S., Ponomareva A.S., Andaev E.I., Balakhonov S.V. Population Genomics of SARS-CoV-2 in the Constituent Entities of Siberian and Far Eastern Federal Districts. *Problemy Osobo Opasnykh Infektsii [Problems of Particularly Dangerous Infections]*. 2023; 1:111–119. (In Russian). DOI: 10.21055/0370-1069-2023-1-111-119

Received 05.08.2022. Accepted 10.08.2022.

Mironova L.V., ORCID: <https://orcid.org/0000-0001-8481-6442>  
Bondaryuk A.N., ORCID: <https://orcid.org/0000-0003-4422-0497>  
Sidorova E.A., ORCID: <https://orcid.org/0000-0003-0279-5831>  
Bochalgin N.O., ORCID: <https://orcid.org/0000-0002-3053-6514>  
Fedotova I.S., ORCID: <https://orcid.org/0000-0001-9890-0960>

Bukin Yu.S., ORCID: <https://orcid.org/0000-0002-4534-3846>  
Ponomareva A.S., ORCID: <https://orcid.org/0000-0002-0674-6159>  
Andaev E.I., ORCID: <https://orcid.org/0000-0002-6612-479X>  
Balakhonov S.V., ORCID: <https://orcid.org/0000-0003-4201-5828>

РНК-содержащие вирусы относятся к группе быстро эволюционирующих патогенов. В организме хозяина адаптивной изменчивости подвержены, как правило, участки генома вируса, обеспечивающие ему уклонение от иммунного ответа. Наряду с этим эффективное распространение возбудителя в популяции восприимчивого хозяина, повышение его трансмиссивности обусловлено изменчивостью поверхностных белков, связывающихся со специфическими рецепторами макроорганизма [1].

На основании байесовского филогенетического анализа SARS-CoV-2 установлено, что средняя скорость накопления мутаций в его геноме составляет  $1 \cdot 10^{-3}$  замен в год [2–4]. Эволюционные преобразования возбудителя приводят к возникновению и закреплению мутаций в ключевых участках генома коронавируса и формированию новых генетических вариантов с определенными селективными преимуществами.

Так, известно о формировании в ходе пандемии вариантов SARS-CoV-2, обозначенных ВОЗ как вызывающие беспокойство варианты, по принятой в РФ номенклатуре – варианты с высоким эпидемическим потенциалом (ВЭП): Alpha, Beta, Gamma, Delta, – характеризующиеся способностью к уклонению от иммунного ответа и более высокой трансмиссивностью [5, 6]. В конце ноября 2021 г. сообщено о появлении нового варианта SARS-CoV-2 – Omicron, который к настоящему времени практически полностью вытеснил существующие ранее варианты коронавируса и демонстрирует высокую вариабельность сублиний.

Высокий потенциал изменчивости возбудителя новой коронавирусной инфекции, смена доминирующих вариантов в динамике пандемии определяют значимость исследований по изучению структурных особенностей геномов SARS-CoV-2 для выяснения закономерностей его эволюционных преобразований и их взаимосвязи с эпидемическим процессом.

В Российской Федерации организована система геномного мониторинга возбудителя новой коронавирусной инфекции, в рамках которой определены научные учреждения Роспотребнадзора, обеспечивающие на постоянной основе молекулярно-

генетические исследования SARS-CoV-2 в прикрепленных субъектах РФ. ФКУЗ Иркутский научно-исследовательский противочумный институт Роспотребнадзора выполняет секвенирование образцов биологического материала от лиц с диагнозом COVID-19 из восьми субъектов, входящих в состав Сибирского и Дальневосточного федеральных округов (СФО и ДФО): Алтайский, Красноярский, Забайкальский края, республики Алтай, Хакасия, Тыва, Бурятия, Иркутская область. Анализ полученных в процессе мониторинга последовательностей геномов SARS-CoV-2 позволяет отследить эволюционные изменения вируса и сопоставить характер этих изменений с эпидемиологическими данными по заболеваемости и распространению SARS-CoV-2 в курируемых субъектах.

**Цель работы** – анализ генетической структуры популяций и изучение закономерностей эволюционной изменчивости возбудителя новой коронавирусной инфекции в Сибирском и Дальневосточном федеральных округах.

## Материалы и методы

**Исследуемый материал.** Объектом исследования служили пробы биологического материала, отобранные от лиц с диагнозом COVID-19 из восьми курируемых институтом территорий (Алтайский, Красноярский, Забайкальский края, республики Алтай, Хакасия, Тыва, Бурятия, Иркутская область) в период с декабря 2020 г. по ноябрь 2021 г. включительно.

**Секвенирование геномов** проводилось на платформе MinION Oxford Nanopore в соответствии с протоколом SARS-CoV-2 v3 (LoCost) консорциума Artic Network. Для подготовки библиотек ампликонов использовалась праймерная панель ARTIC v.3. Результаты анализировались посредством картирования полученных ридов на референсный геном в соответствии со стандартным протоколом (<https://artic.network/nCoV-2019/ncov2019-bioinformatics-sop.html>).

**Выборка нуклеотидных последовательностей.** Всего в работе использовано 1033 нуклеотид-

ные последовательности геномов SARS-CoV-2, полученные из проб биологического материала от лиц с диагнозом COVID-19. Покрытие референсного генома для включенных в исследование нуклеотидных последовательностей составило более 95 %.

**Генотипирование** проводилось при помощи онлайн-версии программы Pangolin (<https://pangolin.cog-uk.io/>).

**Выравнивание нуклеотидных последовательностей** проводилось с использованием онлайн-сервиса MAFFT [7]. Визуализация выравнивания осуществлялась в программе AliView v.1.27 [8].

**Для анализа гомоплазий в белке S** выбрано 11 аминокислотных (а.к.) замен и две делеции, характерные для четырех вариантов с высоким эпидемическим потенциалом: вариант Alpha – Δ69-70, N501Y, A570D, P681H, D1118H; Beta – K417N, E484K, N501Y; Gamma – L18F, D138Y, E484K, N501Y; Delta – E484Q, L452R, P681R, T478K.

**Популяционно-генетический анализ.** На основании территории изоляции исходная выборка поделена на семь предполагаемых популяций (в анализ не включены образцы из Республики Тыва в связи с недостаточным количеством проб в выборке).

Плотность вероятности генетического разнообразия популяций ( $\theta$  – среднее количество попарно несовпадающих нуклеотидов) рассчитана в среде вычислений R на основе выборки из 1000 реплик ультрабыстрого бутстрепа. Достоверность различий в значениях  $\theta$  между каждой парой популяций оценивалась как площадь перекрытия двух распределений плотности вероятности: если площадь перекрытия превышала значение 0,05, то такие пары распределений  $\theta$  не рассматривались нами как достоверно различающиеся. Оценка площади перекрытия проводилась в R с использованием пакета sfsmisc.

Популяционно-генетические тесты – Tajima's D и Ramos-Onsins-Rozas (R2-тест) – на детекцию резкого изменения эффективного размера популяции и нейтральность эволюционного процесса проводились в R с использованием пакета pegas.

Тест на нейтральный характер эволюции Fay and Wu's H, а также расчет показателей  $F_{st}$  между каждой парой популяций выполнены в программе DnaSP v.6 [9].

**Филогенетический анализ** проводился с использованием двух основных подходов: метода максимального правдоподобия (ML), реализованного в программе IQTREE v.1.6.12 [10], и байесовского филогенетического вывода (программа BEAST v.2.6.3 [11]). Для филогенетической реконструкции в IQTREE использовали всю имеющуюся выборку (1033 нуклеотидные последовательности). Выбор эволюционной модели основывался на значении байесовского информационного критерия (BIC), рассчитанного с применением алгоритма ModelFinder [12], реализованного в IQTREE. Статистическая поддержка клад оценивалась при помощи ультрабыстрого бутстрепа (ufbs) с количеством реплик, равным 1000 [13].

Для проведения байесовского филогенетического анализа в BEAST составлено две подвыборки: pop150 (для анализа популяционной динамики) и clock269 (для оценки времени образования устойчивых генетических линий SARS-CoV-2) со 150 и 269 нуклеотидными последовательностями соответственно. Выборка pop150 сформирована случайным образом, выборка clock269 смещена в сторону стабильных генетических линий, циркулирующих преимущественно на территориях Иркутской области и Красноярского края.

Эволюционную модель (матрицу замен) для анализа в BEAST выбирали при помощи ModelFinder (IQTREE). Выбор молекулярных часов основывался на значениях коэффициента вариации (KB) эволюционной скорости, полученных с использованием расслабленных часов с априорным логнормальным распределением скорости нуклеотидных замен между филогенетическими ветвями (UCLD). Если в 95 % интервал максимума плотности апостериорной вероятности (95 % HPD) не входят значения KB от 0 до 0,1, то в этом случае в качестве наиболее подходящей модели рассматривается модель расслабленных молекулярных часов.

Расчет силы временного сигнала (temporal signal) проводился на основе коэффициента детерминации ( $R^2$ ), отражающего взаимосвязь между датой изоляции образцов SARS-CoV-2 и их генетическим расстоянием до корня. Значения  $R^2$  рассчитывали в программе TempEst v.1.5.3. При  $R^2 > 0,1$  сила временного сигнала расценивалась как достаточная для применения молекулярных часов.

Для реконструкции популяционной динамики (изменение относительного генетического разнообразия во времени  $N_e \cdot \tau$ , где  $N_e$  – эффективный размер популяции,  $\tau$  – время генерации) нами составлена дополнительная подвыборка (pop150), в состав которой случайным образом отобраны 150 изолятов из 8 регионов (период изоляции: декабрь 2020 г. – ноябрь 2021 г. включительно). В качестве популяционной модели выбрана непараметрическая модель Bayesian Skyline Plot с 10 композитными коалесцентными интервалами (BSP10).

Количество итераций в запуске алгоритма Монте-Карло по схеме марковской цепи (MCMC) составило  $200 \cdot 10^6$ . Статистическая неопределенность для времени дивергенции и эволюционной скорости оценивалась при помощи 95 % HPD. По итогам запуска значение эффективного размера выборки (ESS) для всех параметров модели превышало 200 независимых образцов. Для построения консенсусного дерева из общего запуска отобрано 36000 деревьев (с учетом отжига 10 % от длины цепи MCMC).

## Результаты и обсуждение

Всего с декабря 2020 г. по ноябрь 2021 г. секвенировано 1819 геномов SARS-CoV-2 из проб биологического материала от лиц с диагнозом COVID-19

из восьми субъектов СФО и ДФО. На начальных этапах исследования в 100 % случаев обнаруживался исходный вариант коронавируса (базальные ветви клады В по номенклатуре PANGO); геновариант Alpha идентифицирован впервые в марте 2021 г. и обнаруживался в отдельных пробах до июня 2021 г., геновариант Beta – в единичных образцах в марте – мае 2021 г. Геновариант Delta, впервые выявленный на территории курируемых субъектов в конце апреля, к июню идентифицирован в 90 % проб и затем полностью вытеснил остальные варианты коронавируса (с августа по ноябрь 2021 г. геновариант Delta идентифицирован в 100 % исследованных проб, <https://doi.org/10.6084/m9.figshare.20453115>). Первые случаи заболевания COVID-19, обусловленные геновариантом Omicron, выявлены в конце декабря 2021 г. в Иркутской области.

Изучение структуры популяций и закономерностей генетической изменчивости SARS-CoV-2 в Сибирском и Дальневосточном регионах проведено на основании анализа секвенированных с декабря 2020 г. по ноябрь 2021 г. 1033 геномов.

**Генотипирование** показало, что 81,2 % исследованных изолятов SARS-CoV-2 отнесены к геноварианту Delta, в структуре которого доминирует линия AY.122 – ее удельный вес составил 77,6 %. Доля геноварианта Alpha в выборке составила 2,4 %, Beta – 0,2 %, исходного варианта SARS-CoV-2 – 16,2 % с распределением на 16 генетических линий, доминирующая среди которых – B.1.1 (61,7 %) (<https://doi.org/10.6084/m9.figshare.20453655>).

**Динамика аминокислотных замен в белке S.**

Для анализа динамики аминокислотных замен в белке S выбрано несколько мутаций, характерных для четырех вариантов ВЭП. Обращает внимание обнаружение в декабре 2020 г. замен, характерных для вариантов ВЭП до их регистрации на исследуемых

территориях (рис. 1). В частности, замены L18F и D138Y, характерные для варианта Gamma (P.1), составили 70 % от общей выборки; замена E484K [варианты Beta (B.1.351) и Gamma (P.1)] – 6 % от выборки; P681H [вариант Alpha (B.1.1.7)] – менее 5 % от общей выборки; P681R [вариант Delta] – 20 % от общей выборки. При этом замена P681R, характерная для варианта Delta, зафиксирована еще до возникновения этого варианта в природе, что является наглядным примером конвергентной эволюции и селективного преимущества обозначенных мутаций.

Наибольшее разнообразие замен в белке S наблюдается в апреле и мае 2021 г. (12 и 10 замен соответственно), когда зафиксировано появление всех трех вариантов на курируемых территориях. Далее, начиная с июня, вследствие нарастания удельного веса варианта Delta разнообразие ВЭП-замен падает.

В сентябре и октябре 2021 г. нам удалось секвенировать несколько образцов, относящихся к варианту Delta (AY.122), содержащих замену E484Q, чаще встречающуюся в линии B.1.617.3: три последовательности в образцах из Красноярского края в сентябре и один образец из г. Иркутска в октябре. В ноябре зафиксирована замена N501Y, которая, по данным outbreak.info, в линии B.1.617.2 встречается лишь в 0,2 % случаев.

**Генетическое разнообразие и популяционная структура SARS-CoV-2.** Проведен анализ степени генетической изоляции SARS-CoV-2, циркулирующего на курируемых территориях, с использованием критерия  $F_{st}$ . Согласно методике, чем выше значения  $F_{st}$ , тем выше изоляция между двумя территориями. Результаты анализа изображены в виде графа (рис. 2, а).

Ожидаемо, максимальное значение  $F_{st}$  (0,18) оказалось характерно для наиболее удаленных территорий (Алтайский край ↔ Республика Бурятия и

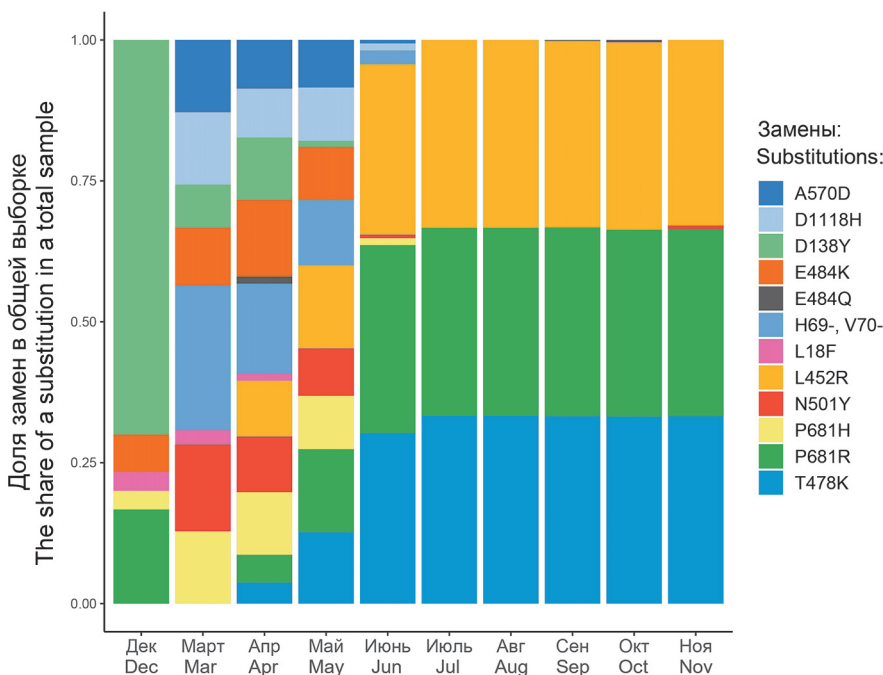


Рис. 1. Динамика аминокислотных замен в белке S, характерных для вариантов SARS-CoV-2 с высоким эпидемическим потенциалом (ВЭП)

Fig. 1. The dynamics of amino acid substitutions specific to the SARS-CoV-2 variants of concern (VOCs) in the S protein

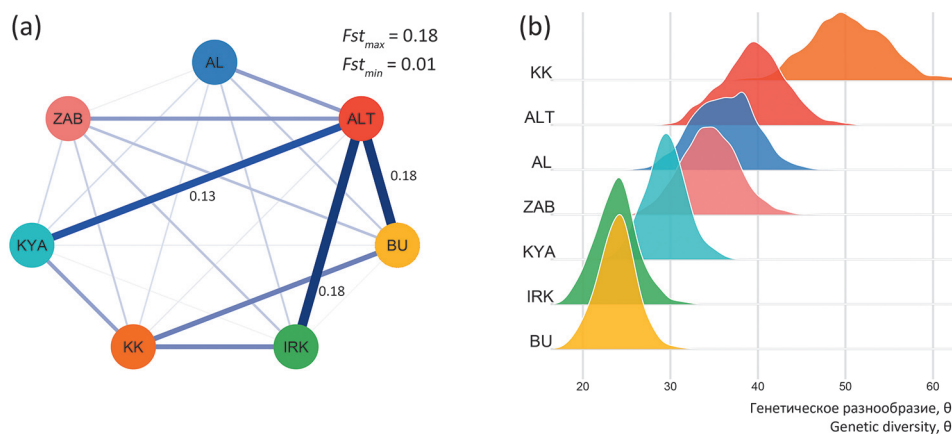


Рис. 2. Результаты анализа степени генетической изоляции SARS-CoV-2 с использованием критерия  $F_{st}$  и расчет генетического разнообразия  $\theta$ :

*a* – попарные значения  $F_{st}$  (толщина ребер графа) для 7 курируемых территорий (вершины графа); *b* – плотность распределения генетического разнообразия ( $\theta$ ) для 7 исследуемых популяций (сокращения регионов даны согласно ISO 3166-2: *AL* – Республика Алтай; *ALT* – Алтайский край; *BU* – Республика Бурятия; *IRK* – Иркутская область; *KK* – Республика Хакасия; *KYA* – Красноярский край; *ZAB* – Забайкальский край)

Fig. 2. The results of the analysis of the degree of SARS-CoV-2 genetic isolation using the  $F_{st}$  criterion and the calculation of genetic diversity  $\theta$ :

*a* – paired  $F_{st}$  values (graph edge thickness) for 7 territories under supervision (graph vertices); *b* – the distribution density ( $\theta$ ) of genetic diversity of 7 analyzed populations (the abbreviations are provided according to the ISO 3166-2: *AL* – Altai Republic; *ALT* – Altai Territory; *BU* – Republic of Buryatia; *IRK* – Irkutsk Region; *KK* – Republic of Khakassia; *KYA* – Krasnoyarsk Territory; *ZAB* – Zabaykalsky Territory)

Алтайский край ↔ Иркутская область). Наиболее изолированным регионом оказался Алтайский край. Между Иркутской областью, Республикой Бурятия и Красноярским краем выявлена относительно свободная циркуляция вируса со значениями  $F_{st}$ , приближающимися к нулю.

Стоит отметить, что в целом значения  $F_{st}$  невелики. Это может быть связано с тем, что даже при высоком уровне изоляции на разных территориях вирус не успел накопить достаточное количество нуклеотидных замен для более точного расчета значений критерия.

Расчет генетического разнообразия ( $\theta$ ) показал, что в Республике Хакасия популяция вируса обладает наивысшим средним  $\theta$  (50,2), что достоверно

отличается ( $p$ -value < 0,05) от показателей  $\theta$  для популяций Красноярского края (28,8), Иркутской области (22,6), Забайкальского края (33,5) и Республики Бурятия (23,1) (рис. 2, b).

Наименьшее генетическое разнообразие оказалось характерно для популяций Иркутской области (22,6) и Республики Бурятия (23,1), что достоверно отличается от  $\theta$  популяций Алтайского края, Республики Алтай и Республики Хакасия, при этом достоверных различий между значениями  $\theta$  для самих популяций Иркутской области и Республики Бурятия нет ( $p$ -value = 0,95, <https://doi.org/10.6084/m9.figshare.20453829>).

**Проверка нейтрального характера эволюции SARS-CoV-2.** Результаты тестов (таблица) показыва-

**Результаты популяционно-генетических тестов**

**The results of population genetic tests**

Регион (популяция) Region (population)	D-тест Таджики Tajima's D-statistics	P-value для D-теста Таджики P-value for Tajima's D-statistics	R2-тест R2-statistics	P-value для R2-теста P-value for R2-statistics	H-тест Фэя и Ву Fay and Wu's H-statistics	P-value для H-теста Фэя и Ву P-value for Fay and Wu's H-statistics
Республика Алтай Altai Republic	-2,79	0,0052	0,039	< 0,001	8,63	0,59
Алтайский край Altai Territory	-2,54	0,011	0,034	< 0,001	19,60	0,94
Республика Бурятия Republic of Buryatia	-3,12	0,0018	0,011	< 0,001	-4,90	0,23
Иркутская область Irkutsk Region	-3,01	0,0026	0,0036	< 0,001	-8,57	0,17
Республика Хакасия Republic of Khakassia	-3,09	0,002	0,0253	< 0,001	19,70	0,84
Красноярский край Krasnoyarsk Territory	-3,06	0,0022	0,0078	< 0,001	0,47	0,34
Забайкальский край Trans-Baikal Territory	-3,01	0,0026	0,0061	< 0,001	2,53	0,36

ли, что на популяционном уровне генетическое разнообразие вируса SARS-CoV-2 во всех курируемых регионах не формировалось по нейтральному сценарию (нейтральный дрейф генов при небольшом изменении эффективного размера популяции).

Тест Tajima's D (таблица) позволил отклонить гипотезу нейтрального характера эволюции ( $p\text{-value} < 0,05$ ) в пользу гипотез быстрого увеличения эффективного размера популяции либо движущего отбора с фиксацией наиболее приспособленных генотипов (отрицательное во всех случаях значение статистики теста). Результаты R2-теста позволяют сделать вывод, что во всех регионах происходило резкое увеличение эффективной численности популяции вируса ( $P < 0,05$ ). Тест Fay and Wu's H не зафиксировал движущий отбор ни в одном из регионов ( $P > 0,05$ ). Таким образом, по совокупности результатов тестирования можно сказать, что определяющим механизмом формирования генетического разнообразия SARS-CoV-2 стал резкий рост в недавнем прошлом эффективного размера популяции вируса. Для вирусов единицей эффективного размера служит инфицированный индивидуум, являющийся источником инфекции для других представителей популяции, поэтому увеличение эффективного размера популяции вируса заключается в росте текущего

числа инфицированных. Результаты популяционного тестирования, показывающие резкий рост числа заболевших за исследуемый промежуток времени, согласуются с реальными эпидемиологическими данными, фиксирующими рост заболеваемости COVID-19 и результатами анализа с помощью метода BSP (см. ниже). Хотя тест Fay and Wu's H и не зафиксировал движущий отбор в каком-либо регионе, полностью отказаться от роли этого процесса в формировании генетического разнообразия вируса нельзя. На начальном этапе появившийся в исследуемых регионах штамм B.1.617.2 SARS-CoV-2 (вариант Delta) и его потомки вытесняли остальные филогенетические линии SARS-CoV-2, так как они имели больший адаптивный потенциал (большую заразность и, соответственно, большее репродуктивное число), реализуя сценарий движущего отбора. Затем движущий отбор сменился резким ростом эффективной численности вируса (сценарий популяционной экспансии), который и определился популяционными тестами в современных генетических данных.

**Филогенетический анализ методом ML** ожидаемо показал соответствие с результатами генотипирования в Pangolin: два варианта ВЭП (Alpha и Delta) образовали на дереве монофилетические кластеры с высокой поддержкой ( $ufb=100$ , рис. 3).

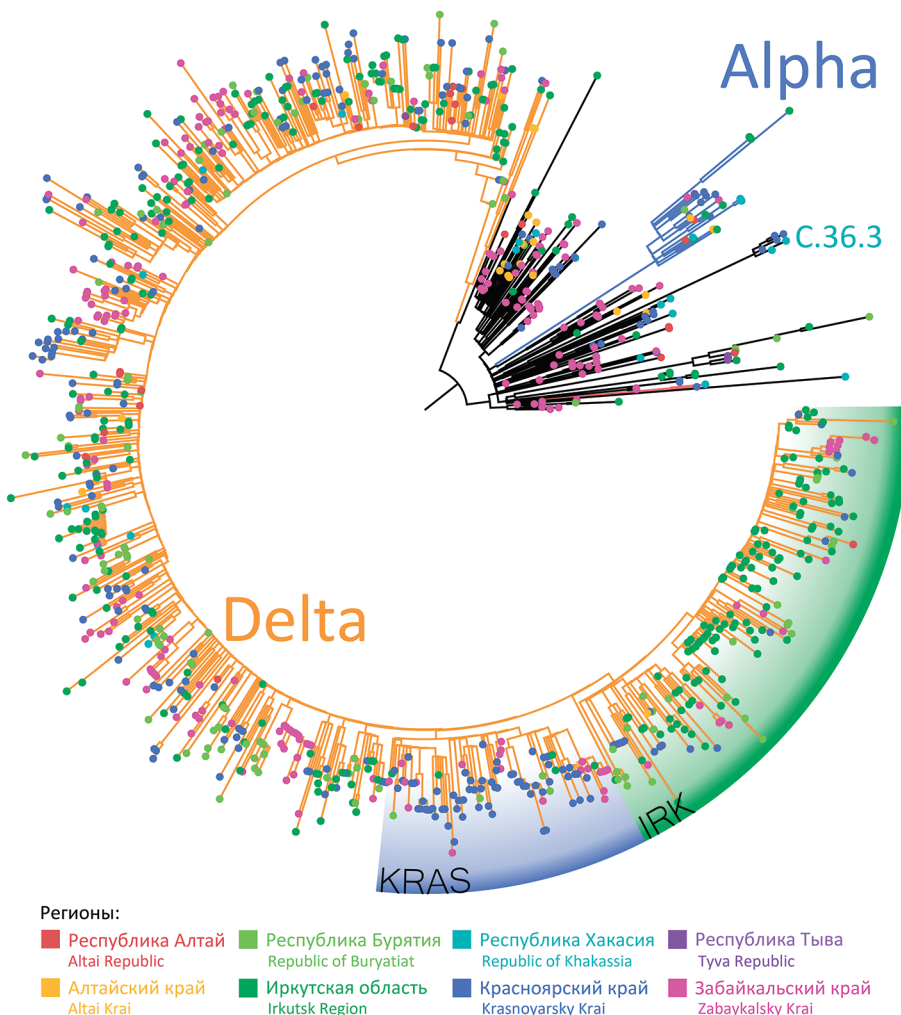


Рис. 3. Филогенетическое дерево, реконструированное методом максимального правдоподобия в программе IQTREE (модель замен с наивысшим значением BIC: GTR+F+I+G4)

Fig. 3. The maximum likelihood phylogenetic tree reconstructed in IQTREE software (the best-fit model according to BIC: GTR+F+I+G4)

Также выявлены отдельные кластеры, сформированные более ранними линиями: В.1.1.189 (Иркутская область,  $n=6$ ), В.1.1.372 (Забайкальский край,  $n=5$ ,  $ufb=98$ ), В.1.1.397 (Забайкальский край,  $n=6$ ,  $ufb=97$ ), С.36 и С36.3 (Красноярский край и Республика Хакасия,  $n=6$ ,  $ufb=100$ ). Данная структура дерева указывает не только на множественные независимые пути проникновения различных вариантов SARS-CoV-2 на территории указанных регионов, но и на их дальнейшее распространение внутри территорий (внутренние узлы субкластеров).

В кластере варианта Delta выявлено наличие двух монофилетических субкластеров, состоящих преимущественно из образцов, выделенных в Иркутской области (кластер Irk,  $n=168$ ,  $ufb=85$ ) и Красноярском крае (кластер Kras,  $n=71$ ,  $ufb=83$ ), что говорит о проникновении, закреплении и дальнейшем распространении на территории субъектов двух независимых генетических линий SARS-CoV-2. Для оценки датировок сформирована дополнительная подвыборка (clock269), включающая в себя помимо изолятов кластеров Irk и Kras более ранние генетические линии SARS-CoV-2.

Предварительно обе анализируемые выборки прошли проверку на наличие временного сигнала.

**Анализ временного сигнала** в программе TempEst показал, что в обеих выборках присутствует выраженная взаимосвязь между датой изоляции образцов и их генетическим расстоянием до корня ( $R^2=0,491$ ,  $R^2=0,723$ ; <https://doi.org/10.6084/m9.figshare.20453931>). Следовательно, применение молекулярных часов в обоих случаях можно считать корректным.

Отклонение генетических дистанций от рассчитанных трендов можно объяснить наличием вариации скорости эволюции между генетическими линиями SARS-CoV-2 ( $KB=0,28$ ; 95 % HPD, 0,15–0,41), поэтому в дальнейшем при анализе обеих выборок в BEAST мы использовали ослабленные молекулярные часы (UCLD).

**Эволюционная скорость.** Значение эволюционной скорости оценивалось на основе анализа несмещенной выборки  $pop150$ . Согласно нашим расчетам, эволюционная скорость SARS-CoV-2 составила  $7,1 \cdot 10^{-4}$  нуклеотидных замен на сайт в год (95 % HPD,  $6,1-8,0 \cdot 10^{-4}$  нуклеотидных замен на сайт в год). В самых ранних работах, посвященных оценке скорости фиксации нуклеотидных замен SARS-CoV-2, получены более высокие значения – порядка  $10^{-3}$  [3, 14]. Однако в ряде других работ исследователи публикуют значения эволюционной скорости, сопоставимые с нашими результатами ( $6,6 \cdot 10^{-4}$ , 95 % HPD,  $6,1-8,0 \cdot 10^{-4}$  [15];  $4,1 \cdot 10^{-4}$  [16];  $3,95 \cdot 10^{-4}$  [17]). Вполне возможно, что наблюдаемые различия в оценках эволюционной скорости существуют в действительности и обусловлены разницей в охвате вакцинацией, скорости распространения вируса, а также спецификой популяционной структуры. Тем не менее стоит иметь в виду и чувстви-

тельность самого метода оценки скорости к выбору априорных распределений параметров.

**Молекулярно-эпидемиологический анализ.** Установлено, что кластер Irk включает первый образец, генотипированный как Delta (10 мая 2021 г.), поэтому время возникновения данного кластера (tMRCA) можно рассматривать как момент появления на территории Иркутской области первого пациента, инфицированного геновариантом Delta (нулевой пациент). tMRCA кластера Kras не может быть интерпретирована схожим образом, так как данная клада не содержит первый положительный образец. Филогенетический анализ в BEAST с применением молекулярных часов позволил рассчитать время возникновения кластеров Irk и Kras с относительно высокой точностью (рис. 4).

**Кластер Irk.** Проведенные расчеты показали, что в Иркутской области нулевой пациент с 95 % вероятностью существовал в период с 8 апреля по 7 мая 2021 г. (по средним оценкам tMRCA – 24 апреля 2021 г., апостериорная поддержка = 0,97). Сравнительный анализ аминокислотных замен выявил две маркерные мутации в рамках считывания ORF1ab (I1215V) и ORF7a (P45S), встречающиеся в изолятах кластера Irk в 100 и 99 % случаев соответственно, что позволяет отследить дальнейшее перемещение вируса на территории других субъектов. Данные перемещения патогена представлены в виде субкластеров с характерной сменой цвета ветвей: субкластер Irk.I – Иркутская область → Забайкальский край → Республика Бурятия → Забайкальский край (реэмиграция); субкластер Irk.II – Иркутская область → Красноярский край → Иркутская область (реэмиграция); субкластер Irk.III – Иркутская область и три независимых заноса в Красноярский край, Республику Бурятия и Забайкальский край с последующим обширным распространением в последнем случае (внутренние узлы); в субкластерах Irk.IV и Irk.V вирус также получил дальнейшее распространение на территориях Красноярского края и Республики Бурятия.

В структуре кластера Irk можно выделить субкластер Irk570, все представители которого несут в белке S аминокислотную замену A570S, однако апостериорная поддержка субкластера относительно низка и равняется 0,2.

**Кластер Kras.** Схожие оценки времени появления ближайшего общего предка получены для кластера Kras (95 % HPD, 17 апреля – 19 мая 2021 г.). Субкластеры Kras.I и Kras.II наглядно иллюстрируют перенос вируса из Красноярского края в Забайкальский край, где SARS-CoV-2 получает свое дальнейшее распространение. В случае Kras.I зарегистрирован пример реэмиграции из Забайкальского края в Красноярский край.

**Популяционная динамика.** В контексте инфекционных болезней изменение относительного генетического разнообразия вируса ( $N_e \cdot \tau$ , где  $N_e$  – эффективный размер популяции,  $\tau$  – время генера-

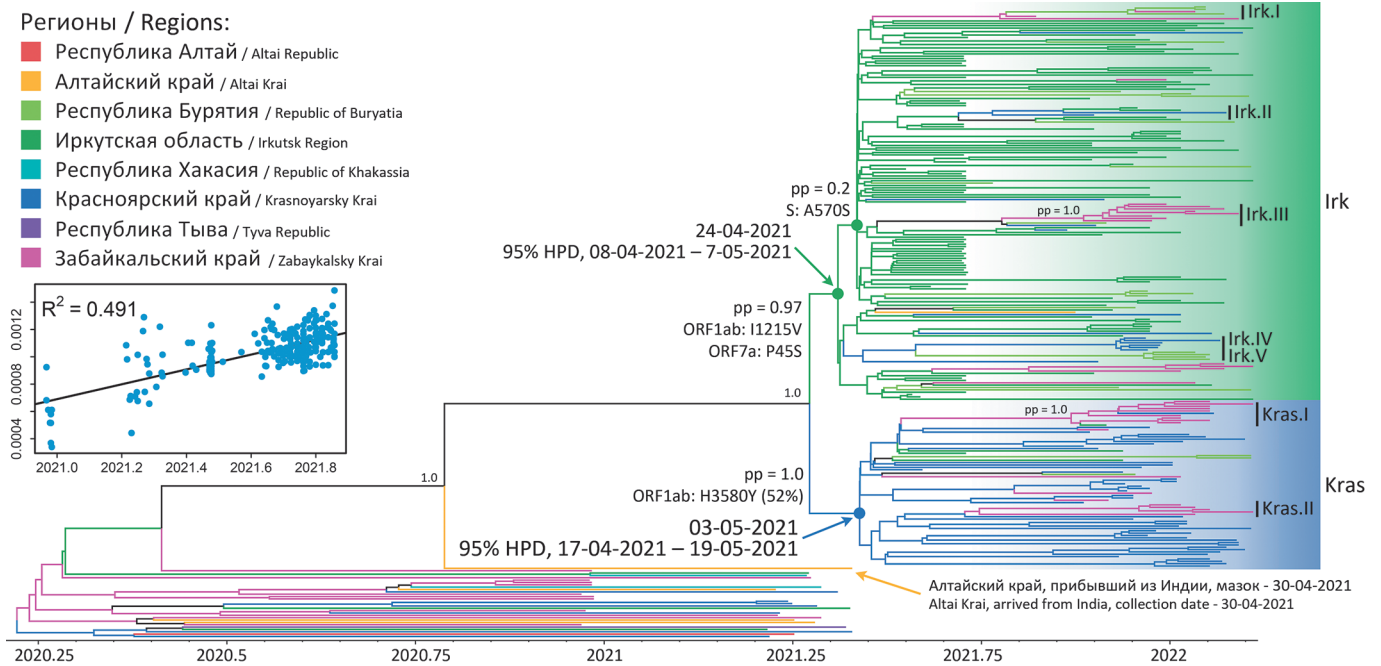


Рис. 4. Филогенетическое дерево, реконструированное в BEAST (модель GTR+F+I+G4+UCLD+BSP10). Регрессионный тренд во внутреннем окне отображает величину временного сигнала

Fig. 4. Maximum clade credibility tree inferred with BEAST (the model combination is GTR+F+I+G4+UCLD+BSP10). In the insertion, the regression analysis demonstrates the strength of temporal signal

ции) можно соотнести с изменением количества активных случаев заболевания [18, 19]. Мы провели реконструкцию динамики генетического разнообразия SARS-CoV-2, используя подвыборку из 150 нуклеотидных последовательностей и сравнили полученные результаты с динамикой активных случаев в 8 курируемых регионах. Анализ показал, что в период с апреля по май 2021 г. произошел спад  $N_e \cdot \tau$  (рис. 5). Далее последовал непродолжительный период стабилизации  $N_e \cdot \tau$  (май – июнь 2021 г.), в течение которого на всех исследуемых территориях регистрировались первые случаи заноса геноварианта Delta

(в Республике Бурятия – 22 апреля, Забайкальском крае – 28 апреля, Алтайском крае – 30 апреля, Красноярском крае – 7 мая, Иркутской области и Республике Алтай – 10 и 11 мая, Республике Тыва – 18 мая, Республике Хакасия – 27 мая). Появление геноварианта Delta спровоцировало дальнейший экспоненциальный рост заболеваемости и увеличение показателя  $N_e \cdot \tau$ .

Полученные данные позволяют предположить, что для реализации эпидемического потенциала на анализируемых территориях варианту Delta потребовался приблизительно один месяц. В случае

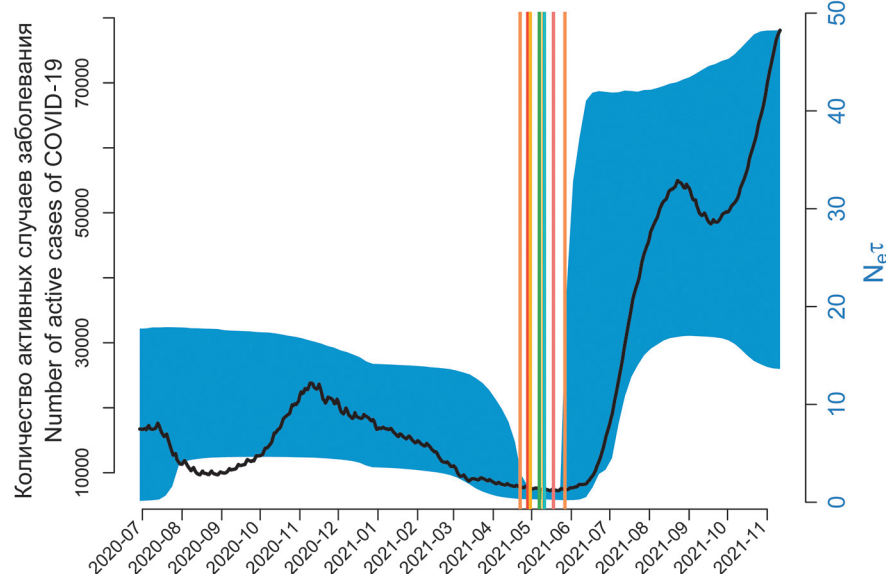


Рис. 5. Сравнение динамики относительного генетического разнообразия ( $N_e \cdot \tau$ ) SARS-CoV-2 (непараметрическая модель Bayesian Skyline Plot с 10 композитными коалесцентными интервалами, BSP10) с количеством активных случаев в 8 курируемых регионах. Закрашенная синяя область отображает 95 % HPD, вертикальные цветные линии – первый зарегистрированный случай заражения вариантом Delta в каждом из регионов. Важно отметить, что кривая активных случаев практически не выходит за пределы 95 % HPD  $N_e \cdot \tau$

Fig. 5. Comparison of SARS-CoV-2 relative genetic diversity dynamics ( $N_e \cdot \tau$ , where  $N_e$  – effective number of infections,  $\tau$  – generation time) (Nonparametric Bayesian Skyline Plot model with 10 composite coalescent intervals, BSP10) with the change in the number of COVID-19 active cases in 8 regions under supervision. The shaded with blue color area is 95 % HPD of  $N_e \cdot \tau$ , vertical colored lines – the first reported case of Delta variant infection in each region. Notably, the active case curve is almost always within the  $N_e \cdot \tau$  95 % HPD



с распространением варианта Omicron на территории России прослеживается аналогичная картина распространения возбудителя.

На основании проведенного анализа показано, что в динамике резкий рост генетического разнообразия SARS-CoV-2 совпадает по времени с интенсивным распространением геноварианта Delta в субъектах. Эти результаты подтверждают охарактеризованный в настоящем исследовании механизм формирования генетического разнообразия возбудителя, заключающийся в действии на начальном этапе движущего отбора за счет селективных преимуществ нового геноварианта с последующим резким ростом эффективного размера популяции патогена. Данный механизм, по всей вероятности, реализован и при распространении в настоящее время геноварианта Omicron.

Таким образом, продемонстрирована смена доминирующего варианта возбудителя новой коронавирусной инфекции в субъектах СФО и ДФО в динамике, что отражает глобальные тренды развития пандемии. При этом установлена относительно свободная циркуляция вируса между отдельными субъектами СФО и ДФО, в частности Иркутской областью, Красноярским краем и Республикой Бурятия. Молекулярно-эпидемиологические данные свидетельствуют о существовании множественных путей проникновения различных генотипов SARS-CoV-2 на территории субъектов с формированием отдельных монофилетических кластеров патогена и дальнейшим внутри- и экстерриториальным распространением входящих в кластеры вариантов.

**Конфликт интересов.** Авторы подтверждают отсутствие конфликта финансовых/нефинансовых интересов, связанных с написанием статьи.

**References / Список литературы**

1. Frost S.D.W., Magalis B.R., Kosakovsky Pond S.L. Neutral theory and rapidly evolving viral pathogens. *Mol. Biol. Evol.* 2018; 35(6):1348–54. DOI: 10.1093/molbev/msy088.
2. Ciotti M., Ciccozzi M., Pieri M., Bernardini S. The COVID-19 pandemic: viral variants and vaccine efficacy. *Crit. Rev. Clin. Lab. Sci.* 2022; 59(1):66–75. DOI: 10.1080/10408363.2021.1979462.
3. Duchene S., Featherstone L., Haritopoulou-Sinanidou M., Rambaut A., Lemey P., Baele G. Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evol.* 2020; 6(2):veaa061. DOI: 10.1093/ve/veaa061.
4. Safari I., Elahi E. Evolution of the SARS-CoV-2 genome and emergence of variants of concern. *Arch. Virol.* 2022; 167(2):293–305. DOI: 10.1007/s00705-021-05295-5.
5. Martin D.P., Weaver S., Tegally H., San J.E., Shank S.D., Wilkinson E., Lucaci A.G., Giandhari J., Naidoo S., Pillay Y., Singh L., Lessells R.J., NGS-SA, COVID-19 Genomics UK (COG-UK), Gupta R.K., Wertheim J.O., Nekturenko A., Murrell B., Harkins G.W., Lemey P., MacLean O.A., Robertson D.L., de Oliveira T., Kosakovsky Pond S.L. The emergence and ongoing convergent evolution of the SARS-CoV-2 N501Y lineages. *Cell.* 2021; 184(20):5189–5200.e5187. DOI: 10.1016/j.cell.2021.09.003.
6. Mascola J.R., Graham B.S., Fauci A.S. SARS-CoV-2 viral variants-tackling a moving target. *JAMA.* 2021; 325(13):1261–2. DOI: 10.1001/jama.2021.2088.
7. Katoh K., Rozewicki J., Yamada K.D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* 2019; 20(4):1160–6. DOI: 10.1093/bib/bbx108.

8. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics.* 2014; 30(22):3276–8. DOI: 10.1093/bioinformatics/btu531.
9. Rozas J., Ferrer-Mata A., Sánchez-DelBarrio J.C., Guirao-Rico S., Librado P., Ramos-Onsins S.E., Sánchez-Gracia A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 2017; 34(12):3299–302. DOI: 10.1093/molbev/msx248.
10. Nguyen L.T., Schmidt H.A., von Haeseler A., Minh B.Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 2015; 32(1):268–74. DOI: 10.1093/molbev/msu300.
11. Bouckaert R., Vaughan T.G., Barido-Sottani J., Duchene S., Fourment M., Gavryushkina A., Heled J., Jones G., Kuhnert D., De Maio N., Matschiner M., Mendes F.K., Muller N.F., Ogilvie H.A., du Plessis L., Poppinga A., Rambaut A., Rasmussen D., Siveroni I., Suchard M.A., Wu C.H., Xie D., Zhang C., Stadler T., Drummond A.J. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 2019; 15(4):e1006650. DOI: 10.1371/journal.pcbi.1006650.
12. Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., von Haeseler A., Jermini L.S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods.* 2017; 14(6):587–9. DOI: 10.1038/nmeth.4285.
13. Minh B.Q., Nguyen M.A.T., von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 2013; 30(5):1188–95. DOI: 10.1093/molbev/mst024.
14. Singh D., Yi S.V. On the origin and evolution of SARS-CoV-2. *Exp. Mol. Med.* 2021; 53(4):537–47. DOI: 10.1038/s12276-021-00604-z.
15. Wang S., Xu X., Wei C., Li S., Zhao J., Zheng Y., Liu X., Zeng X., Yuan W., Peng S. Molecular evolutionary characteristics of SARS-CoV-2 emerging in the United States. *J. Med. Virol.* 2022; 94(1):310–7. DOI: 10.1002/jmv.27331.
16. Motayo B.O., Oluwasemowo O.O., Oluola B.A., Akinduti P.A., Arege O.T., Obafemi Y.D., Faneye A.O., Isibor P.O., Aworunse O.S., Oranusi S.U. Evolution and genetic diversity of SARS-CoV-2 in Africa using whole genome sequences. *Int. J. Infect. Dis.* 2021; 103:282–7. DOI: 10.1016/j.ijid.2020.11.190.
17. Shen S., Zhang Z., He F. The phylogenetic relationship within SARS-CoV-2s: An expanding basal clade. *Mol. Phylogenet. Evol.* 2021; 157:107017. DOI: 10.1016/j.ympev.2020.107017.
18. Pybus O.G., Rambaut A. Evolutionary analysis of the dynamics of viral infectious disease. *Nat. Rev. Genet.* 2009; 10(8):540–50. DOI: 10.1038/nrg2583.
19. Hall M.D., Woolhouse M.E., Rambaut A. The effects of sampling strategy on the quality of reconstruction of viral population dynamics using Bayesian skyline family coalescent methods: A simulation study. *Virus Evol.* 2016; 2(1):vew003. DOI: 10.1093/ve/vew003.

**Authors:**

Mironova L.V., Sidorova E.A., Bochalgin N.O., Fedotova I.S., Ponomareva A.S., Andaev E.I., Balakhonov S.V. Irkutsk Research Anti-Plague Institute of Siberia and Far East. 78, Trilissera St., Irkutsk, 664047, Russian Federation. E-mail: adm@chumin.irkutsk.ru.  
 Bondaryuk A.N. Irkutsk Research Anti-Plague Institute of Siberia and Far East; 78, Trilissera St., Irkutsk, 664047, Russian Federation; e-mail: adm@chumin.irkutsk.ru. Limnological Institute of the Siberian Branch of the Russian Academy Sciences; 3, Ulaanbaatar St., Irkutsk, 664033, Russian Federation.  
 Bukin Yu.S. Limnological Institute of the Siberian Branch of the Russian Academy Sciences; 3, Ulaanbaatar St., Irkutsk, 664033, Russian Federation; e-mail: bukinyura@mail.ru. Irkutsk State University; 1, Karl Marx St., Irkutsk, 664003, Russian Federation.

**Об авторах:**

Миронова Л.В., Сидорова Е.А., Бочалгин Н.О., Федотова И.С., Пономарева А.С., Андаев Е.И., Балахонов С.В. Иркутский научно-исследовательский противочумный институт Сибири и Дальнего Востока. Российская Федерация, 664047, Иркутск, ул. Трилиссера, 78. E-mail: adm@chumin.irkutsk.ru.  
 Бондарюк А.Н. Иркутский научно-исследовательский противочумный институт Сибири и Дальнего Востока; Российская Федерация, 664047, Иркутск, ул. Трилиссера, 78; e-mail: adm@chumin.irkutsk.ru. Лимнологический институт Сибирского отделения РАН; Российская Федерация, 664033, Иркутск, ул. Улан-Баторская, 3.  
 Букин Ю.С. Лимнологический институт Сибирского отделения РАН; Российская Федерация, 664033, Иркутск, ул. Улан-Баторская, 3; e-mail: bukinyura@mail.ru. Иркутский государственный университет; Российская Федерация, 664003, Иркутск, ул. Карла Маркса, 1.