



<http://dx.doi.org/10.35596/1729-7648-2023-21-2-104-113>

Оригинальная статья
Original paper

УДК 004.056

ВЫЧИСЛИТЕЛЬНЫЙ АНАЛИЗ СТРУКТУРНОГО СОСТАВА ГЕНОМОВ КОРОНАВИРУСОВ

М. В. СПРИНДЖУК^{1,2,3}, В. И. БЕРНИК¹, А. С. ВЛАДЫКО⁵, ЧЖОЧЖУАН ЛУ⁴,
Л. П. ТИТОВ⁵

¹Институт математики Национальной академии наук Беларуси (г. Минск, Республика Беларусь)

²Белорусский государственный университет информатики и радиоэлектроники
(г. Минск, Республика Беларусь)

³Объединенный институт проблем информатики Национальной академии наук Беларуси
(г. Минск, Республика Беларусь)

⁴Китайский центр по контролю и профилактике заболеваний (г. Пекин, Китайская Народная Республика)

⁵Республиканский научно-практический центр эпидемиологии и микробиологии
(г. Минск, Республика Беларусь)

Поступила в редакцию 30.01.2023

© Белорусский государственный университет информатики и радиоэлектроники, 2023
Belarusian State University of Informatics and Radioelectronics, 2023

Аннотация. Экологические катастрофы, уменьшение площади лесных насаждений, одомашнивание диких животных, употребление в пищу зараженных животных, загрязнение воды и продуктов питания и их компонентов, эксперименты с вирусами, дефициты и дефекты иммунной системы у современного человека и других млекопитающих стали толчком к развитию новых опасных и особо опасных вирусов. Пандемия коронавируса, отнесенного к категории опасных вирусов, привела к повышению востребованности знаний и навыков вычислительной биологии, эпидемиологии и вирусологии в современном обществе. Существующие секвенаторы способны производить большие объемы биоинформационных данных, которые отображаются в виде геномных текстов. Сравнительный вычислительный анализ такой информации необходим для выяснения вопросов филогенеза, мутационного профилирования, молекулярной эволюции, определения вставок других геномов, аннотирования регионов геномов, поиска мишеней для разработки вакцин и фармакотерапии. В связи с этим авторами статьи проведен вычислительный эксперимент сравнительного анализа геномных текстов белорусских образцов коронавируса с рядом отобранных полных геномов опасных и особо опасных вирусов и коронавирусов различного происхождения. Анализ данных выполнен компьютерной программой YASS, геномные тексты загружали из GISAID, также был использован самостоятельно разработанный конвейер обработки геномных данных на основе биоинформационной платформы Galaxy. В результате анализа данных обнаружено значительное сходство нового коронавируса с рекомбинантным коронавирусом, частичное сходство с вирусами синтетического коронавируса, краснухи, Эбола 1976, ближневосточного респираторного синдрома, ВИЧ-2 (вируса иммунодефицита человека), вируса иммунодефицита обезьян и лихорадки Марбурга.

Ключевые слова: коронавирус, пандемия, биоинформатика, системы медицинского назначения, геномика, особо опасные инфекции, математическое моделирование, иммуноинформатика, противовирусная терапия.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Благодарность. Работа выполнена при поддержке Белорусского республиканского фонда фундаментальных исследований в рамках проектов: Ф21МН-001 «Математическое моделирование передачи и распространения COVID-19 инфекции на основе систем дифференциальных уравнений и алгоритмов обработки данных с применением технологии машинного обучения», ГР 20213518 от 27.09.2021; M21COVID-026 «Ретроспективный анализ клинического и иммунологического статуса групп COVID-19 пациентов с сопутствующим туберкулезом и ВИЧ-инфекцией по данным РНПЦ пульмонологии и фтизиатрии г. Минска», ГР 20210456 от 31.03.2021; M21COVID-001 «Разработка и скрининг мукозной вакцины против COVID-19 на основе векторной платформы кишечного аденовируса», ГР 20210889 от 26.04.2021.

Для цитирования. Вычислительный анализ структурного состава геномов коронавирусов / М. В. Спринд-жук [и др.] // Доклады БГУИР. 2023. Т. 21, № 2. С. 104–113. <http://dx.doi.org/10.35596/1729-7648-2023-21-2-104-113>.

COMPUTATIONAL ANALYSIS OF THE STRUCTURAL COMPOSITION OF CORONAVIRUS GENOMES

MATVEY V. SPRINDZUK^{1,2,3}, VASSILI I. BERNIK¹, ALEXANDER S. VLADYKO⁵, ZHUOZHANG LU⁴, LEONID P. TITOV⁵

¹*Institute of Mathematics of the National Academy of Sciences of Belarus (Minsk, Republic of Belarus)*

²*Belarusian State University of Informatics and Radioelectronics (Minsk, Republic of Belarus)*

³*The United Institute of Informatics Problems of the National Academy of Sciences of Belarus (Minsk, Republic of Belarus)*

⁴*Chinese Center for Disease Control and Prevention (Beijing, People's Republic of China)*

⁵*The Republican Research and Practical Center for Epidemiology and Microbiology (Minsk, Republic of Belarus)*

Submitted 30.01.2023

Abstract. Ecological disasters, wars in regions with microbiological weapons depots, deforestation, domestication of wild animals, consumption of infected animals, contamination of water and food products and their components, experiments with viruses, deficiencies and other defects of the immune system in modern humans and other mammals became the impetus for the evolution of new dangerous and extremely dangerous viruses. Due to the emergence of new dangerous viruses, the importance and demand for knowledge and skills of computational biology, epidemiology and virology in modern society have increased. Modern sequencers are capable of producing large amounts of bioinformatic data that is represented in the form of genomic texts. Comparative computational analysis of this information is necessary to clarify the issues of phylogenesis, mutational profiling, molecular evolution, identification of insertions of other genomes, annotation of genome regions, search for targets for vaccine development and pharmacotherapy. In this context, authors conducted a computational experiment of comparative analysis of the genomic texts of Belarusian coronavirus samples against a number of selected complete genomes of dangerous and extremely dangerous viruses and coronaviruses of various origins. Data analysis was performed using the YASS, genomic texts were downloaded from the GISAID, the custom genomic data processing pipeline based on the Galaxy bioinformatics platform was also applied. The article presents the results of an analysis of the available scientific literature and the computational experiment comparing the genomic texts of Belarusian coronavirus samples with a number of selected complete genomes of dangerous and especially dangerous viruses and coronaviruses of various origin. A significant similarity of the new coronavirus with the recombinant coronavirus, as well as partial similarity with synthetic coronavirus, Rubella, Ebola 1976, HIV-2 (human immunodeficiency virus), Middle East respiratory syndrome, simian immunodeficiency and Marburg fever viruses have been found.

Keywords: coronavirus, pandemic, bioinformatics, medical systems, genomics, especially dangerous infections, mathematical modeling, immunoinformatics, antiviral therapy.

Conflict of interests. The authors declare no conflict of interests.

Gratitude. The work was supported by the Belarusian Republican Foundation for Basic Research within the framework of the projects: F21MN-001 “Mathematical modeling of the transmission and spread of COVID-19 infection based on systems of differential equations and data processing algorithms using machine learning technology”, GR 20213518 dated 27.09.2021; M21COVID-026 “Retrospective analysis of the clinical and immunological status of groups of COVID-19 patients with concomitant tuberculosis and HIV infection according to the Republican Scientific and Practical Center for Pulmonology and Phthisiology in Minsk”, GR 20210456 dated 31.03.2021; M21COVID-001 “Development and screening of a mucosal vaccine against COVID-19 based on the intestinal adenovirus vector platform”, GR 20210889 dated 26.04.2021.

For citation. Sprindzuk M. V., Bernik V. I., Vladyko A. S., Zhuozhuang Lu, Titov L. P. (2023) Computational Analysis of the Structural Composition of Coronavirus Genomes. *Doklady BGUIR*. 21 (2), 104–113. <http://dx.doi.org/10.35596/1729-7648-2023-21-2-104-113> (in Russian).

Введение

По причине значительной летальности и высокого уровня осложнений новая коронавирусная инфекция продолжает оставаться актуальным объектом для изучения с помощью современных методов медицинской кибернетики и прикладной математики на стыке различных наук, в том числе биофизики, биоинформатики и геномики, медицинского приборостроения и научного

программирования, системного анализа, вирусологии и молекулярной эпидемиологии. Вопросы структурного сходства геномных текстов новой коронавирусной инфекции между образцами-штаммами, полученными с одной географической территории, а также с референсными (эталонными) близкородственными геномами сообщаются и обсуждаются в серии недавних научных публикаций (табл. 1, 2) [1–3].

Таблица 1. Генетическая идентичность нового штамма коронавируса SARS-CoV-2 с различными CoV-штаммами [2]

Table 1. Genetic identity of a new SARS-CoV-2 coronavirus strain with different CoV strains [2]

Штамм вируса / Virus strain	Род вируса / Virus Genera	Идентичность, % / Identity, %
HCoV-229E	α	65,04
HCoV-NL63		65,11
HCoV-HKU1	β	67,59
HCoV-OC43		68,93
MERS-CoV		69,58
SARS-CoV		82,45
bat-SL-CoVZC		88,00
bat-SL-CoVZXC		88,00
SARS-HCoV Tor2		82,00
SARS-HCoV BJ01		82,00
SARSr-CoV; RaTG13		96,20

Таблица 2. Матрица процентной идентичности основных белков и доменов нового коронавируса SARS-CoV-2 в сравнении с другими бета-CoV [2]

Table 2. Percentage identity matrix for major proteins and domains of the new SARS-CoV-2 coronavirus compared to other beta-CoV [2]

Белок, домен / Protein, domain	Идентичность штаммов, % / Identity of strains, %			
	SARS-CoV	MERS-CoV	HCoV-HKU1	HCoV-OC43
Протеин S (Spike, шип)	97,71	32,79	30,50	31,26
Протеин E (Envelope, конверт)	96,00	36,00	28,00	20,00
Протеин M (Membrane, мембрана)	89,59	39,27	35,29	38,74
Протеин N (Nucleocapsid, нуклеокапсид)	85,41	48,47	34,28	35,20
Рецептор (ACE-2) binding domain (рецептор-связывающий домен ангиотензинпревращающего фермента (АПФ-2))	74,41	18,75	24,44	22,83
N-terminal domain (N-терминальный домен)	52,55	21,67	21,49	20,26

В ряде статей также излагаются идеи о возможных вставках в SARS-CoV-2 особо опасных вирусов, фрагментов генома человека, ретровирусов иммунодефицита человека и обезьян [4]. Безусловно, интересны гипотезы о структурном подобии (подразумевается на уровне нуклеиновых кислот и аминокислот) SARS-CoV-2 и синтетических и рекомбинантных вирусов. Более того, существуют гипотезы о вкладе в геном нового патогенного коронавируса и бактериальных геномов, как основных и наиболее значимых в эволюции современных опасных для человека микроорганизмов. Таким образом, целью научного исследования авторов статьи было проверить вышеуказанные гипотезы в серии вычислительных экспериментов и по данным современной научной литературы.

Методика проведения вычислительного анализа

Вычислительный эксперимент основывался на использовании программного средства YASS (Yet Another Sequence Similarity Search Software, еще одна программа для поиска сходства последовательностей). Это программное средство представляет собой инструмент для локального выравнивания/картирования (alignment/mapping) ДНК и РНК, основанный на эффективном

и чувствительном алгоритме фильтрации. Он находит наиболее вероятные консервативные мотивы при сравнении между гомологичными последовательностями, используя гибкие критерии совпадения для идентификации групп заданных секвенций [5]. Веб-интерфейс YASS на момент исследования был доступен для загрузки входных последовательностей в формате FASTA, запросов к программе и визуализации полученных результатов в нескольких формах (точечный график, табличный вывод и др.). В табличном выводе выравнивания сортируются в соответствии с их значением E-метрики и связаны с их текстовым представлением. Выходные данные YASS также можно загрузить в текстовом формате для дальнейшего анализа. Автономная версия доступна для загрузки с веб-страницы <https://github.com/laurentnoe/yass> [5].

Веб-сервис Genome detective [6, 7] применялся для идентификации мутационного профиля и гомологичности набора белорусских геномных последовательностей SARS-CoV-2 относительно эталонного китайского штамма, а также для выявления новых на момент исследования в Беларуси штаммов омикрон. Кроме того, применялся специально разработанный авторами конвейер анализа геномных данных, ранее опубликованный в [8].

Результаты исследований и их обсуждение

В ходе вычислительного эксперимента выполнен анализ множества геномных текстов, полученных от пациентов на территории Беларуси. Были отобраны геномы из общедоступной базы данных GISAID (Global initiative on sharing all influenza data – Глобальная инициатива по обмену всеми данными о гриппе) (<https://www.epicov.org/>). Использовали программное обеспечение YASS [5], Genome Detective [6, 7], программные модули облачного сервиса Galaxy (<https://usegalaxy.eu/>).

Гомологичность белорусских образцов с отдельно отобранными геномами вирусов представлена в табл. 3, а также приведены сгенерированные экстремумы баллов идентичности и энтропии, физико-математический базис которых представлен в документации и публикациях программного обеспечения YASS [5]. Сравнение белорусских образцов коронавируса SARS-CoV-2 с различными геномами с отображением на плоскости координат так называемыми точечными графиками (dot plots) приведено на рис. 1–8, где горизонтальные и вертикальные оси представляют собой прямоугольный массив нуклеотидов сравниваемых образцов. Результаты сходства геномов изображены в виде точек, где зеленые сегменты представляют выравнивания (фактически совпадения) прямых прочтений, а красные соответствуют выравниванию между комплементарными прочтениями одной последовательности и прямыми прочтениями другой [5].

Таблица 3. Гомологичность белорусских образцов с отдельно отобранными геномами вирусов из других семейств

Table 3. Comparison of the homology of Belarusian samples with separately selected viruses of different origin and genera

Геном вируса / Virus genome	Балл идентичности /сходства / Identity / Affinity Score	Энтропия /Entropy
GomelSARS-CoV-2 <i>MW674675.1</i>	100919,00	5,78
SARS-CoV <i>AY278488.2</i>	74804,00	5,79
Recombinant CoV <i>FJ211859.1</i>	23018,00	5,66
Synthetic CoV <i>OK560913.1</i>	1584,00	5,54
Marburg <i>OL702894.1</i>	124,00	3,64
Rubella <i>KT962862.1</i>	114,00	4,01
Ebola1976 <i>AF086833.2</i>	108,00	3,92
MERS-CoV <i>JX869059.2</i>	96,98	5,64
HIV-2 <i>KU179861.1</i>	90,00	3,46
SimianID <i>MW815971.1</i>	86,00	3,51

Примечание – Данные получены в результате экспериментов (ID = Immunodeficiency).

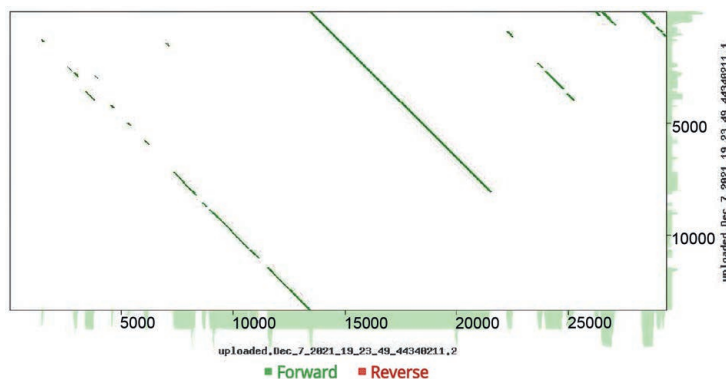


Рис. 1. Сравнение белорусских образцов коронавируса SARS-CoV-2 и генома рекомбинантного коронавируса
Fig. 1. Comparison of Belarusian samples of SARS-CoV-2 coronavirus and the genome of recombinant coronavirus

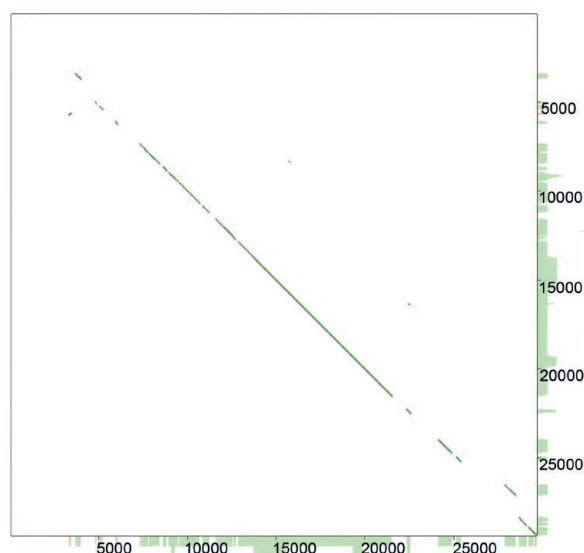


Рис. 2. Сравнение белорусских образцов коронавируса SARS-CoV-2 и генома MERS
Fig. 2. Comparison of Belarusian samples of the SARS-CoV-2 coronavirus and the MERS genome

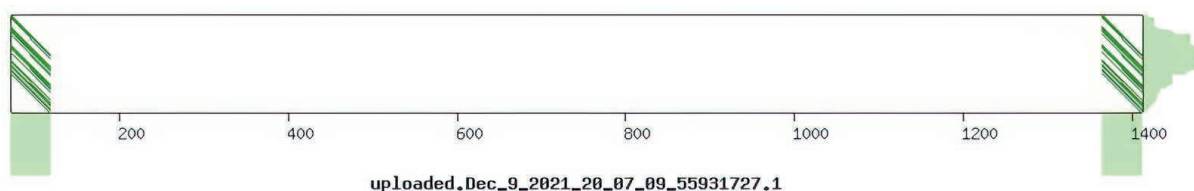


Рис. 3. Сравнение белорусских образцов коронавируса SARS-CoV-2 и генома Simian immunodeficiency virus (вируса иммунодефицита обезьян)
Fig. 3. Comparison of Belarusian samples of the SARS-CoV-2 coronavirus and the genome Simian immunodeficiency virus (monkey immunodeficiency virus)

Точечные графики сравнивают две последовательности путем организации одной последовательности по оси X , а другой – по оси Y . Когда фрагменты обеих последовательностей совпадают в одном и том же месте на графике, в соответствующем месте рисуется точка. Последовательности могут быть записаны назад или вперед, однако последовательности на обеих осях должны быть записаны в одном направлении. Направление последовательностей на осях определяет направление линии на точечной диаграмме. Близость последовательностей по сходству определяет, насколько близка диагональная линия к графику, показывающему кривую, демонстрирующую прямую связь. На эту взаимосвязь влияют определенные особенности последовательности, такие как сдвиг кадров (рамки считывания), прямые повторы и инвертированные повторы. Сдвиги кадра включают вставки, делеции и мутации.

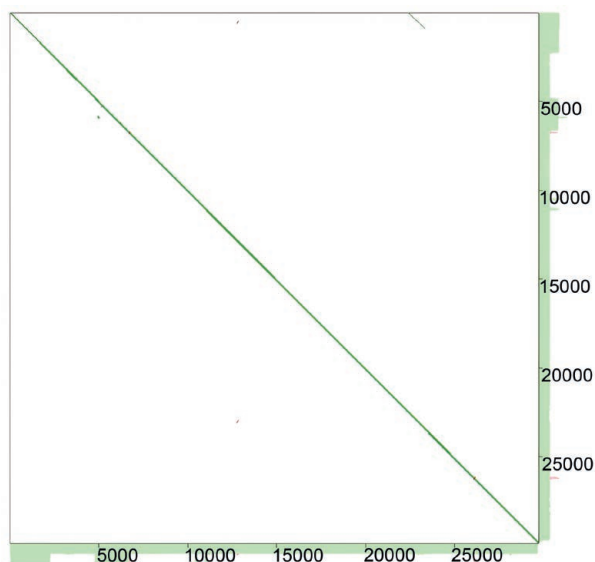


Рис. 4. Сравнение белорусских образцов коронавируса SARS-CoV-2 и генома SARS-CoV (выявлена наибольшая схожесть по сравнению со всеми другими небелорусскими образцами вирусов)
Fig. 4. Comparison of Belarusian samples of SARS-CoV-2 coronavirus and the SARS-CoV genome (the greatest similarity was revealed compared to all other non-Belarusian virus samples)

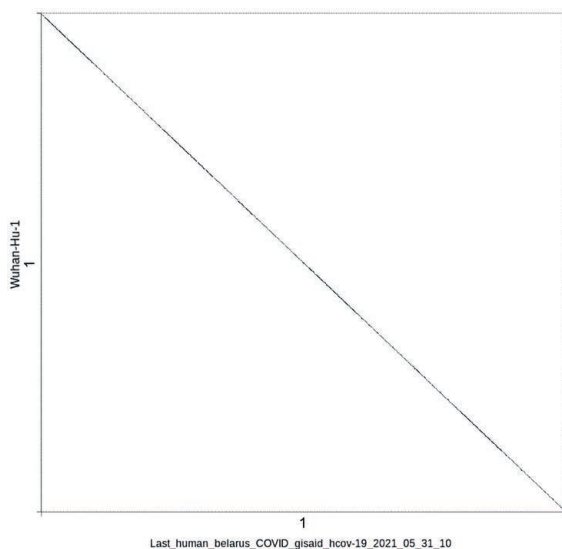


Рис. 5. Сравнение одного из последних просеквенированных геномов со ссылочным геномом-эталонном
Fig. 5. Comparison of one of the latest sequenced genomes with a reference genome

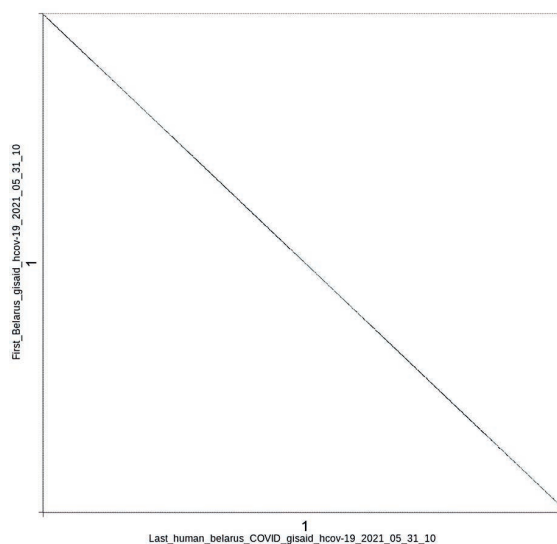


Рис. 6. Сравнение первого и последнего просеквенированных белорусских геномов SARS-CoV-2 между собой
Fig. 6. Comparison between the first and the last sequenced Belarusian SARS-CoV-2 genomes

По результатам вычислительного эксперимента выявлены значительная гомологичность нового коронавируса с рекомбинантным коронавирусом, частичное сходство с вирусами синтетического коронавируса, Марбурга, краснухи, Эбола 1976, ВИЧ-2, вируса иммунодефицита обезьян. Белорусские вирусы оказались предельно идентичными изоляту из метагенома, полученного в Гомеле (данными поделился Е. В. Воропаев, 2021 [9]), и структурно самоподобными, что означает их высокую (более 98 %) структурную гомологию на данной географической территории.

По данным современной научной литературы, результаты сравнительного геномного анализа SARS-CoV-2 и SARS-CoV выявили чрезвычайно высокую гомологию между двумя штаммами на уровне нуклеотидов.

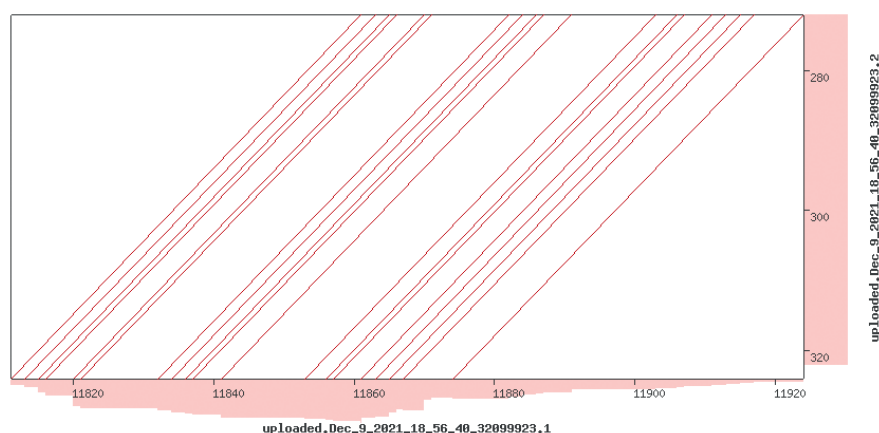


Рис. 7. Сравнение белорусских образцов коронавируса SARS-CoV-2 и генома ВИЧ-2
Fig. 7. Comparison of Belarusian samples of SARS-CoV-2 coronavirus and HIV-2 genome

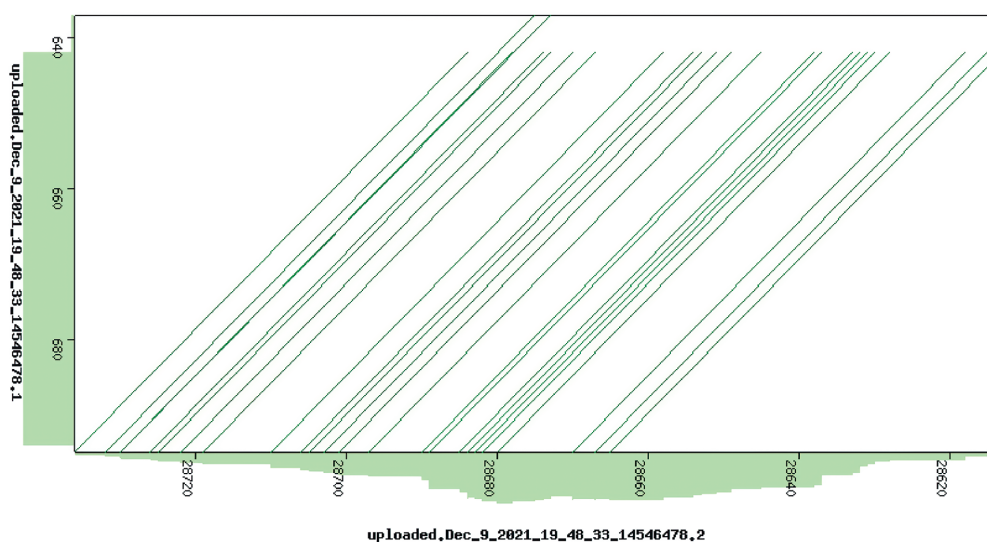


Рис. 8. Сравнение белорусских образцов коронавируса SARS-CoV-2 и генома вируса Эбола 1976
Fig. 8. Comparison of Belarusian samples of SARS-CoV-2 coronavirus and the 1976 Ebola virus genome

Тем не менее геномы этих двух штаммов отличаются друг от друга в нескольких регионах. Шиповидный ген SARS-CoV-2 демонстрирует большую гомологию с геномами коронавирусов летучих мышей, тогда как два дополнительных гена 3А и 8В обладают гомологией с SARS-CoV. Протеомный анализ сходства SARS-CoV и SARS-CoV-2 показал, что большинство белков в высокой степени (95–100 %) гомологичны. Протеазы RdRp и 3CLpro имеют более 95 % сходства последовательностей, хотя на уровне генома эти два штамма сходны только на 82 %. Перечисленные доказательства предполагают общую эволюционную историю обоих вирусных штаммов. Однако SARS-CoV-2 обладает двумя белками (ORF8 и ORF10), которые не имеют гомологии со штаммом SARS-CoV. Консервативные последовательности ORF8 для SARS-CoV отличаются от аминокислотной последовательности ORF8, полученной для SARS-CoV-2 [5].

Ряд исследований предполагают тесную связь между SARS-CoV-2 и SARS-CoV с возбудителем SARS у летучих мышей CoV. Анализ структурно-функциональных различий между SARS-CoV-2 и SARS-CoV может быть полезным ввиду выявления отдельных аминокислотных замен в определенных белках. Всего обнаружено 380 аминокислотных замен у SARS-CoV-2 при сравнении с SARS-CoV. Оболочка, матрикс или вспомогательные белки р6 и 8b, nsp7 и nsp13 не выявили изменений аминокислотной последовательности. В неструктурных белках nsp2 и nsp3 обнаружена одиночная аминокислотная замена в положении 61 и 102 соответственно. Кроме того, в шиповидном белке было обнаружено 27 аминокислотных замен длиной 1273 аминокислот, в том числе аминокислотный регион 357–528 в RBD (рецептор-связывающем домене) с шестью заменами, и еще шесть замен в аминокислотной области 569–655 в основном субдомене (SD). Кро-

ме того, в С-конце домена S1 рецептор-связывающей субъединицы обнаружили четыре замены (Q560L, S570A, F572T и S575A). Рецептор-связывающие мотивы, взаимодействующие с человеческим АПФ-2 рецептором (ангиотензин превращающего фермента), оказались точно гомологичны фрагменту генома штамма SARS-CoV-2 практически без отличий в виде аминокислотных замен. По мнению авторов [2], из-за ограниченных знаний о SARS-CoV-2 разумные объяснения этих аминокислотных замен пока неизвестны.

Прямое зоонозное происхождение SARS-CoV-2 не установлено. Однако теоретические сравнительные исследования генетических последовательностей комплекса коронавируса выявили высокое сродство этого вируса с возбудителем SARS китайских летучих мышей (SARSr-CoV, RaTG13), поскольку он на 96,2 % имеет с ним генетическую гомологию. Несмотря на высокий уровень генетического сходства нового вируса SARS-CoV-2 по сравнению с другими штаммами CoV, обнаружено, что геном SARS-CoV-2 содержит несколько отличительных характеристик, таких как мутации в области S1-S2 рецептор-связывающего домена многоосновного фурина (сайт расщепления фурина, furin cleavage sites) и о-связанные гликаны. Регионы отличия различных штаммов коронавируса имеют значение для разработки лекарств, поскольку эти геномные вариации могут быть причиной повышенной заразности и тяжести заболевания. Таким образом, перепрофилирование лекарств/вакцин в настоящее время является актуальной областью исследований. Ученые стремятся использовать последовательности нуклеотидов и аминокислот антигенов-мишеней ранее известных штаммов коронавируса, чтобы разработать вакцину и лекарственные средства для нового штамма SARS-CoV-2. Гликопротеин S (спайк, шип) отвечает за опосредование проникновения вируса в клетки человека и таким образом определяет вирусную заразность и возможности распространения среди различных млекопитающих. Ограничение его расщепления и манипуляции с экспрессией рецептора АПФ-2 могут стать эффективной терапией. Более того, S-белок является наиболее потенциальным антигенным сайтом-мишенью [2].

Заключение

1. Данные вычислительных экспериментов и научных работ свидетельствуют о том, что новый коронавирус гомологичен другим коронавирусам, особенно SARS-CoV и рекомбинантному коронавирусу, и имеет вставки молекулярных мотивов из геномов опасных и особо опасных вирусов: иммунодефицита обезьян и человека, Эбола, других микроорганизмов. Происхождение такого структурного ансамбля по-прежнему неизвестно. Необходимо дальнейшее изучение (аннотирование, распознавание структурного состава и потенциальных функций, сопоставление с другими массивами данных) вставок геномов опасных и особо опасных вирусов для выяснения происхождения и природы коронавирусной пандемии и разработки стратегий и технологий противостояния вирусным эпидемиям.

2. Предмет изучения идентичности и гомологичности геномных текстов SARS-CoV-2 должен стать темой отдельного научного проекта и основанием для разработки новых методик и алгоритмов анализа биоинформационных данных.

3. Функции программного обеспечения, апробированного в рассмотренном вычислительном эксперименте, можно использовать в разрабатываемой авторами автоматизированной системе анализа биоинформационных данных геномной природы.

Список литературы

1. Behbahani, M. Analysis and Comparison of Physiochemical Properties, Mutations and Glycosylation Patterns between RNA Polymerase and Membrane Protein of SARS-CoV and SARS-CoV-2 / M. Behbahani, P. Rabiei, H. Mohabatkar // *Molecular Biology Research Communications*. 2021. Vol. 10, No 4. P. 171–178. Doi: 10.22099/mbrc.2021.42187.1692.
2. Genetic Comparison Among Various Coronavirus Strains for the Identification of Potential Vaccine Targets of SARS-CoV-2 / N. Kaur [et al.] // *Infection, Genetics and Evolution*. 2021. Vol 89. Doi: 10.1016/j.meegid.2020.104490.
3. Макаров, Л. М. Сравнительный анализ штаммов коронавируса / Л. М. Макаров, Д. О. Иванов, А. В. Поздняков // *European Science*. 2020. Т. 53, № 4. С. 61–66.
4. Харченко, Е. П. Коронавирус SARS-CoV-2: особенности структурных белков, контагиозность и возможные иммунные коллизии / Е. П. Харченко // *Эпидемиология и вакцинопрофилактика*. 2020. Т. 19, № 2. С. 13–30. <https://doi.org/10.31631/2073-3046-2020-19-2-13-30>.

5. Noé, L. YASS: Enhancing the Sensitivity of DNA Similarity Search / L. Noé, G. Kucherov // *Nucleic Acids Research*. 2005. Vol. 33. P. W540–W543. <https://doi.org/10.1093/nar/gki478>.
6. Genome Detective: an Automated System for Virus Identification from High-Throughput Sequencing Data / M. Vilsker [et al.] // *Bioinformatics*. 2019. Vol. 35, Iss. 5. P. 871–873. <https://doi.org/10.1093/bioinformatics/bty695>.
7. Genome Detective Coronavirus Typing Tool for Rapid Identification and Characterization of Novel Coronavirus Genomes / S. Cleemput [et al.] // *Bioinformatics*. 2020. Vol. 36, Iss. 11. P. 3552–3555. Doi: 10.1093/bioinformatics/btaa145.
8. Автоматизированный конвейер анализа данных геномов коронавируса / М. В. Спринджук [и др.] // *Медэлектроника-2022. Средства медицинской электроники и новые медицинские технологии: сб. науч. статей XIII Междунар. науч.-техн. конф., Минск, 8–9 дек. 2022 г. Минск: Белор. госуд. ун-т информ. и радиоэлек., 2022. С. 60–65.*
9. Воропаев, Е. В. Опыт использования современных геномных технологий для изучения микроорганизмов и их сообществ / Е. В. Воропаев, И. О. Стома, Д. В. Тапальский // *Проблемы здоровья и экологии*. 2021. Т. 18, № 3. С. 159–167.

References

1. Behbahani M., Rabiei P., Mohabatkar H. (2021) Analysis and Comparison of Physiochemical Properties, Mutations and Glycosylation Patterns between RNA Polymerase and Membrane Protein of SARS-CoV and SARS-CoV-2. *Molecular Biology Research Communications*. 10 (4), 171–178. Doi: 10.22099/mbrc.2021.42187.1692.
2. Kaur N., Singh R., Dar Z., Bijarnia R. K., Dhingra N., Kaur T. (2021) Genetic Comparison Among Various Coronavirus Strains for the Identification of Potential Vaccine Targets of SARS-CoV-2. *Infection, Genetics and Evolution*. 89. Doi: 10.1016/j.meegid.2020.104490.
3. Makarov L. M., Ivanov D. O., Pozdnyakov A. V. (2020) Comparative Analysis of Coronavirus Strains. *European Science*. 53 (4), 61–66 (in Russian).
4. Kharchenko E. P. (2020) Coronavirus SARS-CoV-2: Features of Structural Proteins, Contagiousness and Possible Immune Collisions. *Epidemiology and Vaccine Prevention*. 19 (2), 13–30. <https://doi.org/10.31631/2073-3046-2020-19-2-13-30> (in Russian).
5. Noé L., Kucherov G. (2005) YASS: Enhancing the Sensitivity of DNA Similarity Search. *Nucleic Acids Research*. 33, W540–W543. <https://doi.org/10.1093/nar/gki478>.
6. Vilsker M., Moosa Y., Nooij S., Fonseca V., Ghysens Y., Dumon K., Pauwels R., Alcantara L. C., Eynden E. V., Vandamme A.-M., Deforche K., de Oliveira T. (2019) Genome Detective: an Automated System for Virus Identification from High-Throughput Sequencing Data. *Bioinformatics*. 35 (5), 871–873. <https://doi.org/10.1093/bioinformatics/bty695>.
7. Cleemput S., Dumon W., Fonseca V., Wasim A. K., Giovanetti M., Alcantara L. C., Deforche K., de Oliveira T. (2020) Genome Detective Coronavirus Typing Tool for Rapid Identification and Characterization of Novel Coronavirus Genomes. *Bioinformatics*. 36 (11), 3552–3555. Doi: 10.1093/bioinformatics/btaa145.
8. Sprindzuk M. V., Bernik V. I., Kalosha N. I., Vladyko A. S., Ulzyibat B., Batgerel B., Titov L. P., Klimuk D. A., Skriahina E. M., Skriahin A. E., Glinskaya T. N. (2022) Automated Data Analysis Pipeline for Coronavirus Genomes. *Medeelectronics-2022. Means of Medical Electronics and New Medical Technologies, Collection of Scientific Articles XIII International Scientific and Technical Conference, Minsk, 8–9 Dec. 2022. Minsk, Belarusian State University of Informatics and Radioelectronics Publ.* 60–65 (in Russian).
9. Voropaev E. V., Stoma I. O., Tapalsky D. V. (2021) Experience of Modern Genomic Technologies Application for the Study of Microorganisms and their Communities. *Problems of Health and Ecology*. 18 (3), 159–167 (in Russian).

Вклад авторов

Спринджук М. В. создал концепцию статьи, выполнил вычислительный эксперимент и написал основной текст.

Берник В. И., Владыко А. С., Чжочжуан Лу, Титов Л. П. консультировали Спринджук М. В. по вопросам математики и иммунологии, проверяли рукопись, написанную Спринджук М. В., участвовали в написании текста заявок научных проектов, при поддержке которых выполнено данное научное исследование.

Authors' contribution

Sprindzuk M. V. created the concept of the article, performed a computational experiment and wrote the core text.

Bernik V. I., Vladyko A. S., Zhuozhuang Lu, Titov L. P. advised Sprindzuk M. V. on questions of mathematics and immunology, checked the manuscript written by Sprindzhuk M. V., participated in writing the text of proposal applications for scientific projects, with the support of which this research was done.

Сведения об авторах

Спринджук М. В., к. т. н., старший научный сотрудник Объединенного института проблем информатики НАН Беларуси, Института математики НАН Беларуси, докторант кафедры электронных вычислительных средств Белорусского государственного университета информатики и радиоэлектроники

Берник В. И., д. ф.-м. н., профессор, главный научный сотрудник отдела теории чисел Института математики НАН Беларуси

Владыко А. С., д. м. н., профессор, главный научный сотрудник Республиканского научно-практического центра эпидемиологии и микробиологии

Чжочжуан Лу, д. м. н., профессор, главный научный сотрудник Китайского центра по контролю и профилактике заболеваний

Титов Л. П., академик НАН Беларуси, д. м. н., профессор, заведующий лабораторией экспериментальной иммунологии Республиканского научно-практического центра эпидемиологии и микробиологии

Адрес для корреспонденции

220012, Республика Беларусь,
г. Минск, ул. Сурганова, 6
Объединенный институт
проблем информатики НАН Беларуси
Тел.: +375 33 682-57-55
E-mail: bioinformatics_bel@yahoo.com
Спринджук Матвей Владимирович

Information about the authors

Sprindzuk M. V., Cand. of Sci., Senior Researcher at the United Institute of Informatics Problems of the NAS of Belarus, the Institute of Mathematics of the NAS of Belarus, Dr. Student at the Department of Electronic Computing Media of the Belarusian State University of Informatics and Radioelectronics

Bernik V. I., Dr. of Sci. (Phys. and Math.), Professor, Principal Researcher at the Department of Number Theory of the Institute of Mathematics of the NAS of Belarus

Vladyko A. S., Dr. of Sci. (Med.), Professor, Principal Researcher at the Republican Research and Practical Center for Epidemiology and Microbiology

Zhuozhuang Lu, Dr. of Sci. (Med.), Professor, Principal Researcher of the Chinese Center for Disease Control and Prevention

Titov L. P., Academic of the NAS of Belarus, Dr. of Sci. (Med.), Professor, Head of Laboratory of Experimental Immunology of the Republican Research and Practical Center for Epidemiology and Microbiology

Address for correspondence

220012, Republic of Belarus,
Minsk, Surganova St., 6
The United Institute
of Informatics Problems of the NAS of Belarus
Tel.: +375 33 682-57-55
E-mail: bioinformatics_bel@yahoo.com
Sprindzuk Matvey Vladimirovich