

Программное обеспечение и возможности современных языков программирования для изучения биоинформатики и вычислительной вакцинологии новой коронавирусной инфекции

М. В. Спринджук, к. т. н., старший научный сотрудник лаборатории математической кибернетики

E-mail: stepanenkomatvei@yandex.ru

Объединенный институт проблем информатики НАН Беларуси,
ул. Сурганова, д. 6, 220012, г. Минск, Республика Беларусь

А. С. Владыко, д. м. н., профессор, главный научный сотрудник лаборатории биотехнологии и иммунодиагностики ООИ РНПЦ эпидемиологии и микробиологии РНПЦ эпидемиологии и микробиологии,
ул. Филимонова, д. 23, 220114, г. Минск, Республика Беларусь

Л. П. Титов, д. м. н., профессор, член – корреспондент НАН Беларуси, заведующий лабораторией клинической и экспериментальной микробиологии РНПЦ эпидемиологии и микробиологии

РНПЦ эпидемиологии и микробиологии, ул. Филимонова, д. 23, 220114, г. Минск, Республика Беларусь

А. П. Кончиц, к. б. н., ведущий научный сотрудник лаборатории лесной селекции и семеноводства

E-mail: konchits@yandex.ru

Институт леса НАН Беларуси, ул. Пролетарская, д. 71, 246001, г. Гомель, Республика Беларусь

Аннотация. Обзорная статья сосредоточена на вопросах применения программного обеспечения для целей геномики, иммуноинформатики, вычислительной вакцинологии, математической эпидемиологии и филогенеза новой коронавирусной инфекции. Приводится разработанная авторами классификация программного обеспечения для изучения COVID-19.

Ключевые слова: SARS-CoV-2, COVID-19, коронавирус, эпидемия, пандемия, автоматизированные системы обработки данных, программное обеспечение, медицинская кибернетика, геномика, транскриптомика, системы медицинского назначения, иммуноинформатика.

Для цитирования: Спринджук, М. В. Программное обеспечение и возможности современных языков программирования для изучения биоинформатики и вычислительной вакцинологии новой коронавирусной инфекции / М. В. Спринджук, А. С. Владыко, Л. П. Титов, А. П. Кончиц // Цифровая трансформация. – 2021. – № 3 (16). – С. 47–57.



© Цифровая трансформация, 2021

Software and Resources of Modern Programming Languages for Bioinformatics and Computational Vaccinology Research of the New Coronavirus Infection

M. V. Sprindzuk, TechSciPhD, Senior Researcher, laboratory of mathematical Cybernetics

E-mail: stepanenkomatvei@yandex.ru

United Institute for Informatics Problems of the NAS of Belarus,
Surganova Str, 6, 220012 Minsk, Republic of Belarus

A.S. Vladyko, Doctor of Sciences (Medical), Professor, chief researcher of the biotechnology and immunodiagnosis laboratory RRPC for Epidemiology and Microbiology, Republic of Belarus,
Filimonova Str., 23, 220114 Minsk, Republic of Belarus

L. P. Titov, Doctor of Sciences (Medical), Professor, Corresponding Member of the NAS of Belarus, Head of the Laboratory for Clinical and Experimental Microbiology

RRPC for Epidemiology and Microbiology, Republic of Belarus,
Filimonova Str., 23, 220114 Minsk, Republic of Belarus

A. P. Konchits, BioSciPhD Leading Researcher, Forest Tree Breeding and Seed Production Laboratory

E-mail: konchits@yandex.ru

Forest Research Institute of the NAS of Belarus, 71 Proletarskaya Str.,
246001 Gomel, Republic of Belarus

Abstract. The review paper focuses on the application of software for the purposes of genomics, immunoinformatics, computational vaccinology, mathematical epidemiology and phylogeny of the new coronavirus infection. Authors provide a classification of software for the investigation of COVID-19.

Key words: SARS-CoV-2, COVID-19, coronavirus, epidemic, pandemic, automated data processing systems, software, medical cybernetics, genomics, transcriptomics, medical systems, immunoinformatics

For citation: Sprindzuk, M. V., Vladyko A.S., Titov L. P., Konchits A. P. Software and resources of modern programming languages for bioinformatics and computational vaccinology research of the new coronavirus infection. *Cifrovaja transformacija* [Digital transformation], 2021, 3 (16), pp. 47–57 (in Russian).

© Digital Transformation, 2021

Введение: актуальность и значение COVID-19 инфекции. По данным www.worldometers.info/coronavirus, на 4 марта 2021 года (6.30, среднее время по Гринвичу) в мире от новой коронавирусной инфекции погибло 2 571 794 человек, из них 531 652 в США, 259 402 в Бразилии, 87 348 в России, 71 711 в Германии, 44 360 в Польше, 3 281 в Литве, и 2002 в Беларуси.

Ущерб COVID-19 (Coronavirus disease 2019, коронавирусная болезнь 2019) для мировой экономики огромен, расходы и потери продолжают.

Основным природным резервуаром инфекции, по-видимому, являются подковоносные рукокрылые (летучие мыши) и панголины (рисунок 1).

По теме COVID-19 в онлайн базе данных На-



Рисунок 1. Панголин – один из предполагаемых природных резервуаров SARS– CoV–2 инфекции
Figure 1. Pangolin is one of the suspected natural reservoirs of SARS-CoV-2 infection

ционального института здоровья США (www.ncbi.nlm.nih.gov/research/coronavirus) доступно 104 872 научных публикаций, и их число каждый день растет.

Разработка эффективной вакцины и новых лекарственных средств для профилактики и лечения инфекции, ассоциированной с новым SARS-CoV-2 (тяжелый острый респираторный синдром, ТОРС) коронавирусом, представляется неотложной актуальной проблемой.

Более 200 вакцин проходят в настоящее время лабораторные и клинические испытания (<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/covid-19-vaccines>).

В Беларуси вакцины (кроме антибактериальных) ранее не разрабатывались. Для изучения проблемы медицинской кибернетики и геномики новой коронавирусной инфекции нами был выполнен сбор и анализ литературы по следующим базам данных:

- 1) Google scholar (<https://scholar.google.ru/>);
- 2) EBSCO (<https://www.ebsco.com/>);
- 3) КиберЛенинка (<https://cyberleninka.ru/>);
- 4) Scielo (<https://scielo.org/>);
- 5) PubMed (<https://pubmed.ncbi.nlm.nih.gov/>);
- 6) Google patents (<https://patents.google.com/>);
- 7) IEEE explore (<https://ieeexplore.ieee.org/Xplore/home.jsp>);
- 8) BioRxiv (<https://www.biorxiv.org/>);
- 9) Научная электронная библиотека диссертаций и авторефератов disserCat (<https://www.dissercat.com/>);
- 10) Электронный каталог Национальной библиотеки Беларуси, <https://e-catalog.nlb.by/>;
- 11) Каталог патентов и изобретений СНГ и Евразии <https://fips.ru/iiss/>;
- 12) Каталог европейских диссертаций <https://oatd.org/>.

Помимо названных веб-серверов для поиска и отбора литературы были использованы технические средства программного обеспечения EndNote, RefNavigator, Mendeley, JabRef.

Как источники наиболее полезной информации, обращают на себя внимание четыре книги по теме современной иммуноинформатики [1,2,3], 1 книга по биоинформатике для вакцинологии [4], 3 монографии по COVID-19 инфекции [5,6,7], учебник [8] и 6 монографий по вакцинологии и генетике коронавирусов [9,10,11,12,13,14], 2 статьи-обзора, затрагивающих вопросы иммуноинформатики [15] и идентификации эпитопов [16] из геномов [17] новой патогенной для человека коронавирусной инфекции, 4 научных статьи, сообщающих об опыте вычисления [18,19] и in

vivo апробирования эпитопов [20,21] из коронавирусных геномов.

Из русскоязычных источников, обратили на себя внимание статьи [22,23,24,25,26,27] и диссертация [28].

Векторные и эпитопные вакцины. Изучение накопленной литературы показывает, что векторные вакцины — это одно из предполагаемых средств профилактики особо опасных вирусных заболеваний. Накоплен неоднозначный мировой опыт разработки вакцин такого типа для вирусов Эболы, Зика, MERS (Ближневосточный респираторный синдром (БВРС, англ. Middle East respiratory syndrome), SARS и опубликован ряд исследований по новой теме SARS-CoV-2 вакцины, в том числе на основе аденовирусного вектора.

Векторные вакцины относят к рекомбинантным вакцинам, которые получают генно-инженерными методами. В качестве векторов используют живые аттенуированные (ослабленные) вирусы, бактерии, дрожжи или эукариотические клетки, в которые встраивают ген, кодирующий образование предполагаемого протективного антигена возбудителя, стимулирующего в организме образование защитных антител. В качестве носителя бактериального вектора используют БЦЖ (сокр. от Бацилла Кальмёта-Герёна, фр. Bacillus Calmette-Guérin, BCG), Vibrio cholera, Salmonella typhimurium, Escherichia coli. Преимущество бактериальных векторных вакцин перед вирусными заключается в возможности контролировать их с помощью антибиотиков. В качестве носителей вирусных векторов используют вирусы осповакцины, бакуловирусы, аттенуированные аденовирусы. Преимущество в использовании вирусов в качестве вектора — более длительная персистенция вирусов в организме по сравнению с бактериями.

Общепринятым условием и утверждением является то, что эффективная антивирусная вакцина должна индуцировать высокие титры нейтрализующих антител, чтобы предотвратить прикрепление вируса к рецепторам клетки-хозяина. Однако для получения защитных уровней антител при вакцинации может потребоваться несколько доз или помощь со стороны других иммуностимулирующих молекул. Кроме того, вакцина должна способствовать выработке защитных уровней антител при наименьшем количестве используемого антигена. Это снижает стоимость вакцины и делает ее более доступной. Включение адъюванта может позволить значительно сократить количество антигена в вакцине [22].

Преимущества и недостатки пептидных вакцин.

Преимущества пептидных (эпитопных вакцин):

1) стоимость разработки меньше, чем полновирионных инактивированных и аттенуированных вакцин;

2) меньше риски анафилаксии, аллергенности, и других побочных эффектов (нарколепсии, бронхиальной астмы, аутоиммунных заболеваний и рассеянного склероза);

3) лучше стабильность и устойчивость растворов и транспортабельность субстанции и препарата вакцины.

Недостатки пептидных вакцин:

1) относительно низкая иммуногенность, что можно компенсировать добавлением так называемых адъювантов: наночастиц и хитозана;

2) отмечается подверженность к ферментативному распаду.

Обобщая информацию из доступной литературы можно классифицировать эпитопы, входящие в состав вакцины следующим образом:

1) линейные (последовательности геномного текста аминокислот) и конформационные (3D модели);

2) по мишени сродства и связывания: а) Т- лимфоцитов (CD4+ и CD8+), представляющих клеточный иммунитет; б) В- лимфоцитов, эффекторов гуморального иммунитета; в) - интерферонов, противовирусных белков.

Методология построения пептидных вакцин имеет свою определенную последовательность: просеквенированный геном подвергается предобработке, анализу, аннотированию и трансляции нуклеотидных последовательностей в аминокислотные. Затем отобранные кандидатные гены вируса анализируются на предмет антигенного потенциала, другими словами, иммуногенности, а также оценивается потенциал их эффективности и безвредности для организма-реципиента вакцины.

Изучение и анализ программного обеспечения и возможностей языков программирования для целей анализа геномов COVID-19. За последние десять лет, с момента распространения эпидемий SARS (тяжёлый острый респираторный синдром, ТОРС), MERS и особенно пандемии SARS-CoV-2, отмечается подъем интереса исследователей к проблемам геномики возбудителей этих инфекций, что сопровождается ростом разработки программного обеспечения для биоинформатики и иммуноинформатики, без которых невозможен углубленный анализ краткосрочной и долговременной эволюции вирусов, молекул иммунной системы в целях иммунодиагностики и иммунопрофи-

лактики. Постоянно обновляемый список доступного программного обеспечения можно найти в источнике [https://www.denbi.de/covid-19/coronavirus-tools].

Классификация программного обеспечения для изучения новой коронавирусной инфекции. Ключевые статьи, в которых наиболее полноценно и детально рассматривается программное обеспечение для многостороннего изучения патогена SARS-CoV-2 – это источники [29,30,31].

Такое программное обеспечение можно классифицировать следующим образом:

1) полностью бесплатные для пользователей веб-серверы для различных задач анализа нуклеотидных последовательностей [https://www.ebi.ac.uk/Tools/emboss/] (см. рисунки 2, 3);

2) бесплатные многоцелевые веб-порталы, специфически разработанные по теме COVID-19 (Coronavirus disease 2019, коронавирусная болезнь 2019) [https://hdsu-bioquant.shinyapps.io/marmycorona/];

3) многоцелевые и кастомизированные облачные серверы с веб-интерфейсом и интерфейсом прикладного программирования на основе конвейерных технологий анализа биоинформационных данных [https://aries.iss.it/];

4) настольные (десктоп) приложения для специализированных задач биоинформатики [http://darlinglab.org/mauve/download.html];

5) пакеты - модули современных скриптовых языков программирования;

6) контейнеризированные программные приложения и их ансамбли для быстрого и надежного развертывания на разных операционных системах;

7) низко- и высокоуровневые библиотеки компилируемых и интерпретируемых языков программирования для разработки программных комплексов;

8) коммерческое программное обеспечение (полностью или частично), которое может иметь веб или настольный интерфейс или интерфейс прикладного программирования посредством ресурсов скриптового языка программирования [https://www.biobam.com/download-omicsbox/].

Биоинформационные задачи, которые возможно решать с помощью программного обеспечения:

1) интерактивный трекинг эпидемиологических данных COVID-19 распространения;

2) математическое моделирование COVID-19 распространения и передачи инфекции;

3) поиск, отбор и загрузка файлов геномов и других данных;

4) тримминг и устранение ошибок в геномных

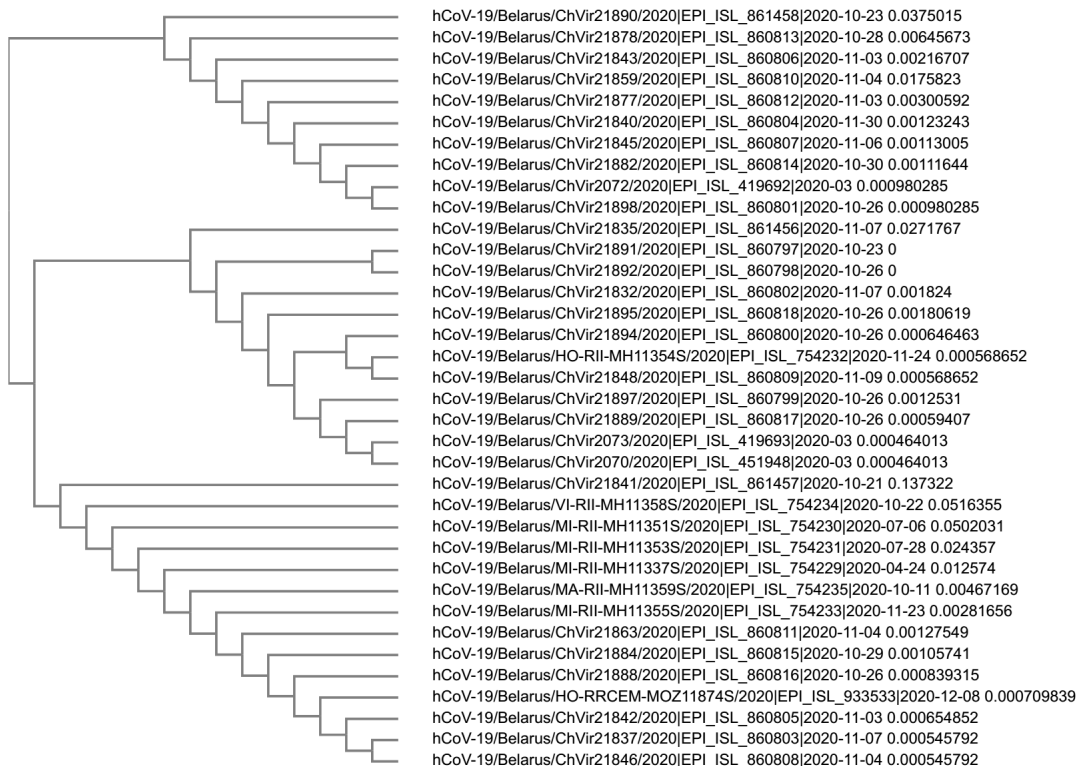


Рисунок 2. Clustal Omega кладограмма 36 образцов геномов SARS-CoV-2 от пациентов в Беларуси
Figure 2. Clustal Omega cladogram of 36 SARS-CoV-2 genome samples from patients in Belarus

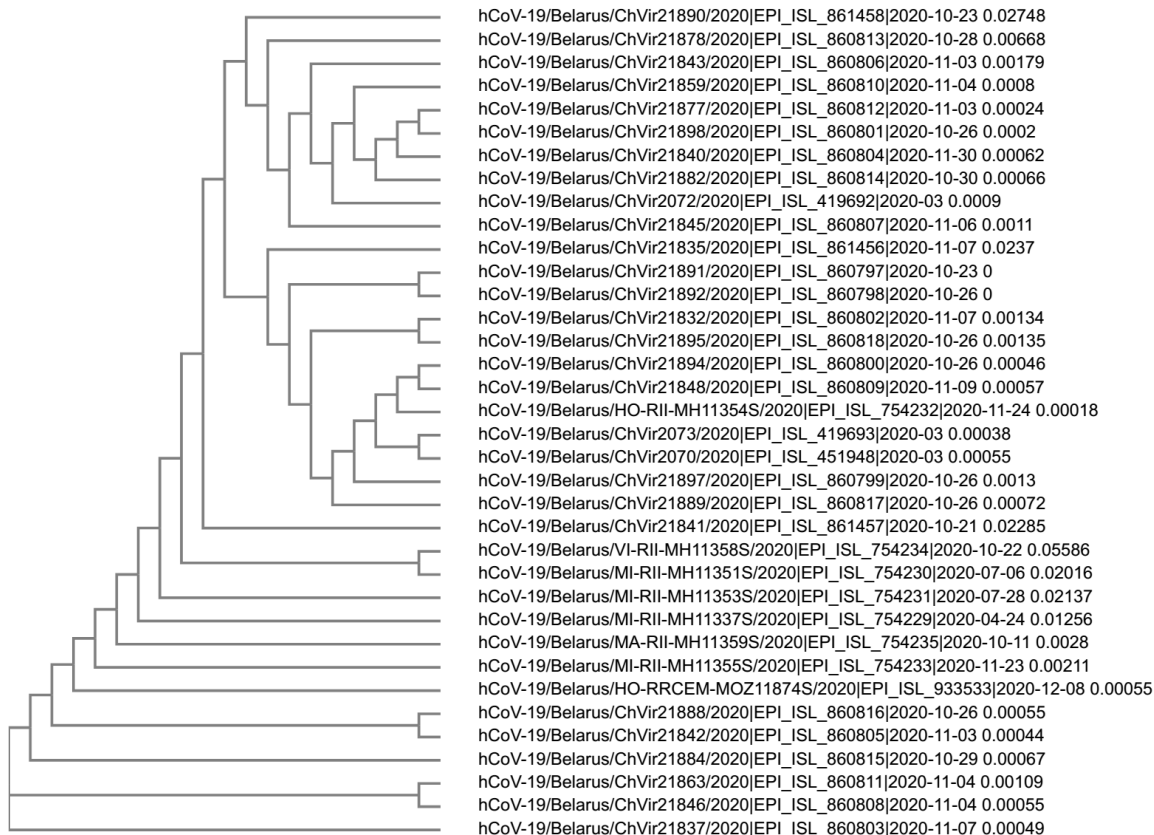


Рисунок 3. Кладограмма 36 образцов геномов SARS- CoV-2 от пациентов в Беларуси, алгоритм дерева, объединяющего соседние элементы без коррекции расстояний
Figure 3. Cladogram of 36 SARS- CoV-2 genome samples from patients in Belarus, tree algorithm combining neighboring elements without distance correction

текстах;

5) одиночное и множественное выравнивание геномов на ссылочные геномы и между собой, получение мутационных профилей и аннотирование генов;

6) поиск и визуализация схожести геномных текстов;

7) распознавание и классификация геномов с помощью методологии современного искусственного интеллекта;

8) де ново сборка геномов из длинных и коротких прочтений, гибридная сборка и сборка с использованием ссылочных геномов и дополнительных геномов с целью деконтаминации данных;

9) эволюционный анализ групп геномов с визуализацией в виде кладограмм и кластерограмм;

10) идентификация участков генома или транскриптома максимальной предполагаемой иммуногенности;

11) in silico анализ аллергенности и токсигенности пептидов;

12) in silico анализ предполагаемой локализации пептида;

13) in silico анализ предполагаемых физических и химических свойств пептида;

14) структурное и функциональное аннотиро-

вание участков генома или транскриптома;

15) прогноз белок-белковых взаимодействий, в том числе искусственного пептида и рецепторов клеток человека;

16) вычисление и прогнозирование биофизических свойств возможных адъювантов и линкеров (связующих элементов) вакцины;

17) организация, представление и визуализация элементов вакцины, собранных воедино.

Программное обеспечение для специфических целей иммуноинформатики и вычислительной биологии. Обзорные статьи, в которых фокус исследования сосредоточен на рассмотрении функциональных возможностей вычислительных инструментов для целей иммуноинформатики, вакцинологии COVID-19 – это источники литературы [16,32,33].

Китайской стороной было рекомендовано использовать программный комплекс Vaxign [34,35,36] как наиболее проверенный опытом разработки анти-вирусных вакцин.

Среди множества доступного программного обеспечения, согласно данным современной литературы, для вычисления антигенных детерминат по данным геномов вирусов рекомендовано использовать следующее программное обеспечение:

- International A_B Diversity —
- B.1.1.7_501Y.V1_20I —
- B.1.351_501Y.V2_20H —
- P1_501Y.V3_20J —
- Y453F.Cluster5_20B —
- B.1.1.70_501Y_20B —

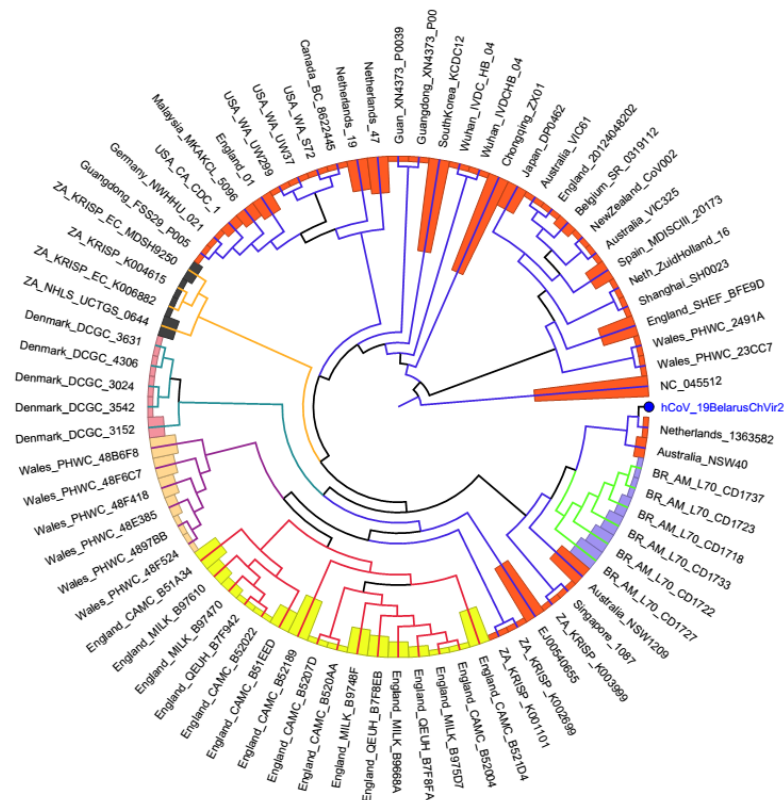


Рисунок 4. Циркулярная филограмма (синим обозначен один из первых образцов генома COVID-19 от пациента в Беларуси)
Figure 4. Circular phylogram (blue indicates one of the first COVID-19 genome samples from a patient in Belarus)

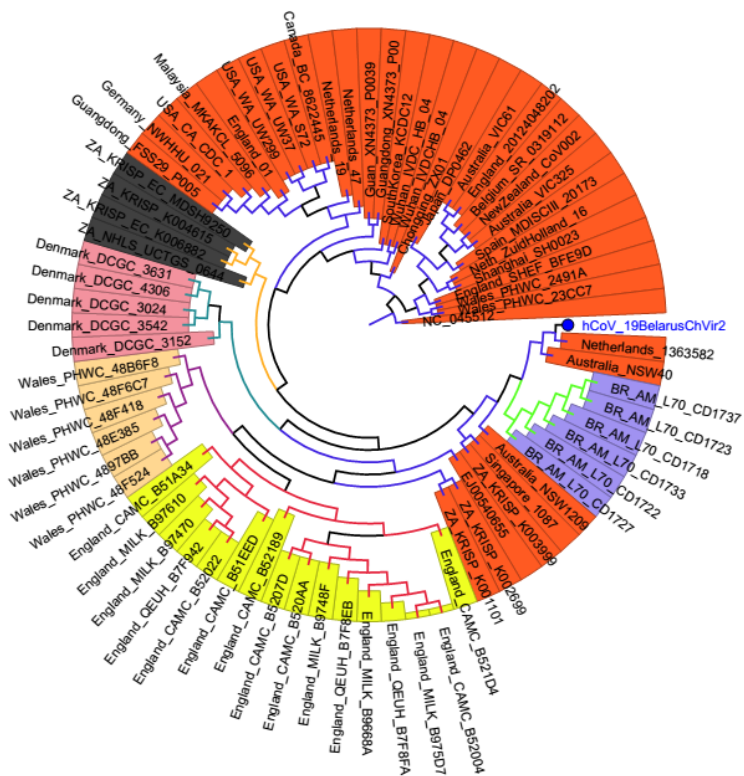


Рисунок 5. Расширенная циркулярная филограмма
 Figure 5. Extended circular phylogram

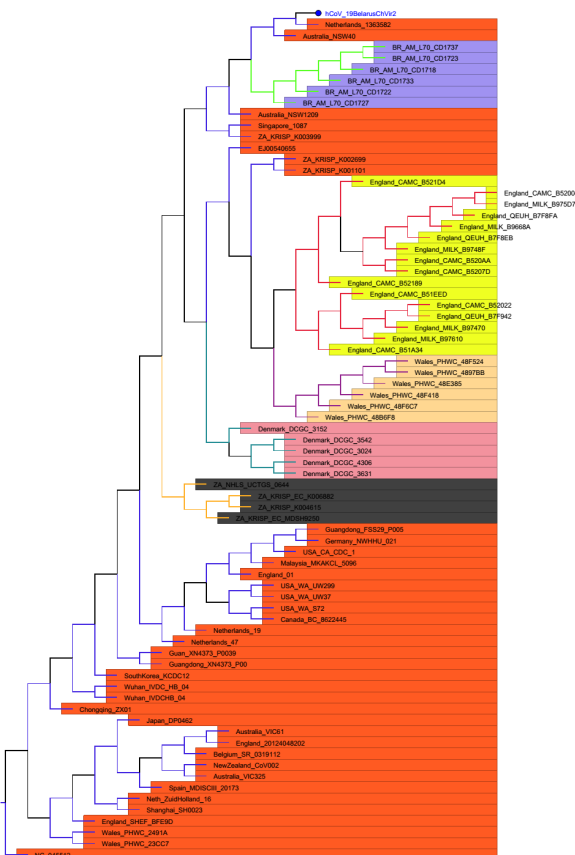


Рисунок 6. Прямоугольное представление филограммы
 Figure 6. Rectangular phylogram representation

1) веб-сервер IEDB (The Immune Epitope Database) с его инструментами для вычисления эпитопов и их необходимых характеристик [37];

2) компьютерная программа VCPred [38];

3) Prokka [39], EggNog [40], Interproscan [41,42,43,44] для аннотирования контигов;

4) Mafft [45,46,47,48], ClustalW [49,50,51,52], Mauve [53,54], Mugsy [55] для множественного выравнивания;

5) Genome detective (см. рисунки 4–6) [56,57] для идентификации мутационного профиля, визуализации филогенеза и изучения эволюции коронавируса

6) ToxinPred [58,59] для проверки токсичности синтетических пептидов [<http://crdd.osdd.net/raghava/toxinpred/>] и AllergCatPro [60,61], AlgPred [62,63,64,65], AllergenFP [<https://ddg-pharmfac.net/AllergenFP/>] для оценки аллергенности пептидов;

7) VaxiJen [<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>] для оценки антигенности эпитопов;

8) веб-серверы для изучения докинга с рецепторами клеток реципиента вакцины [66];

9) веб-сервера [<https://www.ibi.vu.nl/programs/linkerdbwww/>] и программы [67] для изучения выбора линкеров вакцины и оптимизации ее структуры [68,69,70];

10) настольные программные приложения для визуального представления компонентов вакцины [<https://www.snapgene.com/>], собранных в единую систему, и веб-серверы для представления вторичной и третичной структуры вакцины [71].

Объектно-ориентированное, функциональное и процедурное программирование для вычислительной вирусологии

Среди множества современных языков объектно-ориентированного программирования на практике специалиста-биоинформатика чаще всего применяются языки Питон, R и Shell.

Язык программирования Питон за последние десять лет стал популярнее, чем Перл, Руби и МатЛаб, и является основным средством анализа научных данных не только в областях биоинформатики и геномики.

Язык программирования R не так универсален, как Питон, но представляет разработчику значитель-

ный набор пакетов-модулей специфически разработанных для анализа данных и статистических вычислений. Тем не менее, его синтаксис, безусловно, сложнее, чем у Питона, объем документации каждого модуля велик, и программировать им весьма трудно.

Для целей иммуноинформатики Питон предлагает следующие модули:

1) epitopepredict [<https://github.com/dmfarrell/epitopepredict>], функции для вычисления эпитопов;

2) epitopes [<https://pypi.org/project/epitopes/>]: интерфейс Python для IEDB и других данных иммунных эпитопов;

3) EpiDope [72]: глубокая нейронная сеть для предсказания линейных В-клеточных эпитопов.

Для вычислительной вакцинологии в языке R имеются следующие модули:

1) EpitopePrediction [<http://johannes-textor.name/R/epitope-prediction-using-r.html>] для вычисления Т-эпитопов;

2) bbeaR [73]: пакет R с набором функций для профилирования эпитоп-специфических антител;

3) Repitope: прогнозирование иммуногенности эпитопа посредством *in silico* профилирования потенциала взаимодействия TCR-пептида (белка, связывающего Т-клеточный рецептор) [74].

Другие языки программирования, такие как Руби, Яваскрипт, PHP, C#, Java, C++ тоже применяются в геномике, но больше для веб-программирования или для реализации математических алгоритмов (компилируемые, более низкоуровневые языки, такие как C/C++).

Язык Shell необходим для практической биоинформатики для Unix операционных систем, таких как Linux и MacOS.

Как и язык R, Shell можно вызывать из интерфейса Питона, Java или C++, PHP.

Знание языка Shell также необходимо для запуска контейнеров Docker [75,76] и Singularity [77], а также для написания кодов конвейерной обработки данных с применением популярной библиотеки параллельных вычислений GNU Parallel [78] и ряда аналогичных.

References

1. Tomar, Marton. Immunoinformatics. – P [S.I.]: Springer US, 2020. – 409 p.
2. De, R.K., Tomar, N. Immunoinformatics. – New York: Humana Press, 2014. – P xix, 586 pages.
3. Schönbach, C., Ranganathan, S., Brusica, V. Immunoinformatics. – New York: Springer, 2008. – xix, 200 p.
4. Flower, D.R. Bioinformatics for vaccinology. – Chichester, West Sussex, England; Hoboken, NJ: John Wiley & Sons, 2008. – 314 p.
5. Prabhakar, H., Kapoor, I., Mahajan, C. Clinical synopsis of COVID-19: evolving and challenging. – 1 online resource (XIV, 260 pages).
6. Chandra, P., Roy, S. Diagnostic strategies for COVID-19 and other coronaviruses. – Singapore: Springer, 2020. – 1 online resource (viii, 199 pages).
7. Saxena, S.K. Coronavirus disease 2019 (COVID-19): epidemiology, pathogenesis, diagnosis, and therapeutics. – 1 online resource (224 pages).
8. Plotkin, S.A., Orenstein, W.A., Offit, P.A. Plotkin's vaccines. – Philadelphia, PA: Elsevier, 2018. – xxi, 1691 pages.
9. Enjuanes, L. Coronavirus replication and reverse genetics. – Berlin; New York: Springer, 2005. – vii, 257 p.
10. Varshney, D., Singh, M., SpringerLink (Online service). Lyophilized Biologics and Vaccines Modality – Based Approaches. – New York, NY: Springer New York: Imprint: Springer, 2015. – XI, 401 p. 99 illus., 68 illus. in color.
11. Nunnally, B.K., Turula, V.E., Sitrin, R.D., SpringerLink (Online service). Vaccine Analysis: Strategies, Principles, and Control. – Berlin, Heidelberg: Springer Berlin Heidelberg: Imprint: Springer, 2015. – XII, 665 p. 124 illus., 76 illus. in color.
12. Maier, H.J., Bickerton, E., Britton, P. Coronaviruses methods and protocols. – New York: Humana Press; Springer, 2015. – xi, 285 pages.
13. Kiyono, H., Ogra, P.L., McGhee, J.R. Mucosal vaccines. – San Diego: Academic Press, 1996. – xix, 479 p.
14. Vanniasinkam, T., Tikoo, S.K., Samal, S.K. Viral vectors in veterinary vaccine development: a textbook. – Cham, Switzerland: Springer, 2021. – 1 online resource (xi, 230 pages).
15. Hwang, W., Lei, W., Katritsis, N.M. et al. Current and prospective computational approaches and challenges for developing COVID-19 vaccines // *Advanced Drug Delivery Reviews*. – 2021. epub.
16. Sohail, M.S., Ahmed, S.F., Quadeer, A.A., McKay, M.R. In silico T cell epitope identification for SARS–CoV–2: Progress and perspectives // *Adv Drug Deliv Rev*. – 2021. – Vol. 171, – P. 29-47.
17. Noorimotlagh, Z., Karami, C., Mirzaee, S.A. et al. Immune and bioinformatics identification of T cell and B cell epitopes in the protein structure of SARS–CoV–2: A systematic review // *Int Immunopharmacol*. – 2020. – Vol. 86. – P. 106738.
18. Baruah, V., Bose, S. Immunoinformatics aided identification of T cell and B cell epitopes in the surface glycoprotein of 2019 nCoV // *Journal of medical virology*. – 2020. – Vol. 92, No. 5. – P. 495-500.
19. Oliveira, S.C., de Magalhães, M.T., Homan, E.J. Immunoinformatic Analysis of SARS–CoV–2 Nucleocapsid Protein and Identification of COVID–19 Vaccine Targets // *Frontiers in immunology*. – 2020. – Vol. 11. – P. 2758.
20. Feng, Y., Qiu, M., Zou, S. et al. Multi-epitope vaccine design using an immunoinformatics approach for 2019 novel coronavirus in China (SARS–CoV–2) // *BioRxiv*. – 2020. – Vol. 1, No. 1. – P. 1-12.
21. Wang, X., Xu, W., Tong, D. et al. A chimeric multi-epitope DNA vaccine elicited specific antibody response against severe acute respiratory syndrome – associated coronavirus which attenuated the virulence of SARS–CoV in vitro // *Immunology letters*. – 2008. – Vol. 119, No. 1–2. – P. 71-77.
22. Belikova Y., Samsonov Y., Abakushina E. Modern vaccines and coronavirus infections. *Research and Practical Medicine Journal*, 2020, Vol 7, No. 4. pp. 135-154. (In Russian).
23. Kononov A., Mishchenko V., Dumova B. et al. Antigenic properties of bovine coronavirus vaccine with various adjuvants. *Proceedings of the Federal Center for Animal Health*, 2009, Vol 7, pp. 50-55. (In Russian).
24. Kharchenko E. P. Coronavirus SARS-Cov-2: the complexity of pathogenesis, the search for vaccines and future pandemics. *Epidemiology and Vaccine Prevention*, 2020, Vol 19, No. 3. pp. 4-20. (In Russian).
25. Ozharovskaya T., Zubkova O., Dolzhikova I. et al. Immunogenicity of various forms of glycoprotein S of the Middle East respiratory syndrome coronavirus. *Acta Naturae (Russian version)*, 2019, Vol 11, No. 1 (40). pp. 38-47. (In Russian).
26. Chepurinov A. A., Sharshov K. A., Kazachinskaya E. I. et al. Antigenic properties of SARS-CoV-2 / human / RUS / Nsk-FRCFTM-1/202 coronavirus isolate isolated from a patient in Novosibirsk. *Journal of Infectology*, 2020, Vol 12, No. 3. pp. 42-50. (In Russian).
27. Kharchenko E.P. SARS-CoV-2 coronavirus: features of structural proteins, contagiousness and possible immune collisions. *Epidemiology and Vaccine Prevention*, 2020, Vol 19, No. 2. pp. 13-30. (In Russian).
28. Titova M.A. Approaches to modeling immunogenic peptides: author. dis. for a job. learned. step. Cand. chem. Sciences: 02.00.10 / Titova Maya Adolfovna. - M., 2003, 23 p. M., 2003. (In Russian).
29. Hu, T., Li, J., Zhou, H. et al. Bioinformatics resources for SARS–CoV–2 discovery and surveillance // *Briefings in Bioinformatics*. – 2021. – Vol. 22, No. 2. – P. 631-641.
30. Hufsky, F., Lamkiewicz, K., Almeida, A. et al. Computational strategies to combat COVID–19: useful tools to accelerate SARS–CoV–2 and coronavirus research // *Briefings in Bioinformatics*. – 2020. – Vol. 22, No. 2. – P. 642-663.
31. Kangabam, R., Sahoo, S., Ghosh, A. et al. Next– generation computational tools and resources for coronavirus research: From detection to vaccine discovery // *Computers in biology and medicine*. – 2020. – Vol. 1, No. 1. – P. 104-115.

32. Kiyotani, K., Toyoshima, Y., Nemoto, K., Nakamura, Y. Bioinformatic prediction of potential T cell epitopes for SARS–Cov–2 // *Journal of human genetics*. – 2020. – Vol. 65, No. 7. – P. 569-575.
33. Ogishi, M., Yotsuyanagi, H. Quantitative Prediction of the Landscape of T Cell Epitope Immunogenicity in Sequence Space // *Frontiers in Immunology*. – 2019. – Vol. 10, No. 827. – P. 1-20.
34. Ong, E., Wang, H., Wong, M.U. et al. Vaxign– ML: supervised machine learning reverse vaccinology model for improved prediction of bacterial protective antigens // *Bioinformatics*. – 2020. – Vol. 36, No. 10. – P. 3185-3191.
35. Xiang, Z., He, Y. Genome– wide prediction of vaccine targets for human herpes simplex viruses using Vaxign reverse vaccinology // *BMC Bioinformatics*. – 2013. – Vol. 14, No.1. – P. S2.
36. He, Y., Xiang, Z., Mobley, H.L. Vaxign: the first web– based vaccine design program for reverse vaccinology and applications for vaccine development // *J Biomed Biotechnol*. – 2010. – Vol. 10, No. Epub 2010 Jul 4. – P. 297505.
37. Vita, R., Overton, J.A., Greenbaum, J.A. et al. The immune epitope database (IEDB) 3.0 // *Nucleic Acids Res*. –Vol. 43, No. 1. – P. D405-12.
38. El–Manzalawy, Y., Dobbs, D., Honavar, V. Predicting linear B–cell epitopes using string kernels // *J Mol Recognit*. – 2008. – Vol. 21, No. 4. – P. 243-55.
39. Seemann, T. Prokka: rapid prokaryotic genome annotation // *Bioinformatics*. – 2014. – Vol. 30, No. 14. – P. 2068-9.
40. Huerta-Cepas, J., Szklarczyk, D., Heller, D. et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses // *Nucleic Acids Res*. – 2018. – Vol. 47, No. D1. – P. D309-D314.
41. Jones, P., Binns, D., Chang, H.Y. et al. InterProScan 5: genome–scale protein function classification // *Bioinformatics*. – 2014. – Vol. 30, No. 9. – P. 1236-40.
42. Mulder, N., Apweiler, R. InterPro and InterProScan: tools for protein sequence classification and comparison // *Methods Mol Biol*. – 2007. – Vol.396. – P. 59-70.
43. Syed, A., Upton, C. Java GUI for InterProScan (JIPS): a tool to help process multiple InterProScans and perform ortholog analysis // *BMC Bioinformatics*. – 2006. – Vol. 7. – P. 462.
44. Quevillon, E., Silventoinen, V., Pillai, S. et al. InterProScan: protein domains identifier // *Nucleic Acids Res*. – 2005. – Vol. 33, No. 1. – P. W116- 20.
45. Katoh, K., Standley, D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability // *Mol Biol Evol*. – 2013. – Vol. 30, No. 4. – P. 772-80.
46. Katoh, K., Frith, M.C. Adding unaligned sequences into an existing alignment using MAFFT and LAST // *Bioinformatics*. – 2012. – Vol. 28, No. 23. – P. 3144-6.
47. Katoh, K., Toh, H. Parallelization of the MAFFT multiple sequence alignment program // *Bioinformatics*. – 2010. – Vol. 26, No. 15. – P. 1899-900.
48. Katoh, K., Asimenos, G., Toh, H. Multiple alignment of DNA sequences with MAFFT // *Methods Mol Biol*. – 2009. – Vol. 537. – P. 39-64.
49. Hung, C.L., Lin, Y.S., Lin, C.Y. et al. CUDA ClustalW: An efficient parallel algorithm for progressive multiple sequence alignment on Multi– GPUs // *Comput Biol Chem*. – Vol. 58. – P. 62-8.
50. Hung, J.H., Weng, Z. Sequence Alignment and Homology Search with BLAST and ClustalW // *Cold Spring Harb Protoc*. – 2016. – Vol. 16, No. 11. – P. 1-10.
51. Vangala, R.K., Singh, L., Gupta, R.P. BioParishodhana: A novel graphical interface integrating BLAST, ClustalW, primer3 and restriction digestion tools // *Bioinformation*. – 2012. – Vol. 8, No. 13. – P. 639-43.
52. Zaal, D., Nota, B. ADOMA: A Command Line Tool to Modify ClustalW Multiple Alignment Output // *Mol Inform*. – 2016. – Vol. 35, No. 1. – P. 42-4.
53. Darling, A.E., Treangen, T.J., Messeguer, X., Perna, N.T. Analyzing patterns of microbial evolution using the mauve genome alignment system // *Methods Mol Biol*. – 2007. – Vol. 396. – P. 135-52.
54. Darling, A.C., Mau, B., Blattner, F.R., Perna, N.T. Mauve: multiple alignment of conserved genomic sequence with rearrangements // *Genome Res*. – 2004. – Vol. 14, No. 7. – P. 1394-403.
55. Angiuoli, S.V., Salzberg, S.L. Mugsy: fast multiple alignment of closely related whole genomes // *Bioinformatics*. – 2010. – Vol. 27, No. 3. – P. 33-42.
56. Cleemput, S., Dumon, W., Fonseca, V. et al. Genome Detective Coronavirus Typing Tool for rapid identification and characterization of novel coronavirus genomes // *Bioinformatics*. – Vol. 36, No. 11. – P. 3552-3555.
57. Vilsker, M., Moosa, Y., Nooij, S. et al. Genome Detective: an automated system for virus identification from high–throughput sequencing data // *Bioinformatics*. – 2018. – Vol. 35, No. 5. – P. 871-873.
58. Gupta, S., Kapoor, P., Chaudhary, K. et al. Peptide toxicity prediction // *Methods Mol Biol*. – Vol. 1268. – P. 143-57.
59. Gupta, S., Kapoor, P., Chaudhary, K. et al. In silico approach for predicting toxicity of peptides and proteins // *PLoS One*. – 2013. – Vol. 8, No. 9. – P. e. 73957.
60. Krutz, N.L., Winget, J., Ryan, C.A. et al. Proteomic and Bioinformatic Analyses for the Identification of Proteins With Low Allergenic Potential for Hazard Assessment // *Toxicol Sci*. – Vol. 170, No. 1. – P. 210-222.
61. Maurer–Stroh, S., Krutz, N.L., Kern, P.S. et al. AllerCatPro– prediction of protein allergenicity potential from the protein sequence // *Bioinformatics*. – 2019. – Vol. 35, No. 17. – P. 3020-3027.
62. Sharma, N., Patiyal, S., Dhall, A. et al. AlgPred 2.0: an improved method for predicting allergenic proteins and mapping

- of IgE epitopes // *Brief Bioinform.* – 2020. – Vol. 1, № 1. – P. 1-7.
63. Neeharika, D., Sunkar, S. Computational approach for the identification of putative allergens from Cucurbitaceae family members // *J Food Sci Technol.* – 2021. – Vol 58, No. 1. – P. 267-280.
64. Sircar, G., Saha, B., Bhattacharya, S.G., Saha, S. In silico prediction of allergenic proteins // *Methods Mol Biol.* – 2014. – Vol. 1184. – P. 375-88.
65. Saha, S., Raghava, G.P. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes // *Nucleic Acids Res.* – 2006. – Vol. 34, No. Web Server issue. – P. W. 202-9.
66. Pagadala, N.S., Syed, K., Tuszynski, J. Software for molecular docking: a review // *Biophys Rev.* – 2017. – Vol. 9, No. 2. – P. 91-102.
67. Grote, A., Hiller, K., Scheer, M. et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host // *Nucleic Acids Res.* – 2005. – Vol. 33, No. Web Server issue. – P. W526-31.
68. Zheng, W., Zhang, C., Bell, E.W., Zhang, Y. I – TASSER gateway: A protein structure and function prediction server powered by XSEDE // *Future Gener Comput Syst.* – 2019. – Vol. 99, – P. 73-85.
69. Roy, A., Kucukural, A., Zhang, Y. I – TASSER: a unified platform for automated protein structure and function prediction // *Nat Protoc.* – 2010. – Vol. 5, No. 4. – P. 725-38.
70. Zhang, Y. I – TASSER server for protein 3D structure prediction // *BMC Bioinformatics.* – 2008. – Vol. 9. – P. 40.
71. Raborn, R.T., Brendel, V.P. Using RAMPAGE to Identify and Annotate Promoters in Insect Genomes // *Methods Mol Biol.* – 2019. – Vol. 1858, No. 1. – P. 99-116.
72. Collatz, M., Mock, F., Hölzer, M. et al. EpiDope: A Deep neural network for linear B-cell epitope prediction // *bioRxiv.* – 2020. – No. 1. – P. 1-8.
73. Suprun, M., Ellis, R.J., Sampson, H.A., Suárez- Fariñas, M. bbeaR: an R package and framework for epitope – specific antibody profiling // *Bioinformatics.* – 2021. – Vol. 37, No. 1. – P. 131-133.
74. Ogishi, M., Yotsuyanagi, H. Quantitative prediction of the landscape of T cell epitope immunogenicity in sequence space // *Frontiers in Immunology.* – 2019. – Vol .10. – P. 827.
75. Pittard, W.S., Li, S. The Essential Toolbox of Data Science: Python, R, Git, and Docker // *Methods Mol Biol.* – 2020. – Vol. 2104, No. 1. – P. 265-311.
76. Kwon, C., Kim, J., Ahn, J. DockerBIO: web application for efficient use of bioinformatics Docker images // *PeerJ.* – 2018. – Vol. 6, No. 1. – P. e. 5954.
77. Garofoli, A., Paradiso, V., Montazeri, H. et al. PipeIT: A Singularity Container for Molecular Diagnostic Somatic Variant Calling on the Ion Torrent Next- Generation Sequencing Platform // *J. Mol Diagn.* – 2019. – Vol. 21, No. 5. – P. 884-894.
78. Samdani, A., Vetrivel, U. POAP: A GNU parallel based multithreaded pipeline of open babel and AutoDock suite for boosted high throughput virtual screening // *Comput Biol Chem.* – 2018. – Vol. 74, No. 1. – P. 39-48.
79. Красильников, А.П. Микробиологический словарь-справочник. – 2-е изд., доп. и перераб. – Мн.: ООО "Асар", 1999. – 397 с.

Received: 19.06.2021

Поступила: 19.06.2021