



ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

DOI: <https://doi.org/10.36233/0507-4088-179>

© АСАТРИАН М.Н., ТИМОФЕЕВ Б.И., ШМЫР И.С., ХАЧАТРИАН К.Р., ЩЕРБИНИН Д.Н., ТИМОФЕЕВА Т.А., ГЕРАСИМУК Э.Р., АГАСАРИАН В.Г., ЕРШОВ И.Ф., ШАШКОВА Т.И., ИВАНИСЕНКО Н.В., КАРДЫМОН О.Л., СЕМЕНЕНКО Т.А., НАРОДИЦКИЙ Б.С., ЛОГУНОВ Д.Ю., ГИНЦБУРГ А.Л., 2023

Математическая модель для оценки уровня перекрёстного иммунитета между штаммами вируса гриппа подтипа H_3N_2

Асатрян М.Н.¹, Тимофеев Б.И.¹, Шмыр И.С.¹, Хачатрян К.Р.², Щербинин Д.Н.¹, Тимофеева Т.А.¹, Герасимук Э.Р.³, Агасарян В.Г.¹, Ершов И.Ф.¹, Шашкова Т.И.⁴, Иванисенко Н.В.⁴, Кардымон О.Л.⁴, Семенов Т.А.¹, Народицкий Б.С.¹, Логунов Д.Ю.¹, Гинцбург А.Л.¹

¹ФГБУ «Национальный исследовательский центр эпидемиологии и микробиологии имени почетного академика Н.Ф. Гамалеи», 123098, г. Москва, Россия;

²ФГАОУ ВО Национальный исследовательский университет «Высшая школа экономики», 123458, г. Москва, Россия;

³Государственный университет «Дубна», 141982, г. Дубна, Россия;

⁴Институт искусственного интеллекта, 121170, г. Москва, Россия

Введение. Всемирная организация здравоохранения (ВОЗ) регулярно обновляет рекомендации по вакцинам против гриппа с целью достижения их максимального соответствия очередным циркулирующим штаммам. Тем не менее на протяжении нескольких сезонов эффективность вакцины против гриппа А, а именно её компоненты H_3N_2 , определялась как низкая.

Цель исследования – разработка математической модели перекрёстного иммунитета на основании имеющегося массива опубликованных ВОЗ данных реакции торможения гемагглютинации (РТГА).

Материалы и методы. В настоящей работе представлена математическая модель, основанная на нахождении с помощью регрессионного анализа зависимости титров РТГА от замен в антигенных сайтах последовательностей. Разработанная нами компьютерная программа имеет возможность обрабатывать данные (GISAID, NCBI и др.) и формировать в режиме реального времени базы данных согласно поставленным задачам.

Результаты. На основе наших исследований был вычленен дополнительный антигенный сайт F. Разница в 1,6 раза скорректированного R^2 на подмножествах вирусов, выращенных в культуре клеток и культивируемых в куриных эмбрионах, демонстрирует обоснованность нашего решения о разделении первоначального массива данных по пассажным историям. Нами введено понятие степени гомологичности между двумя произвольными штаммами, которая принимает значение функции, зависящей от дистанции Хэмминга, и показано, что результаты регрессии существенно зависят от выбора функции. Проведённый анализ показал, что наиболее значимыми антигенными сайтами являются А, В и Е. Полученные результаты прогноза титров РТГА показали достаточно хороший результат, сопоставимый с аналогичными работами наших коллег.

Заключение. Предложенный метод может послужить хорошим инструментом для будущих прогнозов с дальнейшим изучением для подтверждения его устойчивости.

Ключевые слова: вирус гриппа; подтип H_3N_2 ; титры РТГА; перекрёстный иммунитет; антигенное расхождение; антигенный сайт; дистанция Хэмминга; регрессионный анализ; эпидемиологическая модель; иммунный ландшафт; вакцинный штамм

Для цитирования: Асатрян М.Н., Тимофеев Б.И., Шмыр И.С., Хачатрян К.Р., Щербинин Д.Н., Тимофеева Т.А., Герасимук Э.Р., Агасарян В.Г., Ершов И.Ф., Шашкова Т.И., Иванисенко Н.В., Кардымон О.Л., Семенов Т.А., Народицкий Б.С., Логунов Д.Ю., Гинцбург А.Л. Математическая модель для оценки уровня перекрёстного иммунитета между штаммами вируса гриппа подтипа H_3N_2 . *Вопросы вирусологии*. 2023; 68(3): 252-264. DOI: <https://doi.org/10.36233/0507-4088-179> EDN: <https://elibrary.ru/rqxvea>

Для корреспонденции: Асатрян Марина Норайровна, канд. мед. наук, старший научный сотрудник группы эпидемиологической кибернетики отдела эпидемиологии ФГБУ «Национальный исследовательский центр эпидемиологии и микробиологии имени почетного академика Н.Ф. Гамалеи», 123098, г. Москва, Россия. E-mail: masatryan@gamaleya.org

Участие авторов: все авторы сделали эквивалентный вклад в подготовку публикации.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила 12.05.2023

Принята в печать 22.06.2023

Опубликована 30.06.2023

ORIGINAL STUDY ARTICLE

DOI: <https://doi.org/10.36233/0507-4088-179>

Mathematical model for assessing the level of cross-immunity between strains of influenza virus subtype H₃N₂

Marina N. Asatryan¹, Boris I. Timofeev¹, Ilya S. Shmyr¹, Karlen R. Khachatryan², Dmitrii N. Shcherbinin¹, Tatiana A. Timofeeva¹, Elita R. Gerasimuk³, Vaagn G. Agasaryan¹, Ivan F. Ershov¹, Tatyana I. Shashkova⁴, Nikita V. Ivanisenko⁴, Olga L. Kardymon⁴, Tatyana A. Semenenko¹, Boris S. Naroditsky¹, Denis Yu. Logunov¹, Aleksander L. Gintsburg¹

¹National Research Center for Epidemiology and Microbiology named after Honorary Academician N.F. Gamaleya, 123098, Moscow, Russia;

²National Research University Higher School of Economics, 123458, Moscow, Russia;

³State University "Dubna", 141982, Dubna, Russia;

⁴Artificial Intelligence Research Institute, 121170, Moscow, Russia

Introduction. The WHO regularly updates influenza vaccine recommendations to maximize their match with circulating strains. Nevertheless, the effectiveness of the influenza A vaccine, specifically its H₃N₂ component, has been low for several seasons.

The aim of the study is to develop a mathematical model of cross-immunity based on the array of published WHO hemagglutination inhibition assay (HAI) data.

Materials and methods. In this study, a mathematical model was proposed, based on finding, using regression analysis, the dependence of HAI titers on substitutions in antigenic sites of sequences. The computer program we developed can process data (GISAID, NCBI, etc.) and create "real-time" databases according to the set tasks.

Results. Based on our research, an additional antigenic site F was identified. The difference in 1.6 times the adjusted R², on subsets of viruses grown in cell culture and grown in chicken embryos, demonstrates the validity of our decision to divide the original data array by passage histories. We have introduced the concept of a degree of homology between two arbitrary strains, which takes the value of a function depending on the Hamming distance, and it has been shown that the regression results significantly depend on the choice of function. The provided analysis showed that the most significant antigenic sites are A, B, and E. The obtained results on predicted HAI titers showed a good enough result, comparable to similar work by our colleagues.

Conclusion. The proposed method could serve as a useful tool for future forecasts, with further study to confirm its sustainability.

Keywords: influenza virus; subtype H₃N₂; HAI titers; cross-immunity; antigenic distance; antigenic site; Hamming distance; regression analysis; epidemiological model; immune landscape; vaccine strain

For citation: Asatryan M.N., Timofeev B.I., Shmyr I.S., Khachatryan K.R., Shcherbinin D.N., Timofeeva T.A., Gerasimuk E.R., Agasaryan V.G., Ershov I.F., Shashkova T.I., Ivanisenko N.V., Kardymon O.L., Semenenko T.A., Naroditsky B.S., Logunov D.Yu., Gintsburg A.L. Mathematical model for assessing the level of cross-immunity between strains of influenza virus subtype H₃N₂. *Problems of Virology (Voprosy Virusologii)*. 2023; 68(3): 252-264. DOI: <https://doi.org/10.36233/0507-4088-179> EDN: <https://elibrary.ru/rexvea>

For correspondence: Marina N. Asatryan, PhD (Med.), Senior Researcher, Epidemiological Cybernetics Group of the Epidemiology Department, National Research Center for Epidemiology and Microbiology named after Honorary Academician N.F. Gamaleya, 123098, Moscow, Russia. E-mail: masatryan@gamaleya.org

Information about the authors:

Marina N. Asatryan, <https://orcid.org/0000-0001-6273-8615>

Boris I. Timofeev, <https://orcid.org/0000-0001-7425-0457>

Ilya S. Shmyr, <https://orcid.org/0000-0002-8514-5174>

Karlen R. Khachatryan, <https://orcid.org/0000-0002-1934-532X>

Dmitrii N. Shcherbinin, <https://orcid.org/0000-0002-8518-1669>

Tatiana A. Timofeeva, <https://orcid.org/0000-0002-8991-8525>

Elita R. Gerasimuk, <https://orcid.org/0000-0002-7364-163X>

Vaagn G. Agasaryan, <https://orcid.org/0009-0009-3824-7061>

Ivan F. Ershov, <https://orcid.org/0000-0002-3333-5347>

Tatyana I. Shashkova, <https://orcid.org/0000-0002-8754-8727>

Nikita V. Ivanisenko, <https://orcid.org/0000-0002-0333-8117>

Olga L. Kardymon, <https://orcid.org/0000-0002-4827-8891>

Tatyana A. Semenenko, <https://orcid.org/0000-0002-6686-9011>

Boris S. Naroditsky, <https://orcid.org/0000-0001-5522-8238>

Denis Yu. Logunov, <https://orcid.org/0000-0003-4035-6581>

Aleksander L. Gintsburg, <https://orcid.org/0000-0003-1769-5059>

Contribution: the authors contributed equally to this article.

Conflict of interest. The authors declare no conflicts of interest.

Received 12 May 2023

Accepted 22 June 2023

Published 30 June 2023

Введение

Глобальная сеть эпидемиологического надзора за гриппом и ответных мер Всемирной организации здравоохранения (ВОЗ) WHO GISRS (The World Health Organization Global Influenza Surveillance and Response System) отслеживает и анализирует эволюцию и эпидемиологию вирусов гриппа с основной целью – выбор вакцинного штамма и совершенствование этого процесса посредством исследований, направленных на лучшее понимание эволюционной изменчивости, факторов распространения в сочетании с иммунологическим ландшафтом населения и перекрёстным иммунитетом [1]. При этом ВОЗ регулярно обновляет рекомендации по вакцинам против гриппа с целью достижения их максимального соответствия очередным циркулирующим штаммам.

Тем не менее на протяжении нескольких сезонов эффективность вакцины против гриппа А, а именно её компоненты H_3N_2 , определялась как низкая по сравнению с другими штаммами [2–5]. Причиной низкой эффективности может быть несколько факторов. Например, характерные адаптационные изменения при пассировании рекомендованного штамма в куриных эмбрионах в процессе производства вакцин [6]. Поскольку разработка, масштабное производство и распределение вакцины занимают много месяцев, в конечном счёте к предстоящему сезону преобладание циркулирующих штаммов в некоторых сезонах претерпевает существенные изменения [7]. Но даже если рекомендованная вакцина соответствовала циркулирующим штаммам, на её эффективности могло отрицательно сказаться влияние существующего иммунного ландшафта [8]. Поэтому прогнозирование эволюционной изменчивости вируса гриппа до сих пор представляет большой интерес для общественного здравоохранения [9–12].

Наиболее перспективным направлением в этой области является построение компьютерных моделей, с помощью которых можно комбинировать различные подходы к моделированию, использовать многочисленные источники данных с возможностью интерпретировать результаты для рекомендаций при выборе вакцинного штамма. Для этого необходимо тесное сотрудничество учёных из разных областей и направлений, работающих на всех уровнях эпидемиологического надзора и отбора вакцинных штаммов, а также разработчиков моделей, эпидемиологов и клиницистов [10, 13, 14].

Коллективом Национального исследовательского центра эпидемиологии и микробиологии имени почетного академика Н.Ф. Гамалеи в 2020 г. была разработана и успешно зарегистрирована (свидетельство о регистрации № 2020617965 от 15 июля 2020 г.) **компьютерная программа Influenza IDE** – эпидемиологическая модель (ЭМ) по континентам с упрощённой моделью перекрёстного иммунитета и постоянно обновляющейся базой данных (различных видов и подтипов вируса гриппа) **Influenza DB**. Важной особенностью ЭМ является возможность формирования

(учёта) **нулевого иммунного ландшафта** популяции и далее, после пошаговой имитации (симуляции) распространения вируса гриппа, получение иммунного ландшафта на первый день следующего сезона. ЭМ распространения вируса гриппа среди населения мира на протяжении нескольких сезонов, разработанная с помощью агентного подхода, представлена в виде вложенных моделей: модели **поведения популяции**, модели **инфекционного процесса** и модели **инфицирования** (на основании иммунного ответа в организме отдельного агента (индивидуума) с учётом иммунной памяти и **модели перекрёстного иммунитета**). Компьютерная программа спроектирована с возможностью интегрирования различных моделей перекрёстного иммунитета.

Значительное количество работ по изучению перекрёстного иммунитета основано на обнаружении зависимости между титрами РТГА (реакция торможения гемагглютинации) и различиями в генетических последовательностях вирусов. Для нахождения указанной зависимости исследователи применяют различные математические методы, в том числе регрессионный анализ [15–18]. При этом в качестве меры перекрёстного иммунитета используются различные функции от титров РТГА. Так, в работе F.M. Burnet и D. Lush [19] была введена функция, обозначенная как индикатор эффективности вакцины

$$R_{ij} = c_{ij} / c_{ii}$$

где c_{ij} – значение концентрации (разведения) сыворотки к вирусу i в РТГА с вирусом j , c_{ii} – значение концентрации (разведения) сыворотки i в РТГА с собственным вирусом, где концентрация (разведение) сыворотки в РТГА является величиной, обратной титру РТГА. Исследователи I. Archetti и F.L. Horsfall [20] в своей работе в качестве меры антигенной изменчивости вводят среднее геометрическое значение вышеуказанных соотношений $(R_{ij}R_{ji})^{1/2}$. В более поздних работах, положивших начало широко используемой в настоящее время антигенной картографии, A. Lapedes и R. Farber [21], D.J. Smith и соавт. [22] продемонстрировали, что можно построить пространство формы малой размерности, в которой антитела и антигены рассматриваются как точки, а расстояние между ними обозначается как антигенная дистанция. В качестве меры дистанции D_{ij} использовался как логарифм от титров РТГА по основанию 2, так и величины $\log_2(R_{ij})$, $0,5 \times \log_2(R_{ij}R_{ji})$ [15–17].

В качестве меры различия между последовательностями обычно используется либо простая дистанция Хэмминга, либо функция, зависящая от неё. Наиболее часто в рассматриваемых работах используется модель объединения аминокислотных позиций в антигенные сайты [23–25]. В таких случаях рассматривается дистанция Хэмминга между антигенными сайтами [16–18]. Но бывают и сложные случаи, когда в качестве меры различия между антигенными сайтами выступают физико-химические факторы различия аминокислотных позиций, например гликозилирование [18, 26]. В качестве материалов для построения моделей использовались ограниченные массивы дан-

ных, включающие только референсные штаммы, что снижает объективность. В данной работе мы хотим представить регрессионную модель перекрёстного иммунитета на основании всего имеющегося массива опубликованных ВОЗ данных РТГА, что позволит улучшить точность модели, повысить её объективность и возможность проверки.

Применение адекватной ЭМ с верифицированной моделью перекрёстного иммунитета позволит усовершенствовать процесс подбора необходимых вакцинных штаммов для более успешной борьбы с вирусом гриппа.

Цель исследования – разработка математической модели перекрёстного иммунитета на основании имеющегося массива опубликованных ВОЗ ежесезонных данных серологического тестирования (РТГА).

Материалы и методы

При формировании массива данных **Influenza DB (database)** была использована информация из опубликованных ежесезонных данных ВОЗ по результатам тестирования РТГА и данные (последовательности + сопроводительная информация) из платформы GISAID (Global Initiative on Sharing All Influenza Data). В компьютерной программе **Influenza IDE** [27] также предусмотрен дополнительный модуль **Influenza DP (data processing)**, который представляет собой универсальный обработчик баз данных (GISAID, NCBI и др.).

Для изучения закономерностей в модели перекрёстного иммунитета использовался множественный регрессионный анализ (линейная регрессия) с использованием метода наименьших квадратов (МНК) и метода неотрицательного МНК [28] для оценки параметров регрессии, представленный формулой (1):

$$\text{Modified titre} = c_0 + c_1 \times As_1 + c_2 \times As_2 + c_3 \times As_3 + c_4 \times As_4 + c_5 \times As_5 + c_6 \times As_6, (1)$$

где в качестве значения функции представлены модифицированные титры РТГА, в качестве аргументов $As_1, As_2, As_3, As_4, As_5, As_6$ – значения степени гомологичности антигенных сайтов (Antigenic sites – As) A, B, C, D, E, F соответственно, а $c_0, c_1, c_2, c_3, c_4, c_5, c_6$ – параметры (коэффициенты) модели.

Для аргументов регрессии мы основывались на собственном методе определения антигенных сайтов. Термин «антигенный сайт» был введён Gerhard и Webster в 1978 г., чтобы описать специфические моноклональные антитела. Антитела, которые конкурировали между собой, рассматривались как связывающие один и тот же антигенный сайт. Каждый антигенный сайт может содержать один эпитоп или более – различные наборы аминокислот на антигене, которые контактируют с аминокислотами антител. Конкуренция между антителами, которые связывают тот же самый антигенный сайт, предполагает, что эпитопы в данном сайте физически перекрываются, но могут быть различными и одна молекула антитела экранирует целый антигенный сайт.

Также для учёта различий между штаммами по антигенным сайтам мы ввели понятие **степени гомологичности** и определили её для двух произвольно взятых штаммов как убывающую функцию от дистанции Хэмминга, принимающую значения от 1 до 0. Обоснование и подбор используемых функций подробно описаны в разделе «Результаты». Для обучения регрессионной модели использовался пакет Stats версии 4.0.3 языка программирования R. Для предварительной обработки данных использовался пакет Pandas версии 1.4.2 языка программирования Python. Анализ стабильности прогностической способности модели перекрёстного иммунитета проводился на ретроспективных данных. В качестве меры адекватности модели мы использовали коэффициент детерминации R^2 [29], а мерой точности прогноза – показатель воспроизводимости титров на одно (± 1) и (или) два (± 2) разведения. Мы опирались на многочисленные исследования, в которых при сравнении результатов РТГА для каждого образца внутри одной лаборатории или между несколькими титры считались эквивалентными, если они различались не более чем на одно разведение (т.е. в 2 раза) [30–32].

Формирование массива данных *Influenza DB* для изучения модели перекрёстного иммунитета

Подготовка экспериментальных данных РТГА

Отчёты ВОЗ по гриппу публикуются ежесезонно начиная с 2005 г. и находятся в открытом доступе [33]. Представленные титры антител в РТГА выявляют антигенные свойства референсных и тестовых штаммов на основе перекрёстной реактивности антител хорьков, полученных к референсным штаммам, тормозящих агглютинацию эритроцитов морских свинок тестируемыми вирусами гриппа. В связи с тем что ВОЗ в 2008–2009 гг. обнаружила заметное влияние нейраминидазы и различных источников эритроцитов (индейка, человек, морская свинка) на результаты РТГА, в дальнейших исследованиях использовались и обрабатывались только таблицы РТГА с 2009 по 2022 г., выполненные на эритроцитах морской свинки с добавлением 20nM Oseltamivir, который используется для исключения влияния нейраминидазы. Несмотря на предложенную международную кодировку вирусов, исследователи со всего мира встречаются с неточными аннотациями при депонировании последовательностей в общедоступные базы данных [34]. Поэтому отдельной задачей в нашей работе была обозначена процедура обработки и подготовки опубликованных таблиц РТГА для дальнейшего использования. Также в результате подготовки каждой последовательности присваивался уникальный идентификатор.

В качестве меры перекрёстного иммунитета в нашей модели выступают модифицированные титры РТГА. Для этого значения опубликованных титров РТГА преобразовываются на логарифм от титра по основанию 2. Далее проводится усреднение данных по алгоритму: подмножества объединяются в соответствии с аналогичными парами идентификаторов штаммов

(референсных и тестовых); в каждом полученном подмножестве вычисляется среднее арифметическое значение модифицированных титров по одинаковым парам идентификаторов штаммов.

Отдельно хочется отметить, что культивирование (пассажи) вирусов, выполненное в куриных эмбрионах, может содержать характерные адаптационные замены, меняющие рецепторную специфичность вирусов и характер гликозилирования [6, 35] и, как следствие, влияя на перекрёстную реактивность в РТГА. Поэтому в результате анализа первичных данных нами было принято решение выделить в первоначальном массиве данных РТГА несколько подмножеств:

– с пассажной историей в куриных эмбрионах (Egg-Egg 4226);

– с пассажной историей в культуре клеток (Cell-Cell 28 621);

– со смешанной пассажной историей (Egg_s-Cell_t 16 463): референсные вирусы (для получения контрольных антисывороток), культивированные в куриных эмбрионах в сочетании с тестовыми вирусами, выращенными в клеточной культуре;

– со смешанной пассажной историей (Cell_s-Egg_t 5032): референсные вирусы (для получения контрольных антисывороток), выращенные в клеточной культуре в сочетании с тестовыми вирусами, культивированными в куриных эмбрионах.

Наше решение опиралось на многочисленные проведённые сравнительные исследования антигенных мутаций вируса гриппа H_3N_2 при культивировании в куриных эмбрионах и клеточных пассажах [36–38]. Учитывая, что количество данных с пассажной историей Cell-Cell и Egg_s-Cell_t в несколько раз превышает размерность подмножеств с пассажными историями Egg-Egg и Cell_s-Egg_t, регрессионный анализ мы проводили на данных первых двух подмножеств.

Подготовка данных из платформы GISAID

Международная платформа GISAID была запущена в 2008 г. и с тех пор предлагает надёжный механизм обмена всеми генетическими и связанными с гриппом данными для исследователей, учёных и работников здравоохранения [39]. Как и в случае данных РТГА, при депонировании последовательностей (при вводе данных вручную и т.д.) часто происходит искажение формата скачиваемых данных и их неверная классификация. Это в свою очередь затрудняет анализ и обработку имеющейся информации для усовершенствования подходов (методов) к подбору вакцинных штаммов. Поэтому для дальнейших исследований с помощью компьютерной программы **Influenza IDE** и встроенного в неё универсального обработчика **Influenza DP** мы провели тщательную очистку и согласование доступных данных. После завершения процесса верификации данных аминокислотные последовательности были выравнены на соответствующую референсную последовательность: H_3N_2 : A/Aichi/2/1968, 566 AA, включая сигнальный пептид, с помощью оригинального быстрого лексического алгоритма (максимальное совпадение аминокислот референс-последовательно-

сти и образца с минимизацией числа позиций делеций и вставок) или методом Смита–Ватермана [40]. Образцы с отличиями более 20% от референса отбрасывались. Краткие статистические характеристики данных представлены в **табл. 1**.

Результаты

Определение антигенных сайтов гемагглютинина вируса гриппа H_3N_2

Объединение аминокислотных последовательностей в антигенные сайты проводилось согласно собственному методу с вычленением дополнительного антигенного сайта F. Чтобы определить, к какому именно антигенному сайту вируса гриппа относится конкретная позиция, мы в первую очередь ориентировались на экспериментальные исследования [24]. В дальнейшем мы расширили количество позиций в антигенных сайтах, исследовав вариабельность каждой позиции гемагглютинина начиная с 1968 г. В полученном на первом этапе массиве сиквенсов проводился подсчёт встречаемости каждой аминокислоты в каждой позиции. Параллельно проводился анализ на способность той или иной аминокислоты гемагглютинина взаимодействовать с антителами, определяя погружённость или, наоборот, экспонирование аминокислот на поверхности гемагглютинина. При работе использовали программное обеспечение GETAREA [41], на вход которой подавались третичные структуры молекул гемагглютинина из базы данных PDB (Protein Data Bank) [42]. В результате на основании вариабельности позиций, их экспонирования в третичной структуре белка на поверхности, а также с учётом их максимальной близости к существующим антигенным сайтам были определены вариабельные экспонированные позиции, которые были отнесены к существующим антигенным сайтам. Более того, помимо широко известных антигенных сайтов A, B, C, D и E, нами был вычленен ещё один антигенный сайт, названный соответственно F. Этот антигенный сайт состоит из множества схожих эпитопов в стеблевой части молекулы гемагглютинина. В базе данных третичных структур содержится более 20 комплексов гемагглютининов с моноклональными антителами к данному сайту. Детально об этом сайте изложено в исследовании, проведённом Д.Н. Щербининым и соавт. [43].

Таким образом, для гемагглютинина субтипа H_3 было определено 6 антигенных сайтов. Ниже представлен список аминокислотных позиций, составляющих эти антигенные сайты, всего 109 позиций (**табл. 2**). Нумерация приведена по зрелому гемагглютинуину H_3 [44]. Следует отметить, что антигенный сайт F, в отличие от сайтов A–E, расположен в основном в HA_2 -субъединице гемагглютинина.

Полученная информация была загружена в базу данных **Influenza DB** и послужила лекалом для объединения в антигенные сайты имеющегося массива с выровненными аминокислотными последовательностями. Далее массив с антигенными последовательностями был расширен путём добавления в него информации

Таблица 1. Статистическая характеристика данных

Table 1. Data statistical characteristic

Показатель / Parameter		Значение / Value	
		Cell-Cell	Egg_s-Cell_t
Общее количество наблюдений Total number of observations		28 621	16 463
Общее количество пар штаммов Total number of pairs of strains		21 394	12 444
Количество наблюдений в отчетном периоде Number of observations in the reporting period	2009	21	35
	2010	46	94
	2011	132	461
	2012	2142	1761
	2013	1738	2447
	2014	2891	2201
	2015	1907	1520
	2016	1209	1003
	2017	1948	1279
	2018	659	164
	2019	2236	950
	2020	2496	1064
2021	902	281	
2022	10294	3203	
Значение титра / Titer value, Me [Q ₁ ; Q ₃]		160 [40; 320]	160 [40; 320]
Распределение по значениям титра Distribution by titer values	20	3404 (11,9%)	1858 (11,3%)
	40	4154 (14,5%)	2514 (15,3%)
	80	4232 (14,8%)	3852 (23,4%)
	160	6328 (22,1%)	3794 (23%)
	320	6170 (21,6%)	2446 (14,9%)
	640	3491 (12,2%)	1410 (8,6%)
	1280	772 (2,7%)	511 (3,1%)
	2560	60 (0,2%)	75 (0,5%)
5120	10 (< 0,1%)	3 (< 0,1%)	
Дистанция Хэмминга по антигенным сайтам Hamming distance across antigenic sites	A	2 [0; 4]	2 [0; 3]
	B	2 [1; 4]	3 [2; 4]
	C	1 [0; 2]	1 [0; 2]
	D	1 [0; 1]	1 [1; 2]
	E	1 [0; 2]	1 [0; 1]
	F	1 [0; 2]	1 [0; 2]

о содержании количества аминокислотных замен между любыми двумя штаммами в каждом из 6 антигенных сайтов. В результате вышеприведённых действий мы получили по каждому из 6 антигенных сайтов для произвольно взятой или отдельной (каждой) пары штаммов как значение титров РТГА, так и количество аминокислотных замен (расстояние Хэмминга). При этом замены в антигенных сайтах вне зависимости от типа и конкретной позиции приняты равнозначными. В нашей модели последовательности, не имеющие

замены в антигенных сайтах, но имеющие изменения в других позициях, считались антигенно одинаковыми.

Разработка математической модели перекрестного иммунитета вируса гриппа

Анализ и подбор функции для оценки степени гомологичности

Для проведения регрессионного анализа были выбраны подмножества с наибольшим количеством на-

Таблица 2. Аминокислотные позиции в антигенных сайтах

Table 2. Amino acid positions at antigenic sites

Сайт Site	Номера аминокислотных позиций Amino acid position numbers
A	121, 122, 124, 126, 131, 133, 135, 137, 138, 140, 142, 143, 144, 145, 146
B	128, 155, 156, 157, 158, 159, 160, 163, 186, 188, 189, 190, 192, 193, 194, 196, 197, 198, 199
C	45, 47, 48, 49, 50, 53, 54, 271, 273, 275, 276, 278, 280, 304, 311, 312
D	171, 172, 173, 201, 202, 207, 208, 212, 213, 214, 217, 219, 220, 221, 222, 223, 225, 226, 227, 229, 230, 242, 244, 246, 248
E	57, 58, 62, 63, 75, 78, 82, 83, 88, 92, 94, 96, 260, 261, 262
F	25, 31, 33, 326, 347, 361, 375, 384, 386, 450, 452, 453, 476, 479, 484, 487, 489, 501, 505

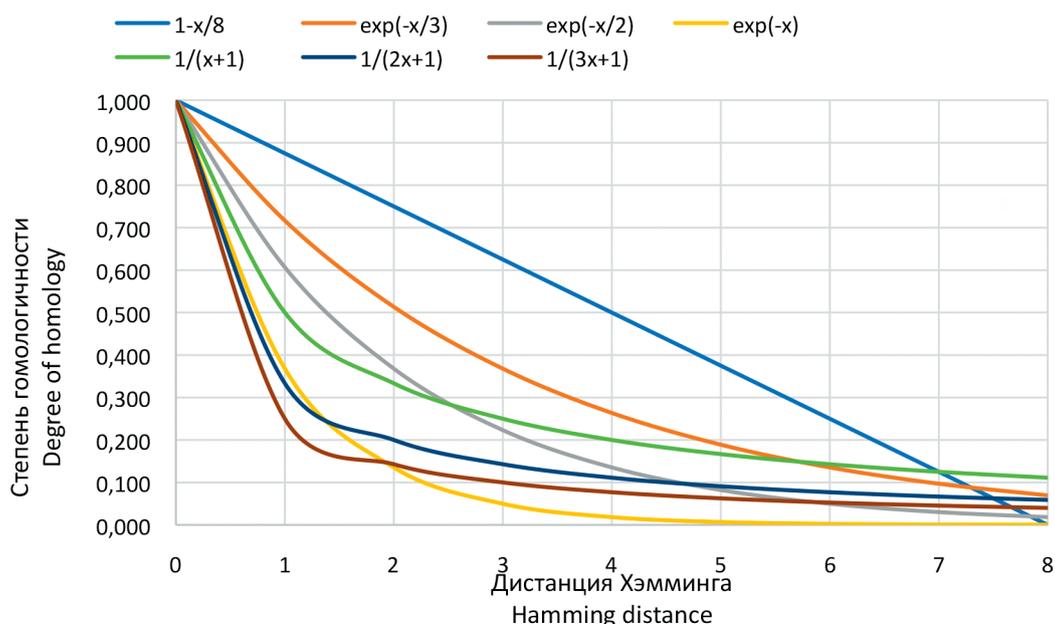


Рисунок. Функции для оценки степени гомологичности.

Figure. Functions for evaluating the degree of homology.

блюдений: Cell-Cell **28 621** и Egg_s-Cell_t **16 463**. Также вместо простой дистанции Хэмминга мы решили использовать функцию от неё, тем самым вводя понятие **степени гомологичности** между двумя произвольными антигенными последовательностями, обозначающее, насколько близки друг к другу антигенные сайты двух различных штаммов с точки зрения антигенных свойств, и принимающее значение этой функции. Необходимо было выбрать функцию, которая могла бы и удовлетворить предъявленным требованиям, и не противоречила бы биологическим процессам: в ситуации полной (абсолютной) гомологичности двух штаммов (при отсутствии замен в антигенных сайтах) функция принимает значение 1, а в случае отсутствия гомологичности (изменения отмечались в каждой аминокислотной позиции, входящей в антигенный сайт) значение функции должно быть равно 0 (либо близко к 0). Очевидно, что функция должна убывать при увеличении дистанции Хэмминга.

Кроме того, как показано в работе [45], функция, зависящая от дистанции Хэмминга и описывающая пере-

крестную реактивность антигенов, является вогнутой. Исходя из этого, мы ввели дополнительное ограничение для рассматриваемой нами функции, а именно то, что она является вогнутой или по крайней мере не выпуклой, т.е. каждая следующая замена вносит не больший вклад в уменьшение гомологичности, чем предыдущая.

Опираясь на результаты исследований коллег [16, 45], которые показали, что при количестве аминокислотных замен в антигенном сайте, равном или больше 7, перекрестная реактивность распознающих антител между штаммами практически не наблюдается, мы ввели дополнительное требование к искомой функции: для 7 замен значение функции должно принимать столь малое значение, которым можно пренебречь. Мы использовали также функцию: $1 - x / 8$, в качестве ограничивающей сверху рассматриваемые функции (для значений $x = \{1; 7\}$), и в качестве порога значимости при 7 заменах мы приняли значение: $1 - 7 / 8 = 0,125$.

С учётом указанных требований были рассмотрены следующие функции (**рисунок**):

Таблица 3. Значения R^2 по пассажным историям для рассматриваемых функцийTable 3. R^2 values by passage histories for the considered functions

Функция Function	Cell-Cell (21 580 пар штаммов 2009–2022) (21 580 pairs of strains 2009–2022)						Egg_s-Cell_t (12 444 пары штаммов 2009–2022) (12 444 pairs of strains 2009–2022)					
	1_20%	2_20%	3_20%	4_20%	5_20%	ср_20%	1_20%	2_20%	3_20%	4_20%	5_20%	ср_20%
$1 - x / 8$	0,457	0,483	0,471	0,486	0,490	0,477	0,273	0,252	0,277	0,282	0,255	0,268
$\exp(-x / 3)$	0,494	0,519	0,504	0,520	0,530	0,513	0,326	0,293	0,317	0,336	0,295	0,313
$\exp(-x / 2)$	0,476	0,498	0,478	0,494	0,509	0,491	0,331	0,294	0,313	0,328	0,293	0,312
$\exp(-x)$	0,366	0,380	0,346	0,363	0,393	0,370	0,253	0,205	0,200	0,242	0,207	0,222
$1 / (x + 1)$	0,441	0,461	0,438	0,453	0,471	0,453	0,312	0,272	0,283	0,315	0,275	0,292
$1 / (2x + 1)$	0,376	0,391	0,358	0,375	0,399	0,380	0,268	0,220	0,216	0,258	0,224	0,237
$1 / (3x + 1)$	0,328	0,340	0,301	0,320	0,348	0,327	0,213	0,162	0,149	0,196	0,168	0,178

- $1 - x / 8$;
- $\exp(-a \times x)$, где $a = 1/3; 1/2; 1$;
- $1 / (a \times x + 1)$, где $a = 1; 2; 3$.

Значения титров РТГА, обозначенные в оригинальных таблицах ВОЗ как $< (< 40)$ в нашей модели были заменены на значение 20. В качестве коэффициента c_0 мы принимаем фиксированное значение наименьшего значения модифицированного титра, т.е. $\log_2(20) \approx 4,322$.

В левой части табл. 3 представлены результаты для подмножества с пассажной историей на культуре клеток (Cell-Cell); в правой части таблицы – со смешанной пассажной историей (Egg_s-Cell_t).

Рассмотрение выбранных функций производилось путём кросс-валидации: весь объём данных 5 раз случайным образом делился в отношении 80% на 20%, где первая часть бралась за обучающую выборку, а вторая – за тестовую. Результаты оценивались по наибольшему среднему значению скорректированного R^2 , полученному для тестовых выборок. Мы также провели ранжирование результатов для всех функций. При сравнении результатов между подмножествами Cell-Cell и Egg_s-Cell_t наблюдалось практически полное соответствие рангов (коэффициент ранговой корреляции Спирмена $\rho = 0,96$) функций применительно к подмножествам. Между тем при сравнении значений коэффициентов детерминации для двух подмножеств соответствующие наивысшие величины отмечались для подмножества Cell-Cell (в 1,6 раза), что подтверждает целесообразность разделения имеющихся данных по пассажирам, а также подтверждает, что пассажи на куриных эмбрионах содержат характерные адаптационные замены, меняющие рецепторную специфичность вирусов и характер гликозилирования [6, 35] и, как следствие, влияют на перекрёстную реактивность в РТГА. Поэтому дальнейшие расчёты мы проводили на подмножестве Cell-Cell. В качестве функции, описывающей степень гомологичности, была выбрана $\exp(-x / 3)$, показавшая наилучший результат для подмножества Cell-Cell и наилучший результат для подмножества Egg_s-Cell_t.

Результаты регрессионного анализа с подмножеством Cell-Cell (21 580 наблюдений после усредне-

ния по парам штаммов) с применением выбранной функции с требованием неотрицательности коэффициентов представлены в табл. 4. В первой колонке таблицы, помимо уже упомянутого выше скорректированного R^2 , представлены обозначения соответствующих антигенных сайтов, далее в колонке Estimate приведены значения искомым коэффициентов С1–6 регрессии, в третьей колонке, Standard error, перечислены/зафиксированы стандартные отклонения этих коэффициентов. В строке Standard deviation представлено стандартное отклонение зависимой переменной (модифицированных титров). О вкладе каждого антигенного сайта можно судить по значению его коэффициента. Чем больше значение коэффициента при определённом антигенном сайте, тем больше его вклад в перекрёстный иммунитет. После определения коэффициентов регрессионной модели проводится проверка статистической гипотезы о равенстве нулю истинных значений коэффициентов по критерию Стьюдента с уровнем значимости 0,05. Результаты проверки представлены в пятой колонке, $\Pr(> |t|)$ p-value, вероятность того, что t-статистика больше модуля t-value (отношение значений коэффициентов к их стандартному отклонению). Если это значение меньше уровня доверия 0,05, гипотеза отвергается, а параметр признаётся значимым. В нашем случае все антигенные сайты, кроме С и F, признаны значимыми. Коэффициенты при антигенном сайте С и F приняли значение 0, а это равносильно незначимости.

Анализ стабильности прогностической способности модели перекрёстного иммунитета на ретроспективных данных

Для более детального понимания влияния фактора времени и количества наблюдений, участвующих в настройке модели на точность прогноза, следующим действием мы разделили подмножество Cell-Cell_2009–2022 на различные временные (1-, 2-, 3-, 4-, 5-летние) настраиваемые периоды с прогнозом на каждый последующий за ними год. Полные результаты регрессионного анализа представлены в приложении.

На основе представленных в **табл. 5.1** и **5.2** результатов можно констатировать, что скорректированный R^2 , принимающий значение не выше 0,3 в настраиваемых периодах, начинает существенно увеличиваться с 2019 г., меняясь от 0,32 до 0,64 (чем больше лет в настраиваемом интервале, тем меньше значение). Из этой тенденции выбивается 2021 г., что связано с низкой заболеваемостью и малым количеством данных вследствие пандемии COVID-19. Динамика величины скорректированного R^2 находит прямое отражение и в прогнозных периодах, принимая значения не выше 0,4 (в том числе отрицательные) и далее увеличиваясь с 2020 г. (от 0,47 до 0,6).

Также мы оценивали прогностическую способность результатов, представленных в **табл. 5.1** и **5.2**, по воспроизводимости титров на одно (± 1) и два (± 2) разведения. При сопоставлении расчётных титров с опубликованными в прогнозных периодах, за ис-

ключением 2016 и 2021 гг. (малое количество наблюдений – 841 и 478 соответственно) и 2015 г. (когда рекомендованные вакцинные штаммы не смогли обеспечить эффективную защиту), была продемонстрирована воспроизводимость титров на одно (± 1) разведение более чем в 50% случаев, а периодами доходила и до 67% случаев. Что касается воспроизводимости титров на два (± 2) разведения, то на всех временных периодах стабильно наблюдается примерно 90% случаев и выше.

Результаты регрессионного анализа по настраиваемым (1–5-летним временным) интервалам продемонстрировали значение вклада каждого антигенного сайта в иммунный ответ. Коэффициенты при антигенных сайтах А и В имеют высокие значения во всех прогнозных периодах. Особенно наглядно это можно увидеть на 3–5-летних настраиваемых периодах. Антигенный сайт С фактически принимает нулевые значения лишь за редкими исключениями. Практически на всех временных интервалах устойчивый вклад, соразмерный со вкладом А и В, также демонстрирует антигенный сайт D, но лишь вплоть до 2018 г. с последующим резким обнулением. Умеренную по сравнению с антигенными сайтами А и В, но при этом устойчивую значимость демонстрирует антигенный сайт E с ощутимым увеличением к 2018–2021 гг. Диаметрально меняется вклад антигенного сайта F в зависимости от временных интервалов – от устойчиво значимого до обнуления. Следует отметить, что в целом результаты регрессии не выявляют существенной зависимости от количества лет, входящих в настраиваемый период. В то же время результаты прогноза существенно хуже в те сезоны, когда в прогнозном периоде мало данных. Кроме того, проведённый нами корреляционный анализ показал, что отсутствует корреляция между значениями коэффициентов при антигенных сайтах и количеством замен в позициях. Учитывая тот факт, что количество наблюдений только за 2022 г. составляет более 1/3 от общего объёма наблюдений за 2009–2022 гг., было принято решение о проведении отдельных расчётов по наблюдениям 2022 г. В **табл. 6** представлены результаты рас-

Таблица 4. Результаты регрессии для функции $\exp(-x / 3)$

Table 4. Regression results for the $\exp(-x / 3)$ function

Антигенные сайты Antigenic sites	Cell-Cell_2009–2022 (21 580 пар штаммов) (21 580 pairs of strains)		
	значение estimate	стандартная ошибка standard error	P-r(> t) p-value
A	1,233	0,034	<0,001
B	2,506	0,034	<0,001
C	0,000		
D	0,083	0,027	0,002
E	0,622	0,039	<0,001
F	0,000		
CR ²	0,514		
Стандартное отклонение Standard deviation	1,161		

Таблица 5.1. R²

Table 5.1. R²

Год Year	1 год year	2 года year	3 года year	4 года year	5 лет year
2014	0,135	0,099	0,099	0,101	0,101
2015	-0,223	-0,204	-0,354	-0,355	-0,348
2016	0,043	-0,225	-0,278	-0,540	-0,546
2017	0,005	0,198	0,158	0,083	0,071
2018	0,092	0,001	0,100	0,071	0,052
2019	0,124	0,244	0,405	0,318	0,277
2020	0,596	0,592	0,577	0,561	0,543
2021	0,044	0,095	0,125	0,135	0,128
2022	0,600	0,472	0,497	0,540	0,561

Таблица 5.2. ± 1 разведение

Table 5.2. ± 1 dilution

Год Year	1 год year	2 года year	3 года year	4 года year	5 лет year
2014	67,84	67,94	67,94	67,24	67,24
2015	46,46	46,69	45,67	45,59	45,75
2016	40,43	38,41	36,50	34,48	34,36
2017	56,40	59,40	63,40	55,83	59,54
2018	65,27	59,62	62,97	60,88	61,09
2019	45,44	50,94	59,41	57,20	57,57
2020	57,70	57,03	58,33	57,32	56,79
2021	46,03	48,12	50,63	53,35	51,88
2022	55,27	50,18	54,16	55,66	60,49

Таблица 6. Результаты регрессии по данным 2022 г.

Table 6. Regression results for 2022 data

Настраиваемые периоды Customizable period	2022 г.	Данные февральского отчета 2022 г. February 2022 report data
Количество наблюдений Number of observations	8183	1994
Скорректированный R ² Adjusted R ²	0,728	0,625
Стандартное отклонение Standard deviation	0,874	0,966
Прогнозные периоды Forecast periods	Данные сентябрьского отчета 2022 г. September 2022 report data	
Количество наблюдений Number of observations	6675	
Скорректированный R ² Adjusted R ²	0,734	
Стандартное отклонение Standard deviation	0,933	
±1 разведение / dilution	69,33%	
±2 разведения / dilution	98,73%	

чётов прогностической способности разработанной модели перекрёстного иммунитета на опубликованных данных по РТГА в 2022 г. (весеннего и осеннего сезонов).

Обсуждение

Из обзора научной литературы следует, что в большинстве работ изучение перекрёстного иммунитета проводилось на ограниченном массиве данных, включающем только референсные штаммы, что снижает объективность исследований и затрудняет возможность проверки прогноза. Проведение расчётов на основании всего имеющегося массива опубликованных ВОЗ данных РТГА позволило обойти эти ограничения. С помощью встроенного дополнительного модуля **Influenza DP (data processing)**, который представляет собой универсальный обработчик баз данных (GISAID, NCBI и др.), нами был разработан алгоритм, позволяющий в режиме реального времени формировать базы данных согласно поставленным задачам для последующего анализа и расчётов. Для изучения закономерностей в модели перекрёстного иммунитета рассматривались подмножества с наибольшим количеством наблюдений – Cell-Cell и Egg_s-Cell_t. Мы также ввели степени гомологичности между двумя произвольными штаммами, которая принимает значение функции, зависящей от дистанции Хэмминга.

При рассмотрении функций для выбора варианта, описывающего степень гомологичности, была отмечена существенная зависимость значений коэффициента детерминации от вида функции. При том что всем функциям предъявлялись одинаковые необходимые условия, соответствующие биологическим

процессам, полученные в результате значения скорректированного R² колебались довольно существенно и принимали значения от 0,327 до 0,513 для подмножества Cell-Cell и 0,178 до 0,313 для подмножества Egg_s-Cell_t соответственно. Столь существенное различие между верхними и нижними границами показателя указывает на то, что использование простой дистанции Хэмминга [16, 18] в качестве меры гомологичности может накладывать существенные ограничения на используемые модели. Также было проведено ранжирование результатов для всех функций. При сравнении результатов с подмножествами Cell-Cell и Egg_s-Cell_t наблюдалось практически полное соответствие рангов функций по значению скорректированного R² применительно к подмножествам. Полученные результаты могут свидетельствовать об определённом соответствии биологическим механизмам нашего подхода к выбору функции.

Сравнение значений скорректированного R² на подмножествах Cell-Cell и Egg_s-Cell_t (0,523 и 0,313 соответственно) демонстрирует обоснованность нашего решения о разделении первоначального массива данных по пассажным историям. Разница скорректированного R² также позволяет предположить, что вирусы, культивированные в куриных эмбрионах, вносят большую погрешность в результаты РТГА.

В своих расчётах мы опирались на результаты собственных исследований объединения аминокислотных позиций в антигенные сайты, описали дополнительный сайт F, тем самым расширив количество антигенных сайтов до 6. Результаты регрессионного анализа на всех подмножествах Cell-Cell продемонстрировали, что наибольший вклад в иммунный ответ внесли антигенные сайты A, B. Их умеренно догоняет по устойчивым значениям антигенный сайт E с ошутимым ростом в 2018–2021 гг. Антигенный сайт C практически не вносит вклад в иммунный ответ во всех рассматриваемых интервалах. В значительной степени уверенно демонстрирующий значимость в начале прогнозных периодов антигенный сайт D начинает сдавать позиции в 2018–2021 гг. Довольно чувствительным к временным интервалам оказался антигенный сайт F со своим сдержанно неустойчивым вкладом в перекрёстный иммунитет.

Кроме того, был произведён анализ стабильности прогностической способности нашей модели для разных временных интервалов (с 2009 по 2022 г.). После разделения подмножества Cell-Cell на 1–5-летние временные интервалы отмечалось, что значение скорректированного R² в настраиваемых периодах возрастает начиная с 2019 г., достигая порядка 0,5 и выше, за исключением 2021 г., в котором количество наблюдений очень мало (478) вследствие пандемии COVID-19. Аналогичная картина наблюдается и в прогнозных периодах. Выраженное высокое значение скорректированного R² для более поздних данных может являться следствием увеличения проводимых исследований (РТГА) на культуре клеток и улучшения их качества.

При сравнении расчётных значений с опубликованными титрами результаты воспроизводимости на

одно (± 1) разведение довольно стабильны по годам прогноза и имеют значения больше 50% (со средним значением около 60%), за исключением 2015, 2016 и 2021 гг. При этом основную причину более низких результатов мы видим в малом количестве данных прогнозных периодов.

Как было показано в результатах, отмечаемое беспрецедентно многочисленное количество наблюдений за 2022 г. предоставило нам возможность провести настройку модели по данным на начало 2022 г. (скорректированный $R^2 = 0,728$) и проверить результаты прогноза на осеннем сезоне. Результаты прогноза (скорректированный $R^2 = 0,734$) продемонстрировали воспроизводимость титров на одно (± 1) разведение в 69,33% случаев и оказались существенно лучше всех ранее рассматриваемых периодов, что можно уже считать достаточным для практического применения. Объяснением такого результата может быть как качество данных, так и прогнозный период, ограниченный сезоном. В работе Т. Bedford и соавт. [15], которые проводили регрессионный анализ на всём массиве данных РТГА для подтипа вируса H_3N_2 , был получен $R^2 = 0,372$, что ниже по сравнению с нашими показателями скорректированного $R^2 = 0,523$ на всём подмножестве Cell-Cell. На результатах коллег могло сказаться отсутствие разделения данных по пассажным историям и на временные интервалы. Также на результаты, возможно, повлияло использование простой дистанции Хэмминга. Важно признать, что, несмотря на все усилия по стандартизации проведения РТГА [32] и независимо от выбранного подхода для исследований зависимости титров РТГА от аминокислотных замен, изначально заложенная высокая погрешность методики РТГА (17%) [46] остаётся фактором, существенно влияющим на результаты.

Заключение

Разработанная нами компьютерная программа, основанная на эпидемиологической модели и с хорошо настроенной моделью перекрёстного иммунитета, предоставляет возможность предсказания наиболее распространённых штаммов с учётом влияния иммунного ландшафта и подбора вакцинного штамма на предстоящий сезон.

Возможность проводить расчёты на основании всего имеющегося массива опубликованных ВОЗ еже-сезонных данных по РТГА, наблюдаемое улучшение результатов за последние годы (начиная с 2019 г.), а также большой объём наблюдений и полученные результаты прогноза за 2022 г. (скорректированный $R^2 = 0,734$; воспроизводимость титров на одно (± 1) разведение в 69,33% случаев) позволяют утверждать, что предложенный метод может служить хорошим инструментом для будущих прогнозов с дальнейшим изучением для подтверждения его устойчивости.

В свою очередь, преимущество применения компьютерной программы **Influenza IDE** с постоянно обновляющейся базой данных различных видов и суб-типов вируса гриппа в перспективе позволит воспроизвести полученные результаты на других вариантах

и подтипах вируса гриппа и тем самым протестировать разработанную модель и расширить витрину рекомендуемых вакцинных штаммов. Полученный опыт и навыки при определении антигенных сайтов в будущем позволят объединить аминокислотные позиции в антигенные сайты по нескольким сценариям и провести расчёты уже с их учётом. Дальнейшее развитие модели перекрёстного иммунитета может быть достигнуто за счёт одностороннего нормирования титров всего массива тестовых штаммов по референсным штаммам, что позволит уменьшить ошибку, связанную с индивидуальными особенностями используемых экспериментальных животных. Также улучшение модели может быть достигнуто за счёт учёта таких процессов, как гликозилирование.

ЛИТЕРАТУРА

- Russell C.A., Jones T.C., Barr I.G., Cox N.J., Garten R.J., Gregory V., et al. Influenza vaccine strain selection and recent studies on the global migration of seasonal influenza viruses. *Vaccine*. 2008; 26(Suppl. 4): 31–4. <https://doi.org/10.1016/j.vaccine.2008.07.078>
- Belongia E.A., Simpson M.D., King J.P., Sundaram M.E., Kelley N.S., Osterholm M.T., et al. Variable influenza vaccine effectiveness by subtype: a systematic review and meta-analysis of test-negative design studies. *Lancet Infect. Dis*. 2016; 16(8): 942–51. [https://doi.org/10.1016/S1473-3099\(16\)00129-8](https://doi.org/10.1016/S1473-3099(16)00129-8)
- Jackson M.L., Chung J.R., Jackson L.A., Phillips C.H., Benoit J., Monto A.S., et al. Influenza vaccine effectiveness in the United States during the 2015–2016 season. *N. Engl. J. Med*. 2017; 377(6): 534–43. <https://doi.org/10.1056/NEJMoa1700153>
- Rolfes M.A., Flannery B., Chung J.R., O'Halloran A., Garg S., Belongia E.A., et al. Effects of influenza vaccination in the United States during the 2017–2018 influenza season. *Clin. Infect. Dis*. 2019; 69(11): 1845–53. <https://doi.org/10.1093/cid/ciz075>
- Doyle J.D., Chung J.R., Kim S.S., Gaglani M., Raiyani C., Zimmerman R.K., et al. Interim estimates of 2018–2019 seasonal influenza vaccine effectiveness – United States. *MMWR. Morb. Mortal. Wkly Rep*. 2019; 68(6): 135–9. <https://doi.org/10.15585/mmwr.mm6806a2>
- Zost S.J., Parkhouse K., Gumina M.E., Kim K., Diaz P.S., Wilson P.C., et al. Contemporary H3N2 influenza viruses have a glycosylation site that alters binding of antibodies elicited by egg-adapted vaccine strains. *Proc. Natl Acad. Sci. USA*. 2017; 114(47): 12578–83. <https://doi.org/10.1073/pnas.1712377114>
- Gouma S., Weirick M., Hensley S.E. Antigenic assessment of the H3N2 component of the 2019–2020 northern hemisphere influenza vaccine. *Nat. Commun*. 2020; 11(1): 2445. <https://doi.org/10.1038/s41467-020-16183-y>
- Cobey S., Gouma S., Parkhouse K., Chambers B.S., Ertl H.C., Schmader K.E., et al. Poor immunogenicity, not vaccine strain egg adaptation, may explain the low H3N2 influenza vaccine effectiveness in 2012–2013. *Clin. Infect. Dis*. 2018; 67(3): 327–33. <https://doi.org/10.1093/cid/ciy097>
- Klingen T.R., Reimering S., Guzmán C.A., McHardy A.C. In silico vaccine strain prediction for human influenza viruses. *Trends Microbiol*. 2018; 26(2): 119–31. <https://doi.org/10.1016/j.tim.2017.09.001>
- Morris D.H., Gostic K.M., Pompei S., Bedford T., Luksza M., Nether R.A., et al. Predictive modeling of influenza shows the promise of applied evolutionary biology. *Trends Microbiol*. 2018; 26(2): 102–18. <https://doi.org/10.1016/j.tim.2017.09.004>
- Тимофеева Т.А., Асатрян М.Н., Альштейн А.Д., Народицкий Б.С., Гицбург А.Л., Каверин Н.В. Прогнозирование эволюционной изменчивости вируса гриппа А. *Acta Naturae*. 2017; 9(3): 104–11. <https://doi.org/10.32607/20758251-2017-9-3-48-54> <https://elibrary.ru/zqitor>
- Боев Б.В. Модель развития эпидемии гриппа А(H1N1) в России в сезон 2009–2010 годов. *Эпидемиология и вакцинопрофилактика*. 2010; (1): 52–8. <https://elibrary.ru/laedxn>
- Huddleston J., Barnes J.R., Rowe T., Kondor R., Wentworth D.E., Whittaker L., et al. Integrating genotypes and phenotypes improves

- long-term forecasts of seasonal influenza A/H3N2 evolution. *eLife*. 2020; 9: e60067. <https://doi.org/10.7554/eLife.60067>
14. CDC. Center for Disease Control and Prevention. Influenza (Flu). Available at: <https://www.cdc.gov/flu/>
 15. Bedford T., Suchard M.A., Lemey P., Dudas G., Gregory V., Hay A.J., et al. Integrating influenza antigenic dynamics with molecular evolution. *eLife*. 2014; 3: e01914. <https://doi.org/10.7554/eLife.01914>
 16. Anderson C.S., McCall P.R., Stern H.A., Yang H., Topham D.J. Antigenic cartography of H1N1 influenza viruses using sequence-based antigenic distance calculation. *BMC Bioinformatics*. 2018; 19(1): 51. <https://doi.org/10.1186/s12859-018-2042-4>
 17. Lee M.S., Chen J.S. Predicting antigenic variants of influenza A/H3N2 viruses. *Emerg. Infect. Dis.* 2004; 10(8): 1385–90. <https://doi.org/10.3201/eid1008.040107>
 18. Lees W.D., Moss D.S., Shepherd A.J. A computational analysis of the antigenic properties of haemagglutinin in influenza A H3N2. *Bioinformatics*. 2010; 26(11): 1403–8. <https://doi.org/10.1093/bioinformatics/btq160>
 19. Burnet F.M., Lush D. The action of certain surface active agents on viruses. *Aust. J. Exp. Biol. Med. Sci.* 1940; 18: 141–50.
 20. Archetti I., Horsfall F.L. Persistent antigenic variation of influenza A viruses after incomplete neutralization in ovo with heterologous immune serum. *J. Exp. Med.* 1950; 92(5): 441–62. <https://doi.org/10.1084/jem.92.5.441>
 21. Lapedes A., Farber R. The geometry of shape space: application to influenza. *J. Theor. Biol.* 2001; 212(1): 57–69. <https://doi.org/10.1006/jtbi.2001.2347>
 22. Smith D.J., Lapedes A.S., de Jong J.C., Bestebroer T.M., Rimmelzwaan G.F., Osterhaus A.D., et al. Mapping the antigenic and genetic evolution of influenza virus. *Science*. 2004; 305(5682): 371–6. <https://doi.org/10.1126/science.1097211>
 23. Wiley D.C., Skehel J.J. The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. *Annu. Rev. Biochem.* 1987; 56: 365–94. <https://doi.org/10.1146/annurev.bi.56.070187.002053>
 24. Wiley D.C., Wilson I.A., Skehel J.J. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. *Nature*. 1981; 289(5796): 373–8. <https://doi.org/10.1038/289373a0>
 25. Wilson I.A., Cox N.J. Structural basis of immune recognition of influenza virus hemagglutinin. *Annu. Rev. Immunol.* 1990; 8: 737–71. <https://doi.org/10.1146/annurev.iy.08.040190.003513>
 26. Liao Y.C., Lee M.S., Ko C.Y., Hsiung C.A. Bioinformatics models for predicting antigenic variants of influenza A/H3N2 virus. *Bioinformatics*. 2008; 24(4): 505–12. <https://doi.org/10.1093/bioinformatics/btm638>
 27. Асатрян М.Н., Агасарян В.Г., Щербинин Д.Н., Тимофеев Б.И., Ершов И.Ф., Шмыр И.С. и др. Influenza IDE. Свидетельство о регистрации № 2020617965; 2020.
 28. Lawson C.L., Hanson R.J. *Solving Least Squares Problems*. New Jersey: Englewood Cliffs; 1974.
 29. Халафян А.А. *Математическая статистика с элементами теории вероятности*. М.: Бином; 2010.
 30. Stephenson I., Gaines Das R., Wood J.M., Katz J.M. Comparison of neutralising antibody assays for detection of antibody to influenza A/H3N2 viruses: an international collaborative study. *Vaccine*. 2007; 25(20): 4056–63. <https://doi.org/10.1016/j.vaccine.2007.02.039>
 31. Wood J.M., Major D., Heath A., Newman R.W., Höscher K., Stephenson I., et al. Reproducibility of serology assays for pandemic influenza H1N1: collaborative study to evaluate a candidate WHO International Standard. *Vaccine*. 2012; 30(2): 210–7. <https://doi.org/10.1016/j.vaccine.2011.11.019>
 32. Zacour M., Ward B.J., Brewer A., Tang P., Boivin G., Li Y. Standardization of hemagglutination inhibition assay for influenza serology allows for high reproducibility between laboratories. *Clin. Vaccine Immunol.* 2016; 23(3): 236–42. <https://doi.org/10.1128/CVI.00613-15>
 33. The Francis Crick Institute. Worldwide Influenza Centre lab. Annual and interim reports. Available at: <https://www.crick.ac.uk/research/platforms-and-facilities/worldwide-influenza-centre/annual-and-interim-reports>
 34. DuPai C.D., McWhite C.D., Smith C.B., Garten R., Maurer-Stroh S., Wilke C.O. Influenza passaging annotations: what they tell us and why we should listen. *Virus Evol.* 2019; 5(1): vez016. <https://doi.org/10.1093/ve/vez016>
 35. Wu N.C., Zost S.J., Thompson A.J., Oyen D., Nycholat C.M., McBride R., et al. A structural explanation for the low effectiveness of the seasonal influenza H3N2 vaccine. *PLoS Pathog.* 2017; 13(10): e1006682. <https://doi.org/10.1371/journal.ppat.1006682>
 36. Park Y.W., Kim Y.H., Jung H.U., Jeong O.S., Hong E.J., Kim H. Comparison of antigenic mutation during egg and cell passage cultivation of H3N2 influenza virus. *Clin. Exp. Vaccine Res.* 2020; 9(1): 56–63. <https://doi.org/10.7774/cevr.2020.9.1.56>
 37. Popova L., Smith K., West A.H., Wilson P.C., James J.A., Thompson L.F. Immunodominance of antigenic site B over site A of hemagglutinin of recent H3N2 influenza viruses. *PLoS One*. 2012; 7(7): e41895. <https://doi.org/10.1371/journal.pone.0041895>
 38. Klein N.P., Fireman B., Goddard K., Zerbo O., Asher J., Zhou J. Vaccine effectiveness of cell-culture relative to egg-based inactivated influenza vaccine during the 2017–2018 influenza season. *PLoS One*. 2020; 15(2): e0229279. <https://doi.org/10.1371/journal.pone.0229279>
 39. GISAID. Eurosurveillance; 2017. Available at: <https://gisaid.org/resources/commentary-on-gisaid/>
 40. Smith T.F., Waterman M.S. Identification of common molecular subsequences. *J. Mol. Biol.* 1981; 147(1): 195–7. [https://doi.org/10.1016/0022-2836\(81\)90087-5](https://doi.org/10.1016/0022-2836(81)90087-5)
 41. GetArea. Available at: <http://curie.utmb.edu/getarea.html>
 42. RCSB PDB: Homepage. Available at: <https://www.rcsb.org/>
 43. Щербинин Д.Н., Алексеева С.В., Шмаров М.М., Смирнов Ю.А., Народицкий Б.С., Гинцбург А.Л. Анализ В-клеточных эпитопов гемагглютинаина вирусов гриппа. *Acta Naturae*. 2016; 8(1): 14–22. <https://doi.org/10.32607/20758251-2016-8-1-13-20> <https://elibrary.ru/vsjiyh>
 44. A standardised numbering for all subtypes of Influenza A hemagglutinin (HA) sequences based on the mature HA sequence. Available at: <https://antigenic-cartography.org/surveillance/evergreen/HANumbering/>
 45. Smith D.J., Forrest S., Hightower R.R., Perelson A.S. Deriving shape space parameters from immunological data. *J. Theor. Biol.* 1997; 189(2): 141–50. <https://doi.org/10.1006/jtbi.1997.0495>
 46. Кильбурн Э.Д., ред. *Вирусы гриппа и грипп*. Пер. с англ. М.: Медицина; 1978.

REFERENCES

1. Russell C.A., Jones T.C., Barr I.G., Cox N.J., Garten R.J., Gregory V., et al. Influenza vaccine strain selection and recent studies on the global migration of seasonal influenza viruses. *Vaccine*. 2008; 26(Suppl. 4): 31–4. <https://doi.org/10.1016/j.vaccine.2008.07.078>
2. Belongia E.A., Simpson M.D., King J.P., Sundaram M.E., Kelley N.S., Osterholm M.T., et al. Variable influenza vaccine effectiveness by subtype: a systematic review and meta-analysis of test-negative design studies. *Lancet Infect. Dis.* 2016; 16(8): 942–51. [https://doi.org/10.1016/S1473-3099\(16\)00129-8](https://doi.org/10.1016/S1473-3099(16)00129-8)
3. Jackson M.L., Chung J.R., Jackson L.A., Phillips C.H., Benoit J., Monto A.S., et al. Influenza vaccine effectiveness in the United States during the 2015–2016 season. *N. Engl. J. Med.* 2017; 377(6): 534–43. <https://doi.org/10.1056/NEJMoa1700153>
4. Rolfes M.A., Flannery B., Chung J.R., O'Halloran A., Garg S., Belongia E.A., et al. Effects of influenza vaccination in the United States during the 2017–2018 influenza season. *Clin. Infect. Dis.* 2019; 69(11): 1845–53. <https://doi.org/10.1093/cid/ciz075>
5. Doyle J.D., Chung J.R., Kim S.S., Gaglani M., Raiyani C., Zimmerman R.K., et al. Interim estimates of 2018–2019 seasonal influenza vaccine effectiveness – United States. *MMWR. Morb. Mortal. Wkly Rep.* 2019; 68(6): 135–9. <https://doi.org/10.15585/mmwr.mm6806a2>
6. Zost S.J., Parkhouse K., Gumina M.E., Kim K., Diaz P.S., Wilson P.C., et al. Contemporary H3N2 influenza viruses have a glycosylation site that alters binding of antibodies elicited by egg-adapted vaccine strains. *Proc. Natl Acad. Sci. USA*. 2017; 114(47): 12578–83. <https://doi.org/10.1073/pnas.1712377114>
7. Gouma S., Weirick M., Hensley S.E. Antigenic assessment of the H3N2 component of the 2019–2020 northern hemisphere influenza vaccine. *Nat. Commun.* 2020; 11(1): 2445. <https://doi.org/10.1038/s41467-020-16183-y>
8. Cobey S., Gouma S., Parkhouse K., Chambers B.S., Ertl H.C., Schmader K.E., et al. Poor immunogenicity, not vaccine strain egg adaptation, may explain the low H3N2 influenza vaccine

- effectiveness in 2012–2013. *Clin. Infect. Dis.* 2018; 67(3): 327–33. <https://doi.org/10.1093/cid/ciy097>
9. Klingon T.R., Reimering S., Guzmán C.A., McHardy A.C. In silico vaccine strain prediction for human influenza viruses. *Trends Microbiol.* 2018; 26(2): 119–31. <https://doi.org/10.1016/j.tim.2017.09.001>
 10. Morris D.H., Gostic K.M., Pompei S., Bedford T., Łuksza M., Neher R.A., et al. Predictive modeling of influenza shows the promise of applied evolutionary biology. *Trends Microbiol.* 2018; 26(2): 102–18. <https://doi.org/10.1016/j.tim.2017.09.004>
 11. Timofeeva T.A., Asatryan M.N., Al'tshiteyn A.D., Naroditskiy B.S., Gintsburg A.L., Kaverin N.V. Predicting the evolutionary variability of the influenza A virus. *Acta Naturae.* 2017; 9(3): 48–54. <https://doi.org/10.32607/20758251-2017-9-3-48-54> <https://elibrary.ru/zqitjr>
 12. Boev B.V. Modeling of the epidemic of influenza A(H1N1) in Russia season 2009–2010. *Epidemiologiya i vaksinoprofilaktika.* 2010; (1): 52–8. <https://elibrary.ru/laedxn> (in Russian)
 13. Huddleston J., Barnes J.R., Rowe T., Kondor R., Wentworth D.E., Whittaker L., et al. Integrating genotypes and phenotypes improves long-term forecasts of seasonal influenza A/H3N2 evolution. *eLife.* 2020; 9: e60067. <https://doi.org/10.7554/eLife.60067>
 14. CDC. Center for Disease Control and Prevention. Influenza (Flu). Available at: <https://www.cdc.gov/flu/>
 15. Bedford T., Suchard M.A., Lemey P., Dudas G., Gregory V., Hay A.J., et al. Integrating influenza antigenic dynamics with molecular evolution. *eLife.* 2014; 3: e01914. <https://doi.org/10.7554/eLife.01914>
 16. Anderson C.S., McCall P.R., Stern H.A., Yang H., Topham D.J. Antigenic cartography of H1N1 influenza viruses using sequence-based antigenic distance calculation. *BMC Bioinformatics.* 2018; 19(1): 51. <https://doi.org/10.1186/s12859-018-2042-4>
 17. Lee M.S., Chen J.S. Predicting antigenic variants of influenza A/H3N2 viruses. *Emerg. Infect. Dis.* 2004; 10(8): 1385–90. <https://doi.org/10.3201/eid1008.040107>
 18. Lees W.D., Moss D.S., Shepherd A.J. A computational analysis of the antigenic properties of haemagglutinin in influenza A H3N2. *Bioinformatics.* 2010; 26(11): 1403–8. <https://doi.org/10.1093/bioinformatics/btq160>
 19. Burnet F.M., Lush D. The action of certain surface active agents on viruses. *Aust. J. Exp. Biol. Med. Sci.* 1940; 18: 141–50.
 20. Archetti I., Horsfall F.L. Persistent antigenic variation of influenza A viruses after incomplete neutralization in ovo with heterologous immune serum. *J. Exp. Med.* 1950; 92(5): 441–62. <https://doi.org/10.1084/jem.92.5.441>
 21. Lapedes A., Farber R. The geometry of shape space: application to influenza. *J. Theor. Biol.* 2001; 212(1): 57–69. <https://doi.org/10.1006/jtbi.2001.2347>
 22. Smith D.J., Lapedes A.S., de Jong J.C., Bestebroer T.M., Rimmelzwaan G.F., Osterhaus A.D., et al. Mapping the antigenic and genetic evolution of influenza virus. *Science.* 2004; 305(5682): 371–6. <https://doi.org/10.1126/science.1097211>
 23. Wiley D.C., Skehel J.J. The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. *Annu. Rev. Biochem.* 1987; 56: 365–94. <https://doi.org/10.1146/annurev.bi.56.070187.002053>
 24. Wiley D.C., Wilson I.A., Skehel J.J. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. *Nature.* 1981; 289(5796): 373–8. <https://doi.org/10.1038/289373a0>
 25. Wilson I.A., Cox N.J. Structural basis of immune recognition of influenza virus hemagglutinin. *Annu. Rev. Immunol.* 1990; 8: 737–71. <https://doi.org/10.1146/annurev.iy.08.040190.003513>
 26. Liao Y.C., Lee M.S., Ko C.Y., Hsiung C.A. Bioinformatics models for predicting antigenic variants of influenza A/H3N2 virus. *Bioinformatics.* 2008; 24(4): 505–12. <https://doi.org/10.1093/bioinformatics/btm638>
 27. Asatryan M.N., Agasaryan V.G., Shcherbinin D.N., Timofeev B.I., Ershov I.F., Shmyr I.S., et al. Influenza IDE. Registration certificate № 2020617965; 2020. (in Russian)
 28. Lawson C.L., Hanson R.J. *Solving Least Squares Problems.* New Jersey: Englewood Cliffs; 1974.
 29. Khalafyan A.A. *Mathematical Statistics with Elements of Probability [Matematicheskaya statistika s elementami teorii veroyatnosti].* Moscow: Binom; 2010. (in Russian)
 30. Stephenson I., Gaines Das R., Wood J.M., Katz J.M. Comparison of neutralising antibody assays for detection of antibody to influenza A/H3N2 viruses: an international collaborative study. *Vaccine.* 2007; 25(20): 4056–63. <https://doi.org/10.1016/j.vaccine.2007.02.039>
 31. Wood J.M., Major D., Heath A., Newman R.W., Höschler K., Stephenson I., et al. Reproducibility of serology assays for pandemic influenza H1N1: collaborative study to evaluate a candidate WHO International Standard. *Vaccine.* 2012; 30(2): 210–7. <https://doi.org/10.1016/j.vaccine.2011.11.019>
 32. Zacour M., Ward B.J., Brewer A., Tang P., Boivin G., Li Y. Standardization of hemagglutination inhibition assay for influenza serology allows for high reproducibility between laboratories. *Clin. Vaccine Immunol.* 2016; 23(3): 236–42. <https://doi.org/10.1128/CVI.00613-15>
 33. The Francis Crick Institute. Worldwide Influenza Centre lab. Annual and interim reports. Available at: <https://www.crick.ac.uk/research/platforms-and-facilities/worldwide-influenza-centre/annual-and-interim-reports>
 34. DuPai C.D., McWhite C.D., Smith C.B., Garten R., Maurer-Stroh S., Wilke C.O. Influenza passaging annotations: what they tell us and why we should listen. *Virus Evol.* 2019; 5(1): vez016. <https://doi.org/10.1093/ve/vez016>
 35. Wu N.C., Zost S.J., Thompson A.J., Oyen D., Nycholat C.M., McBride R., et al. A structural explanation for the low effectiveness of the seasonal influenza H3N2 vaccine. *PLoS Pathog.* 2017; 13(10): e1006682. <https://doi.org/10.1371/journal.ppat.1006682>
 36. Park Y.W., Kim Y.H., Jung H.U., Jeong O.S., Hong E.J., Kim H. Comparison of antigenic mutation during egg and cell passage cultivation of H3N2 influenza virus. *Clin. Exp. Vaccine Res.* 2020; 9(1): 56–63. <https://doi.org/10.7774/cevr.2020.9.1.56>
 37. Popova L., Smith K., West A.H., Wilson P.C., James J.A., Thompson L.F. Immunodominance of antigenic site B over site A of hemagglutinin of recent H3N2 influenza viruses. *PLoS One.* 2012; 7(7): e41895. <https://doi.org/10.1371/journal.pone.0041895>
 38. Klein N.P., Fireman B., Goddard K., Zerbo O., Asher J., Zhou J. Vaccine effectiveness of cell-culture relative to egg-based inactivated influenza vaccine during the 2017–2018 influenza season. *PLoS One.* 2020; 15(2): e0229279. <https://doi.org/10.1371/journal.pone.0229279>
 39. GISAID. Eurosurveillance; 2017. Available at: <https://gisaid.org/resources/commentary-on-gisaid/>
 40. Smith T.F., Waterman M.S. Identification of common molecular subsequences. *J. Mol. Biol.* 1981; 147(1): 195–7. [https://doi.org/10.1016/0022-2836\(81\)90087-5](https://doi.org/10.1016/0022-2836(81)90087-5)
 41. GetArea. Available at: <http://curie.utmb.edu/getarea.html>
 42. RCSB PDB: Homepage. Available at: <https://www.rcsb.org/>
 43. Shcherbinin D.N., Alekseeva S.V., Shmarov M.M., Smirnov Yu.A., Naroditskiy B.S., Gintsburg A.L. The analysis of B-cell epitopes of influenza virus hemagglutinin. *Acta Naturae.* 2016; 8(1): 13–20. <https://doi.org/10.32607/20758251-2016-8-1-13-20> <https://elibrary.ru/vsnklb>
 44. A standardised numbering for all subtypes of Influenza A hemagglutinin (HA) sequences based on the mature HA sequence. Available at: <https://antigenic-cartography.org/surveillance/evergreen/HAnumbering/>
 45. Smith D.J., Forrest S., Hightower R.R., Perelson A.S. Deriving shape space parameters from immunological data. *J. Theor. Biol.* 1997; 189(2): 141–50. <https://doi.org/10.1006/jtbi.1997.0495>
 46. Kilbourne E.D., ed. *The Influenza Viruses and Influenza.* London: Academic Press Inc.; 1975.