

MEDICAL SCIENCES

COMPARATIVE ANALYSIS OF CORONAVIRUS STRAINS Makarov L.M.¹, Ivanov D.O.², Pozdnyakov A.V.³ (Russian Federation) Email: Makarov453@scientifictext.ru

¹Makarov Leonid Mikhailovich - PhD in System analysis, Associate Professor,
DEPARTMENT OF DESIGNING AND PRODUCTION OF RADIO-ELECTRONIC MEANS,
ST. PETERSBURG STATE UNIVERSITY OF TELECOMMUNICATIONS
OF THE PROF. M.A. BONCH-BRUYEVICH;

²Ivanov Dmitriy Olegovich - Doctor of Medical Sciences, Professor, Rector;

³Pozdnyakov Alexander Vladimirovich - Doctor of Medical Sciences, Professor,
DEPARTMENT OF MEDICAL BIOPHYSICS,
ST. PETERSBURG STATE PEDIATRIC MEDICAL UNIVERSITY,
ST. PETERSBURG

Abstract: there is presented a systematic approach to analysis of different by structure and etiology of SARS COV, SARS MERS and SARS COV2 coronaviruses. A technology for calculating the complexity of a viral tuple created on four nucleotides has been implemented, taking into account the possibility of numerous modifications to the genome. On the basis of in-information resources of the Internet, a computational procedure for comparing various strains of coronavirus is shown, which contributes to the formation of an operational summary of events in various territories, as well as contributes to the formation of a timely and objective forecast of events in society.

Keywords: computer assessment, SARS coronavirus.

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ШТАММОВ КОРОНАВИРУСА Макаров Л.М.¹, Иванов Д.О.², Поздняков А.В.³ (Российская Федерация)

¹Макаров Леонид Михайлович – кандидат технических наук, профессор,
кафедра конструирования и производства радиоэлектронных средств,
Санкт-Петербургский государственный университет телекоммуникаций
им. проф. М.А. Бонч-Бруевича;

²Иванов Дмитрий Олегович - доктор медицинских наук, профессор, ректор;

³Поздняков Александр Владимирович – доктор медицинских наук, профессор,
кафедра медицинской биофизики,

Санкт-Петербургский государственный педиатрический медицинский университет,
г. Санкт-Петербург

Аннотация: представлен системный подход к анализу различных по строению и этиологии штаммов коронавирусов SARS COV, SARS MERS и SARS COV2. Реализована технология вычисления оценки сложности вирусного кортежа, созданного на четырех нуклеотидах, учитывающая возможность многочисленных модификаций генома. На основе информационных ресурсов сети Интернет показана вычислительная процедура сравнения различных штаммов коронавируса, способствующая формированию оперативной сводки событий на различных территориях, также способствующая формированию своевременного и объективного прогноза развития событий в социуме.

Ключевые слова: компьютерная оценка, коронавирус SARS.

УДК 539.1+519.61
DOI: 10.24411/2410-2865-2020-10402

Массовое распространение коронавируса актуализирует проблему формирования противовирусных мероприятий, основой для которых является научная концепция. Основная идея проведения сравнительного анализа разных штаммов коронавирусов формируется на представлении о наличии разных по сложности фрагментов генома. Формально этот тезис хорошо согласуется с фундаментальными положениями биологической термодинамики о строении живых объектов. Используя результаты натуральных исследований различных штаммов коронавирусов, представленных в информационном Internet ресурсе [4], представляется возможным априорно осуществить вычисление оценки сложности набора нуклеотидов. Фактически такой концептуальный взгляд на пандемию позволяет выявить слабые фрагменты генома вируса, количественно определить уровень агрессивности вируса и сформировать медицинский фармакологический пул мероприятий, способствующих понижению риска распространения вирусной пандемии.

Принимая во внимание, что размер коронавируса составляет примерно 100 нм, констатируем отсутствие возможности осуществить оперативный мониторинг посредством оптического микроскопа.

Традиционно оценка вирусной инфекции создается на основе представлений о вирулентности, некотором показателе, характеризующим финишный процесс развития событий. При взаимодействии с живым организмом штамм вируса может либо воспроизвести нарушение работы некоторых функциональных систем организма. Выделяют относительно простые нарушения, которые идентифицируют как инфицирующие и как обладающие способностью вызвать летальный исход. И в том и другом случае оперируют статистическими показателями жизненной стойкости живых организмов. Такие представления, сформированные в далеком историческом прошлом, наделенные современными понятиями и определениями воспроизвели несколько очень близких методов «констатации проявления активности штаммов вирусов с разной этиологий».

В настоящее время сама идея о наличии вирусной инфекции формируется на понятиях о генотипе, представленном набором нуклеотидов. Формально любой штамм вируса обладает уникальным набором нуклеотидов. Такой набор характеризуется кортежем, в котором установлен определенный порядок следования нуклеотидов. Формат позиционирования кортежа позволяет создавать построчную запись достаточно длинных цепочек из нуклеотидов. При длине строк в 50 – 150 элементов, воспроизводится построчная запись генома штамма вируса, который позиционируется несколькими десятками тысяч пар нуклеотидов (bp). Такая практика фиксации генома вируса позволяет создавать информационные ресурсы, доступные для исследователей из разных областей науки. Хорошо известен информационный ресурс геномной информации в Национальном центре биоинформатики Национального института здравоохранения (NCBI) в штате Мэриленд (США).

В современном социуме интенсивно изучается семейство коронавирусов, терминология для которых происходит от латинского словосочетания *coronaviridae* – корона, венец и вирус. Натурные исследования вируса позволили воспроизвести как сферическую форму вируса, так и линейную цепочку нуклеотидов в количестве примерно 30 000 bp. Предварительные исследования генома [1] указывают на преобладание трех типов фрагментов (вирусных белков): S протеинов, соотносимых с «шипами на поверхности», мембранного белка и белка оболочки. Изучая различные виды вирусов этот набор расширяется до десяти фрагментов, в которых важным элементом являются S протеины.

В историческом отношении вирус SARS-CoV (Severe acute respiratory syndrome coronavirus - 2002) – представляет *стартовую версию* коронавируса тяжелого острого респираторного синдрома (SARS) с инкубационным периодом в четыре дня[1].

Модификация генома вируса отмечена в 2012 г., когда регистрируется новое массовое поражение социума. **Вирус впервые обнаружен в июне 2012 года в городе**

Джидда, Саудовская Аравия. Этот вирус получил название MERS-CoV (английский термин - *Middle-East respiratory syndrome* - «ближневосточный синдром»).

Новое проявление вируса отмечается в 2019 г. Сохраняя таксономические правила наименования версий вируса, принимается название SARS-CoV-2, обладающего инкубационным периодом в 14 дней [1].

Наличие данных о геноме SARS-CoV-2 активизируют серию исследований по созданию тест систем, формированию филогенетического дерева, разработке методов информационного описания нуклеотидной цепи событий. Для оценки достоверности событий на филогенетическом дереве использовался метод Симодайра - Хасэгава (Shimodaira — Hasegawa), который позволил выявить связь зоонозного вируса SARS-Cov2 с вирусом SARS-Cov, который сегодня распространен в социуме разных странах.

По мнению современных исследователей, геномная цепь SARS-CoV-2 незначительно отличается от генома SARS-CoV и генома MERS-CoV, который характеризуется меньшей агрессивностью. Однако, SARS-CoV-2 демонстрирует высокую степень адаптации в социуме, сохраняя агрессивность длительное время. Вероятно, продолжительный период агрессии вируса реализуется сложным процессом жизнеобеспечения, а также большой массой, примерно в 30 000 вр. Наличие этих факторов, в семействе коронавирусов, способствует формированию устойчивых к профилактическим методам штаммов. Для организации мероприятий по профилактике распространения штаммов вируса, актуализируются задачи по применению информационных технологий, обеспечивающих как проведение сравнительного анализа, так и определение наиболее слабых фрагментов генома.

Проведем сравнительный анализ геномных последовательностей с использованием технологии создания оценок информационной энтропии [2,3]. С учетом исторического развития штаммов вируса выберем геномы: SARS-CoV, MERS-CoV, и SARS-CoV-2. Типичный геном вируса представлен нуклеотидами: глицином (G), треонином (T), цистеином (C) и аланином (A). Линейная последовательность нуклеотидов вирусного генома записана построчно в информационном источнике [4] в формате fasta в виде прямоугольного массива: AAGCTTA..... . Установим количественные показатели для нуклеотидов в геноме вируса. Введем обозначения:

$$m_G = G; m_T = T; m_C = C; m_A = A \quad (1)$$

$$l_G = M - m_G; l_T = M - m_T; l_C = M - m_C; l_A = M - m_A$$

Где M – полное количество нуклеотидов в геноме.

Очевидно, что количественные показатели отдельных нуклеотидов в геноме создаются на основе набора данных:

$$m = [G, T, C, A] \rightarrow m_G; m_T; m_C; m_A \rightarrow m_i \quad (2)$$

Обозначим

$$a_i = \frac{m_i}{M}; b_i = \frac{l_i}{M} \quad (3)$$

$$i = [G, T, C, A]$$

Линейный кортеж из нуклеотидов вируса обладает порождающей функцией вида [3,5]:

$$\begin{aligned} x(\tau) &= a_i + b_i(\tau_{m_i}) * \cos(\tau_{m_i}) \\ y(\tau) &= a_i + b_i(\tau_{m_i}) * \sin(\tau_{m_i}) \end{aligned} \quad (4)$$

Где

$$\begin{aligned} \tau_{m_i} &= m_i \Delta\varphi_{m_i} / \text{радиан/} \\ \Delta\varphi_{m_i} &= \frac{2\pi}{b_i} / \text{радиан/} \end{aligned} \quad (5)$$

Наличие порождающей функции позволяет создать оценку D_c – сложности строения вирусного кортежа из нуклеотидов, вычисляемую по выражению:

$$D_e = f(d_G; d_T; d_C; d_A) \quad /\text{радиан}/ \quad (6)$$

Где f – функция расстояния, задаваемая набором кортежей $d_G; d_T; d_C; d_A$. Очевидно, что при условии $f_1(d_{G1}; d_{T1}; d_{C1}; d_{A1}) = f_2(d_{G2}; d_{T2}; d_{C2}; d_{A2})$ оценка сложности строения вирусного кортежа соответствует значению $D_e=0$. Это свойство порождающей функции используем для оценки разных по этиологии штаммов коронавирусов SARS COV [6] (основа) и SARS MERS [7]. Проведем расчеты D_e при $M=136$ для 19 и 20 строки кортежа.

Таблица 1. Идентификаторы сравниваемых кортежей вируса

Наименование	Локус	Размер	Дата
Коронавирус SARS-COV, Homo sapiens	LOCUS FJ959407 /19-20 стр./	29694 bp	26-09-2009
Коронавирус MERS-COV, Homo sapiens	LOCUS NC 019843 /19-20 стр./	30119 bp	13-06-2012

Для данных кортежей имеем: $D_e= 943,489$ (радиан). Проведем расчет оценки D_e на уровне 100 и 101 строки кортежа – таблица 2.

Таблица 2. Идентификаторы сравниваемых кортежей вируса

Наименование	Локус	Размер	Дата
Коронавирус SARS-COV, Homo sapiens	LOCUS FJ959407 /100-101 стр./	29694 bp	26-09-2009
Коронавирус MERS-COV, Homo sapiens	LOCUS NC 019843 /100-101 стр./	30119 bp	13-06-2012

Для данных кортежей имеем: $D_e =918,940$ /радиан/.

Оценим сложность строения штамма вируса SARS COV 2 (LOCUS LC 546038) [8] по отношению к штамму SARS COV (LOCUS FJ959407). Проведем расчеты D_e при $M=136$ для 19 и 20 строки кортежа – таблица 3.

Таблица 3. Идентификаторы сравниваемых кортежей вируса

Наименование	Локус	Размер	Дата
Коронавирус SARS-COV, Homo sapiens	LOCUS FJ959407 /19-20 стр./	29694 bp	26-09-2009
Коронавирус SARS-COV 2 Япония Homo sapiens	LC 546038 /19-20 стр./	29876 bp	23-03-2020

Для данных кортежей имеем: $D_e =947,397$ /радиан/. Проведем расчет оценки D_e на уровне 100 и 101 строки кортежа – таблица 4.

Таблица 4. Идентификаторы сравниваемых кортежей вируса

Наименование	Локус	Размер	Дата
Коронавирус SARS-COV, Homo sapiens	LOCUS FJ959407 /100-101 стр./	29694 bp	26-09-2009
Коронавирус SARS-COV 2 Япония Homo sapiens	LC 546038 /100-101 стр./	29876 bp	23-03-2020

Для данных кортежей имеем: $De = 852,222$ /радиан/

Современные медицинские исследования используют прогрессивные технологии формирования новых знаний, в том числе и о вирусных инфекциях. В этом направлении широко и повсеместно используются компьютерные информационные технологии, которые требуют наличия базовой концепции. Декларация о наличии базовой концепции, а аннотированном виде представленной набором выражений (1 - 6), применительно к области вирусологии, актуализируют проблему получения оперативной информации о строении вирусов, на основании которой могут разрабатываться и совершенствоваться методы профилактики, обеспечивающие понижение риска распространения инфекции.

Рассмотренная в материале статьи методология формирования оперативной информации о строении коронавирусных штаммов оперирует типовыми понятиями и определениями генетики, которые дополняются обнаруженными математическими закономерностями, обеспечивающими формирование концептуального образа, посредством которого создается количественный показатель сложности конкретного штамма, естественно по отношению к некоторому базовому образу.

Наличие информационных ресурсов о генетических данных в сети Интернет, практически представленных со всех континентов, позволяет проводить сравнительный количественный анализ различных штаммов, отслеживать динамику показателей агрессивности – способности к массовому распространению в социуме.

Проведенный анализ типовых штаммов коронавирусов: SARS-COV – SARS MERS и SARS-COV2 свидетельствует о восходящей сложности строения геномов. Априорно выбираемые фрагменты генома штаммов убедительно свидетельствуют о возрастании сложности строения внутренней части генома вируса, в то время как наружная часть практически мало меняется. Представляется целесообразным рассматривать наличие такого феномена как показателя агрессивности штамма вируса. Очевидно, что формирование такого суждения создается на основе компьютерной технологии вычисления сложности строения вирусного генома. Такой показатель, вычисляемый по реальным данным генотипа вируса, обладает вариабельностью, что может служить основанием для построения прогнозов развития событий в социуме.

Список литературы / References

1. *Gorbalenya A.E., Baker S.C., Baric R.S. et al.* The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2 // *Nat. Microbiol.* 02.03.2020. DOI: 10.1038/s41564-020-0695-z.
2. *Макаров Л.М.* Информационная энтропия. International scientific review of the problems and prospects of modern science and education Collection of scientific articles LXVII International correspondence scientific and practical conference, 2020. DOI: 10.24411/2542-0798-2020-16702.
3. *Макаров Л.М.* Формализм вычисления оценки эмерджентности. Наука, техника и образование, 2020. № 1 (65). DOI: 10.24411/2312-8267-2020-10101.

4. [Электронный ресурс]. Режим доступа: <https://www.ncbi.nlm.nih.gov/genome/gdv/> (дата обращения: 17.06.2020).
5. Макаров Л.М., Поздняков А.В. Фрактальный образ генома коронавируса Covid -19 В сборнике: International scientific review of the problems and prospects of modern science and education Collection of scientific articles LXIX International correspondence scientific and practical conference, 2020. С. 6-10. DOI: 10.24411/2542-0798-2020-16902.
6. [Электронный ресурс]. Режим доступа: <https://www.ncbi.nlm.nih.gov/nuccore/FJ959407.1/> (дата обращения: 17.06.2020).
7. [Электронный ресурс]. Режим доступа: https://www.ncbi.nlm.nih.gov/nuccore/NC_019843.3?report=fasta/ (дата обращения: 17.06.2020).
8. [Электронный ресурс]. Режим доступа: <https://www.ncbi.nlm.nih.gov/nuccore/LC546038/> (дата обращения: 17.06.2020).