

Научная статья

<https://doi.org/10.36233/0372-9311-275>

Нуклеотидные тетрамеры TCGA и CTAG: вирусные ДНК и генетический код (гипотеза)

Филатов Ф.П.

Научно-исследовательский институт вакцин и сывороток им. И.И. Мечникова, Москва, Россия;
Национальный исследовательский центр эпидемиологии и микробиологии имени почетного академика Н.Ф. Гамалеи, Москва, Россия

Аннотация

Введение. Литературные и наши собственные данные показывают, что в частотных профилях (ЧП) герпесвирусных ДНК тетрануклеотиды CTAG и, в меньшей степени, TCGA выделяются среди других полных, билатерально симметричных тетрануклеотидов заметно более низкими значениями концентраций.

Цель работы — сравнительный анализ ЧП тетрануклеотидов CTAG и TCGA в вирусных ДНК.

Материалы и методы. Проанализированы ЧП и другие особенности указанных двух тетрамеров в ДНК не менее одного вида вирусов каждого рода (или субсемейства, если оно не классифицировано по родам) в соответствии с ограничениями по размеру (не ниже 100 000 пар оснований) — всего свыше 200 видов вирусов. Для анализа использованы инструменты GenBank.

Результаты. Описаны две группы формальных особенностей тетрамеров TCGA и CTAG. Одна из них относится к результатам анализа ЧП этих тетрануклеотидов в вирусных ДНК и показывает, что в ДНК с GC:AT > 2 имеют место определённые симметрии ЧП nCGn при частом нарушении таких симметрий в ЧП nTAn из-за недопредставленности CTAG. Другая группа особенностей этих тетрамеров демонстрирует различия их ЧП в полных ДНК вирусов и в их геномах (кодирующей части, которая у некоторых исследованных вирусов достигает 80%, делая анализ их ДНК более убедительным, нежели анализ ДНК клеточных форм жизни) и указывает на возможную роль этих тетрамеров в происхождении универсального генетического кода.

Обсуждение. Предполагается, что генетический код первоначально формировался на основе некоторого преобладания C+G в «до-кодовых» ДНК-полимерах с последующей эволюцией стартовых форм кода до конечной фиксированной структуры, в которой тетрамеры TCGA и CTAG занимают центральное место, отражая исходные этапы этой эволюции. Симметрии ЧП nCGn, характерные для «полной» ДНК герпесвирусов рода Simplex, исчезают в цепи вторых кодонных букв генома этих вирусов, косвенно указывая на отличия их функций от функций других букв и подчёркивая целесообразность представления генетического кода в формате каллиграммы, в которой вторая строка не симметрична.

Ключевые слова: вирусная ДНК, функции тетрамеров TCGA и CTAG, частотный профиль nCGn и nTAn, симметрии генетического кода

Источник финансирования. Автор заявляет об отсутствии внешнего финансирования при проведении исследования.

Конфликт интересов. Автор декларирует отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Для цитирования: Филатов Ф.П. Нуклеотидные тетрамеры TCGA и CTAG: вирусные ДНК и генетический код (гипотеза). *Журнал микробиологии, эпидемиологии и иммунобиологии*. 2022;99(4):478–493. DOI: <https://doi.org/10.36233/0372-9311-275>

Original article

<https://doi.org/10.36233/0372-9311-275>

Nucleotide tetramers TCGA and CTAG: viral DNA and the genetic code (hypothesis)

Felix P. Filatov

I. Mechnikov Research Institute of Vaccines and Sera, Moscow, Russia;
National Research Center for Epidemiology and Microbiology named after Honorary Academician N.F. Gamaleya, Moscow, Russia

Abstract

Introduction. The published and our own data show that CTAG and, to a lesser extent, TCGA tetra-nucleotides have significantly lower concentrations in frequency profiles (FPs) of herpesvirus DNAs compared to other complete, bilaterally symmetrical tetra-nucleotides.

The aim of the study is to present a comparative analysis of CTAG and TCGA tetra-nucleotide FPs in viral DNAs.

Materials and methods. We have analyzed FPs and other characteristics of the two above tetramers in DNAs of at least one species of viruses of each genus (or each subfamily, if the classification into genera was not available), complying with the size limit requirements (minimum 100,000 base pairs) — a total of more than 200 species of viruses. The analysis was performed using the GenBank database.

Results. Two groups of characteristics of TCGA and CTAG tetramers have been described. One of them covers the results of the FP analysis for these tetranucleotides in viral DNAs and shows that DNAs with GC:AT > 2 are characterized by nCGn FP symmetries while these symmetries are frequently distorted in nTAn FP due to CTAG underrepresentation. The other group of tetramer characteristics demonstrates differences in their FPs in complete viral DNAs and in their genomes (a coding part, which can reach 80% in some studied viruses, thus making the analysis of their DNAs more significant than the analysis of DNAs of cellular live forms) and suggests that these tetramers may have participated in the origin of the universal genetic code.

Discussion. Assumedly, the genetic code started evolving amid C+G prevailing in "pre-code" DNA polymers; then the initial code forms evolved further to their final structure where TCGA and CTAG tetramers hold a central position, encapsulating the previous stages of this evolution. The nCGn FP symmetries typical of the "complete" DNA of Herpes simplex viruses disappear in the sequence of the second codon letters of the genome of these viruses, implying that their functions differ from functions of other letters and emphasizing the reasonableness of presenting the genetic code as a calligram where the second line is not symmetrical.

Keywords: *viral DNA, functions of TCGA and CTAG tetramers, frequency profile of nCGn and nTAn, symmetries of the genetic code*

Funding source. This study was not supported by any external sources of funding.

Conflict of interest. The author declares no apparent or potential conflicts of interest related to the publication of this article.

For citation: Filatov F.P. Nucleotide tetramers TCGA and CTAG: viral DNA and the genetic code (hypothesis). *Journal of microbiology, epidemiology and immunobiology = Zhurnal mikrobiologii, èpidemiologii i immunobiologii.* 2022;99(4):478–493.

DOI: <https://doi.org/10.36233/0372-9311-275>

Введение

Ранее мы описали частоту встречаемости билатерально симметричных, полных (состоящих из 4 оснований) тетрануклеотидов (ТН) в геномах герпесвирусов [1]. Обнаружив тогда, что частотные профили (ЧП) двух ТН — СТАГ и, в меньшей степени, ТСГА — герпесвирусных ДНК выделяются заметно низкими значениями концентраций, на что указывают и данные литературы [2–4], мы более внимательно проанализировали другие особенности этих двух ТН и вывели такой анализ за пределы герпесвирусов.

Предполагается, что действие СТАГ связано с нарушением оптимальной структуры створочных петель нуклеиновой кислоты, что приводит к торможению репликации ДНК («термодинамическая модель»). Кроме того, последовательность СТАГ более подвержена химическим воздействиям [5, 6]. Пониженная концентрация ТСГА связана с его центральным димером СrG, который часто метилируется и отличается высокой частотой мутаций [7–10].

Мы приводим здесь многочисленные ссылки на источники (на самом деле работ, посвящённых этой теме, намного больше), чтобы показать разнообразие последствий наличия обсуждаемых олигонуклеотидов в ДНК и геномах живых организмов [11]. Нежелательное ингибирование биологических синтезов компенсируется снижением концентрации в

ДНК обоих ТН. Нас в данной работе интересовали в первую очередь формальные характеристики обоих ТН, которые, в отличие от остальных, обладают биологическими функциями, какими бы ни были функции и механизмы, их определяющие. Эти характеристики демонстрируют неожиданные особенности, объяснение которых изложено здесь в порядке весьма предварительной гипотезы.

Цель работы — сравнительный анализ ЧП СТАГ и ТСГА в вирусных ДНК и в геномных участках этих ДНК.

Мы анализировали ближайший контекст центральных пар СG и ТА нуклеотидных тетрамеров, среди которых — ТСГА и СТАГ, в ДНК вирусов различных классификационных групп. Такой подход несколько уравнивает оценку сравнения ЧП динуклеотида СG и СТАГ, сближая их размеры и рассматривая их одновременно и как тетрамеры, и как димеры (тем более что в литературе отмечаются также сходные по функции, пусть и в значительно меньшей степени, характеристики центрального димера СТАГ [12, 13]). Минус этого подхода заключается в значительно меньшем различии плотности симметричных пар тетрамеров с общей функцией (ТСГА и АCGT) по сравнению с существенной разницей плотности симметричных пар тетрамеров, имеющих (СТАГ) и практически не имеющих (GTAC) такую функцию.

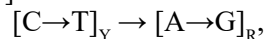
Материалы и методы

Аналізу подвергались физически несегментированные ДНК, полноразмерный сиквенс которых представлен в GenBank¹ на 2021 г. Третий ограничивающий фактор — размер ДНК, который в первом приближении должен быть не ниже 100 т.п.н., как и в случае со Вторым правилом Чаргаффа [5, 14, 15], и не выше 300–400 т.п.н. ДНК последних отличается слишком сложные вирусы и содержит преимущественно А+Т. Самые большие известные вирусные РНК — геномы коронавирусов — составляют не более 32–35 т.п.н.

Указанным выше условиям удовлетворяют геномы вирусов только двух больших доменов общего надцарства *Vira*: *Duplodnaviria* (царство *Heunggongvirae*) и *Varidnaviria* (царство *Bamfordvirae*). Мы проанализировали ДНК не менее одного вида вирусов каждого рода (или семейства, если оно не классифицировано по родам); всего таких родов — свыше 200 (20 семейств). Исследованные нами вирусы первого домена принадлежат к типам *Uroviricota* царства *Heunggongvirae* (отряд *Caudovirales*) и *Peploviricota* того же царства (отряд *Herpesvirales*). Исследованные вирусы второго домена принадлежат к типу *Nucleocytoviricota* классов *Megaviricetes* и *Pokkesviricetes*. Кроме того, мы проанализировали ДНК вирусов без установленных промежуточных доменов: 9 представителей семейств *Baculoviridae*, *Nudiviridae* и суперсемейства *Nimaviridae*, а также 6 представителей неклассифицированных видов вирусов архей и 3 неклассифицированных вида семейств *Pytho-* и *Hytrosaviridae* (Приложение).

В качестве инструментов анализа использовали пакеты программ GenBank.

Графики частотного распределения изучаемых ТН строили перебором вариантов ближайшего контекста центральных пар пТАн и пСГн (СТАГ) с последовательным повышением молекулярной массы п [16, 17]:



где С, Т — пиримидины (_Y); А, G — пурины (_R).

Результаты

Плотность минимальных пТАн и пСГн в вирусных ДНК

Общий анализ плотности пТАн и пСГн показал, что в ДНК большей части (75 из 128) исследованных представителей типа *Uroviricota* минимальный ТН — это СТАГ_{min}. Минимальным мы называем здесь полный самокомплементарный тетрамер, плотность которого ниже, чем плотность симметричного ему по контекстным основаниям тетра-

мера в общем ЧП вирусной ДНК. В нашем случае СТАГ_{min} < GTAC и TCGA_{min} < ACGT.

ДНК ни одного вида типа *Uroviricota* не содержит TCGA в качестве минимального тетрамера. В ДНК вирусов типа *Nucleocytoviricota*, вирусов неутонченной классификации (*Baculoviridae*, *Nudiviridae* и *Nimaviridae*) и вирусов архей TCGA_{min}, хотя и встречается, но редко и без видимой связи с классификационными группами.

Один из иридовирусов — альфа-иридовирус, а также вирус инфекционного некроза селезенки и почек человека содержат в своей ДНК оба названных ТН в качестве минимальных и примерно в равных концентрациях (СТАГ_{min} ~ TCGA_{min}). Та же особенность относится и к вирусу *Ranid 1* семейства *Alloherpesviridae*. ДНК значительной части альфа-герпесвирусов человека рода *Simplex* содержит эти тетрамеры в качестве минимальных при СТАГ_{min} < TCGA_{min}.

TCGA не является минимальным в ДНК розеоновирусов. В то же время большая часть герпесвирусов ДНК (26 из 35) — кроме гамма-герпесвирусов — имеет в качестве минимального СТАГ. Раздел 2 детализирует особенности ЧП пТАн и пСГн в ДНК герпесвирусов.

Минимальной концентрацией СТАГ отличаются все исследованные *Nucleocytoviricota* — за исключением поксвирусов, среди которых только 3 хордопоксвируса (из 19 анализированных) имеют СТАГ_{min}. Поксвирусы, как было сказано, отличаются также ДНК с преобладающим типом А+Т и с высоким (> 2) соотношением [Т+А] : [G+C].

Подытоживая эту часть работы, отметим следующее:

1) ДНК герпесвирусов выделяется среди ДНК других исследованных здесь вирусов соотношением [G+C] > [A+T];

2) среди изученных вирусных ДНК тетрамер TCGA характерен как минимальный, почти исключительно для ДНК герпесвирусов — с более чем двойным преобладанием G+C над А+Т. Это в первую очередь герпесвирусы рода *Simplexviruses* субсемейства *Alpha* и частично рода *Lymphocryptoviruses* субсемейства *Gamma*. ДНК многих герпесвирусов имеют в качестве минимального тетрамера ACGT, но он не уникален, т.е. встречается в этом качестве в ДНК других классификационных групп вирусов.

Частотный профиль тетрапуклеотидов в вирусных ДНК

Яркой характеристикой количественного распределения геномных тетрамеров пСГн некоторых вирусов типа *Peploviricota* является симметрия их ЧП. **Рисунок 1** демонстрирует её для вируса простого герпеса-1, чей геном организован по типу D [18]. При этом ЧП пТАн асимметричен в силу «минимальности» СТАГ (СТАГ_{min}). Более строгим, чем

¹ URL: https://www.ncbi.nlm.nih.gov/genomes/Genomes_Group.cgi?taxid=10239&sort=taxonomy

симметрия, но также её отражающим является соотношение $CTAG : GTAC$ и $TCGA : ACGT$. Для ДНК вируса простого герпеса-1 — это $CTAG_{min} < GTAC$ (21%) и $TCGA_{min} < ACGT$ (86%).

Достаточно заменить центральный димер тетрамера $CTAG$ на обратный — $CATG$, как симметрия соответствующего ЧП восстанавливается. Асимметрия ЧП $nCGn$ выражена намного слабее, хотя для ЧП многих симплексовирусов соотношение $TCGA < ACGT$ весьма характерно. Любые замены центрального димера CG приводят к сильной деформации и нарушению симметрии ЧП.

На **рис. 2** показаны ЧП ТН $nTAn$ и $nCGn$ в *Human Cytomegalovirus (HHV5)* ДНК, организованной, как и в случае герпесвирусов рода *Simplex*, по

типу D. Здесь симметрии обоих ЧП практически не выражены, однако соотношения $CTAG_{min} : GTAC$ и $TCGA_{min} : ACGT$ вполне отчётливы (36 и 52% соответственно). Очевидно, что симметрия ЧП $nCGn$ связана с GC-типом ДНК герпесвируса рода *Simplex*.

Количественное выражение $GC : AT$ и связь его с симметриями ЧП герпесвирусовых ДНК иллюстрирует **таблица**. Выраженные симметрии ЧП $nCGn$ обеспечивает соотношение $GC : AT > 2$, что характерно в первую очередь для ДНК герпесвирусов рода *Simplex*. Соотношения пар $TCGA < ACGT$ и $CTAG < GTAC$ характерны в основном также для ДНК вирусов рода *Simplex*. AT-тип вирусных ДНК при соотношении $AT : GC > 2$ (поксвирусы) тоже придаёт симметрию ЧП обсуждаемых ТН.

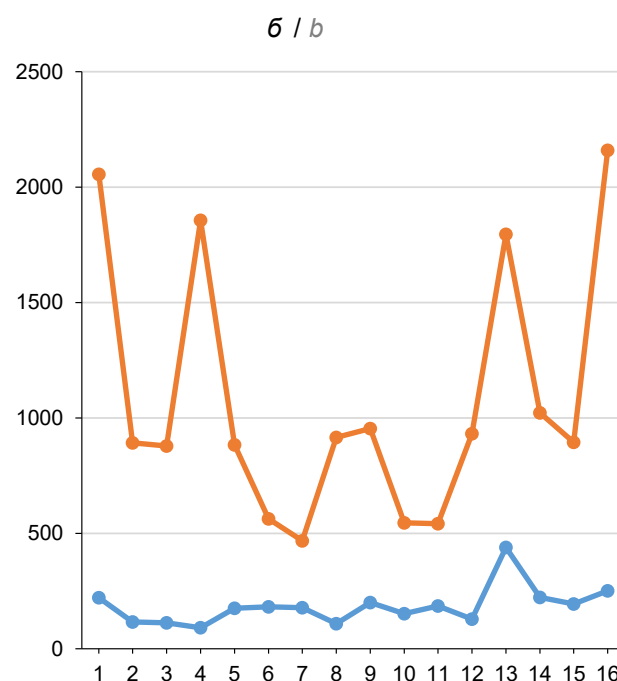
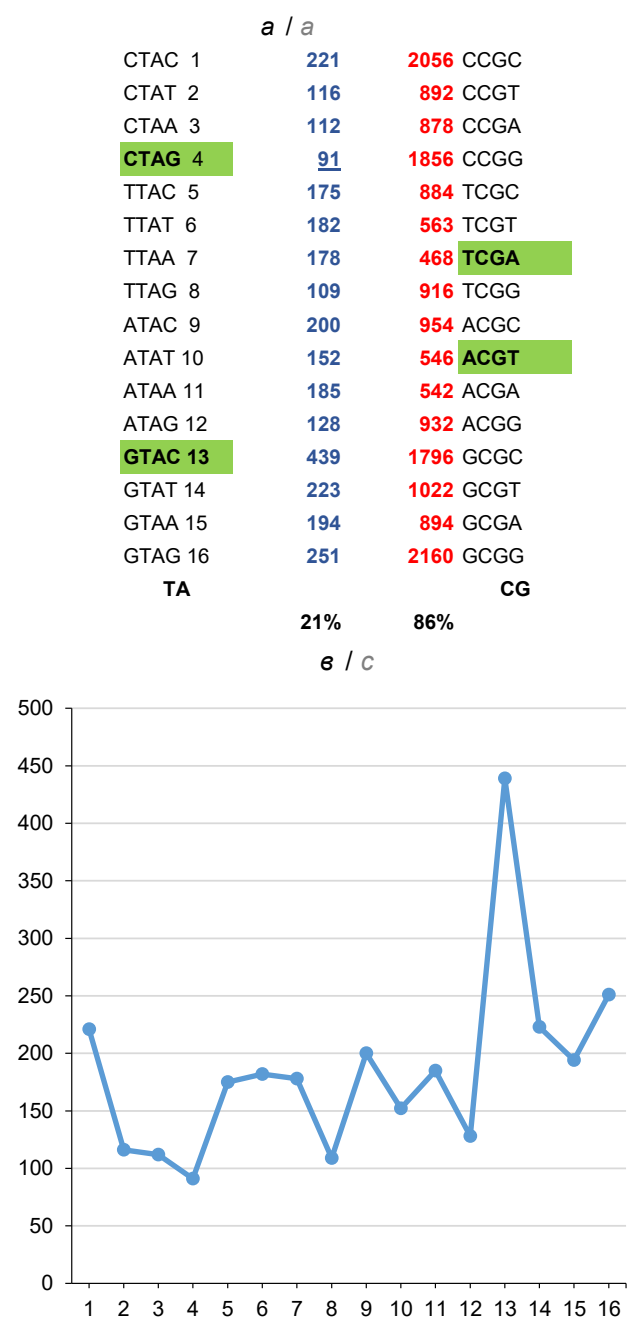


Рис. 1. ЧП тетрануклеотидов $nTAn$ (синий цвет) и $nCGn$ (красный цвет) в *Human Simplexvirus 1 DNA*.

a — абсолютные значения, приведено процентное соотношение симметричных по шкале пар двух полных тетрамеров — $TCGA_{min} : ACGT$ и $CTAG_{UNP} : GTAC$ (окрашены зелёным и выделены жирным шрифтом); б, в — графическое выражение абсолютных значений; в — ЧП $nTAn$, больший масштаб по вертикали.

Fig. 1. FPs of $nTAn$ (blue) and $nCGn$ (red) tetra-nucleotides in *Human Simplexvirus 1 DNA*.

a — absolute values; the percentage ratio of symmetric pairs of two complete tetramers — $TCGA_{min} : ACGT$ and $CTAG_{UNP} : GTAC$ (shown in green and bold); b, c — a graphic representation of absolute values; c — $nTAn$ FP; vertical scale up.

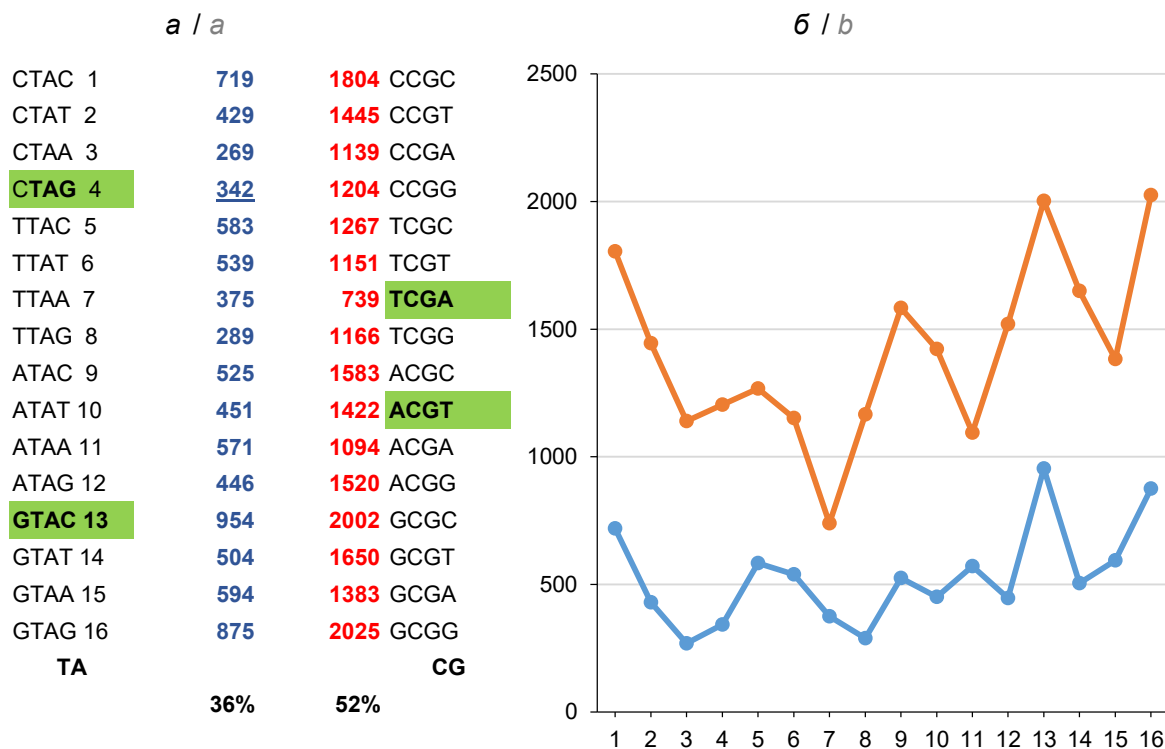


Рис. 2. Частотные профили тетрануклеотидов nTAn (синий цвет) и nCGn (красный цвет) в *Human Cytomegalovirus (HHV5) DNA*.

a — абсолютные значения; б — их графическое выражение. Показано процентное соотношение симметричных по шкале пар двух полных тетрамеров, ACGT : TCGA и GTAC : CTAG (окрашены зелёным и выделены жирным шрифтом).

Fig. 2. Frequency profiles of nTAn (blue) and nCGn (red) tetra-nucleotides in the *Human Cytomegalovirus (HHV5) DNA*.

a — absolute values; b — their graphic representation. The percentage ratio of symmetrical pairs of two complete tetramers, ACGT:TCGA and GTAC:CTAG (shown in green and in bold).

Рисунок 3 показывает ЧП CTAG и TCGA в ДНК *Ranid herpesvirus 1, (Batrachovirus)*, в которой оба ТН находятся в минимальных позициях своих ЧП, а геном организован по типу В Ройзмана. ЧП обоих ТН в ДНК этого вируса симметрий не демонстрируют, но весьма отчётливо выделяют снижением плотности так называемые неполные тетрамеры — тримеры СТА/TAG и CCG/CGA. В определённой мере это характерно и для ряда других вирусных ДНК. Плотность таких тринуклеотидов, как правило, не достигает значений полных CTAG или TCGA.

Подытоживая эту часть работы, отметим две выраженные, не упоминавшиеся прежде формальные особенности CTAG и TCGA.

ЧП nCGn в ДНК $[G+C] : [A+T] > 2$ отличаются определённой симметрией, в то время как симметрия ЧП nATn в таких ДНК часто нарушена (CTAG < GTAC). Эта симметрия не тождественна Второму правилу Чаргаффа, которое, по крайней мере, не имеет таких ограничений и даже не является следствием этого правила, как это может показаться на первый взгляд. Она также отлична от типа симметрий ДНК-последовательностей, описаного в работах [19–21].

Не только CTAG, но и его тринуклеотидные взаимоперекрывающиеся участки nTAG и СТАn имеют, как правило, более или менее выраженную тенденцию к снижению плотности в ЧП соответствующих тетрамеров (рис. 2). Тенденцию к снижению плотности демонстрируют также тримеры nCGA и TCGn. Если учитывать в качестве минимальных тримеры nTAG или СТАn, перекрывающиеся 5'- или 3'-участки $CTAG_{min}$, то упомянутое в разделе 1 в отношении ДНК исследованных представителей типа *Uroviricota* число 75 $CTAG_{min}$ из 128 исследованных увеличивается до 93. Такие тринуклеотиды также имеют тенденцию к снижению частоты, которая, как правило, не достигает минимальных значений полного $CTAG_{min}$. На наш взгляд, это требует некоторого уточнения упомянутой выше гипотезы термодинамической модели, которая относится к полному ТН CTAG.

Для этих четырех тринуклеотидов (СТА/TAG, частично перекрывающихся CTAG, и TCG/CGA, частично перекрывающихся TCGA) характерна ещё одна совершенно неожиданная особенность, которая, на первый взгляд, не связана с их известными функциями; фактически она относится к следующему разделу и здесь используется лишь как

Характеристики 1-х (А) и 3-х (G) строк вирусных (в основном герпесвирусных) геномов с соотношением GC : AT > 1
 Characteristics of the 1st (A) and 3rd (G) lines of virus (mainly herpesvirus) genomes with ratio GC : AT > 1

Род / Genus	Вид / Species	GC : AT	ДНК / DNA	Геном / Genome	
				1-й ТН / 1 st nt	3-й ТН / 3 rd nt
<i>Simplexvirus</i>	<i>Papiine HV2</i>	3,18	acgt	TCGA	–
<i>Simplexvirus</i>	<i>Cercopithec aHV2</i>	3,16	acgt	TCGA	TCGA
<i>Simplexvirus</i>	<i>Ateline aHV1</i>	3,08	acgt	TCGA	TCGA
<i>Simplexvirus</i>	<i>Macacine aHV1</i>	2,92	acgt	TCGA	TCGA
<i>Simplexvirus</i>	<i>Human aHV2</i>	2,38	acgt	TCGA	TCGA
<i>Simplexvirus</i>	<i>Human aHV1</i>	2,15	TCGA	TCGA	TCGA
<i>Simplexvirus</i>	<i>Chimpanzee aHV</i>	2,13	TCGA	TCGA	TCGA
<i>Simplexvirus</i>	<i>Saimiriine aHV1</i>	2,04	TCGA	TCGA	TCGA
<i>Quwivirus</i>	<i>Tupaia bHV1</i>	2,00	acgt	acgt	acgt
<i>Simplexvirus</i>	<i>Leporid aHV4</i>	1,97	TCGA	TCGA	acgt
<i>Shapirovirus</i>	<i>Caulobacter phage CcrKarma</i>	1,96	acgt	TCGA	TCGA
<i>Phicbkvirus</i>	<i>Caulobacter virus Rogue</i>	1,95	acgt	acgt	TCGA
<i>Bertelyvirus</i>	<i>Caulobacter phage CcrBL9</i>	1,68	acgt	TCGA	TCGA
<i>Cytomegalovirus</i>	<i>Rat Maastricht</i>	1,56	acgt	TCGA	acgt
<i>Simplexvirus</i>	<i>Fruit bat aHV1</i>	1,55	acgt	TCGA	TCGA
<i>Lymphocryptovirus</i>	<i>Human bHV4</i>	1,47	TCGA	TCGA	TCGA
<i>Muromegalovirus</i>	<i>Murid bHV1</i>	1,42	TCGA	TCGA	acgt
<i>Cytomegalovirus</i>	<i>Human bHV5</i>	1,35	TCGA	TCGA	acgt
<i>Rhadinovirus</i>	<i>Dolphin gHV1</i>	1,32	acgt	acgt	TCGA
<i>Varicellovirus</i>	<i>Equid aHV1</i>	1,31	TCGA	TCGA	TCGA
<i>Quiwivirus</i>	<i>Caviid bHV2</i>	1,22	TCGA	TCGA	acgt
<i>Percavirus</i>	<i>Equid gHV5</i>	1,21	TCGA	TCGA	TCGA
<i>Batrachovirus</i>	<i>Ranid HV1</i>	1,20	TCGA	TCGA	TCGA
<i>Rhadinovirus</i>	<i>Human gHV8</i>	1,16	acgt	acgt	acgt
<i>Mardivirus</i>	<i>Gallid aHV3</i>	1,16	acgt	TCGA	TCGA
<i>Simplexvirus</i>	<i>Macropodid HV1</i>	1,12	acgt	TCGA	TCGA
<i>Batrachovirus</i>	<i>Ranid HV2</i>	1,12	TCGA	TCGA	TCGA
<i>Cyprinivirus</i>	<i>Cyprinid HV1</i>	1,05	acgt	acgt	TCGA

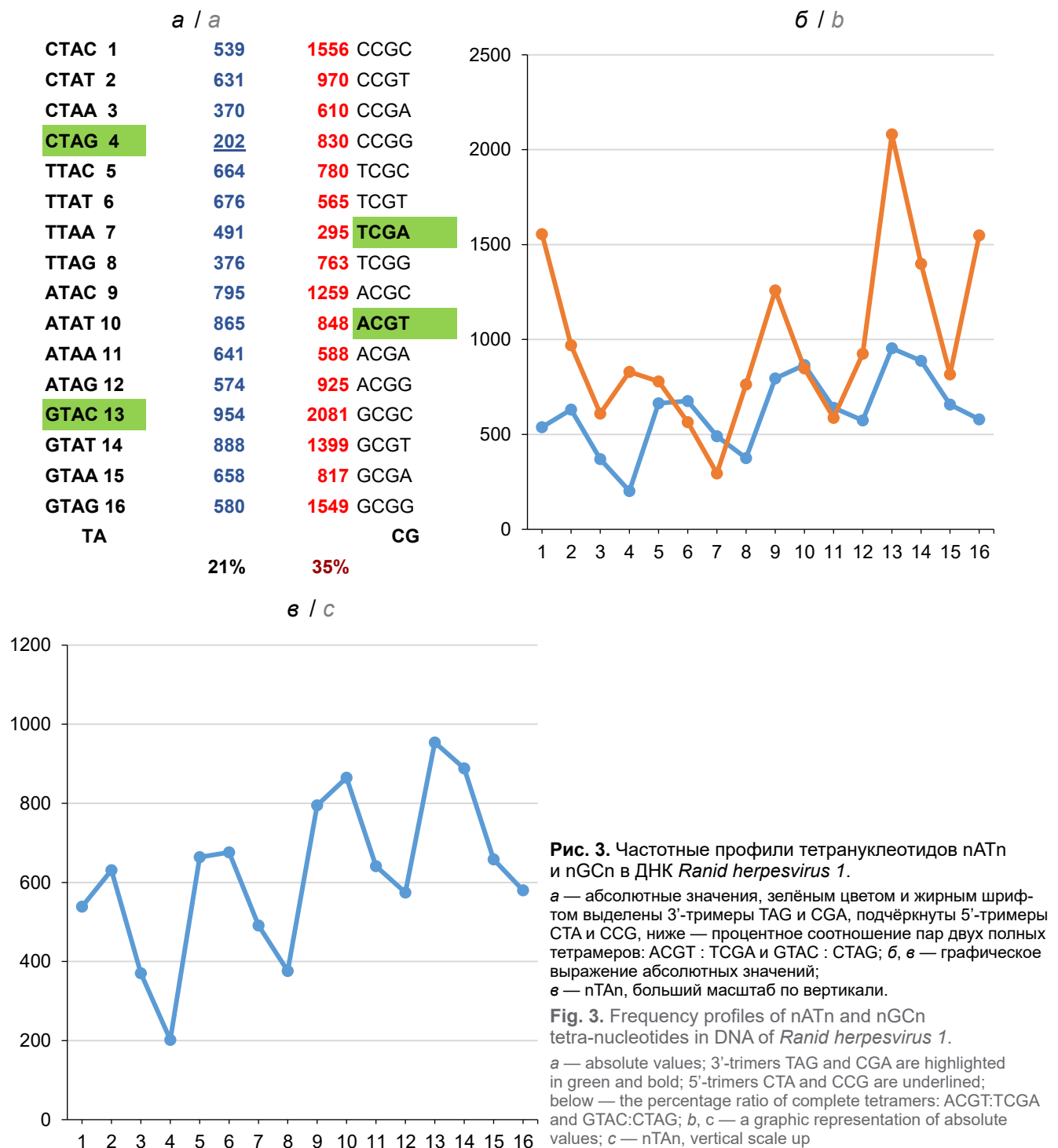
Примечание. Минимальные тетра-нуклеотиды ЧП nCGn показаны как ACGT или TCGA. Наличие симметрии ЧП обозначено серым цветом ячеек (пояснения в тексте).

Note. Tetra-nucleotides (nCGn) of the minimum concentration are shown as ACGT or TCGA. The FP symmetry is shown in grey color (see the explanation in the text).

переход к нему: рассматриваемые как взаимно перекрывающиеся кодоны, эти 4 тримера полностью исчерпывают избыточность универсального генетического кода (выделено жирным шрифтом; римские цифры указывают на группу их вырожденности): **CTA_{IV} = TTR_{II}(L)** и **TAG_{II} = TGA_{III}(stop)**; **TCG_{IV} = AGY_{II}(S)** и **CGA_{IV} = AGR_{II}(R)**. При этом группа их вырожденности всегда выше группы вырожденности альтернативных кодонов той же аминокислоты.

Альтернативные кодоны составляют симметричную центральную строку А-Т-Т-Т-А (первые

буквы) или S_{II}-stop-L_{II}-stop-R_{II} (продукты кодирования) так называемой матрицы кода [17]. Другой такой, как TCGA и СТАG, пары полных самокомплементарных ТН, которые перекрывались бы три-нуклеотидами, обладающими подобными свойствами, в генетическом коде нет. Эта их особенность, возможно случайная, тем не менее заставляет внимательнее пересмотреть в описанном контексте структуру универсального генетического кода и поделиться столь же неожиданными наблюдениями, не всегда объяснимыми, но на наш взгляд, заслуживающими интереса.



Универсальный генетический код и геномные симметрии тетрануклеотидов TCGA и CTAG

Генетический код имеет помехоустойчивую структуру, основой чего является, в частности, симметрия его элементов. Один из наиболее демонстрационных вариантов такой симметрии (исторически первый) представляет таблица Ю. Румера [22, 23], позднее преобразованная В. Щербаком в каллиграмму, которая связывает симметрией первые

буквы кодирующих триплетов и упорядоченные по молекулярной массе продукты кодирования [16]. Такая связь до сих пор не имеет внятного объяснения. В представленной здесь несколько модифицированной таблице (каллиграмма А; рис. 4, а) мы отдали приоритет этапам формирования кода — в противовес уже сформированным, фиксированным группам вырожденности, которые показывает исходная каллиграмма. Это подчёркивает преимущественное содержание G+C в первом

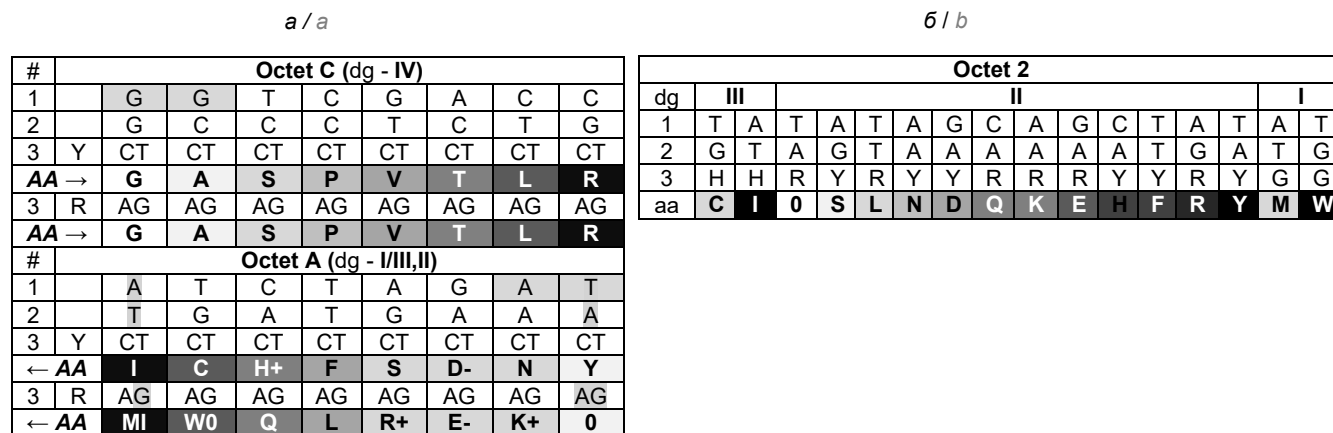


Рис. 4. Каллиграмма А универсального генетического кода (а) и октет 2 каллиграммы универсального генетического кода ([16]; б).

Третий нуклеотид кодона представлен пурином (R) или пиримидином (Y). Стартовый кодон ATG и стоп-кодон TA_R выделены серым по вертикали. Центральные тетра-нуклеотиды первых букв кодонов каждого октета выделены жирным шрифтом. Тетрануклеотид стыка октетов А и С также выделен серым. Последовательное нарастание молекулярной массы продуктов кодирования (AC, аминокислоты) показано нарастанием плотности фона клеток от белого к чёрному и стрелками. Римские цифры — группы вырожденности кода, dg. Три пары продуктов кодирования, часть которых может нести заряд (отчего их позиции в строке неустойчивы и фиксированы по доминирующему правилу — симметрии первых букв кодонов), выделены общим светло-серым фоном.

Fig. 4. Calligram A of the universal genetic code (a) and octet 2 of the calligram of the universal genetic code ([16]; b).

The third nucleotide of the codon is represented by purine (R) or pyrimidine (Y). The start codon ATG and the stop codon TA_R are highlighted vertically in grey. The first letters of central tetra-nucleotide codons of each octet are shown in bold. The tetra-nucleotide of the junction of octets A and C is also highlighted in grey. The successive increase in the molecular weight of the coding products (AC, amino acids) is shown by the increasing background density from white to black and by arrows. Roman numerals are used to denote degeneracy groups of the code, dg. Three pairs of coding products, some of which can bear a charge (that is why their positions in the lines are not stable and fixed following the dominant rule — the symmetry of the first letters of codons), highlighted in light grey.

октете, вероятно, отражающем преимущественное содержание более термостабильных пар G≡C в «до-кодовом» наборе полинуклеотидов, и A+T — во втором, в котором закреплены также направления чтения генов и другие особенности, которые делают второй октет более сложным, т.е., скорее всего, эволюционно более поздним. В нашей версии октет 2 каллиграммы (рис. 4, б) «уплотнён» за счёт того, что в основу последовательности аминокислот положена общая масса продуктов, кодируемых триплетами с третьим пиримидином Y или пурином R. Этот октет обозначен как октет А (рис. 4, а) — по преимущественному суммарному содержанию нуклеотидов А в 1-й и 2-й кодирующих строках. На том же основании октет 1 каллиграммы обозначен здесь как октет С. Количество всех четырех нуклеотидов в первых строках каждого октета равно и подчёркивает таким образом межоктетную симметрию.

Основу организации октета С каллиграммы А составляет последовательное изменение (нарастание) значений молекулярных масс кодируемых аминокислот, которое придаёт октету симметрию нуклеотидов верхней строки (первых букв кодонов). Центром этой симметрии является тетрамер TCGA (выделен жирным шрифтом на сером фоне). Мы опускаем здесь значения молекулярных масс продуктов кодирования, их можно найти в работах [15, 16]. Октеты С и I обеих каллиграмм полностью совпадают.

Основу организации октета А нашей каллиграммы также составляет последовательное изменение (но убывание) значений молекулярных масс кодируемых аминокислот, которое придаёт этому октету симметрию нуклеотидов верхней строки (первых букв кодонов). Некоторую неопределённость весьма близким значениям масс аминокислот, составляющим три пары — R⁺S, E⁻D⁻ и K⁺N (на рис. 4 выделены общим светло-серым цветом), придаёт их колебание за счёт способности нести заряд (протонирования). Однако доминирующий принцип организации октетов, а именно симметрия первых букв кодирующих триплетов, фиксирует центральный тетрамер СТАG в составе верхней строки октета А, помещая в четвертую позицию тетрамера глутаминовую кислоту. Возможно, некоторую роль играет здесь симметрия зарядов гистидина (H⁺) и глутамина (D⁻) при нейтральном pH, в кодонах которых третьи буквы — пиримидины. Два эти фактора — последовательное изменение молекулярной массы кодируемых продуктов и симметрия первых букв кодонов — определяют направление чтения этого октета и направление чтения генов — от триплета ATG (старт-кодон) к триплетам TGA и TA_R (стоп-кодоны).

Оба октета генетического кода могут отражать также его предполагаемую эволюцию [7] — от случайных старт/стоп-кодонов октета С к фиксированным (октет А) кодонам и от преобладания в октете С нуклеотидов G и C к выравниванию их наличия за

счёт доминирования нуклеотидов А и Т в кодонах октета А и к общему усложнению этого октета.

Подход, упомянутый в конце предыдущего раздела, т.е. взаимное перекрывание тетрамеров тримерами, делает каллиграмму несколько более информативной, выделяя также нечётные группы вырожденности. Линейный четырехнуклеотидный «стык» первых строк октетов А и С, т.е. АТ|GG, можно рассматривать как частично перекрывающийся кодонами АТG и TGG группы вырожденности I (обозначено серым на рис. 4, а). Наличие такого стыка подчёркивает разнонаправленность организации октетов А и Cght, формирующей их симметрию — снижение или нарастание нуклонных масс продуктов кодирования при однонаправленных центральных октетных тетрамерах.

Анализ ЧП тетрамеров nCGn и nTAn в условных цепях 1, 2 и 3-х строк вирусных геномов указывает на определённое сходство с симметриями этих ЧП в 1, 2 и 3-х строках генетического кода. Первая из этих цепей начинается с нуклеотида А, вторая — с Т, третья — с G, а гены следуют один за другим без промежутков, независимо от того, в какой цепи вирусных ДНК они локализируются, перекрывают ли друг друга и содержат ли интроны. Примеры этого анализа показаны только для геномов тех же вирусов, что и в предыдущем разделе: HHV1 (рис. 5), HHV5 (рис. 6) и HV Randid 1 (рис. 7).

Рисунок 5, а демонстрирует результат такого анализа, а именно симметрию профиля цепи первых нуклеотидов генома HHV1, в котором тетрамер TCGA остается минимальным, уступая тетрамеру ACGT. Симметрия вторых нуклеотидов отсутствует, как и во 2-й строке каллиграммы кода. Оба этих факта соответствуют функциям 1-х и 2-х нуклеотидов кодона, а их характер (наличие и отсутствие симметрии) следует организации универсального генетического кода. Выраженная симметрия профиля цепи 3-х нуклеотидов генома, которая, на первый взгляд, не является необходимой, поскольку выбор нуклеотидов для неё произволен, наводит на мысль о необходимой компенсации симметрии ЧП цепи первых нуклеотидов генома и ЧП nCGn реальной ДНК HHV1 (рис. 1). Кроме того, сходной симметрией могли характеризоваться нуклеотидные полимеры на этапе, предшествующем формированию генетического кода, или отобранные для его формирования. Анализ полной герпесвирусной ДНК, разделённой на три цепочки по тому же принципу, что и геном, возвращает ЧП nCGn статистический характер, т.е. сходную симметрию по изучаемым тетрамерам для всех 3 цепочек, не привязанных к генам.

Рисунок 5, б показывает сильно искажённую симметрию ЧП nTAn цепочки 1-х букв кодонов — явно в силу небольшого числа тетрамеров nTAn. В принципе то же можно было бы сказать и о TCGA,

но его функциональный димер (CG) даже в прерывистой цепи встречается намного чаще, чем функциональный тетрамер СТАG, и может сохранять иллюзию функции в случае тетрамера скорее, чем в случае декамера CnnTnnAnnG. Надо отметить, что минимальное значение сохраняет и декамер TnnCnnGnnA. Дефицит СТАG в цепочке 1-х букв кодирующих триплетов исчезает, хотя значение СТАG несколько уступает симметричному по шкале GTAC. Как и в случае nCGn (рис. 5, а), ЧП nTAn вторых букв не показывает симметрии, а значения перебираемых тетрамеров 3-й цепочки настолько низки, что ими можно было бы пренебречь, тем не менее они следуют порядку значений 1-х букв и, возможно, тоже участвуют в формировании общей симметрии ЧП nTAn реальной ДНК HHV1.

Рисунок 6 показывает, что ЧП цепочки 1-х и 3-х букв тетрамера nCGn (в меньшей степени nATn) генома бета-герпесвируса HHV5 «восстанавливают» симметрию, отсутствующую в реальной ДНК этого вируса, но теряют характеристики СТАG и TCGA как минимальных. Как и в геноме HHV1, в геноме HHV5 ЧП 2-х букв обоих тетрамеров несимметричен.

Рисунок 7 демонстрирует ЧП обсуждаемых тетрамеров в цепочках 1, 2 и 3-х букв генома аллогерпесвируса *Randid herpesvirus 1*. В какой-то мере и здесь мы видим «восстановление» симметрии ЧП цепочек 1-х и 3-х букв тетрамера nCGn и в меньшей степени — nATn генома по сравнению с этими характеристиками в реальной вирусной ДНК. Характеристики тетрамеров СТАG и TCGA как минимальных при этом сохраняются, хотя их значения сильно снижены.

Полученные данные мы свели в итоговую таблицу с обобщёнными данными по ЧП nCGn в ДНК вирусов, у которых соотношение $[G+C] : [A+T] > 1,0$. В первую очередь это герпесвирусы. Выделены две характеристики: род герпесвирусов *Simplex*, для ДНК вирусов которого характерен $TCGA_{min}$, и симметрия соответствующего профиля. В значительно большей мере эти характеристики отличаются геномы симплексов, точнее цепочки их 1-х (и 3-х) кодонных нуклеотидов.

Возникает вопрос о ДНК с типом АТ и с подобным высоким соотношением $[A+T] : [G+C]$. Среди исследованных нами вирусов такое соотношение большей частью относится к ДНК и геномам поксвирусов. Сходство с герпесвирусами обнаружено только у рисунка симметрий ЧП анализируемых тетрамеров, да и то только когда порядок перебора краевых оснований четвёрки меняется с СТАG на TCGA.

Подытоживая эту часть работы, отметим несколько формальных особенностей тетрамеров TCGA и СТАG в связи со структурой генетического кода.

Центральными тетрамерами 1-х строк октетов каллиграммы А являются TCGA (октет С) и СТАG (октет А).

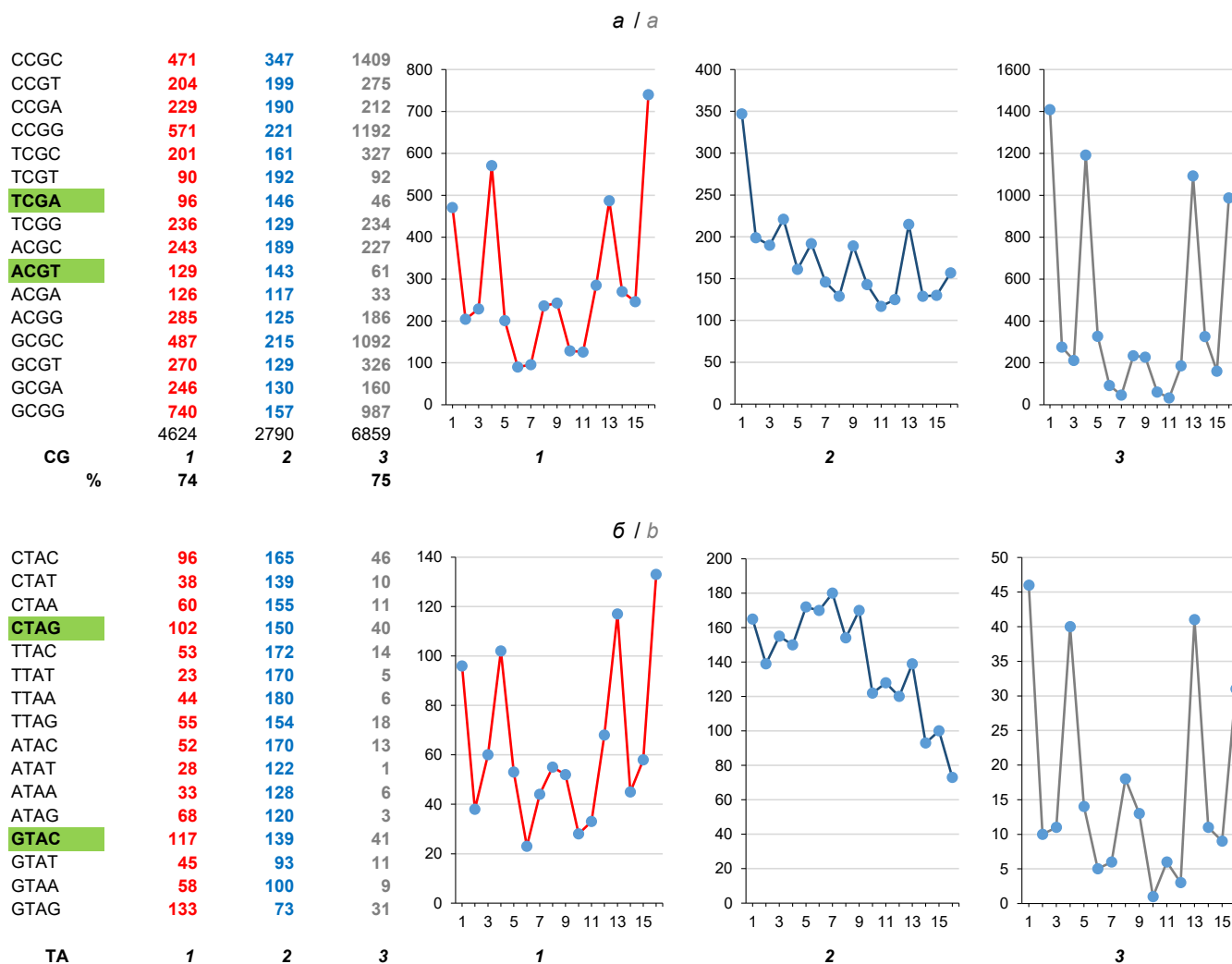


Рис. 5. ЧП тетрамеров nCGn (а) и nTAn (б) в цепочках 1, 2 и 3-х нуклеотидов генома *Herpes simplex virus 1*.

Здесь и на рис. 6, 7: слева — абсолютные значения, справа — их графическое выражение в цепочках 1, 2 и 3-х нуклеотидов, представленное для демонстрации пропорций профиля (но не его масштабов, которые читатель может представить самостоятельно, пользуясь абсолютными значениями цифровой части рисунка).

Fig. 5. FP of nCGn (a) and nTAn (b) tetramers in chains of the 1st, 2nd and 3rd nucleotides of the *Herpes simplex virus 1* genome.

Here and in Fig. 6, 7: on the left — absolute values, on the right — their graphic representation in chains of the 1st, 2nd and 3rd nucleotides to demonstrate proportions of the profile (but not its scale, which can be estimated with the help of absolute values presented in the numeric section of the figure).

ЧП тетрамеров nCGn и nTAn в случае GC-типа генома демонстрирует билатеральную симметрию 1-х и 3-х строк нуклеотидов геномов ряда вирусов и отсутствие такой симметрии в строках 2-х нуклеотидов. Такая построчная особенность характерна и для самого генетического кода. При этом ЧП 3-х строк (G) геномов демонстрируют симметрию при меньших ограничениях $[G+C] : [A+T] > 1$. ЧП 1, 2 и 3-й цепей полной (не ограниченной геномом) ДНК, гены которой не выделены, нивелирует отмеченное различие между цепями генома.

Построчный ЧП nCGn герпесвирусных ДНК выделяет тетрамер TCGA (напомним, что он не является ТН в конечной версии кода) в качестве минимального в строках А и G в большей части ис-

следованных случаев, а построчный ЧП nTAn не выделяет герпесвирусы среди других как группу, уникальную по $STAG_{min}$.

При естественном равенстве размеров всех трех строк вирусного генома (GC тип ДНК) сумма числа тетрамеров nCGn ЧП 1-й и 2-й строк примерно равна числу таких тетрамеров в 3-й строке.

Таким образом, особенности обеих групп функциональных ТН TCGA и STAG, описанных в разделах 2 и 3, объединяет общее свойство — симметрия, выявляемая как в полных вирусных ДНК, так и в отдельных кодонных строках геномов этих вирусов. В первом случае она относится к ДНК «сегодняшних» вирусов, во втором — к их геномам и к самому генетическому коду. Обе группы симмет-

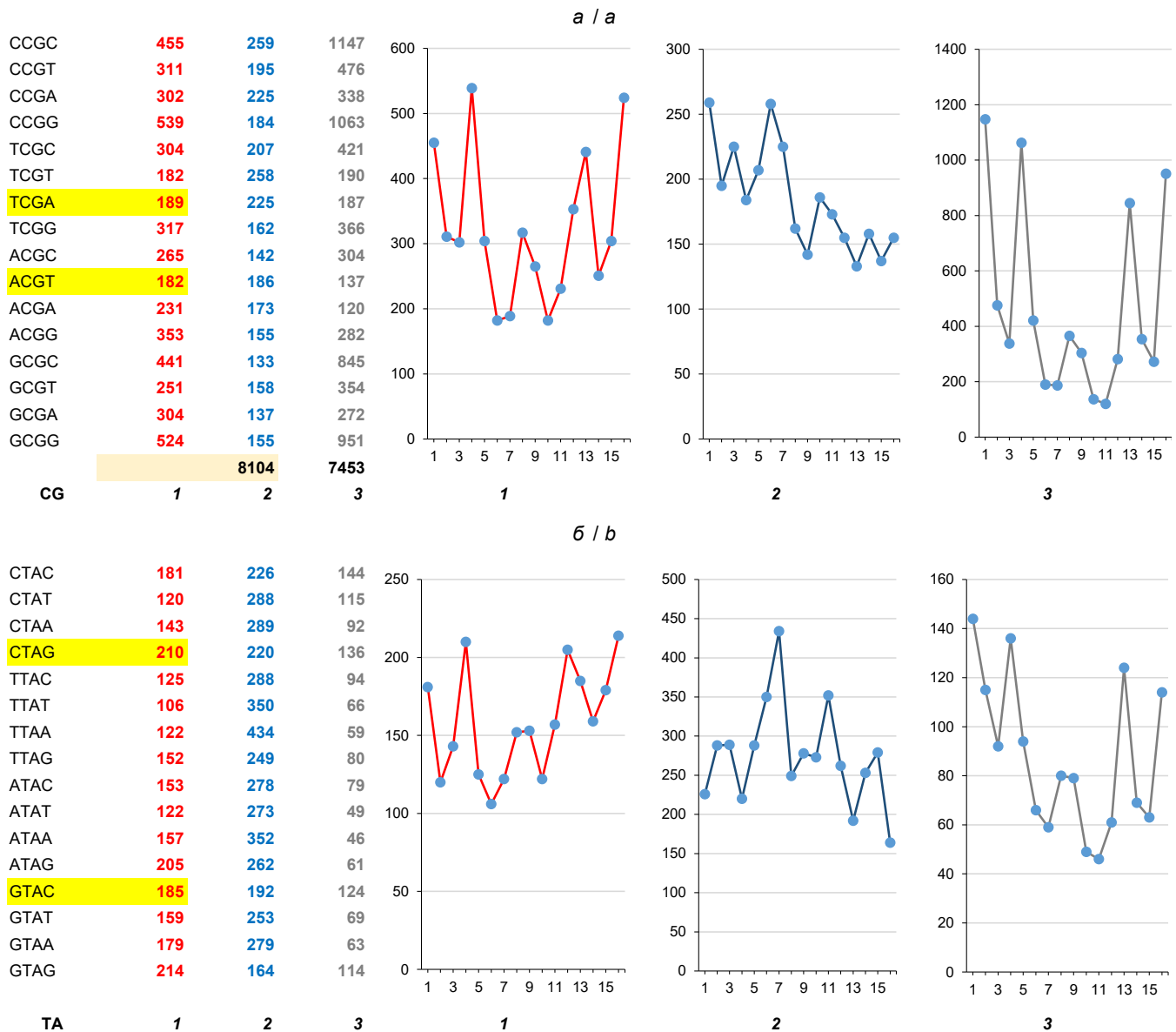


Рис. 6. ЧП тетрамеров nCGn (а) и nTAn (б) в цепочках 1, 2 и 3-х нуклеотидов генома *Human cytomegalovirus* (HHV5).

Fig. 6. FPs of nCGn (a) and nTAn (b) tetramers in chains of the 1st, 2nd and 3rd nucleotides of *Human cytomegalovirus* (HHV5) genome.

рий — и самый их набор — ставят вопрос о происхождении вирусов или хотя бы некоторых из них.

Обсуждение

Жизнь на Земле начиналась с гликозилирования и фосфорилирования пуринов и пиримидинов с последующим отбором однотипных оптических изомеров и их нематричной полимеризацией. Ни один из этих этапов — в известных нам современных естественных условиях на поверхности нашей планеты — практически невозможен без участия ферментов, хотя на ранних этапах абиогенеза в качестве замены такого фермента могли использоваться, например, различные глины [24]. Какое событие и какие условия подготовили здесь (или даже за пре-

делами Земли), запустили и масштабировали этот абиогенетический процесс свыше 4 млрд лет назад — этот вопрос, который остаётся предметом многочисленных спекуляций; вопрос, могло ли это быть случайностью, также пока не имеет ответа [25, 26]. Дальнейшая эволюция могла определяться формированием кластеров микроскопических компартментов (также при участии упомянутых глин), внутри которых происходила конкуренция растущих гетерополимеров за ограниченные ресурсы. «Проигравшие» разрушались и использовались «победителями» либо вытеснялись за пределы компартмента через его полупроницаемую мембрану. И если они успевали, не разрушившись в агрессивной внешней среде, внедриться в ближайший компартмент или оказаться в

CCGC	322	260	494
CCGT	215	216	287
CCGA	269	226	299
CCGG	429	182	513
TCGC	209	191	260
TCGT	151	225	180
TCGA	164	215	190
TCGG	253	178	272
ACGC	294	214	296
ACGT	173	192	218
ACGA	250	188	200
ACGG	341	174	293
GCGC	417	144	402
GCGT	278	158	295
GCGA	301	156	244
GCGG	481	143	422
	1	2	3

CTAC	207	287	205
CTAT	134	254	155
CTAA	128	344	163
CTAG	250	248	227
TTAC	127	239	170
TTAT	112	302	97
TTAA	151	363	122
TTAG	189	250	169
ATAC	186	306	175
ATAT	140	231	127
ATAA	134	319	120
ATAG	210	240	139
GTAC	253	238	228
GTAT	140	200	163
GTAA	167	250	163
GTAG	274	164	215
	1	2	3

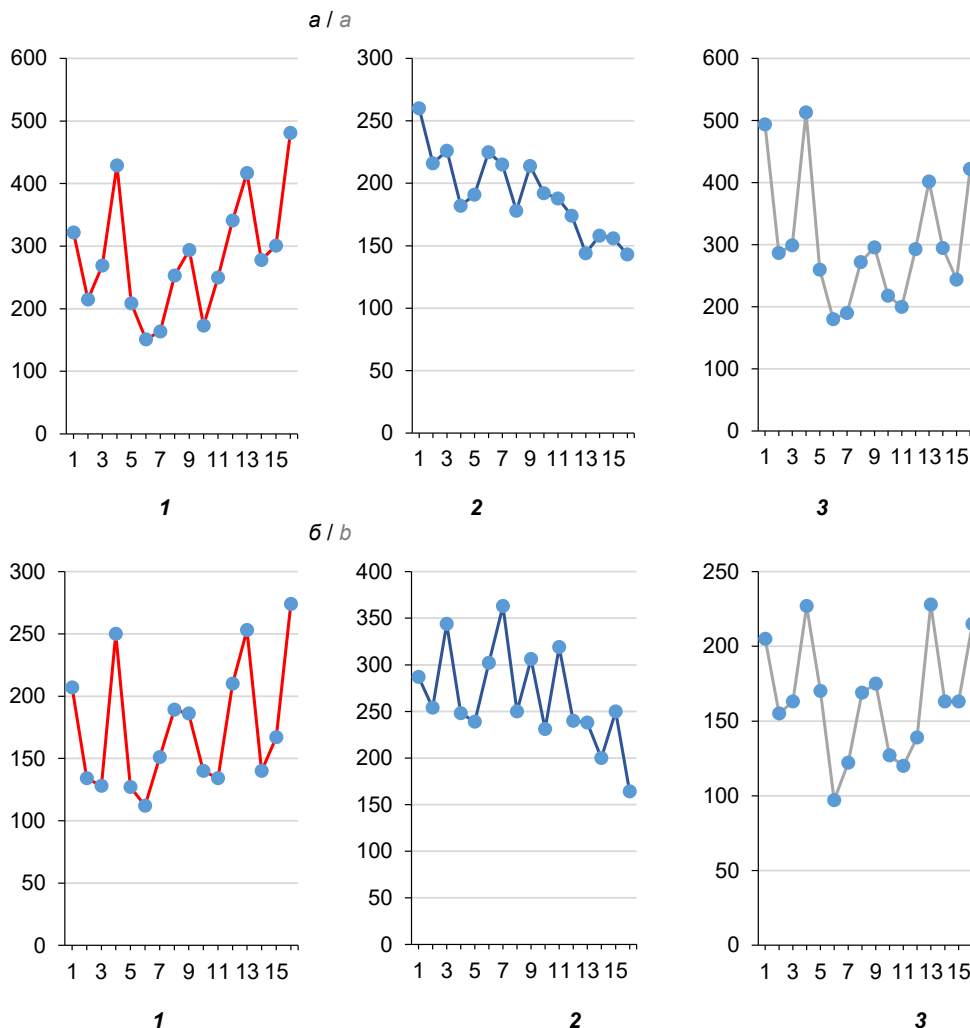


Рис. 7. ЧП тетрамеров nCGn (а) и nTAn (б) в цепочках 1, 2 и 3-х нуклеотидов генома *Ranid herpesvirus 1*.
Fig. 7. FPs of nCGn (a) and nTAn (b) tetramers in chains of the 1st, 2nd and 3rd nucleotides of *Ranid herpesvirus 1* genome.

нём после слияния, то продолжали борьбу за ресурс с новыми конкурентами, которую могли и выиграть. Относительно компартамента поведение этих конкурентов практически не отличалось от поведения сегодняшних вирусов, хотя сам компартимент значительно больше отличался от современной клетки. Преимущество «победителя» определялось в первую очередь скоростью роста, ограниченного допустимым размером, а также формирующейся в этих условиях матричной репликацией, катализируемой рибозимами, продуктами РНК-мира [27, 28], протометаллополипротеинами [29] или случайными факторами.

Описываемые события уже на этом этапе имели две выраженные характеристики жизни: жёсткую конкуренцию участников за ресурсы роста и эволюцию системы, которой требовала такая конкуренция. Устойчивость полимеров могла поддерживать их структура, которая в межрепликационный период делала цепь минимум двойной [30], что — при сохранении её общей длины — укорачивало более чувствительные к повреждениям одноцепо-

точные участки и формировало множественные повторы, придававшие этой цепи элементы симметрии. Возможно, такая система возникала ненадолго и неоднократно в различных местах планеты, но в конечном счёте она подошла к фундаментальному эволюционному скачку — созданию трансляционной машинерии и генетического кода, которые фиксировали кооперацию нуклеотидных и аминокислотных гетерополимеров и существенно снизили случайность дальнейших процессов на молекулярном уровне.

В нуклеотидных полимерах, способных к росту и репликации, возникла информация, определяющая аминокислотные последовательности, способные катализировать синтетические и репликационные процессы намного эффективней, чем случайные факторы предшествующих этапов.

Генетический код стабилизировал химию жизни и многократно ускорил её эволюцию, приведшую к организации первых клеток и разделившую позиции геномных нуклеиновых кислот на клеточные и

внеклеточные и, таким образом, закрепившую два первых центральных биологических звена, способных к эффективному взаимодействию (конкуренции или кооперации) — клетку и вирус, размеры которых получили возможность существенно вырасти. Вирусы, скорее всего, продолжали образовываться и позднее, и другими путями [31–34].

По мнению некоторых специалистов, генетический код формировался поэтапно [35–38]. Мы предполагаем, что первоначально код сохранял характеристики «до-кодовых» гетерополимеров, именно некоторый избыток G и C, а также элементы симметрии (за счет повторов), придававшие ему помехоустойчивость. В основу симметрии кода легла не только комплементарность, но и другой простой параметр, объединявший кодоны и кодируемые продукты — молекулярная масса (размер) участников. Димер CpG, который, благодаря распространённости, вероятнее всего, стал исходным структурным элементом кода, характеризуется комплементарностью $C \equiv G$ и соотношением $C < G$ ($Y < R$) молекулярных масс мономеров. Вероятно, наличие определённых уникальных функций этого динуклеотида в синтезе биополимеров выделяло его среди других и было как-то причастно к выбору его в качестве исходного. По некоторым предположениям, первые кодоны и были дуплетными [35]. Позднее соотношение $Y < R$ было сохранено, а набор первых нуклеотидов кода был расширен до полной четвёрки — TCGA.

Соотношение $Y < R$ несколько позднее легло в основу сборки другого тетрамера — СТАG, который (уже как ТН) также обладает уникальной биологической функцией. Этот тетрамер детализировал однонаправленность соотношения нуклеотидов с пиримидин-пуринового уровня до уровня нуклеотидов ($C < T < A < G$).

Эволюция размера кодонов привела к возможной на первых этапах взаимной перекрываемости, которая позднее сменилась триплетностью кода с различием функций 1, 2 и 3-х букв кодона. Первые стали обеспечивать стабильность кода на основе симметрии, основанной на последовательном изменении молекулярной массы продуктов кодирования. Аминокислоты, имеющие общий путь биосинтеза, как правило, имеют и общую первую позицию кодонов [25]. Вторые буквы кодирующих триплетов контролируют функции аминокислот на основе их полярности; кодоны аминокислот со схожими физико-химическими свойствами также, как правило, похожи, что смягчает последствия точечных мутаций и нарушений трансляции. Третьи буквы кодонов разделяют кодирующие дуплеты пуринами или пиримидинами (октет А) либо произвольным выбором тех или других (октет С) [23]. Всё это закрепило за обоими тетрамерами принадлежность к различным группам вырожденности, которая сохра-

нялась при расширении за пределы тетрамеров.

Результатом эволюции кода стал предшественник октета С (преобладание С и G), а позднее (или одновременно), когда в код были введены дополнительные характеристики — направление чтения гена и различие кодонов по третьим буквам, пурины или пиримидину, сформировался октет А (компенсаторное преобладание А и T; рис. 4).

Сегодняшние «живые» одноцепочечные нуклеиновые кислоты (отдельные цепи геномов) также отличаются определённой симметрией, включая симметрию описываемых нами ЧП функциональных ТН. Вирусные ДНК — хороший предмет для изучения этой симметрии, поскольку их геном, т.е. набор собственно генов, кодирующих последовательностей, занимает большую часть ДНК (у герпесвирусов, например, свыше 80%).

Мы показали, что в геноме герпесвирусов с большим содержанием GC ЧП nCGn в цепочках 1-х и 3-х нуклеотидов имеет тот же характер симметрий, что и в цепочках 1-х и 3-х нуклеотидов генетического кода. При этом ЧП nCGn в цепочке 2-х кодонных нуклеотидов такой симметрии нет, хотя они имеют те же общие характеристики, что и остальные две цепи, а также общую неразделенную цепь ДНК: тип GC и соотношение $[G+C] : [A+T] > 2$. Различия ЧП nCGn в цепочках 1, 2 и 3-х букв генома соответствуют функциям нуклеотидов кодона и формальной структуре генетического кода. Общее число тетрамеров nЦГн 3-й цепочки в геноме вирусов с большим содержанием GC приблизительно равно сумме общего числа тетрамеров nЦГн 1-й и 2-й цепочек (кодоны октета С — это выбор из 2 в случае первых двух нуклеотидов и выбор из 4 — в случае третьего).

Всё это подчёркивает функциональную характеристику каллиграмм генетического кода, делая их более содержательными, нежели стандартная таблица, варианты которой используются всеми учебниками.

Мы предполагаем, что симметрия ЧП nCGn нуклеотидов третьей цепи — как и общей симметрии ЧП nCGn — может являться атавистической характеристикой «до-кодового» пула полинуклеотидов. С другой стороны, характеристика третьей цепи может требоваться как некий «резерв» для обеспечения симметрии первой. Разумеется, описанные здесь симметрии могли формироваться любыми, даже сгенерированными, как случайные, полимерами ДНК достаточной длины. Однако в таком случае разделение их на три цепочки по описанным выше принципам не могло выделять по таким симметриям вторую цепочку.

Сохранение симметрии ЧП nCGn хотя бы в одной из трех цепочек вирусного генома при его разделении означает, что ограничение на размер генома может — в определённых условиях — быть снижено

примерно в 2–3 раза относительно того, которое мы приняли в начале данной работы (100 т.п.н.).

Разумеется, акцент на герпесвирусах в нашем сообщении (и на аденовирусах также) не означает, что именно с этих вирусов начиналась жизнь на Земле. Для такого предположения эти вирусы, их компоненты (и их хозяева) слишком сложны в структурном и функциональном отношении [18], а их ДНК слишком велика, что показывает, что они прошли длительную эволюцию, которая должна была касаться в том числе их типа (GC) и высокого соотношения GC/AT [39]. Эта эволюция относилась не только к ДНК, но и к продуктам кодирования — белкам, более стабильным компонентам жизни [40, 41]. Именно отдельные белки при сравнении их у различных вирусов показывают эволюционную близость герпесвирусов и хвостатых фагов [42] — при серьёзном эволюционном расхождении структуры ДНК по обсуждаемым здесь параметрам. Роль концевых или внутренних повторов этих ДНК не так очевидна для формирования симметрий, но и они также могут носить атавистический, реликтовый характер.

Симметрии генетического кода обсуждались и прежде, к этой теме исследователи — с самых разных позиций (не только с упомянутых выше) — возвращаются вновь и вновь [43, 44]. Мы рассматриваем здесь частную, но не менее интересную сторону этих симметрий.

Публикуя приведённые данные, мы хотели обратить внимание читателя на характеристики и сходство двух биологических объектов, казалось бы, далеких друг от друга, но обладающих общим весьма выразительным маркером — тетрамерами TCGA и СТАГ и общим свойством ЧП этих тетрамеров — симметриями. Первый такой объект — вирусные (в нашем случае) ДНК, второй — универсальный генетический код. С точки зрения представленных здесь данных, предполагается эволюционная связь между этими объектами, в основе которой лежат не до конца изученные биологические функции этих тетрамеров. Хотя эти функции в биосинтезе ДНК и в процессе формировании кода, на первый взгляд, очень различны, такие различия могут определяться условиями их проявления на различных этапах биологической эволюции.

СПИСОК ИСТОЧНИКОВ

1. Филатов Ф.П., Шаргунов А.В. Тетрануклеотидный профиль герпесвирусных ДНК. *Журнал микробиологии, эпидемиологии и иммунобиологии*. 2020; 97(3): 216–26. <https://doi.org/10.36233/0372-9311-2020-97-3-3>
2. Tang L., Zhu S., Mastriani E., Fang X., Zhou Y.J., Li Y.G., et al. Conserved intergenic sequences revealed by CTAG-profiling in *Salmonella*: thermodynamic modeling for function prediction. *Sci. Rep.* 2017; 7: 43565. <https://doi.org/10.1038/srep43565>
3. Lundberg P., Welander P., Han X., Cantin E. *Herpes simplex* virus type 1 DNA is immunostimulatory *in vitro* and *in vivo*. *J. Virol.* Oct. 2003; 77(20): 11158–69. <https://doi.org/10.1128/JVI.77.20.11158-11169.2003>
4. Sharawy M., Louyakis A., Gogarten J.P., May E.R. CTAG vs. GATC: structural basis for representational differences in reverse palindromic DNA tetranucleotide sequences. *Biophys. J.* 2021; 120(3): 222a.
5. Albrecht-Buehler G. Asymptotically increasing compliance of genomes with Chargaff's second parity rules through inversions and inverted transpositions. *Proc. Natl Acad. Sci. USA.* 2006; 103(47): 17828–33. <https://doi.org/10.1073/pnas.0605553103>
6. Albrecht-Buehler G. The three classes of triplet profiles of natural genomes. *Genomics.* 2007; 89(5): 596–601. <https://doi.org/10.1016/j.ygeno.2006.12.009>
7. Zhang S.H., Wang L. A novel common triplet profile for GC-rich prokaryotic genomes. *Genomics.* 2011; 97(5): 330–1. <https://doi.org/10.1016/j.ygeno.2011.02.005>
8. Stevens M., Cheng J., Li D., Xi M., Hong C., Maire C., et al. Estimating absolute methylation levels at single-CpG resolution from methylation enrichment and restriction enzyme sequencing methods. *Genome Res.* 2013; 23(9): 1541–53. <https://doi.org/10.1101/gr.152231.112>
9. Krieg A.M., Yi A.K., Matson S., Waldschmidt T.J., Bishop G.A., Teasdale R., et al. CpG motifs in bacterial DNA trigger direct B-cell activation. *Nature.* 1995; 374(6522): 546–9. <https://doi.org/10.1038/374546a0>
10. Fatemi M., Pao M.M., Jeong S., Gal-Yam E.N., Egger G., Weisenberger D.J., et al. Footprinting of mammalian promoters: use of a CpG DNA methyltransferase revealing nucleosome positions at a single molecule level. *Nucleic Acids Res.* 2005; 33(20): e176. <https://doi.org/10.1093/nar/gni180>
11. Woellmer A., Hammerschmidt W. Epstein–Barr virus and host cell methylation: regulation of latency, replication and virus reactivation. *Curr. Opin. Virol.* 2013; 3(3): 260–5. <https://doi.org/10.1016/j.coviro.2013.03.005>
12. Burge C., Campbell A.M., Karlin S. Over- and under-representation of short oligonucleotides in DNA sequences. *PNAS.* 1992; 89(4) 1358–62. <https://doi.org/10.1073/pnas.89.4.1358>
13. Duret L., Galtier N. The covariation between TpA deficiency, CpG deficiency, and G+C content of human isochores is due to a mathematical artifact. *Mol. Biol. Evol.* 2000; 17(11): 1620–5. <https://doi.org/10.1093/oxfordjournals.molbev.a02621>
14. Gori F., Mavroedis D., Jetten M.S.M., Marchiori E. The importance of Chargaff's second parity rule for genomic signatures in metagenomics. *bioRxiv*. Preprint. <https://doi.org/10.1101/146001>
15. Rudner R., Karkas J.D., Chargaff E. Separation of *B. subtilis* DNA into complementary strands, 3 Direct Analysis. *Proc. Natl Acad. Sci. USA.* 1968; 60(3): 921–2. <https://doi.org/10.1073/pnas.60.3.921>
16. Makukov M.A., Shcherbak V.I. The “Wow! signal” of the terrestrial genetic code. *Icarus.* 2013; 224(1): 228–42. <https://doi.org/10.1016/j.icarus.2013.02.017>
17. Filatov F. A molecular mass gradient is the key parameter of the genetic code organization. In: Blaho J., Baines J., eds. *From the Hallowed Halls of Herpesvirology: A Tribute to Bernard Roizman*. World Scientific Publishing Co.; 2012: 155–68. https://doi.org/10.1142/9789814338998_0006
18. Pellett P., Roizman B. Herpesviridae. In: Knipe D.M., Howley P.M., eds. *Fields Virology*. Philadelphia: Lippincott Williams & Wilkins; 2013: 1802–2
19. Prabhu V.V. Symmetry observations in long nucleotide sequences. *Nucleic Acids Res.* 1993; 21(12): 2797–800. <https://doi.org/10.1093/nar/21.12.2797>
20. Forsdyke D.R. Symmetry observations in long nucleotide sequences: a commentary on the discovery note of Qi and Cuticchia. *Bioinformatics.* 2002; 18(1): 215–7. <https://doi.org/10.1093/bioinformatics/18.1.215>

21. Baisnee P.F., Hampson S., Baldi P. Why are complementary strands symmetric? *Bioinformatics*. 2002; 18(8): 1021–33. <https://doi.org/10.1093/bioinformatics/18.8.1021>
22. Румер Ю.Б. О систематизации кодонов в генетическом коде. *Доклады Академии наук СССР*. 1966; 167(6): 1393–4.
23. Волькенштейн М.В., Румер Ю.Б. О систематике кодонов. *Биофизика*. 1967; 12(1): 10–3.
24. Kim H.Y., Cheon J.H., Lee S.H., Min J.Y., Back S.Y., Song J.G., et al. Ternary nanocomposite carriers based on organic clay-lipid vesicles as an effective colon-targeted drug delivery system: preparation and *in vitro/in vivo* characterization. *J. Nanobiotechnology*. 2020; 18(1): 17. <https://doi.org/10.1186/s12951-020-0579-7>
25. Koonin E.V., Novozhilov A.S. Origin and evolution of the genetic code: the universal enigma. *IUBMB Life*. 2009; 61(2): 99–111. <https://doi.org/10.1002/iub.146>
26. Marlaire R., ed. *Ames Research Center. NASA Ames Reproduces the Building Blocks of Life in Laboratory*. Moffett Field, CA: NASA; 2015.
27. Herbert K.M., Nag A. A tale of two RNAs during viral infection: how viruses antagonize mRNAs and small non-coding RNAs in the host cell. *Viruses*. 2016; 8(6): 154. <https://doi.org/10.3390/v8060154>
28. Tjhung K.F., Shokhirev M.N., Horning D.P., Joyce G.F. An RNA polymerase ribozyme that synthesizes its own ancestor. *Proc. Natl Acad. Sci. USA*. 2020; 117(6): 2906–13. <https://doi.org/10.1073/pnas.1914282117>
29. Kim J.D., Senn S., Harel A., Jelen B.I., Falkowski P.G. Discovering the electronic circuit diagram of life: structural relationships among transition metal binding sites in oxidoreductases. *Philos. Trans. R Soc. Lond. B. Biol. Sci.* 2013; 368(1622): 20120257. <https://doi.org/10.1098/rstb.2012.0257>
30. Yakovchuk P., Protozanova E., Frank-Kamenetskii M.D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* 2006; 34(2): 564–74. <https://doi.org/10.1093/nar/gkj454>
31. Forterre P. The origin of viruses and their possible roles in major evolutionary transitions. Review. *Virus Res.* 2006; 117: 5–16.
32. Mughal F., Nasir A., Caetano-Anollés G. The origin and evolution of viruses inferred from fold family structure. *Arch. Virol.* 2020; 165(10): 2177–91. <https://doi.org/10.1007/s00705-020-04724-1>
33. Brussow H., Kutter E. Genomics and evolution of tailed phages. In: Kutter E., Sulakvelidze A. eds. *Bacteriophages: Biology and Applications*. Boca Raton, London, New York, Washington: CRC press; 2005: 129–64.
34. Abedon S.T. Phage evolution and ecology. *Adv. Appl. Microbiol.* 2009; 67: 1–45. [https://doi.org/10.1016/s0065-2164\(08\)01001-0](https://doi.org/10.1016/s0065-2164(08)01001-0)
35. Altstein A.D. The progene hypothesis: the nucleoprotein world and how life began. *Biol. Direct*. 2015; 10: 67. <https://doi.org/10.1186/s13062-015-0096-z>
36. Di Giulio M. The origin of the genetic code: theories and their relationships, a review. *Biosystems*. 2005; 80(2): 175–84. <https://doi.org/10.1016/j.biosystems.2004.11.005>
37. Gilis D., Massar S., Cerf N.J., Rooman M. Optimality of the genetic code with respect to protein stability and amino-acid frequencies. *Genome Biol.* 2001; 2(11): RESEARCH0049. <https://doi.org/10.1186/gb-2001-2-11-research0049>
38. Wetzel R. Evolution of the aminoacyl-tRNA synthetases and the origin of the genetic code. *J. Mol. Evol.* 1995; 40(5): 545–50. <https://doi.org/10.1007/bf00166624>
39. McGeoch J., Rixon F.J., Davison A.J. Topics in herpesvirus genomics and evolution. *Virus Res.* 2006; 117(1): 90–104. <https://doi.org/10.1016/j.virusres.2006.01.002>
40. Wang N., Baldi P.F., Gaut B.S. Phylogenetic analysis, genome evolution and the rate of gene gain in the *Herpesviridae*. *Mol. Phylogenet. Evol.* 2007; 43(3): 1066–75. <https://doi.org/10.1016/j.ympev.2006.11.019>
41. Wertheim J.O., Smith M.D., Smith D.M., Scheffler K., Kosakovsky Pond S.L. Evolutionary origins of human herpes simplex viruses 1 and 2. *Mol. Biol. Evol.* 2014; 31(9): 2356–64. <https://doi.org/10.1093/molbev/msu185>
42. Baker M.L., Jiang W., Rixon F.J., Chiu W. Common ancestry of herpesviruses and tailed DNA bacteriophages. *J. Virol.* 2005; 79(23): 14967–70. <https://doi.org/10.1128/JVI.79.23.14967-14970.2005>
43. Гупал А.М., Гупал Н.А., Островский А.В. Симметрия и свойства записи генетической информации в ДНК. *Проблемы управления и информатики*. 2011; 5(3): 120–7.
44. Сергиенко И.В., Гупал А.М., Вагис А.А. Симметричный код и генетические мутации. *Кибернетика и системный анализ*. 2016; (2): 73–80.

REFERENCES

1. Filatov F.P., Shargunov A.V. Tetranucleotide profile of herpesvirus DNA. *Zhurnal mikrobiologii, epidemiologii i immunobiologii*. 2020; 97(3): 216–26. <https://doi.org/10.36233/0372-9311-2020-97-3-3> (in Russian)
2. Tang L., Zhu S., Mastriani E., Fang X., Zhou Y.J., Li Y.G., et al. Conserved intergenic sequences revealed by CTAG-profiling in *Salmonella*: thermodynamic modeling for function prediction. *Sci. Rep.* 2017; 7: 43565. <https://doi.org/10.1038/srep43565>
3. Lundberg P., Welander P., Han X., Cantin E. *Herpes simplex virus type 1 DNA is immunostimulatory in vitro and in vivo*. *J. Virol.* Oct. 2003; 77(20): 11158–69. <https://doi.org/10.1128/JVI.77.20.11158-11169.2003>
4. Sharawy M., Louyakis A., Gogarten J.P., May E.R. CTAG vs. GATC: structural basis for representational differences in reverse palindromic DNA tetranucleotide sequences. *Biophys. J.* 2021; 120(3): 222a.
5. Albrecht-Buehler G. Asymptotically increasing compliance of genomes with Chargaff's second parity rules through inversions and inverted transpositions. *Proc. Natl Acad. Sci. USA*. 2006; 103(47): 17828–33. <https://doi.org/10.1073/pnas.0605553103>
6. Albrecht-Buehler G. The three classes of triplet profiles of natural genomes. *Genomics*. 2007; 89(5): 596–601. <https://doi.org/10.1016/j.ygeno.2006.12.009>
7. Zhang S.H., Wang L. A novel common triplet profile for GC-rich prokaryotic genomes. *Genomics*. 2011; 97(5): 330–1. <https://doi.org/10.1016/j.ygeno.2011.02.005>
8. Stevens M., Cheng J., Li D., Xi M., Hong C., Maire C., et al. Estimating absolute methylation levels at single-CpG resolution from methylation enrichment and restriction enzyme sequencing methods. *Genome Res.* 2013; 23(9): 1541–53. <https://doi.org/10.1101/gr.152231.112>
9. Krieg A.M., Yi A.K., Matson S., Waldschmidt T.J., Bishop G.A., Teasdale R., et al. CpG motifs in bacterial DNA trigger direct B-cell activation. *Nature*. 1995; 374(6522): 546–9. <https://doi.org/10.1038/374546a0>
10. Fatemi M., Pao M.M., Jeong S., Gal-Yam E.N., Egger G., Weisenberger D.J., et al. Footprinting of mammalian promoters: use of a CpG DNA methyltransferase revealing nucleosome positions at a single molecule level. *Nucleic Acids Res.* 2005; 33(20): e176. <https://doi.org/10.1093/nar/gni180>
11. Woellmer A., Hammerschmidt W. Epstein-Barr virus and host cell methylation: regulation of latency, replication and virus reactivation. *Curr. Opin. Virol.* 2013; 3(3): 260–5. <https://doi.org/10.1016/j.coviro.2013.03.005>
12. Burge C., Campbell A.M., Karlin S. Over- and under-representation of short oligonucleotides in DNA sequences. *PNAS*. 1992; 89(4): 1358–62. <https://doi.org/10.1073/pnas.89.4.1358>
13. Duret L., Galtier N. The covariation between TpA deficiency, CpG deficiency, and G+C content of human isochores is due to a mathematical artifact. *Mol. Biol. Evol.* 2000; 17(11): 1620–5. <https://doi.org/10.1093/oxfordjournals.molbev.a02621>

14. Gori F., Mavroicidis D., Jetten M.S.M., Marchiori E. The importance of Chargaff's second parity rule for genomic signatures in metagenomics. *bioRxiv*. Preprint. <https://doi.org/10.1101/146001>
15. Rudner R., Karkas J.D., Chargaff E. Separation of *B. subtilis* DNA into complementary strands, 3 Direct Analysis. *Proc. Natl Acad. Sci. USA*. 1968; 60(3): 921–2. <https://doi.org/10.1073/pnas.60.3.921>
16. Makukov M.A., Shcherbak V.I. The “Wow! signal” of the terrestrial genetic code. *Icarus*. 2013; 224(1): 228–42. <https://doi.org/10.1016/j.icarus.2013.02.017>
17. Filatov F. A molecular mass gradient is the key parameter of the genetic code organization. In: Blaho J., Baines J., eds. *From the Hallowed Halls of Herpesvirology: A Tribute to Bernard Roizman*. World Scientific Publishing Co.; 2012: 155–68. https://doi.org/10.1142/9789814338998_0006
18. Pellett P., Roizman B. Herpesviridae. In: Knipe D.M., Howley P.M., eds. *Fields Virology*. Philadelphia: Lippincott Williams & Wilkins; 2013: 1802–2
19. Prabhu V.V. Symmetry observations in long nucleotide sequences. *Nucleic Acids Res*. 1993; 21(12): 2797–800. <https://doi.org/10.1093/nar/21.12.2797>
20. Forsdyke D.R. Symmetry observations in long nucleotide sequences: a commentary on the discovery note of Qi and Cuticchia. *Bioinformatics*. 2002; 18(1): 215–7. <https://doi.org/10.1093/bioinformatics/18.1.215>
21. Baisnee P.F., Hampson S., Baldi P. Why are complementary strands symmetric? *Bioinformatics*. 2002; 18(8): 1021–33. <https://doi.org/10.1093/bioinformatics/18.8.1021>
22. Rumer Yu.B. On codon systematization in the genetic code. *Doklady Akademii nauk SSSR*. 1966; 167(6): 1393–4. (in Russian)
23. Vol'kenshtein M.V., Rumer Yu.B. Systematics of codons. *Biofizika*. 1967; 12(1): 10–3.
24. Kim H.Y., Cheon J.H., Lee S.H., Min J.Y., Back S.Y., Song J.G., et al. Ternary nanocomposite carriers based on organic clay-lipid vesicles as an effective colon-targeted drug delivery system: preparation and *in vitro/in vivo* characterization. *J. Nanobiotechnology*. 2020; 18(1): 17. <https://doi.org/10.1186/s12951-020-0579-7>
25. Koonin E.V., Novozhilov A.S. Origin and evolution of the genetic code: the universal enigma. *IUBMB Life*. 2009; 61(2): 99–111. <https://doi.org/10.1002/iub.146>
26. Marlaire R., ed. *Ames Research Center. NASA Ames Reproduces the Building Blocks of Life in Laboratory*. Moffett Field, CA: NASA; 2015.
27. Herbert K.M., Nag A. A tale of two RNAs during viral infection: how viruses antagonize mRNAs and small non-coding RNAs in the host cell. *Viruses*. 2016; 8(6): 154. <https://doi.org/10.3390/v8060154>
28. Tjhung K.F., Shokhiev M.N., Horning D.P., Joyce G.F. An RNA polymerase ribozyme that synthesizes its own ancestor. *Proc. Natl Acad. Sci. USA*. 2020; 117(6): 2906–13. <https://doi.org/10.1073/pnas.1914282117>
29. Kim J.D., Senn S., Harel A., Jelen B.I., Falkowski P.G. Discovering the electronic circuit diagram of life: structural relationships among transition metal binding sites in oxidoreductases. *Philos. Trans. R Soc. Lond. B. Biol. Sci*. 2013; 368(1622): 20120257. <https://doi.org/10.1098/rstb.2012.0257>
30. Yakovchuk P., Protozanova E., Frank-Kamenetskii M.D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res*. 2006; 34(2): 564–74. <https://doi.org/10.1093/nar/gkj454>
31. Forterre P. The origin of viruses and their possible roles in major evolutionary transitions. *Review. Virus Res*. 2006; 117: 5–16.
32. Mughal F., Nasir A., Caetano-Anollés G. The origin and evolution of viruses inferred from fold family structure. *Arch. Virol*. 2020; 165(10): 2177–91. <https://doi.org/10.1007/s00705-020-04724-1>
33. Brussow H., Kutter E. Genomics and evolution of tailed phages. In: Kutter E., Sulakvelidze A. eds. *Bacteriophages: Biology and Applications*. Boca Raton, London, New York, Washington D.C.: CRC press; 2005: 129–64.
34. Abedon S.T. Phage evolution and ecology. *Adv. Appl. Microbiol*. 2009; 67: 1–45. [https://doi.org/10.1016/s0065-2164\(08\)01001-0](https://doi.org/10.1016/s0065-2164(08)01001-0)
35. Altstein A.D. The progene hypothesis: the nucleoprotein world and how life began. *Biol. Direct*. 2015; 10: 67. <https://doi.org/10.1186/s13062-015-0096-z>
36. Di Giulio M. The origin of the genetic code: theories and their relationships, a review. *Biosystems*. 2005; 80(2): 175–84. <https://doi.org/10.1016/j.biosystems.2004.11.005>
37. Gilis D., Massar S., Cerf N.J., Rooman M. Optimality of the genetic code with respect to protein stability and amino-acid frequencies. *Genome Biol*. 2001; 2(11): RESEARCH0049. <https://doi.org/10.1186/gb-2001-2-11-research0049>
38. Wetzel R. Evolution of the aminoacyl-tRNA synthetases and the origin of the genetic code. *J. Mol. Evol*. 1995; 40(5): 545–50. <https://doi.org/10.1007/bf00166624>
39. McGeoch J., Rixon F.J., Davison A.J. Topics in herpesvirus genomics and evolution. *Virus Res*. 2006; 117(1): 90–104. <https://doi.org/10.1016/j.virusres.2006.01.002>
40. Wang N., Baldi P.F., Gaut B.S. Phylogenetic analysis, genome evolution and the rate of gene gain in the *Herpesviridae*. *Mol. Phylogenet. Evol*. 2007; 43(3): 1066–75. <https://doi.org/10.1016/j.ympev.2006.11.019>
41. Wertheim J.O., Smith M.D., Smith D.M., Scheffler K., Kosakovsky Pond S.L. Evolutionary origins of human herpes simplex viruses 1 and 2. *Mol. Biol. Evol*. 2014; 31(9): 2356–64. <https://doi.org/10.1093/molbev/msu185>
42. Baker M.L., Jiang W., Rixon F.J., Chiu W. Common ancestry of herpesviruses and tailed DNA bacteriophages. *J. Virol*. 2005; 79(23): 14967–70. <https://doi.org/10.1128/JVI.79.23.14967-14970.2005>
43. Gupal A.M., Gupal N.A., Ostrovskiy A.V. Symmetry and properties of recording genetic information in DNA. *Problemy upravleniya i informatiki*. 2011; 5(3): 120–7. (in Russian)
44. Sergienko I.V., Gupal A.M., Vagis A.A. Symmetric code and genetic mutations. *Kibernetika i sistemnyy analiz*. 2016; (2): 73–80. (in Russian)

Информация об авторе

Филатов Феликс Петрович[✉] — д.б.н., в.н.с. лаб. молекулярной биотехнологии отдела вирусологии НИИВС им. И.И. Мечникова, Москва, Россия; в.н.с. отдела эпидемиологии НИЦЭИМ им. Н.Ф. Гамалеи, Москва, Россия, felix001@gmail.com, <https://orcid.org/0000-0001-6182-2241>

Статья поступила в редакцию 07.05.2022;
принята к публикации 30.06.2022;
опубликована 30.07.2022

Information about the authors

Felix P. Filatov[✉] — D. Sci. (Biol.), leading researcher, Laboratory of molecular biotechnology, Department of virology, I. Mechnikov Research Institute of Vaccines and Sera, Moscow, Russia; leading researcher, Department of epidemiology, National Research Center for Epidemiology and Microbiology named after Honorary Academician N.F. Gamaleya, Moscow, Russia, felix001@gmail.com, <https://orcid.org/0000-0001-6182-2241>

The article was submitted 07.05.2022;
accepted for publication 30.06.2022;
published 30.07.2022