

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

### Филатов Ф.П.<sup>™</sup>

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 СТАС и, в меньшей степени, ТССА выделяются среди других полных, билатерально симметричных тетрануклеотидов заметно более низкими значениями концентраций. Цель работы — сравнительный анализ ЧП тетрануклеотидов СТАС и ТССА 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 ТСGA и СТAG. 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 СТАG. 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:online-first. 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<sup>™</sup>

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:online-first.

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

### Введение

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

Предполагается, что действие СТАG связано с нарушением оптимальной структуры стволовых петель нуклеиновой кислоты, что приводит к торможению репликации ДНК («термодинамическая модель»). Кроме того, последовательность СТАG более подвержена химическим воздействиям [5, 6]. Пониженная концентрация TCGA связана с его центральным димером СрG, 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 СТАС и ТССА в вирусных ДНК и в геномных участках этих ДНК.

Мы анализировали ближайший контекст центральных пар CG и TA нуклеотидных тетрамеров, среди которых — TCGA и CTAG, в ДНК вирусов различных классификационных групп. Такой подход несколько уравнивает оценку сравнения ЧП динуклеотида CG и CTAG, 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 СТАБ [12, 13]). Минус этого подхода заключается в значительно меньшем различии плотности симметричных пар тетрамеров с общей функцией (TCGA и ACGT) по сравнению с существенной разницей плотности симметричных пар тетрамеров, имеющих (CTAG) и практически не имеющих (GTAC) такую функцию.

### Материалы и методы

Анализу подвергались физически несегментированные ДНК, полноразмерный сиквенс которых представлен в GenBank<sup>1</sup> 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 отличает 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 субсемейства, 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 Caudoviirales) 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 TH строили перебором вариантов ближайшего контекста центральных пар nTAn и nCGn (CTAG) с последовательным повышением молекулярной массы n [16, 17]:

 $[C \rightarrow T]_{Y} \rightarrow [A \rightarrow G]_{R}$ , где С, Т — пиримидины (<sub>y</sub>); А, G — пурины (<sub>R</sub>).

### Результаты

## Плотность минимальных nTAn и nCGn в вирусных ДНК

Общий анализ плотности nTAn и nCGn показал, что в ДНК большей части (75 из 128) исследованных представителей типа *Uroviricota* минимальный TH — это CTAG<sub>min</sub>. 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  $CTAG_{min} < GTAC$  и  $TCGA_{min} < ACGT$ .

ДЁК ни одного вида типа Uroviricota не содержит TCGA в качестве минимального тетрамера. В ДНК вирусов типа Nucleocytoviricota, вирусов неуточнённой классификации (Baculoviridae, Nudiviridae и Nimaviridae) и вирусов архей TCGA<sub>min</sub>, хотя и встречается, но редко и без видимой связи с классификационными группами.

Один из иридовирусов — альфа-иридовирус, а также вирус инфекционного некроза селезёнки и почек человека содержат в своей ДНК оба названных ТН в качестве минимальных и примерно в равных концентрациях (СТАG<sub>min</sub>~TCGA<sub>min</sub>). 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 СТАG<sub>min</sub> < TCGA<sub>min</sub>.

ТССА не является минимальным в ДНК розеоловирусов. В то же время большая часть герпесвирусных ДНК (26 из 35) — кроме гамма-герпесвирусов — имеет в качестве минимального СТАС. Раздел 2 детализирует особенности ЧП пТАп и пССП в ДНК герпесвирусов.

Минимальной концентрацией СТАG отличаются все исследованные *Nucleocytoviricota* — за исключением поксвирусов, среди которых только 3 хордопоксвируса (из 19 анализированных) имеют СТАG<sub>min</sub>. Поксвирусы, как было сказано, отличаются также ДНК с преобладающим типом A+T и с высоким (> 2) соотношением [T+A] : [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 ТСGA характерен как минимальный, почти исключительно для ДНК герпесвирусов — с более чем двойным преобладанием G+C над A+T. Это в первую очередь герпесвирусы рода *Simplexviruses* субсемейства *Alpha* и частично рода *Lymphocriptoviruses* субсемейства *Gamma*. ДНК многих герпесвирусов имеют в качестве минимального тетрамер ACGT, но он не уникален, т.е. встречается в этом качестве в ДНК других классификационных групп вирусов.

### Частотный профиль тетрануклеотидов в вирусных ДНК

Яркой характеристикой количественного распределения геномных тетрамеров nCGn 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 nTAn асимметричен в силу «минимальности» CTAG (CTAG<sub>min</sub>). Более строгим, чем

<sup>&</sup>lt;sup>1</sup> URL: https://www.ncbi.nlm.nih.gov/genomes/Genomes Group. cgi?taxid=10239&sort=taxonomy

симметрия, но также её отражающим является соотношение СТАС : GTAC и TCGA : ACGT. Для ДНК вируса простого герпеса-1 — это СТАС GTAC (21%) и TCGA<sub>min</sub> < ACGT (86%).

Достаточно заменить центральный димер тетрамера СТАС на обратный — САТС, как симметрия соответствующего ЧП восстанавливается. Асимметрия ЧП пССп выражена намного слабее, хотя для ЧП многих симплексвирусов соотношение TCGA < ACGT весьма характерно. Любые замены центрального димера СС 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

	<b>a</b> / a	
CTAC 1	221	2056 CCGC
CTAT 2	116	892 CCGT
CTAA 3	112	878 CCGA
CTAG 4	<u>91</u>	1856 CCGG
TTAC 5	175	884 TCGC
TTAT 6	182	563 TCGT
TTAA 7	178	468 TCGA
TTAG 8	109	916 TCGG
ATAC 9	200	954 ACGC
ATAT 10	152	546 ACGT
ATAA 11	185	542 ACGA
ATAG 12	128	932 ACGG
GTAC 13	439	1796 GCGC
GTAT 14	223	1022 GCGT
GTAA 15	194	894 GCGA
GTAG 16	251	2160 GCGG
ТА		CG
	21%	86%





типу D. Здесь симметрии обоих ЧП практически не выражены, однако соотношения CTAG<sub>min</sub>: GTAC и TCGA<sub>min</sub>: ACGT вполне отчётливы (36 и 52% соответственно). Очевидно, что симметрия ЧП пCGn связана с GC-типом ДНК герпесвируса рода *Simplex*.

Количественное выражение GC : АТ и связь его с симметриями ЧП герпесвирусных 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*. АТ-тип вирусных ДНК при соотношении AT : GC > 2 (поксвирусы) тоже придаёт симметрию ЧП обсуждаемых TH.



**б** / b

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

а — абсолютные значения, приведено процентное соотношение симметричных по шкале пар двух полных тетрамеров — TCGA<sub>min</sub>: ACGT и CTAG<sub>UNP</sub>: 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 — TCGAmin:ACGT and CTAGUNP: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*.

а — 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 СТАС и ТССА 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/ТАС и ССС/ССА. В определённой мере это характерно и для ряда других вирусных ДНК. Плотность таких тринуклеотидов, как правило, не достигает значений полных СТАС или TCGA.

Подытоживая эту часть работы, отметим две выраженные, не упоминавшиеся прежде формальные особенности СТАG 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 СТАС, но и его тринуклеотидные взаимоперекрывающиеся участки nTAG и CTAn 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 CTAn, перекрывающие 5'- или 3'-участки СТАG<sub>тіп</sub>, то упомянутое в разделе 1 в отношении ДНК исследованных представителей типа Uroviricota число 75 СТАG<sub>тіп</sub> из 128 исследованных увеличивается до 93. Такие тринуклеотиды также имеют тенденцию к снижению частоты, которая, как правило, не достигает минимальных значений полного CTAG<sub>min</sub>. На наш взгляд, это требует некоторого уточнения упомянутой выше гипотезы термодинамической модели, которая относится к полному ТН СТАС.

Для этих четырех тринуклеотидов (CTA/TAG, частично перекрывающих CTAG, и TCG/CGA, частично перекрывающих 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

<b>D</b> ( 0	5 /0 /			Геном /	Genome
Род / Genus	Вид / Species	GC : AI	ДНК / DNA	<b>1-й TH /</b> 1 <sup>st</sup> nt	<b>3-й TH /</b> 3 <sup>rd</sup> nt
Simplexvirus	Papiine HV2	3,18	acgt	TCGA	_
Simplexvirus	Cercopithec aHV2	3,16	acgt	TCGA	TCGA
Simplexvirus	Ateline aHV1	3,08	acgt	TCGA	TCGA
Simplexvirus	Macacine aHV1	2,92	acgt	TCGA	TCGA
Simplexvirus	Human aHV2	2,38	acgt	TCGA	TCGA
Simplexvirus	Human aHV1	2,15	TCGA	TCGA	TCGA
Simplexvirus	Chimpanzee aHV	2,13	TCGA	TCGA	TCGA
Simplexvirus	Saimiriine aHV1	2,04	TCGA	TCGA	TCGA
Quwivirus	Tupaiid bHV1	2,00	acgt	acgt	acgt
Simplexvirus	Leporid aHV4	1,97	TCGA	TCGA	acgt
Shapirovirus	Caulobacter phage CcrKarma	1,96	acgt	TCGA	TCGA
Phicbkvirus	Caulobacter virus Rogue	1,95	acgt	acgt	TCGA
Bertelyvirus	Caulobacter phage CcrBL9	1,68	acgt	TCGA	TCGA
Cytomegalovirus	Rat Maastricht	1,56	acgt	TCGA	acgt
Simplexvirus	Fruit bat aHV1	1,55	acgt	TCGA	TCGA
Lymphocriptovirus	Human bHV4	1,47	TCGA	TCGA	TCGA
Muromegalovirus	Murid bHV1	1,42	TCGA	TCGA	acgt
Cytomegalovirus	Human bHV5	1,35	TCGA	TCGA	acgt
Rhadinovirus	Dolphin gHV1	1,32	acgt	acgt	TCGA
Varicellovirus	Equid aHV1	1,31	TCGA	TCGA	TCGA
Quiwivirus	Caviid bHV2	1,22	TCGA	TCGA	acgt
Percavirus	Equid gHV5	1,21	TCGA	TCGA	TCGA
Batrachovirus	Ranid HV1	1,20	TCGA	TCGA	TCGA
Rhadinovirus	Human gHV8	1,16	acgt	acgt	acgt
Mardivirus	Gallid aHV3	1,16	acgt	TCGA	TCGA
Simplexvirus	Macropodid HV1	1.12	acgt	TCGA	TCGA
Batrachovirus	Ranid HV2	1,12	TCGA	TCGA	TCGA
Cyprinivirus	Cyprinid HV1	1,05	acgt	acgt	TCGA

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

**Примечание.** Минимальные тетрануклеотиды 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-А (первые

буквы) или  $S_{II}$ -stop- $L_{II}$ -stop- $R_{II}$  (продукты кодирования) так называемой матрицы кода [17]. Другой такой, как TCGA и CTAG, пары полных самокомплементарных TH, которые перекрывались бы тринуклеотидами, 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.

	a/a		
CTAC 1	539	1556 CCGC	25
CTAT 2	631	970 CCGT	
CTAA 3	370	610 CCGA	
CTAG 4	<u>202</u>	830 CCGG	20
TTAC 5	664	780 TCGC	
TTAT 6	676	565 TCGT	
TTAA 7	491	295 TCGA	15
TTAG 8	376	763 TCGG	10
ATAC 9	795	1259 ACGC	
ATAT 10	865	848 ACGT	10
ATAA 11	641	588 ACGA	10
ATAG 12	574	925 ACGG	
GTAC 13	954	2081 GCGC	_
GTAT 14	888	1399 GCGT	5
GTAA 15	658	817 GCGA	
GTAG 16	580	1549 GCGG	
ТА		CG	
	21%	35%	





в / с

### 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 **Рис. 3.** Частотные профили тетрануклеотидов nATn и nGCn в ДНК *Ranid herpesvirus 1*.

а — абсолютные значения, зелёным цветом и жирным шрифтом выделены З'-тримеры ТАG и СGA, подчёркнуты 5'-тримеры СТА и ССG, ниже — процентное соотношение пар двух полных тетрамеров: АСGT : TCGA и GTAC : CTAG; б, в — графическое выражение абсолютных значений;

в — nTAn, больший масштаб по вертикали.

**Fig. 3.** Frequency profiles of nATn and nGCn tetra-nucleotides in DNA of *Ranid herpesvirus 1*.

*a* — absolute values; 3'-trimers TAG and CGA are highlighted in green and bold; 5'-trimers CTA and CCG are underlined; below — the percentage ratio of complete tetramers: ACGT:TCGA and GTAC:CTAG; *b*, *c* — a graphic representation of absolute values; *c* — nTAn, vertical scale up

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 A; **рис.** 4, *a*) 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

бIb

**ORIGINAL RESEARCHES** 

а/	а
----	---

#

1

2

1

Octet C (dg - IV) G G G С С т С А G С С С т С т G CT CT CT 3 СТ СТ СТ СТ CT AA G S Ρ ν R Α п 3 R AG AG AG AG AG AG AG AG AA  $\rightarrow$ G Α S Ρ V R # - 1/111.11) Octet A (dg A Т С G A Α т 2 Т G Α т G А A А 3 CT СТ СТ CT СТ CT СТ Υ CT С H+ F Dγ AA S Ν AG 3 R AG AG AG AG AG AG AG W0 **→ AA** MI Q R+ E-K+ 0

Octet 2 III dg П 1 ΤA Т А Т А G CA GC Т А т А т Т Α Т Т 2 G G А А A A А А G А Т G Υ 3 H H R Y R Υ R R R Y YR Y G G С 0 S D Q aa L Ν Κ Е Μ

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

Третий нуклеотид кодона представлен пурином (R) или пиримидином (Y).

Стартовый кодон ATG и стоп-кодон TA<sub>в</sub> выделены серым по вертикали.

Центральные тетрануклеотиды первых букв кодонов каждого октета выделены жирным шрифтом. Тетрануклеотид стыка октетов А и С также выделен серым. Последовательное нарастание молекулярной массы продуктов кодирования (АС, аминокислоты) показано нарастанием плотности фона клеток от белого к чёрному и стрелками. Римские цифры — группы вырожденности кода, dq. 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<sub>o</sub> 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 А+Т во втором, в котором закреплены также направление 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 Ү или пурином R. Этот октет обозначен как октет A (рис. 4, a) — 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 1 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<sup>+</sup>S, E<sup>-</sup>D<sup>-</sup>и K<sup>+</sup>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 CTAG в составе верхней строки октета А, помещая в четвертую позицию тетрамера глутаминовую кислоту. Возможно, некоторую роль играет здесь симметрия зарядов гистидина (H<sup>+</sup>) и глутамина (D<sup>-</sup>) при нейтральном рН, 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<sub>в</sub> (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, *a*). 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 локализуются, 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 Ranid 1 (**рис. 7**).

Рисунок 5, а демонстрирует результат такого анализа, а именно симметрию профиля цепи первых нуклеотидов генома *HHV1*, в котором тетрамер ТСGА остается минимальным, уступая тетрамеру АСGТ. 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 ННV1 (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 CTAG, и может сохранять иллюзию функции в случае тетрамера скорее, чем в случае декамера CnnTnnAnnG. Надо отметить, что минимальное значение сохраняет и декамер TnnCnnGnnA. Дефицит СТАС в цепочке 1-х букв кодирующих триплетов исчезает, хотя значение СТАС несколько уступает симметричному по шкале GTAC. Как и в случае nCGn (рис. 5, a), 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 ННV1.

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 *Ranid herpesvirus 1*. В какой-то мере и здесь мы видим «восстанавление» 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 пСGп 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 ТСGA<sub>min</sub>, и симметрия соответствующего профиля. В значительно большей мере эти характеристики отличают 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 CTAG в связи со структурой генетического кода.

Центральными тетрамерами 1-х строк октетов каллиграммы A являются TCGA (октет C) и CTAG (октет A).

ORIGINAL RESEARCHES



**Рис. 5.** ЧП тетрамеров nCGn (*a*) 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.** FPs of nCGn (*a*) and nTAn (*b*) tetramers in chains of the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> 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 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> 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 TH в конечной версии кода) в качестве минимального в строках A и G в большей части ис-

следованных случаев, а построчный ЧП nTAn не выделяет герпесвирусы среди других как группу, уникальную по CTAG<sub>min</sub>.

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 ТСGА и СТАG, 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 симметDOI: https://doi.org/10.36233/0372-9311-275

ОРИГИНАЛЬНЫЕ ИССЛЕДОВАНИЯ



Рис. 6. ЧП тетрамеров nCGn (*a*) 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 1<sup>st</sup>, 2ns and 3<sup>rd</sup> 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 сегодняшних 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 пределами 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

ORIGINAL RESEARCHES



**Рис. 7.** ЧП тетрамеров nCGn (*a*) 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 1<sup>st</sup>, 2ns and 3<sup>rd</sup> nucleotides of *Ranid herpesvirus* 1 gnome.

нём после слияния, то продолжали борьбу за ресурс с новыми конкурентами, которую могли и выиграть. Относительно компартмента поведение этих конкурентов практически не отличалось от поведения сегодняшних вирусов, хотя сам компартмент 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 одноцепочечные 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 СрG, который, благодаря распространённости, вероятнее всего, стал исходным структурным элементом кода, характеризуется комплементарностью C=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 TH) 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 сохранялась 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, пурину или пиримидину, сформировался октет А (компенсаторное преобладание А и Т; 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 бо́льшую 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 нЦГн 3-й цепочки в геноме вирусов с большим содержанием GC приблизительно равно сумме общего числа тетрамеров нЦГн 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 фукциональном 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 СТАС 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

- 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
- 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

- 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.
- 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
- 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
- Zhang S.H., Wang L. A novel common triplet profile for GCrich prokaryotic genomes. *Genomics*. 2011; 97(5): 330–1. https://doi.org/10.1016/j.ygeno.2011.02.005
- 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
- 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
- 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
- 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
- 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., Mavroeidis 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
- 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
- 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
- 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
- Pellett P., Roizman B. Herpesviridae. In: Knipe D.M., Howley P.M., eds. *Fields Virology*. Philadelphia: Lippincott Williams & Wilkins; 2013: 1802–2
- 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
- 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 claylipid 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
- 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.
- 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
- 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 oxido-reductases. *Philis. Trans. R Soc. Lond. B. Biol. Si.* 2013; 368(1622): 20120257. https://doi.org/10.1098/rstb.2012.0257
- Yakovchuk P., Protozanova E., Frank-Kamenetskii M.D. Basestacking 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
- Forterre P. The origin of viruses and their possible roles in major evolutionary transitionsa. Review. Virus Res. 2006; 117: 5–16.
- 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.
- Abedon S.T. Phage evolution and ecology. Adv. Appl. Microbiol. 2009; 67: 1–45. https://doi.org/10.1016/s0065-2164(08)01001-0
- 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
- 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
- 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
- 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/IVI.79.23.14967\_14970.2005
  - 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)
- 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
- 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
- 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.
- 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

- 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
- Zhang S.H., Wang L. A novel common triplet profile for GCrich prokaryotic genomes. *Genomics*. 2011; 97(5): 330–1. https://doi.org/10.1016/j.ygeno.2011.02.005
- 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
- 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
- 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
- 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
- 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.

- Gori F., Mavroeidis 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
- 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
- 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
- Pellett P., Roizman B. Herpesviridae. In: Knipe D.M., Howley P.M., eds. *Fields Virology*. Philadelphia: Lippincott Williams & Wilkins; 2013: 1802–2
- 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
- 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
- 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

### Информация об авторе

Филатов Феликс Петрович<sup>№</sup> — 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 30.05.2022; опубликована 30.07.2022

- 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. *Philis. Trans. R Soc. Lond. B. Biol. Si.* 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
- Forterre P. The origin of viruses and their possible roles in major evolutionary transitionsa. Review. Virus Res. 2006; 117: 5–16.
- 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.
- Abedon S.T. Phage evolution and ecology. Adv. Appl. Microbiol. 2009; 67: 1–45. 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
- 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
- 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
- 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
- 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

- 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)
- Sergienko I.V., Gupal A.M., Vagis A.A. Symmetric code and genetic mutations. *Kibernetika i sistemnyy analiz*. 2016; (2): 73–80. (in Russian)

### Information about the authors

*Felix P. Filatov*<sup>⊠</sup> — 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 30.05.2022; published 30.07.2022

## Приложение к статье Ф.П. Филатова «Нуклеотидные тетрамеры TCGA и CTAG: вирусные ДНК и генетический код (гипотеза)» (Журнал микробиологии, эпидемиологии и иммунобиологии. 2022;99:online-first. DOI: https://doi.org/10.36233/0372-9311-275)

Appendix to the article 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:online-first. DOI: https://doi.org/10.36233/0372-9311-275

### Вирусы, ДНК которых исследована в настоящей работе

Viruses whose DNAs were studied in this research

VIRUSES Надцарство / SuperKingdom

DUPLODNAVIRIA Клада, Домен, Реалм / Clade, Realm, Domain

**HEUNGGONGVIRAE** Клада, Царство / Clade, Kingdom

**UROVIRICOTA** Тип / Phylum

Caudoviricetes Класс / Class

Caudovirales Отряд / Order

<b>Семейство</b> Family(-nae)	Подсемейство Subfamily(-dae)	<b>Род (вирус)</b> Genus (virus)	Вид Species	NC	<b>т.п.н</b> . kbp	S	т	G/A	A/G
Ackermann	-		Erwinia phage phiEa2809	027340	162	F	GC		
			Serratia phage phiMAM1	020083	158	F	GC		
	Aglim	Agtrevirus	Enterobacter phage EspM4VN	049384	161	F	GC		
		Limestonevirus	Dickeya virus Limestone	019925	152	F	AT		
	Cvi	Kuttervirus	Salmonella phage S118	049437	157	F	AT		
			Escherichia virus CBA120	016570	157	F	AT		
			Salmonella phage 38	029042	157	F	AT		
		Taipeivirus	Klebsiella virus UPM 2146	049472	161	F	AT		
	Unclassified		Ralstonia phage RSP15	030948	168	F	AT		
Herelle	Bastilli		Bacillus phage phiAGATE	020081	150	A	AT		
	Brock		Enterococcus phage EFP01	047796	155	F	AT		
	Jasinka		Listeria phage LMTA-148	024787	133	F	AT		
	Spouna		Bacillus phage CP-51	025423	131	F	AT		
	Twort		Lactobacillus phage Bacchae	047924	141	F	AT		
Муо		Acionnavirus	Synechococcus phage S-MbCM100	023584	170	F	AT		1,54
			Synechococcus phage ACG-2014c	019444	176	F	AT		1,56
			Synechococcus phage S-CAM8	021530	171	F	AT		1,54
		Agricanvirus	Erwinia phage Ea35-70	023557	271	F	~		
		Ahtivirus	Synechococcus phage S-ShM2	015281	180	F	AT		
		Alcyoneusvirus	Klebsiella phage K64-1	027399	347	F	AT		
		Alexandravirus	Dickeya virus AD1	048054	262	F	AT		
		Anaposvirus	Synechococcus phage S-CAM1	020837	198	F	AT		
		Aokuangvirus	Synechococcus phage S-CBWM1	048106	139	F	GC		
		Aphroditevirus	Vibrio phage Aphrodite1	042100	238	F	AT		

ORIGINAL RESEARCHES

<b>Семейство</b> Family(-nae)	Подсемейство Subfamily(-dae)	<b>Род (вирус)</b> Genus(-virus)	<b>Вид</b> Species	NC	<b>т.п.н.</b> kbp	S	т	G/A	A/G
	1	Asteriusvirus	Escherichia phage 121Q	025447	349	F	AT		
		Aurunvirus	Cyanophage S-TIM5	019516	161	F	AT		
		Baikalvirus	Pseudomonas phage PaBG	022096	258	F	AT		
		Bellamyvirus	Synechococcus phage Bellamy	047838	205	F	AT		
		Biquartavirus	Aeromonas virus 44RR2	005135	205	F	AT		
		Bixzunavirus	Mycobacterium phage Alice	028986	153	F	GC	1,83	
		Busanvirus	Acidovorax phage ACP17	041997	156	F	AT		
		Chakrabartyvirus	Pseudomonas phage pf16	041881	158	F	AT		
		Charybdisvirus	Synechococcus phage S-CAM3	031906	198	F	AT		
		Chiangmaivirus	Ralstonia phage RSF1	028899	223	F	GC		
		Cymopoleiavirus	Synechococcus phage S-WAM2	031935	186	F	AT		
		Derbicusvirus	Erwinia phage Derbicus	048173	224	F	AT		
		Elvirus	Pseudomonas phage EL	007623	211	F	AT		
		Emdodecavirus	Sinorhizobium phage phiM12	027204	195	F	AT		
		Eneladusvirus	Yersinia phage fHe- Yen9-04	042116	354	F	~		
		Erskinevirus	Erwinia phage phiEaH2	019929	243	F	AT		
		Eurybiavirus	Prochlorococcus phage MED4-213	020845	181	F	AT		
		Goslarvirus	Escherichia phage vB_EcoM_Goslar	048170	237	F	AT		
		lapetusvirus	Erwinia phage PhiEaH1	023610	218	F	GC		
		Ionavirus	Delftia phage PhiW-14	013697	157	F	GC		
		Kanaloavirus	Synechococcus phage S-CAM9	031922	175	F	AT		
		Leucotheavirus	Synechococcus phage S-P4	048102	158	F	AT		
		Libanvirus	Cyanophage P-TIM40	028663	189	F	AT		
		Llyrvirus	Synechococcus phage S-SKS1	020851	208	F	AT		
		Machinavirus	Erwinia phage Machina	042056	242	F	GC		
		Metrivirus	Acinetobacter phage vB_AbaM_ME3	041884	235	F	AT		
		Mieseafarmvirus	Ralstonia phage phiRSL1	010811	231	F	GC		
		Mimasvirus	Cronobacter phage vB_CsaM_GAP32	019401	359	F	AT		
		Neptunevirus	Cyanophage S-RIM50	031242	174	F	AT		
		Nerrivikvirus	Synechococcus phage S-RIM2 R1	020859	175	F	AT		
		Noxifervirus	Pseudomonas phage Noxifer	041994	278	F	GC		
		Palaemonvirus	Prochlorococcus phage P-SSM7	015290	182	F	AT		
		Petsuvirus	Edwardsiella phage pEt-SU	048182	277	F	AT		

					·				
Семейство Family(-nae)	Подсемейство Subfamily(-dae)	<b>Род (вирус)</b> Genus(-virus)	Вид Species	NC	т.п.н. kbp	S	т	G/A	A/G
	I	Phapecoctavirus	Escherichia phage ESCO13	047770	150	F	AT		
		Phikzvirus	Pseudomonas phage SL2	042081	290	F	AT		
		Plaesancevirus	Pseudomonas phage phiPMW	041880	103	F	AT		
		Polybotosvirus	Agrobacterium phage Atu_ph07	042013	490	F	AT		
		Pontusvirus	Synechococcus phage Syn19	015286	175	F	AT		
		Punavirus	Escherichia phage RCS47	042128	115	F	AT		
		Ripduovirus	Ralstonia phage RP12	041911	280	F	GC		
		Risingsunvirus	Erwinia phage vB_ EamM_RisingSun	042018	235	F	AT		
		Saclayvirus	Acinetobact. phage vBAbaMB09Aci	048074	104	F	AT		
		Salacisavirus	Prochlorococcus phage P-SSM2	006883	252	F	AT		
		Salmondvirus	Dickeya phage vB_ DsoM_JA11	048077	255	F	AT		
		Sasquatchvirus	Erwinia phage vB_ EamM_Y3	047880	261	F	AT		
		Seoulvirus	Salmonella phage SPN3US	027402	240	F	AT		
		Shalavirus	Bacillus phage Shbh1	030925	138	F	AT		
		Tamkunvirus	Synechococcus phage S-T4	048049	181	F	AT		
		Tefnutvirus	Synechococcus phage S-IOM18	021536	172	F	AT		
		Tegunavirus	Yersinia phage phiR1- RT	019909	169	F	AT		
		Thaumasvirus	Cyanophage S-TIM4	048015	176	F	AT		
		Thetisvirus	Synechococcus phage S-SM1	015282	174	F	AT		
		Thornevirus	Bacillus phage SP-15	031245	222	F	AT		
		Tidunavirus	Vibrio phage pTD1	041916	239	F	AT		
		Tulanevirus	Aeromonas virus 25	008208	161	F	AT		
		Vellamovirus	Prochlorococcus phage Syn1	015288	191	F	AT		
		Viunavirus	Escherichia phage ECML-4	025446	157	F	AT		
		Wellingtonvirus	Erwinia phage Wellington	048016	245	F	AT		
		Winklervirus	Serratia phage CHI14	041996	171	F	AT		
		Yoloswagvirus	Erwinia phage vB_ EamM_Yoloswag	047815	260	F	AT		
	Eucampy	Firehammervirus	Campylobacter virus CP21	019507	183	F	AT		
	"	Fletchervirus	Campylobacter phage CP30A	018861	134	F	AT		
	Teven	Dhakavirus	Enterobacteria phage IME08	014260	172	F	AT		
	u	Gaprivervirus	Escherichia phage vB_EcoM_VR20	028894	170	F	AT		

ORIGINAL RESEARCHES

	1	1							
<b>Семейство</b> Family(-nae)	Подсемейство Subfamily(-dae)	<b>Род (вирус)</b> Genus(-virus)	Вид Species	NC	<b>т.п.н.</b> kbp	S	Т	G/A	A/G
	ű	Gelderland	Salmonella phage Melville	042044	159	F	AT	[]	
	u	Jiaodavirus	Klebsiella phage JD18	028686	166	F	AT		
	и	Karamvirus	Enterobacter phage CC31	014662	166	F	AT		
	u	Krischvirus	Enterobacteria phage GEC-3S	025425	163	F	AT		
	"	Moonvirus	Citrobacter phage Moon	027331	170	F	AT		
	и	Mosigvirus	Escherichia coli O157 typing phage 3	041863	169	F	AT		
	и	Schizote– quatrovir	Vibrio phage nt-1	021529	248	F	AT		
	ű	Stopekvirus	Enterobacter phage phiEap-3	041980	176	F	AT		
	"	Tequatrovirus	Escherichia phage AR1	027983	167	F	AT		
	Vequinta	Avunavirus	Escherichia phage Av-05	025830	121	F	AT		
	u	Certevirus	Cronobacter phage CR3	017974	149	F	GC		
	u	Seunavirus	Cronobacter phage vB_CsaM_GAP31	019400	148	F	AT		
	u	Vequintavirus	Escherichia phage APECc02	041869	135	F	AT		
Demerec	Ermolyeva	Cetovirus	Vibrio phage Thalassa	042095	129	F	AT		
		Jesfedecavirus	Vibrio phage JSF10	042074	112	F	AT		
		Vipunavirus	Vibrio phage pVp-1	019529	112	F	AT		
	Markadams	Epseptimavirus	Escherichia phage saus132	047884	122	А	AT		
	Mccorquodale	Myunavirus	Pectobacterium phage My1	018837	122	А	AT		
		Novosibvirus	Proteus phage PM135	042090	104	F	AT		
		Pogseptimavirus	Vibrio phage VspSw_1	048151	114	F	AT		
		Shenzhenvirus	Aeromonas phage AhSzw-1	047950	116	F	AT		
		Sugarlandvirus	Klebsiella phage vB_ Kpn_IME260	041899	123	A	AT		
Sipho	Dolichocephalo	Bertelyvirus	Caulobacter phage CcrBL9	048047	322	A	GC	1,68	
	и	Colossusvirus	Caulobacter phage CcrColossus	019406	280	F	GC	1,64	
	u	Poindexter	Caulobacter phage CcrBL10	048045	221	A	GC	1,91	
	"	Shapirovirus	Caulobacter virus Karma	019410	222	А	GC	1,96	
		Franklinbayvirus	Colwellia phage 9A	018088	105	F	AT		
		Godonkavirus	Gordonia phage GodonK	048176	123	F	AT		
		Jenstvirus	Brevibacillus phage Jenst	028805	126	A	AT		
		Lacusarxvirus	Sphingobium phage Lacusarx	041927	130	F	AT		
		Nickievirus	Pseudomonas phage nickie	042091	112	F	~		
		Omegavirus	Mycobacterium virus Baka	042316	112	F	GC	1,55	

<b>Семейство</b> Family(-nae)	Подсемейство Subfamily(-dae)	<b>Род (вирус)</b> Genus(-virus)	Вид Species	NC	<b>т.п.н</b> . kbp	S	т	G/A	A/G
		_ " _	Mycobacterium phage Ariel	028876	110	F	GC	1,56	
		Phicbkvirus	Caulobacter virus Rogue	019408	224	А	GC	1,95	
		Priunavirus	Providencia phage vB_PreS_PR1	041913	119	А	AT		
		Samistivirus	 Streptomyces phage Jay2Jay	029098	134	F	~		
		Spbetavirus	Bacillus virus SPbeta	001884	134	F	AT		
		Trinavirus	Rhodococcus phage Trina	042040	139	F	AT		
		Weaselvirus	Rhodococcus phage Weasels2	041887	135	F	AT		
		Wilnyevirus	Streptomyces phage BillNye	042105	127	F	GC		
		PEP	LOVIRICOTA Тип / Phylum						
		Hei	<b>rviviricetes</b> Класс / Class						
		Her	<b>pesvirales</b> Отряд / Order						
AlloHV		Batrachovirus	Ranid herpesvirus 1, Lucke tumor	008211	221	В	GC		
			Ranid herpesvirus 2	008210	232	М	GC		
			Ranid herpesvirus 3	034618	208	F	GC		
		Cyprinivirus	Cyprinid herpesvirus 1	019491	291	А	GC		
		Ictalurivirus	Ictalurid herpesvirus 1	001493	134	А	GC		
HerpesV	Alphaherpes	lltovirus	Gallid alphaherpesvirus 1	006623	149	Е	GC		
		Mardivirus	Gallid alphaherpesvirus 3	002577	164	D	GC		
			Gallid alphaherpesvirus 2	002229	178	D	GC		
		Simplexvirus	Fruit bat alphaherpesvirus 1	024306	157	D	GC	1,55	
			Human alphaherpesvirus 1	001806	152	D	GC	2,15	
			Human alphaherpesvirus 2	001798	155	D	GC	2,38	
			Ateline alphaherpesvirus 1	034446	147	Е	GC	3,08	
			Leporid Herpesvirus 4	029311	124	D	GC	1,97	
			Macropodid Herpesvirus 1	029132	140	М	GC	1,12	
			Cercopithecine alphaherpesvirus 2	006560	151	D	GC	3,16	
			Chimpanzee alphaherpesvirus	023677	153	D	GC	2,13	
			Macacine alphaherpesvirus 1	004812	157	D	GC	2,92	
			Papiine herpesvirus 2	007653	156	D	GC	3,18	
			Saimirine alphaherpesvirus 1	014567	157	Е	GC	2,04	
		Varicellovirus	Human alphaherpesvirus 3	001348	125	Е	AT		
			Canid alphaherpesvirus 1	030117	125	Е	AT		

ORIGINAL RESEARCHES

<b>Семейство</b> Family(-nae)	Подсемейство Subfamily(-dae)	<b>Род (вирус)</b> Genus(-virus)	Вид Species	NC	<b>т.п.н.</b> kbp	S	т	G/A	A/G
	-	1	Equid alphaherpesvirus	001491	150	F	GC		
			Saimiriine alphaherpesvirus 1	014567	157	Е	GC		
		Scutavirus	Testudinid alphaherpesvirus 3	027916	160	Е	GC		
	Betaherpes	Cytomegalovirus	Caviid betaherpesvirus 2	020231	234	А	GC		
			Human betaherpesvirus 5	006273	236	D	GC		
		Muromegalovirus	Murid betaherpesvirus 1	004065	230	F	GC		
			Rat CMV Maastricht	002512	230	А	GC	1,56	
		Proboscivirus	Elephantid betaherpesvirus 1	020474	180	A	GC		
		Roseolovirus	Human betaherpesvirus 6A	001664	159	А	AT		
		u	Human betaherpesvirus 6B	000898	162	A	AT		
		u	Human betaherpesvirus 7RK	001716	153	A	AT		
		ű	Human betaherpesvirus 7JI	U43400	145	A	AT		
		Cytomegalovirus	Suid betaherpesvirus 2	022233	128	А			
		Roseolovirus	Murine roseolovirus	033620	174	F	GC		
		"	Macaca nemestrina	030200	138	А	GC		
		"	Tupaiid betaherpesvirus 1	002794	160	F	GC		
	Gammaherpes	Lympho- criptovirus	Callitrichine gammaherpesvirus 3	004367	150	F	GC		
		u	Human gammaherpesvirus 4	007605	172	В	GC		
		Macavirus	Alcelaphine gammaherpesvirus 1	002531	131	М	GC		
		Percavirus	Equid gammaherpesvirus 5	026421	182	М	GC		
		Rhadinovirus	Ateline GHV3	001987	108	F	GC		
			Dolphin GHV1	035117	167	М	GC		
			Human gammaherpesvirus 8	009333	138	С	GC		
MalacoHV		Aurivirus	Abalone herpesvirus Victoria/AUS	018874	212	М	GC		
		Ostreavirus	Ostreid herpesvirus 1	005881	207	М	GC		
	VA	A <b>RIDNAVIRIA</b> Клад	ца, Домен, Реалм / Clade, F	Realm, Doma	ain				
		BAMFORDVIRA	<b>\Е</b> Клада, Царство / Clade	Kingdom					
		NUCLEO	CITOVIRICOTA Тип / Phyl	um					
		Мед	<b>javiricetes</b> Класс / Class						
		Alg	<b>gavirales</b> Отряд / Order						
Phycodna		Chlorovirus	Paramecium bursaria Chlorella 158	009899	345	F	AT		
		Coccolivirus	Emiliania huxleyi virus 86	007346	407	М	AT		

<b>Семейство</b> Family(-nae)	Подсемейство Subfamily(-dae)	<b>Род (вирус)</b> Genus(-virus)	Вид Species	NC	<b>т.п.н</b> . kbp	S	т	G/A	A/G
		Prasinovirus	Micromonas pusilla virus 12T	020864	206	F	AT		
		Prymnesiovirus	Phaeocystis globosa virus	021312	460	F	AT		
		Raphidovirus	Heterosigma akashiwo virus 01	038553	275	F	AT		
		Unclass	Aureococcus anophagefferens virus	024697	371	F	AT		
		Im	<b>itervirales О</b> тряд / Order						
Mimi		Mimivirus	Cafeteria roenbergensis virus	014637	617	F	AT		
		Pima	<b>ascovirales</b> Отряд / Order						
Iridoviridae	Alphairido	Ranavirus 3	Bohle iridovirus (FROG)	038507	104	F	GC		
		Megalocitivirus	Spleen/kidney necrosis virus	003494	111	М	GC		
	Betairido	Chloriridovirus	Invertebrate iridescent virus 22	023615	196	F	AT		
Asco		Ascovirus	Heliothis virescens ascovirus 3e	009233	186	F	AT		
			Spodoptera frugiperda ascovirus 1a	008361	157	F	~		
			Trichoplusia ni ascovirus 2c	008518	174	F	AT		
		Toursvirus	Diadromus pulchellus ascovirus 4a	011335	119	F	~		
Marseille		Marseillevirus	Golden Marseillevirus	031465	360	F	AT		
			Melbournevirus	025412	369	F	AT		
		Pokl	<b>kesviricetes</b> Класс / Class s <b>fuvirales</b> Отряд / Order						
Asfar		Asfivirus	African swine fever virus E75	044958	181	D	AT		
			African swine fever virus Benin 97/1	044956	182	F	AT		
		Ch	<b>itovirales</b> Отряд / Order						
Pox	Chordopox		Eptesipoxvirus strain Washington	035460	177	D	AT		3,24
			Ecromelia virus strain Moscow	004105	210	F	AT		2,01
			Fowlpoxvirus	002188	289	А	AT		2,23
			NY_014 poxvirus	035469	200	D	AT		2,39
			Rabbit fibroma virus	001266	160	М	AT		1,53
			Sea otter poxvirus	037656	128	D	AT		2,20
			Seal parapoxvirus	035188	128	D	GC		
			Vaccinia virus	006998	195	F	AT		
			Variola virus	001611	186	F	AT		
			Penguinpox virus	024446	307	F	AT		
			Canarypox virus	005309	360	F	AT		
			Pteropox virus	030656	133	F	AT		
			Squirrelpox virus Berlin_2015	035797	143	F	AT		

ORIGINAL RESEARCHES

<b>Семейство</b> Family(-nae)	Подсемейство Subfamily(-dae)	<b>Род (вирус)</b> Genus(-virus)	<b>Вид</b> Species	NC	<b>т.п.н.</b> kbp	S	т	G/A	A/G
	Entomopox		Amsacta moorei entomopoxvirus "L"	002520	232	Μ	AT		4,62
			Anomala cuprea entomopoxvirus	023426	246	F	AT		4,01
			Melanoplus sanguinipes	001993	236	А	AT		4,47
	Вирусы без	промежуточных	доменов / Viruses witho	ut intermedia	ate doma	ains			
Baculo		Alphabaculovirus	Agrotis ipsilon nucleopolyhedrovirus	007921	148	F	AT		
			Mamestra configurata NPV-A	003529	155	М	AT		
		Betabaculovirus	Epinotia aporema granulovirus	018875	119	F	AT		
			Pseudalatia unipuncta granulovirus	013772	177	Μ	AT		
		Gammabaculo	Neodiprion sertifer nucleopolyhedro	005905	86	Μ	AT		
			Neodiprion abietis NPV	008252	84	F	AT		
		Deltabaculovirus	Culex nigripalpus nucleopolyhedro	003084	108	F	GC		
Nima		Whispovirus	White spot syndrome virus	003225	309	F	AT		
Nudi		Alphanudivirus	Oryctes rhinoceros nudivirus	011588	128	F	AT		
		Betanuduvirus	Helicoverpa zea nudivirus 2	004156	232	Μ	AT		
	Неклассифициров	анные архейные	вирусы DSDNA / Unclas	sified archa	eal DSDI	NA viru	ses		
		Halovirus	Halovirus HGTV-1	021328	144	F	GC		
			Halovirus HCTV-5	021327	102	F	AT		
			Halovirus HVTV-1	020158	102	F	AT		
			Halovirus HCTV-1	021330	103	F	AT		
	Неклас	ссифицированны	ые архейные вирусы / U	nclassified v	iruses				
Pitho		Pithovirus	Pithovirus sibericum	023423	610	М	~		
Hytrosa		Glossinavirus	Pallidipes salivary gland hypertrophy	010356	190	F	AT		
		Muscavirus	Musca salivary gland hypertrophy	010671	124	М	AT		

Примечание. S — макроструктура генома (M — структура, выходящая за пределы данной классификации). G/A или A/G — соотношение [G+T] : [A+C] или [A+C] : [G+T] свыше 1. Темным выделены соотношения выше 1,5. NC — номер генома в GenBank. Note. S — the macrostructure of the genome (M — the structure falling out of scope of this classification). G/A or A/G — the ratio [G+T] : [A+C] or [A+C] : [G+T] greater than 1. Ratios greater than 1.5. NC — the genome number in the GenBank – are highlighted in dark.