



Тетрануклеотидный профиль герпесвирусных ДНК

Филатов Ф.П.^{1,2✉}, Шаргунов А.В.¹

¹ФГБНУ «Научно-исследовательский институт вакцин и сывороток им. И.И. Мечникова», 105064, Москва, Россия;

²ФГБУ «Национальный исследовательский центр эпидемиологии и микробиологии имени почетного академика Н.Ф. Гамалеи», 123098, Москва, Россия

Введение. Герпесвирусные ДНК (около 90% всех полногеномных последовательностей семейства *Herpesvirales*, представленных в GenBank) содержат в минимальной концентрации один из двух тетрануклеотидов — CTAG или TCGA. «Недопредставленность» CTAG ранее наблюдалась только в ДНК некоторых бактерий и фагов. Ранее выявленная «недопредставленность» метилируемого димера CpG находит свое выражение в низкой концентрации TCGA в ДНК герпесвирусов.

Цель работы — продолжение анализа формальных характеристик герпесвирусных ДНК, а также сопоставление их с плотностью ДНК-микргомологии вирус/хозяин и с геномной макроструктурой герпесвирусов.

Материалы и методы. Проанализированы по 20 штаммов и изолятов каждого из пяти типов вирусов герпеса человека (HHV1, HHV2, HHV3, HHV4, HHV5), 10 штаммов HHV8, 5 штаммов HHV6A, 4 штамма HHV6B и 3 штамма HHV7. Для определения частоты тетрануклеотидов использовали инструменты GenBank, а для сравнения — фрагменты ДНК человека размером с ДНК герпесвирусов.

Результаты. Минимальная концентрация CTAG в ДНК герпесвирусов в основном характерна для двух- и односегментных геномов с прямыми или инвертированными концевыми повторами (классов А, D и E), тогда как минимальная плотность TCGA — главным образом для значительно менее структурированной ДНК (классов В, С и F). По нарастанию плотности CTAG геномы герпесвирусов человека образуют последовательность, близкую к последовательности 20 нт-гомологии ДНК герпесвирус/человек, организованной по нарастанию плотности, что также коррелирует с макроструктурой ДНК. Параллель этой минимизации со структурой ДНК вирусов герпеса или с их принадлежностью к тому или иному подсемейству в литературе не отмечена. Хотя герпесвирусные ДНК довольно велики (125–295 Кб), некоторые из них (например, ДНК HHV4, HHV5 и HHV7) демонстрируют заметные отклонения от второго правила четности ДНК и, таким образом, могут служить компонентом вирусных молекулярных сигнатур.

В **Обсуждении** предлагаются возможные гипотезы происхождения некоторых из отмеченных явлений.

Ключевые слова: герпесвирусная ДНК; тетрануклеотидный профиль; «недопредставленность» CTAG/TCGA; второе правило четности ДНК.

Источник финансирования. Авторы заявляют об отсутствии финансирования при проведении исследования.

Конфликт интересов. Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Для цитирования: Филатов Ф.П., Шаргунов А.В. Тетрануклеотидный профиль герпесвирусных ДНК. *Журнал микробиологии, эпидемиологии и иммунобиологии*. 2020; 97(3): 216–226.
DOI: <https://doi.org/10.36233/0372-9311-2020-97-3-3>

Поступила 22.04.2020

Принята в печать 25.05.2020

Tetranucleotide Profile of Herpesvirus DNA

Felix P. Filatov^{1,2✉}, Alexander V. Shargunov¹

¹Mechnikov Federal Research Institute of Vaccines and Sera, 105064, Moscow, Russia;

²National Research Centre for Epidemiology and Microbiology named after the honorary academician N.F. Gamaleya, 123098, Moscow, Russia

Introduction. Herpesvirus DNAs (about 90% of the total genomic sequences of the *Herpesvirales* family presented in GenBank) contain at a minimum concentration one of the two tetranucleotides, CTAG or TCGA. The “underrepresentation” of CTAG was previously observed only in the DNA of some bacteria and phages. The **aim** of the study was the further analysis of the formal characteristics of herpesvirus DNA, as well as their comparison with the density of the virus/host DNA microhomology and with the genomic macrostructure of herpes viruses.

Materials and methods. Twenty strains and isolates of each of the five types of human herpes viruses (HHV1, HHV2, HHV3, HHV4, HHV5), 10 strains of HHV8, 5 strains of HHV6A, 4 strains of HHV6B and 3 strains of HHV7 were analyzed. GenBank tools were used to determine the frequency of tetranucleotides, and human DNA fragments with size matched herpesvirus DNA were used for comparison.

Results. Minimum CTAG concentration in DNA of herpes viruses is mainly characteristic of two- and single-segment genomes with direct or inverted terminal repeats (classes A,D,E), while the minimum TCGA density is characteristic mainly for DNA that is significantly less structured (classes B,C,F). By increasing CTAG density, human herpes viruses form a sequence close to the sequence of increasing the homology density of 20 nt with human DNA, which also correlates with the macrostructure of DNA. A parallel of this minimization with the DNA structure of herpes viruses or with their belonging to one or another subfamily — as well as the context of the “minimal” CpG (that is, TCGA) — is not noted in the literature. Although herpesvirus DNA is quite large (125–295 Kb), some of them (for example, HHV4, HHV5 and HHV7 DNA) show noticeable deviations from the second DNA parity rule, and can thus serve as a component of the molecular signature. The **Discussion** suggests possible hypotheses for the origin of some of the observed phenomena.

Keywords: *Herpesvirus DNA; tetranucleotide profile; CTAG/TCGA deficiency; Chargaff Second Parity Rule.*

Acknowledgments. The study had no sponsorship.

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

For citation: Filatov F.P., Shargunov A.V. Tetranucleotide profile of herpesvirus DNA. *Journal of microbiology, epidemiology and immunobiology = Zhurnal mikrobiologii, èpidemiologii i immunobiologii.* 2020; 97(3): 216–226. (In Russ.).

DOI: <https://doi.org/10.36233/0372-9311-2020-97--3-3>

Received 22 April 2020
 Accepted 25 May 2020

Введение

Герпесвирусы семейства *Herpesviridae*, включая HHV, делятся на три подсемейства: альфа-HV, бета-HV и гамма-HV [1]. Другой классификацией HV является классификация по макроструктуре ДНК (рис. 1). Она не совсем совпадает с делением на подсемейства и, в соответствии с общепринятыми взглядами [2], образует 6 классов — от А до F.

Альфа-HHV (HHV1, HHV2 и HHV3), а также бета-HHV (HHV5) содержат двухсегментную ДНК; каждый сегмент ограничен взаимно инвертированными мономерными концевыми повторами TR₁ (ДНК классов D и E). ДНК HHV класса А (бета-HHV: HHV6А, HHV6В и HHV7) представляет собой несегментированную уникальную линейную последовательность, ограниченную прямыми мономерными концевыми повторами TR₁, содержащими по два «островка» теломероподобных гексануклеотидов. Гамма-HHV содержат ДНК классов В и С, которые имеют уникальную последователь-

ность, ограниченную прямыми тандемно организованными короткими (т.е. не мономерными) повторами, TR₂.

Сегменты и концевые повторы ДНК класса F не структурированы, хотя иногда они просто не показаны в GenBank. Существуют герпесвирусные ДНК с более экзотической макроструктурой (например, скутавирусы), но их очень немного. Данные, полученные нами в предлагаемой работе, позволили нам объединить классы А, D и E в одну группу (сегменты ДНК, ограниченные мономерными концевыми повторами), а В, С и F — в другую (отсутствие макроструктуры или нефиксированное число тандемно организованных коротких концевых повторов).

Ранее мы заметили, что молекулы ДНК герпесвируса и его хозяина содержат короткие (20–29 нт) идентичные последовательности — микрогомологии, концентрация которых не случайна и, как мы полагаем, объясняется длительными (в эво-

DNA macrostructure	class	TR	HHV	subfamily
	D	TR ₁	1/2, 5	alpha, beta
	E	TR ₁	3	alpha
	A	TR ₁	6A/B, 7	beta
	B	TR ₂	8	gamma
	C	TR ₂	4	gamma
	F	TR ₀	—	—

Рис. 1. Основные макроструктурные классы ДНК HHV (пропорции длин геномных фрагментов произвольны).

TR₁ — мономерные терминальные повторы (одиночные прямоугольники); TR₂ — тандемные повторы (с нефиксированным числом повторов); TR₀ — терминальных повторов нет. Группа BC[F] ДНК HHV выделена серым цветом.

Fig. 1. Basic macrostructural classes of HHV DNA (the proportions of the lengths of the genome fragments are arbitrary).

TR₁ — monomeric terminal repeats (single rectangles), TR₂ — tandem organized repeats (non-fixed number of repeats), TR₀ — no terminal repeats. BC[F] group of the HHV DNA (see text) is highlighted in gray.

люционном масштабе) близкими межгеномными отношениями между партнерами [3]. Позже [4, 5] мы обнаружили, что такие микрогомологии имеют характерные особенности распределения в геноме герпесвируса, концентрируясь в основном в его концевых (прямых или инвертированных) повторах, особенно в тех областях TR, в которых нет генов. Самое интересное — это последовательность видов HHV по нарастанию плотности геномных микрогомологий вирус/хозяин, которая согласуется с макроструктурой ДНК: меньшая плотность — в двухсегментных ДНК, большая — в несегментированных. В качестве рабочей гипотезы мы предположили, что двухсегментные вирусные геномы, у которых терминальные повторы взаимно инвертированы, склонны в латентном состоянии замыкаться, скорее, на себя (подобно эписомам) и меньше взаимодействовать с хозяйской ДНК, в то время как односегментные, имеющие прямые повторы, могут вытягиваться вдоль хозяйской ДНК, что облегчает межгеномное взаимодействие.

В качестве подхода для анализа мы использовали сравнение частот нуклеотидов в молекулах ДНК и второе правило четности ДНК Чаргаффа, CPR2 [6], которое становится наглядным в ДНК размером более 100 000 нт [7, 8]. CPR2 формулируется так же, как первое (CPR1), но относится только к одной цепи ДНК. Оно имеет приблизительную точность, которая увеличивается по мере удлинения анализируемой цепи. Оно относится не только к моно-, но и к олигонуклеотидам до 10–15 нт — с уменьшением строгости по мере удлинения анализируемого олигонуклеотида [7, 9]. В метагеномике часто используется тетра-нуклеотидный анализ для формирования молекулярных сигнатур [10]. Частота тетра-нуклеотидов (TN) в геномах герпесвирусов достоверно соответствует CPR2 и обеспечивает более детальную характеристику ДНК, чем моно-, ди- и тринуклеотиды [7, 11]. В принципе, симметрии TN генома HHV были описаны ранее [12], но они только подтвердили соответствие CPR2. Наш подход обнаруживает другие необычные свойства этих геномов.

Цель настоящей работы — продолжение анализа формальных характеристик герпесвирусных ДНК, а также сопоставление их с плотностью ДНК-микрогомологий вирус/хозяин и с геномной макроструктурой герпесвирусов.

Материалы и методы

Мы проанализировали около 90% нуклеотидных последовательностей полноразмерных молекул вирусной ДНК каждого рода всех трех семейств герпесвирусов позвоночных и беспозвоночных, содержащихся в GenBank. Проанализировав по 20 штаммов и изолятов каждого из 5 типов HHV (HHV1, HHV2, HHV3, HHV4, HHV5), 10 штаммов

HHV8, 5 штаммов HHV6A, 4 штамма HHV6B и 3 штамма HHV7, мы убедились в практической идентичности внутривидовых результатов и поэтому приводим в таблицах данные только по ДНК референс-штаммов каждого вида герпесвирусов.

Для сравнения использовали ДНК человека длиной 1,5 мегануклеотида (5 фрагментов по 300 000 нт каждый):

- фрагмент Chr 03 163229646–163529646;
- фрагмент Chr 05 29372672–29672672;
- фрагмент Chr 14 64016329–64316329;
- фрагмент Chr 21 15306102–15606102;
- фрагмент Chr 21 33931862–34231862.

Для определения частоты TN мы использовали инструменты GenBank.

Результаты

Мы проанализировали TN-состав полностью секвенированных ДНК практически всех герпесвирусов отряда *Herpes virales*, содержащихся в GenBank. Класс ДНК, т.е. преобладание G + C или A + T в одной из ее цепей, не дает слишком много в этом отношении, разделяя HHV на две группы по классам:

- класс AT — HHV3 (альфа) и HHV6A, 6B, 7 (бета);
- класс GC — HHV1,2 (альфа), 5 (бета) и 4,8 (гамма).

Вместе с тем динуклеотидный анализ хорошо иллюстрирует CPR2 [8], по которому $A \approx T$, $C \approx G$, $C + T \approx A + G$ и $C + A \approx T + G$ для одной нити ДНК. Это определяется размером герпесвирусных ДНК — 125–295 Кб.

Общее количество TN составляет 256 (4⁴). Чтобы избежать влияния класса ДНК (GC или AT) на результаты подсчета, показанного ранее [8, 13, 14], мы проанализировали только те TN, которые содержат все 4 различных основания, т.е. «полные» TN — 4TN. В ДНК HHV1 (класс GC) наименьшим («недопредставленным») является именно такой 4TN — СТАГ (91 нт на геном из ~150 тыс. пар оснований). В ДНК HHV6A (класс AT) число СТАГ также близко к наименьшему (303 нт) среди всех тетрамеров и является наименьшим из 4TN. Меньше только АGGG (287) и GGCT (296) — в соответствии с классом ДНК. Это обстоятельство еще раз обосновывает выбор именно «полных» наборов для анализа TN.

Из 256 TN только 24 состоят из всех 4 нуклеотидов ($P_4 = 4! = 24$). Эти 24, в свою очередь, делятся на две группы: 8 из них (октет А) при инверсии не меняются, например СТАГ|СТАГ, остальные (октет В) составляют пары В1 и В2 взаимноинвертированных неидентичных TN, например, СТАГ|ТСАГ. В таблицах и на рисунках октеты А и В показаны отдельно. Для корректного сравнения данные представлены в процентах от сумм частот TN каж-

дого октета, А и В. В **табл. 1** сравниваются данные по всем 24 обсуждаемым TN октетов А и В (референс-штаммы) всех известных типов HHV. Данные табл. 1 показывают, что тетрамер СТАГ «недопредставлен» в геномах всех HHV, за исключением HHV7; в последнем «недопредставлен» тетрамер АСТГ. В ДНК HHV4 «недопредставлен» тетрамер ТСГА (как и в геноме человека).

В соответствии со снижением «недопредставленности» СТАГ геномы HHV образуют последова-

тельность, которая напоминает последовательность ДНК-микрогомологий вирус/хозяин по нарастанию их плотности (**рис. 2**): наибольшая «недопредставленность» СТАГ характерна для двухсегментных ДНК, наименьшая — для односегментных.

В то же время ДНК каждого HHV содержит «сверхпредставленные» TN, которые также характерны для геномов определенной макроструктуры: АСГТ — для двухсегментной ДНК (классы D, E),

Таблица 1. Профиль идентичных (октет А) и неидентичных (октеты В1 и В2) тетрамеров, содержащих 4ТНТ 8 типов HHV (референс-штаммы), выраженный в процентах от общего числа результатов по октетам А и В раздельно

Table 1. Profile of identical (octet A) and non-identical (octets B1 and B2) tetramers containing four different nucleotides (4TN) of eight types of HHV (reference strains), expressed as a percentage of the total number of the octets A and B separately

HHV	1	2	3	5	6A	6B	7	4	8	4TN Human
subfamily	alpha			beta			gamma			
TR	TR_1						TR_2			
4TN octet A										
СТАГ	2,8	2,6	4,9	5,0	7,5	7,3	13,9	9,2	8,3	12,2
ТСГА	14,5	15,1	10,6	10,9	15,0	14,7	9,1	6,6	8,6	2,4
AGCT	12,7	13,7	8,2	11,2	10,3	11,2	17,7	18,4	15,1	18,9
GATC	13,2	13,5	12,9	11,6	13,6	14,0	10,8	10,3	10,0	11,4
CATG	15,1	13,1	15,7	13,7	13,8	14,4	14,1	19,6	16,2	21,6
TGCA	11,2	11,7	15,6	12,6	15,0	14,9	17,9	17,9	15,9	22,3
АСГТ	16,9	17,5	18,7	20,9	14,8	14,1	8,1	9,6	13,7	3,0
GTAC	13,6	12,8	13,4	14,1	10,0	9,4	8,4	8,4	12,2	8,2
4TN octet B1										
CTGA	5,3	5,4	3,9	5,9	6,9	7,1	7,6	8,9	7,4	11,2
TACG	7,2	7,0	8,3	7,6	5,9	5,3	3,4	2,6	4,8	1,0
GCAT	7,2	6,8	8,5	5,3	6,8	6,8	7,3	7,6	6,6	7,6
AGTC	5,5	5,4	4,3	5,9	5,7	5,7	5,9	7,9	5,8	6,3
CAGT	5,2	4,6	6,1	5,9	6,0	6,3	7,8	8,9	8,0	9,6
ATCG	8,2	7,5	8,3	6,5	7,7	7,2	4,9	3,0	4,5	1,1
GCTA	3,4	3,2	4,4	4,4	4,1	4,0	4,9	4,8	5,1	6,3
TGAC	6,9	7,1	5,7	8,0	6,5	6,6	6,9	8,2	8,0	6,7
4TNs octet B2										
TCAG	6,3	6,0	4,9	6,0	6,7	7,3	7,9	10,6	7,3	10,9
CGTA	7,5	8,0	7,9	7,7	6,1	6,0	4,0	3,3	4,7	1,0
ATGC	7,7	8,8	8,1	5,6	6,1	6,7	7,5	7,0	6,7	7,5
GACT	5,0	5,7	5,2	5,0	5,4	5,7	4,9	5,4	6,3	6,7
ACTG	5,0	4,5	6,3	5,9	6,9	6,7	7,3	6,2	8,7	10,0
CGAT	7,2	6,8	8,5	5,3	6,8	6,8	5,0	3,0	6,6	1,1
TAGC	3,7	3,1	4,7	5,2	5,3	5,1	7,8	4,8	4,4	6,3
GTCA	7,6	7,2	5,6	8,3	6,6	6,2	6,9	8,4	7,7	6,7

Примечание. Жирными буквами обозначены тетрамеры СТАГ и димеры CpG в обсуждаемых в тексте тетрамерах (соответствующие ячейки обоих октетов выделены серым цветом).

Note. Bold letters are tetramers CTAG and dimers CpG in the tetramers discussed in text (corresponding cells of the both octets are highlighted in gray).

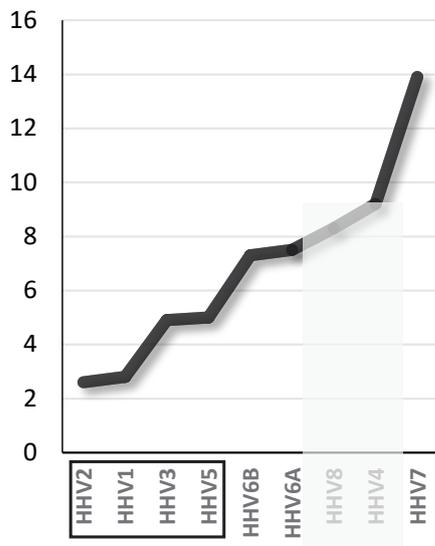


Рис. 2. Частота (%) СТАГ среди 4ТН октета А в ДНК герпесвирусов человека.

HHV4 и HHV8 (классы ДНК ВС[F]) отмечены серым. В прямоугольнике — четыре HHV класса DE.

Fig. 2. Frequency (%) of CTAG among other 4TNs of the octet A in human herpesvirus DNA.

HHV4 and HHV8 (BC[F] DNA classes) are marked in gray. In the rectangle there are four HHVs of DE classes.

TGCA — для односегментной (класс А, розеоловирусы) и CATG — для односегментных (классы В, С). Однако, поскольку ДНК вирусов герпеса, отличных от человека, очень слабо представлена для видов-хозяев в GenBank, мы не приводим результаты анализа по «максимальным» TN.

Колонки чисел, относящихся к ДНК каждого вируса, представляют собой TN-профили ДНК,

и они — в случае HHV4, 8 и 7 — демонстрируют определенное сходство с профилем ДНК человека. В некоторых случаях (HHV4, 5, 7) представленные попарно TN октетов В1 и В2 демонстрируют характерные отклонения от CPR2, которые, вероятно, связаны с недостаточными размерами ДНК этих вирусов или с недостаточным числом штаммов в GenBank, которое не обеспечивает достоверность соответствующих данных. Позитивная сторона таких отклонений заключается в том, что они могут быть использованы в качестве компонентов молекулярных сигнатур этих вирусов.

Обращает на себя внимание, что разница между максимальными и минимальными значениями в октете А существенно больше, чем в октете В. В тех случаях, когда показатели плотности TN октета В меньше, чем октета А, их «недопредставленность» напрямую связана с классом ДНК, т.е. они имеют вид [TA/AT]GC/CG]; левая и правая пары тетрамера могут меняться местами, а знак «/» означает «или». Таких TN — 8, и их причастность к формированию класса ДНК не имеет значения для использования — вместе с другими полными тетрамерами — в качестве молекулярных сигнатур. В табл. 2 суммированы данные по «минимальным» («недопредставленным») 4TN ДНК герпесвирусов человека. Анализ серий штаммов (до 20) одного и того же типа HHV показал почти полную идентичность результатов, что в первом приближении позволило считать полученные результаты достаточно надежными.

Далее мы провели TN-анализ полностью секвенированных ДНК почти всех других видов вирусов суперсемейства *Herpesvirales* (табл. 3).

Таблица 2. Обобщенная версия данных о ДНК HHV (семейство *Herpesviridae*)

Table 2. A generalized version of the data on the human herpesviruses DNA (family *Herpesviridae*)

Подсемейство Subfamily	Род Genus	Вид Species	Номер Reference	Размер ДНК, Кб Size of the DNA, Kb	Класс Class	Тип Type	«Недопредставленные» 4TN 4TN _{min}	Количество изученных штаммов Number of studied strains
alpha	Herpes simplex virus	HHV1	NC_001806	155	E	GC	CTAG	20
	Herpes simplex virus	HHV2	NC_001798	155	E	GC	CTAG	20
	Varicella-Zoster virus	HHV3	NC_001348	125	D	AT	CTAG	20
beta	Cytomegalovirus	HHV5	NC_006273	236	E	GC	CTAG	20
	Roseoloviruses	HHV6A	NC_001664	159	A	AT	CTAG	5
		HHV6B	NC_000898	162	A	AT	CTAG	4
		HHV7	NC_001716	153	A	AT	ACGT	3
gamma	Lymphocryptovirus	HHV4	NC_007605	172	C	GC	TCGA	20
	Rhadinovirus	HHV8	NC_009333	138	B	GC	CTAG	10

Примечание. Не-CTAG_{min} ДНК выделены курсивом (CpG — жирным курсивом). HHV, в ДНК которых CpG>GpC, выделены серым цветом.

Note. Non-CTAG_{min} DNAs are highlighted in italics (CpG — in bold italics). HHVs in the DNA with CpG>GpC are highlighted in gray.

Таблица 3. «Непредставленные» TN в ДНК герпесвирусов животных (семейства *Herpesviridae*, *Alloherpesviridae* и *Malacoherpesviridae*)**Table 3.** A generalized version of the data on the DNA of animal herpesviruses (families *Herpesviridae*, *Alloherpesviridae* and *Malacoherpesviridae*)

Подсемейство Subfamily	Род Genus	Вид Species	Номер Reference	Размер ДНК, Кб Size of the DNA, Kb	Класс Class	Тип Type	«Непредстав- ленные» 4TN 4TN _{min}	
Семейство: <i>Herpesviridae</i> (герпесвирус животных) Family: <i>Herpesviridae</i> (animal HV)								
alpha	Iltovirus	Gallid AHV1	NC_006623	149	D	AT	CTAG	
		Psittacid AHV1	NC_005264	163	D	GC	CTAG	
	Mardivirus	Anatid AHV1	NC_013036	158	F	AT	CTAG	
		Columbid AHV1	NC_034266	204	E	GC	CTAG	
		Falconid AHV1	NC_024450	204	E	GC	CTAG	
		Gallid AHV2	NC_002229	178	E	AT	CTAG	
		Gallid AHV2	MF431495	178	E	AT	CTAG	
		Gallid AHV3	NC_002577	164	E	GC	CTAG	
		Meleagrid AHV1	NC_002641	159	E	AT	CTAG	
		Sphenicid AHV1	NC_033464	165	D	AT	CTAG	
		Scutavirus	Testudinid HV3	NC_002794	196	D*	AT	TGCA
			Simplex virus	Ateline AHV1	NC_034446	147	D	GC
	Cercopithecine AHV2	NC_006560		151	E	GC	CTAG	
	Panine HV3	NC_023677		153	E	GC	CTAG	
	Leporidae AHV4	NC_029311		124	E	GC	CTAG	
	Macacine AHV1	NC_004812		157	E	GC	CTAG	
	Macropodid AHV1	NC_029132		140	D	GC	CTAG	
	Papiine AHV2	NC_007453		156	E	GC	CTAG	
	Saimiriine AHV1	NC_014567		157	D	GC	TGCA	
	Fruit bat AHV1	NC_024306		149	E	GC	GTAC	
	Varicella virus	Bovine AHV1		NC_001847	135	D	GC	CTAG
		Bovine AHV5		NC_005261	138	F	GC	CTAG
		Bubaline AHV1		NC_043054	137	F	GC	CTAG
		Cercopithecine AHV9		NC_002686	125	D	AT	CTAG
		Equid AHV3		NC_024771	184	E	GC	CTAG
		Suid AHV1	NC_006151	143	D	GC	CTAG	
		Canid AHV1	NC_030117	125	D	AT	TCGA	
		Equid AHV4	NC_001844	146	D	GC	TCGA	
Felid AHV1		NC_013590	136	D	AT	TCGA		
Equid AHV1		NC_001491	150	D	GC	GATC		
Equid AHV8	NC_017826	149	F	GC	GATC			
Equid AHV9	NC_011644	148	D	GC	GATC			
beta	Cytomegalovirus	Aotine BHV1	NC_016447	219	E	GC	CTAG	
		Caviid BHV2	NC_020231	234	A	GC	CTAG	
		Cercopithecine BHV5	NC_012783	226	A	GC	CTAG	
		Papio ursinus CMV	NC_027016	226	F	GC	CTAG	
		Cynomolgus CMV	NC_033176	224	A	AT	CTAG	
		Macacine BHV3	NC_006150	221	F	AT	CTAG	
		Panine BHV2	NC_003521	241	D	GC	CTAG	
		Saimiriine BHV4	NC_016448	197	E	AT	CTAG	

Окончание табл. 3 / End of Table 3

Подсемейство Subfamily	Род Genus	Вид Species	Номер Reference	Размер ДНК, Кб Size of the DNA, Kb	Класс Class	Тип Type	«Недопредстав- ленные» 4TN 4TN _{min}
gamma	Muromegalovirus	Murid BHV1	NC_004065	230	F	GC	CTAG
		Murid BHV8	NC_019559	203	F	AT	CTAG
		Rat CMV Maastricht	NC_002512	230	A	GC	CTAG
	Proboscivirus	Elephant BHV4	NC_028379	206	F	GC	CTAG
		Elephant BHV5	NC_024696	181	A	AT	CTAG
		Elephantid BHV1	NC_020474	180	A	AT	CTAG
	Roseolovirus	Murine roseolovirus	NC_033620	174	F	AT	CTAG
		Macaca nemestrina	NC_030200	137	A	AT	CTAG
		Suid BHV2	NC_022233	128	A	AT	CTAG
	Macavirus	Alcelaphine GHV1	NC_002531	131	F	AT	TCGA
		Alcelaphine GHV2	NC_024382	137	F	AT	TCGA
		Bovine GHV6	NC_024303	145	F	AT	TCGA
		Ovine GHV2	NC_007646	135	F	AT	TCGA
	Percavirus	Felis catus GHV1	NC_028099	123	F	AT	TCGA
		Equid GHV5	NC_026421	182	B	GC	TCGA
		Equid GHV2	NC_001650	184	A	GC	TCGA
	Rhadinovirus	Ateline GHV3	NC_001987	108	F	AT	TCGA
		Cricetid GHV2	NC_015049	124	F	AT	TCGA
		Murid GHV4	NC_001826	119	F	AT	TCGA
		Saimiriine GHV2	NC_001350	113	F	AT	TCGA
Dolphin GHV1		NC_035117	167	F	AT	CTAG	
Lymphocryptovirus	Macacine GHV5	NC_003401	134	F	GC	CTAG	
	Callitrichine GHV3	NC_004367	150	F	AT	TCGA	
Unclassified gamma	Macacine GHV4	NC_006146	171	F	GC	TCGA	
	Rhinolophus GHV1	NC_040539	148	A	AT	TCGA	
Unclassified gamma	Eptesicus fuscus GHV	NC_040615	167	F	GC	CTAG	
Семейство: <i>Alloherpesviridae</i> (герпесвирус рыб и земноводных) <i>Family: Alloherpesviridae (pisces and amphibia HV)</i>							
Cyprinivirus		Anguillid HV1	NC_013668	249	A	GC	CTAG
		Cyprinid HV2	NC_019495	290	A	GC	CTAG
		Cyprinid HV3	NC_009127	295	A	GC	CTAG
		Cyprinid HV1	NC_019491	291	A	GC	TCGA
Ictalurivirus		Ictalurid HV1	NC_001493	134	F	GC	CTAG
		Ictalurid HV2	NC_036579	143	A	GC	CTAG
Batrachovirus		Ranid HV1	NC_008211	221	A	GC	CTAG
		Ranid HV2	NC_008210	232	A	GC	CTAG
Семейство: <i>Malacoherpesviridae</i> (герпесвирус беспозвоночных) <i>Family: Malacoherpesviridae (invertebrates HV)</i>							
Aurivirus		Haliotid HV1	NC_018874	212	E	AT	CTAG
Ostreavirus		Ostreid HV1	NC_005881	207	F	AT	CTAG
Неклассифицированный герпесвирус / Unclassified <i>Herpesvirales</i>							
Unclassified		Buфонid HV1	NC_040681	158	F	AT	TCGA

Примечание. Серые ячейки — ДНК класса BC[F] и «минимальные» не-CTAG TN. D* — необычная макроструктура ДНК вируса Testudinid HV3, в которой два примерно равных сегмента, ограниченные концевыми повторами, разделены короткой уникальной последовательностью.

Note. Gray cells — BC [F] class DNA and “minimal” non-CTAG TN. Asterix in D* denotes the unusual macrostructure of the Testudinid HV3 DNA, in which two approximately equal segments bounded by terminal repeats are separated by a short unique sequence.

Таблица 3 показывает, что все герпесвирусы разделены на две группы в соответствии с двумя основными «недопредставленными» TN — CTAG или TCGA. Разница между этими двумя группами следует из их геномной макроструктуры. Минимальный CTAG (CTAG_{min}) характерен для классов A, D, E структурированной ДНК с большими мономерными концевыми повторами — TR₁, TCGA_{min} характерен для менее строго структурированных классов ДНК B, C[F] с нефиксированными тандемными концевыми повторами — TR₂.

Обсуждение

«Недопредставленность» TN CTAG (CTAG_{min}) в геномах *эшерихий* и *сальмонелл*, а также некоторых фагов известна давно [15] и продолжает изучаться [16]. Мы впервые систематически показываем эту особенность для всех известных герпесвирусных ДНК и предполагаем ее возможную связь с их структурой. CTAG_{min} характерен для ДНК большей группы герпесвирусов (классы A, D, E) с наличием одного или двух сегментов, ограниченных мономерными концевыми повторами (TR₁, прямыми или взаимно инвертированными). TCGA_{min} характерен для ДНК меньшей группы герпесвирусов (классы B, C[F]) с односегментным геномом, ограниченным неопределенным числом тандемно организованных прямых концевых повторов (TR₂).

На мысль о более общей природе параллелей CTAG_{min}|ADE и TCGA_{min}|BC[F] среди вирусов животных наводит «недопредставленность» CTAG за рамками отряда *Herpesvirales*, в частности в ДНК вирусов африканской чумы свиней (*Asfarviridae*) и фибромы Shope у кроликов (*Poxviridae*), структурированных аналогично герпесвирусам A, D, E. В то же время геномы вирусов оспы и оспавакцины (*Poxviridae*) не структурированы подобным образом, в них нет «недопредставленности» CTAG. Эти наблюдения требуют серьезного расширения обсуждаемых исследований в других таксономических группах вирусов животных.

Термодинамическая модель CTAG в составе РНК показывает, что этот тетрамер нарушает оптимальную структуру стволовых петель молекулы, которые контролируют экспрессию генов, увеличивая их свободную энергию. Авторы этой гипотезы [15] предположили также, что общий предок *Salmonella* и *Escherichia* имел значительно более высокую плотность CTAG, но эволюционное вырождение привело к замене CTAG у его потомков, и эта тенденция в настоящее время сохраняется. В серии генов и в межгенных пространствах у *Escherichia* и *Salmonella* это вырождение выразилось в эволюционной замене CTAG, в первую очередь на CTGG.

В этом отношении наиболее уместно сравнить филогенетически родственные (одного рода) розеоновирусы человека HHV6 и HHV7. В ДНК

обоих вирусов — по сравнению с другими вирусами герпеса — частоты CTAG и CTGG наиболее различны. Сравнение показывает, что если в HHV7 (NC_001716) отношение частот CTAG:CTGG составляет 530:301 соответственно, то в HHV6A (NC_001664) оно даже противоположное и составляет 303:391 — при близких размерах ДНК обоих вирусов. Если наблюдение Le Tang с соавт. [16] в какой-то мере применимо и к обсуждаемым герпесвирусам, то HHV7, очевидно, ближе к эволюционному предшественнику обоих розеоновирусов, чем HHV6, в котором многие CTAG были заменены на CTGG. В то же время HHV6 приобрел способность интегрировать свой геном в геном хозяина, что, как правило, не является обязательным условием для более тесных отношений с ДНК хозяина [17], о чем свидетельствует сходство профиля TN HHV7 (но не HHV6) и ДНК человека (рис. 3), а также более высокий уровень ДНК-микргомологии вирус/хозяин в HHV7, чем в HHV6, или более низкий уровень такой микргомологии у мардивирусов с выраженными теломерными островками в концевых повторах сегментов ДНК [5, 18].

На рис. 3 дополнительно показаны некоторые особенности проанализированных профилей 4TN вирусных ДНК. В соответствии с CPR2 сходство между B1 и B2 в ДНК человека намного больше, чем в ДНК вирусов, поскольку фрагменты ДНК человека имеют здесь длину 300 Кб, а геномы HHV намного короче. В HHV7 различия между B1 и B2 достаточно характерны и могут использоваться в качестве элемента молекулярной сигнатуры ДНК этого вируса (то же относится к профилю 4TN HHV4; рис. 4). Тот факт, что GenBank представляет полные (почти полные) последовательности ДНК только трех штаммов HHV7, позволяет применять статистические методы для подтверждения представленных здесь данных с большими оговорками. По этой причине мы не использовали здесь эти методы, отметив, что сегодня это лишь похоже на факт.

На рис. 4 приведено сравнение 4TN-профиля ДНК другой пары вирусов — HHV1 и HHV4. В случае HHV1 низкое содержание CTAG позволяет вирусу вызывать острую продуктивную инфекцию и накапливаться в клетках входных ворот (в тех же фибробластах), а затем переходить в нейроны, где он будет оставаться на всю жизнь — в частности, из-за ингибирующего действия эпигенетических механизмов хозяина, одним из которых является метилирование вирусной ДНК. Концентрация динуклеотидов CpG в геноме HHV1 существенно превышает среднее значение (табл. 1).

Низкие уровни CTAG могут играть роль в обострении латентных инфекций. В случае HHV4 первичная литическая инфекция не характеризуется высоким уровнем вирусных синтезов, а после перехода в хроническую фазу она также регулируется

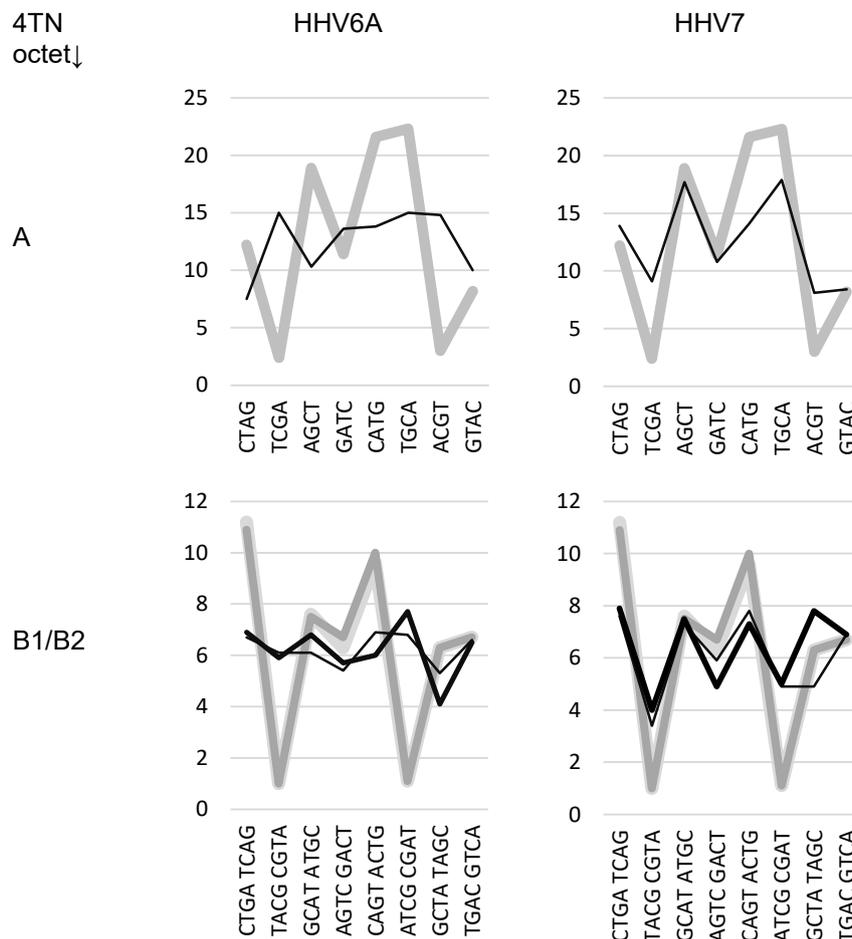


Рис. 3. 4TN-профиль ДНК HHV6A и HHV7 по сравнению с профилем 4TN ДНК человека.

Октет А: ДНК человека выделена серым, вирусная ДНК — черным. Октет В: график В1 ДНК человека выделен светло-серым, В2 — темно-серым; график В1 ДНК вируса выделен жирным черным, В2 — тонким черным.

Fig. 3. 4TN profile of HHV6A and HHV7 DNA compared to human DNA 4TN profile.

Octet A: human DNA is highlighted in gray, viral DNA — in black.

Octet B: human B1 is highlighted in light gray, human B2 is highlighted in dark gray, virus B1 is highlighted in bold black, B2 — in thin black.

эпигенетическими инструментами, включая метилирование цитозина в составе CpG [18, 19]. В то же время очевидная близость 4TN-профилей генома HHV4 и хозяина указывает на сходную реакцию на эти регулирующие инструменты. То же самое можно сказать о HHV8 и эпигенетической регуляции его генов [19, 20]. Из многих эпигенетических механизмов, которые модифицируют экспрессию генов вируса и хозяина, мы рассматриваем здесь только метилирование ДНК, точнее, метилирование цитозина в CpG, поскольку этот димер является частью тетрамера TCGA, что позволяет сравнивать его с другим тетрамером, CTAG, в предложенном здесь аспекте.

Гипотеза о низкой плотности тетрамера CTAG из-за его эволюционного вырождения не объясняет очевидных ограничений его использования и вовсе не касается причин низкой плотности другого тетрамера октета А, TCGA, в ДНК представителей

этого же суперсемейства. Около 40% CpG, центральной пары этого тетрамера, находится в промоторных зонах млекопитающих [21, 22] и имеет гораздо более низкую плотность в полных последовательностях геномов позвоночных, чем можно было бы ожидать [23, 24]. Эта «недопредставленность» является следствием высокой частоты мутаций метилированных сайтов CpG в геномах хозяев и их вирусов, особенно тех, которые тесно взаимодействуют с ДНК хозяина.

Причины пониженного содержания CpG неоднократно обсуждались и прежде [25], однако вопрос заключается не столько в низкой плотности CpG, сколько в контексте этой пары, т.е. в составе TCGA, поскольку этот тетрамер представлен в герпесвирусных ДНК в значительно меньшей концентрации, нежели ACGT.

Данные Le Tang и соавт. [16] показывают, что само по себе минимальное содержание CTAG (и TCGA) не ограничивается герпесвирусной ДНК.

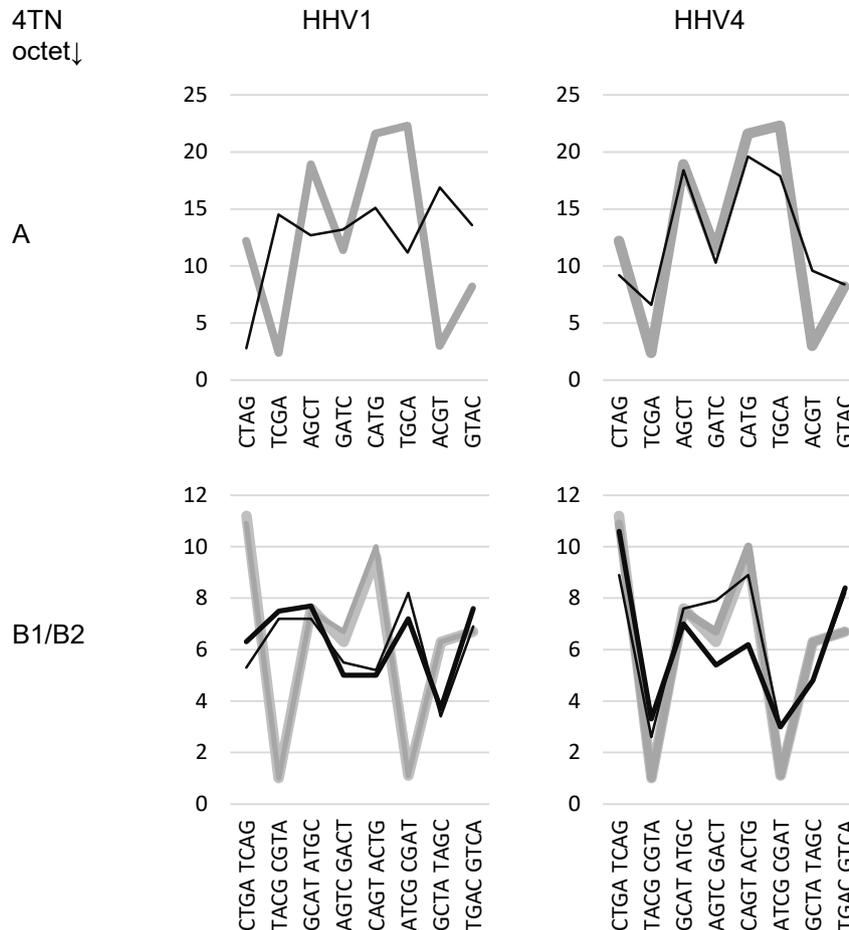


Рис. 4. 4TN-профиль HHV1 и HHV4 ДНК по сравнению с профилем 4TN ДНК человека.

Октет А: ДНК человека выделена серым, вирусная ДНК — черным. Октет В: график В1 ДНК человека выделен светло-серым, В2 — темно-серым; график В1 ДНК вируса выделен жирным черным, В2 — тонким черным.

Fig. 4. 4TN profile of HHV1 and HHV4 DNA compared to human DNA 4TN profile.

Octet A: human DNA is highlighted in gray, viral DNA — in black.

Octet B: human B1 is highlighted in light gray, human B2 is highlighted in dark gray, virus B1 is highlighted in bold black, B2 — in thin black.

Мы проанализировали TN-профиль больших ДНК с концевыми повторами некоторых других вирусов. Обнаружено, что СТАГ является «минимальным» у вирусов африканской чумы свиней (семейство *Asfarviridae*) и вируса фибромы Shope (семейство *Poxviridae*), но не у вирусов оспы и осповакцины (также семейства *Poxviridae*), ДНК которых не имеет терминальных повторов. Это означает, что при построении филогенетических деревьев необходимо учитывать не только изменения в генах и белках, но и эволюцию молекулы ДНК, включая ее обсуждаемые здесь характеристики.

В первом приближении для анализа плотности потенциально метилируемого цитозина в геномах герпесвирусов достаточно оценить соотношение CpG:GpC (т.е. проанализировать динуклеотидный профиль ДНК), которое не связано с типом генома (АТ или GC). Эту оценку можно проследить в табл. 2: ДНК HHV с CpG>GpC (в серых ячейках). В этом случае результаты, представленные здесь,

будут касаться только концентрации и соотношения СТАГ/CpG в герпесвирусных ДНК, которые могут влиять на уровень вирусных синтезов. В наиболее общем (не строгом) виде это соотношение имеет зеркальный характер: наиболее низкая концентрация СТАГ сопровождается самой высокой концентрацией CpG (табл. 1). Тем не менее соотношение СТАГ/CpG обедняет получаемую информацию, которая указывает на различие именно в TN-профиле герпесвирусных ДНК, СТАГ/TCGA. Другими словами, составляющей этого отношения является TCGA/ACGT, четко выраженное в рамках классов DE/A/BC[F] (табл. 1). В свою очередь, это указывает на необходимость учитывать контекст, который определяет функциональную ценность димера CpG. Возможно, этот контекст выходит за рамки тетрамера. Для надежных выводов необходимо расширить исследования за рамки герпесвирусов — при серьезном пополнении GenBank новыми полными последовательностями вирусных ДНК. Но в любом

случае результаты, продемонстрированные здесь, указывают на то, что биологический смысл макро-структуры герпесвирусной ДНК гораздо глубже, чем принято считать.

ЛИТЕРАТУРА / REFERENCES

- Whitley R., Kimberlin D., Prober C. Pathogenesis and disease. In: Arvin A., Campadelli-Fiume G., Mocarsky E., Moore P.S., Roizman B., Whitley R., eds. *Human Herpesviruses: Biology, Therapy and Immunoprophylaxis. Chapter 32*. Cambridge: Cambridge University Press; 2007.
- Pellet P., Roizman B. Herpesviridae. In: Knipe D.M., Howley P.M., eds. *Fields Virology*. Philadelphia: Lippincott Williams & Wilkins; 2013: 1802-2.
- Zabolotneva A., Tkachev V., Filatov F., Buzdin A. How many antiviral small interfering RNAs may be encoded by the mammalian genomes? *Biol. Direct*. 2010; 5: 62. DOI: <http://doi.org/10.1186/1745-6150-5-62>
- Filatov F., Shargunov A. Short nucleotide sequences in herpesviral genomes identical to the human DNA. *J. Theor. Biol.* 2015; 372: 12-21. DOI: <http://doi.org/10.1016/j.jtbi.2015.02.019>
- Filatov F., Shargunov A. Microhomology of Viral/Host DNAs and macrostructure of herpesviral genome. *Int. J. Virol. AIDS*. 2018; 5(1): 042. DOI: <http://doi.org/10.23937/2469-567X/1510042>
- 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. DOI: <http://doi.org/10.1073/pnas.60.3.921>
- 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. DOI: <http://doi.org/10.1093/bioinformatics/18.1.215>
- Albrecht-Buehler G. Asymptotically increasing compliance of genomes with Chargaff's second parity rules through inversions and inverted transpositions. Version 2. *Proc. Natl. Acad. Sci. USA*. 2006; 103(47): 17828-33. DOI: <http://doi.org/10.1073/pnas.0605553103>
- Baisnee P.F., Hampson S., Baldi P. Why are complementary strands symmetric? *Bioinformatics*. 2002; 18(8): 1021-33. DOI: <http://doi.org/10.1093/bioinformatics/18.8.1021>
- Gori F., Mavroeidis D., Jetten M.S.M., Marchiori E. The importance of Chargaff's second parity rule for genomic signatures in metagenomics. Available at: <http://www.biorxiv.org/content/biorxiv/early/2017/06/04/146001.full.pdf>
- Pride D.T., Blaser M.J. Identification of horizontally acquired genetic elements in *Helicobacter pylori* and other prokaryotes using oligonucleotide difference analysis. *Genome Lett.* 2002; 1(1): 2-15. DOI: <http://doi.org/doi.org/10.1166/gl.2002.003>
- Prabhu V.V. Symmetry observations in long nucleotide sequences. *Nucleic Acids Res.* 1993; 21(12): 2797-800. DOI: <http://doi.org/10.1093/nar/21.12.2797>
- Albrecht-Buehler G. The three classes of triplet profiles of natural genomes. *Genomics*. 2007; 89(5): 596-601. DOI: <http://doi.org/10.1016/j.ygeno.2006.12.009>
- Zhang S.H., Wang L. A novel common triplet profile for GC-rich prokaryotic genomes. *Genomics*. 2011; 97(5): 330-1. DOI: <http://doi.org/10.1016/j.ygeno.2011.02.005>
- Burge C., Campbell A.M., Karlin S. Over- and under-representation of short oligonucleotides in DNA sequences. *Proc. Natl. Acad. Sci. USA*. 1992; 89(4): 1358-62. DOI: <http://doi.org/10.1073/pnas.89.4.1358>
- 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. DOI: <http://doi.org/10.1038/srep43565>
- Bhende P.M., Seaman W.T., Delecluse H.J., Kenney S.C. The EBV lytic switch protein, Z, preferentially binds to and activates the methylated viral genome. *Nat. Genet.* 2004; 36(10): 1099-104. DOI: <http://doi.org/10.1038/ng1424>
- Kaufer B.B., Flamand L. Chromosomally integrated HHV-6: impact on virus, cell and organismal biology. *Curr. Opin. Virol.* 2014; 9: 111-8. DOI: <http://doi.org/10.1016/j.coviro.2014.09.010>
- 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. DOI: <http://doi.org/10.1016/j.coviro.2013.03.005>
- Lim C., Lee D., Seo T., Choi C., Choe J. Latency associated nuclear antigen of Kaposi's sarcoma-associated herpesvirus functionally interacts with heterochromatin protein 1. *J. Biol. Chem.* 2003; 278(9): 7397-405. DOI: <http://doi.org/10.1074/jbc.M211912200>
- Pantry S.N., Medveczky P.G. Epigenetic regulation of Kaposi's sarcoma associated herpesvirus replication. *Semin. Cancer Biol.* 2009; 19(3): 153-7. DOI: <http://doi.org/10.1016/j.semcancer.2009.02.010>
- 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. DOI: <http://doi.org/10.1093/nar/gni180>
- Lander E.S., Linton L.M., Birren B., Nusbaum C., Zody M.C., Baldwin J., et al. Initial sequencing and analysis of the human genome. *Nature*. 2001; 409(6822): 860-921. DOI: <http://doi.org/10.1038/35057062>
- 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. DOI: <http://doi.org/10.1101/gr.152231.112>
- Nicholas J. Evolutionary aspects of oncogenic herpesviruses. *Mol. Pathol.* 2000; 53(5): 222-37. DOI: <http://doi.org/10.1136/mp.53.5.222>

Информация об авторах:

Филатов Феликс Петрович[✉] — к.м.н., д.б.н., в.н.с. лаб. молекулярной биотехнологии ФГБНУ НИИВС им. И.И. Мечникова, 105064, Москва, Россия; ведущий научный сотрудник ФГБУ «НИЦ эпидемиологии и микробиологии им. почетного академика Н.Ф. Гамалеи», 123098, Москва, Россия. ORCID ID: <https://orcid.org/0000-0001-6182-2241>. E-mail: felix001@gmail.com

Шаргунов Александр Валерьевич — ведущий инженер, лаб. генетики ДНК-содержащих вирусов ФГБНУ НИИВС им. И.И. Мечникова, 105064, Москва, Россия. ORCID ID: <https://orcid.org/0000-0001-5536-1557>.

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

Information about the authors:

Felix P. Filatov[✉] — PhD (Med.), D. Sci. (Biol.), leading researcher, Laboratory of molecular biotechnology, Mechnikov Federal Research Institute of Vaccines and Sera, 105064, Moscow, Russia; leading researcher, National Research Centre for Epidemiology and Microbiology named after the honorary academician N.F. Gamaleya, 123098, Moscow, Russia. ORCID ID: <https://orcid.org/0000-0001-6182-2241>. E-mail: felix001@gmail.com

Alexander V. Shargunov — leading engineer, Laboratory of DNA-containing viruses genetics, Mechnikov Federal Research Institute of Vaccines and Sera, 105064, Moscow, Russia. ORCID ID: <https://orcid.org/0000-0001-5536-1557>.

Contribution: the authors contributed equally to this article.