

*В. В. Давыдов, С. В. Жаворонок*

## ИСПОЛЬЗОВАНИЕ СИНОНИМИЧНЫХ КОДОНОВ В ГЕНОМЕ ВИРУСА ГЕПАТИТА E

*УО «Белорусский государственный медицинский университет»*

*Вирус гепатита E (ВГЕ) является малоизученным патогеном, имеющий широкое распространение среди людей и животных. В геномах всех организмов существует предвзятость в использовании синонимичных кодонов. Мутации, генетический дрейф и естественный отбор являются основными факторами, определяющим смещение использования кодонов в геномах различных вирусов. Целью настоящего исследования явилось изучение систематической ошибки в использовании синонимичных кодонов в последовательностях РНК ВГЕ 1-го и 3-го генотипов, выделенных из организма разных хозяев. Исследованы 309 полных геномов ВГЕ, депонированных в базе данных GenBank NCBI (in silico). В последовательностях РНК открытых рамок считывания ORC1, ORC2 и ORC3 изучен нуклеотидный состав, состав динуклеотидов, относительное использование синонимичных кодонов, проведен ENC-Plot анализ, анализ правила парности, рассчитано dN/dS-отношение. Установлено: GC состав последовательностей ВГЕ составляет  $59,3 \pm 0,52$  %; в ORC1 и ORC2 в третьем положении кодона более предпочтительными являются 3T(U) или 3C, а в ORC3 – 3C или 3G; в ORC1 и ORC2 динуклеотиды CpG и TpC недопредставлены, а TpG перепредставлены; в синонимичных кодонах ORC2 выявлено U/A концевое смещение, в ORC1 и ORC2 – 3G или 3C; влияние трансляционного отбора на ВГЕ-1 более выражено, чем ВГЕ-3; предпочтения в использовании синонимичных кодонов в ORC3 не зависят от действия трансляционного отбора; ведущим фактором эволюции ORC1 и ORC2 ВГЕ является отрицательный (очищающий) отбор; в ORC3 ВГЕ-1 преобладают нейтрально отобранные сайты, что является следствием устоявшихся взаимоотношений между вирусом и хозяином.*

**Ключевые слова:** *Вирус гепатита E, использование синонимичных кодонов, естественный отбор.*

*V. V. Davydov, S. V. Zhavoronok*

## SYNONYMIC CODON'S USAGE IN THE HEPATITIS E VIRUS GENOME

*Hepatitis E virus (HEV) is a poorly understood pathogen that is widespread in humans and animals. In the genomes of all organisms, there is a bias in the use of synonymous codons. Mutations, genetic drift and natural selection are the main factors that determine codon usage bias in the genomes of viruses. The aim of this study was to study the systematic error in the use of synonymous codons in whole genome sequences of HEV RNA genotypes 1 and 3 isolated from different hosts. We studied 309 complete HEV genomes deposited in the GenBank NCBI database (in silico). In the sequences of open reading frames ORF1, ORF2, and ORF3, the nucleotide composition, the composition of dinucleotides, the relative use of synonymous codons were studied, ENC-Plot analysis and parity rule analysis was performed, the dN/dS ratio was calculated. Found: GC composition of HEV sequences is  $59.3 \pm 0.52$  %; in ORF1 and ORF2 in the third codon position, 3T(U) or 3C is more preferred, and in ORF3 – 3C or 3G; in ORF1 and ORF2, CpG and TpC dinucleotides are underrepresented, while TpG are overrepresented; in the synonymous codons of ORF2, a U/A terminal shift was detected, in ORF1 and ORF2 – 3G or 3C; the effect of translational*

*selection on HEV-1 is more pronounced than HEV-3; preference for the use of synonymous codons in ORF3 does not depend on the action of translational selection; the leading factor in the evolution of ORF1 and ORF2 HEV is negative (purifying) selection; neutrally selected sites predominate in ORF3 of HEV-1, which is a consequence of the established relationship between the virus and the host.*

**Key words:** *Hepatitis E virus, use of synonymous codons, natural selection.*

Аминокислоты являются мономерами белковых молекул. В стандартном генетическом коде набор из 61 кодона кодирует 20 аминокислот. За исключением триптофана и метионина, все аминокислоты представлены более чем одним кодоном (от двух до шести), что является еще одним свойством генетического кода – избыточностью (вырожденностью), а кодоны, имеющие один генетический смысл, называются синонимичными. Для кодирования одной аминокислоты в теоретической модели все синонимичные кодоны должны использоваться равновероятно, однако это не так. Смещение или предвзятость использования кодонов (от англ. «Codon Usage Bias, CUB») – это явление, при котором в геноме организма или в различных генах одного организма определенные кодоны более предпочтительно используются в кодирующих областях, вместо их синонимичных альтернатив. В настоящее время является общепризнанным представление о том, что CUB может быть обусловлено различными факторами, а для объяснения возникновения CUB были предложены различные гипотезы.

В теории нейтральной эволюции (Kimura, 1968) мутационное давление в вырожденных положениях кодона приводит, к неравномерному использованию синонимичных кодонов для конкретной аминокислоты, возникающему при отсутствии естественного отбора. Также было показано, что уровень экспрессии генов связан с CUB и гены, имеющие высокую экспрессию, имеют более выраженное смещение. Другими важными детерминантами CUB являются нуклеотидный состав последовательностей (GC-насыщенность), длина гена, отбор, на-

правленный на эффективность трансляции. Определенные кодоны могут улучшать точность трансляции и снижать вероятность ошибок, тогда как использование других кодонов, наоборот, может ее ускорять. Существует предположение, что использование редких кодонов замедляет скорость трансляции, чтобы обеспечить оптимальные уровни экспрессии белка и его правильную укладку.

Мутации и генетический дрейф являются основными факторами, определяющим конфигурацию паттернов использования кодонов в геномах различных вирусов. Ма и др. предположили, что трансляционный отбор и мутации являются основными эволюционными силами, управляющими образованием CUB у вируса гепатита В [1]. Исследования, проведенные в Испании в 2018 г., продемонстрировали значительное смещение использования кодонов вируса гепатита А (ВГА), которое оказалось деоптимизированным по сравнению с организмом хозяина [2].

Геном вируса гепатита Е (ВГЕ) имеет размер около 7,2 тыс. п. н. и состоит из 5'-UTR, кэпированного 7-метилгуанозином, за которым следуют три открытые рамки считывания (ОРС1, ОРС2 и ОРС3). 3'-UTR заканчивается поли(А)-хвостом, как эукариотическая структура мРНК. Репликация вируса начинается с трансляции неструктурного полипротеина, кодируемого ОРС1. РНК-зависимая РНК-полимераза впоследствии транскрибирует полноразмерную антисмысловую РНК. Эта РНК служит матрицей для синтеза двух вирусных РНК с положительным смыслом в инфицированных клетках, одну полноразмерную геномную РНК и другую субгеномную РНК, содержащую ОРС2, ко-

дирующую капсидный белок, и OPC3, кодирующую полифункциональный белок.

Исследование CUB в геноме вируса гепатита E (ВГЕ) представляет большой интерес с точки зрения его эволюции. В отличие от других вирусов, вызывающих острый гепатит у человека, ВГЕ является зооантропооонозом и может реплицироваться в организме как человека, так и других млекопитающих. Анализ опубликованных данных по тематике CUB в последовательностях ВГЕ свидетельствует о недостаточной глубине проведенных исследований и остается мало изученной.

**Целью настоящего исследования** явилось изучение предвзятости использования синонимичных кодонов в последовательностях OPC1, OPC2 и OPC3 РНК ВГЕ-1 и ВГЕ-3, выделенных из организма разных хозяев.

### Материалы и методы

В этом исследовании мы провели актуальный всесторонний анализ особенностей состава и использования кодонов полных геномов ВГЕ-1 и ВГЕ-3, о которых сообщалось в период с 1982 по 2022 год. Полные кодирующие последовательности (CDS) OPC1, OPC2 и OPC3 геномов ВГЕ 1-го ( $n = 61$ ) и 3-го генотипов, выделенных из организма человека ( $n = 148$ ) и животных ( $n = 100$ ) были извлечены из базы данных GenBank в Национальном центре биотехнологической информации (NCBI) (<http://www.ncbi.nlm.nih.gov>). Все последовательности были выровнены с помощью программы MEGA 11.0 с использованием алгоритма CLUSTAL W. Всего в этом исследовании было проанализировано 309 последовательностей генома ВГЕ.

Анализ использования кодонов выполняли с помощью программы CodonW, интегрированной в on-line сервис Galaxy Pasteur, доступной по адресу <https://galaxy.pasteur.fr>. Был рассчитан нуклеотидный состав РНК ВГЕ, кодирующих OPC1, OPC2 и OPC3. Затем были оценены частоты нуклеотидов

в первом, втором и третьем положении кодонов. Была рассчитана общая GC насыщенность и частоты нуклеотидов G+C в первом, втором и третьем положениях кодона и RSCU.

Synonymous Codon Usage», RSCU – это показатель, который используется для измерения степени смещения использования синонимичных кодонов в генах. Для расчёта RSCU для каждого кодона определяли отношение его фактической частоты использования в каждой OPC к ожидаемой. Значения RSCU более 1 указывают на предпочтительное использование данного кодона, в то время как значения менее 1 указывают на меньшее предпочтение. Значение  $RSCU > 1,6$  указывает на то, что данный кодон является чрезмерно представленным (перепредставленным), в то время как значение  $RSCU < 0,6$  указывает на недопредставленность кодона.

Анализ относительного содержания динуклеотидов выполняли при помощи специализированного on-line сервиса «EMBOSS Explorer» доступного в сети Интернет по адресу <https://www.bioinformatics.nl/cgi-bin/emboss/compseq>. Отношение фактической частоты динуклеотидов к прогнозируемой представляет собой отношение шансов. Отношение шансов ниже 0,78 указывает на недопредставленные пары динуклеотидов, тогда как отношение шансов выше 1,25 указывает на чрезмерно представленные динуклеотиды.

Для оценки разнообразия использования кодонов в генах использовали графический метод анализа использования кодонов (от англ. «Effective Number of Codons Plot Analysis», ENC-Plot Analysis). В ENC-Plot анализе, значения ENC для последовательностей каждой OPC отображали на графике, где по оси X – содержание GC в третьей позиции синонимичных кодонов (% 3GC), а по оси Y – ENC. На графике для сравнения также строили стандартную кривую ожидаемых значений ENC (на основе 3GC). Если точки на графике ENC-Plot сгруппи-

рованы вдоль стандартной кривой ожидаемых значений ENC, это указывает на то, что мутационное давление является главной причиной неравномерности использования кодонов. Если точки отклоняются вниз от стандартной кривой ожидаемых значений ENC, это указывает на существование дополнительных факторов, которые могут быть причиной неравномерности использования кодонов, например, как трансляционный отбор.

Анализ правила парности (от англ. «Parity Rule 2», PR2) использовали для оценки влияния мутации и отбора при использовании кодонов. На графике PR2 смещение GC [ $3G/(3G + 3C)$ ] в третьем положении кодона и смещение AU [ $3A/(3A + 3U)$ ] в третьем положении кодона представляют собой абсциссу и ординату соответственно. Центр графика (0,5; 0,5) соответствует точке, где использование AU- и GC-богатых кодонов в геноме равномерно, т. е.  $A = U$  и  $G = C$ , что указывает на отсутствие отклонения между мутационным давлением и естественным отбором. Если значения на графике смещены в сторону большей частоты использования AU-богатых кодонов, это может указывать на доминирование в то время как смещение в сторону большей частоты использования GC-богатых кодонов может указывать на доминирование естественного отбора.

Отношение между числом несинонимичных замен ( $dN$ ) и числом синонимичных замен ( $dS$ )  $dN/dS$ -отношение, является мерой давления отбора на гены в белок кодирующей последовательности. Значения  $dN$  и  $dS$  были рассчитаны для последовательностей ВГЕ-1 и ВГЕ-3 в трех рамках считывания ORF1, ORF2 и ORF3. Для каждого кодона во всех последовательностях было рассчитано количество предполагаемых синонимичных ( $s$ ) и несинонимичных ( $n$ ) замен, а также количество сайтов, которые оцениваются как синонимичные ( $S$ ) и несинонимичные ( $N$ ). Эти оценки были получены с использованием совместных рекон-

струкций максимального правдоподобия предковых состояний в соответствии с моделью Muse-Gaut замены кодонов и моделью Felsenstein замены нуклеотидов. Для обнаружения кодонов, подвергшихся положительному, отрицательному и нейтральному отбору вычисляли разность  $dN - dS$  (где  $dS$  – количество синонимичных замен на сайт ( $s/S$ ), а  $dN$  – количество несинонимичных замен на сайт ( $n/N$ )). Расчеты максимального правдоподобия  $dN$  и  $dS$  проводились с использованием программного пакета HyPhy, интегрированного в MEGA7.

### Результаты и обсуждение

Отношения в системе «паразит-хозяин» являются сложным многоуровневым процессом, на который может оказывать влияние множество взаимодействующих факторов. Вирусы являются облигатными паразитами и их процесс эволюции неразрывно связан со своим хозяином. Конкретный фрагмент генома или небольшой участок гена вируса может быть вовлечен в специфичные для хозяина адаптации и приводить к различным последствиям, от минимальных, позволяющих уклоняться от иммунной защиты хозяина, вплоть до гостального прыжка, приводящего к расширению круга хозяев для вируса.

Самой сильной детерминантой различий в нуклеотидной композиции у разных видов является содержание GC в геноме. Большинство вирусов с положительной цепью РНК характеризуются высокой частотой нуклеотида А в нуклеотидной композиции их геномов [3]. Проведенные нами исследования показали, что GC насыщенность генома изученных последовательностей ВГЕ в среднем составляла  $59,3 \pm 0,52$  % (рисунок 1), что согласуется с данными, полученными В. Li и др., 2022 [4]. Однако, содержание нуклеотидов GC в различных положениях кодона, установленное в нашем исследовании, значительно отличается от данных, опубликованных в работе китайских исследователей.

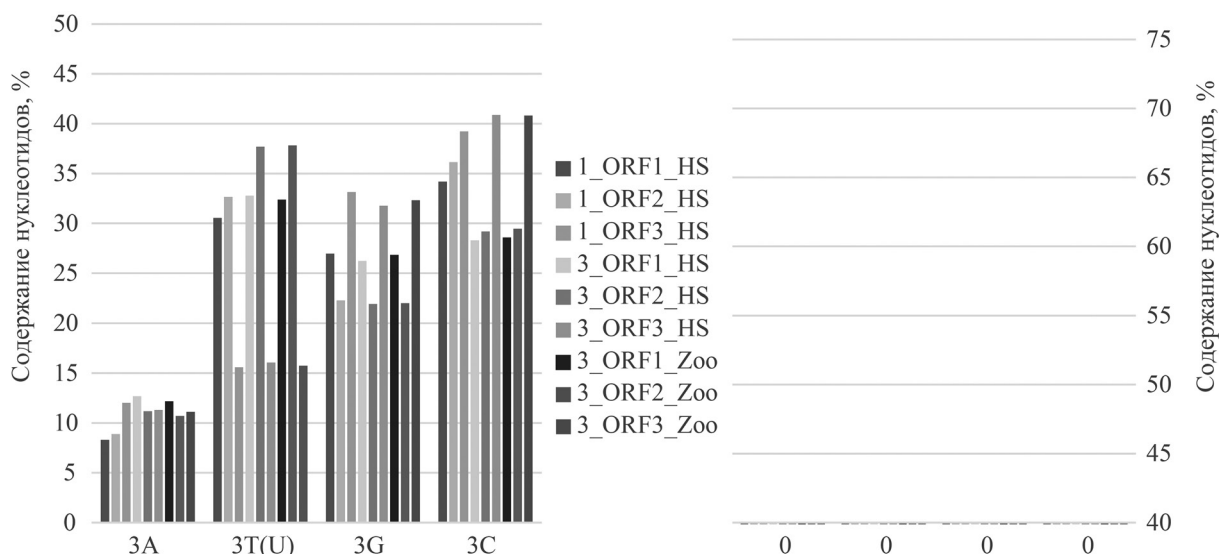


Рисунок 1. Нуклеотидная композиция последовательностей ORF1, ORF2 и ORF3 1-го и 3-го генотипов генома ВГЕ, выделенных из организма разных хозяев

В ORF1 и ORF2 ВГЕ в третьем положении кодонов чаще используются нуклеотиды T(U) либо C, чем A и G, поэтому содержание 3GC в этих рамках составляющее  $54,7 \pm 0,48 \%$  и  $55,9 \pm 0,49 \%$  соответственно, значительно ниже, чем в ORF3, где оно составляет в среднем  $67,5 \pm 0,58 \%$ . Теоретически, когда на использование синонимичных кодонов влияет только мутационное давление, частоты нуклеотидов U и A в третьем положении кодона должны быть равны частотам нуклеотидов G и C. Полученные нами результаты не соответствуют этой парадигме и указывают на значительную роль естественного отбора в формировании нуклеотидного состава ВГЕ. Установленное GC смещение в нуклеотидной композиции генома ВГЕ может быть результатом адаптации общего предка современных штаммов ВГЕ к требованиям хозяина в отношении состава нуклеотидов, как это сообщалось ранее [5]. Композиционные свойства разных генов ВГЕ находятся под влиянием разных направлений естественного отбора и мутационного давления.

Использование кодонов обусловлено разными факторами, такими, как содержание GC, преобладание определённых динуклеотидов в геноме вируса, профила-

ми тРНК хозяина, уровнями экспрессии белков и иммунными стратегиями хозяина. Избегание динуклеотидов CpG и UpA является основным фактором предпочтения динуклеотидов в РНК-вирусах. Клетки хозяина имеют множество механизмов для нацеливания на чужеродные нуклеиновые кислоты, опосредованное противовирусным белком цинковых пальцев (ZAP). У хозяев ВГЕ также имеется система врожденного иммунитета (например, пути Toll и JAK/STAT). Они активно используют интерферон в качестве эффектора, стимулируемого через рецепторы распознавания образов (PRR) с помощью патоген-ассоциированных молекулярных структур (PAMP).

Исследовательская группа из Австралии установила, что в большинстве семейств РНК-содержащих вирусов динуклеотиды CpG и TpA являются недопредставленными, а CpA и TpG наоборот, представленными избыточно. В то время как в семействе *Hepeviridae* динуклеотиды UpA или CpG недопредставленными не являются [6]. Полученные нами результаты лишь отчасти подтверждают результаты этих исследований (рисунок 2).

Среднее значение ОШ динуклеотида CpG во всех изученных последовательностях ВГЕ составляло 0,843, что указывает на то, что его содержание соответствует CG составу

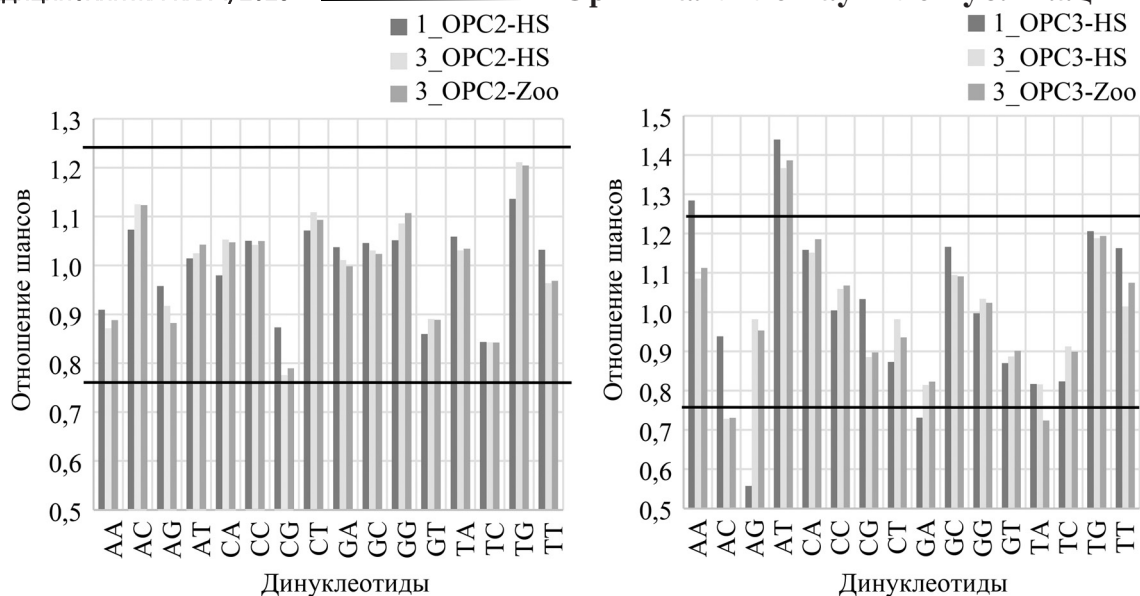


Рисунок 2. Отношение шансов динуклеотидов в последовательностях OPC2 и OPC3 ВГЕ-1 и ВГЕ-3, выделенных из организма человека и животных. Горизонтальные линии обозначают границы значимости содержания динуклеотидов

гена и эта нуклеотидная пара не является недопредставленной. В последовательностях OPC1 и OPC2 ВГЕ динуклеотид CpG действительно был недопредставлен ( $OШ < 0,78$ ) или близок к этому, в то время как в OPC3 его содержание было более высоким ( $OШ \approx 1,0$ ).  $OШ$  динуклеотида TpA наоборот, в последовательностях OPC1 и OPC2 имело значение близкое к 1, а в OPC3 этот динуклеотид был недопредставленным.  $OШ$  нуклеотидной пары TrG в геноме ВГЕ установлено на уровне более 1,2, что является близким к тому, чтобы охарактеризовать ее как перепредставленную. Для динуклеотида CpA  $OШ$  установлено лишь на уровне 1,1–1,15, что не может свидетельствовать о ее избыточной представленности.

Состав динуклеотидов обуславливает определенные предпочтения в использовании синонимичных кодонов, которые зависят от различных факторов, включая естественный отбор, давление мутаций, нуклеотидный состав области генома и длину генов.

Анализ качественного состава избыточно представленных синонимичных кодонов последовательностей ВГЕ разных генотипов (хозяев) выявил предпочтение использования кодонов, содержащих G/C нуклео-

тиды в третьем положении кодона, которое было более выраженным в последовательностях ВГЕ-1, чем ВГЕ-3. В результате анализа систематической ошибки использования кодонов в последовательностях разных OPC обнаружена более выраженная предвзятость использования кодонов в последовательностях OPC3. При этом, за исключением единственного кодона UAU, все перепредставленные кодоны в этой OPC содержали нуклеотид G/C в третьем положении. Последовательности OPC1 ВГЕ имели схожую, но менее выраженную тенденцию. Все перепредставленные кодоны ВГЕ-1 также имели предпочтение в G/C завершении синонимичных кодонов. Последовательности OPC2 характеризовались иной моделью использования кодонов. Несмотря на высокое содержание цитозина в третьем положении, в последовательностях ВГЕ OPC2 мы не выявили предпочтения в 3G/C завершении ее кодонов. Оно было характерно только для перепредставленных кодонов в последовательностях ВГЕ-1 (8/11). Перепредставленные кодоны OPC2 ВГЕ-3 (6/7) имели U-завершение (таблица 1).

Проведенный анализ значений RSCU выявил отличающиеся у ВГЕ-1 и ВГЕ-3 модели использования синонимичных кодонов,

Таблица 1. Количество перепредставленных и недопредставленных кодонов в полных последовательностях ОРС ВГЕ разных генотипов (хозяев)

ОРС	ОРС1						ОРС2						ОРС3					
	ВГЕ-1_HS		ВГЕ-3_HS		ВГЕ-3_Zoo		ВГЕ-1_HS		ВГЕ-3_HS		ВГЕ-3_Zoo		ВГЕ-1_HS		ВГЕ-3_HS		ВГЕ-3_Zoo	
Генотип (хозяин)																		
Тип кодона																		
Нуклеотид в 3-м положении кодона	3G(C)	3A(U)	3G(C)	3A(U)	3G(C)	3A(U)	3G(C)	3A(U)	3G(C)	3A(U)	3G(C)	3A(U)	3G(C)	3A(U)	3G(C)	3A(U)	3G(C)	3A(U)
Перепредставленные кодоны (RSCU > 1,6)	9		5		4		12		9		7		12		11		11	
Недопредставленные кодоны (RSCU < 0,6)	9	0	4	1	3	1	7	5	4	5	2	5	12	0	11	0	10	1
	14		11		10		17		15		18		21		22		23	
	11	3	9	3	9	1	4	13	4	11	5	13	18	3	3	19	4	19

которые свидетельствуют о независимой эволюции этих генотипов ВГЕ. Выявлено 3G/C смещение в ОРС1 и особенно в ОРС3, а также специфичное для ОРС2 U/A-концевое смещение использования синонимичных кодонов, которое свидетельствует о преобладающей роли в эволюции ОРС2 естественного отбора, а не мутаций.

Для установления силы влияния трансляционного отбора на последовательности ВГЕ был проведен ENC-Plot анализ и анализ правила парности (рисунок 3).

Результаты исследований, полученные разными методами в целом, согласуются и соответствуют ранее опубликованным данным исследователей из Китая (S. Вага и др.). Полученные результаты позволяют резюмировать, что влияние трансляцион-

ного отбора на смещение использования синонимичных кодонов более выражено в последовательностях ВГЕ-1. Смещение CUB в последовательностях ОРС3 значительно меньше зависит от действия трансляционного отбора и не дифференцировано в зависимости от генотипа вируса. CUB в последовательностях ВГЕ-3 в меньшей степени, чем ВГЕ-1 находится под давлением селективного влияния, причем это влияние меньше в последовательностях ОРС1. Менее значительное смещение ENC в ОРС1, также характерное для вируса гепатита С и вируса Зика [7], говорит о снижении конкуренции между вирусом и хозяином, улучшает выживаемость и эффективную репликацию в среде хозяина, снижает энергию, необходимую для биосинтеза вируса,

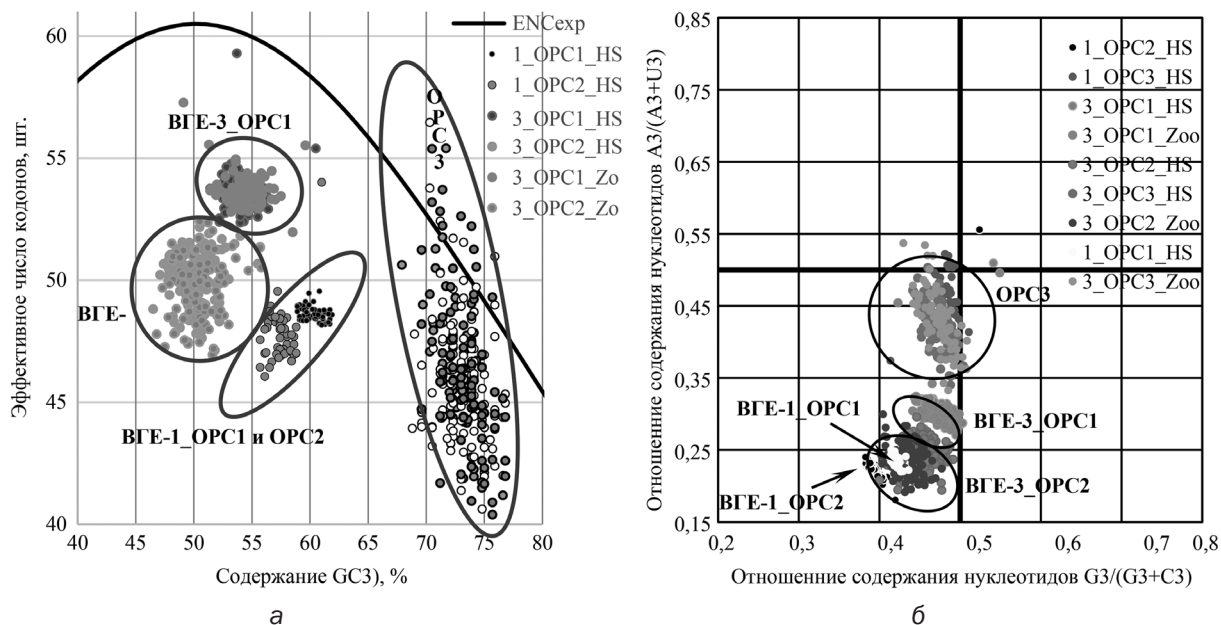


Рисунок 3. Анализ влияния естественного отбора на последовательности ОРС ВГЕ-1 и ВГЕ-3, выделенных из организма человека и животных: а – ENC-Plot анализ, б – анализ правила парности

и позволяет избежать конкуренции с синтезом белка хозяина, как это было показано F. van Hemert и др. Это может быть критически важным для генов, имеющих высокий уровень экспрессии в самом начале развития инфекции, что согласуется с данными о медленной репликации вирусной РНК ВГЕ [8]. Кроме того, нам не удалось выявить какую-либо зависимость смещения использования синонимичных кодонов под действием трансляционного отбора в последовательностях ВГЕ-3, выделенных из разных хозяев.

Анализ влияния отбора на белок кодирующие последовательности ОРС ВГЕ на основе расчета разности  $dN-dS$ , в целом, позволил подтвердить результаты, полученные исследователями из США [9]. Поскольку ВГЕ-1 и ВГЕ-3 адаптированы к передаче в организме разных хозяев, а эффективное заражение ВГЕ-3 большого круга видов-хозяев должно требовать вовлечения многих участков генома ВГЕ для адаптации к различным внутриклеточным средам, логично предположить, что ОРС ВГЕ-3 должны быть в большей степени подвержены изменчивости. Последовательности ОРС1 и ОРС2 ВГЕ-3 преимущественно испытывают действие очищающего отбора, с незначительным числом положительно отобранных сайтов в ОРС1 ВГЕ-1. В последовательностях ОРС3 ВГЕ выявлены более интенсивные эволюционные процессы, прежде всего

проявляющиеся во влиянии позитивного отбора на данный участок генома вируса, приводящие к высокой вариабельности аминокислотного состава продукта ОРС3. Действие движущего отбора на ОРС3 ВГЕ-3 более сильное, приводящее к более значительному аминокислотному полиморфизму. Количество позитивно отобранных сайтов в последовательностях ОРС3 ВГЕ-3, имеющих антропонозное происхождение на 5 % больше, чем в зоонозных, что говорит о более напряженном процессе адаптации этого генотипа к существованию в организме человека, чем в организме животных.

Использование синонимичных кодонов для построения ОРС не является случайным, существуют механистические или эволюционные ограничения, которые определяют степень свободы для построения кодирующей последовательности. Другими словами, каждый организм, в том числе и вирус, использует определенный набор правил для построения последовательностей своих ОРС, которые ограничивают общее количество вариантов, предоставляемых вырожденностью генетического кода. Использование кодонов и предвзятость контекста пар нуклеотидов отражают действие двух основных эволюционных сил: отбора по эффективности трансляции мРНК и мутационного дрейфа, действующего без разбора на кодирующую и не кодирующую последовательность [10].

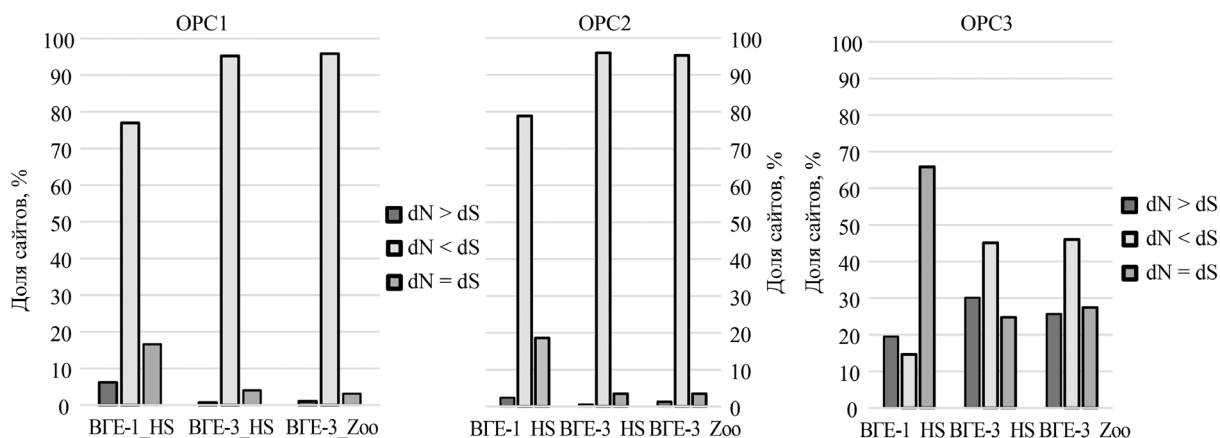


Рисунок 4. Эволюционное давление на последовательности ОРС ВГЕ-1 и ВГЕ-3, выделенные из разных хозяев



### Выводы

1. GC состав последовательностей ВГЕ, в отличие от большинства (+s)РНК вирусов превышает 50 % и в среднем составляет  $59,3 \pm 0,52$  %. Композиционные свойства разных генов ВГЕ находятся под влиянием разных направлений естественного отбора и мутационного давления. В третьем положении кодона в ОРС1 и ОРС2 предпочтительными являются нуклеотиды Т(У)/С, а в ОРС3 – С/Г.

2. ОРС1 и ОРС2 всех изученных генотипов (хозяев) ВГЕ характеризуются сбалансированным динуклеотидным составом, за исключением недопредставленности СрG и TrC, а также перепредставленностью TrG нуклеотидных пар, что коррелирует с динуклеотидным составом хозяев ВГЕ. Паттерн динуклеотидного состава ОРС3 ВГЕ характеризуется большей разнородностью, по сравнению с ОРС1 и ОРС2. Генотип специфические особенности динуклеотидной композиции свидетельствуют об адаптации последовательностей ОРС3 к внутриклеточной среде разных хозяев.

3. ВГЕ-1 и ВГЕ-3 характеризуются разными моделями использования синонимичных кодонов, что свидетельствует о независимой эволюции этих генотипов вируса. У/А концевое смещение в синонимичных кодонах ОРС2, в отличие от 3G/C, характерного для ОРС1 и ОРС3, свидетельствует о преобладающей роли естественного отбора в эволюции этого гена, особенно сильно выраженное в последовательностях ВГЕ-1.

4. Влияние трансляционного отбора на эволюцию последовательностей разных ОРС ВГЕ различно. Селективное влияние на последовательности ВГЕ-1 более выражено, чем ВГЕ-3. В последовательностях ОРС2 ВГЕ использование синонимичных кодонов смещено в большей степени, чем в ОРС1. Предпочтения в использовании синонимичных кодонов в последовательностях ОРС3 не зависят от действия трансляционного отбора и не дифференци-

ровано в зависимости от генотипа вируса. В последовательностях ВГЕ-3, выделенных из разных хозяев, отличия в смещении в использовании синонимичных кодонов под действием трансляционного отбора не выявлены.

5. Ведущим фактором эволюции последовательностей ОРС1 и ОРС2 ВГЕ является отрицательный (очищающий) отбор. Аминокислотная последовательность продукта ОРС3 ВГЕ-3 характеризуется значительной вариабельностью, возникшей под действием позитивного (движущего) отбора в процессе адаптации этого генотипа к существованию в организме разных хозяев. Последовательности ОРС3 антропонозного ВГЕ-1 более консервативны, в них преобладают нейтрально отобранные сайты, что является следствием устоявшихся взаимоотношений между вирусом и хозяином.

### Литература

1. *The characteristics of the synonymous codon usage in hepatitis B virus and the effects of host on the virus in codon usage pattern* / M. Ma [et al.] // *Virology*. – 2011. – Vol. 8. – С. 544.
2. *Hepatitis A Virus Codon Usage: Implications for Translation Kinetics and Capsid Folding* / R. M. Pintó [et al.] // *Cold Spring Harb Perspect Med.* – 2018. – Vol. 8, № 10. – P. a031781.
3. *Auewarakul, P. Composition bias and genome polarity of RNA viruses* / P. Auewarakul // *Virus Research.* – 2005. – Vol. 109, № 1. – С. 33–37.
4. *Using codon usage analysis to speculate potential animal hosts of hepatitis E virus: An exploratory study* / B. Li [et al.] // *Infection, Genetics and Evolution.* – 2022. – Vol. 101. – С. 105284.
5. *Bouquet, J. Genetic characterization and codon usage bias of full-length Hepatitis E virus sequences shed new lights on genotypic distribution, host restriction and genome evolution* / J. Bouquet, P. Chereh, N. Pavo // *Infection, Genetics and Evolution.* – 2012. – Vol. 12, № 8. – С. 1842–1853.
6. *Dinucleotide Composition in Animal RNA Viruses Is Shaped More by Virus Family than by Host Species* / F. di Giallonardo [et al.] // *Journal of Virology.* – 2017. – Vol. 91, № 8. – P. 10. – doi: 1128/jvi.02381-16.
7. *Evolution of codon usage in Zika virus genomes is host and vector specific* / A. M. Butt [et al.] // *Emerg Microbes Infect.* – 2016. – Т. 5, № 10. – С. e107.
8. *In Vitro Replication of Hepatitis E Virus (HEV) Genomes and of an HEV Replicon Expressing Green*

Fluorescent Protein / S. U. Emerson [et al.] // J Virol. – 2004. – T. 78, № 9. – С. 4838–4846.

9. Purdy, M. A. Evolutionary History and Population Dynamics of Hepatitis E Virus / M. A. Purdy, Y. E. Khudyakov // PLOS ONE. – 2010. – Vol. 5, № 12. – P. e14376.

10. Fedorov, A. Regularities of context-dependent codon bias in eukaryotic genes / A. Fedorov, S. Saxonov, W. Gilbert // Nucleic Acids Research. – 2002. – Vol. 30, № 5. – С. 1192–1197.

## References

1. *The characteristics of the synonymous codon usage in hepatitis B virus and the effects of host on the virus in codon usage pattern* / M. Ma [et al.] // Virol J. – 2011. – Vol. 8. – S. 544.

2. *Hepatitis A Virus Codon Usage: Implications for Translation Kinetics and Capsid Folding* / R. M. Pintó [et al.] // Cold Spring Harb Perspect Med. – 2018. – Vol. 8, № 10. – P. a031781.

3. *Auewarakul, P. Composition bias and genome polarity of RNA viruses* / P. Auewarakul // Virus Research. – 2005. – Vol. 109, № 1. – S. 33–37.

4. *Using codon usage analysis to speculate potential animal hosts of hepatitis E virus: An exploratory study* / B. Li [et al.] // Infection, Genetics and Evolution. – 2022. – Vol. 101. – S. 105284.

5. *Bouquet, J. Genetic characterization and codon usage bias of full-length Hepatitis E virus sequences shed new lights on genotypic distribution, host restriction and genome evolution* / J. Bouquet, P. Chereh, N. Pavio // Infection, Genetics and Evolution. – 2012. – Vol. 12, № 8. – S. 1842–1853.

6. *Dinucleotide Composition in Animal RNA Viruses Is Shaped More by Virus Family than by Host Species* / F. di Giallonardo [et al.] // Journal of Virology. – 2017. – Vol. 91, № 8. – S. 10. – doi: 1128/jvi.02381-16.

7. *Evolution of codon usage in Zika virus genomes is host and vector specific* / A. M. Butt [et al.] // Emerg Microbes Infect. – 2016. – Vol. 5, № 10. – S. e107.

8. *In Vitro Replication of Hepatitis E Virus (HEV) Genomes and of an HEV Replicon Expressing Green Fluorescent Protein* / S. U. Emerson [et al.] // J Virol. – 2004. – Vol. 78, № 9. – S. 4838–4846.

9. Purdy, M. A. Evolutionary History and Population Dynamics of Hepatitis E Virus / M. A. Purdy, Y. E. Khudyakov // PLOS ONE. – 2010. – Vol. 5, № 12. – P. e14376.

10. Fedorov, A. Regularities of context-dependent codon bias in eukaryotic genes / A. Fedorov, S. Saxonov, W. Gilbert // Nucleic Acids Research. – 2002. – Vol. 30, № 5. – S. 1192–1197.

Поступила 14.09.2023 г.