

Ж.С. Тюлько^{1,2}, В.В. Якименко²

Анализ использования кодонов в кодирующих последовательностях подсемейства *Coronavirinae*.

¹ ФГБОУ ВО ОмГМУ Минздрава России,

² ФБУН «Омский научно-исследовательский институт природно-очаговых инфекций»

Проведены анализ и сравнение особенностей использования синонимичных кодонов у вирусов подсемейства *Coronavirinae* родов *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, *Deltacoronavirus*, в том числе у вирусов SARS-CoV-2 в открытых рамках считывания ORF1ab и ORF2. Для всех изучаемых кодирующих последовательностей рассчитывались показатели относительного использования синонимичных кодонов которые в дальнейшем использовались при проведении дискриминантного анализа для выявления характерных особенностей каждой изучаемой группы последовательностей. В результате были выявлены достоверные различия в использовании синонимичных кодонов, как разными родами, так и отдельными видами коронавирусов причем SARS-CoV-2 значительно дистанцируется от большинства других бетакоронавирусов. Показано, также, что коронавирусы летучих мышей и птиц, часто занимают промежуточное положение по отношению к видам, относящимся к разным родам коронавирусов.

Ключевые слова: коронавирусы, COVID-19, показатели относительного использования синонимичных кодонов, дискриминантный анализ, эволюция.

Введение. Коронавирусы принадлежат отряду *Nidovirales*, семейство *Coronaviridae*, подсемейство *Coronavirinae*. В состав отряда *Nidovirales* входят оболочечные вирусы с инфекционной односегментной линейной одноцепочечной РНК позитивной полярности, которые имеют ряд общих черт организации генома. Помимо семейства *Coronaviridae*, отряд *Nidovirales* включает еще два семейства: *Arteriviridae* и *Roniviridae*. Первое объединяет вирусы млекопитающих, но не человека (включая вирус артериита лошадей и вирус репродуктивно-респираторного синдрома свиней), второе – вирусы членистоногих [1]. В составе подсемейства *Coronavirinae* выделяют роды *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, *Deltacoronavirus*, представители которых способны сосуществовать с широчайшим кругом позвоночных хозяев от рыб до млекопитающих [1,2]. Например, в род *Betacoronavirus* входят виды вирусов, выделенных от рукокрылых, парнокопытных и в частности, новый коронавирус (SARS-CoV-2), вызвавший пандемию в человеческой популяции в 2020 г.

Коронавирусы позвоночных способны к сравнительно быстрой адаптации к новому хозяину. Выявлена возможность приспособления коронавирусов к новым хозяевам, в том числе и людям, при возникновении мутаций у вирусов из природных резервуаров: так предполагается, что природным резервуаром SARS-CoV служат летучие мыши, промежуточные хозяева – циветты [3, 4], природным резервуаром MERS служат летучие мыши, промежуточные хозяева – верблюды [5,6]. По результатам серологических исследований населения Китая, предполагается также возможность прямого инфицирования человека от летучих мышей [7], что делает изучение коронавирусов необходимым ввиду возможности возникновения новых вариантов, опасных для человека и животных. Изменчивость коронавирусов, как и других РНК-вирусов в основном определяется двумя факторами: накоплением точечных мутаций и процессами рекомбинации, которые в этой группе вирусов происходят достаточно часто [8, 9, 10, 11, 12].

Рекомбинация, типичная для коронавирусов, может быть способом адаптации. Например, было показано, что бетакоронавирусы различных линий могут рекомбинировать, получая при этом нуклеотидную последовательность, кодирующую рецептор-связывающий домен белка S и, следовательно, возможность проникнуть в

клетки, обладающие соответствующим рецептором [13]. Нестабильность коронавируса генома может проявляться также в утрате фрагментов отдельных генов, например, в 7 и 8-й рамках считывания у SARS-CoV-2. Распространение подобной делеции связывается с иммунным отбором [14].

Ранее были представлены результаты исследований, в которых рассматривалось влияние многократного пассирования коронавирусов на различных клеточных культурах теплокровных. Они показали, что адаптация вирусов к клеткам хозяев может происходить, например, за счет накопления точечных мутаций. Так из SARS-CoV, не вызывавшего ранее заболевания мышей, путем серийных пассажей в их дыхательных путях был получен летальный для мышей вариант, имеющий шесть аминокислотных замен, связанных с адаптацией и повышенной вирулентностью. Причем авторы работы выявили смешанные вирусные популяции в летальном пуле вирусов [15]. Точечные замены, которые при этом возникают в нуклеотидных последовательностях вирусов, могут быть как синонимичными, так и несинонимичными, и их наличие и расположение зависит от многих факторов. На уровне нуклеотидной последовательности генома это может проявиться в различиях использования синонимичных кодонов. Условия окружающей среды, включая типы тканей, в которых экспрессируются гены, и конкретные клеточные условия в этих тканях, влияют на предпочтения кодонов. Исследование генов, экспрессируемых во множественных тканях человека, показало, что использование кодонов отличается для наборов генов, экспрессируемых в разных тканях, и напрямую зависит от фактического количества молекул тРНК в каждой ткани [16]. Попытки изучения особенностей использования синонимичных кодонов уже делалась для SARS-CoV-2, SARS, MERS [17], и полученные результаты желательно сопоставить с особенностями других геновидов коронавирусов, входящих подсемейство *Coronavirinae*. В настоящее время, учитывая постоянный рост объемов генетических баз данных, полный анализ этого подсемейства вирусов становится доступным.

Целью данного исследования был анализ и сравнение особенностей использования синонимичных кодонов у вирусов подсемейства *Coronavirinae* и оценка наличия или отсутствия особенностей данного процесса у SARS-CoV-2.

Материалы и методы. В данной работе анализировалось использование синонимичных кодонов в рамках считывания ORF1ab (кодирует полипротеины PP1ab и PP1a, которые расщепляются с образованием 16 неструктурных белков, NSP1-16) и ORF2 (кодирует S-белок) различных коронавирусов. Для изучения брались полноразмерные геномы, представленные в GenBank, содержащие описание кодирующей части генома. Рассмотренные виды коронавирусов представлены в Таблице 1.

Табл.1 Виды коронавирусов, полноразмерные нуклеотидные последовательности которых представлены в GenBank использовались для анализа.

Род	Вид	Количество последовательностей
Alphacoronavirus	Bat alphacoronavirus	46
	Swine enteric alphacoronavirus	6
	Alphacoronavirus 1	76
	Camel alphacoronavirus	30
	Mink coronavirus	18
	Porcine epidemic diarrhea virus	86
	Human coronavirus 229E	71
	Wencheng Shrew coronavirus	8
	Canine coronavirus strain	15
	Swine acute diarrhea syndrome coronavirus	39

	Human coronavirus NL63	64
	Transmissible gastroenteritis virus	42
	Feline infectious peritonitis virus	5
Betacoronavirus	Betacoronavirus HKU24	8
	Human coronavirus HKU1	48
	Human coronavirus OC43	163
	Murine coronavirus MHV	33
	Murine hepatitis virus	25
	Middle East respiratory syndrome-related coronavirus	118
	SARS-CoV-2	380
	Hedgehog coronavirus	3
Gammacoronavirus	Canada goose coronavirus	2
	Duck coronavirus	3
	Beluga whale coronavirus SW1	2
	Turkey coronavirus	9
	Bottlenose dolphin coronavirus	2
Deltacoronavirus	Quail coronavirus UAE-HKU30	2
	Bulbul coronavirus	2
	Porcine coronavirus HKU15	17
	Porcine deltacoronavirus	12
	Sparrow deltacoronavirus	4
	Thrush coronavirus HKU12-600	1
	Houbara coronavirus UAE-HKU28	1
	White-eye coronavirus HKU16	3
	Pigeon coronavirus UAE-HKU29	1
	Falcon coronavirus UAE-HKU27	1

Для всех изучаемых полноразмерных кодирующих последовательностей рассчитывались показатели относительного использования синонимичных кодонов, обозначаемые как $RSCU_k$ (Relative Synonymous Codon Usage), для каждого кодона k (стоп кодоны и однозначно кодируемые кодоны не рассматривались). Показатель $RSCU_k$ применяют для проведения корректных сравнений частот использования синонимичных кодонов в различных сериях [17, 18, 19, 20]. Он оценивает неслучайность появления каждого триплета k при кодировании аминокислоты, а также позволяет сравнить схемы кодирования в разных последовательностях. Большие значения $RSCU_k$ соответствуют более частому использованию кодона.

Расчет значений $RSCU_k$ может быть выполнен многими видами программного обеспечения. В нашей работе значения $RSCU_k$ получены с помощью программ, которые были созданы авторами на базе пакета статистического анализа R. Корректность работы, этих программ протестирована при сравнении получаемых результатов с результатами работы программного пакета Mega 4.0.1.

Полученные значения $RSCU_k$ для всех типов кодонов у каждого вируса сравнивались с соответствующими значениями с помощью модулей «Групповой анализ» (cluster analysis) и «Общее модели дискриминантного анализа» (general discriminant analysis) программы STATISTICA 6.

Рассчитывались также значения CUB (codon usage bias) и CG-смещение, которые также оценивают неслучайность при неравномерном использовании синонимичных кодонов. Расчет значений CUB, и содержания CG проводились на базе пакета статистического анализа R.

Считается, что смещение в использовании кодонов (CUB) может происходить в ходе эволюции или при репликации вирусной РНК как результат предпочтения некоторых кодонов из-за различий во внутриклеточном содержании транспортных РНК (тРНК), специфичных для этих кодонов (трансляционное смещение) [21, 22].

Результаты.

После проведения дискриминантного анализа сгруппированных по родам и видам нуклеотидных последовательностей коронавируса, были рассчитаны классифицирующие функции f_1 и f_2 , которые позволяют выявлять особенности использования синонимичных кодонов каждой группой последовательностей. При анализе кодирующей последовательности ORF1ab (рис.1а) выявлены достоверные различия в использовании кодонов ($P < 0,001$) разными родами коронавируса. Из полученной диаграммы видно, что функция f_1 описывает особенности использования синонимичных кодонов, отличающие род гаммакоронавирусов от других родов, а функция f_2 эффективно описывает особенности, отличающие друг от друга альфа, бета и дельтакоронавирусы.

При использовании для анализа кодирующей последовательности белка S (рис. 1б) качество дискриминации хуже и расстояние между центроидами групп меньше, что вероятно объясняется как высокой вариабельностью рецептор-связывающего домена (RBD) в шпильке шипа белка S, так и меньшей длиной нуклеотидной последовательности, кодирующей белок S, поскольку при меньшей длине дискриминация последовательностей РНК вирусов ухудшается [20].

Для бетакоронавирусов летучих мышей (рис. 1б), таких как коронавирус ночных крыланов BtCoV-HKU9 и коронавирус подковоноса Блазиуса BtCoV/BM48-31/BGR/2008), а также для гаммакоронавируса утки (коронавирус уток DK/GD/27/2014) возникают

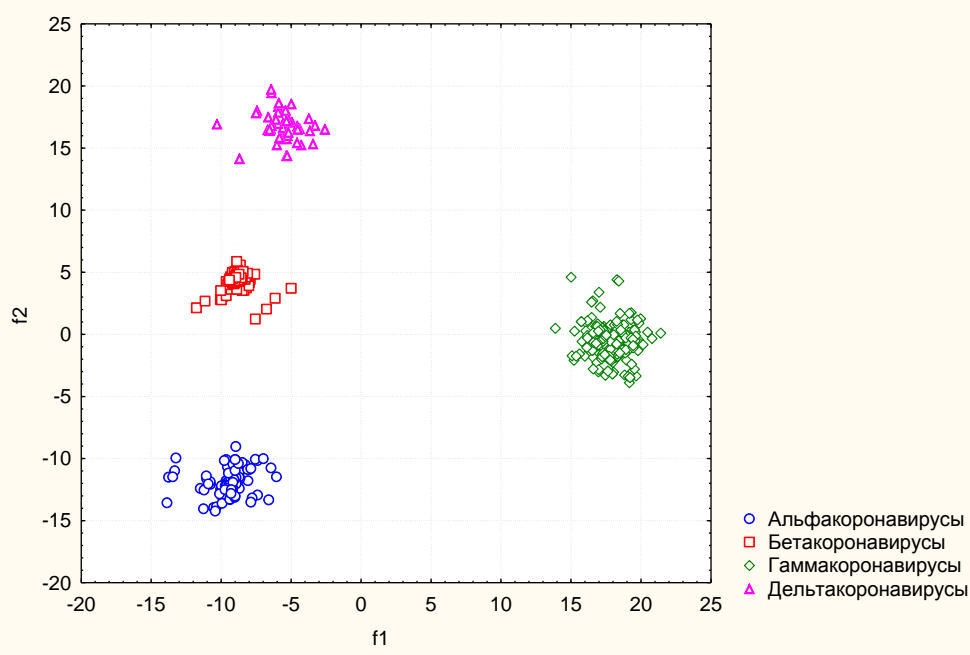


Рис.1а

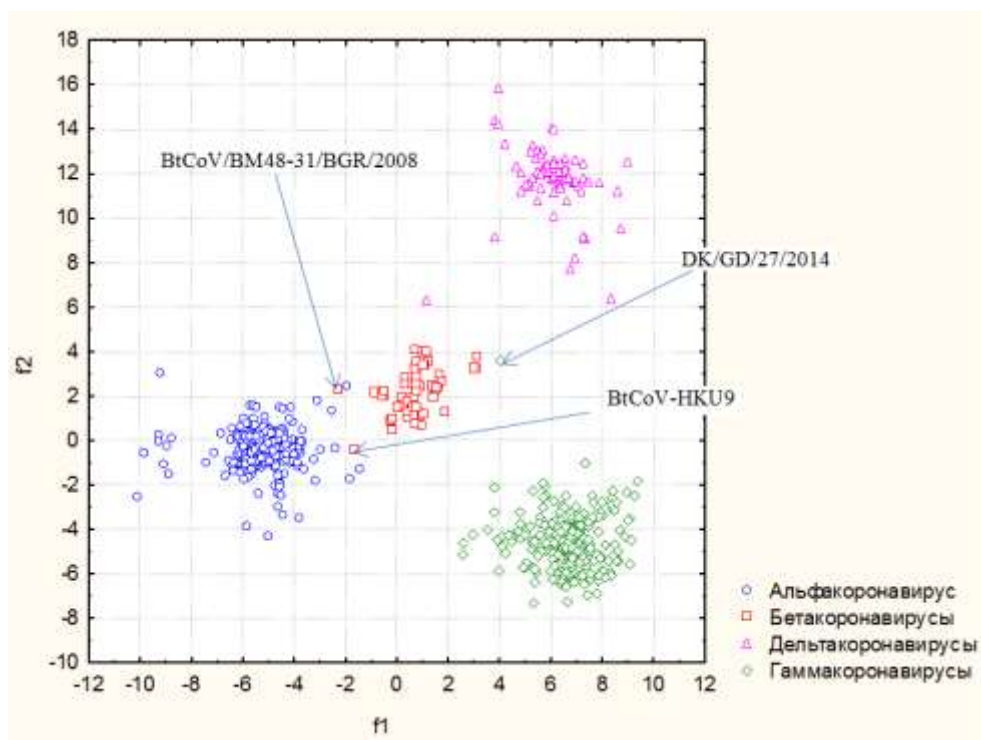


Рис.1б

Рис.1 Диаграммы рассеяния значений дискриминантных функций f_1 и f_2 (по оси абсцисс f_1 , по оси ординат - f_2 .), рассчитанных для кодирующих последовательностей коронавируса различных родов: а) рамки считывания ORF1ab; б) рамки считывания белка S. Каждый маркер на схеме соответствует последовательности из банка данных, для которой были рассчитаны значения функций. Стрелками с подписями указаны последовательности для которых классификация при помощи функций f_1 и f_2 оказалась ошибочной.

случаи недостоверной классификации. Следует заметить, что BtCoV-HKU9 относят к особой подгруппе среди бетакоронавирусов [23]. BtCoV/BM48-31/BGR/2008 относится к европейской подгруппе SARS CoV и отличается от других бетавирусов отсутствием у него рамки ORF8 [24]. РНК-изолят коронавируса уток DK/GD/27/2014 также сильно отличается от других птичьих коронавирусов и предположительно произошел вследствие рекомбинации [25, 26]. Однако общая картина разделения на группы, соответствующие родам не меняется и аналогична схеме на рис. 1а. Таким образом, проведенный анализ демонстрирует значимое различие в использовании синонимичных кодонов основными родами *Coronavirinae* (рис.1 а, б).

Следующим шагом был анализ использования синонимичных кодонов разными геновидами коронавирусов (рис.2,3). Полученная диаграмма (рис.2) показывает, что дискриминация на уровне видов оказывается менее выраженной, чем на уровне родов, хотя остается достаточно надежной для дифференцировки по использованию кодонов ($P < 0,001$) у большинства видов альфа и бетакоронавирусов.

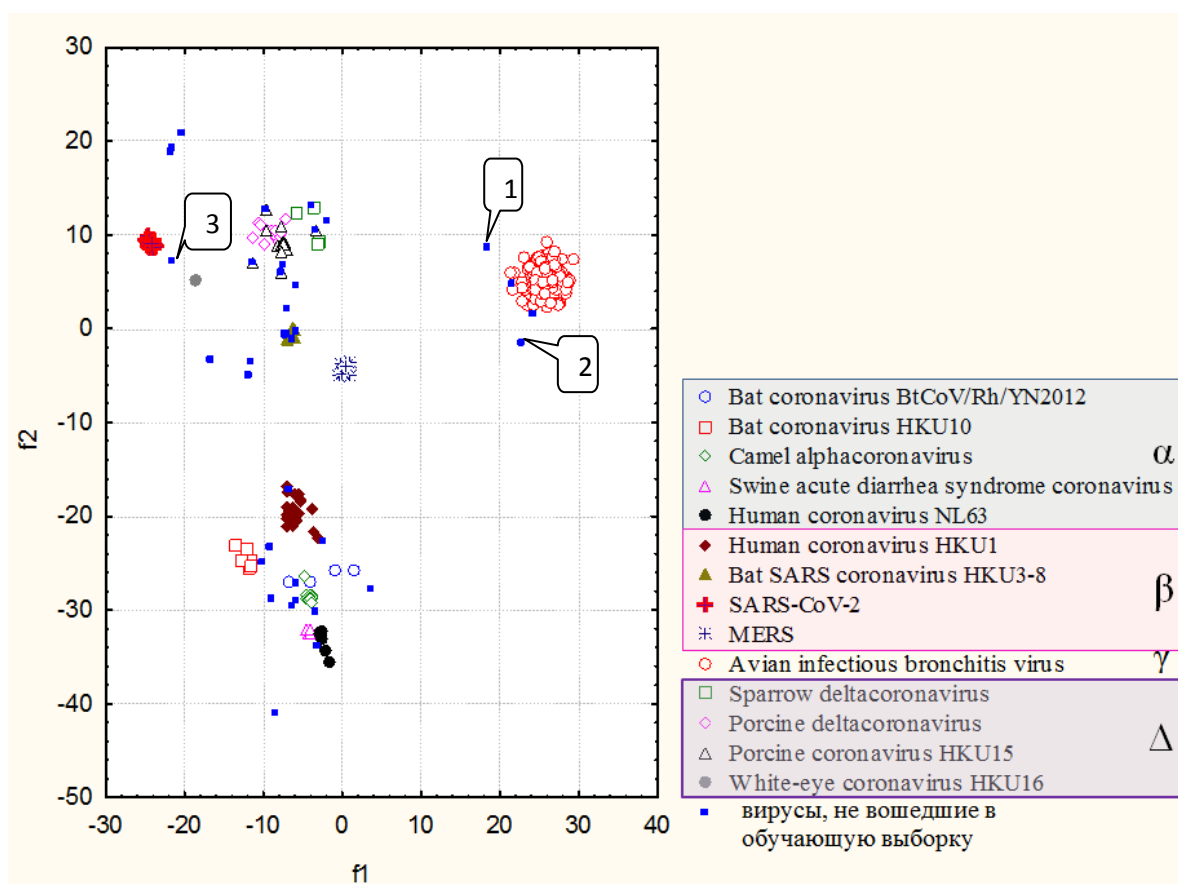


Рис.2 Диаграмма рассеяния значений дискриминантных функций f_1 и f_2 , рассчитанная по рамке считывания ORF1ab для кодирующих последовательностей коронавирусов различных видов из родов Alphacoronavirus, Betacoronavirus, Deltacoronavirus, Gammacoronavirus, Deltacoronavirus. Каждый маркер на схеме соответствует последовательности из банка данных, для которой были рассчитаны значения функций.

В случае гаммакоронавирусов большая часть имеющихся полноразмерных последовательностей относится к вирусу инфекционного бронхита птиц, остальные виды представлены одной или двумя последовательностями, например на рис. 2 цифрой 1 обозначено положение коронавируса индейки (код доступа EU022525), цифрой 2 положение коронавируса уток (KM454473). Поэтому делать определенные выводы о дифференциации видов рода дельтакоронавирусов по использованию синонимичных кодонов можно будет при дальнейшем пополнении банка данных для них.

В случае дельтакоронавирусов, как показано на рис. 2 для дельтакоронавируса свиней, вируса HKU15 и дельтакоронавируса воробьев, не получено достаточно хорошего качества дискриминации. В то же время как центроид вирусов HKU16 (№3 на рис. 2) удален от всех остальных дельтакоронавирусов и расположен ближе к некоторым бетакоронавирусам (SARS-CoV-2 и RaTG13 (MN996532)).

Вирусы летучих мышей и птиц, не вошедшие в обучающую выборку, часто занимают промежуточное положение по отношению к отдельным видам (см. рис. 2). Возможно, это связано с тем, что многие из коронавирусов эффективно циркулируют между разными видами летучих мышей (как фруктоядными, так и насекомоядными), а также среди других групп животных [27, 28], а у некоторых отмечена высокая частота рекомбинаций [8], что может приводить к возникновению геновариантов с усредненными частотными характеристиками $RSCU_k$.

Для видов из рода Betacoronavirus, представленных на рис.3, наибольшие различия наблюдаются при использовании синонимичных кодонов, кодирующих аргинин, цистеин,

глутаминовую кислоту, треонин. Наиболее сильно в этой группе отличается использование синонимичных кодонов у SARS-Cov-2, хотя общая картина предпочтения конкретных кодонов из тех, что кодируют определенную аминокислоту сохраняется для всех этих вирусов.

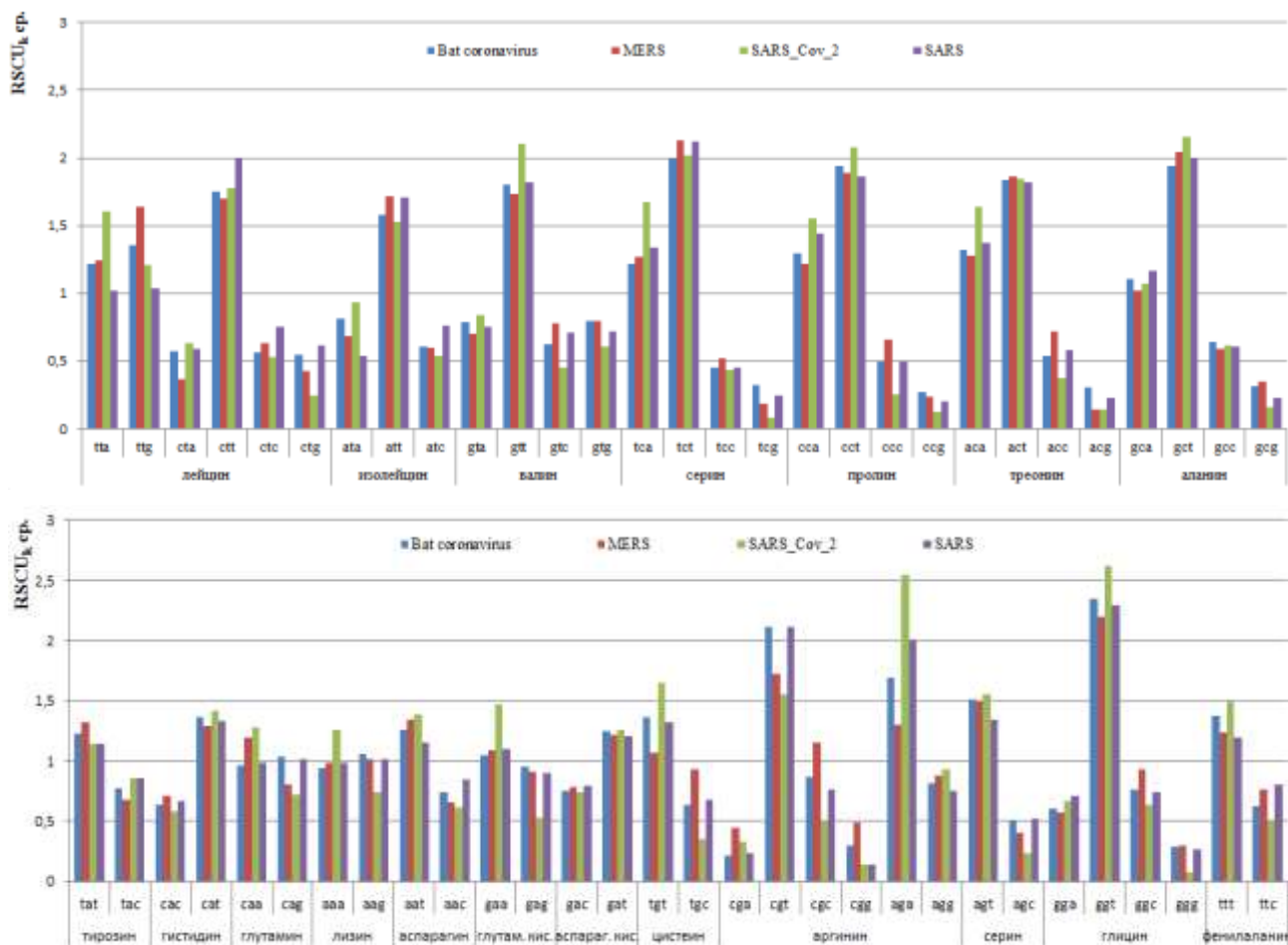


Рис.3 Средние значения $RSCU_{kcp.}$, полученные при анализе кодирующей последовательности ORF1ab для вирусов видов SARS-2, MERS и бетакоронавирусов летучих мышей. По горизонтальной оси указаны кодоны типа k и соответствующая аминокислота.

Обсуждение. SARS-CoV-2, ставший причиной мировой пандемии 2019-20 гг., относится к коронавирусам, которые способны вызывать заболевания у человека с респираторными проявлениями разной степени тяжести (MERS-CoV, SARS-CoV, HKU1, NL63, OC43, 229E). В случае коронавирусов актуальным вопросом является вопрос о процессах видообразования и видовых границах в этой группе. Эволюция таких вирусов происходит путем накопления точечных замен и осуществления рекомбинаций, а также, возможна потеря значительных участков генома в кодирующих и не кодирующих областях [17, 24, 29]. SARS-CoV-2 имеет высокую степень гомологии между нуклеотидными последовательностями географически удаленных изолятов, что подтверждает его недавнее происхождение. В то же время при сравнении отдельных участков нуклеотидной последовательности SARS-CoV-2 с гомологичными участками генома некоторых коронавирусов животных выявляется высокий - >96% - уровень гомологии [29, 30]. Зоонозное происхождение коронавирусов предполагает высокую способность их адаптации к организму хозяина, которая происходит, в том числе путем накопления синонимичных и несинонимичных компенсаторных замен.

Причинами неодинакового использования синонимичных кодонов могут быть многочисленные факторы и влияния, например процесс оптимизации трансляции [31], т.е. использование кодонов, соответствующих наиболее распространенным тРНК клеток хозяина, что увеличивает скорость и точность трансляции [32]. Или это может быть результатом влияния высокоэкспрессируемых генов, которое приводит к увеличению числа тРНК, соответствующих их триплетному составу и, как следствие – к смещению состава кодонов в остальных генах [33]. На частоту использования кодонов также может оказывать влияние спектр тканей, в которых экспрессируется ген [16]. Показано, что CUB сильнее выражено у более длинных генов, чем у более коротких, т.к. цена ошибки при трансляции увеличивается с удлинением вновь синтезируемого пептида [34, 35]. Это подтверждается при расчете значений оценок смещения в использовании кодонов для разных генов у большинства исследованных коронавирусов, где для ORF1ab оно достоверно выше, чем для более коротких генов S, E, M. Как видно из рисунка 1а,б, дискриминация по более длинным генам ORF1ab является более достоверной, чем по более короткому гену S, что детерминировано более быстрым накоплением изменений в последовательности гена S [17].

Наиболее дисперсным оказался род *Betacoronavirus* (см. рис. 2), виды, принадлежащие к этому роду значительно дистанцированы друг от друга (SARS-CoV-2, HKU1, MERS).

По результатам [17] для SARS-CoV-2 не было найдено каких-либо выделяющих его существенных отличий в использовании синонимичных кодонов по сравнению с геномами других β -коронавирусов, при том, что признавалось наличие индивидуальной схемы использования кодонов у каждого вида коронавирусов человека. Анализ показал высокое содержание нуклеотидов AU и низкое содержание GC и привел авторов к выводу о преимущественной адаптации к хозяину с помощью случайно выбранных замен, а не рекомбинации. По результатам, полученным нами, с использованием большего количества нуклеотидных последовательностей, различия в использовании синонимичных кодонов разными видами коронавирусов подтверждаются, как и имеющийся у них сдвиг в сторону высокого содержания нуклеотидов AU. При этом географическое положение места изоляции того или иного вируса не оказывает значительного влияния на характер использования синонимичных кодонов. SARS-CoV-2 формирует отдельную группу, достаточно удаленную, как от других бетакоронавирусов, так и от вирусов не включенных в обучающую выборку (см. рис. 2).

В нескольких исследованиях было высказано предположение, что летучие мыши могут быть потенциальным естественным хозяином прототипа SARS-CoV-2 [30]. Полногеномная нуклеотидная последовательность SARS-CoV-2 на 79,6% идентична последовательности SARS-CoV BJ01 (AY278488.2) коронавируса летучей мыши, изолят коронавируса летучей мыши RaTG13 (MN996532) идентичен SARS-CoV-2 на 96,2%. [30]. В другом исследовании сравниваются полногеномные образцы SARS-CoV-2, полученные от людей в 2019г.: уровень гомологии при сравнении между собой составлял 99-98%, с SARS-CoV GZ02 (AY390556) - 79-93%, с коронавирусами летучих мышей bat-SL-CoVZC45 (MG772933)- 87-98% [36]. При этом выдвигается предположение, что коронавирусы летучих мышей не являются непосредственными прототипами SARS-CoV-2, и предполагается наличие неизвестного промежуточного хозяина. Иногда роль такого хозяина отводится панголину (*Manis javanica*), т.к. SARS-CoV-2 был изолирован и от панголинов, и эти полногеномные нуклеотидные последовательности имеют 85,5–92,4% сходства с SARS-CoV-2 [29], причем наибольшее сходство наблюдается в рецептор-связывающем домене (RBD) в последовательности белка S. Это позволяет предположить, что панголин мог быть потенциальным промежуточным хозяином или в истории этого вируса наблюдалось явление рекомбинации. Однако авторы не исключают возможности, что аминокислотное сходство между RBD коронавирусов панголинов провинции Гуандун и SARS-CoV-2 обусловлено конвергентной эволюцией, а не рекомбинацией, т.к. этот

вывод согласуется тем фактом, что сходство последовательностей рецептора ACE2 выше между людьми и ящерами (84,8%), чем между людьми и летучими мышами (80,8–81,4% для *Rhinolophus* sp.) [29]. Т.е. приспособление могло происходить путем накопления точечных замен на тех участках генома, которые испытывали влияние отбора при взаимодействии с организмом-хозяином.

Предпочтения при отборе кодонов формируются мутационными отклонениями и естественным отбором для оптимизации трансляции, и помогают повысить ее эффективность и точность, однако этот фактор влияния не является единственным и действует совместно со многими другими. Исследования нуклеотидного состава выявляют такие предпочтения, но не объясняют причины их появления. Для получения однозначных выводов требуется увеличение количества секвенированных последовательностей, как известных, так и новых видов коронавирусов, которые продолжают обнаруживаться, по мере удешевления процедуры секвенирования вирусного генома, а также появляются *de novo* в процессе эволюции.

Литература.

1. Львов Д.К., Щелканов М.Ю. Коронавирусы (Coronaviridae) // Руководство по вирусологии. Вирусы и вирусные инфекции человека и животных. М.: МИА, 2013. С. 211–218.
2. Бучацкий Л.П., Макаров В.В. Нидовирусы, ассоциированные с водными животными. Ветеринария сегодня. 2020; 2 (33): 115–121.
3. Cheng V.C., Lau S.K., Woo P.C., Yuen K.Y. Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. Clin. Microbiol. Rev. 2007; 20(4):660-94. <https://cmr.asm.org/content/20/4/660>
4. Liu L., Fang Q., Fei Deng, Hanzhong Wang, Christopher E. Yi, Lei Ba, Wenjie Yu, Richard D. Lin, Taisheng Li, Hu Zh., Ho D.D., Zhang L., Chen Zh. Natural Mutations in the Receptor Binding Domain of Spike Glycoprotein Determine the Reactivity of Cross-Neutralization between Palm Civet Coronavirus and Severe Acute Respiratory Syndrome Coronavirus. Journal of Virology. 2007; 81(9): 4694–4700.
5. Cotton M., Watson S.J., Kellam P., et al. Transmission and evolution of the Middle East respiratory syndrome coronavirus in Saudi Arabia: a descriptive genomic study. Lancet. 2013; 382: 1993–2002; [https://doi.org/10.1016/S0140-6736\(13\)61887-5](https://doi.org/10.1016/S0140-6736(13)61887-5)
6. Reusken C.B., Raj V.S., Koopmans M.P., Haagmans B.L. (2016). Cross host transmission in the emergence of MERS coronavirus. *Current opinion in virology*. 2016; 16, 55–62. <https://doi.org/10.1016/j.coviro.2016.01.004>
7. Wang N, Li S.Y., Yang X.L., Huang H.M., Zhang Y.J., Guo H., Luo C.M., Miller M., Zhu G., Chmura A.A., Hagan E., Zhou J.H., Zhang Y.Z., Wang L.F., Daszak P., Shi Z.L. Serological Evidence of Bat SARS-Related Coronavirus Infection in Humans, China. Virol Sin. 2018; 33(1):104-107. <https://pubmed.ncbi.nlm.nih.gov/29500691/>
8. Lai M.M.C., Baric R.S., Makino S., Keck J.G., Egbert J., Leibowitz J., Stohlman S.A. Recombination between Nonsegmented RNA Genomes of Murine Coronaviruses. Journal of Virology. 1985; 56(2):449-456. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC252599/>
9. Lai M.M.C., Brayton P.R., Armen R.C., Patton C.D., Pugh C., Stohlman S.A. Mouse Hepatitis Virus A59: mRNA Structure and Genetic Localization of the Sequence Divergence from Hepatotropic Strain MHV-3. Journal of Virology. 1981; 39(4):823-834. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC171315/>
10. Yu W.-B., Tang G.-D., Zhang L., Corlett R.T. Decoding the evolution and transmissions of the novel pneumonia coronavirus (SARS-CoV-2) using the whole genomic data. Zoological Research. 2020; 41(3): 247–257. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7231477/>

11. Zhao Zh., Li H., Wu X., Zhong Y., Zhang K., Zhang Y.-P., Boerwinkle E., Fu Y.-X. Moderate mutation rate in the SARS coronavirus genome and its implications. *BMC Evolutionary Biology*. 2004;4:21. <http://www.biomedcentral.com/1471-2148/4/21>
12. Yang Y, Liu C, Du L, Jiang S, Shi Z, Baric RS, Li F. Two mutations were critical for bat-to-human transmission of Middle East respiratory syndrome coronavirus. *Journal of Virology*. 2015; 89:9119–9123. <https://jvi.asm.org/content/89/17/9119>
13. Letko M., Miazgowiec K., McMinn R., Seifert S.N., Sola I., Enjuanes L., Carmody A., van Doremalen N., Munster V.. Adaptive Evolution of MERS-CoV to Species Variation in DPP4. *Cell Rep*. 2018; 24 (7):1730-1737.
14. Su Y.C.F., Anderson D.E., et al. Discovery and Genomic Characterization of a 382-Nucleotide Deletion in ORF7b and ORF8 during the Early Evolution of SARS-CoV-2. *mBio* Jul. 2020;11 (4). <https://mbio.asm.org/content/11/4/e01610-20>
15. Roberts A., Deming D., Paddock C.D., Cheng A., Yount B., et al. A mouse-adapted SARS-coronavirus causes disease and mortality in BALB/c mice. *PLoS Pathog*. 2007; 3(1): e5. doi:10.1371/journal.ppat.0030005.
16. Kotlar, D., Lavner, Y. The action of selection on codon bias in the human genome is related to frequency, complexity, and chronology of amino acids. *BMC Genomics* 7, 67 (2006). <https://doi.org/10.1186/1471-2164-7-67>.
17. Dilucca M., Forcelloni S., Georgakilas A.G., Giansanti A., Pavlopoulou A. Codon Usage and Phenotypic Divergences of SARS-CoV-2 Genes. *Viruses*. 2020;12(5):498. doi:10.3390/v12050498.
18. Бутвиловский, А.В. Бутвиловский В.Э., Черноус Е.А. Изучение стратегии кодирования белков. *Медицинский журнал*. 2009; (2): 24-7.
19. Perriere, G., Thioulouse J. Use and misuse of correspondence analysis in codon usage studies. *Nucleic Acids Research*. 2002; 30(20): 4548-55. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC137129/>
20. Якименко В.В., Тюлько Ж.С. Классификация флавивирусных геномов по результатам дискриминантного анализа показателей относительного использования синонимичных кодонов: информационно-методическое письмо. ООО Издательский центр «Омский научный вестник», 2015.-20 с.
21. Лукашев, В.В. Молекулярная эволюция и филогенетический анализ / В.В. Лукашев. -М.: БИНОМ, 2009. -256 с.
22. Cardinale, D.J., DeRosa K., Duffy S. Base Composition and Translational Selection are Insufficient to Explain Codon Usage Bias in Plant Viruses. *Viruses*. 2013; 5(1):162-81. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3564115/>
23. Woo P.C., Wang M., Lau S.K., et al. Comparative analysis of twelve genomes of three novel group 2c and group 2d coronaviruses reveals unique group and subgroup features. *Journal of Virology*. 2007; 81(4):1574-1585. <https://jvi.asm.org/content/81/4/1574>
24. Drexler JF, Gloza-Rausch F, Glende J, et al. Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. *Journal of Virology*. 2010; 84(21):11336-11349. <https://jvi.asm.org/content/84/21/11336>
25. Zhuang Q-Y, Wang K-C, Liu S, Hou G-Y, Jiang W-M, Wang S-C, et al. (2015) Genomic Analysis and Surveillance of the Coronavirus Dominant in Ducks in China. *PLoS ONE* 10(6): e0129256. <https://doi.org/10.1371/journal.pone.0129256>
26. Zhuang Q., Liu S., Zhang X., et al. Surveillance and taxonomic analysis of the coronavirus dominant in pigeons in China. *Transbound Emerg Dis*. 2020; 67(5):1981-1990. <https://doi.org/10.1111/tbed.13541>
27. Lau S.K., Poon R.W., Wong B.H., et al. Coexistence of different genotypes in the same bat and serological characterization of Rousettus bat coronavirus HKU9 belonging to a novel Betacoronavirus subgroup. *Journal of Virology*. 2010; 84(21):11385-11394. <https://jvi.asm.org/content/84/21/11385>

28. Lau S.K., Li K.S., Tsang A.K., et al. Recent transmission of a novel alphacoronavirus, bat coronavirus HKU10, from Leschenault's rousettes to pomona leaf-nosed bats: first evidence of interspecies transmission of coronavirus between bats of different suborders. *Journal of Virology*. 2012; 86(21):11906-11918. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3486284/>
29. Lam T.T., Jia N., Zhang Y.W., Shum M.H. et.al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature*. 2020 Jul; 583(7815):282-285. doi: 10.1038/s41586-020-2169-0.
30. Zhou P., Yang X.L., Wang X.G., et.al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020; 579 (7798):270-273. doi: 10.1038/s41586-020-2012-7.
31. Drummond, D.A., Raval A. , Wilke C.O. A single determinant dominates the rate of yeast protein evolution. *Molecular Biology and Evolution*. 2006; 23(2): 327–337. <https://pubmed.ncbi.nlm.nih.gov/16237209/>
32. Ingvarsson P.K. Gene Expression and Protein Length Influence Codon Usage and Rates of Sequence Evolution in *Populustremula* // *Molecular Biology and Evolution*, Volume 24, Issue 3, March 2007, Pages 836–844.
33. Kahali B., Basak S., Ghosh T.C. Reinvestigating the codon and amino acid usage of *S. cerevisiae* genome: a new insight from protein secondary structure analysis// *BiochemBiophys Res Commun*. 2007;354(3):693-699.
34. Qin H., Wu W.B., Comeron J.M., Kreitman M., Li W.H. Intragenic spatial patterns of codon usage bias in prokaryotic and eukaryotic genomes. *Genetics*. 2004; 168(4):2245-2260.
35. Marais G, Piganeau G. Hill-Robertson interference is a minor determinant of variations in codon bias across *Drosophila melanogaster* and *Caenorhabditis elegans* genomes. *Molecular Biology and Evolution*. 2002;19(9):1399-1406. <https://academic.oup.com/mbe/article/23/2/327/1118974>
36. Lu R., Zhao X., Li J., et. al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*. 2020; 22;395 (10224) :565-574. [https://www.thelancet.com/article/S0140-6736\(20\)30251-8/fulltext](https://www.thelancet.com/article/S0140-6736(20)30251-8/fulltext)