

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
имени М.В. ЛОМОНОСОВА  
БИОЛОГИЧЕСКИЙ ФАКУЛЬТЕТ

*На правах рукописи*

**Вакуленко Юлия Александровна**

**Закономерности рекомбинации и эволюции у ряда социально значимых  
(+)РНК-вирусов млекопитающих  
1.5.10 «Вирусология»**

**ДИССЕРТАЦИЯ**

на соискание ученой степени  
кандидата биологических наук

Научный руководитель:  
д.м.н., член-корр. РАН Лукашев Александр Николаевич

Москва – 2023

## Оглавление

СПИСОК СОКРАЩЕНИЙ .....	6
ВВЕДЕНИЕ .....	8
1 ОБЗОР ЛИТЕРАТУРЫ .....	17
1.1 Эволюция (+)РНК-вирусов.....	17
1.1.1. Мутации.....	17
1.1.2. Естественный отбор и генетический дрейф .....	19
1.1.3. Рекомбинация.....	21
1.1.4. Миграция.....	23
1.2 Общая характеристика калицивирусов .....	24
1.2.1 Классификация .....	24
1.2.2 Физико-химические свойства и структура вириона.....	28
1.2.3 Структура генома .....	30
1.2.4 Репликация .....	31
1.2.5 Рекомбинация калицивирусов .....	33
1.3 Общая характеристика пикорнавирусов .....	36
1.3.1 Классификация .....	36
1.3.2 Физико-химические свойства, структура вириона.....	38
1.3.3 Структура генома .....	39
1.3.4 Репликация энтеровирусов .....	41
1.3.5 Рекомбинация энтеровирусов.....	42
1.3.6 Патогенность неполиомиелитных энтеровирусов .....	42
1.3.7 Методы типирования энтеровирусов .....	45
1.3.8 Особенности эволюции типов энтеровирусов.....	46
1.3.9 Возможные механизмы происхождения типов энтеровирусов .....	50

1.4	Общая характеристика астровирусов.....	54
1.4.1	Классификация .....	54
1.4.2	Физико-химические свойства и структура вириона.....	56
1.4.3	Структура генома .....	57
1.4.4	Репликация .....	59
1.4.5	Рекомбинация астровирусов.....	61
1.5	Общая характеристика коронавирусов .....	63
1.5.1	Классификация .....	63
1.5.2	Физико-химические свойства и структура вириона.....	66
1.5.3	Структура генома .....	67
1.5.4	Репликация .....	68
1.5.5	Рекомбинация коронавирусов .....	69
2	МАТЕРИАЛЫ И МЕТОДЫ .....	73
2.1	Данные .....	73
2.1.1	Подготовка выравниваний полных геномов для анализа рекомбинации.....	73
2.1.2	Изоляты CVA2, выделенные от больных с острыми вялыми параличами .....	77
2.1.3	Подготовка выравниваний для филогенетического анализа типов энтеровирусов .....	77
2.1.4	Подготовка выравнивания полной последовательности VP1 подтипа EV-A71.....	82
2.2	Определение нуклеотидной последовательности типизирующего фрагмента VP1 изолятов CVA2, выделенных от больных с ОВП .....	83
2.3	Анализ рекомбинации.....	83

2.4 Методы сокращения выборок последовательностей из референсного выравнивания полных генов VP1.....	88
2.5 Филогенетический анализ.....	90
2.5.1 Построение деревьев максимального правдоподобия .....	90
2.5.2 Байесовский филогенетический анализ норовирусов .....	90
2.5.3 Байесовский филогенетический анализ типизирующего фрагмента VP1 типов энтеровирусов.....	91
2.5.4 Байесовский филогенетический анализ полной последовательности VP1 типа EV-A71 .....	92
2.6 Симуляция ошибок секвенирования и ошибок в аннотации в выравниваниях .....	92
2.7 Вычисление времени полужизни рекомбинантных форм норовирусов	93
3 РЕЗУЛЬТАТЫ.....	95
3.1 Метод для детекции и визуализации рекомбинации.....	95
3.2 Рекомбинация у калицивирусов .....	97
3.2.1 Рекомбинация у норовирусов .....	97
3.2.2 Рекомбинация у саповирусов .....	106
3.2.3 Рекомбинация у лаговирусов.....	110
3.3 Рекомбинация у астровирусов.....	115
3.4 Рекомбинация у коронавируса .....	120
3.5 Влияние способа формирования выборки нуклеотидных последовательностей и артефактов выборки на результат Байесовского филогенетического анализа .....	134
3.5.1 Влияние размера набора данных и способа формирования выборки на оценки эволюционных параметров .....	136

3.5.2	Влияние ошибок секвенирования на оценки эволюционных параметров.....	140
3.5.3	Влияние ошибок аннотации дат выделения вируса на оценки эволюционных параметров.....	143
3.6	Молекулярная эволюция типов энтеровирусов человека .....	146
3.6.1	Характеристика выборки последовательностей энтеровирусов. Оценка влияния неравномерности выборки на результаты филогенетического анализа.....	146
3.6.2	Особенности топологии филогенетических деревьев для отдельных типов энтеровирусов.....	152
3.6.3	Филогенетический анализ видов <i>Enterovirus A</i> и <i>Enterovirus B</i> .....	156
3.6.4	Молекулярно-генетический анализ клинических изолятов коксакивируса A2, выделенных от детей с острым вялым параличом...	158
4	ОБСУЖДЕНИЕ .....	161
4.1	Общие закономерности естественной рекомбинации вирусов семейств <i>Caliciviridae</i> , <i>Astroviridae</i> и <i>Coronaviridae</i> .....	161
4.2	Влияние способа формирования выборки нуклеотидных последовательностей и артефактов выборки на результат Байесовского филогенетического анализа.....	169
4.3	Молекулярная эволюция типов энтеровирусов.....	173
	ВЫВОДЫ.....	178
	СПИСОК ЛИТЕРАТУРЫ.....	179

## СПИСОК СОКРАЩЕНИЙ

- BCoV – коронавирус быков  
BF – коэффициент Байеса  
BoAstV – астровирус быков  
CaAstV – астровирус собак  
CCoV – коронавирус собак  
CVA2 – коксакивирус A2  
EBHSV – European brown hare syndrome virus  
FAstV – астровирус кошек  
FCV – калицивирус кошек  
FeCoV – коронавирус кошек  
HaCV – калицивирус зайцев  
HAstV – астровирус человека  
HPD – наибольшие значения плотности апостериорной вероятности  
IBV – Вирус инфекционного бронхита  
ICTV – Международный комитет по таксономии вирусов  
IRES – внутренний сайт посадки рибосом (internal ribosome entry site)  
ML – максимальное правдоподобие  
MERS-CoV – вирус ближневосточного респираторного синдрома  
MHV – вирус мышинного гепатита  
MNV – норовирус мышей  
MuAstV – астровирус мышей  
PAstV – астровирус свиней  
PEDV – вирус эпидемической диареи свиней  
RCV – калицивирус кроликов  
RHDV – вирус геморрагической болезни кроликов  
SARS-CoV – вирус тяжёлого острого респираторного синдрома  
TCoV – коронавирус индейки  
TGEV – вирус трансмиссионного гастроэнтерита  
TRS – transcriptional regulatory sequence  
 $\alpha$ -CoV – род *Alphacoronavirus*  
 $\beta$ -CoV – род *Betacoronavirus*  
 $\gamma$ -CoV – род *Gammacoronavirus*  
 $\delta$ -CoV – род *Deltacoronavirus*  
БИК – Байесовский информационный критерий  
график СПГР – график соответствия попарных генетических расстояний  
гРНК – геномная РНК

дерево МСС – дерево максимального правдоподобия клад  
ДНК – дезоксирибонуклеиновая кислота  
з/с/г – замен/сайт/год  
матрица ОНР – матрица отклонений попарных расстояний  
млн – миллион  
НСО – неструктурная область  
нт – нуклеотид  
НТО – нетранслируемая область  
НТФ – нуклеотидтрифосфат  
ОРС – открытая рамка считывания  
оцРНК – одноцепочечная РНК  
оц(+)-РНК-вирус – вирус с одноцепочечным РНК-геномом положительной полярности относительно мРНК  
ПЦР – полимеразная цепная реакция  
(+)РНК-вирус – вирус с РНК-геномом положительной полярности относительно мРНК  
РзРп – РНК-зависимая РНК полимераз  
РНК – рибонуклеиновая кислота  
РСД – рецептор-связывающий домен  
сгРНК – субгеномная РНК  
СКО – среднее квадратическое отклонение  
СО – структурная область  
УФ-излучение – ультрафиолетовое излучение  
ЭПР – эндоплазматический ретикулум

## **ВВЕДЕНИЕ**

### *Актуальность*

Вирусы млекопитающих с РНК-геномом положительной полярности относительно мРНК, или (+)РНК-вирусы, вызывают широкий спектр заболеваний человека и животных, например, полиомиелит, гастроэнтерит, болезни органов дыхания, ящур и др. Кроме того, (+)РНК-вирусы являются частым источником новых инфекций, таких как менингоэнцефалит, вызываемый энтеровирусом 71 типа, SARS, MERS и COVID-19. Рекомбинация широко распространена у (+)РНК-вирусов и, наряду с высокой скоростью накопления замен, является фактором, обеспечивающим их генетическое разнообразие. Несмотря на то, что большое количество исследований посвящено рекомбинации и эволюции конкретных патогенов человека и животных, знания об эволюции (+)РНК-вирусов в целом остаются фрагментарными. Последний систематический анализ рекомбинации у пяти основных семейств (+)РНК-вирусов был опубликован в 2006 году. За последние пятнадцать лет количество вирусных последовательностей в базе данных GenBank выросло в сотни раз, что позволяет провести комплексный анализ на новом уровне, а также требует разработки нового биоинформатического инструментария для обработки геномных последовательностей. Рекомбинация может привести к более быстрым по сравнению с накоплением мутаций изменениям фенотипа вируса, что позволяет вирусам менять клеточный тропизм, хозяев и осваивать новые экологические ниши, поэтому анализ закономерностей рекомбинации необходим для понимания закономерностей возникновения новых вирусов в частности и принципов существования генетической информации в мире РНК-вирусов в целом.

Одной из наиболее актуальных групп (+)РНК-вирусов как источника возникающих заболеваний являются энтеровирусы человека. Как правило, энтеровирусная инфекция проходит без клинических проявлений, но иногда энтеровирусы могут вызывать тяжелые заболевания, такие как полиомиелит,



серозный менингит, энцефалит, сепсис-подобное заболевание новорожденных и др. Регулярно открывают новые серотипы, раз в несколько десятилетий возникают новые клинические формы энтеровирусной инфекции, однако понимания механизмов происхождения и возможности исчезновения серотипов на данный момент нет.

Таким образом, исследование является комплексным анализом в высококонкурентной области вирусной генетики, нацеленным на анализ рекомбинации и проблемы возникновения новых вирусов как в разрезе генома, так и в динамике по времени, которое объединяет и обобщает в единой системе многочисленные описательные исследования, выполненные ранее.

### ***Степень разработанности темы***

Рекомбинация была описана практически у всех (+)РНК-вирусов. Наиболее значимый систематический анализ рекомбинации у пяти основных семейств (+)РНК-вирусов был опубликован в 2006 году [1]. При этом подавляющее большинство работ было сосредоточено на очень узких группах вирусов, как правило, уровня рода или вида. Несмотря на сотни таких сообщений, аналогии между профилями рекомбинации у разных семейств проводились очень редко. Тем более, не было исследований рекомбинации у разных семейств с использованием идентичной методологии.

Мощным инструментом для решения фундаментальных и прикладных задач вирусологии являются Байесовские филогенетические методы, которые вошли в практику около 15 лет назад [2]. Байесовские методы позволяют восстановить филогенетическую историю вирусов во времени и пространстве [3,4]. Вопрос времени возникновения современного разнообразия вирусов также остается открытым. Для некоторых типов энтеровирусов (из более чем 100) было показано, что общий предок вирусов, принадлежащих одному типу, возник не более 150 лет назад [5,6]. Систематического анализа динамики и закономерностей эволюции типов энтеровирусов в глобальном масштабе не проводилось. Кроме того, требовалось исключить влияние артефактов

анализа на результаты молекулярного датирования событий в эволюции энтеровирусов.

Таким образом, имеется систематический пробел в наших знаниях о генетике циркулирующих вирусов.

### **Цели и задачи**

Цель работы – провести анализ генетического разнообразия, изменчивости и рекомбинации у ряда социально значимых (+)РНК-вирусов млекопитающих, принадлежащих семействам *Picornaviridae*, *Caliciviridae*, *Astroviridae* и *Coronaviridae*.

Задачи:

- 1) Изучить закономерности естественной рекомбинации у вирусов, принадлежащих семействам *Caliciviridae*, *Astroviridae* и *Coronaviridae* на разных таксономических уровнях и с учетом динамики во времени.
- 2) Изучить принципы происходящей за счет рекомбинации модульной эволюции геномов вирусов, принадлежащих семействам *Caliciviridae*, *Astroviridae* и *Coronaviridae*.
- 3) Систематически изучить филодинамику наиболее распространенных типов неполиомиелитных энтеровирусов человека видов А-D.
- 4) Определить нуклеотидную последовательность типизирующего фрагмента VP1 и изучить филогенетические взаимоотношения изолятов коксакивируса А2 (CVA2), выделенных от пациентов с острыми вялыми параличами, в сопоставлении с другими последовательностями CVA2, доступными в базе GenBank.

Для достижения целей работы также было необходимо решить следующие вспомогательные задачи:

- 1) Разработать инструментарий для автоматизации работы с нуклеотидными и аминокислотными последовательностями вирусов.

- 2) Разработать инструментарий для детекции и визуализации рекомбинационных событий в условиях частой рекомбинации, разрушающей филогенетический сигнал.
- 3) Оценить влияние артефактов в геномных данных и их метаданных на результаты молекулярного датирования событий в эволюции энтеровирусов.

### **Научная новизна**

Проведен систематический анализ рекомбинации в семействах (+)РНК-вирусов *Caliciviridae*, *Astroviridae* и *Coronaviridae*. Показано, что профили рекомбинации внутри участков генома, кодирующих структурные и неструктурные белки, различаются между семействами, но сходны внутри одного семейства РНК-вирусов. Для калицивирусов характерен обмен целыми блоками генов неструктурных и структурных белков и крайне редкая рекомбинация внутри участков, кодирующих структурные и неструктурные белки. Для астровирусов характерен обмен целыми блоками генов неструктурных и структурных белков и умеренная рекомбинация внутри структурных и неструктурных генов. У коронавирусов наиболее часто рекомбинация происходит между геном структурного белка S и другими участками генома. Рекомбинация внутри гена белка S происходит чаще, чем в неструктурной области, причем с большей вероятностью происходит обмен целыми доменами белка S. Гены остальных структурных белков (N, E, M) по профилю естественной рекомбинации соответствуют генам неструктурных белков.

Впервые выполнено систематическое изучение филодинамики типов (фрагментов генома, кодирующих капсидные белки) энтеровирусов человека в глобальном масштабе. Показано, что глобальная популяция энтеровирусов разных типов регулярно подвергалась «бутылочным горлышкам», что приводило к вымиранию линий. Отдельные типы могут регулярно подвергаться почти полному исчезновению практически в глобальном масштабе даже в течение десятков лет, однако не зафиксировано ни одного примера безвозвратного исчезновения типа энтеровируса. Определены последовательности

типизирующих фрагментов VP1 коксакивирусов A2, ассоциированных со случаями острого вялого паралича, выделенных в рамках надзора за неполиомиелитными энтеровирусами с 2001 по 2020 год. Опровергнута гипотеза о возникновении варианта коксакивируса A2 с измененными патогенетическими характеристиками.

### ***Теоретическая и практическая значимость***

Теоретическая значимость работы заключается в том, что были выявлены особенности естественной рекомбинации трех семейств (+)РНК-вирусов – *Caliciviridae*, *Astroviridae* и *Coronaviridae*, а также закономерности эволюции в глобальном масштабе участков, кодирующих капсидный белок VP1 неполиомиелитных энтеровирусов (семейство *Picornaviridae*).

Разработаны и опубликованы в журнале «Инфекция и иммунитет» НИИ эпидемиологии и микробиологии имени Пастера Роспотребнадзора рекомендации по подготовке и аннотации последовательностей геномов вирусов для филогенетического анализа.

### ***Методология и методы исследования***

При выполнении исследований использовались специализированные программы, для некоторых задач были разработаны программные сценарии на языках R и Python.

Для получения информации о вирусных последовательностях из записей GenBank, выделения необходимых участков полногеномных вирусных последовательностей, полуавтоматического формирования выравниваний последовательностей полных геномов вирусов автором были разработаны скрипты на языке Python. Для множественного выравнивания последовательностей использовалась программа mafft v.7.304.

Оценка временного сигнала в наборах вирусных последовательностей осуществлялась в программе TempEst v.1.5.3. Филогенетический анализ нуклеотидных последовательностей вирусов был выполнен с помощью программ IQ-TREE v.1.6.12, BEAST v.1.8.3 и v.1.10.4. Для визуализации филогенетических деревьев использовались возможности программы FigTree v.1.4.4

и пакета *ggtree* (язык R). Визуализация результатов Байесовского филогенетического анализа для разных выборок энтеровирусных последовательностей осуществлялась средствами пакета *matplotlib* языка Python.

Для анализа рекомбинации в вирусных последовательностях использовались общепринятые методы, реализованные в программе RDP4. Визуализация результатов программы RDP4 была осуществлена с помощью пакета *ggplot2* языка R. Также автором был разработан метод для детекции и визуализации рекомбинации с помощью матриц отклонений попарных расстояний и графиков соответствия попарных генетических расстояний. Метод был реализован на языке R в качестве пакета *RecDplot* и интерактивного приложения *ShinyApp*.

Статистические тесты выполнялись в программе *GraphPad Prism v.8*.

Последовательности типизирующего фрагмента энтеровирусов типа коксакивирус A2, выделенных от пяти пациентов с острыми вялыми параличами, были определены с помощью секвенирования по Сэнгеру.

### ***Объект и предмет исследования***

Объект исследования – нуклеотидные последовательности геномов вирусов, принадлежащих семействам *Picornaviridae*, *Caliciviridae*, *Astroviridae*, *Coronaviridae*; энтеровирусы типа коксакивирус A2, выделенные от больных с острыми вялыми параличами.

Предмет исследования – закономерности естественной рекомбинации представителей семейств *Caliciviridae*, *Astroviridae* и *Coronaviridae*, эволюция участков, кодирующих капсидные белки энтеровирусов (семейство *Picornaviridae*).

### ***Положения, выносимые на защиту***

1. Общим свойством эволюции вирусов с несегментированным (+)РНК-геномом семейств *Caliciviridae*, *Astroviridae*, *Coronaviridae* является высокая частота рекомбинации между участками, кодирующими структурные и неструктурные белки. Профили рекомбинации внутри участков генома, коди-

рующих структурные и неструктурные белки, различаются между семействами, но сходны внутри одного семейства вирусов.

2. У коронавирусов наиболее часто рекомбинация происходит между геном структурного белка S и другими участками генома. Рекомбинация внутри гена белка S происходит чаще, чем в неструктурной области, причем с большей частотой происходит обмен целыми доменами белка S.

3. Участки генома, кодирующие капсидные белки энтеровирусов, могут регулярно проходить «бутылочное горлышко», т.е. подвергаться почти полному исчезновению и замещаться новыми вариантами практически в глобальном масштабе даже в течение десятков лет.

4. Выявлено отсутствие связи вирусов типа коксакивирус A2, выделенных от больных с острыми вялыми параличами, с конкретной филогенетической группой, что опровергает гипотезу о возникновении варианта вируса с измененными патогенетическими характеристиками.

### ***Степень достоверности***

Достоверность результатов определяется применением общепринятых методов обработки и анализа геномных данных, автоматизированных методов подготовки данных, исключающих «человеческий фактор», статистических критериев при анализе данных, проведением симуляционных исследований для оценки влияния возможных ошибок в исходных данных на результат. Достоверность результатов и выводов также подтверждена процедурой независимого рецензирования публикаций по теме диссертации, согласованностью с независимыми исследованиями в данной области, значительным уровнем цитирования публикаций.

### ***Связь с государственными программами***

Диссертационная работа выполнена при поддержке грантов РНФ 19-15-00055 (руководитель А.Н. Лукашев), РФФИ 19-115-50403 (руководитель А.А. Девяткин), РНФ 22-15-00230 (руководитель А.Н. Лукашев).

### ***Личный вклад***

Соискателем были получены основные результаты представленной работы. Соискателем были выполнены экспериментальные исследования по определению и анализу частичной нуклеотидной последовательности геномов коксакивирусов А2. Соискателем были разработаны методы детекции и визуализации рекомбинации, используемые в данной работе. Интерпретация результатов, представление результатов на научных конференциях и подготовка их к публикации в рецензируемых научных журналах также составляют личный вклад автора.

### ***Публикации***

По материалам диссертационной работы опубликовано 5 экспериментальных и 2 обзорные статьи в журналах, индексируемых в RSCI, WoS или Scopus:

1. Lukashev A., **Vakulenko Y.** Molecular evolution of types in non-polio enteroviruses // J Gen Virol. Microbiology Society, 2017. Vol. 98, № 12. P. 2968–2981. Q2, IF Scopus=2.78
2. Lukashev, A.N.; **Vakulenko, Y.A.**; Turbabina, N.A.; Deviatkin, A.A.; Drexler, J.F. Molecular epidemiology and phylogenetics of human enteroviruses: Is there a forest behind the trees? // Rev. Med. Virol. 2018. Vol. 28, № 6. P. e2002. Q1, IF Scopus=3.943
3. **Vakulenko Y.**, Deviatkin A., Lukashev A. The effect of sample bias and experimental artefacts on the statistical phylogenetic analysis of picornaviruses // Viruses. 2019. Vol. 11, № 11. Q1, IF Scopus=4.013
4. **Вакуленко Ю.А.**, Лукашев А.Н., Девяткин А.А. Использование методов статистической филогенетики в вирусологии // ИНФЕКЦИЯ И ИММУНИТЕТ. 2021. Vol. 11, № 1. P. 42–56. Q4, IF Scopus=0.598
5. **Vakulenko Y.**; Deviatkin A.; Drexler J.F.; Lukashev A. Modular Evolution of Coronavirus Genomes // Viruses. 2021. Vol. 13, № 7. P. 1270. Q1, IF Scopus=5.708

6. Ivanova O.E., Shakaryan A.K., Morozova N.S., **Vakulenko Y.A.**, Eremeeva T.P., Kozlovskaya L.I., Baykova O.Y., Shustova E.Y., Mikhailova Y.M., Romanenkova N.I., Rozaeva N.R., Dzhaparidze N.I., Novikova N.A., Zverev V.V., Golitsyna L.N., Lukashev A.N. Cases of Acute Flaccid Paralysis Associated with Coxsackievirus A2: Findings of a 20-Year Surveillance in the Russian Federation // *Microorganisms*. 2022. Vol. 10, № 1. P. 112. Q2, IF Scopus = 4.782
7. **Vakulenko Y.**, Orlov A., Lukashev A. Patterns and temporal dynamics of natural recombination in noroviruses // *Viruses* 2023. Vol. 15, № 2. P. 372. Q1, IF Scopus = 5.712

### ***Апробация работы***

Апробация основных результатов диссертации проведена в 2022 году в рамках защиты научно-квалификационной работы, подготовленной соискателем при обучении в аспирантуре биологического факультета МГУ. Результаты работы были представлены соискателем на российских и международных конференциях «Europic», 2018, Эгмонд-ан-Зе (устный доклад); «14th International Conference on Molecular Epidemiology and Evolutionary Genetics of Infectious Diseases», 2018, Ситжес (устный доклад); «II Student Conference Life Sciences in the 21st Century: Looking into the Future (SCLS 2019)», 2019, Москва (устный доклад); «Студенческая конференция», 2021, Москва (стендовый доклад). Основные результаты диссертации были опубликованы в 5 статьях в научных рецензируемых журналах, входящих в Q1, Q2 и Q4. Публикации по теме диссертации проходили процедуру независимого рецензирования.

### ***Структура и объем НКР***

Работа состоит из введения, обзора литературы, описания материалов и методов, результатов и их обсуждения, выводов и списка цитируемой литературы, содержащего 348 ссылок. Работа изложена на 200 страницах текста, содержит 9 таблиц и 48 рисунков.



# 1 ОБЗОР ЛИТЕРАТУРЫ

## 1.1 Эволюция (+)РНК-вирусов

Популяции (+)РНК-вирусов эволюционируют под действием тех же факторов, что и другие живые организмы, – мутации, рекомбинация, естественный отбор, дрейф генов и миграция [7]. Однако из-за особенностей биологии РНК-вирусов относительная сила действия этих факторов отличается от таковой от организмов с ДНК-геномом.

### 1.1.1. Мутации

РНК-зависимые РНК-полимеразы (РзРп) (+)РНК-вирусов обладают самой высокой частотой мутаций среди полимераз всех живых организмов [8]. Это объясняется тем, что, в отличие от ДНК-полимераз клеточных организмов и некоторых групп вирусов, они не обладают (за редким исключением) 3'–5'-экзонуклеазной активностью, позволяющей исправлять ошибочные нуклеотиды [9]. Определенные экспериментально частоты мутаций (+)РНК-вирусов варьируются от  $10^{-3}$  до  $10^{-6}$  замен на нуклеотид на инфекцию клетки, что приблизительно соответствует одной мутации на геном на один цикл репликации [8,10]. Высокая частота мутаций (+)РНК-вирусов накладывает ограничения на длину их генома, которая в среднем составляет около 10 тысяч нуклеотидов. Коронавирусы выделяются на фоне других (+)РНК-вирусов самым длинным геномом (до 32 тысяч нт) и низкой частотой мутаций –  $1.3 \times 10^{-6}$  [11]. Это возможно благодаря наличию 3'–5'-экзонуклеазного домена (ExoN) в одном в белке репликативного комплекса nsp14, который участвует в коррекции ошибок репликации [12,13].

Важным параметром, описывающим скорость эволюции вирусов на более длительных промежутках времени, чем один цикл репликации, является скорость накопления замен – количество мутаций, которые вирус накапливает в единицу времени. Скорость накопления замен гораздо более изменчива, чем частота мутаций. Она может существенно различаться как внутри генома, так и между разными видами вирусов, даже со сходными характери-

стиками полимеразы [14]. Скорости накопления замен у РНК-вирусов находятся в пределах от  $10^{-2}$  до  $10^{-5}$  замен/сайт/год (з/с/г) и превышают таковые у организмов с дцДНК геномами [15]. Оценка скорости накопления замен зависит от временного промежутка, на котором рассматривается молекулярная эволюция вируса. Например, при анализе давно дивергировавших вирусов наличие множественных замен в одной позиции может приводить к недооценке скорости накопления замен [16].

Высокая частота мутаций у РНК-вирусов приводит к тому, что они могут быстро накапливать спектр варибельных геномов внутри хозяина. Одной из определяющих концепций в краткосрочной эволюции РНК-вирусов являются квазивиды. Изначально теория квазивидов была разработана для объяснения эволюции биологических макромолекул, которые самореплицируются с большим количеством ошибок [17], а потом применена к РНК-вирусам [18]. Концепция квазивида предполагает, что популяция вирусов представляет собой набор близкородственных геномов, отличающихся друг от друга на несколько мутаций, или спектра мутантов, с доминирующей «главной» последовательностью, имеющей максимальную приспособленность, или успешность репликации. Основная идея концепции заключается в том, что частота мутаций в РНК-вирусах так высока, что частота варианта зависит не только от его собственной скорости репликации, но и от вероятности того, что он может появиться в результате мутации из других вариантов в популяции, которые с ним связаны в пространстве последовательностей. Такая взаимосвязанность вирусных геномов означает, что спектр мутантов эволюционирует как единое целое, и естественный отбор действует не на отдельные варианты вируса, а на популяцию вирусных геномов в целом. Поэтому квазивид эволюционирует, максимизируя свою среднюю приспособленность, а не приспособленность отдельных вариантов [19]. Одним из следствий такой групповой формы отбора является то, что вирусы с низкой скоростью репликации могут иметь высокую частоту в популяции, если имеют мутационные связи с более приспособленными вариантами [20].

Вопрос о том, насколько хорошо концепция квазивида описывает эволюцию РНК-вирусов, остается открытым. Первые экспериментальные свидетельства того, что вирусы эволюционируют как квазивиды, подвергаются критике [21]. К ним относятся данные о существовании спектра мутантов у бактериофага Q $\beta$ , а также наблюдение того, что вариант вируса везикулярного стоматита с высокой приспособленностью при изначально низкой доле в популяции подавлялся вариантом с более низкой приспособленностью [19]. Однако эти явления можно объяснить не только квазивидами. Образование спектра мутантов у бактериофага Q $\beta$  может являться следствием высокой частоты мутаций. Редкие варианты вирусов с более высокой приспособленностью могут теряться в популяциях небольшого размера из-за генетического дрейфа. Позже появились исследования эволюции вирусов *in vitro*, которые показывают, что вирусные популяции могут претерпевать групповую форму естественного отбора [22–25]. Тем не менее, всё ещё неясно, применима ли концепция квазивидов к РНК-вирусам, циркулирующим в природе [26].

#### 1.1.2. Естественный отбор и генетический дрейф

Мутации оказывают влияние на приспособленность вируса. Большинство случайных мутаций РНК-вирусов вредные или смертельные. Мутации, которые не влияют или оказывают незначительное влияние на приспособленность вирусной популяции, называют нейтральными или почти нейтральными (вопрос возможности истинно нейтральных мутаций выходит за пределы данной работы). Мутации, дающие вирусам адаптивное преимущество, относительно редки [27–30].

В результате действия естественного отбора в популяции увеличивается число особей, обладающих более высокой приспособленностью к условиям среды. Положительный отбор приводит к закреплению в популяции полезных мутаций, а особи, несущие вредные мутации, удаляются из популяции под действием отрицательного отбора. Вероятность того, что аллель, дающий адаптивное преимущество, закрепится в популяции зависит от его

частоты, влияния на приспособленность и эффективного размера популяции. Естественный отбор имеет направленное действие на эволюцию вирусных популяций. Генетический дрейф, напротив, приводит к случайному изменению частот аллелей в популяции [7].

Вопрос об относительном действии естественного отбора и генетического дрейфа на популяции вирусов имеет и фундаментальное, и практическое значение. В частности, от интенсивности действия отбора при взаимодействии вируса с хозяином зависит то, насколько быстро вирус может адаптироваться к противовирусным препаратам или менее восприимчивым к нему хозяевам. Для оценки относительной интенсивности отбора и дрейфа в вирусной популяции, реплицирующейся в анализируемой системе, необходимо определить её размер. Это связано с тем, что действие естественного отбора наиболее выражено, когда размер популяции велик. Воздействие генетического дрейфа, напротив, наиболее сильно проявляется в популяциях небольшого размера, в частности, при резком сокращении численности популяции (эффект бутылочного горлышка) [7]. Оценивают не реальный размер популяции, а так называемый эффективный размер популяции, то есть размер идеальной популяции, на которую дрейф действует с такой же скоростью, как на изучаемую популяцию. Он эквивалентен числу особей популяции, которые участвуют в воспроизводстве последующего поколения. Эффективный размер популяции чувствителен к колебаниям размера реальных популяций и, как правило, ниже, чем реальный [31].

Численность вирусных популяций претерпевает изменения на разных этапах цикла инфекции, поэтому соотношение эффекта естественного отбора и генетического дрейфа также меняется. В целом, для РНК-вирусов характерен большой размер популяций: число вирусных частиц может достигать  $10^{12}$  в одном организме [21,32]. Это связано с высокой эффективностью их репликации. Количество копий вируса, которые может воспроизвести одна вирусная частица, и время одного цикла репродукции отличается даже у вирусов, принадлежащих одному роду. (+)РНК-вирусы эукариот в среднем продуци-

руют от  $10^4$  до  $10^{10}$  вирусных частиц на зараженную клетку в ходе одного цикла репродукции, который может длиться от нескольких часов до нескольких дней [33]. Вирусные популяции проходят через бутылочные горлышки на разных стадиях своего жизненного цикла – как при передаче от одного хозяина к другому, так и внутри зараженного хозяина. В частности, такие факторы, как ответ иммунной системы хозяина, ограниченная доступность восприимчивых к вирусу клеток или рецепторов на их поверхности, приводят к уменьшению количества вирусов, участвующих в дальнейшем размножении. Также для некоторых РНК-вирусов показан процесс исключения суперинфекции. При этом процессе вирус, попавший в клетку, включает молекулярные механизмы, которые ограничивают попадание новых вирионов в клетку, что также сокращает эффективный размер популяции. Число вирусных геномов, которые могут попасть в клетку, или множественность инфекции клетки, определяет узость бутылочного горлышка и возможность рекомбинации. Множественность инфекции отличается у разных вирусов и скорее всего является их адаптивной стратегией [34].

### 1.1.3. Рекомбинация

Рекомбинацию в РНК-содержащих вирусах можно определить как обмен генетическим материалом, как минимум, между двумя отдельными вирусными геномами. Для рекомбинации необходимо, чтобы два или более вируса инфицировали одну клетку, и после репликации образовался жизнеспособный (способный к репликации) гибридный геном. Рекомбинационные события были обнаружены у вирусов с РНК-геномом как положительной [35], так и отрицательной полярности [36], однако у последней группы рекомбинация встречается крайне редко [37]. Также описаны случаи приобретения РНК-вирусами последовательностей хозяина, что влияет на их функциональность [38]. Благодаря тому, что рекомбинация приводит к более обширным генетическим изменениям, чем мутации, она может привести к быстрым изменениям фенотипа вируса, включая уклонение от иммунного ответа или

противовирусной терапии, изменение тропизма к клеткам и изменение патогенности [35].

Рекомбинацию разделяют на три типа на основании природы участвующих в ней молекул РНК и сайтов, где она происходит: гомологичная, абберрантная гомологичная рекомбинация (этот термин был предложен недавно) и негомологичная рекомбинация [38]. Гомологичная рекомбинация происходит между двумя близкородственными молекулами РНК, а кроссинговеры происходят в местах, идеально совпадающих между двумя РНК, так что рекомбинантные РНК сохраняют точную последовательность и структурную организацию родительских молекул РНК. Большинство рекомбинантных событий относятся к этому типу. Абберрантная гомологичная рекомбинация происходит между гомологичными молекулами РНК, однако кроссинговеры происходят не в совпадающих сайтах, хотя обычно это близкие участки на родительских РНК. Негомологичная рекомбинация происходит между последовательностями, не имеющими гомологию.

#### *Модели рекомбинации*

На сегодняшний день две модели рекомбинации РНК-вирусов получили широкое признание: репликативная модель «смена матрицы» и нерепликативная модель «разрыв-лигирование». Модель «смена матрицы» была изначально предложена для полиовирусов [39], а потом была изучена и на других системах [40]. Считается, что в большинстве случаев рекомбинация у РНК-вирусов происходит с помощью этого механизма. Модель включает следующие этапы: 1) остановка репликации при синтезе (-)цепи РНК; остановка возможна при встрече полимеразы с элементом вторичной структуры РНК-вируса, или же при вставке неправильного нуклеотида; 2) диссоциация полимеразного комплекса с первой молекулой РНК; 3) опознавание репликационным комплексом гомологичного участка на другой молекуле РНК; 4) продолжение репликации [41].

Альтернативная модель нерепликативной рекомбинации «разрыв-лигирование» предполагает, что рекомбинация может проходить в результате разрыва и лигирования молекул РНК. В отличие от репликативной модели рекомбинант состоит из фрагментов родительских молекул. Впервые этот механизм рекомбинации был предложен для бактериофага Q $\beta$  [42], а потом подтвержден для полиовирусов [43], вируса гепатита С [44] и вируса диареи крупного рогатого скота [45]. Также в отличие от репликативной рекомбинации, нерепликативная рекомбинация не требует участия вирусных белков и опосредуется клеточными факторами [46].

#### 1.1.4. Миграция

Под миграцией, или потоком генов, в популяционной генетике понимают перенос аллелей генов из одной популяции в другую [7]. Поток генов является одним из механизмов увеличения генетического разнообразия в популяции. При изучении эволюции вирусов миграцию можно рассматривать на макроскопическом уровне как передачу вирусов между хозяевами или между разными видами хозяев, а также на микроскопическом уровне – при изучении распределения вируса в тканях зараженного индивидуума.

Вирусы, реплицирующиеся в разных тканях одного хозяина, представляют собой субпопуляции. Поток генов между такими субпопуляциями ограничен, но может осуществляться за счет кровотока хозяина. Экспериментально показано, что разные типы клеток налагают разные селективные ограничения, определяющие степень и специфичность адаптации вирусов [47–49]. Влияние неравномерного распределения вирусной популяции в организме на её приспособленность изучалось гораздо меньше, чем другие факторы эволюции. В экспериментах с вирусом везикулярного стоматита, имитирующих *in vitro* его миграцию между разными типами клеток, показана положительная корреляция между скоростью миграции и величиной средней приспособленности вирусной популяции [50]. Увеличение скорости миграции между разными типами клеток приводило к отбору вирусов, которые

эффективно реплицировались во всех типах клеток. В отсутствии миграции адаптация вируса везикулярного стоматита становилась тканеспецифичной и генетическое разнообразие популяции уменьшалось [51]. Таким образом, для адаптации к специфическому типу клеток миграция должна быть достаточно низкой по сравнению с силой естественного отбора.

В связи с развитием методов филогенетики появилась возможность выводить историю географических перемещений хозяев вирусов, используя их геномные последовательности [52–54], но, как правило, эти вопросы изучаются в контексте молекулярной эпидемиологии, а не популяционной генетики [55,56].

Таким образом, механизмы эволюции, универсальные для всех форм жизни, имеют специфические особенности у РНК-вирусов. При этом в большинстве работ уровень изучения проблемы был ограничен организмом хозяина, популяцией вирусов и т.п. В последние годы распространение недорогого и высокопроизводительного определения нуклеотидной последовательности вирусных геномов привело к экспоненциальному росту объема доступных данных и сделало возможным изучение глобальных закономерностей эволюции для наиболее изученных вирусов.

## **1.2 Общая характеристика калицивирусов**

### **1.2.1 Классификация**

Семейство *Caliciviridae* включает небольшие безоболочечные вирусы (от 27 до 40 нм) с линейным одноцепочечным РНК-геномом положительной полярности относительно мРНК и икосаэдрическим капсидом, имеющим чашеобразные углубления (лат. *calix* - чаша). Семейство включает одиннадцать родов, из которых представители семи родов заражают млекопитающих (*Lagovirus*, *Norovirus*, *Nebovirus*, *Recovirus*, *Sapovirus*, *Valovirus* и *Vesivirus*), представители двух родов заражают птиц (*Bavovirus*, *Nacovirus*), представители двух родов заражают рыб (*Minovirus*, *Salovirus*). Внутри каждого рода выделяют один или более видов (Таблица 1). Спектр хозяев калицивирусов



широк: в последние годы были обнаружены новые калицивирусы у гусей, желтоперых морских лещей, большой зеленой змеи, арктических миног, лягушек и различных австралийских птиц [57–59].

Таблица 1. Классификация вирусов семейства *Caliciviridae*, включающая роды и виды

Род	Виды
<i>Norovirus</i> (NoV)	<i>Norwalk virus</i> (NV)
<i>Sapovirus</i> (SaV)	<i>Sapporo virus</i> (SV)
<i>Lagovirus</i> (LaV)	<i>Rabbit hemorrhagic disease virus</i> (RHDV) <i>European brown hare syndrome virus</i> (EBHSV)
<i>Vesivirus</i> (VeV)	<i>Vesicular exanthema of swine virus</i> (VESV) <i>Feline calicivirus</i> (FCV) <i>San Miguel sea lion virus</i> (SMSV)
<i>Nebovirus</i> (NeV)	<i>Newbury-1 virus</i> (NBV)
<i>Recovirus</i> (ReV)	<i>Tulane virus</i>
<i>Valovirus</i> (VaV)	<i>St Valérian virus</i>
<i>Bavovirus</i> (BaV)	<i>Bayern virus</i>
<i>Nacovirus</i> (NaV)	<i>Novel avian calicivirus</i>
<i>Salovirus</i> (SaV)	<i>Atlantic salmon calicivirus</i> <i>Minovirus</i> (MiV) <i>Fathead minnow calicivirus</i>

## Норовирусы

Род *Norovirus* получил название от города Норфолк в штате Огайо, США, где в 1968 году произошла вспышка острого гастроэнтерита, вызванного вирусом Norwalk [60]. Род включает один вид – *Norwalk virus*, который объединяет генетически разнообразные вирусы, заражающие широкий спектр видов млекопитающих. Человеческие норовирусы являются главной причиной острого гастроэнтерита во всех возрастных группах [61].

В 2013 году была предложена классификация вирусов внутри вида, основанная на филогенетическом группировании аминокислотных последовательностей главного капсидного белка VP1 [62]. Геном норовирусов кодирует три открытых рамки считывания – ORF1 (неструктурные белки), ORF2 (капсидный белок VP1) и ORF3 (капсидный белок VP2). Для того, чтобы учесть частую рекомбинацию вирусов в месте перекрывания ORF1 и ORF2,

была введена двойная номенклатура, основанная на полной последовательности ORF2 и частичной последовательности РНК-зависимой РНК полимеразы (РзРп), кодируемой ORF1 [62]. Согласно предложенной номенклатуре по капсидному белку норовирусы разделяют на десять геногрупп (GI-GX), которые в свою очередь подразделяются на 49 генотипов. Также обсуждается две «предполагаемые» (не окончательно утвержденные) геногруппы (GNA1 и GNA2). По частичной последовательности РзРп выделяют две Р-группы, 60 общепринятых и еще 14 предполагаемых Р-типов. Генотип норовирусной последовательности обозначается как «геногруппа.генотип[Р-группа.Р-тип]», например, GII.3[GII.P12] [63].

Вирусы, принадлежащие геногруппам GI, GII, GIV, GVIII и GIX, заражают человека и вызывают острый гастроэнтерит [63,64]. Также норовирусы были выявлены у свиней (GII), скота (GIII), мышей и крыс (GV), собак (GVI) и летучих мышей (GX). Вирусы из двух предполагаемых геногрупп были выделены из морских свиней (GNA1) и морского льва (GNA2) [65].

#### Саповирусы

Вирусы рода *Sapovirus* вызывают острый гастроэнтерит у человека и животных. Саповирусная инфекция имеет важное медицинское значение, так как саповирусы вызывают гастроэнтерит у людей всех возрастов по всему миру, причем как в виде вспышек, так и спорадических случаев. Клинические симптомы саповирусного гастроэнтерита практически неотличимы от гастроэнтерита, вызванного норовирусами. Саповирусы были открыты в Великобритании в 1976 году с помощью электронной микроскопии образцов стула людей с диареей [66], а прототипный штамм был выделен во время вспышки городе Саппоро в Японии в 1982 году [67].

Род *Sapovirus* включает единственный вид – *Sapporo virus*, однако саповирусы очень разнообразны генетически. В настоящее время внутри вида выделяют 19 геногрупп на основе последовательности участка, кодирующего главный белок капсида VP1 [68,69]. При построении распределения попарных генетических расстояний между последовательностями VP1 выделяется

три четких пика, соответствующих сравнениям вирусов, принадлежащих разным геногруппам, генотипам и штаммам [68]. Вирусы геногрупп GI, GII, GIV и GV заражают человека и подразделяются на по крайней мере 17 генотипов. Вирусы других геногрупп могут заражать свиней (GIII, GV-GXI), морских львов (GV), норок (GXII), собак (GXIII), летучих мышей (GXIV, GXVI-GXIX) и крыс (GXV) [70].

### Лаговирuсы

Род *Lagovirus* включает вирусы, вызывающие серьезные заболевания у европейских кроликов (*Oryctolagus cuniculus*) и некоторых видов зайцев (*Lepus spp.*). Первые эпидемии, вызванные лаговирuсами, произошли в конце 1970-х-начале 1980-х. Эпидемия синдрома зайца-русака привела к смертям в популяции зайца-русака в Швеции [71], а в 1984 сходное заболевание – геморрагическая болезнь кроликов – было выявлено у кроликов на фермах в Китае и затем распространилось по всему миру [72]. Геморрагическая болезнь кроликов – крайне заразное заболевание, протекает молниеносно, смертельно для кроликов возрастом более 8 недель. Оба заболевания характеризуются острым некротическим гепатитом и кровоизлияниями во многие органы, в частности, в легкие, сердце и почки. Вирус геморрагической болезни кроликов (rabbit hemorrhagic disease virus, RHDV) и Вирус синдрома зайца-русака (European brown hare syndrome virus, EBHSV), вызывающие эти заболевания, были открыты в конце 1980-х годов. Позже в Италии был открыт RCV (rabbit calicivirus) – непатогенный вирус, близкий к RHDV [73]. В дальнейшем штаммы RCV были найдены в других странах Европы и Австралии у диких животных. В 2010 году во Франции описали новый более патогенный вариант RHDV – RHDV2/b, который в дальнейшем стал вытеснять старые варианты RHDV. Также были описаны непатогенные штаммы лаговирuсов, заражающих зайцев, которым дали название HaCV (от англ. «hare caliciviruses») [74].

Международный комитет по таксономии вирусов (ICTV) выделяет в роде *Lagovirus* два вида – RHDV и EBHSV [61]. Считалось, что RHDV зара-

жает только кроликов, а EBHSV – зайцев, однако в последнее время появляется всё больше свидетельств того, что вирусы могут проходить кросс-видовой барьер внутри семейства *Leporidae*. Так как RHDV и EBHSV генетически очень близки и нет надежных биологических критериев для демаркации этих видов, была предложена новая номенклатура лаговировусов, основанная на их филогенетических отношениях (Рис. 1). Согласно предложенной таксономии, род *Lagovirus* включает один вид, который в свою очередь объединяет две геногруппы – GI (RHDV и RCV) и GII (EBHSV и HaCV). Внутри геногруппы предлагается выделять генотипы на основе нуклеотидных генетических дистанций (до 15% внутри генотипа) в гене VP1 [75].

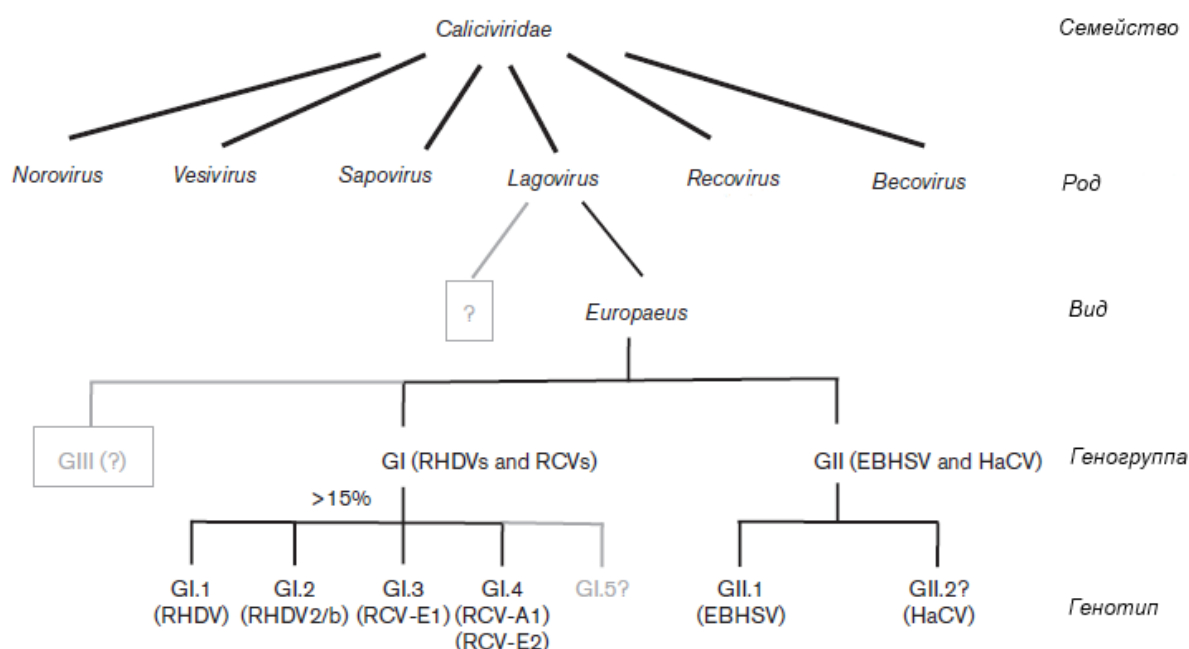


Рис. 1. Предложенная в [75] классификация лаговировусов.

### 1.2.2 Физико-химические свойства и структура вириона

Вирион калицивирусов не имеет оболочки, диаметр составляет примерно 35–40 нм. Внутри капсида норовирусов находятся геномная и субгеномная РНК (сгРНК). Капсид имеет икосаэдральную симметрию (T=3) и состоит из 90 димеров главного структурного белка VP1 – капсомеров (Рис. 2). Капсомеры имеют аркообразную форму и выступают на осях симметрии второго порядка. Капсомеры расположены таким образом, что в местах осей

симметрии 3-го и 5-го порядка образуются 32 углубления, которые видны как чашеобразные структуры на поверхности капсида («calici» происходит от латинского слова «calyx», или «чаша») (Рис. 2). Белок VP1 способен собираться в вирусоподобные частицы без геномной РНК или минорного капсидного белка VP2. Это свойство оказалось очень полезным при изучении калицивирусов, так как их долго не получалось выращивать в культуре клеток. Рекомбинантные вирусоподобные частицы, экспрессированные в бакуловирусной системе, служили заменителями нативных вирионов [76].

Структурные исследования капсидов норовирусов показали, что белок VP1 имеет два домена – S (от англ. “shell”) и P (от англ. “protruding”). Домен S образует внутреннюю часть капсида, которая окружает геном и обеспечивает контакты, образующие икосаэдрическую структуру. Домен P образует аркообразные выступы, которые содержат области контактов димеров (Рис. 2б). P-домен состоит из двух субдоменов – P1 и P2. Субдомен P2 расположен в верхней части «арки». Он содержит потенциальные эпитопы нейтрализующих антител и взаимодействует с антигенами группы крови.

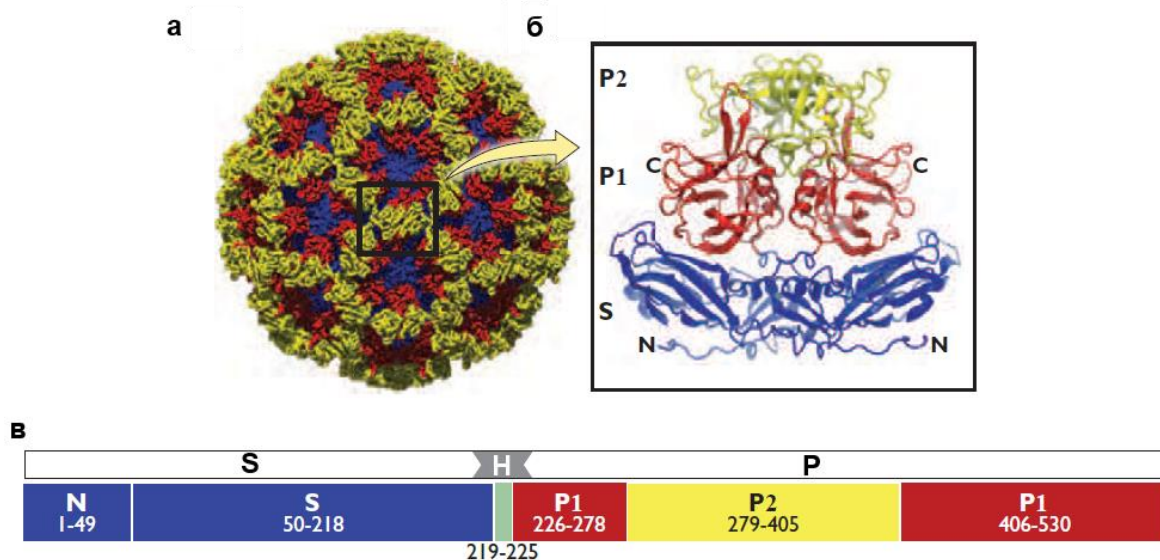


Рис. 2. Организация главного капсидного белка норовирусов VP1. а – модель полного капсида вируса Norwalk, определенная при разрешении 3.4Å. Капсид состоит из 90 димеров VP1 с доменами S, P1 и P2, показанными синим, красным и желтым цветами соответственно. б – трехмерное представление димера VP1, полученное с помощью рентгеноструктурного анализа рекомбинантных ВПЧ вируса Norwalk с разрешением 3.4 Å. Домен P2 находится в верх-

ней части аркообразной структуры и поддерживается двумя «плечами» домена Р1. Указаны N-конец (N) и С-конец белка VP1. в – главный капсидный белок VP1 вируса Norwalk состоит из двух основных доменов – S и Р – соединенных шарнирной (H) областью. Номера аминокислот, входящих в домены, указаны на схеме. Рисунок адаптирован из [76].

Для некоторых представителей семейства установлены физико-химические свойства. Как правило, калицивирусы стабильны в окружающей среде, многие штаммы относительно устойчивы к инаktivации нагреванием и некоторыми химическими веществами (эфир, хлороформ). Кишечные калицивирусы кислотоустойчивы [76].

### 1.2.3 Структура генома

Калицивирусы обладают линейным одноцепочечным геномом положительной полярности длиной от 7.3 до 8.5 тысяч нт. С 5'-концом генома ковалентно связан белок VPg, на 3'-конце генома располагается поли-А последовательность. Калицивирусы обладают короткими консервативными нетранслируемыми областями (НТО): 3'-НТО имеет длину 4–19 нт, а 5'-НТО – 46-108 нт. Небольшой консервативный участок на 5'-конце генома повторяется внутри генома рядом с началом участка, соответствующего субгеномной РНК (сгРНК). Неструктурные белки калицивирусов закодированы ближе к 5'-концу генома, а структурные белки (главный белок капсида VP1 и минорный белок капсида VP2) – ближе к 3'-концу генома в регионе, соответствующем сгРНК. Геном калицивирусов в зависимости от рода содержит две или более основных открытых рамок считывания (ОРС) (Рис. 3). У норовирусов, везивирусов и рековирусов структурный белок VP1 кодирует отдельная ОРС (ORF2), а у саповирусов, лаговирусов, небовирусов и других родов VP1 является продолжением неструктурного полипротеина в ORF1. У последней группы родов белок VP1 может транслироваться как в составе полипротеина, так и с сгРНК [69]. Все калицивирусы имеют относительно небольшую ОРС вблизи 3'-конца, кодирующую минорный структурный белок VP2. VP2 различается среди калицивирусов по размеру (от 12000 до 29000 Да) и по после-

довательности. В геномах мышиноного норовируса (MNV) была найдена ORF4, которая кодирует фактор вирулентности VF1 и располагается внутри ORF2 [77]. Также наличие ORF3, эквивалентной ORF4 MNV, было предположено для саповирусов [78].

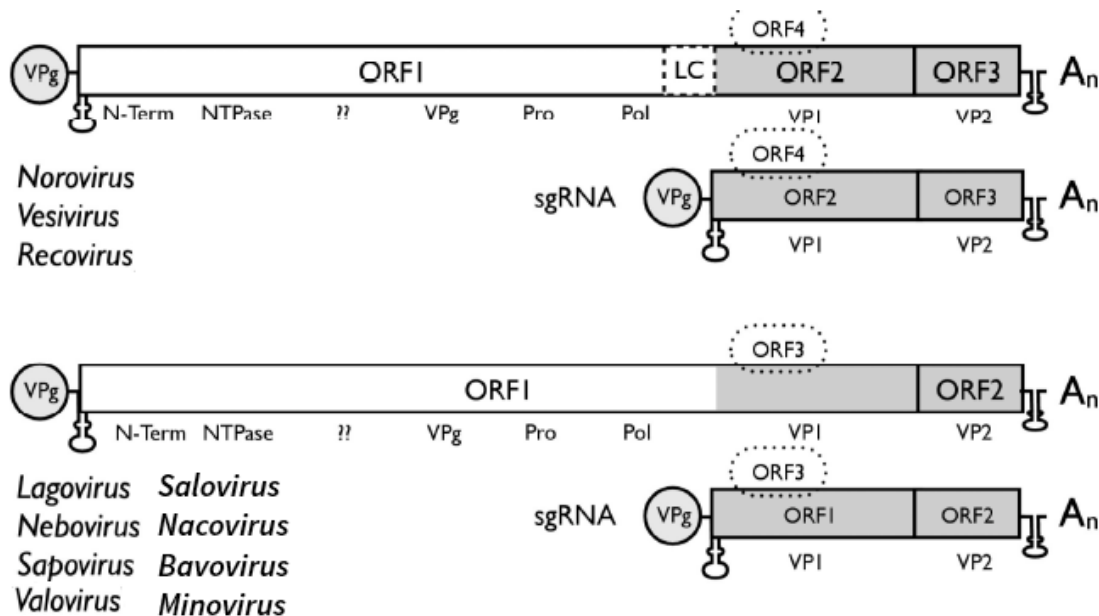


Рис. 3. Схематическая организация генома вирусов разных родов семейства *Caliciviridae*. Рисунок из [79] с небольшими изменениями.

#### 1.2.4 Репликация

Калицивирусы связываются с рецепторами на поверхности хозяйской клетки, после чего вирус проникает в клетку. Рецепторами для калицивирусов могут служить различные молекулы, включая белки, углеводы и гликолипиды. Так, например, показано, что углеводы, входящие в состав антигенов группы крови, служат рецепторами для многих штаммов калицивирусов. Единственным известным белковым рецептором семейства является иммуноглобулин-подобный мембранный белок JAM-1 (junction adhesion molecule-1), с которым связывается калицивирус кошек (FCV). FCV использует клатрин-зависимый эндоцитоз для проникновения в клетки, в то время как для MNV известно, что он не использует ни клатрин-, ни кавеолин-зависимый эндоцитоз.

После проникновения в клетку, геном вируса освобождается и попадает в цитоплазму, где происходит его трансляция. 5'-конец геномной РНК

(гРНК) и субгеномной РНК (сгРНК) калицивирусов ковалентно связан с белком VPg, который служит для рекрутирования клеточных факторов трансляции и рибосом. Помимо калицивирусов, белок VPg есть также у астровирусов и пикорнавирусов. VPg калицивирусов больше похож на таковой у астровирусов [80]. Факторы трансляции, связывающиеся с VPg, отличаются у разных калицивирусов. У FCV VPg функционирует идентично кэпу клеточных мРНК и связывается с eIF4E комплекса eIF4F. Для VPg норовирусов существенным для трансляции является связывание с eIF4G, а взаимодействие с eIF4E может регулировать ответ хозяина на инфекцию, но необязательно для трансляции вируса [76].

После проникновения вируса в клетку в первую очередь происходит трансляция большого полипротеина, кодируемого ORF1, сопровождающаяся ко- и посттрансляционным разрезанием вирусной протеазой NS6 и образованием промежуточных продуктов и финальных белков. Между родами калицивирусов есть существенные различия в процессинге полипротеина. У MNV хозяйские каспазы участвуют в дополнительном разрезании белка NS1. У везивирусов VP1 экспрессируется в виде белка-прекурсора, содержащего N-терминальный лидерный белок, который в дальнейшем разрезается вирусной протеазой и хозяйскими каспазами. У лаго-, реко- и саповирусов VP1 экспрессируется как с полипротеина, так и с сгРНК, что, по-видимому, необходимо для обеспечения большего синтеза VP1. Белки NS6 и NS7 экспрессируются в виде слитого белка NS6-7 у вези- и саповирусов.

После полного разрезания полипротеина образуется до семи неструктурных белков. Функции отдельных белков не до конца известны, но для части белков они были предсказаны на основе анализа аминокислотных последовательностей или определены с помощью биохимических исследований. NS3 обладает НТФазной и хеликазной активностями, NS5 кодирует VPg и отвечает за инициацию трансляции и привлечение рибосом, NS6 является химотрипсин-подобной протеазой, а для NS7 показана активность РНК-зависимой РНК полимеразы (РзРп) [81].



Вирусные белки и РНК образуют связанные с мембраной репликационные комплексы, в которых репликация вирусной РНК осуществляется вирусной РзРп и происходит через промежуточную стадию дцРНК. дцРНК служит матрицей для синтеза гРНК и сгРНК, связанных с VPg. гРНК синтезируется за счет связывания полимеразы с 3' концом РНК отрицательной полярности. Синтез сгРНК начинается с промотора, который у всех калицивирусов представляет собой шпильку, начинающуюся за 6 нт до стартового сайта сгРНК. С вновь синтезированных гРНК и сгРНК транслируются новые вирусные белки. СгРНК всегда полицистронна: с неё идет трансляция капсидных белков VP1 и VP2. У норовируса быка (GIII) трансляция VP1 (ORF2) может идти за счет дополнительного механизма – терминации и реинициации трансляции ORF1 [82]. В количественном соотношении синтезируется больше сгРНК, чем гРНК, что необходимо для большего синтеза структурных белков.

После продукции достаточного количества структурных белков и вирусной РНК происходит энкапсидация вирусного генома и освобождение вирионов из клетки. Процесс энкапсидации недостаточно изучен, но известно, что у норовирусов и FCV белок VPg взаимодействует с VP1, что, возможно, указывает на процесс отбора связанных с VPg геномов. После энкапсидации вирусной РНК происходит выход вирусных частиц из клетки. Все калицивирусы, способные к репликации в клеточной культуре, индуцируют клеточную смерть: для FCV и MNV известно, что в процесс выхода вовлечена активация каспаз, а также MNV отрицательно регулирует ингибитор апоптоза сурвивин. Неизвестно, есть ли нелитические механизмы выхода вирусных частиц калицивирусов из зараженных клеток, однако персистирующая инфекция в В-лимфоцитах была описана [76].

#### 1.2.5 Рекомбинация калицивирусов

Рекомбинация широко распространена у калицивирусов и является одним из механизмов увеличения их разнообразия.

Первые данные о естественной рекомбинации норовирусов были получены в 1997 году для геногруппы GII на основе анализа генетических последовательностей циркулирующих штаммов [83,84]. С тех пор норовирусы, появившиеся в результате рекомбинации, были обнаружены повсеместно. Первое экспериментальное свидетельство рекомбинации норовирусов было задокументировано у MNV (GV) в месте перекрытия ORF1/ORF2 [85]. В другом экспериментальном исследовании проанализировали MNV у коинфицированных мышей и обнаружили частую рекомбинацию на стыке ORF1/ORF2, а также редкие события рекомбинации в VPg, протеазе и 3'-конце кодирующей области PзPп, а также в S-домене VP1 [86].

Чаще всего точка рекомбинации лежит в высококонсервативной области генома на перекрытии ORF1/ORF2, которое соответствует стыку последовательностей, кодирующих полимеразу и VP1. Большинство рекомбинантов являются результатом внутригеногрупповой рекомбинации и были описаны для GI-V (подробный обзор приведен в [87]). К настоящему моменту известно только четыре события межгеногрупповой рекомбинации: между вирусами из GI.P8 и GII.4 [88], GII и GIV [89], GIV и GVI [90], GIV и GVI.2 [91]. Помимо типичной точки рекомбинации, ряд событий рекомбинации между вирусами из GII был детектирован в области, кодирующей VP1, с точками рекомбинации на стыках доменов S/P1, P1-1/P2 [92], внутри домена P1 [89] и P2 [92,93]. Также есть единичные сообщения о рекомбинации вирусов геногруппы GII внутри ORF1 [94,95] и на стыке ORF2/ORF3 [96–98], а также о рекомбинации вирусов GV внутри ORF3, кодирующей капсидный белок VP2 [99,100].

Практически все случаи рекомбинации норовирусов, описанные в литературе, происходили между вирусами, заражающими один и тот же вид хозяина. Вирусы из некоторых геногрупп (GII, GIII, GIV, GV и GVI) могут заражать несколько видов хозяев, что может быть основой для появления новых рекомбинантных вирусов и увеличения генетического разнообразия норовирусов. Тем не менее, рекомбинация между вирусами, заражающими раз-

ных хозяев, встречается редко. На сегодняшний день описан единственный норовирус кошек, который был возможным рекомбинантом норовируса кошек (GIV) и собак (GVI) [91].

Все события естественной рекомбинации саповирусов, описанные в литературе, происходили в месте стыка участков генома, кодирующих полимеразу и капсидный белок VP1. Впервые рекомбинация была описана в 2004 году между саповирусами человека, принадлежащими геногруппе GII [101]. В дальнейшем появились сообщения о рекомбинации между вирусами других геногрупп. Большинство описанных событий происходило между вирусами, принадлежащими одной и той же геногруппе, а именно GI [102–104] и GII [105–108], заражающим человека, и GIII [109–111], вирусы которой заражают свиней. Также было обнаружено несколько событий рекомбинации, происходивших между саповирусами человека разных геногрупп: GIV и GI, GII и GIV [112], а также между саповирусами свиней, принадлежащих GVII и GXI [113]. И у свиней [68,114], и у людей во время вспышек пищевого гастроэнтерита [115,116] были показаны случаи коинфекции вирусами разных геногрупп.

Как и у других калицивирусов, у лаговировусов рекомбинация наиболее часто происходит между участками генома, кодирующими неструктурные и структурные белки (неструктурные и структурные участки генома соответственно) (Таблица 2). Большинство рекомбинационных событий было показано для геногруппы GI, включающей RHDV и RCV, что связано с ветеринарным значением этих вирусов и большим количеством известных последовательностей генома. Рекомбинация способствовала возникновению нового более патогенного генотипа GI.2 (RHDV2/b). Показано, что первые изоляты GI.2 (RHDV2/b), которые ранее считались нерекомбинантными, возникли в результате рекомбинации между GI.3 (RCV) и GI.2 (RHDV2/b), причем GI.3 был донором неструктурной части, а GI.2 – структурной части генома [117]. Внутри геногрупп GI и GII рекомбинация в месте стыка последовательностей, кодирующих полимеразу и капсидный белок VP1, была показана как

между вирусами, принадлежащими одному генотипу, так и разным. Рекомбинация могла происходить только между патогенными, только между непатогенными или между патогенными и непатогенными генотипами (Таблица 2). Практически любой генотип мог быть донором и неструктурного, и структурного участка. Недавно у кролика в Германии был выделен первый лаговиром, являющийся межгрупповым рекомбинантом. Донором неструктурной области его генома был вирус геногруппы GII.1 (EBHSV), а структурной – GI.2 (RHDV2/b) [118].

Помимо «традиционной» точки рекомбинации, несколько рекомбинантов внутри гена капсидного белка были описаны для GI.1 (RHDV) [119–122]. Также в недавнем исследовании были описаны тройные рекомбинанты, у которых участок, соответствующий белку р16, был близок к непатогенному вирусу GI.4, а остальная неструктурная и структурная области соответствовали GI.1 (RHDV) и GI.2 (RHDV2/b) соответственно [123].

Таблица 2. События рекомбинации в месте соединения участков, кодирующих полимеразу и VP1, описанные в литературе у лаговиров

Генотип по полимеразе	Генотип по VP1	Источник
GI.1 (RHDV)	GI.2 (RHDV2/b)	[118,123–126]
GI.2 (RHDV2/b)	GI.1(RHDV)	[127]
GI.2 (RHDV2/b)	GI.4 (RCV)	[127]
GI.3 (RCV)	GI.2 (RHDV2/b)	[117]
GI.4 (RCV)	GI.4 (RCV)	[128]
GI.4 (RCV)	GI.2 (RHDV2/b)	[123]
GII.2 (HaCV)	GII.2 (HaCV)	[129]
GII.1 (EBHSV)	GI.2 (RHDV2/b)	[118]

### 1.3 Общая характеристика пикорнавирусов

#### 1.3.1 Классификация

Семейство *Picornaviridae* (pico gra – маленькие РНК-содержащие) объединяет небольшие безоболочечные икосаэдрические вирусы с оцРНК-геномом положительной полярности. Семейство *Picornaviridae* было выделе-

но в 1963 году и изначально включало лишь два рода – *Enterovirus* и *Rhinovirus*, однако потом оно постоянно расширялось за счет открытия новых вирусов, и в настоящее время насчитывает 63 рода и 147 видов. Большинство известных пикорнавирусов заражают млекопитающих и птиц, недавно были обнаружены пикорнавирусы у рептилий, земноводных и рыб [130].

Энтеровирусы человека относятся к роду *Enterovirus*. Изначально по патогенным свойствам энтеровирусы подразделяли на 4 группы: вирусы полиомиелита, Коксаки А, Коксаки В и ЕСНО (Enteric Cytopathogenic Human Orphan), которые в свою очередь разделялись на серотипы – группы, между которыми практически отсутствует перекрестная серологическая реакция в реакции нейтрализации инфекционности.

Современная классификация энтеровирусов основана преимущественно на генетических свойствах и филогенетических взаимоотношениях вирусов [131]. Род *Enterovirus* на основании филогенетических характеристик разделен на 15 видов: *Enterovirus A-L* и *Rhinovirus A-C* (Таблица 3) [132]. Виды *Enterovirus A-D* и *Rhinovirus A-C* способны заражать человека. Раньше для обозначения видов *A-D* использовалось название Human enterovirus, однако оно больше не применяется, так как представители этих видов регулярно обнаруживаются и у приматов. Виды энтеровирусов подразделяются на более 300 серотипов (типов). Так как в настоящее время при определении новых типов обычно используются только молекулярные критерии, рекомендуется пользоваться термином «тип», а не «серотип» [133]. Исторически некоторые типы называют коксакивирусами А и В (например, КВА2/СВА2 и КВБ3/СВВ3) и эховирусами (например, Е11).

Таблица 3. Таксономическая структура рода *Enterovirus* [132].

Вид *	Серотипы **
<i>Enterovirus A</i> (25)	Коксакивирусы CVA2 – CVA8, CVA10, CVA12, CVA14, CVA16 Энтеровирусы EV-A71, EV-A76, EV-A89 – EV-92, EV-A114, EV-A119, EV-A120, EV-A121; SV19, SV43, SV46 (энтеровирусы обезьян); VA13 (энтеровирус бабуинов)
<i>Enterovirus B</i> (63)	Коксакивирус CVA9

	Коксакивирусы CVB1-В6 Эховирусы E1 – E7, E9, E11 – E21, E24 – E27, E29 – E33 Энтеровирусы EV-B69, EV-B73 – EV-75, EV-B77, EV-B78 – EV-88, EV-B93, EV-B97 – EV-101, EV-B106, EV-B107, EV-B110 – EV-B113, SA5 (энтеровирус обезьян).
<i>Enterovirus C</i> (23)	Полиовирусы PV1-3 Коксакивирусы CVA1, CVA11, CVA13, CVA17 – CVA22, CVA24 Энтеровирусы EV-C95, EV-C96, EV-C99, EV-C102, EV-C104, EV-C105, EV-C109, EV-C113, EV-C116 – EV-C118
<i>Enterovirus D</i> (5)	Энтеровирусы EV-D68, EV-D70, EV-D94, EV-D111, EV-D120 (энтеровирус горилл)
<i>Enterovirus E</i> (4)	Энтеровирусы E1-E4
<i>Enterovirus F</i> (6)	Энтеровирусы F1-F6
<i>Enterovirus G</i> (16)	Энтеровирусы G1-G16
<i>Enterovirus H</i> (1)	Энтеровирус H1
<i>Enterovirus I</i> (1)	Энтеровирус I1
<i>Enterovirus J</i> (6)	Simian enterovirus 6, Энтеровирусы J103, J108, J112, J115, J121
<i>Enterovirus K</i> (2)	Энтеровирусы K1-2
<i>Enterovirus L</i> (1)	Энтеровирус L1
<i>Rhinovirus A</i> (80)	Риновирусы RV-A1 – RV-A13, RV-A15, RV-A16, RV-A18 – RV-A25, RV-A28 – RV-A34, RV-A36, RV-A38 – RV-A43, RV-A45 – RV-A47, RV-A49 – RV-A51, RV-A53 – RV-A68, RV-A71, RV-A73 – RV-A78, RV-A80 – RV-A82, RV-A85, RV-A88 – RV-A90, RV-A94, RV-A96, RV-A100 – RV-A109
<i>Rhinovirus B</i> (32)	Риновирусы RV-B3 – RV-B6, RV-B14, RV-B17, RV-B26, RV-B27, RV-B35, RV-B37, RV-B42, RV-B48, RV-B52, RV-B69, RV-B70, RV-B72, RV-B79, RV-B83, RV-B84, RV-B86, RV-B91 – RV-B93, RV-B97, RV-B99 – RV-B106
<i>Rhinovirus C</i> (55)	Риновирусы RV-C1 – RV-C55
Пока не выделены в отдельные виды (2)	Энтеровирусы EV-122, EV-123

\*В скобках указано количество типов для вида

\*\*Для некоторых типов видов *Enterovirus A-D*, которые заражают не человека, в скобках указаны хозяева

### 1.3.2 Физико-химические свойства, структура вириона

Вирион энтеровирусов обладает сферической формой, его диаметр составляет около 30 нм. Вирусы не имеют липидной оболочки, и поэтому устойчивы к обработке органическими растворителями. Также энтеровирусы устойчивы к понижению и повышению pH (от 3 до 10). Устойчивость к низ-

ким значениям рН помогает им сохранять инфекционность при воздействии желудочного сока.

Капсид энтеровирусов состоит из 60 блоков, называемых протомерами, и имеет икосаэдральную симметрию (Рис. 4). Протомер содержит по одной копии каждого из четырех структурных белков: VP1, VP2, VP3 и VP4, причем поверхность вириона образована только белками VP1-VP3, а белок VP4 находится на внутренней поверхности капсида и взаимодействует с РНК-геномом вируса [134].

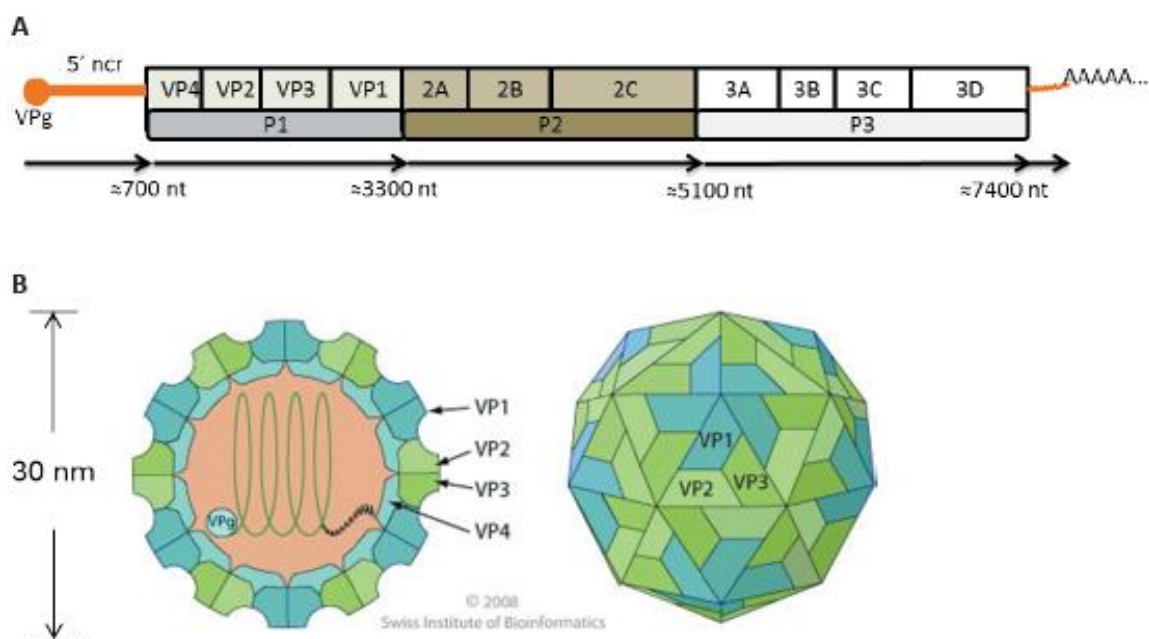


Рис. 4. Структура генома (А) и капсида (В) энтеровирусов [135].

### 1.3.3 Структура генома

Геном энтеровирусов представляет собой одноцепочечную молекулу РНК положительной полярности относительно мРНК длиной от 7200 до 8500 нт (Рис. 4). Геном представлен 5'-НТО, единственной открытой рамкой считывания и 3'-НТО [134].

К 5'-концу вирусной РНК ковалентно присоединен небольшой вирусный белок VPg (virion protein, genome-linked), состоящий из 22 аминокислот [136,137]. Уридилированный VPg служит праймером при репликации РНК. Длина 5'-НТО составляет около 750 нт. 5'-НТО содержит участок длиной около 90 нт, образующий вторичную структуру типа клеверного листа,

которая участвует в регуляции репликации вирусного генома [137], а также внутренний сайт посадки рибосом (internal ribosome entry site, IRES) длиной около 450 нт. IRES – это комплекс элементов вторичной структуры, который обеспечивает инициацию трансляции вирусной РНК по кэп-независимому механизму [138].

Единственная открытая рамка считывания имеет длину около 6600 нт и транслируется как один полипептид, который во время трансляции разрезается вирусными протеазами 2А и 3СD на 11 белков. Многие промежуточные продукты протеолиза могут выполнять определенные функции во время репликации: так, например, 3СD выполняет функцию протеазы наряду с 3С. Полипептид сначала разрезается на три полипептида: Р1, Р2 и Р3, а затем на 4 структурных и 7 неструктурных белков. Область Р1 кодирует структурные белки VP1, VP2, VP3 и VP4. Участок Р2 кодирует следующие белки: цистеиновую протеазу 2А, гидрофобный белок 2В, участвующий в повышении проницаемости внутриклеточных мембран, и гомологичный РНК-хеликазам белок 2С, обладающий АТФ- и ГТФ-азной активностью. 2А осуществляет разрезание полипептида между белками VP1 и 2А, а также выключение клеточной кэп-зависимой трансляции. Участок Р3 кодирует небольшой белок 3А, пептид VPg (3В), протеазу 3С, которая наряду со своим предшественником 3СD разрезает все белки, кроме VP4-VP2 и VP1-2А, и вирусную РНК-зависимую РНК-полимеразу 3D [134].

В отличие от большей части клеточных мРНК, 3'-НТО энтеровирусов имеет длину всего лишь около 50-100 нт и несет элементы вторичной структуры, необходимые для репликации. Показано, что удаление 3'-НТО у полио- и риновирусов не летальна для них [139]. 3'-конец генома имеет поли-А последовательность, которая стабилизирует и защищает вирусный геном от клеточных нуклеаз, а также связывается с различными факторами инициации трансляции клетки-хозяина. Удаление поли-А последовательности ведет к потере инфекционности вирусной РНК [134].



### 1.3.4 Репликация энтеровирусов

Первая стадия цикла репликации – взаимодействие вирусной частицы с рецепторами на поверхности хозяйской клетки. Рецепторами энтеровирусов могут быть белки семейства иммуноглобулинов, интегрины, рецепторы липопротеинов низкой плотности и др. Так, например, полиовирус связывается с иммуноглобулин-подобным рецептором CD155, также известным как рецептор полиовируса (poliovirus receptor, PVR). Белок ICAM-1 (intercellular adhesion molecule-1), принадлежащий семейству иммуноглобулинов, является рецептором для большинства риновирусов А и В. Коксакивирусы связываются с рецепторами CAR (coxsackievirus and adenovirus receptor), ICAM-Фактор ускорения распада комплемента (DAF), или CD55, является рецептором для коксакивирусов и некоторых эховирусов [140]. Долгое время главный рецептор эховирусов был неизвестен, и только в 2019 году показали, что им является неонатальный Fc-рецептор. Это объяснило повышенную восприимчивость младенцев к эховирусной инфекции [141].

Связывание вируса с рецептором приводит к сильным конформационным изменениям капсида и выходу вирусной РНК в цитоплазму, где она служит мРНК и транслируется с помощью клеточной системы трансляции. Далее вирусные протеазы 2A, 3C и 3CD разрезают полипептид на 11 белков. Затем синтезируется (-)РНК цепь, на матрице которой синтезируются новые копии (+)РНК. Репликация осуществляется вирусной РНК-зависимой РНК-полимеразой 3D на мембранных микродоменах при участии некоторых клеточных и вирусных белков [142]. Уридилированный белок VPg (3В), связанный с 5'-концом РНК-вируса во время репликации, играет роль праймера во время синтеза новой цепи РНК. Считается, что белок 3АВ участвует в закреплении VPg на мембранных микродоменах. Репликации способствуют *cis*-действующие репликативные элементы CRE, представляющие собой РНК-шпильки длиной около 60 нт. Полиовирус имеет 4 CRE-элемента, которые находятся в 5'-НТО, открытой рамке считывания, 3'-НТО [143].

При сборке капсида сначала происходит сборка незрелых протомеров из белков VP0 (предшественник VP2 и VP4), VP1, VP3, которые в свою очередь собираются в пентамеры. Пентамеры формируют икосаэдрический капсид. Процессинг незрелого VP0 происходит автокаталитически после связывания РНК с капсидом. Выход вируса из клетки происходит после её разрушения.

Цикл репликации вирусов занимает от 5-10 часов и зависит от многих параметров, включая температуру, рН, тип хозяйской клетки и множественность инфекции [134].

### 1.3.5 Рекомбинация энтеровирусов

Как правило, рекомбинация у энтеровирусов происходит в области генома, кодирующей неструктурные белки. Наиболее часто рекомбинация происходит внутри генов 2А и 3С, на конце гена 2С, а также на границах участков 5'-НТО/VP4, 3С/3D [144,145]. Было показано, что циркулирующие энтеровирусы рекомбинируют внутри видов каждые несколько лет [146,147], поэтому вид энтеровирусов можно рассматривать как популяцию неструктурных генов [148]. В то же время рекомбинация в участках, кодирующих белки капсида, происходит крайне редко и является скорее исключением, чем правилом [149], поэтому структурные гены разных типов ведут себя как отдельные популяции. Таким образом, эволюцию капсид-кодирующих генов можно исследовать без рассмотрения неструктурных участков генома.

### 1.3.6 Патогенность неполиомиелитных энтеровирусов

Энтеровирусы распространены повсеместно. Обычно в течение первого года жизни человек переносит неоднократную инфекцию энтеровирусами. Главным путем передачи энтеровирусной инфекции является фекально-оральный путь, реже респираторный путь. Фекально-оральный путь превалирует в регионах с плохими санитарными условиями, тогда как респираторный путь играет большее значение в более развитых странах. В случае EV-D70 и CVA24, вызывающих острый геморрагический конъюнктивит, переда-

ча инфекции также возможна при контакте с секретом глаза, а при везикулярной экзантеме, вызываемой некоторыми типами – при прямом или непрямом контакте с везикулярной жидкостью [140].

Первичное размножение энтеровирусов происходит в тканях желудочно-кишечного и респираторного трактов, после чего проходит вирусемическая фаза и распространение вирусов во вторичные органы-мишени [140].

Большинство энтеровирусных инфекций проходит бессимптомно, однако энтеровирусы могут вызывать большой спектр разных синдромов (Таблица 4). Один тип может вызывать разные симптомы и заболевания, а сходные клинические проявления могут быть вызваны различными типами. Но есть и заболевания, вызываемые определенными специфическими типами. Так, например, острый геморрагический конъюнктивит обычно связан с CVA24 и EV-D70, а острый вялый паралич вызывают полиовирусы и EV-A71 [140].

Таблица 4. Заболевания и синдромы, вызываемые энтеровирусами [140]

Вирусы	Вызываемые заболевания и синдромы
Полиовирусы PV1 – PV3	Параличи, асептический менингит, фебрильное заболевание
Вирусы Коксаки А CVA1 – CVA24	Герпангина, острый фарингит, асептический менингит, параличи, экзантема, ящуроподобное заболевание (CVA10, CVA16), пневмониты у грудных детей, простуда, гепатит, плевродиния, диарея у новорожденных, острый геморрагический конъюнктивит (CVA24)
Вирусы Коксаки В CVB1 – CVB5	Плевродиния, асептический менингит, параличи (редко), острое системное заболевание у детей, протекающее с миокардитом и менингоэнцефалитом, перикардит, миокардит, ОРВИ, пневмония, сыпь, гепатит, фебрильное заболевание
Эховирусы E1 – E34	Асептический менингит, параличи, энцефалиты, атаксия, синдром Гийена-Барре, экзантема, респираторные заболевания, диарея, эпидемическая миалгия, перикардиты и миокардиты, гепатиты
Энтеровирусы EV-68 – EV-116	Пневмония и бронхолит, острый геморрагический конъюнктивит (EV-D70), менингоэнцефалит (EV-D70, EV-A71), параличи (полиоподобное заболевание) (EV-D70, EV-A71), ящуроподобное заболевание (EV-A71)

Энтеровирусы являются основными возбудителями асептического менингита детей и взрослых в развитых странах и могут быть причиной 85-95%

его случаев [150]. Коксакивирусы В являлись этиологическим агентом в 62% случаев асептического менингита детей младше 3 лет. Заболевание обычно проходит доброкачественно (без осложнений) и разрешается самостоятельно. Также в редких случаях энтеровирусы могут вызывать энцефалит. Наиболее часто встречаются и тяжело протекают неврологические инфекции, вызванные EV-A71 [151].

Инфекция коксакивирусами В считается главной причиной вирусного миокардита. Она чаще всего поражает людей от 20 до 39 лет. Полагают, что коксакивирусы группы В могут быть причиной около 70% случаев миокардита, причем явные симптомы острого вирусного миокардита могут наблюдаться только у половины инфицированных. Однако даже те люди, у которых миокардит проходит бессимптомно, находятся в зоне риска: дисфункция электрической проводимости в сердце, являющаяся следствием вирусного миокардита, может привести к коллапсу и даже смерти при тяжелых физических нагрузках [152].

Энтеровирусы являются причиной респираторных заболеваний примерно в 1,5% случаев в США [140]. Особенно тяжело энтеровирусные инфекции протекают у новорожденных и младенцев. Также было показано, что энтеровирусы могут играть роль в развитии инсулин-зависимого диабета [153].

В данной работе особенное внимание уделяется вирусам типа коксакивирус А2 (CVA2). Чаще всего CVA2 вызывают герпетическую ангину, ящуроподобное заболевание и плеврдинию [140,154–156]. Также были описаны редкие случаи миокардита [157,158] и фульминантного диабета 1-го типа [159], связанные с инфекцией CVA2. CVA2 могут вызывать инфекционные заболевания нервной системы – менингит, энцефалит и миелит [156]. Миелит может проявляться как острый вялый паралич (ОВП), сопровождающийся стойким нарушением активных движений в одной или нескольких конечностях. По клиническим проявлениям без лабораторной диагностики он неотличим от паралитического полиомиелита, вызванного полиовирусом [160].

Обнаружение CVA2 у больных с ОВП, по-видимому, является редким событием [161–163], однако в исследовании, проведенном в Бразилии, CVA2 был одним из ведущих типов неполиомиелитных энтеровирусов, выделенных от больных с ОВП [164]. Таким образом, имелись основания предполагать, что CVA2 отличается от большинства других типов неполиомиелитных энтеровирусов способностью чаще вызывать поражение ЦНС. Среди 7280 случаев ОВП, зарегистрированных в Российской Федерации за 20 лет (2001–2020 гг.) [165], CVA2 был выделен только в пяти случаях. Было выдвинуто предположение о возможном возникновении варианта CVA2 с измененными патогенетическими свойствами. Глава данной работы посвящена проверке этой гипотезы.

### 1.3.7 Методы типирования энтеровирусов

Традиционным методом типирования энтеровирусов является выделение вирусов из клинических образцов в клеточной культуре с последующей реакцией нейтрализации инфекционности вируса с набором семи поливалентных сывороток [166]. Данный метод очень длительный, трудоемкий и дорогостоящий, и с помощью реакции нейтрализации со стандартным набором сывороток можно типировать менее половины известных на сегодняшний день типов.

В настоящее время определение типа производят с помощью секвенирования полного структурного гена VP1 (~900 нт). В 1999 году группой S. Oberste было показано, что последовательность РНК гена VP1 коррелирует с серотипом. Оказалось, что в распределении попарных сходств нуклеотидных и аминокислотных последовательностей всех серотипов есть четкая граница между пиками, соответствующими попарным сходствам между последовательностями разных серотипов и между последовательностями, принадлежащим одному серотипу, – 75 % сходства нуклеотидных последовательностей и 85 % аминокислотных последовательностей (Рис. 5) [167].

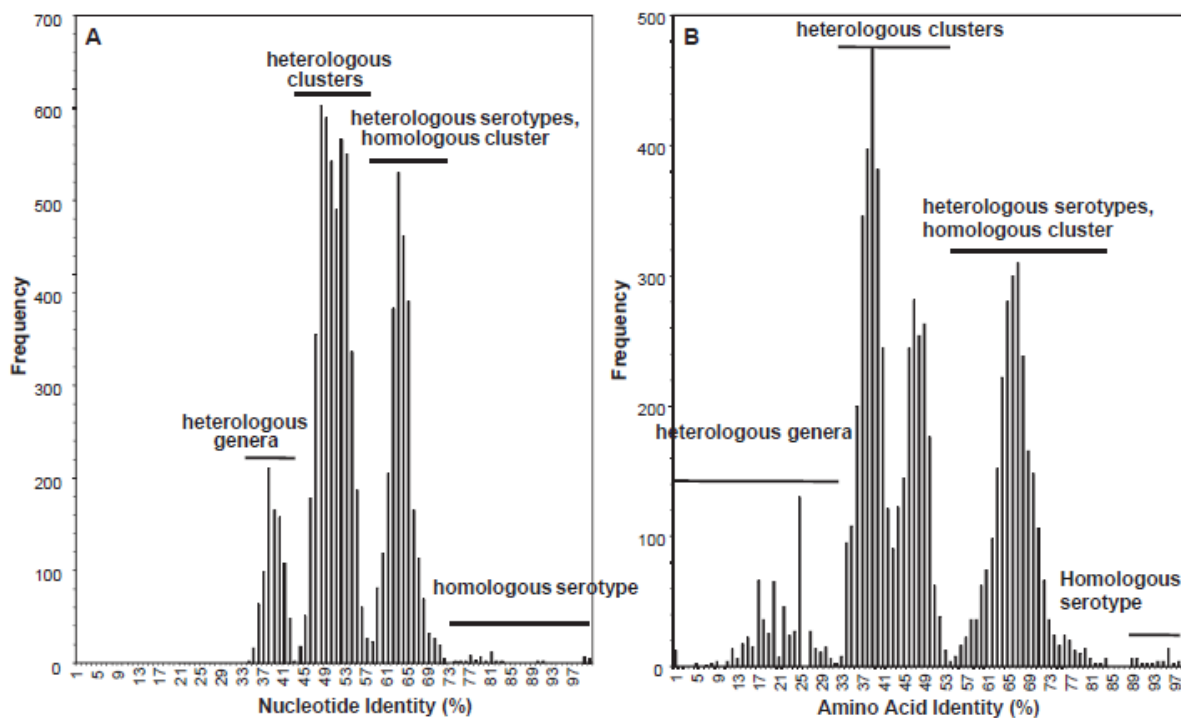


Рис. 5. Распределение частот попарных сходств всех известных на 1999 год (А) нуклеотидных и (В) аминокислотных последовательностей гена VP1 серотипов [168].

Благодаря результатам данной работы стало возможно применение молекулярного типирования энтеровирусов на практике. Считают, что если сходство между нуклеотидными и аминокислотными последовательностями полного гена VP1 двух вирусов более 75 % и 85 % соответственно, то они принадлежат одному типу. Как правило, в большинстве лабораторий молекулярное типирование проводят не по полной последовательности гена VP1, а по его так называемому «типизирующему» фрагменту длиной около 370 нт, который может быть амплифицирован с помощью универсальных праймеров (позиции 2627–2951 по геному штамма Mahoney PV1) [169].

### 1.3.8 Особенности эволюции типов энтеровирусов

Накопление мутаций и рекомбинация играют главную роль в эволюции РНК-вирусов. Геном полиовируса накапливает мутации со скоростью  $10^{-2}$  з/с/г, для других энтеровирусов скорость накопления замен варьирует от  $0.32 \times 10^{-2}$  до  $1.2 \times 10^{-2}$  з/с/г (Таблица 5). Такая высокая скорость накопления замен связана с отсутствием 3'–5'-экзонуклеазной активности вирусной РНК-зависимой РНК-полимеразы и другими факторами [170]. Благодаря вы-

сокой скорости фиксации мутаций возможно исследовать молекулярную эпидемиологию энтеровирусов с высоким разрешением.

Таблица 5. Скорость накопления замен для области VP1 у наиболее распространенных типов энтеровирусов

Тип	Скорость накопления замен $\times 10^{-2}$ , з/с/г	Скорость накопления замен $\times 10^{-2}$ , 95% ДИ, з/с/г	Источник
CV-A4	0.81	[0.60, 1.05]	[171]
CV-A16	0.66	-	[172]
CV-B2	0.53	[0.36, 0.7]	[173]
	0.54	[0.34, 0.70]	[174]
CV-B4	0.57	[0.42, 0.72]	[173]
E-11	0.61	[0.52, 0.70]	[175]
	0.48	[0.36, 0.61]	[176]
E-30	0.70	[0.64, 0.76]	[175]
	0.88	[0.74, 1.04]	[177]
E-6	1.05–1.2*	[0.81, 1.45]	[178]
E-9	0.58	[0.37, 0.81]	[147]
EV-D68	0.49	[0.40, 0.59]	[179]
EV-A71	0.32	-	[180]
	0.36–0.53*	-	[181]
	0.72	[0.62, 0.83]	[176]
	0.42–0.46**	-	[5]

\* - в зависимости от выбранной эволюционной модели

\*\* - в зависимости от генотипа и выбранной эволюционной модели

В ряде работ были описаны некоторые закономерности молекулярной эпидемиологии и эволюции энтеровирусов. Как было сказано ранее, циркулирующие энтеровирусы рекомбинируют внутри видов в среднем каждые несколько лет. Популяции серотипов характеризуются постоянными циклами возникновения, доминирования и исчезновения рекомбинантных форм с определенной периодичностью. Для разных типов периоды отличаются: так, например, период полужизни рекомбинантных форм E9 составляет около 1 года, у E30 смена рекомбинантных форм происходила каждые 3–5 лет, а у

Е11 период полужизни равен 10 годам (Рис. 6) [147,177]. У EV-A71 времена полужизни рекомбинантных форм было разным у различных генотипов. Для генотипов В и С среднее время полужизни рекомбинантной формы составляло около 6 и 10 лет соответственно, однако некоторые самые распространенные рекомбинантные формы циркулировали без рекомбинации в течение нескольких десятилетий [176]. Также для перечисленных типов было характерно увеличение вероятности рекомбинации с ростом количества отличий в последовательностях генов VP1 [147,176,177].

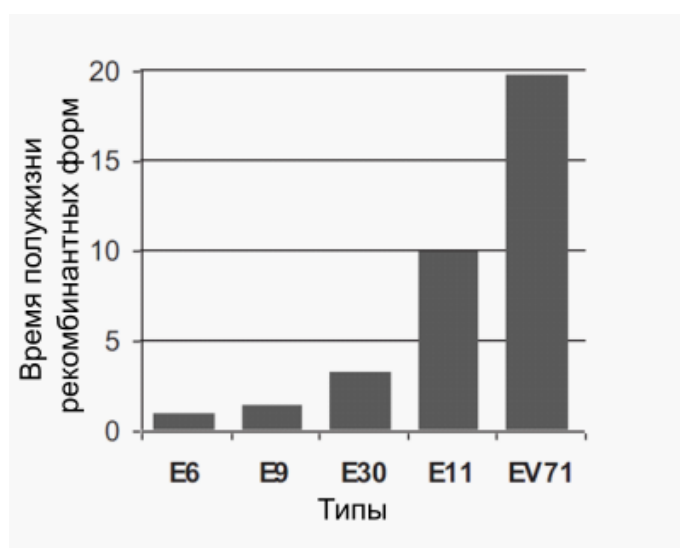


Рис. 6. Время полужизни рекомбинантных форм для некоторых серотипов [178].

Анализ рекомбинации и эволюционной динамики изолятов Е6, циркулировавших в Испании в 2004–2010 годах, показал несколько другую картину: время полужизни рекомбинантных форм составляло всего 0.87 года, а также встречалось довольно много изолятов с разными неструктурными генами, но очень близкими последовательностями VP1. Вероятно, совместная эндемичная циркуляция разных линий Е6 давала больше возможностей для рекомбинации, чем могло бы быть у E30 и EV-A71, для которых характерна эпидемическая циркуляция [178,182].

Анализ эпидемиологии E30 выявил, что в определенный промежуток времени в популяции типа на территории всей Европы преобладали одна–две рекомбинантные формы (Рис. 7) [177]. Для нескольких типов показано, что появление новой рекомбинантной формы предшествовало образованию и



экспансии новых вариантов вируса, а также вспышкам инфекций, вызываемых данным типом [147,176,177]. Неизвестно, однако, было ли появление новой рекомбинантной формы условием возникновения вспышки, или же стохастическим событием с последующей экспансией в популяции хозяина.

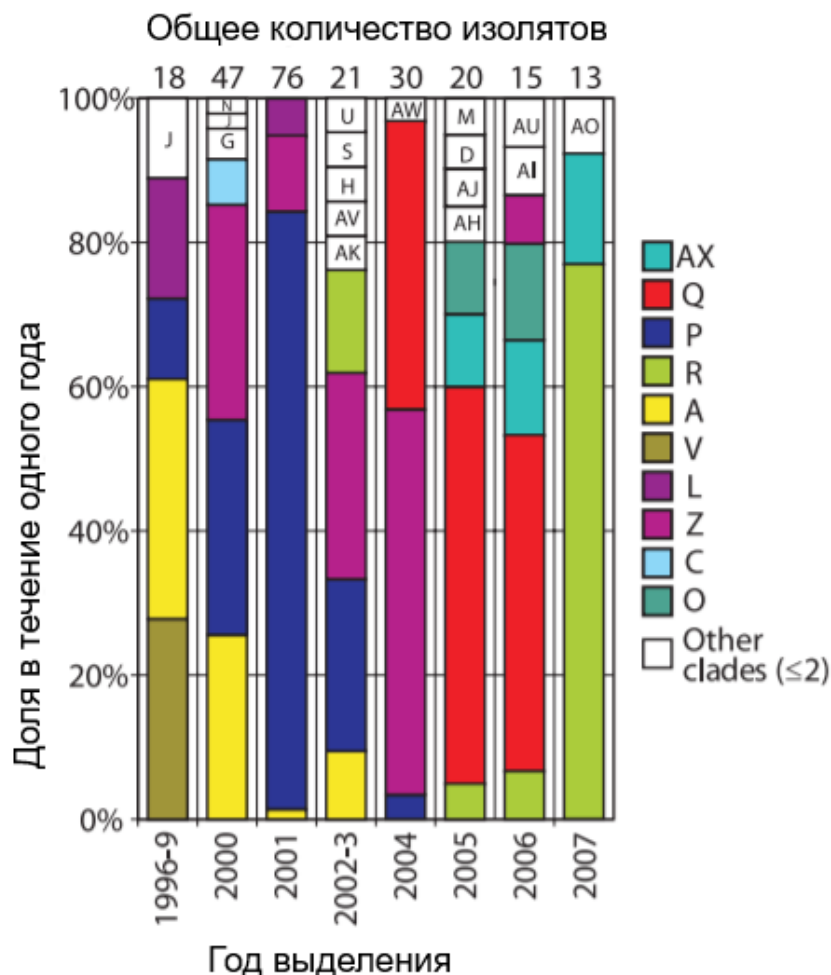


Рис. 7. Частоты разных рекомбинантных форм E30 в Европе за 1996–2007 гг. Клады 3D полимераз покрашены разными цветами [177].

Филогенетические деревья, построенные по VP1, имеют структуру, подобную лестнице: чем позже вирус был выделен, тем дальше он расположен от корня дерева (Рис. 8). Интересно, что, несмотря на высокую скорость накопления мутаций, энтеровирусы, выделенные в течение небольшого промежутка времени (несколько лет), характеризовались небольшим генетическим разнообразием участка, кодирующего VP1. Также наблюдалась смена преобладающих вариантов VP1 в глобальном масштабе [5,175,182,183].

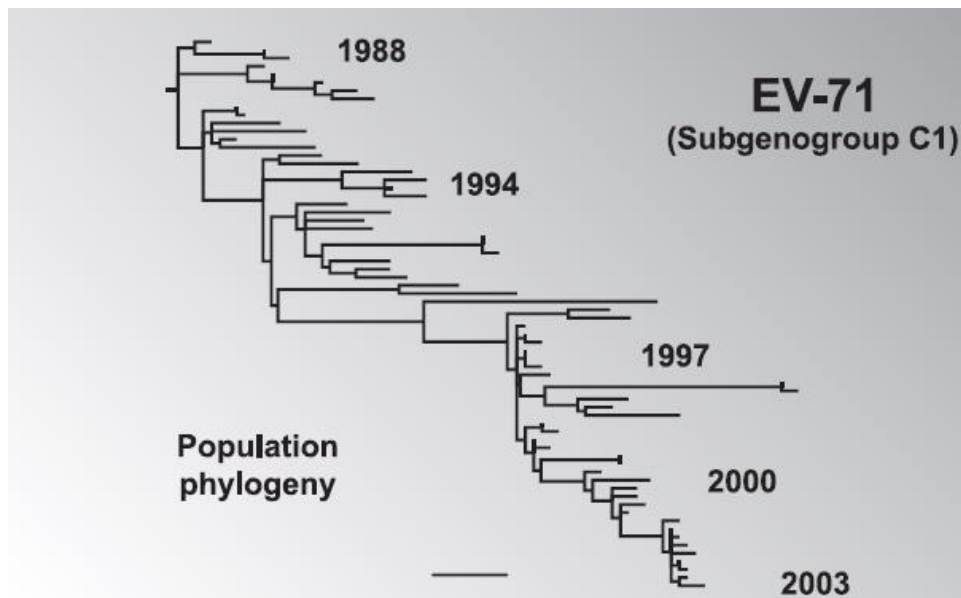


Рис. 8. Филогенетическое дерево субгенотипа C1 EV-A71, построенное по гену VP1 [5].

### 1.3.9 Возможные механизмы происхождения типов энтеровирусов

Применение методов молекулярного типирования энтеровирусов привело к открытию большого числа новых типов. Позднее оказалось, что многие открытые типы на самом деле циркулировали в популяции за несколько десятков лет до их открытия, но не могли быть точно типированы с помощью серологических методов. Таким образом, вопрос о происхождении новых типов остается открытым. В настоящее время предлагают две возможные модели возникновения новых типов: постепенная дивергентная эволюция от общего предка в рамках одного вида хозяина и ускоренная («квантовая») дивергенция, вероятно, за счет межвидовой передачи.

Эволюция линий энтеровирусов может идти за счет плавных изменений последовательности генома или за счет более быстрых, как, например, быстрое увеличение разнообразия последовательностей за счет адаптации к новой окружающей среде. Сравнение последовательностей энтеровирусов внутри типов показало, что отношение частот несинонимичных мутаций к частотам синонимичных по геному очень низкое, что свидетельствует том, что на гены действует отрицательный отбор [184]. В исследованиях полиовирусов и E30 было показано, что все сайты белков капсида подвержены

сильному отрицательному отбору [185,186]. Таким образом, по-видимому, различия между генетическими линиями одного типа являются результатом нейтральных мутаций. Сравнение же последовательностей фрагментов VP1 разных типов показало другие результаты: всего лишь 59 из 124 аминокислот были консервативны для CV-B2 и CV-B4, 17 из 124 аминокислот были полиморфными у обоих типов. Интересно, что 17 из 124 аминокислот были консервативны у одного типа, но полиморфны у другого. Это говорит о том, что в вирусах разных типов отбор мог по-разному действовать на эти сайты, что приводит к различным между двумя типами структурным ограничениям эволюционной лабильности капсида в этих сайтах [187]. Есть мнение, что при возникновении нового типа более вероятна полная перестройка капсида, а не плавные изменения в некоторых сайтах [184].

Очевидные различия между разными типами энтеровирусов можно объяснить адаптацией к новым экологическим нишам. Существование генетической гетерогенности популяций энтеровирусов, достигаемой за счет высокой частоты мутаций, дает им большой потенциал для адаптаций [188]. Так, например, субпопуляция с измененным фенотипом (измененная аффинность к рецепторам, альтернативный рецептор, измененный тропизм к тканям и др.) могла бы случайно занять другую экологическую нишу, где действует другое давление отбора, что могло бы привести к закреплению новых мутаций, повышающих приспособленность. В конечном итоге, вариант вируса с закрепившимися мутациями может быть идентифицирован как новый тип.

В ряде экспериментальных исследований на мышинных моделях было показано, что небольшие изменения последовательности капсида достаточны для изменения различных свойств вируса: специфичности к определенным типам клеток, тропизма к тканям и видам хозяев, цитопатогенности, вирулентности и др. [187]. В то же время экспериментов по смене серологической специфичности энтеровирусов неизвестно.

Одним из примеров эволюции путем адаптивной радиации является эволюция полиовируса. Полиовирусы формируют монофилетическую группу

внутри вида *Enterovirus C* и используют в качестве рецептора CD155, тогда как самые филогенетические близкие типы (CVA13, CVA15, CVA18, CVA20) используют рецептор ICAM-1 (CD54). Полагают, что образование полиовируса произошло за счет перехода от ICAM-1 к новому рецептору (CD155) и последующей адаптации к новой экологической нише [189]. Предполагают также, что другой социально значимый вирус, EV-A71 (прототипный штамм выделен в 1971 году), мог возникнуть путем отделения от CVA16 (прототипный штамм датируется 1951 годом). Эти типы близки как филогенетически, так и по антигенным свойствам. Набор используемых ими рецепторов сильно перекрывается, оба типа являются этиологическими агентами вирусной экзантемы, но только EV-A71 вызывает серьезные неврологические заболевания. Несмотря на близкие биологические свойства, последовательности прототипных штаммов этих типов, кодирующие структурные белки, сильно отличаются (20.7 % в P1, 29 % различий в гене VP1) [190] и подвержены сильному отрицательному отбору внутри типов, причем замены распределены равномерно по геному, а не в нескольких антигенных сайтах. Согласно времени возникновения EV-A71 датируется 1929–1952 годом, однако нельзя исключать того, что современные филогенетические методы могут недооценивать время возникновения. Возможно, в момент возникновения типов скорость накопления замен при адаптационном процессе была такой быстрой, что даже самые первые выделенные последовательности этих типов имели существенные различия [5].

Другим способом возникновения нового типа может быть переход вируса от одного вида хозяина к другому. Вирус везикулярной болезни свиней (SVDV, swine vesicular disease virus) был впервые выделен в Италии в 1966 году и вызывал серьезные вспышки заболевания в Европе и Восточной Азии. Было показано, что он и филогенетически, и по серологическим свойствам очень близок к CVB5. Оба типа в качестве основного рецептора используют CAR (coxsackie-adenovirus receptor), а также могут использовать DAF человека как ко-рецептор. Тем не менее, CVB5 не может заражать клетки свиньи

через DAF и CAR, а вирусы SVDV, выделенные после 1990 года, утратили способность связываться с DAF человека. Сравнительный анализ последовательностей и структур капсидных белков вирусов показал, что при адаптации к новому виду хозяина большинство мутаций затронуло CAR-связывающий сайт белка VP1. При Байесовском филогенетическом анализе этих типов было выявлено, что ветки филогенетического дерева, ведущие к SVDV, имеют самую большую скорость замен по дереву [187].

Наибольшее число случаев межвидовой передачи описано среди вирусов вида *Enterovirus A*. Типы энтеровирусов обезьян этого вида SV-19, SV-43, SV-46, Ва-EV, EV-A92 были впервые выделены в 1950 – 1960-х годах из диких обезьян, часто используемых в биомедицинских исследованиях. Интересно, что все типы филогенетически (по участку VP1) близки к недавно описанным типам человека EV-A76, EV-A89, EV-A90 и EV-A91. Вероятно, имела место неоднократная передача вирусов от человека к обезьянам и наоборот. Также EV-A76, EV-A89, EV-A90 и EV-A91 формируют отдельную группу на филогенетических деревьях, построенных по неструктурным генам и, по-видимому, не могут рекомбинировать с другими энтеровирусами человека, принадлежащими виду А [191]. Связи энтеровирусов человека и обезьян посвящен ряд работ [192–194], но, к сожалению, знания в этой области фрагментарны, и данные вопросы требуют дальнейшего изучения.

Таким образом, исходя из систематического анализа известных филогенетических закономерностей внутри типов энтеровирусов и установления возможных сроков и механизмов их возникновения можно сделать следующие выводы. Вопрос о времени возникновения различных серотипов, известных в настоящее время, остается нерешенным. Благодаря современным молекулярно-биологическим методам и применению Байесовского подхода в филогенетике возможно достаточно точно датировать эволюционные события у вирусов в недавнем прошлом. Для некоторых серотипов показано, что общий предок вирусов, принадлежащих одному серотипу, возник не более 150 лет назад [5,6]. Не ясно, насколько этот результат достоверен и насколько

ко универсален для всех серотипов, и как объем и качество выборки известных последовательностей генома влияет на оценку датирования эволюционных событий вирусов.

## 1.4 Общая характеристика астровирусов

### 1.4.1 Классификация

Семейство *Astroviridae* объединяет безоболочечные вирусы сферической формы с оц(+)РНК-геномом. Астровирусы были открыты в 1975 году с помощью электронной микроскопии образцов стула младенцев, госпитализированных с диареей [195]. Своё название семейство получило из-за звездчатой формы вириона с пятью или шестью «лучами». Семейство включает два рода – *Avastrovirus* и *Mamastrovirus* – вирусы которых заражают птиц и млекопитающих соответственно. По состоянию на 2021 год ICTV выделяет 22 вида астровирусов на основе аминокислотной последовательности капсидных белков, три из которых принадлежат роду *Avastrovirus* [132]. С тех пор были обнаружены новые вирусы, которые пока ещё не классифицированы.

Астровирусы человека вызывают острый гастроэнтерит во всех возрастных группах, а также они были выделены от детей с иммунодефицитом и энцефалитом. Астровирусы человека, открытые до 2008 года, называют «классическими», разделяют на 8 серотипов и относят в виду MAstV1. Позже в образцах стула детей с гастроэнтеритом были найдены новые группы человеческих астровирусов – HAstV-MLB и HAstV-VA/HMO, которые относят к видам MAstV6 и MAstV8–MAstV9 (Рис. 9). Астровирусы имеют также ветеринарное значение: они вызывают энтерит у индеек, гепатит у уток, нефрит у кур и энцефаломиелит у крупного рогатого скота и норок. Также астровирусы были найдены у кошек, собак, морских львов, мышей, землероек, кроликов, овец, оленей и дикобразов [196].

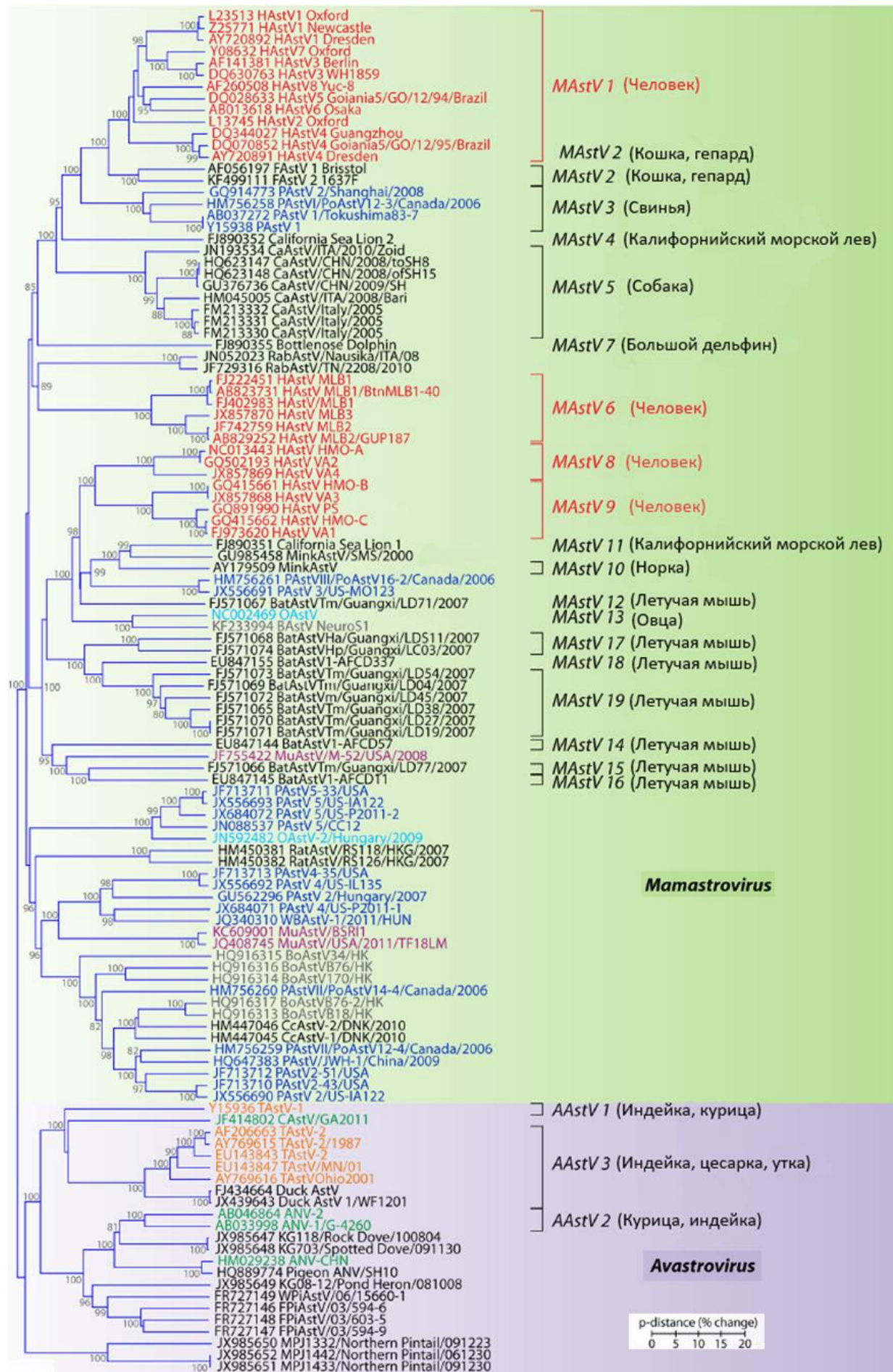


Рис. 9. Филогенетическое дерево вирусов, принадлежащих семейству *Astroviridae*, построенное по аминокислотной последовательности капсидного полипротеина алгоритмом Neighbor-joining. На узлах дерева указаны бутстрэп-поддержки в процентах (1000 репликатов), длины ветвей соответствуют расстояниям. Справа от дерева указаны виды, признанные ICTV, в скобках указаны хозяева. Сокращения: HAstV, human astrovirus; FAstV, feline astrovirus; PAstV, porcine astrovirus; CaAstV, canine astrovirus; RabAstV, rabbit astrovirus; OAstV, ovine astrovirus; BoAstV, bovine astrovirus; MuAstV, murine astrovirus; WBastV, wild boar astrovirus; CcAstV, deer astrovirus; TAstV, turkey astrovirus; CAstV, chicken astrovirus; ANV, avian nephritis virus; WPIAstV, wild pigeon astrovirus. Адаптировано из [197].

#### 1.4.2 Физико-химические свойства и структура вириона

Вирион астровирусов представляет собой частицу икосаэдрической формы ( $T=3$ ), примерно 10% вирионов имеют форму 5-6 конечных звезд. Вирусные частицы сначала состоят из прекурсорного капсидного полипротеина VP90 (молекулярная масса 90 кДа), у которого клеточными каспазами в дальнейшем отсекается С-концевой домен. Для того чтобы частицы из полученного белка VP70 (70 кДа) созрели, необходимо разрезание трипсином, которое происходит вне хозяйской клетки. Созревшие вирусные частицы крайне заразны и состоят из трех капсидных белков с молекулярной массой 32 – 34, 27 – 29, 25 – 26 кДа (VP34, VP27/29 и VP25/26 соответственно) (Рис. 10).

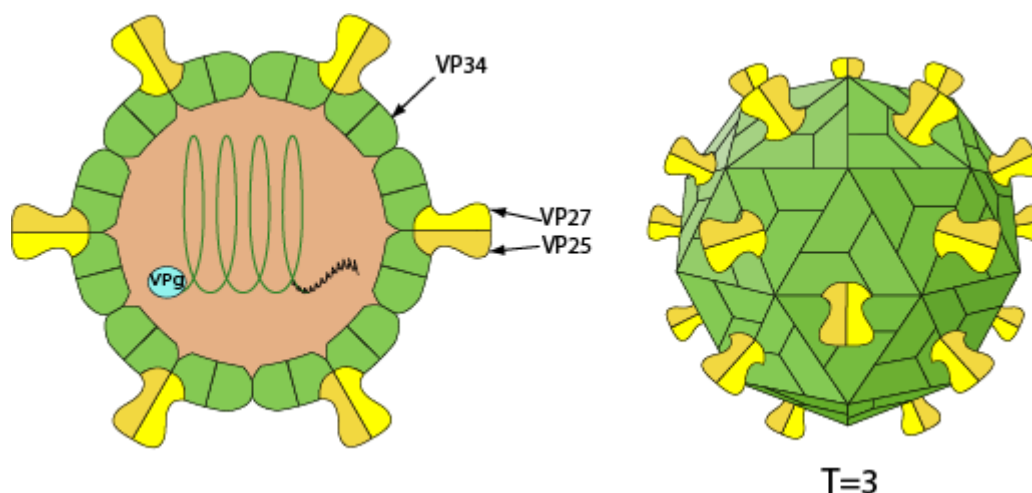


Рис. 10. Строение вириона астровирусов [198].

VP34 образуется из очень консервативного N-концевого участка полипротеина и образует оболочку капсида, а VP27/29 и VP25/26 происходят из



одного и того же вариабельного С-концевого терминального участка, но имеют разные N-концы и образуют димерные шипы. Диаметр астровирусных частиц различается в зависимости от источника вируса, обработки образца и может составлять от 28 до 41 нм. Реконструкции с помощью криоэлектронной микроскопии показывают, что оболочка капсида имеет диаметр 35 нм и состоит из 180 субъединиц капсидного белка с 30 (в зрелых частицах) либо 90 (в незрелых частицах) глобулярными димерными шипами. Таким образом, общий диаметр частицы составляет 44 нм [199].

Вирионы астровирусов устойчивы к кислой среде (рН=3), нагреванию до 50 градусов в течение часа или до 60 градусов в течение пяти минут, хлороформу, неионным, анионным и цвиттерионным детергентам.

#### 1.4.3 Структура генома

Геном астровирусов представляет собой оцРНК положительной полярности длиной от 6.2 до 7.8 тыс. нт. К 5'-концу генома ковалентно прикреплен белок VPg, а на 3'-конце есть поли-А хвост длиной примерно 30 нт. Геном содержит три открытые рамки считывания: ORF1a, ORF1b и ORF2 (Рис. 11). ORF1a кодирует возможный домен хеликазы, несколько трансмембранных доменов и домен coiled-coil, домен протеазы, VPg, гипервариабельный участок, сигнал ядерной локализации, возможный домен смерти (death domain, DD) (Рис. 12). ORF1b кодирует РзРп. ORF2 кодирует белки капсида, которые экспрессируются с сгРНК. В астровирусах человека и некоторых астровирусах млекопитающих описана новая ОРС – ORFX – которая перекрывается с ORF2. Вероятно, инициация её трансляции происходит по механизму «сканирование с протечкой» («leaky scanning»). Показано, что продукт трансляции ORFX обладает виropорин-подобной активностью [200].

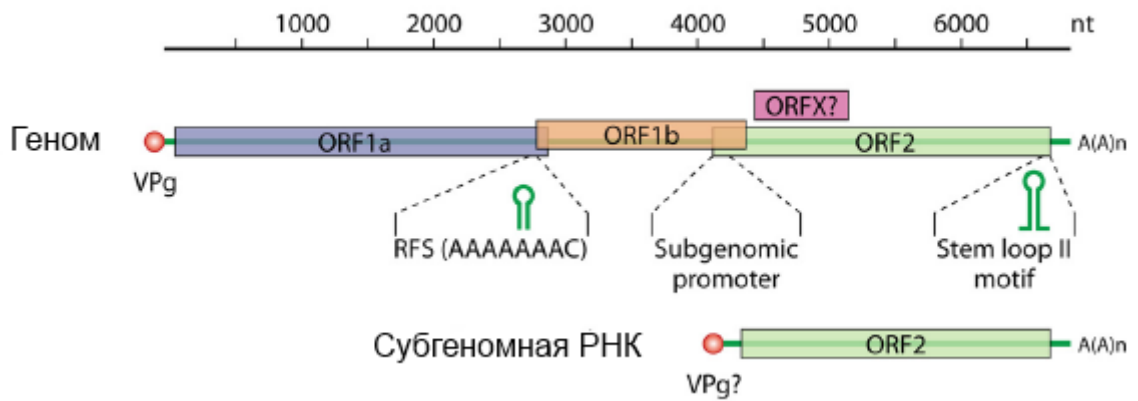


Рис. 11. Схема строения генома астровирусов. Рисунок взят из [197].

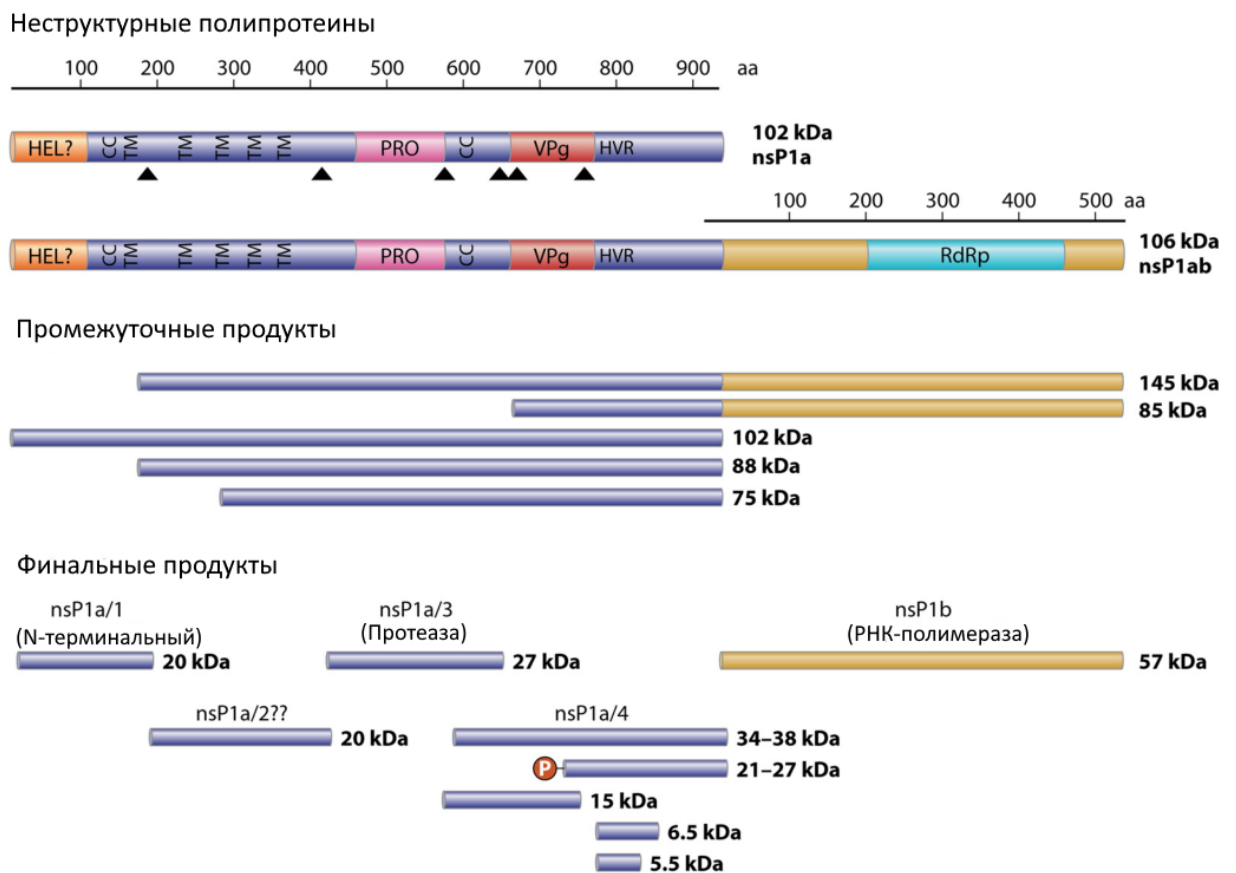


Рис. 12. Схема строения неструктурных полипротеинов астровирусов. Промежуточные и финальные продукты расщепления протеазами астровирусных неструктурных белков, экспрессируемых ORF1a и ORF1b. HEL, предполагаемый хеликазный домен; CC, домен coiled-coil; TM, трансмембранный домен; PRO, протеазный домен; HVR, гипервариабельная область; RdRp, РНК-зависимая РНК полимераз; P, посттрансляционное фосфорилирование. Треугольники изображают места протеолитического расщепления, описанные в литературе. Рисунок взят из [197].

Длины ORF1a, ORF1b и ORF2, а также длина перекрывающегося участка у ORF1a и ORF1b отличаются у разных штаммов. Например, у астровирусов млекопитающих длина перекрывающегося участка варьируется от 10 до 148 нт, а у астровирусов птиц – от 12 до 45 нт. В перекрывающемся участке находится сигнал рибосомального сдвига рамки считывания, который необходим для синтеза РзРп, закодированной в ORF1b. Перекрывающийся участок ORF1b и ORF2 также отличается или может вовсе отсутствовать. В нем находится очень консервативная последовательность, которая, возможно, является частью промотора сГРНК [199].

На 5'- и 3'-концах генома есть НТО длиной 11–85 и 80–85 нт соответственно. Для астровирусов не показано наличие IRES. На 3'-конце у классических астровирусов человека, а также астровирусов кошек, овец, свиней и птиц присутствует консервативный мотив вторичной структуры – stemloop II motif. Эта структура также была найдена в геномах некоторых коронавирусов, норовирусов и риновирусов. Функции шпильки неизвестны, но предполагается, что она может обеспечивать стабильность вторичной структуры РНК и взаимодействовать с вирусными и клеточными белками, необходимыми для репликации генома [201].

#### 1.4.4 Репликация

Цикл репликации лучше всего изучен для «классических» астровирусов человека. Главный клеточный рецептор HАstV остается неизвестным. Различные штаммы HАstV демонстрируют разный тропизм к клеточным культурам, поэтому, вероятно, астровирусы имеют более одного рецептора или молекулы для связывания. Предполагается, что полисахариды могут быть общей рецепторной молекулой. Для входа в клетку астровирусы используют клатрин-зависимый эндоцитоз. Проникновение вируса в клетку зависит от степени кислотности и зрелости эндосомы, из которой вирусный геном высвобождается в цитоплазму. Во время входа вируса в клетку активируются клеточные сигнальные каскады. В частности, взаимодействие HАstV

с клеткой приводит к активации ERK1/2 (extracellular signal-regulated kinase) и PI3K (phosphoinositide 3-kinase) путей, которые необходимы для эффективного проникновения вируса и установления продуктивной инфекции [199].

После выхода РНК в цитоплазму происходит трансляция геномной РНК, в которой ключевую роль играет белок VPg, привлекающий клеточные факторы инициации трансляции аналогично VPg калицивирусов и пикорнавирусов. Трансляция генома приводит к образованию двух полипротеинов nsP1a1b и nsP1a, которые котрансляционно разрезаются вирусной сериновой протеазой из nsP1a и некоторыми клеточными протеазами (Рис. 12). Трансляция nsP1a1b происходит за счет рибосомального сдвига рамки считывания в месте перекрывания ORF1a и ORF1b. Протеолитическое расщепление полипротеина nsP1a1b приводит к образованию полимеразы nsP1b и белка nsP1a, который в дальнейшем расщепляется на несколько белков (Рис. 12). Некоторые неструктурные белки подвергаются посттрансляционному фосфорилированию, за счет чего модулируются их взаимодействия с другими белками.

Точный процесс репликации астровирусов не был изучен экспериментально и выводится из данных, полученных для других (+)РНК-вирусов. Полногеномная РНК отрицательной полярности синтезируется вирусной РзРп и служит шаблоном для синтеза сгРНК и гРНК. В полипротеине nsP1a были найдены несколько трансмембранных доменов, которые могут участвовать в закреплении неструктурных репликационных комплексов на клеточных мембранах. Как *in vitro*, так и *in vivo* было показано, что в клетках, инфицированных HAstV, происходят крупные перестройки мембран. Вирусные неструктурные белки и вирусная РНК колокализованы с эндоплазматическим ретикулумом (ЭПР). Также сигнал удержания в ЭПР был предсказан на конце nsP1a. Поэтому считается, репликация генома HAstV может происходить на мембранах, происходящих из ЭПР.

Структурные белки экспрессируются с сгРНК в виде полипротеина VP90, причем, вероятно, для этого процесса необходим белок VPg. Синтез

структурных белков происходит в тех же клеточных компартментах, где формируются репликационные комплексы. Изначально из полипротеина VP90 собирается незрелый вирион, связанный с внутриклеточными мембранами С-концом. Затем некоторые клеточные каспазы разрезают VP90 недалеко от С-конца, и получившиеся капсиды, состоящие из белка VP70, диссоциируют от мембран. Этот процесс необходим для выхода вируса из клетки, и освобождение вирусных частиц, скорее всего, происходит без лизиса клеток. После выхода из клетки капсиды подвергаются протеолитической обработке трипсином, что повышает их инфекционность [197].

#### 1.4.5 Рекомбинация астровирусов

Рекомбинация является одним из главных механизмов, играющих роль в эволюции астровирусов и появлении новых штаммов. В отличие от других семейств небольших РНК-вирусов рекомбинация астровирусов не изучалась *in vitro*, поэтому все известные данные касаются естественной рекомбинации и получены в ходе биоинформатического анализа генетических последовательностей астровирусов.

Можно выделить несколько типичных точек рекомбинации у астровирусов. Наиболее часто рекомбинация происходит в консервативном участке на стыке ORF1b и ORF2, причем было детектировано множество рекомбинантов между одинаковыми и разными «классическими» серотипами HAstV (MAstV1) [202–208], новыми астровирусами человека (MAstV6, MAstV8) [209,210], астровирусами свиней [211–213], собак [214]. Также в литературе описана рекомбинация между вирусами, заражающими разных хозяев: астровирусами кошек MAstV2 (донор ORF1b) и астровирусами свиней MAstV3 (ORF2) [215], астровирусами свиней MAstV3 (донор ORF1b) и оленей [216]. Считается, что астровирус Калифорнийского морского льва CslAstV3 появился в результате рекомбинации астровируса человека HAstV4 (ORF1b) и CslAstV2 (ORF2) [217]. Недавно был отсеквенирован астровирус яка, который, вероятно, отнесут к новому виду. Оказалось, что на филогенетическом

дереве, построенном по ORF1ab, новый астровирус группируется с бычими астровирусами, в то время как ORF2 этого астровируса наиболее близка к таковой астровируса оленя CcAstV2. Это может также свидетельствовать о рекомбинации вирусов, заражающих разные виды [218]. Помимо домашних кошек, астровирусы были идентифицированы у гепардов с диареей, выращенных в неволе. ORF2 астровируса гепарда кластеризовалась ближе к астровирусам человека, чем ORF1b. Одним из возможных объяснений этого могла быть рекомбинация генов капсида астровирусов, подобных человеческим, и полимеразы другого неизвестного вируса [219]. Неясно, преодолел ли астровирус гепарда видовой барьер от домашних кошек, или он появился у гепардов независимо [220].

Кроме рекомбинации на границе ORF1 и ORF2, часто детектировали рекомбинацию внутри ORF2 между «классическими» HАstV [202,221], астровирусами собак [214,222], свиней [211,212] и верблюдов [223]. Было показано, что рекомбинация в вариабельной области ORF2 происходит между HАstV и астровирусами свиней [224] и кошек [209]. При этом астровирусы свиней, являющиеся результатом рекомбинации HАstV и астровирусов свиней, были собраны в одном и том же экономически депрессивном географическом районе, где люди вынуждены жить со своими животными под одной крышей [224]. Была описана рекомбинация между астровирусом быков и астровирусом Европейской косули [225]. В этом же исследовании обнаружили коинфекцию разными астровирусами быков. Это согласуется с другими сообщениями о коинфекциях различными астровирусами, заражающими одного и того же хозяина, как было описано для людей [205] и свиней [226], и является необходимым условием для рекомбинации между различными штаммами.

Другие точки рекомбинации являются менее распространенными. Рекомбинация на стыке ORF1a/ORF1b была детектирована между астровирусами свиней [211,212,227] и между разными «классическими» серотипами HАstV (MAstV1) [207,208,228]. Рекомбинация внутри ORF1a была найдена у

астровирусов свиней [212]. В недавнем исследовании были отсеквенированы новые астровирусы жвачных животных, среди которых были два рекомбинантных астровируса коз, появившихся в результате рекомбинации внутри ORF1a [229]. Первый астровирус являлся рекомбинантом астровирусов овец (OvAstV) и коз (CapAstV), а второй – результатом рекомбинации двух астровирусов быков. Рекомбинацию внутри ORF1b показали для «классических» HAstV [208,230] и астровирусов собак [214].

## 1.5 Общая характеристика коронавирусов

### 1.5.1 Классификация

Коронавирусы – это самая большая группа в порядке *Nidovirales*, который включает семейства *Coronaviridae*, *Arteriviridae* и *Roniviridae*. Нидовирусы – это оболочечные РНК-вирусы с геномом положительной полярности. Этот порядок выделяется среди других РНК-вирусов. Для нидовирусов характерны следующие особенности:

- инвариантная общая организация генома, в котором большой ген полимеразы находится перед структурными генами;
- экспрессия полипротеина-предшественника неструктурных белков осуществляется за счет -1 рибосомального сдвига рамки считывания;
- уникальные для порядка ферментативные активности неструктурных белков (например, домен нуклетидилтрансферазы NiRAN и домен уридилат-специфической эндорибонуклеазы, EndoU);
- экспрессия структурных белков через транскрипцию 3'-котерминальных вложенных сгРНК [231].

Коронавирусы относятся к подсемейству *Orthocoronavirinae* семейства *Coronaviridae*. Помимо коронавирусов, в семейство *Coronaviridae* сейчас включают ещё два подсемейства – *Letovirinae* и *Pitovirinae* [232]. Классификация коронавирусов основана на филогении и расчете попарных эволюционных расстояний между следующими консервативными доменами в полипротеине полимеразы: ADRP, nsp5 (3CLpro), nsp12 (PзРп), nsp13 (Hel), nsp14

(EcoN), nsp15 (NendoU) и nsp16 (O-MT). Вирусы относят к новому роду, если они не кластеризуются ни с одним из текущих родов и имеют менее 46% идентичности последовательности этих доменов с любым другим установленным членом семейства. Вирусы, схожие по последовательности консервативных доменов более чем на 90%, считаются принадлежащими к одному и тому же виду [233].

Выделяют 4 рода коронавирусов – *Alphacoronavirus* ( $\alpha$ -CoV), *Betacoronavirus* ( $\beta$ -CoV), *Gammacoronavirus* ( $\gamma$ -CoV) и *Deltacoronavirus* ( $\delta$ -CoV) (Таблица 6). Члены  $\alpha$ -CoV и  $\beta$ -CoV заражают преимущественно млекопитающих, в то время как  $\gamma$ -CoV и  $\delta$ -CoV в основном заражают птиц, но также были найдены у китообразных и свиней.

Таблица 6. Классификация коронавирусов согласно отчету ICTV за 2021 год [234]

Род <i>Alphacoronavirus</i>	Род <i>Betacoronavirus</i>
подрод <i>Amalacovirus</i> <i>Alphacoronavirus AMALF</i>	подрод <i>Embecovirus</i> <i>Betacoronavirus 1</i> (Bovine coronavirus, Canine respiratory coronavirus, Dromedary camel coronavirus HKU23, Equine coronavirus, Human coronavirus OC43, Porcine hemagglutinating encephalomyelitis virus)
подрод <i>Colacovirus</i> <i>Bat coronavirus CDPHE15</i>	<i>China Rattus coronavirus HKU24</i> <i>Human coronavirus HKU1</i> <i>Murine coronavirus</i> (Murine hepatitis virus (MHV)) <i>Myodes coronavirus 2JL14</i>
подрод <i>Decacovirus</i> <i>Alphacoronavirus CHB25</i> <i>Alphacoronavirus WA3607</i> <i>Bat coronavirus HKU10</i> <i>Rhinolophus ferrumequinum alphacoronavirus HuB-2013</i>	подрод <i>Hibecovirus</i> <i>Bat Hp-betacoronavirus Zhejiang2013</i>
подрод <i>Duvinacovirus</i> <i>Human coronavirus 229E</i>	подрод <i>Merbecovirus</i> <i>Hedgehog coronavirus 1</i> <i>Middle East respiratory syndrome-related coronavirus</i> <i>Pipistrellus bat coronavirus HKU5</i> <i>Tylonycteris bat coronavirus HKU4</i>
подрод <i>Luchacovirus</i> <i>Lucheng Rn rat coronavirus</i>	подрод <i>Nobecovirus</i> <i>Eidolon bat coronavirus C704</i> <i>Rousettus bat coronavirus GCCDC1</i> <i>Rousettus bat coronavirus HKU9</i>
подрод <i>Minacovirus</i> <i>Mink coronavirus 1</i>	подрод <i>Sarbecovirus</i> <i>Severe acute respiratory syndrome-related</i>
подрод <i>Minunacovirus</i> <i>Miniopterus bat coronavirus 1</i> <i>Miniopterus bat coronavirus HKU8</i>	
подрод <i>Myotacovirus</i> <i>Myotis ricketti alphacoronavirus Sax-2011</i>	



	<i>coronavirus</i>
	Род <i>Gammacoronavirus</i>
подрод <i>Nyctacovirus</i> <i>Alphacoronavirus HKU33</i> <i>Alphacoronavirus WA2028</i> <i>Nyctalus velutinus alphacoronavirus SC-2013</i> <i>Pipistrellus kuhlii coronavirus 3398</i>	подрод <i>Brangacovirus</i> <i>Goose coronavirus CB17</i>
подрод <i>Pedacovirus</i> <i>Alphacoronavirus BT020</i> <i>Alphacoronavirus WA1087</i> <i>Porcine epidemic diarrhea virus (PEDV)</i> <i>Scotophilus bat coronavirus 512</i>	подрод <i>Cegacovirus</i> <i>Beluga whale coronavirus SW1</i>
подрод <i>Rhinacovirus</i> <i>Rhinolophus bat coronavirus HKU2</i>	подрод <i>Igacovirus</i> <i>Avian coronavirus</i> <i>Avian coronavirus 9203</i> <i>Duck coronavirus 2714</i>
подрод <i>Setracovirus</i> <i>Human coronavirus NL63</i> <i>NL63-related bat coronavirus strain BtKYNL63-9b</i>	Род <i>Deltacoronavirus</i>
подрод <i>Soracovirus</i> <i>Sorex araneus coronavirus T14</i>	подрод <i>Andecovirus</i> <i>Wigeon coronavirus HKU20</i>
подрод <i>Sunacovirus</i> <i>Suncus murinus coronavirus X74</i>	подрод <i>Buldecovirus</i> <i>Bulbul coronavirus HKU11</i> <i>Common moorhen coronavirus HKU21</i> <i>Coronavirus HKU15</i> <i>Munia coronavirus HKU13</i> <i>White-eye coronavirus HKU16</i>
подрод <i>Tegacovirus</i> <i>Alphacoronavirus 1 (Canine coronavirus, Feline coronavirus, Swine enteric coronavirus, Transmissible gastroenteritis virus)</i>	подрод <i>Herdecovirus</i> <i>Night heron coronavirus HKU19</i>

Коронавирусы вызывают у животных респираторные, желудочно-кишечные, неврологические заболевания, а также заболевания печени. Многие коронавирусы животных были открыты в первой половине 20-го века. Коронавирус кошек (*Feline coronavirus*, FeCoV,  $\alpha$ -CoV) вызывает кишечные инфекции у кошек и в 5-12% может привести к инфекционному перитониту – заболеванию со смертельным исходом. Вирус инфекционного бронхита (*Infectious bronchitis virus*, IBV,  $\gamma$ -CoV) вызывает инфекционный бронхит у домашних птиц, затрагивает репродуктивную систему и почки. Вирус трансмиссивного гастроэнтерита (*Transmissible gastroenteritis virus*, TGEV,  $\alpha$ -CoV), вирус эпидемической диареи свиней (вид *Porcine epidemic diarrhea virus*, PEDV, род  $\alpha$ -CoV), дельтакоронавирус свиней (вид *Coronavirus HKU15*,

PDCoV, рол  $\delta$ -CoV) могут вызывать диарею и рвоту различной степени тяжести. Вирус гепатита мышей (Murine hepatitis virus, MHV,  $\beta$ -CoV) вызывает у мышей гепатит, а также неврологические заболевания и энтерит. Инфекции коронавирусом быков (Bovine coronavirus, BCoV,  $\beta$ -CoV) вызывают различные клинические симптомы, включая диарею телят, геморрагическую диарею у взрослых животных и респираторные заболевания. У собак гастроэнтерит и респираторные симптомы обычно возникают из-за коронавируса собак (CCoV, Canine coronavirus,  $\alpha$ -CoV) и респираторного коронавируса собак (Canine respiratory coronavirus,  $\beta$ -CoV).

Коронавирусы людей были неизвестны до конца 1960-х. Болезни человека могут вызывать семь коронавирусов. Коронавирусы NL63 и 229E ( $\alpha$ -CoV), а также HKU1 и OC43 ( $\beta$ -CoV) вызывают симптомы простуды. За последние 20 лет в результате зоонозной передачи появилось три новых коронавируса – SARS-CoV [235], MERS-CoV [236] и SARS-CoV-2 [237] – которые вызывают тяжелые респираторные инфекции с более высокой летальностью и послужили причиной вспышек пневмонии и пандемии COVID-19. Все эти коронавирусы принадлежат роду  $\beta$ -CoV.

Филогенетические исследования показали, что летучие мыши являются основным резервуаром  $\alpha$ -CoV и  $\beta$ -CoV и их источником для других млекопитающих [182,183]. Дикие птицы являются резервуарными хозяевами  $\gamma$ - и  $\delta$ -CoV [240].

### 1.5.2 Физико-химические свойства и структура вириона

Вирионы коронавирусов плеоморфны, имеют приблизительно сферическую форму. Их диаметр составляет 120 – 160 нм, а на поверхности наблюдаются выступы примерно 20 нм, которые образованы тримерами шиповидного гликопротеина S. Помимо белка S, в липидную оболочку вириона встроены матриксный гликопротеин (M) и пентамеры гликопротеина внешней оболочки (E) (Рис. 13). В зависимости от вида коронавируса в вирион могут быть включены дополнительные вспомогательные белки, например,

некоторые  $\beta$ -CoV кодируют дополнительный гомодимерный гликопротеин оболочки – гемагглютинин-эстеразу (HE). РНК-геном связан с белком N и образует спиральный нуклеокапсид.

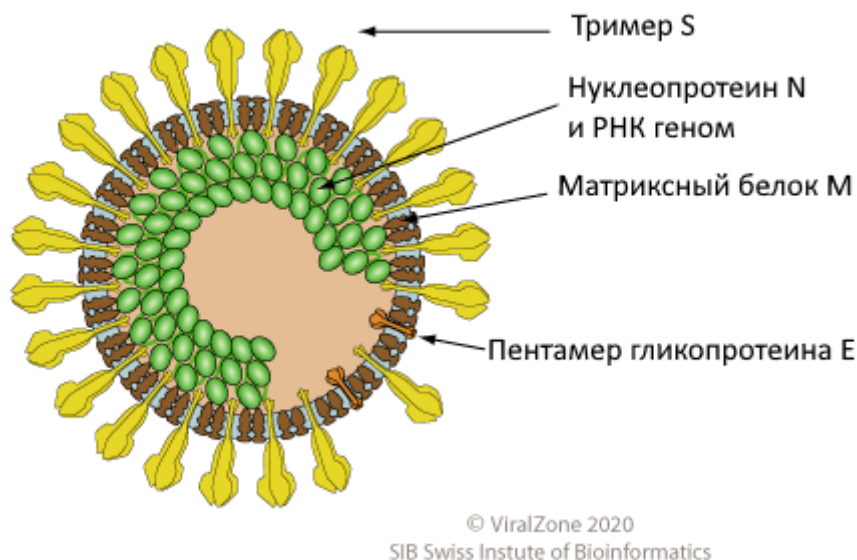


Рис. 13. Строение вириона коронавируса (SARS-CoV) [198].

Частицы коронавируса чувствительны к нагреванию, липидным растворителям, неионогенным детергентам, формальдегиду, окислителям и УФ-излучению.

### 1.5.3 Структура генома

Геном коронавируса представляет собой одноцепочечную РНК положительной полярности длиной 26–32 тыс. нт и является самым длинным среди всех РНК-вирусов. Он имеет стандартный эукариотический кэп на 5'-конце и поли-А хвост на 3'-конце, а также содержит 5'- и 3'-НТО длиной от 200 до 600 нт и от 200 до 500 нт соответственно. Геном включает один длинный ген репликазы длиной около 20 Кб, разделенный на две ОРС – ORF1a и ORF1b, в результате трансляции которых через сдвиг рамки считывания образуются два полипротеина, pp1a и pp1ab. За ним следуют четыре ОРС, которые кодируют базовые белки, содержащиеся в геномах всех коронавируса: шиповидный белок S, гликопротеин внешней оболочки E, матриксный белок M и белок нуклеокапсида N. Порядок базовых ОРС в геномах

коронавирусов инвариантен. Помимо базовых белков геном может содержать от одной до восьми ОРС, которые кодируют дополнительные белки, причем они могут попасть в любой из межгенных интервалов ниже гена репликазы, за исключением интервала между генами E и M [231] (Рис. 14). Дополнительные белки необходимы для репликации вируса, но различаются по количеству и часто негомологичны даже среди близкородственных вирусов, что указывает на события горизонтального переноса во время их эволюции [241].

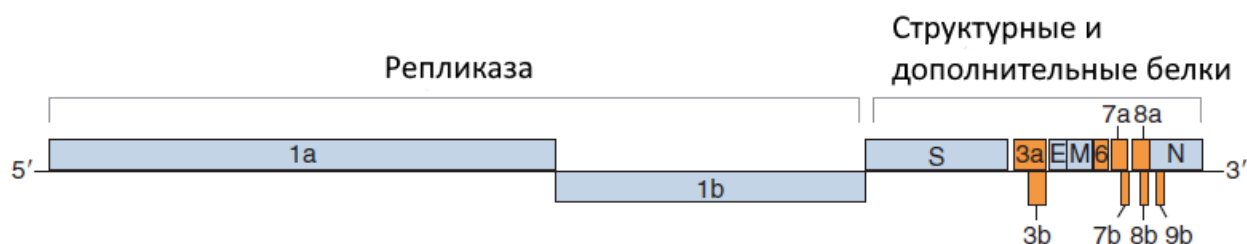


Рис. 14. Схема генома коронавирусов на примере IBV. Рисунок из [242] с небольшими изменениями.

#### 1.5.4 Репликация

Коронавирусная инфекция начинается со связывания вирионов с клеточными рецепторами и последующим эндоцитозом. Шиповидный белок является основным рецептор-связывающим белком коронавируса. Он состоит из двух основных доменов: апикального домена S1, который несет сайты связывания рецепторов, и базального домена S2. S1 дополнительно разделяют на N-концевой и C-концевой субдомены (S1-NTD и S1-CTD). Область S1, связывающаяся с рецептором, отличается у коронавирусов. Например, у MHV рецептор-связывающим доменом служит S1-NTD, а у SARS-CoV, HCoV NL63, TGEV и HCoV-229 рецептор-связывающий домен находится в S1-CTD. После проникновения вируса в клетку происходит слияние мембраны вируса с эндосомальной мембраной, опосредованное S2, и нуклеокапсид высвобождается в цитоплазму.

После попадания нуклеокапсида в цитоплазму происходит трансляция гена репликазы, который состоит из двух ОРС, имеющих небольшой перекрывающийся участок. Трансляция ORF1b происходит за счет рибосомального сдвига рамки считывания, которому способствуют «скользящая» после-

довательность 5'-UUUAAAC-3' и РНК-псевдоузел, находящиеся в участке перекрытия ORF1a и ORF1b. Получившиеся в результате трансляции гена репликазы полипротеины pp1a и pp1b, которые последовательно аутопротеолитически расщепляются на 16 зрелых белков – nsp1–nsp16. Некоторые продукты частичного расщепления также играют функциональную роль. Неструктурные белки формируют репликативный комплекс, ассоциированный с мембраной.

Репликативный комплекс осуществляет репликацию геномной РНК и синтез сгРНК. На 3'-конце лидерной последовательности, кодируемой в начале генома, и перед генами структурных белков находятся консервативные участки длиной 6–9 нт, называемые TRS (transcriptional regulatory sequence). При синтезе (-)цепи на каждом участке TRS с определенной вероятностью либо происходит продолжение синтеза (-)РНК, либо «перескок» на TRS лидерной последовательности и достройка анти-лидерной последовательности. Набор анти-сгРНК используется для синтеза сгРНК, имеющих одинаковую 5'-лидерную последовательность.

С сгРНК транслируются структурные белки. Синтез белков S, E, M с последующим гликозилированием происходит в ЭПР. Из ЭПР они транспортируются в место сборки вирионов – промежуточный компартмент ЭПР-аппарата Гольджи (ERGIC), где происходит сборка вирионов, в которой ключевую роль играют белки E и M, и отпочковывание вирионов от мембран. Высвобождение новых вирионов происходит за счет экзоцитоза.

#### 1.5.5 Рекомбинация коронавируса

Гомологичная рекомбинация коронавируса была впервые показана у МНУ в 1985 году в эксперименте с коинфекцией клеток диким и термочувствительным штаммами вируса [243]. В дальнейших экспериментальных исследованиях показали, что частота рекомбинации у МНУ крайне высока и за один пассаж может достигать 25% [244]. На коротких промежутках времени рекомбинация происходит практически случайным образом по всему геному,

однако после серийных пассажей рекомбинантных вирусов в культуре тканей точки рекомбинации кластеризуются в определенных областях – «горячих точках рекомбинации», что скорее всего вызвано давлением отбора [245]. Более поздние исследования показали, что частота рекомбинации постепенно увеличивается от 5'-конца к 3'-концу генома. Это связывают с наличием транскрипционно активных анти-сРНК, которые увеличивают количество мРНК, способной участвовать в рекомбинации по «соры-choice» механизму [246].

Отдельные события рекомбинации были найдены у многих коронавирусов. Наибольшее число событий рекомбинации во всех четырех родах коронавирусов было обнаружено в гене белка *S*. Часть этих событий относится к практически полной передаче гена *S*, когда точка рекомбинации находится прямо перед геном шиповидного белка и после него. Подобная рекомбинация была найдена как между TGEV [247], так и между вирусами свиней TGEV и вирусом эпидемической диареи свиней (PEDV), принадлежащими разным под родам  $\alpha$ -CoV [248]. Коронавирус индейки TCoV близок по последовательности птичьему вирусу инфекционного бронхита (IBV, вид *Avian Coronavirus*) во всех участках генома, кроме гена шиповидного белка, происхождение которого неизвестно [249]. Новый дельтакоронавирус воробья имел ген *S* более близкий к дельтакоронавирусу свиньи, чем к известным дельтакоронавирусам воробьев [250]. Есть случаи, когда точка рекомбинации находится перед геном шиповидного белка, и таким образом все структурные белки передаются рекомбинанту. Например, в результате рекомбинации коронавирусов собак (CCoV) и кошек (FeCoV) перед геном *S* появились несколько штаммов кошачьего коронавируса [251]. Подобная рекомбинация описана между штаммами человеческих коронавирусов OC43 [252]. Однако наиболее часто происходит обмен участками гена *S*. Обмен целыми доменами S1 [253–256], а также рекомбинация внутри доменов S1 и S2 [257–262] были описаны у штаммов IBV, циркулирующих в разных странах мира, включая рекомбинацию между вакцинными и циркулирующими штаммами

[253]. В роде  $\alpha$ -CoV описаны случаи рекомбинации внутри S2 между PEDV [263,264], а также между вирусами, заражающими разных хозяев. Рекомбинация внутри S происходит между коронавирусом собак CCoV и TGEV ( $\alpha$ -CoV) [265–267], коронавирусом человека 229E и коронавирусом альпаки [268]. Участки гена S с рекомбинантным происхождением были детектированы во всех новых коронавирусах человека. Обмен как полным геном S, так и его доменом S1 был показан между SARS-подобными вирусами летучих мышей [235], между MERS-подобными вирусами верблюдов [236], а также MERS-подобными вирусами летучих мышей и верблюдов [269]. Рецептор-связывающий домен SARS-CoV-2 был приобретен в результате рекомбинации с коронавирусом панголина [237]. Рекомбинация внутри S была выявлена между различными  $\delta$ -CoV, заражающими свиней и птиц. Считается, что она играет роль в облегчении кросс-видовых переходов  $\delta$ -CoV [270].

В ряде работ по реконструкции рекомбинантных коронавирусов показано, что отдельные фрагменты шиповидного белка, включая домены S1, S2, могут быть взаимозаменяемы и влияют на диапазон инфицируемых хозяев [271]. Например, удалось заменить в SARS-подобном коронавирусе летучих мышей не только рецептор-связывающий домен, но и его 3'-половину на эквивалент человеческого коронавируса, причем оба мутантных вируса были инфекционны и могли заражать клетки приматов [272]. Поэтому говорят о том, что шиповидный белок имеет модульную структуру.

Рекомбинация в других областях генома детектировалась существенно реже. Есть сообщения о рекомбинации внутри ORF1ab между CCoV и FeCoV ( $\alpha$ -CoV) [251], разными штаммами PDCoV [273], внутри гена M между PEDV и коронавирусом человека NL63 [274], в гене гемагглютинин-эстеразы между коронавирусом верблюда HKU23 и коронавирусом кроликов [275], в генах 3, 4 между IBV [276].

Помимо гомологичной рекомбинации между коронавирусами, было предположено, что могла происходить негомологичная рекомбинация между коронавирусами и вирусами других семейств. Например, схожие вторичные структуры РНК были найдены в 3'-НТО сапеловирусов, энтеровирусов и ко-

ронавирусов, заражающих летучих мышей [277]. Так как дополнительные коронавирусные белки отличаются по количеству и бывают негомологичны даже у близкородственных коронавирусов, предполагается, что они могли быть приобретены за счет рекомбинации [241]. Один из генов нового коронавируса ночного крылана Ro-BatCoV GCCDC1 не имел гомологов среди известных коронавирусов, но достигал 55% сходства с белком p10 ортореовируса летучих мышей [278]. Аминокислотная последовательность ORF2 MHV A59 оказалась гомологична гемагглютинуину HA1 вируса гриппа С [279].

Рекомбинация играет важную роль в эволюции РНК-вирусов, являясь механизмом для создания генетической изменчивости и уменьшения мутационного груза. Благодаря рекомбинации вирусы могут приобретать новые фенотипические признаки, например, менять тропизм к клеткам и круг хозяев, менять патогенность, уклоняться от иммунного ответа. Для того чтобы рекомбинация между вирусами произошла, необходимо выполнение ряда условий: вирусы должны заразить одну и ту же клетку и успешно реплицироваться в ней. В случае гомологичной репликативной рекомбинации вирусные последовательности должны быть достаточно похожи, чтобы полимераза села на новую матрицу. Более того, полученный рекомбинант должен оказаться жизнеспособным, чтобы закрепиться в вирусной популяции. Те рекомбинантные последовательности, которые детектируются у циркулирующих вирусов, являются результатом естественного отбора. Последний систематический анализ естественной рекомбинации РНК-вирусов ряда семейств (*Picornaviridae*, *Alphaviridae*, *Astroviridae*, *Caliciviridae*, *Flaviviridae*) был проведен в 2006 году, когда в рамках одного семейства было известно не более 200 вирусных последовательностей [184]. За последние 20 лет количество вирусных последовательностей в базах данных увеличилось в сотни раз. Изолированные события рекомбинации были описаны для разных видов семейств РНК-вирусов, однако понимания о совместимости геномных фрагментов и частоты рекомбинации пока нет. Несколько глав данной работы посвящены систематическому изучению естественной рекомбинации в семействах *Caliciviridae*, *Astroviridae* и *Coronaviridae*.



## 2 МАТЕРИАЛЫ И МЕТОДЫ

### 2.1 Данные

#### 2.1.1 Подготовка выравниваний полных геномов для анализа рекомбинации

Все доступные полные геномные последовательности вирусов, принадлежащих семействам *Caliciviridae* (рода *Norovirus*, *Lagovirus*, *Sapovirus*), *Astroviridae* (род *Mamastrovirus*) и *Coronaviridae* (рода *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, *Deltacoronavirus*) (Таблица 7) были загружены из базы данных нуклеотидных последовательностей GenBank (апрель 2019-август 2020). Для анализа рекомбинации были подготовлены выравнивания нуклеотидных последовательностей вирусов, принадлежащих родам, указанным в Таблице 7. Для калицивирусов и астровирусов были подготовлены выравнивания слитых нуклеотидных последовательностей всех ОРС. Для коронавирусов были выбраны только последовательности пяти основных генов (полная ORF1ab, S, E, M и N), которые присутствуют в геномах всех родов коронавирусов.

Для автоматизации подготовки выравниваний был разработан набор скриптов на языке Python. Процесс подготовки выравнивания нуклеотидных последовательностей включал следующие шаги:

- 1) Информация о названии изолята (квалификатор «isolate» поля «FEATURES»), хозяине вируса («host»), дате («collection date») и стране сбора («country») изолята была извлечена из файла с нуклеотидными последовательностями в формате «genbank», а также последовательности полных геномов были переведены в формат «fasta» (скрипт «parser\_gb.py»).
- 2) Извлечение информации о координатах ОРС из файла с нуклеотидными последовательностями в формате «genbank» (скрипт «get\_orfs.py»).
- 3) Последовательности ОРС были вырезаны из полных геномов по координатам, полученным в п.2 (скрипт «split\_genome.py»).

- 4) Нуклеотидные последовательности отдельных ОРС были транслированы в аминокислотные, выровнены с использованием программы mafft [280] и обратно транслированы (скрипт «trans\_alignment.py»).
- 5) Выравнивания отдельных ОРС были слиты (скрипт «join\_al.py»).
- 6) Нуклеотидные последовательности, схожие более чем на 99% для коронавирусов и 99.5% для других вирусов, были исключены из выравнивания (скрипт «remove\_similar.py»).
- 7) Колонки выравнивания, содержащие более 10% гэпов, были удалены из выравнивания с помощью программы trimAl [281].

Координаты ключевых белков вирусов были взяты из аннотаций Swiss-Prot [282]. Для последовательностей в финальном выравнивании недостающие данные о названии изолята, хозяине вируса, дате и стране сбора изолята были взяты из литературы, если было возможно. В таблице 7 обобщена информация о полученных выравниваниях для всех таксонов.

По группированию последовательностей на филогенетическом древе было выполнено уточнение таксономического положения вирусов, вошедших в выборку. У норовирусов и саповирусов уточнялись геногруппы, а у лаговирусов – генотип.

Таблица 7. Характеристика выборок последовательностей, использованных для анализа рекомбинации

Семейство	Род	Вид	Число доступных последовательностей в GenBank	Число последовательностей в итоговом выравнивании	Длина выравнивания, нт
<i>Caliciviridae</i>					
	<i>Norovirus</i>	<i>Norwalk virus</i>	3439 (сентябрь 2020)	1084	7431
	<i>Sapovirus</i>	<i>Sapporo virus</i>	216 (апрель 2019)	148	7191
	<i>Lagovirus</i>	<i>European brown hare syndrome virus,</i> <i>Rabbit hemorrhagic disease virus</i>	507 (июнь 2021)	338	7401
<i>Astroviridae</i>					
	<i>Mamastrovirus</i>	<i>Mamastrovirus I-19</i>	382 (август 2020)	319	4764
<i>Coronaviridae</i>					
	<i>Alphacoronavirus</i>	<i>Alphacoronavirus 1,</i> <i>Miniopterus bat coronavirus HKU8,</i> <i>Miniopterus bat coronavirus 1,</i> <i>Mink coronavirus 1,</i> <i>Porcine epidemic diarrhea virus,</i> <i>Human coronavirus NL63,</i> <i>Hipposideros bat coronavirus HKU10,</i> <i>Human coronavirus 229E-related,</i> <i>Bat coronavirus CDPHE15,</i> <i>Rhinolophus ferrumequinum alphacoronavirus HuB-2013,</i> <i>NL63-related bat coronavirus,</i> <i>Bat coronavirus CDPHE16,</i> <i>Scotophilus bat coronavirus 512,</i> <i>Miniopterus schreibersii bat coronavirus 1-related</i>	1260 (июль 2020)	164	24585
	<i>Betacoronavirus</i>	<i>Murine coronavirus,</i>	17083 (июль 2020)	122	24690

Семейство	Род	Вид	Число доступных последовательностей в GenBank	Число последовательностей в итоговом выравнивании	Длина выравнивания, нт
		<i>Betacoronavirus 1</i> , <i>Severe acute respiratory syndrome-related coronavirus</i> , <i>Human coronavirus HKU1</i> , <i>Tylonycteris bat coronavirus HKU4</i> , <i>Pipistrellus bat coronavirus HKU5</i> , <i>Rousettus bat coronavirus HKU9</i> , <i>Betacoronavirus 1 related</i> , <i>Middle East respiratory syndrome-related coronavirus</i> , <i>Hedgehog coronavirus 1</i> , <i>China Rattus coronavirus HKU24</i> , <i>Bat Hp-betacoronavirus Zhejiang2013</i> , <i>Rousettus bat coronavirus GCCDC1</i>			
	<i>Gammacoronavirus</i>	<i>Avian coronavirus</i> , <i>Avian coronavirus 9203</i> , <i>Beluga whale coronavirus SW1</i> , <i>Duck coronavirus 2714</i> <i>Goose coronavirus CB17</i>	452 (июль 2020)	260	24870
	<i>Deltacoronavirus</i>	<i>Bulbul coronavirus HKU11</i> , <i>Munia coronavirus HKU13</i> , <i>Coronavirus HKU15</i> , <i>White-eye coronavirus HKU16</i> , <i>Common moorhen coronavirus HKU21</i>	162 (июль 2020)	56	23840

### 2.1.2 Изоляты CVA2, выделенные от больных с острыми вялыми параличами

Впервые были исследованы 5 изолятов энтеровирусов типа коксакивирус А2 (CVA2), выделенных на территории Российской Федерации в 2001-2020 годах от больных с ОВП. Материалы для исследования были любезно предоставлены Ивановой О.Е. (ФГАУ «Федеральный научный центр исследований и разработки иммунобиологических препаратов им. М.П. Чумакова РАН» (Институт полиомиелита)).

### 2.1.3 Подготовка выравниваний для филогенетического анализа типов энтеровирусов

Для анализа молекулярной эволюции были выбраны 44 типа энтеровирусов человека, для каждого из которых было доступно более 50 последовательностей в базе данных GenBank. Данные о вирусах, выбранных для филогенетического анализа, представлены в Таблице 8. Для каждого из типов все доступные последовательности изолятов были выгружены из базы GenBank, чтобы составить выравнивания области генома VP1 и провести филогенетический анализ. Так как число последовательностей коротких типизирующих фрагментов VP1 (длиной около 300 нт) значительно превышало число полных последовательностей VP1 (длиной 840–930 нт), для каждого типа было составлено по два выравнивания: выравнивание типизирующих фрагментов и полных генов VP1.

Было составлено отдельное выравнивание последовательностей прототипных штаммов, которое в дальнейшем использовались для определения точных границ участка VP1 в выравниваниях для каждого типа. Последовательности полных геномов (или областей P1, если полный геном недоступен) прототипных штаммов каждого типа (Таблица 8) [168,193] были выровнены с помощью программы mafft v.7.304 [283]. Полные участки VP1 были вырезаны вручную согласно предсказанным границам, указанным в аннотациях GenBank-файлов, и с учетом последовательностей сайтов разрезания протеаз [284]. Короткие типизирующие фрагменты были вручную вырезаны по после-

довательностям универсальных праймеров AN232 и AN233, опубликованным в [169].

Выборка последовательностей энтеровирусов, депонированных в GenBank, неравномерна: во многих странах не существует системы рутинного надзора за энтеровирусами, поэтому большая часть доступных последовательностей энтеровирусов получена в результате научных исследований. Также в GenBank нередко депонируют множества практически идентичных последовательностей, полученных в ходе расследования вспышек заболеваний. Больше всего изолятов отсеквенировано в Китае, Индии, странах Европы, США и России.

В большинстве современных филогенетических исследований энтеровирусов выборки нуклеотидных последовательностей изолятов составляются вручную. Такой подход является недостаточно объективным для всеобъемлющего анализа и слишком трудоемким: для некоторых типов число известных последовательностей превышает сотни (Таблица 8). С целью повышения оперативности исследований и уменьшения количества ошибок, вносимых при ручной работе, был разработан метод, позволяющий из всех имеющихся данных в базе выбрать последовательности, относящиеся к области VP1, исключить неинформативные последовательности и составить выравнивание. Алгоритм, реализующий разработанный метод, включает следующие шаги.

1. Чтение из базы данных GenBank всех записей в формате «genbank», относящихся к определенному типу; запись последовательностей в файл в формате «fasta», причем идентификатор последовательности в новом файле формировался следующим образом: идентификатор в GenBank + “\_” + страна выделения + “\_” + год выделения изолята. Страна и год выделения последовательности включались в её новый идентификатор, если в её записи в GenBank поле «FEATURES» содержало квалификаторы «country» и «collection date».

Таблица 8. Данные об энтеровирусах, выбранных для филогенетического анализа

Тип	Название прототипного штамма	Код прототипного штамма в GenBank	Число последовательностей типа в GenBank	Число последовательностей типизирующих фрагментов после обработки	Число последовательностей полных VP1 после обработки
CVA2	Fleetwood	AY421760	302	67	18
CVA4	High Point	AY421762	475	115	51
CVA5	Swartz	AY421763	165	38	11
CVA6	Gdula	AY421764	4156	228	89
CVA8	Donovan	AY421766	110	31	11
CVA9	Griggs	D00627	607	124	43
CVA10	Kowalik	AY421767	1809	147	71
CVA13	Flores	AF499637	199	70	29
CVA14	G-14	AY421769	70	20	15
CVA16	G-10	U05876	3818	303	252
CVA20	Tulane	EF015021	118	58	34
CVA21	Kuykendall	AF546702	170	31	27
CVA24	Joseph	EF026081	1527	100	46
CVB1	Conn-5	M16560	558	100	43
CVB2	Ohio-1	AF081485	428	96	34
CVB3	Nancy	JX312064	945	177	106
CVB4	JVB	X05690	838	166	102
CVB5	Faulkner	AF114383	1226	230	136
CVB6	Schmitt	AF105342	102	28	17
E1	Farouk	AF029859	129	67	29
E2	Cornelis	AY302545	71	38	10
E3	Morrisey	AY302553	286	80	35
E4	Pesaseck	AY302557	126	26	15
E6	D'Amori	AY302558	1727	336	185

Тип	Название прототипного штамма	Код прототипного штамма в GenBank	Число последовательностей типа в GenBank	Число последовательностей типизирующихся фрагментов после обработки	Число последовательностей полных VP1 после обработки
E7	Wallace	AY302559	346	139	87
E9	Hill	X84981	814	32	28
E11	Gregory	AY903644	1360	209	111
E13	Del Carmen	AY302539	754	137	50
E14	Tow	AY302540	205	92	43
E16	Harrington	AY302542	72	23	7
E17	CHNE-29	AY302543	57	17	10
E18	Metcalf	AF317694	335	66	34
E19	Burke	AY302544	285	87	32
E20	JV-1	AY302546	136	54	22
E21	Farina	AY302547	101	42	17
E24	DeCamp	AY302548	96	35	19
E25	JV-4	AY302549	346	108	45
E29	JV-10	AY302552	113	45	27
E30	Bastianni	AF162711	4200	377	270
E32	PR-10	AY302555	51	17	12
E33	Toluca-3	AY302556	182	47	29
EV-D68	Fermon	AY426531	1667	85	33
EV-D70	J 670/71	DQ201177	214	9	9
EV-A71	BrCr	AB204853	9517	629	512

\* На август 2017 года.



2. Добавление к выборке последовательности прототипного штамма (полный участок VP1 или его типизирующего фрагмент), полученной как описано выше.

3. Удаление слишком коротких и слишком длинных последовательностей. Минимальная длина для полного VP1 – 800 нт, а для его типизирующего фрагмента – 250 нт. Для максимальной длины был выбран порог 8000 нт. Выборы порогов обоснованы длинами соответствующих фрагментов генома.

4. Удаление последовательностей, которые отличаются от последовательности прототипного штамма и любых других менее чем на 0%. Это позволяет убрать часть неинформативных последовательностей и значительно сократить объем выборки для облегчения дальнейших шагов.

5. Выравнивание последовательностей с помощью программы `mafft` v.7.304 [283] со следующими параметрами: алгоритм выравнивания – FFT-NS-2 (прогрессивный метод), штраф за открытие гэпа – 10, штраф за продолжение гэпа – 0.5.

6. Повторное удаление последовательностей, отличающихся от прототипа и любых других менее чем на 2% и более чем на 40%. Это условие позволяет исключить последовательности, не относящиеся к области VP1.

7. Вырезание полного или типизирующего фрагмента VP1 согласно последовательности прототипного штамма.

Реализация алгоритма выполнена в виде программы на языке Python. На вход программе подается файл с параметрами:

- путь к файлу с последовательностями в формате GenBank;
- путь к файлу с последовательностью прототипного штамма;
- тип выравнивания – выравнивание для полного участка VP1 или для его типизирующего фрагмента;
- порог минимального и максимального процентов сходства нуклеотидных последовательностей (используется в шагах 4 и 6).

На выходе программа выдает выравнивание полных или типизирующих фрагментов VP1 в формате «fasta». В новых идентификаторах последовательностей указаны страна и год выделения изолята, если они известны. Программа доступна по ссылке <https://github.com/v-julia/GenAlignment>.

После обработки программой полученное выравнивание инспектируется вручную:

- вырожденные позиции разрешаются согласно консенсусной последовательности выравнивания;
- последовательности с явными ошибками, например, сдвигом рамки, удаляются.

Для обработки вручную использовалась программа BioEdit v.7.2 [285].

Таким образом, были подготовлены выравнивания полных VP1 и его типизирующих фрагментов для 44 типов (Таблица 8). В выравниваниях типизирующего фрагмента, которые в дальнейшем использовались для филогенетического анализа, количество гэпов в последовательности не превышало 10% её длины.

Для типа CVA2 в выборку последовательностей типизирующего фрагмента VP1 были отдельно добавлены три последовательности вирусов, выделенных от пациентов с ОВП, клиническая картина которых соответствовала паралическому полиомиелиту. Эти последовательности были определены в рамках диссертационной работы. Таким образом, в выравнивании типизирующих фрагментов CVA2 содержалось 305 последовательностей.

#### 2.1.4 Подготовка выравнивания полной последовательности VP1 под-типа EV-A71

Для анализа влияния способа формирования выборки последовательностей, а также влияния наличия ошибок в последовательностях и их аннотациях на результаты Байесовского филогенетического анализа, использовалось выравнивание полного участка VP1 вирусов, принадлежащих типу EV-A71. Выравнивание полного VP1 было подготовлено, как описано выше,

за исключением того, что не удалялись схожие последовательности. Итоговое выравнивание содержало 7026 последовательностей. После удаления последовательностей, для которых отсутствовала дата выделения изолята, в выравнивании осталось 6902 полные последовательности VP1.

## **2.2 Определение нуклеотидной последовательности типизирующего фрагмента VP1 изолятов CVA2, выделенных от больных с ОВП**

Типизирующие фрагменты VP1 изолятов, выделенных от пяти пациентов с ОВП, были секвенированы согласно [167]. Последовательность типизирующего фрагмента VP1 была амплифицирована с помощью полимеразной цепной реакции (ПЦР) с обратной транскрипцией со специфическими праймерами к типизирующему фрагменту. Продукты амплификации были визуализированы с помощью электрофореза в 1% агарозном геле на трис-ацетатном буфере и окрашены бромистым этидием. Полоски, соответствующие ПЦР-продуктам, были вырезаны из геля, очищены набором «QIAquick Gel Extraction kit» (фирма QIAGEN). Определение нуклеотидной последовательности ПЦР-продуктов осуществлялось методом Сэнгера на секвенаторе Applied Biosystems 3100.

## **2.3 Анализ рекомбинации**

Для выявления общих закономерностей рекомбинации были рассчитаны матрицы филогенетической совместимости [1,286], реализованные в программе RDP4 [287]. Этот подход заключается в построении филогенетических деревьев для разных областей генома с использованием скользящего окна и вычисляет расстояние Робинсона-Фулдса между ними (Рис. 15а). Для всех исследуемых вирусов размер скользящего окна составлял 600 нт, размер шага — 100 нт для коронавирусов и 50 нт для остальных семейств.

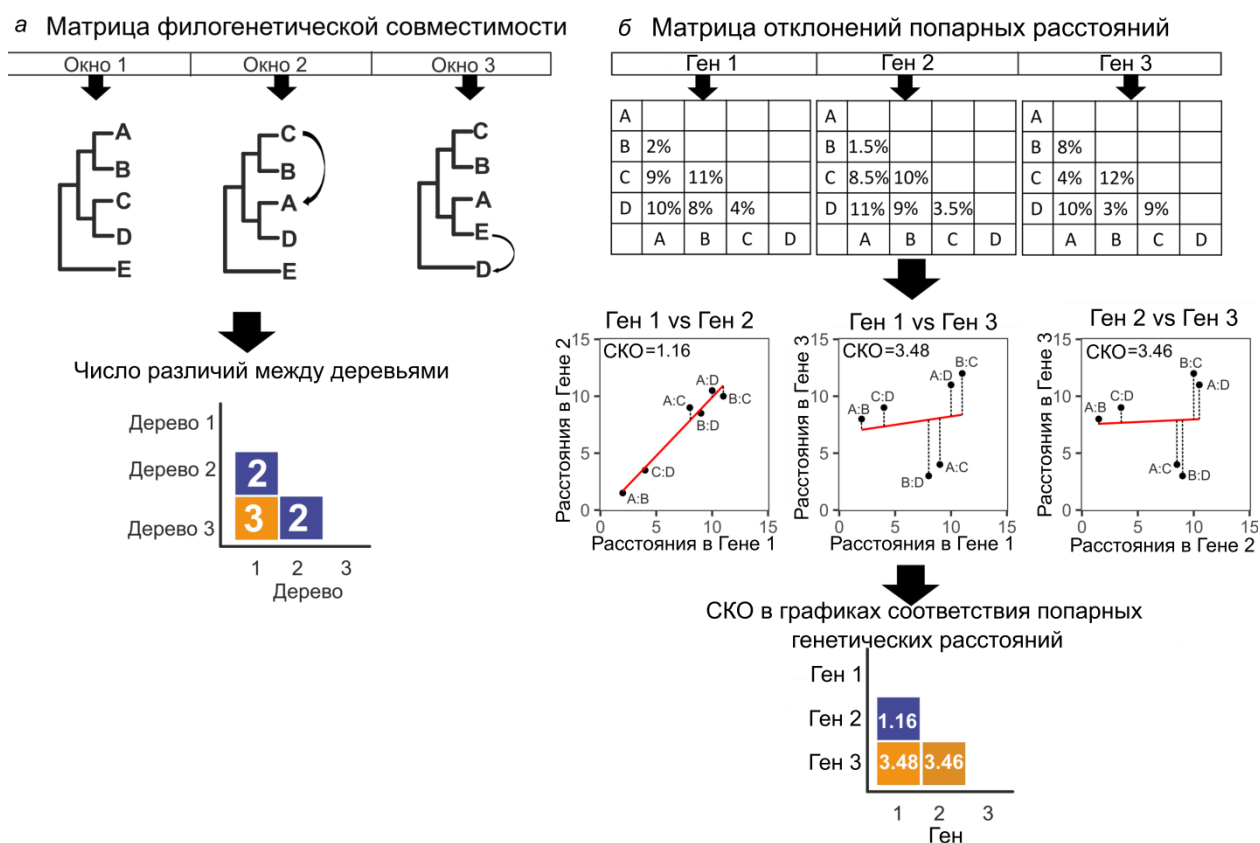


Рис. 15. Принципы работы методов обнаружения рекомбинации. (а) Иллюстрация алгоритма построения матрицы филогенетической совместимости [1,286]. Выравнивание последовательностей разделяется на окна определенной длины. Для каждого окна в выравнивании строится филогенетическое дерево, затем количество различий между полученными деревьями рассчитывается и сохраняется в матрице. Количество различий нормализуют и визуализируют в виде тепловой карты. (б) Алгоритм построения графиков соответствия попарных генетических расстояний (СПГР) и матриц отклонений попарных расстояний (ОПР). Попарные генетические расстояния рассчитываются для интересующих областей генома. Для построения графика соответствия попарных генетических расстояний генетические расстояния между каждой парой последовательностей в области генома 1 (Ген 1) и области генома (Ген 2) отмечаются на осях x и y соответственно. Общее отклонение всех попарных расстояний от линии регрессии оценивается как среднеквадратическое отклонение (СКО). СКО на графиках для различных областей генома составляют матрицы ОПР и визуализируются в виде тепловой карты.

В рамках данной работы был реализован подход для обнаружения и визуализации рекомбинационных событий – график соответствия попарных генетических расстояний (график СПГР). Он основан на проверке того, как накапливаются замены в разных частях генома [100, 233]. В этом подходе строятся матрицы генетических расстояний между последовательностями

для двух участков генома, а затем попарные генетические расстояния наносятся на график. Каждая точка на графике представляет собой расстояния между парой последовательностей в рассматриваемых участках генома. При отсутствии рекомбинации попарные генетические расстояния между парами последовательностей в двух разных регионах следуют линейной зависимости (Рис. 15б, ген 1 против гена 2). Наклон линии регрессии будет зависеть от скорости эволюции в областях генома.

В случае рекомбинации между двумя участками попарные расстояния, принадлежащие рекомбинантной области, будут значительно отклоняться от линии регрессии (Рис. 15б, ген 1 против гена 3, ген 2 против гена 3). Если было много событий рекомбинации, несколько точек будут отклоняться от линии тренда. Степень несогласованности попарных расстояний между двумя областями генома может быть выражена как среднеквадратическое отклонение (СКО) всех попарных расстояний от линии регрессии. Высокое СКО может указывать на относительно большее количество событий рекомбинации с участием двух проанализированных областей генома. Чтобы проиллюстрировать закономерности рекомбинации в геноме в наборе последовательностей вирусов, можно построить матрицы расстояний для всех возможных пар геномных областей с использованием скользящего окна, а СКО для каждой возможной пары этих окон можно визуализировать в виде тепловой карты (Рис. 15б, матрица ОНР).

Построение графиков СПГР и вычисление матриц ОНР было реализовано в виде пакета R «recDplot» (<https://github.com/v-julia/recDplot>) и онлайн-приложения Shiny ([https://v-julia.shinyapps.io/recdplot\\_app/](https://v-julia.shinyapps.io/recdplot_app/)). Обе реализации принимают на вход выравнивание последовательностей вирусов в формате «fasta». На графике СПГР могут быть отображены названия пар последовательностей, соответствующих различным точкам.

В качестве дополнительного метода анализа общих закономерностей рекомбинации были рассчитаны матрицы ОНР со скользящим окном 600 нт, размером шага — 100 нт для коронавируса и 50 нт для остальных семейств.

Во всех расчетах использовались р-дистанции (пропорция отличающихся нуклеотидов у двух последовательностей). Информация о видах вируса и хозяине, полученная из GenBank, использовалась для интерпретации результатов ГСПГР.

Детекция единичных рекомбинационных событий также проводилась с помощью 9 методов, реализованных в программе RDP4: RDP [289], GENECONV [290], Bootscan [291], Maxchi [292], Chimaera [293], SiSscan [294], PhylPro [295], LARD [296], 3Seq [297]. При анализе учитывались только те события рекомбинации, которые поддержаны как минимум пятью методами. Результат работы программы RDP4 содержит названия потенциальных рекомбинантных последовательностей, «главного» и «минорного» родителей и координаты точек рекомбинации. Часть генома рекомбинантной последовательности, полученная от «второстепенного» родителя согласно RDP4, далее именуется рекомбинантным фрагментом. Рекомбинантные фрагменты и количество точек рекомбинации по геному у норовирусов были визуализированы в RStudio (пакет ggplot2 [298]).

Анализ графика СПГР для выявления рекомбинационных событий осуществляется визуально. С целью выявления отличительных особенностей типичных рекомбинантных событий и их визуальной оценки по графикам СПГР была проведена имитация типовых событий рекомбинации и построены СПГР (Рис. 16). В качестве шаблона использовали выравнивания последовательностей вирусов, принадлежащих роду *Alphacoronavirus*. Выравнивание без рекомбинантных последовательностей («отрицательный контроль») моделировали путем объединения всех нечетных позиций выравнивания, а затем всех четных позиций. Таким образом, эффект разной скорости накопления замен в разных областях генома был сведен на нет. Для того чтобы проиллюстрировать влияние длины выравнивания на СКО случайных отклонений попарных расстояний от линии регрессии были созданы выравнивания «отрицательного контроля» длиной 24500, 3500 и 500 нт, что примерно соответствует длине полного коронавирусного генома, полного гена S и его ре-

цептор-связывающего домена (Рис. 16а). Так как самый большой разброс точек наблюдался на графике СПГР для выравнивания длиной 500 нт, это выравнивание использовалось для дальнейшего моделирования событий рекомбинации.

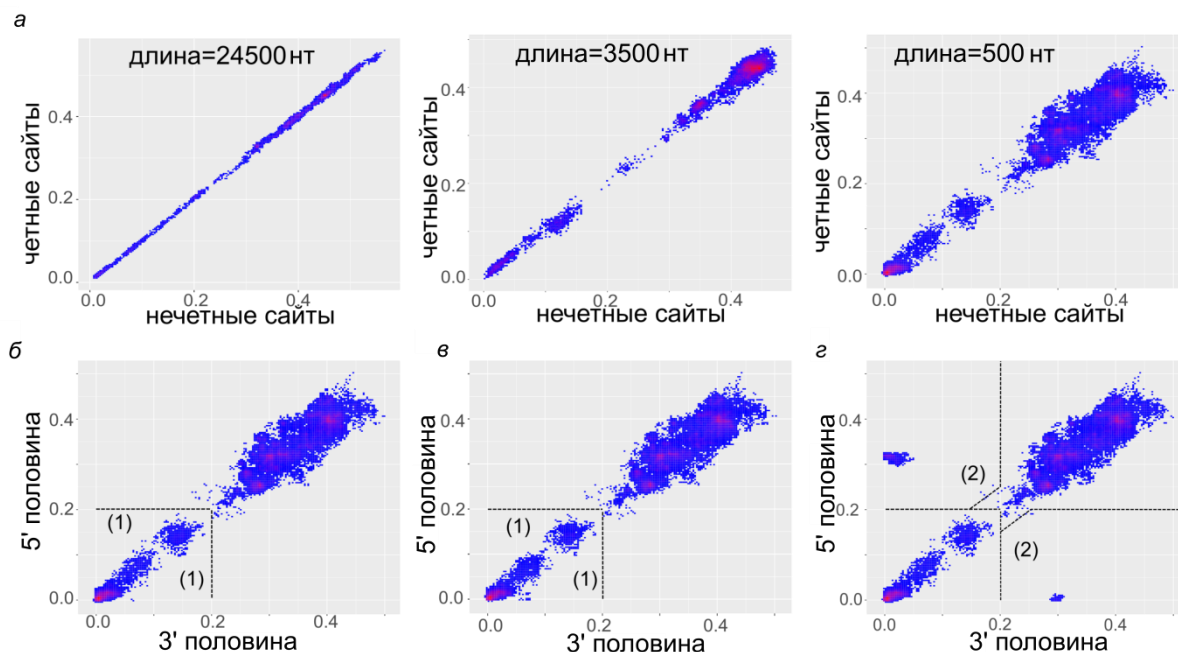


Рис. 16. Типовые графики СПГР. Смоделированы выравнивания без рекомбинантных последовательностей разной длины (а), выравнивания длиной 500 нт с одной рекомбинантной последовательностью (б), несколькими событиями рекомбинации между вирусами из близких клад (в), событиями рекомбинации между отдаленными кладами (г). Пунктирные линии указывают области графика, соответствующие событиям рекомбинации между вирусами, отличающимися менее чем на 20% нуклеотидной последовательности в обоих участках генома (1) («недавние» рекомбинационные события) и событиям, между вирусами, схожими менее чем на 20% («недавние» события между сильно дивергировавшими вирусами).

Далее были сгенерированы следующие выравнивания нуклеотидных последовательностей, содержащие рекомбинантные последовательности:

- выравнивание, содержащее одну последовательность, которая является результатом рекомбинации между близкородственными последовательностями (р-дистанция  $\sim 5\%$ ) (Рис. 16б);
- выравнивание, с несколькими событиями рекомбинации между близкими кладами, отличающимися на 5-15% нуклеотидной последовательности (Рис. 16в);

- выравнивание с событием рекомбинации, которое произошло между вирусами, отличающимися на ~30% нуклеотидов (что соответствует рекомбинации между разными подродами коронавируса) (Рис. 16г).

Для создания единичных событий рекомбинации произвольно выбирали две последовательности, близкие на филогенетическом дереве рода *Alphacoronavirus*, и половину последовательности «реципиента рекомбинации» заменяли соответствующей последовательностью «донора». Несколько событий рекомбинации были смоделированы путем выбора на филогенетическом дереве двух близких клад, принадлежащих одному и тому же виду вируса, и замены половины последовательностей из клады-«реципиента» соответствующими половинами из клады-«донора». Тот же алгоритм применялся при создании событий рекомбинации между кладами вирусов, принадлежащих разным подкладам.

#### **2.4 Методы сокращения выборок последовательностей из референсного выравнивания полных генов VP1**

Одной из задач работы являлось определение влияния алгоритмов сокращения выборок на результаты Байесовского филогенетического анализа. Для решения этой задачи к референсному набору последовательностей VP1 EV-A71 были применены следующие алгоритмы сокращения выборок:

1. Случайная выборка последовательностей, которые были отсеквенированы в отдельных исследованиях. Все последовательности в референсном выравнивании были разделены на группы на основе первых пяти символов идентификатора GenBank. Затем выбиралась случайная группа, и все последовательности из этой группы были добавлены к выравниванию. Этот шаг повторялся до тех пор, пока количество последовательностей не достигало заданного значения. Этот алгоритм воспроизводит ситуацию со взятием проб энтеровирусов 15 лет назад, когда было проведено всего несколько исследований, или имеющуюся в настоящее время выборку последовательностей менее распространенных типов или видов энтеровирусов.



2. Случайная выборка последовательностей («random single»). Определенное количество последовательностей случайно выбирается из исходного набора данных.

3. Фильтрация последовательностей по схожести. Последовательности, которые отличались от любой другой последовательности в наборе данных менее чем на выбранный процент нуклеотидной последовательности, исключались.

4. «Smart picking». Все последовательности в исходном выравнивании разделены на группы на основе первых пяти символов идентификатора GenBank. Все последовательности из групп, размер которых не превышал установленный пользователем порог, включались в окончательный набор данных; из больших групп к сокращенной выборке добавлялась одна последовательность или определенная доля случайно выбранных последовательностей. Этот алгоритм составления выборки позволяет включить уникальные последовательности из небольших исследований и сократить количество последовательностей из массовых эпидемиологических исследований. Для данного исследования использовались следующие параметры алгоритма: группы, содержащие 2 последовательности, включали целиком, из остальных групп случайно выбирался 1% последовательностей. Это позволило значительно сократить количество последовательностей в исходной выборке.

Десять выравниваний были сгенерированы с помощью алгоритма «случайные группы», чтобы оценить влияние количества последовательностей в выравнивании на высоту дерева и скорость накопления замен, выводимые в Байесовском филогенетическом анализе. Также десять выравниваний были сгенерированы остальными методами для анализа влияния алгоритма подготовки выборки на реконструкцию динамики численности вирусной популяции.

Скрипты, которые были использованы для генерирования выравниваний, доступны по ссылке [https://github.com/v-julia/sample\\_bias](https://github.com/v-julia/sample_bias).

## 2.5 Филогенетический анализ

### 2.5.1 Построение деревьев максимального правдоподобия

Деревья нуклеотидных последовательностей ORF1a, ORF1b и S коронавируса, участка, кодирующего капсидный белок VP1, у калицивирусов, участка, кодирующего капсидный белок VP34 у астровирусов были построены методом максимального правдоподобия (ML), реализованного в программе IQ-TREE [299]. Поддержка узлов деревьев рассчитывалась с помощью сверхбыстрой аппроксимации бутстрепа UFBoot (10 000 реплик) [300]. Для каждого дерева выбиралась наиболее подходящая модель нуклеотидных замен согласно Байесовскому информационному критерию (БИК). Выбор модели замен был осуществлен с помощью модуля ModelFinder [301] программы IQ-TREE.

Окрашивание названий таксонов в деревьях ORF1a, ORF1b и S на рисунке 34 было выполнено с использованием скрипта на языке Python ([https://github.com/v-julia/coronavirus\\_recombination](https://github.com/v-julia/coronavirus_recombination)).

Филогенетические деревья были визуализированы с помощью программы FigTree v1.4.4 [302] и пакета ggtree [303] языка программирования R.

### 2.5.2 Байесовский филогенетический анализ норовирусов

Деревья максимального правдоподобия клад (MCC) по ORF2 (VP1) для геногрупп GI (N=71 последовательность) и GII (N=915 последовательностей) норовирусов были построены с помощью программы BEAST v.1.10.4 [304]. Лучшие схемы разбиения (GI: (1+2)(3), GII: (1,2,3)) и модели замен (GTR+I+G+X) для Байесовского филогенетического анализа были выбраны в соответствии с БИК с использованием программы PartitionFinder 2 [305]. Для каждой геногруппы были рассчитаны маргинальные правдоподобия для комбинаций коалесцентных моделей ветвления дерева (постоянный размер популяции, экспоненциальный рост) и моделей молекулярных часов (строгие, ослабленные логнормальные) процедурой Path sampling/Stepping stone, реализованной в BEAST [14]. На основании значения коэффициента Байеса

(BF) была выбрана комбинация коалесцентной модели с постоянным размером популяции и расслабленных логнормальных молекулярных часов для обеих геногрупп ( $\log BF > 10$ ).

Для GI и GII генерировалось 50 млн и 800 млн состояний цепи Маркова, каждые 5000 и 10000 состояний дерева записывались в лог-файл соответственно. Деревья MCC были аннотированы с помощью программы TreeAnnotator v1.8.2, в качестве отжига использовались первые 10% сгенерированных деревьев.

### 2.5.3 Байесовский филогенетический анализ типизирующего фрагмента VP1 типов энтеровирусов

Для Байесовского филогенетического анализа были выбраны 32 типа энтеровирусов, для которых число последовательностей коротких типизирующих фрагментов в выравнивании превышало 35 (Таблица 8). Филогенетический анализ был выполнен с помощью программы Beast v.1.8.3 [306]. Для анализа отдельных типов использовали выравнивания коротких типизирующих последовательностей, а для анализа видов HEV-A и HEV-B – выравнивания полных VP1. Те последовательности, для которых не был указан год выделения ни в файле GenBank, ни в научных статьях, были исключены из выравниваний.

Модель нуклеотидных замен (SRD06), логнормальные молекулярные часы, коалесцентная модель ветвления дерева с постоянным размером популяции были выбраны для филогенетического анализа типов и видов в соответствии с проведенными ранее исследованиями [175]. Модель SRD06 [307] предполагает, что частоты замен происходят согласно шестипараметрической модели НКУ, но с различными параметрами для первой-второй и третьей позиций кодона.

В каждом филогенетическом анализе генерировалось 50 миллионов состояний цепи Маркова, каждые 10000 состояний дерева записывались в лог-файл. Деревья MCC были аннотированы с помощью программы TreeAnnota-

tor v1.8.2, в качестве отжига использовались первые 5 миллионов деревьев. Визуализация деревьев осуществлялась с помощью программы FigTree v1.4.2 [302].

Для статистического анализа была использована программа GraphPad Prism [308].

#### 2.5.4 Байесовский филогенетический анализ полной последовательности VP1 типа EV-A71

Филогенетический анализ выравниваний участка, кодирующего полный VP1, энтеровирусов типа EV-A71, с разными количествами последовательностей, а также полученных с помощью разных методов сокращения выборки (раздел 2.4) был проведен в программе BEAST v.1.10.4 [304]. В анализе использовалась модель замен SRD06, расслабленные логнормальные молекулярные часы, модель ветвления дерева Bayesian Skygrid. Модель Bayesian Skygrid является непараметрической коалесцентной моделью и позволяет выводить изменение эффективного размера популяции во времени. Каждый анализ был проведен до тех пор, пока эффективный размер выборки для всех параметров был не меньше 200. Число состояний в цепи Маркова варьировалось от 50 до 200 млн для разных выравниваний. Деревья МСС были аннотированы с помощью программы TreeAnnotator v1.8.2 с 20% отжигом. Деревья были визуализированы в программе FigTree v1.4.4 [302]. Скорости накопления замен, высоты деревьев и изменения численности популяции были визуализированы с помощью скриптов на языке R.

## **2.6 Симуляция ошибок секвенирования и ошибок в аннотации в выравниваниях**

Для анализа влияния ошибок секвенирования и ошибок в аннотации использовалось выравнивание полных последовательностей VP1 вирусов, принадлежащих генотипу B1 (n=56) типа EV-A71. Последовательности были взяты из референсного выравнивания VP1 EV-A71. Были выбраны две последовательности – близкая к корню AB524092 и последовательность из по-

верхностной клады AF135866. С помощью скрипта на языке Python в последовательности AB524092 и AF135866 были введены 5, 10, 20 мутаций. Также даты выделения AB524092 и AF135866 были увеличены или уменьшены на 5, 10, 15, 20 лет.

Филогенетические деревья максимального правдоподобия были построены программой IQ-TREE v.1.6.12 [299] со сверхбыстрой аппроксимацией бутстрепа UFBoot (10 000 реплик) [300]. Выбор модели замен был осуществлен с помощью модуля ModelFinder [301] программы IQ-TREE. Наличие временного сигнала в исходном и измененных выравниваниях было проверено с помощью программы TempEst v.1.5.3 [309].

Байесовский филогенетический анализ исходного выравнивания, а также выравниваний с введенными мутациями и измененными датами выделения изолятов был проведен в программе BEAST v.1.10.4 [304] с такими же параметрами модели, как для выравнивания всех последовательностей типа EV-A71, за исключением того, что в качестве модели ветвления дерева использовалась коалесцентная модель с постоянной численностью.

Скорости накопления замен на ветвях дерева были выделены из деревьев МСС, распределения скоростей накопления замен были визуализированы с помощью скриптов на языке Python.

Скрипты, которые были использованы для введения мутаций и изменения дат выделения изолятов, визуализации распределений скоростей накопления замен доступны по ссылке [https://github.com/v-julia/sample\\_bias](https://github.com/v-julia/sample_bias).

## **2.7 Вычисление времени полужизни рекомбинантных форм норовирусов**

Для оценки времен полужизни рекомбинантных форм геногрупп GI и GII норовирусов было использовано два подхода. Первый подход описан в работе [146]. Он заключался в том, что сравнивали два филогенетических дерева с разверткой по времени, построенные по нуклеотидным последовательностям VP1 и PзPп (описано в разделе 2.5.2), и отмечали такие узлы, в поддеревьях которых группирование последовательностей совпадает в двух

деревьях. Затем среди всех отмеченных узлов выделялись самые близкие к корню, и рассчитывалась медиана возраста таких узлов. Это значение принималось за время полужизни рекомбинантных форм. Подход применялся для вычисления времени полужизни рекомбинантных форм как GI, так и GII.

Второй подход был введен в работах [176,177] и использовался для вычисления времен полужизни рекомбинантных форм в геногруппе GII. Для всех пар последовательностей геногруппы GII с одинаковым генотипом VP1 вычислили MCL-расстояние между последовательностями VP1 в программе MEGA v.7 [310]. С помощью скрипта на языке Python рассчитывались доли сравнений, в которых изоляты имели одинаковые генотипы по VP1 и разные P-типы (генотипы по PзPп) для диапазонов MCL-дистанций (Рис. 21). С помощью регрессионного анализа, выполненного средствами модуля sklearn [311] языка Python, вычислялось MCL-расстояние, при котором 50% пар сравниваемых вирусов имеют одинаковый генотип по VP1 и разные P-типы ( $MCL\ distance_{50\%}$ ). Период полужизни рекомбинантных форм ( $T$ ) вычислялся по формуле:

$$T = \frac{MCL\ distance_{50\%}}{2 * rate},$$

где  $rate$  – скорость накопления замен, рассчитанная при Байесовском филогенетическом анализе ( $4,34 \times 10^{-3}$  з/с/г) (раздел 2.5.2). Деление MCL-расстояния на скорость накопления замен соответствует времени дивергентной эволюции двух последовательностей, а деление на два производится для того, чтобы вычислить время дивергенции двух последовательностей от общего предка.

## 3 РЕЗУЛЬТАТЫ

### 3.1 Метод для детекции и визуализации рекомбинации

Общепринятые инструменты анализа рекомбинации позволяют анализировать общую распространенность событий рекомбинации путем детекции и визуализации консервативных филогенетических групп в геноме (матрицы филогенетической совместимости [1] и сканирование TreeOrder [312]), а также обнаруживать отдельные события рекомбинации (например, различные алгоритмы детекции рекомбинации, реализованные в программе RDP4 [287]). В данной работе матрицы филогенетической совместимости (Рис. 15а) использовались в качестве референсного метода для общего анализа распространенности рекомбинации. Недостатком этого метода является то, что матрицы не указывают «глубину» событий рекомбинации, то есть насколько близкородственны вирусы, между которыми проходила рекомбинация, а показывают только относительную распространенность рекомбинации в геноме. Программа TreeOrder позволяет визуализировать события рекомбинации на основе predetermined групп последовательностей, поэтому не подходит для исследования рекомбинации без гипотезы и не применялась в данном исследовании. Основной причиной, обусловившей необходимость разработки нового подхода, было то, что перечисленные методы основаны на филогенетическом группировании, которое выявляет изолированные события рекомбинации, но может не работать в случае многочисленных событий рекомбинации внутри сравниваемых фрагментов, как будет показано ниже на примере рода *Gammacoronavirus*.

Одной из задач данной работы была разработка метода детекции и визуализации рекомбинации, который бы позволил выявлять многочисленные события, а также показывал бы «глубину» этих событий. Для этого был разработан и реализован метод детекции рекомбинации при сравнении двух участков генома, использованный в [146]. Метод заключается в вычислении попарных расстояний между последовательностями вирусов в разных регионах генома. В отсутствие рекомбинации замены будут пропорционально накап-

ливаться в двух областях генома (примеры симуляций на Рис. 16а). На графиках, построенных по реальным нерекомбинантным последовательностям, может быть небольшое отклонение от линии тренда из-за различного давления отбора в разных областях генома и стохастического накопления мутаций.

Даже одно событие рекомбинации приводит к отклонению нескольких точек от линии регрессии, которые соответствуют сравнениям рекомбинанта с его «родителями», а также близких к ним последовательностей (Рис. 16б). Множественные недавние события рекомбинации создают группы точек на близких генетических расстояниях (Рис. 16в), в то время как давние события рекомбинации отражаются во многих последовательностях потомков и создают «группы облаков» точек (Рис. 16г). В рамках работы было разработано веб-приложение ([https://v-julia.shinyapps.io/recdplot\\_app](https://v-julia.shinyapps.io/recdplot_app)), которое позволяет строить графики соответствия попарных генетических расстояний (СПГР) и отображать названия пар вирусов, соответствующих конкретным точкам, для быстрой идентификации потенциальных рекомбинантов. Координаты (генетические расстояния) отклоняющихся от линейного тренда точек могут указывать на время рекомбинационных событий в прошлом. Например, частота замен у коронавирусов варьируется в зависимости от таксона вируса, его хозяина и участка генома от  $10^{-4}$  до  $6,0 \times 10^{-4}$  з/с/г [313]. Грубая средняя оценка скорости накопления замен  $5 \times 10^{-4}$  з/с/г предполагает, что самый недавний общий предок коронавирусов, отличающихся на 1% по нуклеотидной последовательности, существовал около десяти лет назад. Эту оценку следует использовать с осторожностью, поскольку разные вирусы могут иметь разную скорость накопления замен [314], а при больших генетических расстояниях возможны ошибки молекулярного датирования на несколько порядков из-за насыщения замен [315].

Отдельные графики СПГР могут выявить рекомбинацию между двумя участками генома. Для того чтобы глобально оценить несогласованность попарных генетических расстояний, можно отразить множество графиков СПГР для всех возможных пар областей генома в виде матрицы отклонений попарных расстояний (ОПР). В рамках этого подхода матрицы генетических



расстояний строятся для всех участков генома с использованием скользящего окна, затем для каждой возможной пары участков (один график СПГР) вычисляется СКО точек от линии регрессии. Затем СКО для всех возможных пар регионов генома визуализируются в виде тепловой карты (Рис. 15б). Высокое СКО указывает на более низкую общую согласованность расстояний в двух областях генома, что может быть вызвано рекомбинацией. Метод нечувствителен к потере филогенетического сигнала и позволяет обнаруживать частые события рекомбинации, что обычно ограничивает чувствительность многих классических методов. Данный метод был разработан в рамках диссертационной работы и использовался для глобальной оценки частоты рекомбинации в геномах вирусов наряду с матрицами филогенетической совместимости.

Необходимость разработки этого метода была обусловлена очень высокой частотой рекомбинации у многих РНК-вирусов, ограничивавшей применение классических подходов для анализа. Затем новый метод был использован в комбинации с классическими подходами для систематического изучения роли рекомбинации в эволюции вирусов, принадлежащих нескольким широко распространенным и социально значимым семействам РНК-вирусов – *Caliciviridae*, *Astroviridae* и *Coronaviridae*.

## **3.2 Рекомбинация у калицивирусов**

### **3.2.1 Рекомбинация у норовирусов**

Для того чтобы изучить закономерности рекомбинации норовирусов по геному и геногруппам, были использованы все полногеномные нуклеотидные последовательности норовирусов, доступные в базе Genbank. После удаления почти идентичных последовательностей окончательный набор данных содержал 1084 полных генома. Распределение последовательностей по геногруппам было неравномерным (Рис. 17). Большинство последовательностей рода *Norovirus* (915 из 1084) принадлежало геногруппе GII, вирусы которой заражают исключительно человека. Более половины последовательностей GII принадлежали к генотипу GII.4 (N=564), который был причиной

большинства вспышек норовирусов на протяжении более двух десятилетий [65]. Чтобы учесть неравномерность выборки, были проведены три анализа – вирусов всех геногрупп, всех геногрупп, кроме GII и вирусов геногруппы GII.

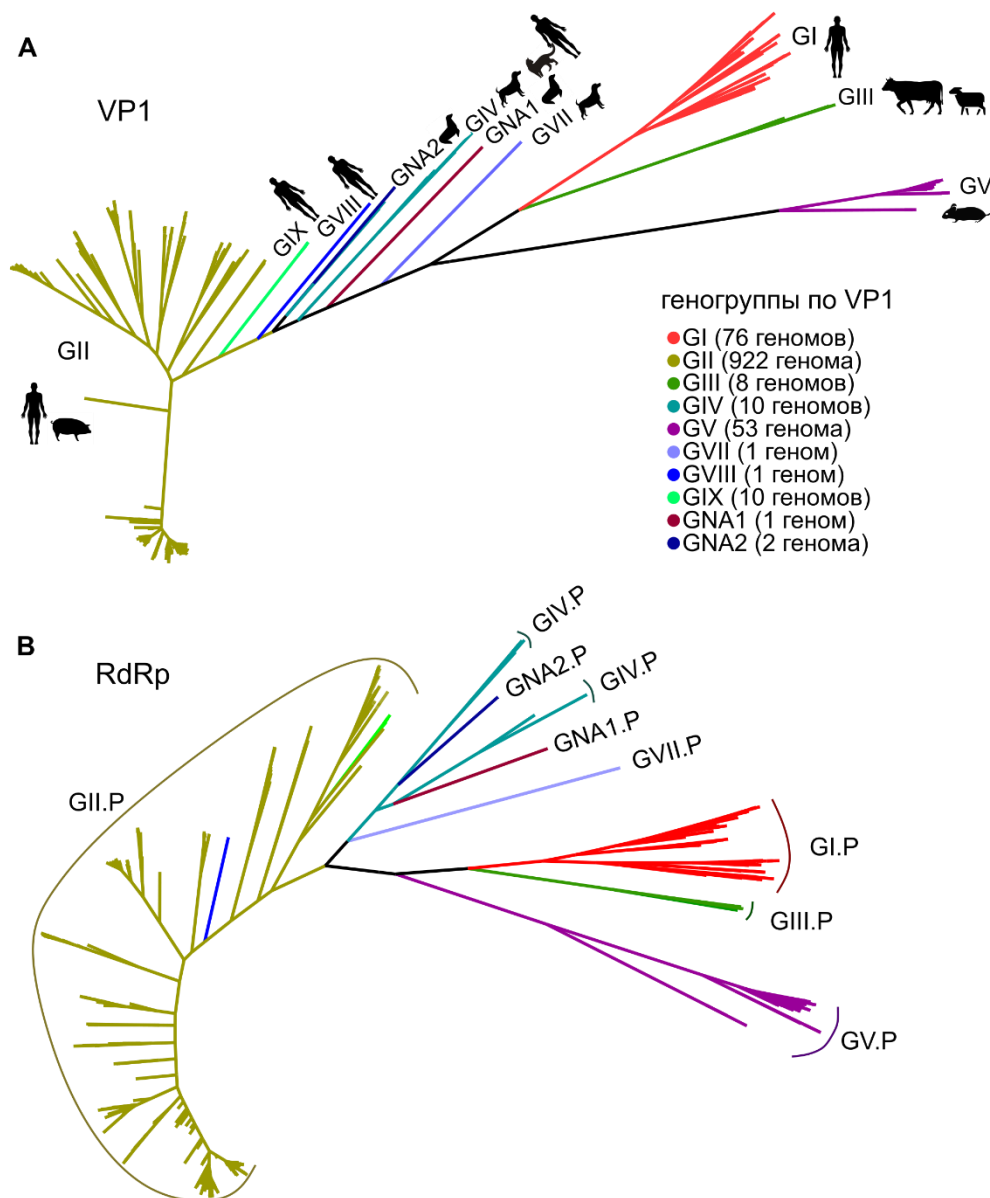


Рис. 17. Филогенетическое дерево, построенное по нуклеотидной последовательности капсидного белка VP1 (а) и РзРп (б) вирусов рода *Norovirus*, для которых проводился анализ рекомбинации. Дерево построено алгоритмом ML в программе IQ-TREE [299]. Силуэты основных хозяев вирусов разных геногрупп показаны рядом с соответствующими листьями дерева. Таксономия соответствует обновлению ICTV 2021. Количество последовательностей для каждой геногруппы указано в скобках в легенде рисунка. Ветви дерева покрашены по геногруппе VP1 в обоих деревьях, Р-группы отмечены на дереве для РзРп.

Для анализа глобальных закономерностей обмена областями генома были рассчитаны матрицы филогенетической совместимости и матрицы ОНР. Матрица филогенетической совместимости показала, что филогенетические деревья, построенные по ORF1, были наиболее несовместимы с областями ORF2-ORF3 (Рис. 18а). Это подтверждает, что стык ORF1/ORF2 является «горячей точкой» рекомбинации (стрелки на Рис. 18а) [87]. Матрицы ОНР в целом согласовывались матрицам филогенетической совместимости (Рис. 18б).

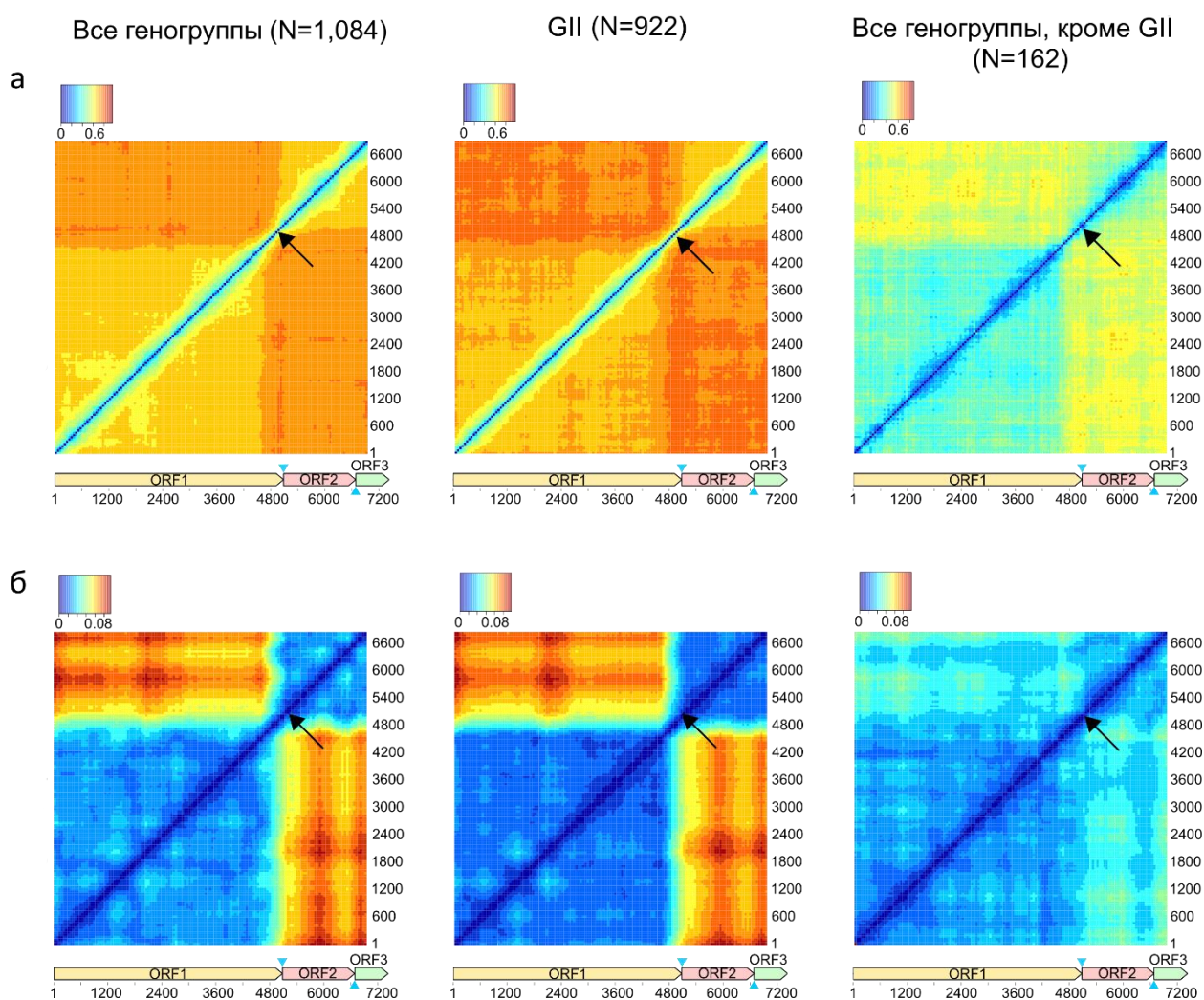


Рис. 18. а – матрицы филогенетической совместимости на основе нормализованных расстояний Робинсона-Фулдса, б – матрицы ОНР для вирусов рода *Norovirus*, принадлежащих всем геногруппам, только геногруппе GI, всем геногруппам, кроме GI. Цвета отражают значения нормированных расстояний Робинсона-Фулдса и СКО в графиках СПР соответственно. Голубые треугольники показывают перекрытия ORF1 и ORF2 (17 нт), ORF2 и ORF3 (1 нт), которые были дублированы при выравнивании конкатенированных ОНР. Черные стрелки указывают на стык ORF1/ORF2.

Поскольку матрицы филогенетической совместимости указывают только на относительную распространенность рекомбинации, для подробного исследования событий рекомбинации были использованы все доступные методы анализа рекомбинации в программе RDP4. Большинство рекомбинантных фрагментов составляли почти полные области, кодирующие VP1 и VP2. Подобные события рекомбинации целого участка ORF2-ORF3 были обнаружены во всех проанализированных геногруппах (Рис. 19). Во всех ОРС также присутствовали рекомбинантные фрагменты меньшего размера от 55 до 1500 нт. Рекомбинационные события были обнаружены в разных частях структурного участка генома и затрагивали вирусы, принадлежащие одинаковым геногруппам: события в VP1 (ORF2) были обнаружены в вирусах GI, GII и GV, а в VP2 – в вирусах из геногрупп GI, GII, GIV и GV.

Большинство рекомбинантов были результатом внутригеногрупповых рекомбинаций, и все события рекомбинации, кроме одного, произошли между вирусами, заражающими один и тот же вид хозяина. Исключением были вирусы GIV кошек и собак, у которых VP1 был наиболее близок к норовирусу GIV, выделенному от человека. Также было детектировано несколько событий межгеногрупповой рекомбинации, большинство из которых происходило между норовирусами человека из GII, GI, GVIII и GIX. Только два события межгеногрупповой рекомбинации произошло на стыке ORF1/2. То, что вирусы GVIII являются результатом рекомбинации норовирусов из GII и неизвестных норовирусов, согласовывалось с разным группированием изолятов GVIII на филогенетических деревьях VP1 и РзРп (Рис. 17). Большинство событий межгеногрупповой рекомбинации происходили внутри ОРС (Рис. 19). Вирусы, участвовавшие в межгеногрупповой рекомбинации, отличались на 30-50% по нуклеотидной последовательности. Например, был детектирован межгеногрупповой рекомбинант, ORF1 которого ближе всего к норовирусам GNA2, а ORF2-3 – к норовирусам GIV. Он отличался от GIV на 35% и от GNA2 на 45% по нуклеотидной последовательности ORF2-3. Таким образом, наблюдаемые межгеногрупповые события рекомбинации скорее всего

произошли до того, как вирусы дивергировали и стали относиться к разным геногруппам.

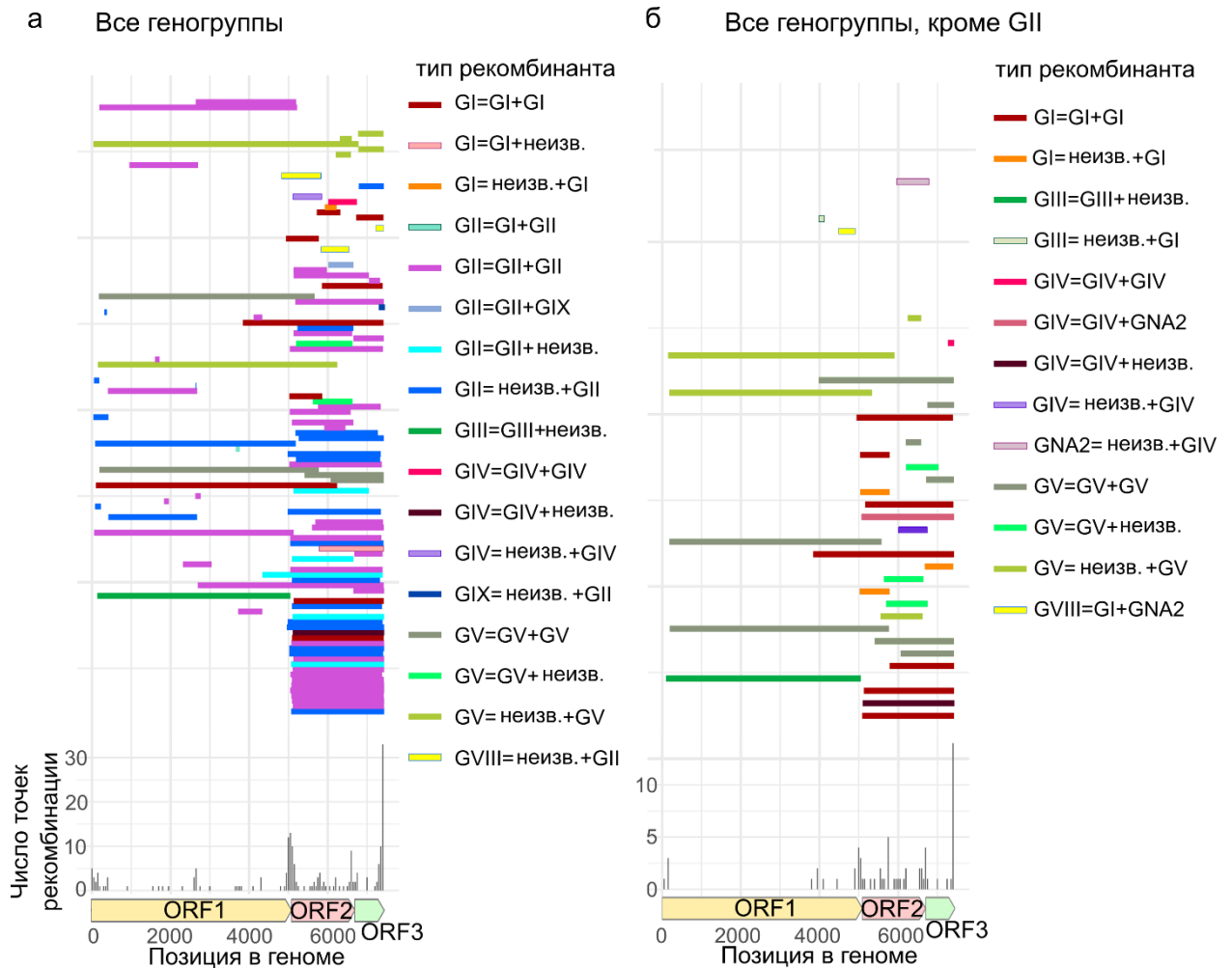


Рис. 19. Рекомбинантные фрагменты, обнаруженные с помощью программы RDP4, в двух наборах данных: а – набор данных, включающий все геногруппы норовирусов, б – набор данных, включающий все геногруппы, кроме GII. Тип рекомбинанта определяется следующим образом: «геногруппа рекомбинанта = минорный родитель + главный родитель».

Несмотря на то, что методы, реализованные в RDP4, позволяют обнаружить события рекомбинации, они не показывают, насколько давно эти события произошли. Для того чтобы оценить пределы генетических различий между вирусами, вовлеченными в рекомбинацию, были построены графики СПГР, которые визуализируют отношения генетических расстояний в двух областях генома.

Как матрицы филогенетической совместимости, так и анализ RDP4 указывают на умеренный уровень рекомбинации в GII и низкий уровень ре-

комбинации в других геногруппах в неструктурных генах (ORF1). Действительно, на графиках СПГР в двух половинах ORF1 все отклоняющиеся попарные расстояния принадлежали вирусам GII, что полностью согласуется с результатами анализа в RDP4 (Рис. 19). Среди них можно отметить недавние рекомбинационные события между вирусами из GII.4\_GII.P4, GII.4\_GII.P31 (Рис. 20а, симметричные красные круги). Также от общего тренда отклонялись точки, соответствующие сравнениям вирусов, принадлежащих GII.3, GII.13, GII.21 и GII.P21 с вирусами из GII.4\_GII.P31 (Рис. 20а, зеленый круг), но для них не было найдено симметричных точек, что говорит о том, что для этой группы неизвестен второй партнер рекомбинации.

Графики СПГР в неструктурных (ORF1) и структурных (ORF2, ORF3) областях генома характеризуются наиболее заметным несоответствием попарных расстояний (Рис. 20б). Попарные генетические дистанции между вирусами, принадлежащими GII, образовывали симметричное облако в пределах до 40% различий в нуклеотидной последовательности, что может свидетельствовать о многочисленных событиях рекомбинации. Пары GII, чьи дистанции отклонялись от линии регрессии, принадлежали различным генотипам. В GII многие события рекомбинации произошли недавно, поскольку вирусы различались всего на несколько процентов в нуклеотидной последовательности в одной из геномных областей. При предполагаемой скорости накопления замен  $4,34 \times 10^{-3}$  з/с/г (см. ниже) один процент генетического расстояния соответствует примерно 2.3 годам. В других геногруппах рекомбинация была менее распространена, что согласуется с результатами RDP4 (Рис. 19б). В рекомбинации участвовали как отдаленно родственные вирусы с расстояниями по нуклеотидной последовательности от 20% до 40% (представители геногрупп GI, GIII, GIV), так и умеренно родственные вирусы с нуклеотидными дистанциями менее 20% (GV, GI) (Рис. 20б). Таким образом, структурные и неструктурные белки достаточно дивергентных норовирусов разных генотипов совместимы между собой, а ко-инфекция одной клетки та-

кими вирусами происходила достаточно регулярно для того, чтобы обеспечить наблюдаемую высокую частоту рекомбинации.

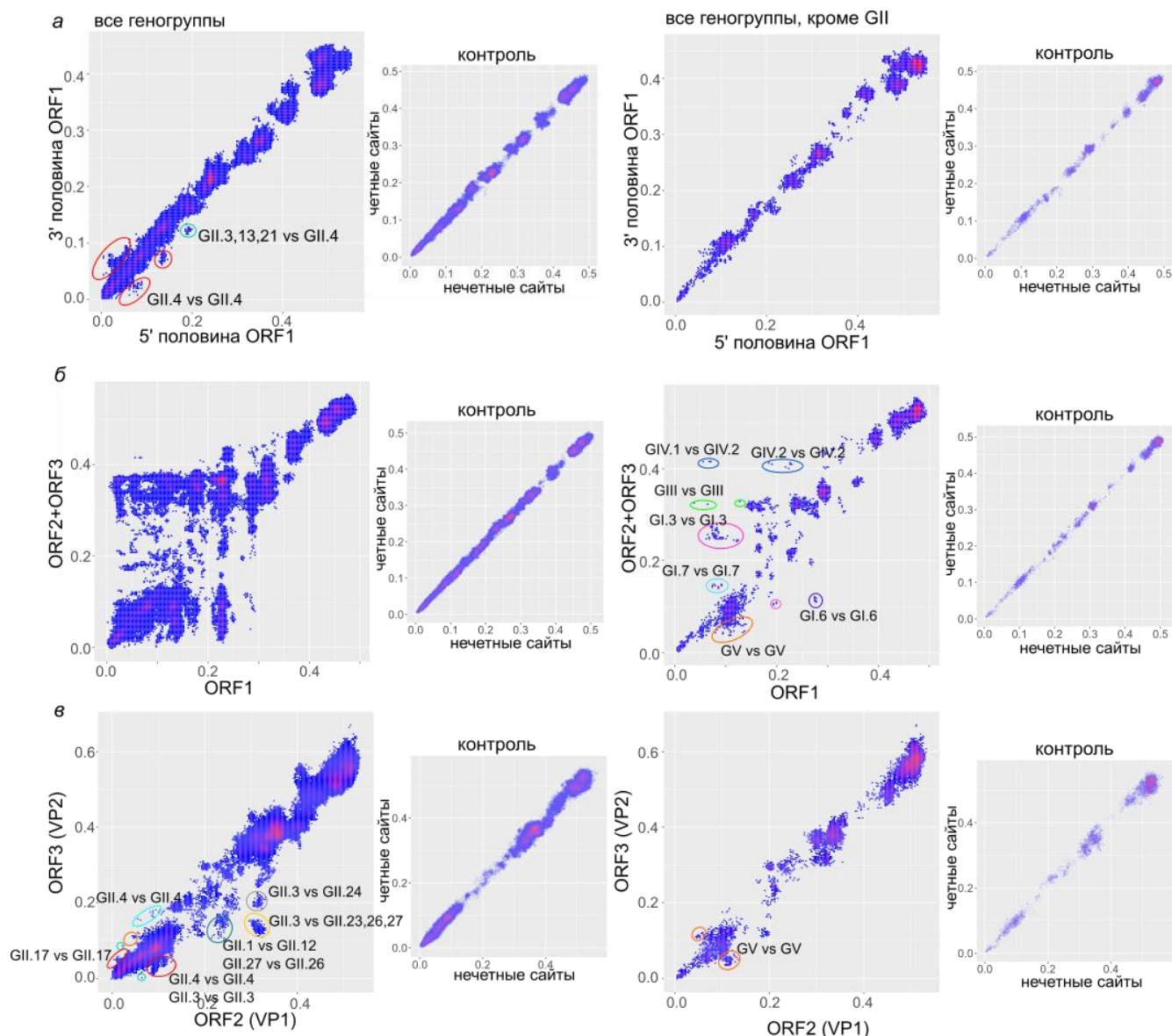


Рис. 20. Соответствие попарных нуклеотидных расстояний в 5'- и 3'-половинах ORF1 (а), ORF1 и слитых ORF2-ORF3 (б), а также между ORF2 и ORF3 (в), кодирующими структурные белки VP1 и VP2 соответственно, у норовирусов. Точки, которые соответствуют парам вирусов, подвергшимся рекомбинации, и обсуждаются в тексте, отмечены кружками. Контрольные графики соответствия между нуклеотидными расстояниями в четных и нечетных позициях выравниваний имитируют графики СПГР в отсутствии рекомбинантных последовательностей.

Рекомбинация между VP1 и VP2 обнаружена у вирусов, принадлежащих GII, GV и GI (Рис. 20в). В отличие от рекомбинации между ORF1 и ORF2, которая, как правило, происходит на стыке ORF1-2, рекомбинация между ORF2 и ORF3 происходила где угодно внутри этих ОРС (Рис. 19б).

Графики СПГР в областях, кодирующих VP1 и VP2, приведены в качестве примера. Практически все события рекомбинации, выделяемые на графике СПГР, совпали с результатами программы RDP4. Среди них можно выделить группу рекомбинантных норовирусов мышей GV (Рис. 20в, оранжевые круги), отличающихся от одного «родителя» примерно на 10% нуклеотидной последовательности в одном участке и на 5% в другом. В GII в рекомбинации между ORF2 и ORF3 участвовали как генетически близкие вирусы с генетическими расстояниями менее 3% (Рис. 20в, красные круги), так и более дивергировавшие вирусы с расстояниями от 20 до 35% (Рис. 20в, темно-зеленый, желтый, серый кружки). Однако для GII, в отличие от GI и GV, по графикам СПГР нельзя было точно определить, какие именно последовательностями являлись рекомбинантными, потому что на графиках отсутствовали симметричные отклоняющиеся точки.

Генетические расстояния дают лишь приблизительную оценку временной динамики рекомбинации. Для более точной оценки времени полужизни рекомбинантных форм в геногруппе GII была определена связь между дивергенцией последовательностей VP1 и вероятностью рекомбинации так, как это было сделано ранее для энтеровирусов [176,177]. Для пар изолятов геногруппы GII с одинаковым генотипом VP1 вычисляли MCL-расстояние между нуклеотидными последовательностями VP1. Геномы с одинаковым генотипом VP1, но разными генотипами по РзРп (Р-типами) считались рекомбинантными. Доля сравнений, в которых изоляты имели разные Р-типы, увеличивалась с увеличением дивергенции VP1, достигая 100% при расстояниях MCL более 0.2 (Рис. 21). Регрессионный анализ дистанций и доли рекомбинантных вирусов показал, что расстояние MCL, равное 0.0982, соответствует 50% рекомбинантных форм среди пар с соответствующим расстоянием в VP1. Время полужизни рекомбинантных форм было вычислено путем деления этого значения на скорость накопления замен в гене VP1, полученной с помощью BEAST ( $4.34 \times 10^{-3}$  з/с/г), и на два. Предсказанное время полужизни рекомбинантных форм для GII составило 11.3 года.



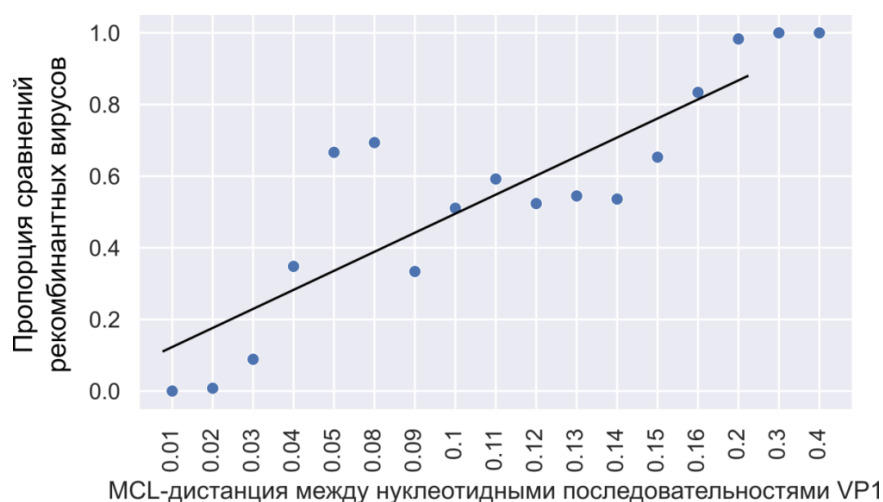


Рис. 21. Регрессионный анализ пропорции попарных сравнений норовирусов GII с одинаковым генотипом VP1, но разными Р-типами, и MCL-расстояния между нуклеотидными последовательностями VP1 этих вирусов.

В GI большинство событий рекомбинации приводили к изменению положения целых клад на дереве. Поскольку предыдущий метод этого не учитывает, период полужизни рекомбинантных форм определяли как медианный возраст узлов от корня, которые ведут к кладам с нерекомбинантными вирусами (с идентичной группировкой в РзРп и VP1), как описано в [146]. Время полужизни рекомбинантных форм, измеренное таким способом, для GI составило 10.44 лет (Рис. 22), а для GII – 8.4 лет, что согласуется с результатами первого метода. Возраст корня деревьев и скорости накопления замен GI и GII, выведенные в результате Байесовского филогенетического анализа, были в тех же пределах, что и в других исследованиях [316–318].

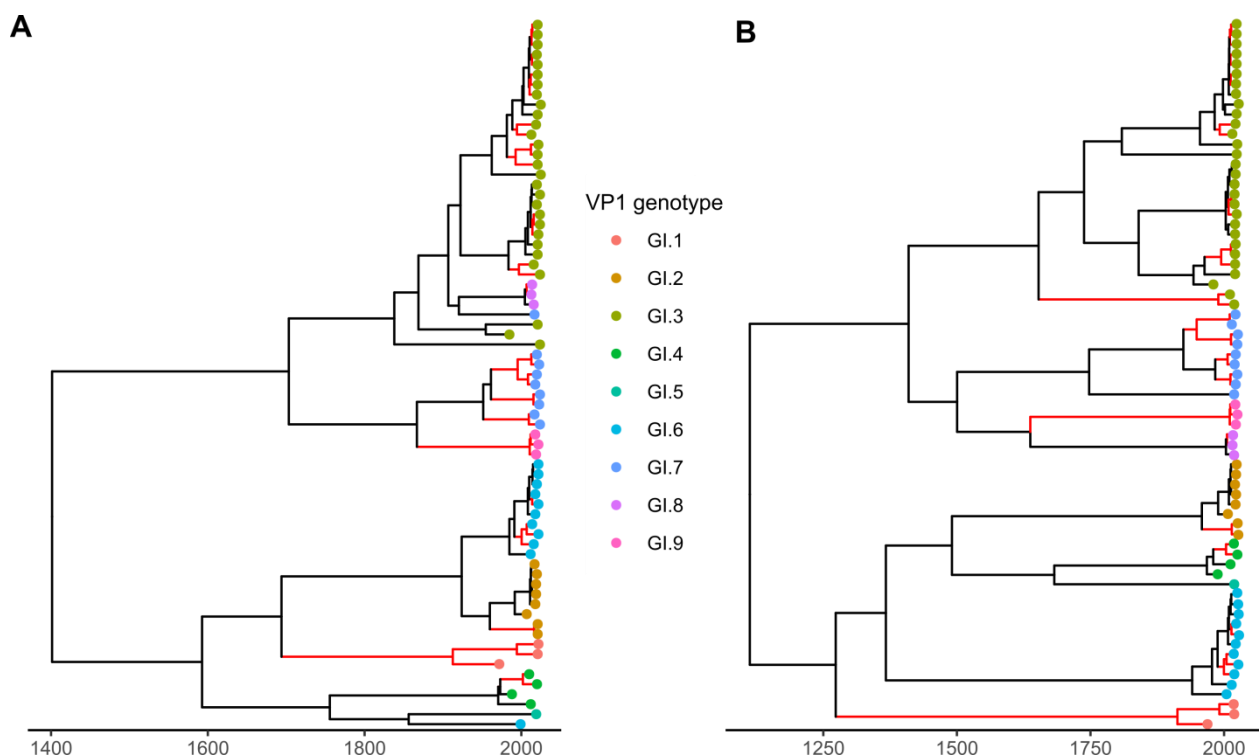


Рис. 22. Филогенетические деревья максимального правдоподобия клад для геногруппы GI, построенные по нуклеотидной последовательности VP1 (а) и полимеразы (б). Красным отмечены совпадающие клады, возраст которых использовался для вычисления времени полужизни рекомбинантных форм. Так как все узлы имели апостериорную вероятность более 0.9, поддержка ветвей не указана на филогенетических деревьях. Листья дерева покрашены по генотипу капсида.

### 3.2.2 Рекомбинация у саповирусов

В анализе естественной рекомбинации вирусов рода *Sapovirus* было задействовано 148 полногеномных последовательностей вирусов, принадлежащих геногруппам GI-GVIII, GX, GXI, GXIII, GXIV (Рис. 23). Больше всего последовательностей было доступно для геногрупп GI и GII, вирусы которых заражают человека, в то время как для некоторых геногрупп в базе GenBank доступны единицы последовательностей (GVIII, GX).

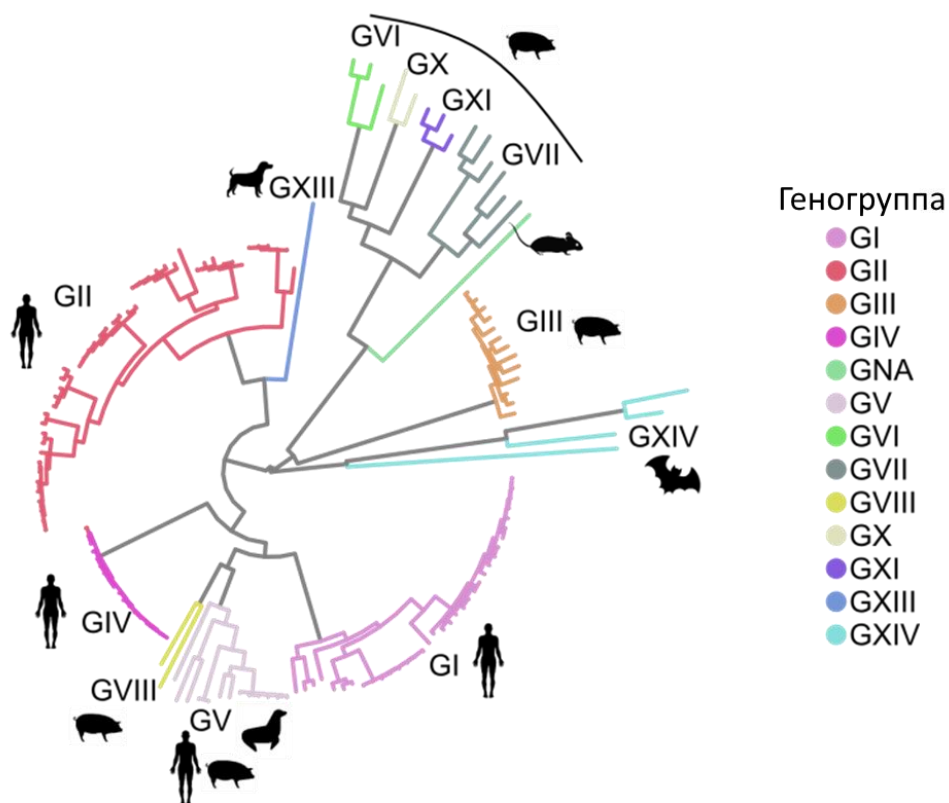


Рис. 23. Филогенетическое дерево капсидного белка VP1 вирусов рода *Sapovirus*, использованных для анализа рекомбинации. Дерево построено по нуклеотидной последовательности VP1 алгоритмом ML в программе IQ-TREE [299]. Ветки дерева покрашены согласно геногруппам саповирусов. Силуэты основных хозяев вирусов разных геногрупп показаны рядом с соответствующими листьями дерева. Таксономия соответствует обновлению ICTV 2021.

Как матрица филогенетической совместимости, так и матрица ОПР указывали на высокую совместимость филогенетических деревьев и нуклеотидных дистанций, построенных по участкам внутри неструктурной (часть ORF1) или структурной областей генома (часть ORF1 и ORF2), и низкую совместимость деревьев и нуклеотидных дистанций при сравнении структурной и неструктурной областей (Рис. 24).

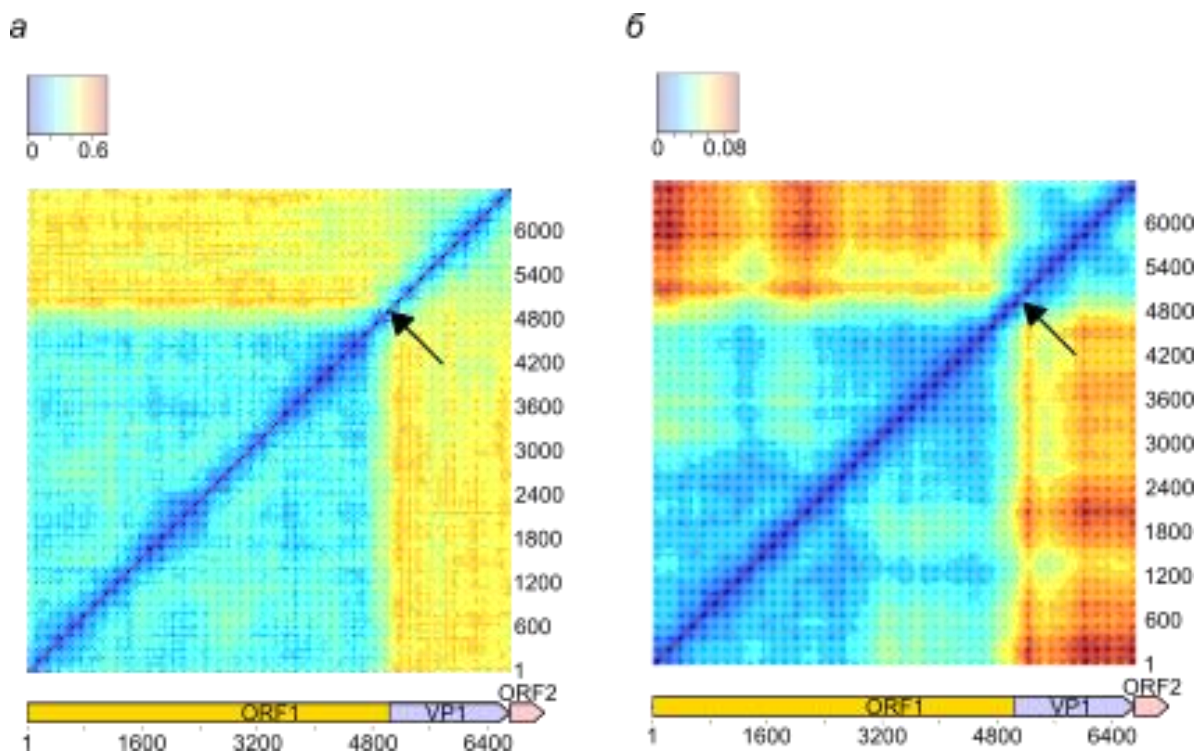


Рис. 24. а – матрицы филогенетической совместимости на основе нормализованных расстояний Робинсона-Фулдса, б – матрицы ОПР для вирусов рода *Sapovirus*. Цвета отражают значения нормированных расстояний Робинсона-Фулдса и СКО в графиках СПГР соответственно. Черные стрелки указывают на стык НСО и СО генома.

В графиках СПГР в двух половинах неструктурной области (НСО) и двух половинах структурной области (СО) генома наблюдается линейная зависимость между попарными нуклеотидными дистанциями, близкая к таковой на контрольных графиках (Рис. 25а,в). Напротив, при сравнении нуклеотидных дистанций, рассчитанных по НСО и СО, наблюдаются пары вирусов, попарные дистанции которых отклоняются от линейного тренда (Рис. 25б). Рекомбинации между очень близкими вирусами у саповирусов не наблюдалось, что резко отличало их от норовирусов. Самое недавнее событие рекомбинации произошло между вирусами GIII, заражающими свиней: вирусы отличались на 6-7% нуклеотидной последовательности в НСО и на 16-17% в СО (Рис. 25б, голубой кружок). Среди относительно близких вирусов можно отметить пары вирусов, принадлежащие GI (хозяин – человек), отличающиеся на 5-12 % в НСО и 26-28% в СО (Рис. 25б, фиолетовый круг).

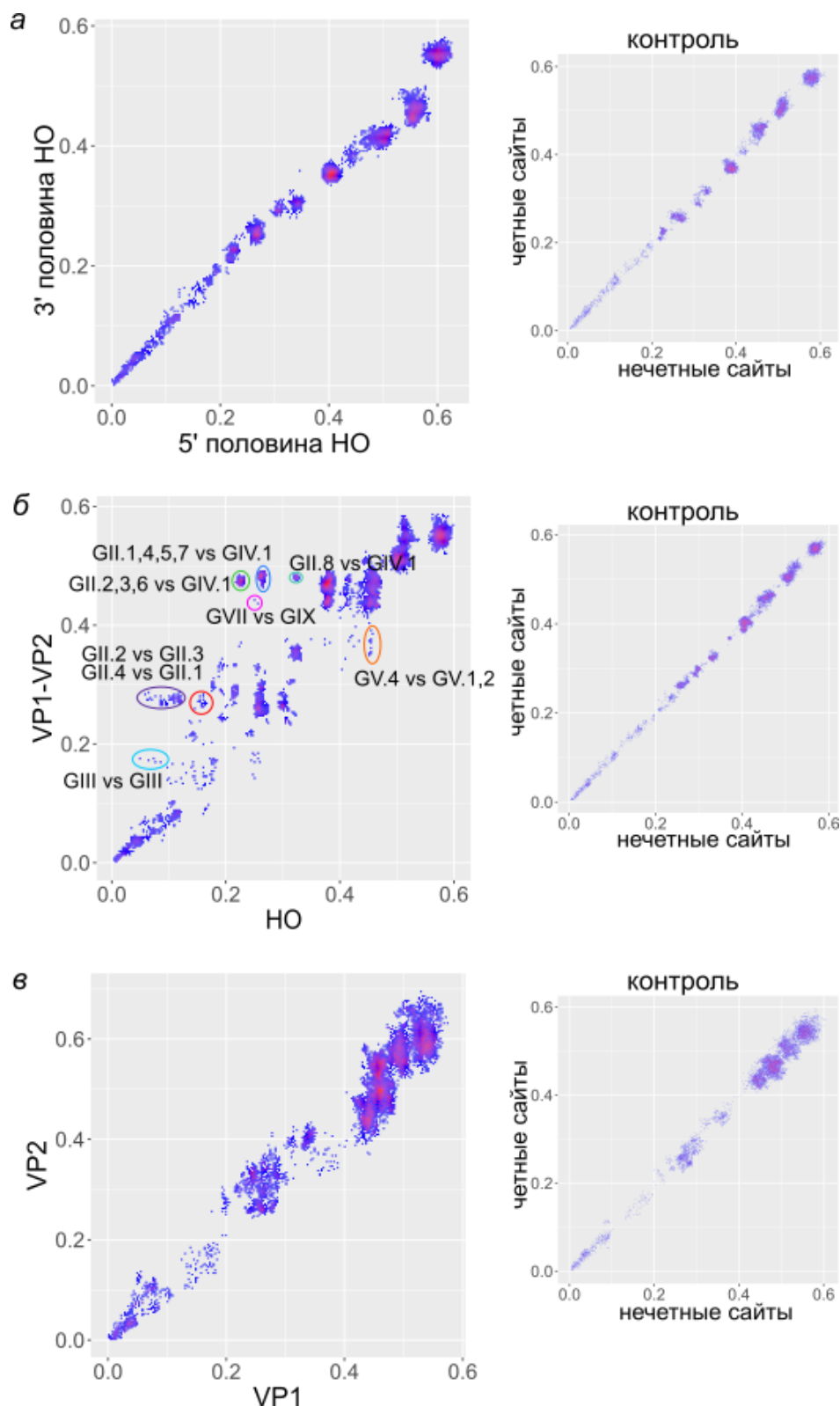


Рис. 25. Соответствие попарных нуклеотидных расстояний между 5'- и 3'-половинами неструктурной области генома (а), неструктурной и структурной областей генома (б), а также между VP1 и VP2 (в) у саповирусов. Точки, которые соответствуют парам вирусов, подвергшимся рекомбинации, и обсуждаются в тексте, отмечены кружками. Контрольные графики соответствия между нуклеотидными расстояниями в четных и нечетных позициях

выравниваний имитируют графики СПГР в отсутствии рекомбинантных последовательностей.

Также отклонения от регрессионной прямой давали пары вирусов, принадлежащие разным геногруппам: GII и GIV (Рис. 25б, салатовый, зеленый, голубой круги), заражающим людей, а также GVII и GIX (Рис. 25б, розовый круг), заражающим свиней. Саповирусы, участвовавшие в рекомбинации, отличались на 20-30% по нуклеотидной последовательности НО и на 40-50% по последовательности СО, что свидетельствует о том, что рекомбинация скорее всего произошла до дивергенции вирусов на разные геногруппы, как и у норовирусов.

Таким образом, все пары вирусов с отклоняющимися от линейного тренда дистанциями имели более сходные неструктурные области, чем при пропорциональном накоплении замен в двух участках генома. Единственное исключение – сравнения вируса JN420370 из морского льва с другими вирусами геногруппы GV, выделенными из свиньи и человека.

### 3.2.3 Рекомбинация у лаговирuсов

В базе данных GenBank было доступно 507 полногеномных последовательностей лаговирuсов. После исключения практически идентичных последовательностей в выборке осталось 338 полногеномных последовательностей, принадлежащих двум геногруппам – GI, включающей вирусы геморрагической болезни кроликов (RHDV) и калицивирусы кроликов (RCV), и GII, включающей вирусы EBHSV и HaCV (Рис. 26). В связи с бóльшей ветеринарной значимостью, GI лучше представлена в выборке, чем другие геногруппы.

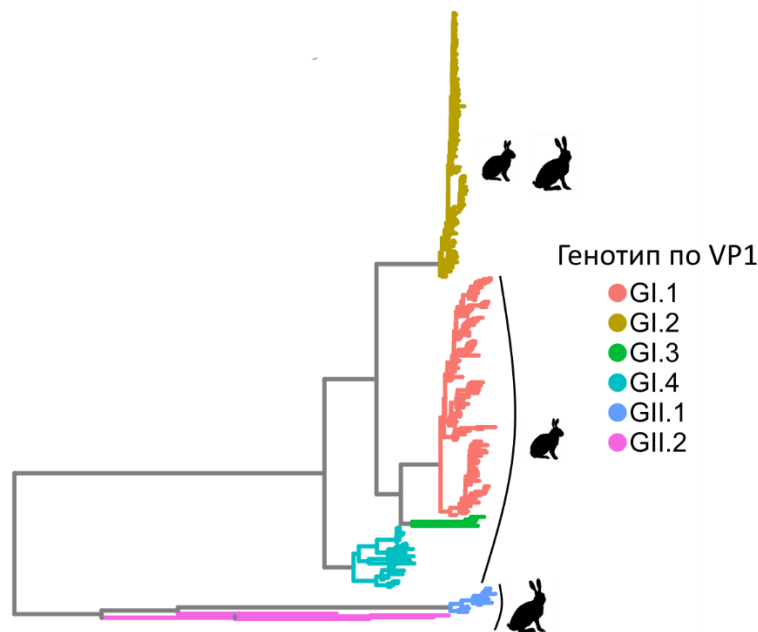


Рис. 26. Филогенетическое дерево капсидного белка VP1 вирусов рода *Lagovirus*, использованных для анализа рекомбинации. Дерево построено по нуклеотидной последовательности VP1 алгоритмом ML в программе IQ-TREE [299]. Ветки дерева покрашены согласно генотипам лаговировусов. Силуэты основных хозяев вирусов разных генотипов показаны рядом с соответствующими листьями дерева. Таксономия соответствует обновлению ICTV 2021.

Матрицы филогенетической совместимости и матрица ОНР лаговировусов больше напоминают таковые у норовировусов, чем у саповировусов (Рис. 27). Наибольшая филогенетическая несовместимость наблюдается между неструктурной (часть ORF1) и структурной (часть ORF1, кодирующая белок VP1, и ORF2) областями генома. Согласованность филогенетических деревьев и генетических дистанций внутри неструктурных и структурных областей выше, чем между ними, но не такая высокая, как у саповировусов.

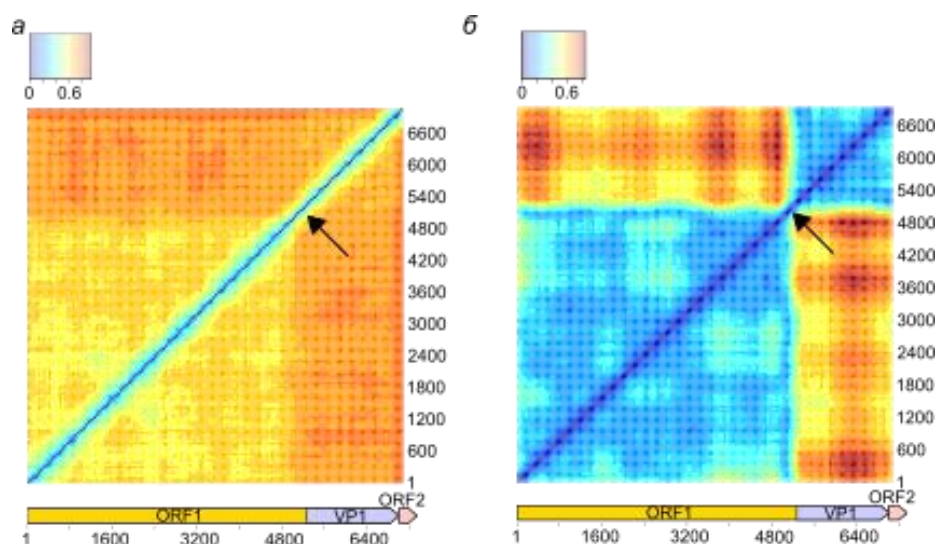


Рис. 27. а – матрицы филогенетической совместимости на основе нормализованных расстояний Робинсона-Фулдса, б – матрицы ОПР для вирусов рода *Lagovirus*. Цвета отражают значения нормированных расстояний Робинсона-Фулдса и СКО в графиках СПГР соответственно. Черные стрелки указывают на стык НСО и СО генома.

Наибольшая несогласованность попарных генетических дистанций наблюдалась между неструктурной и структурной областями, как и было предсказано с помощью матрицы филогенетической совместимости и матрицы ОПР (Рис. 28б). Практически для всех событий рекомбинации удалось найти рекомбинантную последовательность, и, как правило, отклоняющиеся точки соответствовали сравнениям одной-двух рекомбинантных последовательностей со своими «родителями» и близкими к «родителям» последовательностями.



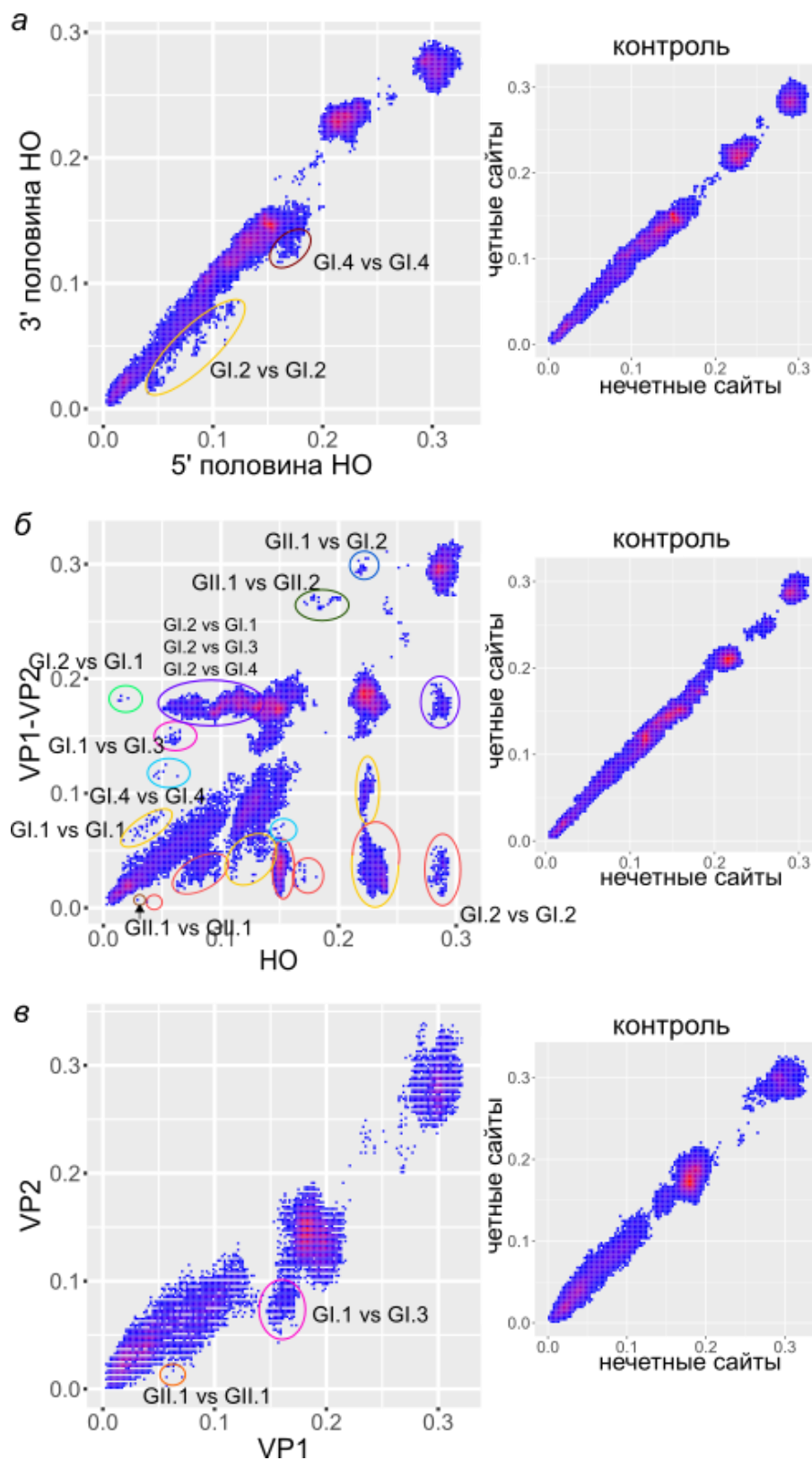


Рис 28. Соответствие попарных нуклеотидных расстояний между 5'- и 3'-половинами неструктурной области генома (а), неструктурной и структурной областей генома (б), а также между VP1 и VP2 (в) у лаговировусов. Точки, которые соответствуют парам вирусов, подвергшимся рекомбинации, и обсуждаются в тексте, отмечены кружками. Контрольные графики соответствия между нуклеотидными расстояниями в четных и нечетных позициях

выравниваний имитируют графики СПГР в отсутствии рекомбинантных последовательностей.

Удалось детектировать только два события рекомбинации между одноименными генотипами. Одна последовательность являлась результатом рекомбинации непатогенных вирусов GI.4 (RCV) (Рис. 28б, голубые круги), разошедшихся от рекомбинанта примерно на 4% по нуклеотидной последовательности, и уже была описана в литературе [128]. Второе событие затрагивало вирусы GI.1 (EBHSV) и, по-видимому, произошло недавно, так как вирусы были практически идентичны в НСО и отличались на 3% в НСО (Рис. 28б, коричневый кружок).

Большинство событий рекомбинации происходило между вирусами разных генотипов GI. Самые многочисленные события соответствовали рекомбинации патогенной геногруппы GI.2 (RHDV2/b) с патогенной GI.1 (RHDV) и непатогенными GI.3 и GI.4 (RCV). В этих событиях GI.1, GI.3 и GI.4 являлись донорами НСО, и расстояние между вирусами из этих геногрупп и GI.2 составляло от 7 до 11% в НСО и от 17 до 20% в СО (Рис. 28б, фиолетовый круг). Отдельные случаи рекомбинации между GI.2 и GI.1, GI.3 и GI.4 ранее упоминались в литературе [117,118,123–126,128], а также было высказано предположение о том, что ранние изоляты GI.2 появились в результате рекомбинации между GI.2 и GI.3, где GI.3 была донором НСО [117]. Учитывая, что скорость накопления замен для GI составляет около  $7.7 \times 10^{-4}$  з/с/год [319], рекомбинация между GI.2 и GI.1, GI.3 и GI.4 могла произойти порядка 90-150 лет назад. Можно выделить ещё одно чуть более раннее событие рекомбинации между патогенным генотипом GI.1 и непатогенным GI.3, в котором донором НСО выступал непатогенный генотип (Рис 28б, розовый круг, дистанция между GI.3 и GI.1 около 5%). Все перечисленные выше события рекомбинации образуют на графиках СПГР группы точек, соответствующие сравнениям рекомбинантных вирусов с вирусами одноименного генотипа, симметричные сравнениям с разноименными генотипами (Рис 28б, красные круги - сравнения GI.2 и GI.2, оранжевые круги - сравнения GI.1 и

GI.1). Две последовательности GI.2 были результатом рекомбинации между структурными генами GI.2 (Рис 28б, красный круг, дистанции около 2-4% в СО и около 30% в НСО) и неструктурными генами GII.1 (EBHSV) (Рис 28б, синий круг). Рекомбинация между GI.2 и EBHSV повлияла на расстояния рекомбинантных последовательностей от других вирусов (Рис 28б, фиолетовый круг с 30% в НСО и 20% в СО). Это событие было недавно описано в литературе как первый случай рекомбинации между лаговирusernameами разных генo-групп [118].

Генетические расстояния в двух половинах неструктурной области гeнома в целом коррелировали, однако были группы точек, отклоняющихся от линейного тренда, соответствующие сравнениям вирусов из GI.2 (RHDV2/b) и GI.4 (RCV) (Рис 28а). При сравнении генетических расстояний в VP1 и VP2 также наблюдалась линейная зависимость между расстояниями, и было лишь несколько групп отклоняющихся точек. Первая группа соответствует сравнениям последовательности МК440615, принадлежащей GII.1 (EBHSV), с двумя другими последовательностями из GII.1: последовательности отличаются на 0.2-1% последовательности в VP2 и примерно на 5% в VP1 (Рис 28в, оранжевый круг). Так как симметричной группы отклоняющихся точек не было, нельзя точно определить рекомбинантную последовательность. Вторая группа точек соответствовала сравнениям вирусов из патогенного генотипа GI.1 (RHDV) и непатогенного GI.3 (RCV-E1), которые отличались на 14-17% в VP1 и 4-8% в VP2 (Рис 28в, отмечено красным) – в RDP4 эта находка не воспроизвелась.

### 3.3 Рекомбинация у астровирусов

В анализе рекомбинации рода *Mamastrovirus* было использовано 319 полногеномных последовательностей, принадлежащих 13 видам и нескольким группам ещё не классифицированных астровирусов, заражающих различных хозяев (Рис. 29).

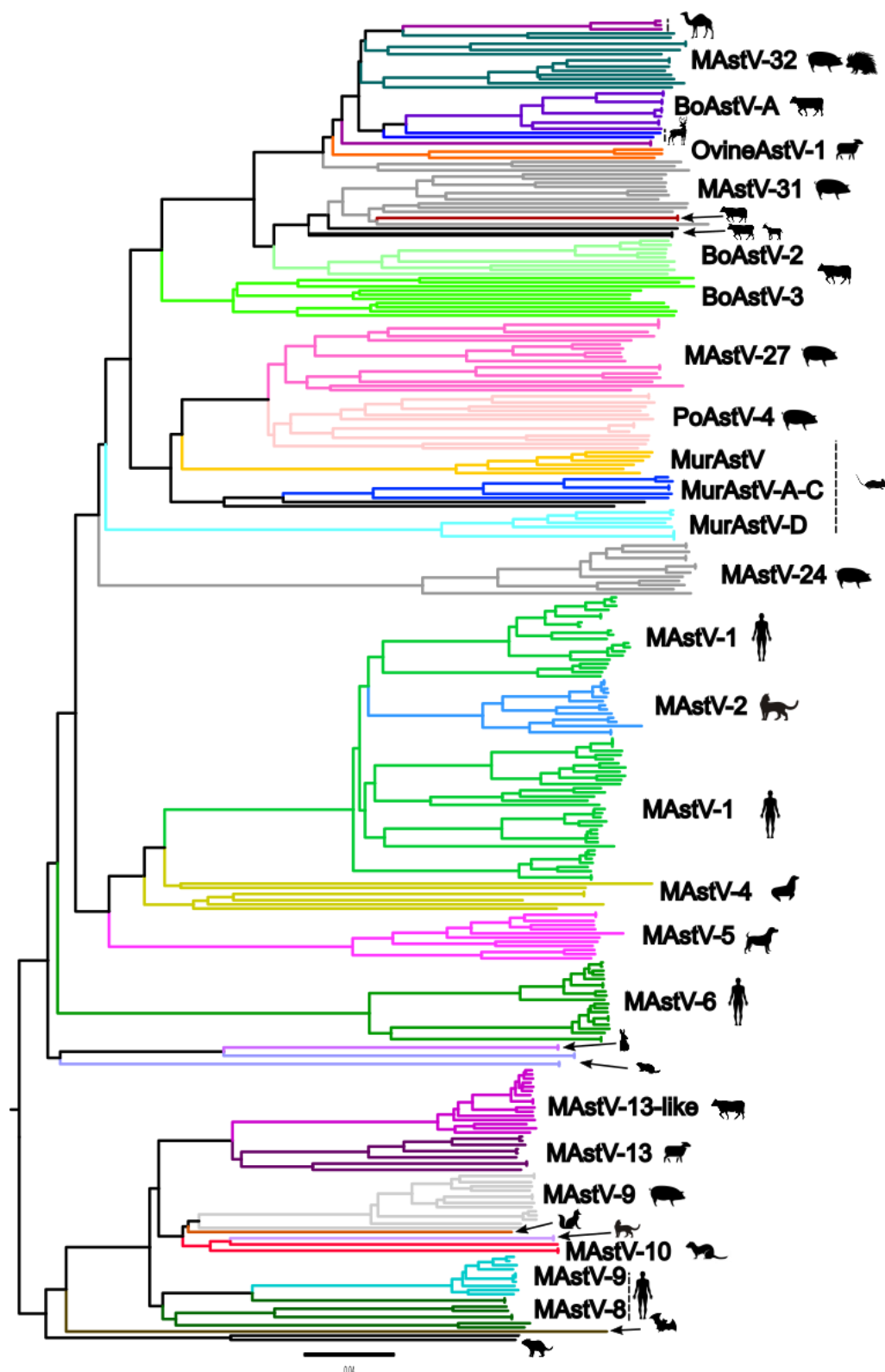


Рис. 29. Филогенетическое дерево, построенное по ORF2, кодирующей капсидные белки, для последовательностей рода *Mamastrovirus*, использованных для анализа рекомбинации. Дерево построено алгоритмом ML в программе IQ-TREE [299]. Ветки дерева покрашены согласно видовой принадлежности астровирусов. Названия видов и силуэты основных хозяев показаны рядом с соответствующими листьями дерева.

На графиках филогенетической совместимости астровирусов млекопитающих можно выделить области с ограниченной рекомбинацией, совпадающие с половинами ORF1a, а также с ORF1b и ORF2 (Рис. 30а). Внутри этих областей расстояния между филогенетическими деревьями меньше, чем при сравнении с другими регионами. Наибольшие расстояния Робинсона-Фулдса наблюдаются при сравнении деревьев, построенных по ORF2, кодирующей капсидные белки, и ORF1, кодирующей неструктурные белки. Аналогичная закономерность наблюдается в матрицах ОПР (Рис. 30б). При этом на обоих графиках не наблюдается четких границ между этими участками генома, что отличает астровирусы от норовирусов, лаговирусов и саповирусов.

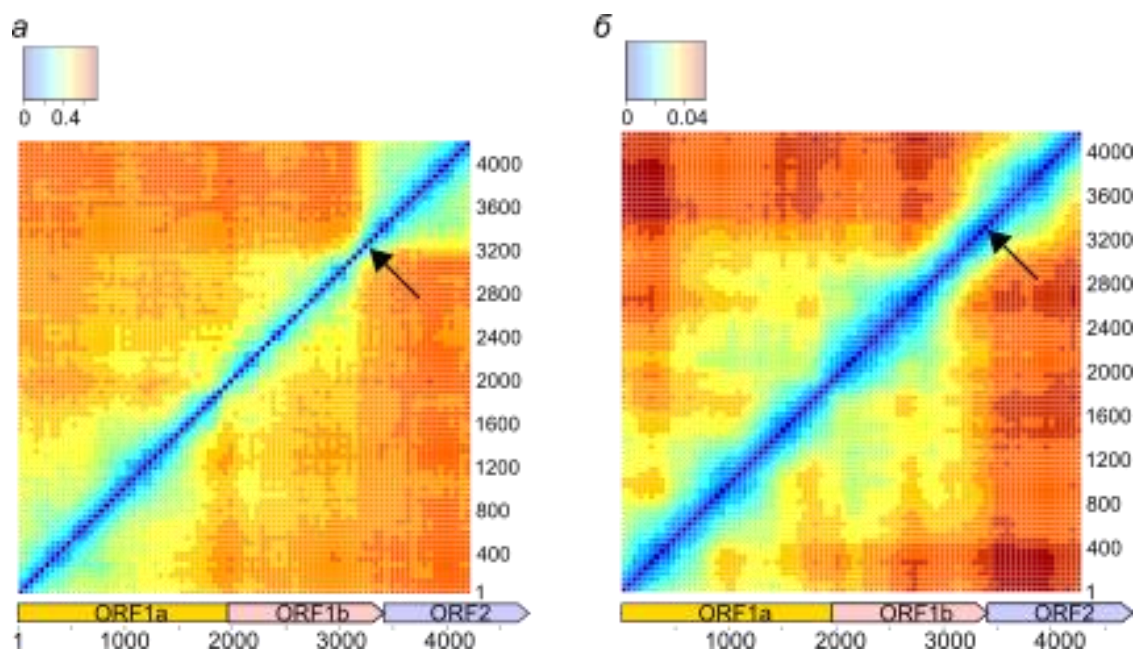


Рис. 30. а – матрицы филогенетической совместимости на основе нормализованных расстояний Робинсона-Фулдса, б – матрицы ОПР для вирусов рода *Mamastrovirus*. Цвета отражают значения нормированных расстояний Робинсона-Фулдса и СКО в графиках СПГР соответственно. Черные стрелки указывают на стык НСО и СО генома.

Относительно более низкая частота рекомбинации внутри ORF1b и ORF2 подтверждается графиками СПГР (Рис. 31в,д). График попарных генетических дистанций в двух половинах ORF1b практически совпадает с контролем (Рис. 31д). Это контрастирует с графиком СПГР двух половин ORF1a, где есть группы точек, существенно отклоняющиеся от линии линейного тренда (Рис. 31г). Несколько последовательностей, образовавшихся в резуль-

тате рекомбинации между очень близкими вирусами своего вида, относились к астровирусам человека MAsV1, астровирусам кошек MAsV2 и астровирусам мышей MurAsV (Рис. 31г, фиолетовый, зеленый, оранжевый круги соответственно). Более многочисленные рекомбинационные события можно выделить между разными видами астровирусов свиней: MAsV27 и PoAsV4, MAsV31 и MAsV32. Этим событиям соответствуют две симметричные группы отклоняющихся точек (Рис. 31г, розовые круги). Астровирус коз (МК404645) представлял собой единственную последовательность, явившуюся результатом рекомбинации в ORF1a между относительно сильно дивергировавшими астровирусами коров BoAsV (Рис. 31г, черные круги).

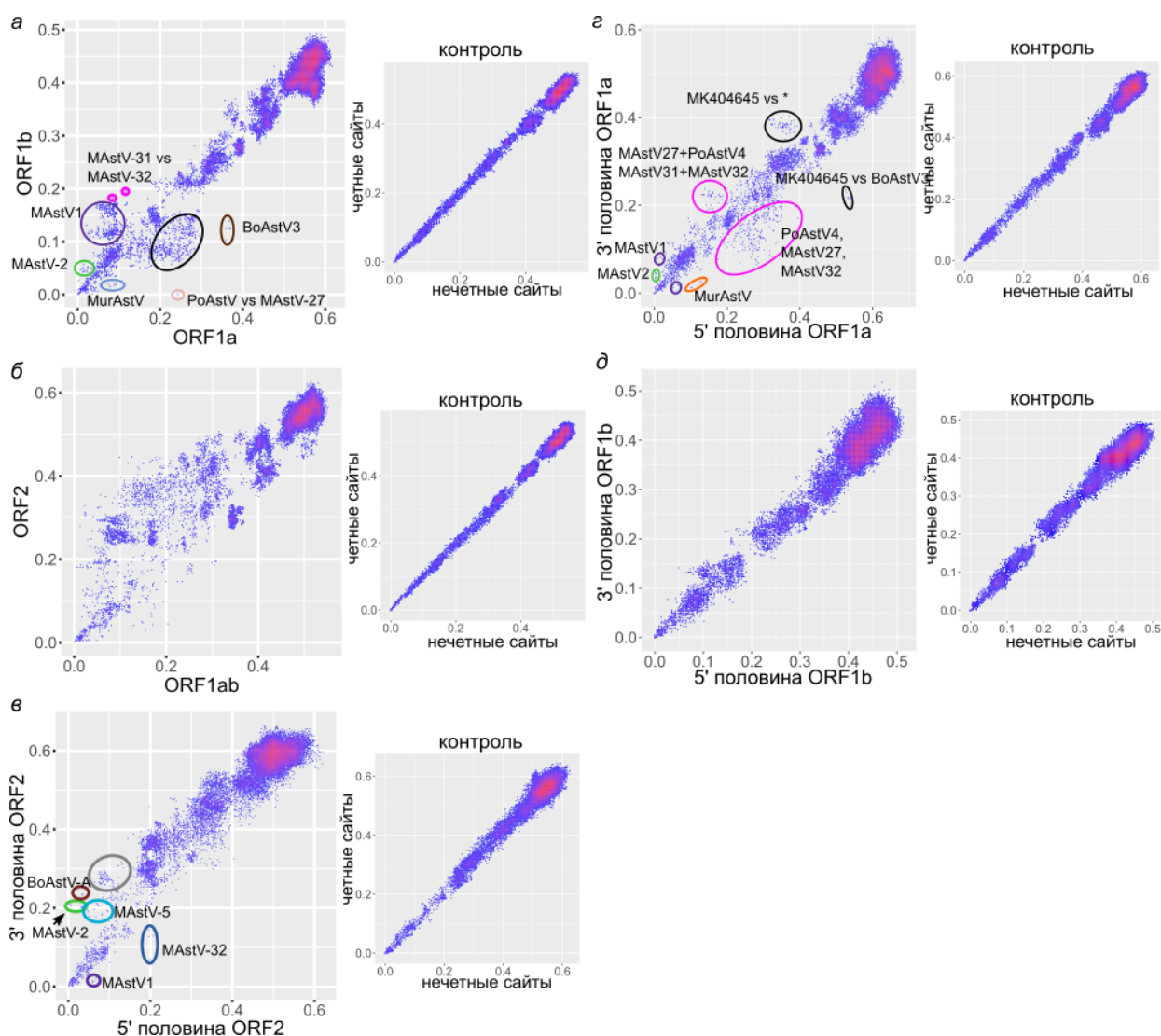


Рис. 31. Соответствие попарных нуклеотидных расстояний между ORF1a и ORF1b (а), ORF1ab и ORF2 (б), а также между 5'- и 3'-половинами ORF3 (в) у астровирусов. Точки, которые соответствуют парам вирусов, подвергшимся

рекомбинации, и обсуждаются в тексте, отмечены кружками. Контрольные графики соответствия между нуклеотидными расстояниями в четных и нечетных позициях выравниваний имитируют графики СПГР в отсутствии рекомбинантных последовательностей.

График СПГР ORF1a и ORF1b характеризовался умеренной корреляцией между попарными расстояниями и наличием небольших групп точек, соответствующих событиям рекомбинации между близкими или умеренно дивергировавшими вирусами (Рис. 31а). Рекомбинация между близкими вирусами видов MAstV2 и MurAstV внутри ORF1a (Рис. 31г) отразилась и на графике СПГР ORF1a и ORF1b (Рис. 31а, зеленый и голубой круги соответственно). Рекомбинация между ORF1a и ORF1b происходила между астровирусами человека MAstV1 разных серотипов (Рис. 31а, фиолетовый и черный круги). Астровирус коз, относящийся к BoAstV-3, давал отклоняющиеся попарные сравнения с другими BoAstV-3, заражающими коров, которые являлись донором его ORF1b (Рис. 31а, коричневый круг, 35% отличий в ORF1a и 10% в ORF1b).

В двух половинах ORF2 попарные генетические дистанции также были в целом согласованы, но наблюдались отклоняющиеся группы точек, соответствующие событиям рекомбинации. Самое недавнее событие рекомбинации внутри ORF2 произошло между астровирусами человека MAstV1, имеющими практически идентичные 3'-половины ORF2 и 5'-половины, отличающиеся на 4-6% нуклеотидной последовательности (Рис. 31в, фиолетовый круг). Также недавняя рекомбинация между близкими вирусами, отличающимися менее чем на 5% в 5'-половине ORF2 и на 25-30% во второй половине, произошла между астровирусами кошек MAstV2 (Рис. 31в, зеленый круг), астровирусами коров BoAstV (Рис. 31в, зеленый круг). Помимо этого, можно отметить рекомбинацию между астровирусами собак MAstV5 (Рис. 31в, зеленый круг). Отклонения от линейного тренда дают многочисленные сравнения вирусов внутри видов MAstV1, MAstV2 (в том числе сравнения между астровирусами тигра и кошек), BoAstV-A, MurAstV, MAstV24 (сви-

ньи), MAstV27 (свиньи), MAstV5 (собаки) (Рис. 31в, серый круг), что может также свидетельствовать о рекомбинации.

И в матрицах филогенетической совместимости, и в матрицах ОПП наименьшая согласованность наблюдалась между деревьями и попарными дистанциями, построенными по ORF1, кодирующей неструктурные белки, и ORF2, кодирующей белки капсида. График СПГР ORF1ab и ORF2 отличался от других отклонением большого количества попарных генетических расстояний, образующих «облако». Эти расстояния соответствовали попарным сравнениям астровирусов, принадлежащим одному и тому же виду: MAstV1 (человек), MAstV2 (кошка), MAstV5 (собака) BoAstV-A (крупный и мелкий рогатый скот), MurAstV (мышь), а также сравнениям астровирусов свиней, принадлежащих разным видам или группам: MAstV27, PoAstV4, MAstV31 и MAstV32. «Облако» генетических дистанций было несимметрично, поэтому нельзя было определить, какие именно последовательности являются рекомбинантами, но можно сделать вывод о том, что среди астровирусов встречаются вирусы с очень близкими неструктурными белками и дивергировавшими белками капсида, но не наоборот. Это наблюдение соответствует большей вариабельности рекомбинационно-мобильных структурных генов у астровирусов и гена S у коронавирусов, как будет показано далее.

В отличие от калицивирусов, у астровирусов не было четкой границы между блоками генов CO и HCO в плане рекомбинационной мобильности. Наблюдалось много событий рекомбинации не на границе CO и HCO. Очевидно, что с одной стороны имеет место та же модульная динамика генома, но закономерности биологической совместимости генов отличаются – нет признаков строгой несовместимости фрагментов внутри CO или HCO между разными вирусами.

### **3.4 Рекомбинация у коронавирусов**

Количество доступных полногеномных последовательностей для разных родов коронавирусов было неравномерным (Таблица 7). В пределах  $\gamma$ -



CoV и  $\delta$ -CoV для некоторых видов было доступно очень много последовательностей вирусов, заражающих только один вид хозяев – кур и свиней, соответственно (Рис. 32). Некоторые виды, в частности, из рода  $\alpha$ -CoV, отсутствовали в наборе данных, так как для них не было доступных полных геномов.

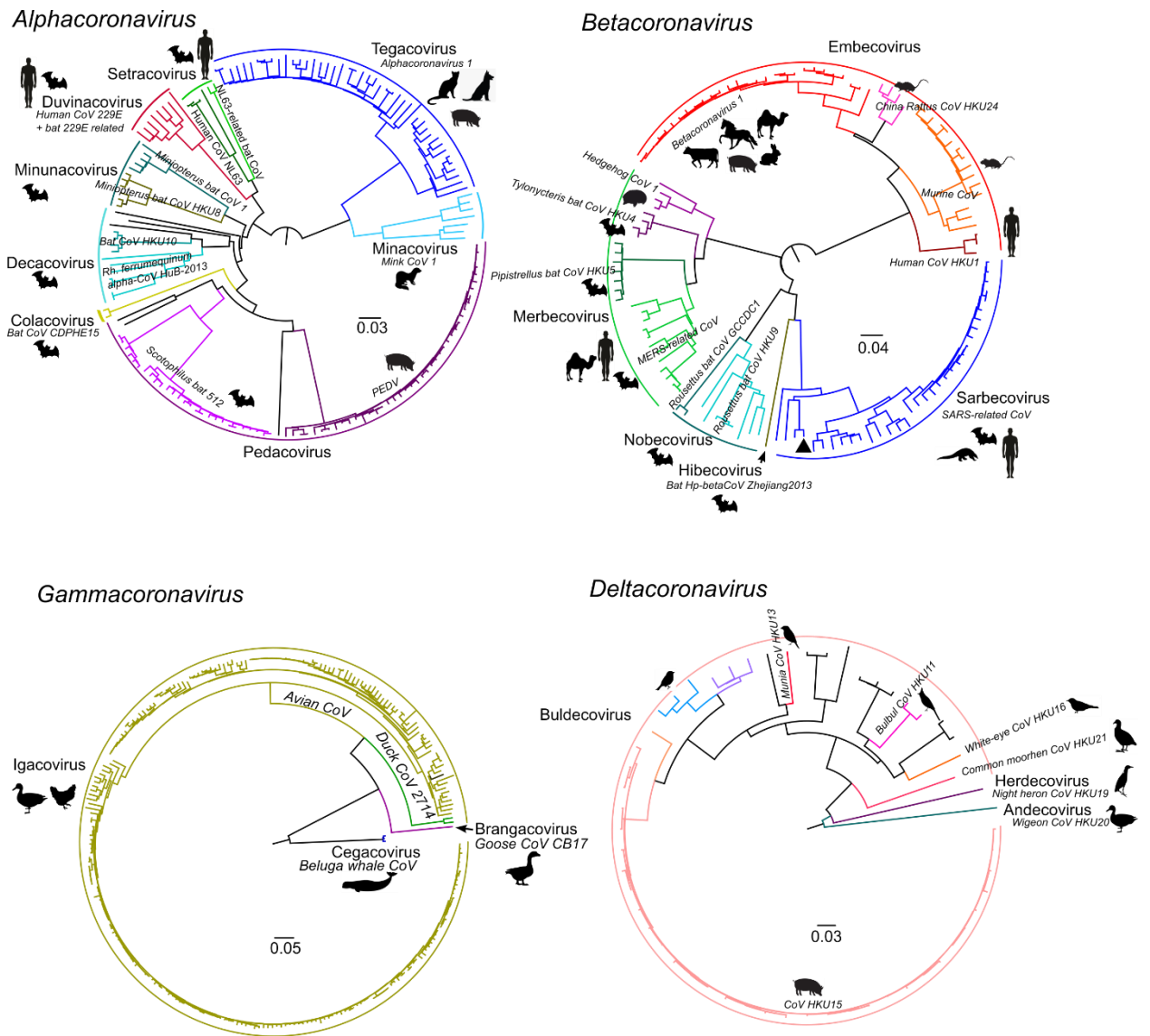


Рис. 32. Филогенетические деревья, построенные последовательности Р3Р1 (ORF1b) четырех родов коронавирусов, использованных для анализа естественной рекомбинации. Деревья построены алгоритмом ML в программе IQ-TREE [299]. Приведены названия подродов и видов. Силуэты основных хозяев вирусов разных видов показаны рядом с соответствующими листьями дерева. Ветки дерева покрашены согласно видовой принадлежности коронавирусов. Таксономия соответствует обновлению ICTV 2021. Не все вирусы формально отнесены к определенным видам, для некоторых вирусов отсутствуют данные о хозяевах.

Анализ рекомбинации у коронавирусов проводили независимо для  $\alpha$ -*CoV* (N=164),  $\beta$ -*CoV* (N=122),  $\gamma$ -*CoV* (N=260) и  $\delta$ -*CoV* (N=56). Для того чтобы выявить общие закономерности рекомбинации коронавирусов, были рассчитаны матрица филогенетической совместимости и матрица ОПР. Матрица филогенетической совместимости (Рис. 33а) показала, что рекомбинация происходила во всем геноме коронавирусов. Ген шиповидного белка S чаще был рекомбинантным по сравнению с остальным геномом во всех четырех родах коронавирусов. В остальных частях генома у  $\alpha$ -*CoV*,  $\beta$ -*CoV* и  $\delta$ -*CoV* распространенность рекомбинации была ниже, а у  $\gamma$ -*CoV* практически равномерно высока по всему геному, о чем свидетельствует большее количество филогенетических конфликтов между всеми областями генома в матрицах филогенетической совместимости (Рис. 33а).

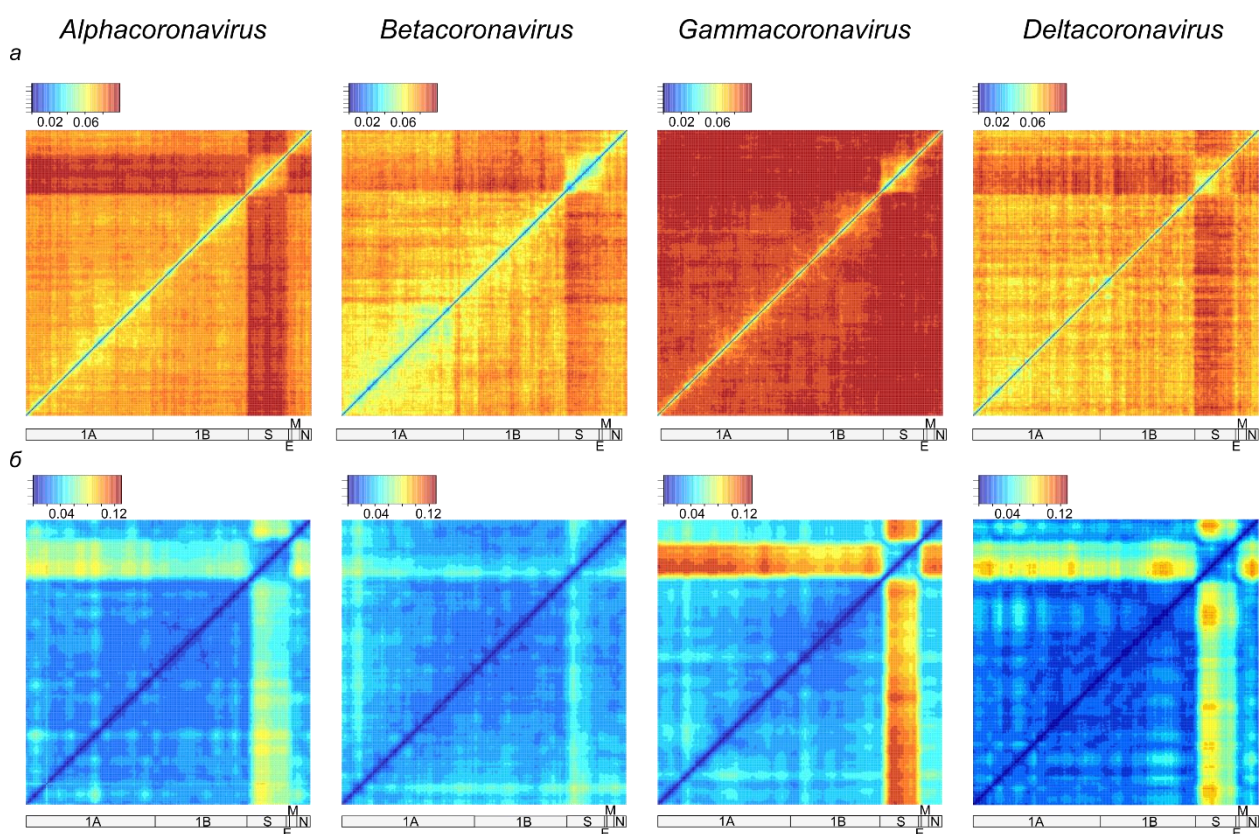


Рис. 33. Ген белка S визуально более вовлечен в рекомбинацию по сравнению с остальной частью генома в четырех родах коронавирусов. а – матрицы филогенетической совместимости на основе нормализованных расстояний Робинсона-Фулдса, б – матрицы ОПР для четырех родов коронавирусов. Цвета отражают значения нормированных расстояний Робинсона-Фулдса и СКО в графиках СПГР соответственно.

Было выдвинуто предположение, что общая распространенность рекомбинации по геному у  $\gamma$ -CoV настолько высока, что матрица совместимости не может обнаружить дополнительную более частую рекомбинацию с участием гена *S*. Для того чтобы проверить это предположение, были построены филогенетические деревья по ORF1a, ORF1b и *S* для анализируемых последовательностей (Рис. 34). В дереве по ORF1a названия таксонов были

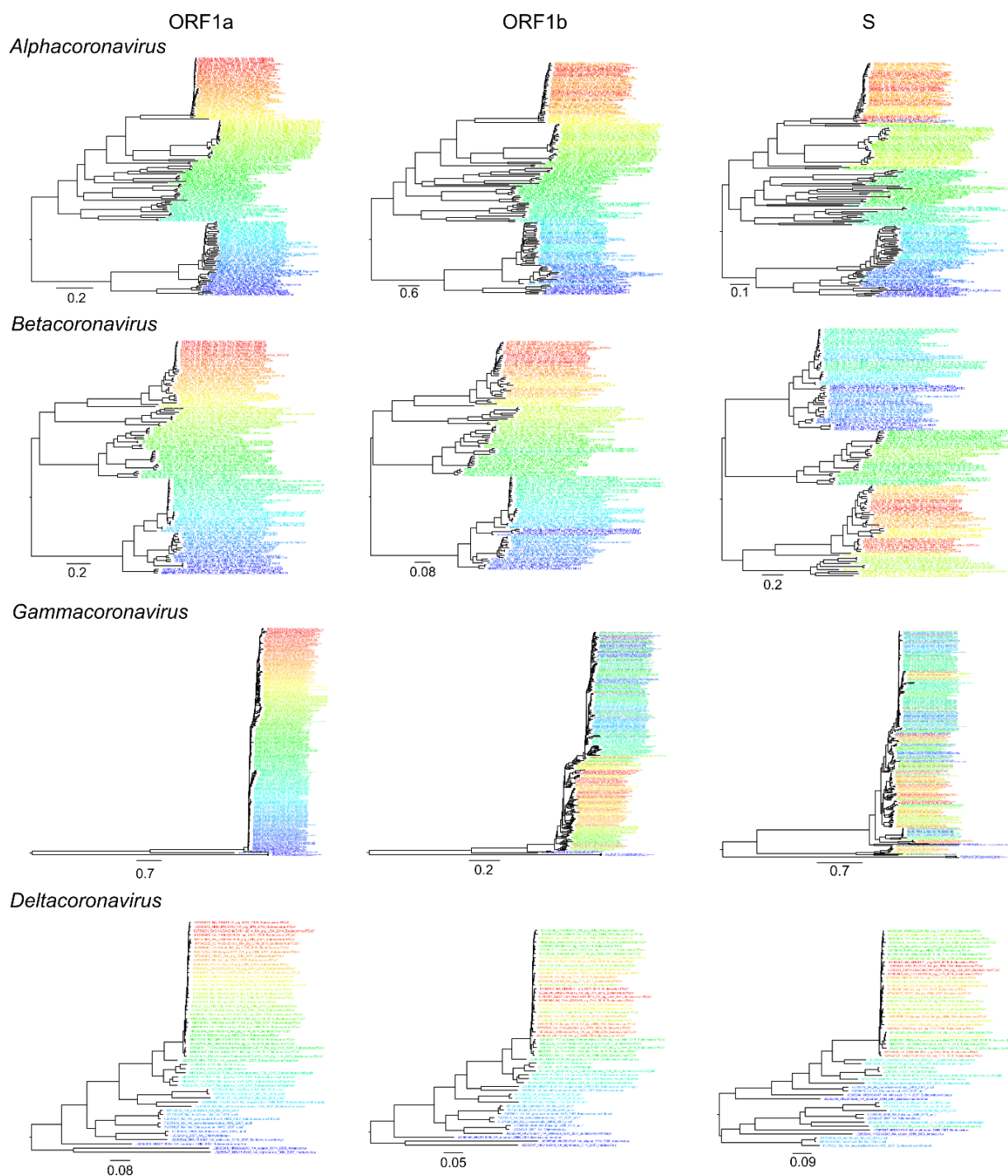


Рис. 34. Перемешивание таксонов в деревьях ML для генов ORF1a, ORF1b и *S* для четырех родов коронавирусов. Цвета названий таксонов совпадают на

деревьях для каждого рода. Отрезки под деревьями соответствуют генетическим расстояниям, рассчитанным в рамках модели GTR+F+R5.

покрашены по градиенту цветов, и те же самые цвета были использованы для каждого таксона на других деревьях. Степень смешения таксонов  $\gamma$ -CoV была заметно выше между деревьями для ORF1a и S, чем между деревьями для ORF1a и ORF1b. Количественная оценка различия перемешивания таксонов не проводилась поскольку только около половины узлов имели бустрэп-поддержки более 0.8. Таким образом, благодаря классическому методу общего анализа рекомбинации – матрицам филогенетической совместимости – удалось подтвердить ранее описанную повсеместную рекомбинацию между S и другими генами [216, 261] и предположить более распространенную рекомбинацию в  $\gamma$ -CoV, но не удалось выявить более детальные закономерности.

Матрицы ОПР в целом соответствовали матрицам филогенетической совместимости (Рис. 33б). Можно выделить две группы генов, которые характеризуются относительно высокой согласованностью генетических расстояний внутри этих групп и низкой согласованностью между группами – (1) S и (2) все остальные гены, кроме S. Это предполагает менее распространенную рекомбинацию внутри групп и более частую рекомбинацию между S и остальной частью генома, что согласуется с матрицами филогенетической совместимости. Границы групп генов (S против остального генома) очень четкие у  $\gamma$ -CoV, а в остальных трех родах (особенно у  $\beta$ -CoV) в 5'-части S наблюдалась бóльшая, чем в 3'-части, степень несогласованности попарных дистанций по отношению к другим участкам генома. Таким образом, матрицы ОПР предоставили дополнительную информацию о границах областей рекомбинации, а также в отличие от матриц филогенетической совместимости позволили выявить более частую рекомбинацию между геном S и остальным геномом у  $\gamma$ -CoV. Далее для проверки закономерностей рекомбинации, выведенных с помощью матриц филогенетической совместимости и ОПР, были проанализированы графики СПГР в разных областях генома.

Матрицы ОПП (Рис. 33б) предполагают умеренную распространенность рекомбинации внутри ORF1 во всех четырех родах с некоторыми потенциальными «горячими точками» рекомбинации в ORF1a. Умеренный уровень рекомбинации в  $\alpha$ -CoV и  $\beta$ -CoV подтверждается графиками СПГР в ORF1a и ORF1b (Рис. 35а), что также видно на филогенетических деревьях (Рис. 34). С помощью графиков СПГР был обнаружен ряд событий рекомбинации внутри ORF1ab, например, между изолятами вида *Alphacoronavirus 1*, заражающими разные виды животных (Рис. 35а, черные кружки), а также между представителями коронавируса летучих мышей *Miniopterus HKU8* (Рис. 35а, красный кружок). Кроме того, было несколько событий рекомбинации в пределах вида *Murine coronavirus* (Рис. 35а, зеленые кружки) и *SARS-related coronavirus* (Рис. 35а, черные кружки), включая рекомбинацию между новым SARS-CoV-2 и коронавирусами из летучих мышей (жирный черный круг). Однако в целом наблюдалась линейная зависимость генетических расстояний в ORF1a и ORF1b. В  $\gamma$ -CoV наблюдалось облако пар генетических расстояний до 10%, что свидетельствует о частой рекомбинации среди близкородственных вирусов (Рис. 35а). Это согласуется с перемешиванием родственных таксонов в филогенетических деревьях  $\gamma$ -CoV, построенных по ORF1a и ORF1b (Рис. 34). На графиках СПГР в ORF1a и ORF1b для  $\delta$ -CoV события рекомбинации не определялись, в то время как на филогенетических деревьях ORF1a и ORF1b группирование близких вирусов, в частности, принадлежащих виду *Porcine coronavirus HKU15*, отличалось (Рис. 34). Это связано с тем, что на матрицах ОПП и графиках СПГР отклонения расстояний, связанные с этими рекомбинационными событиями, не отличались от шума. События рекомбинации в ORF1ab во всех родах коронавирусов затрагивали вирусы, принадлежащие одному виду.

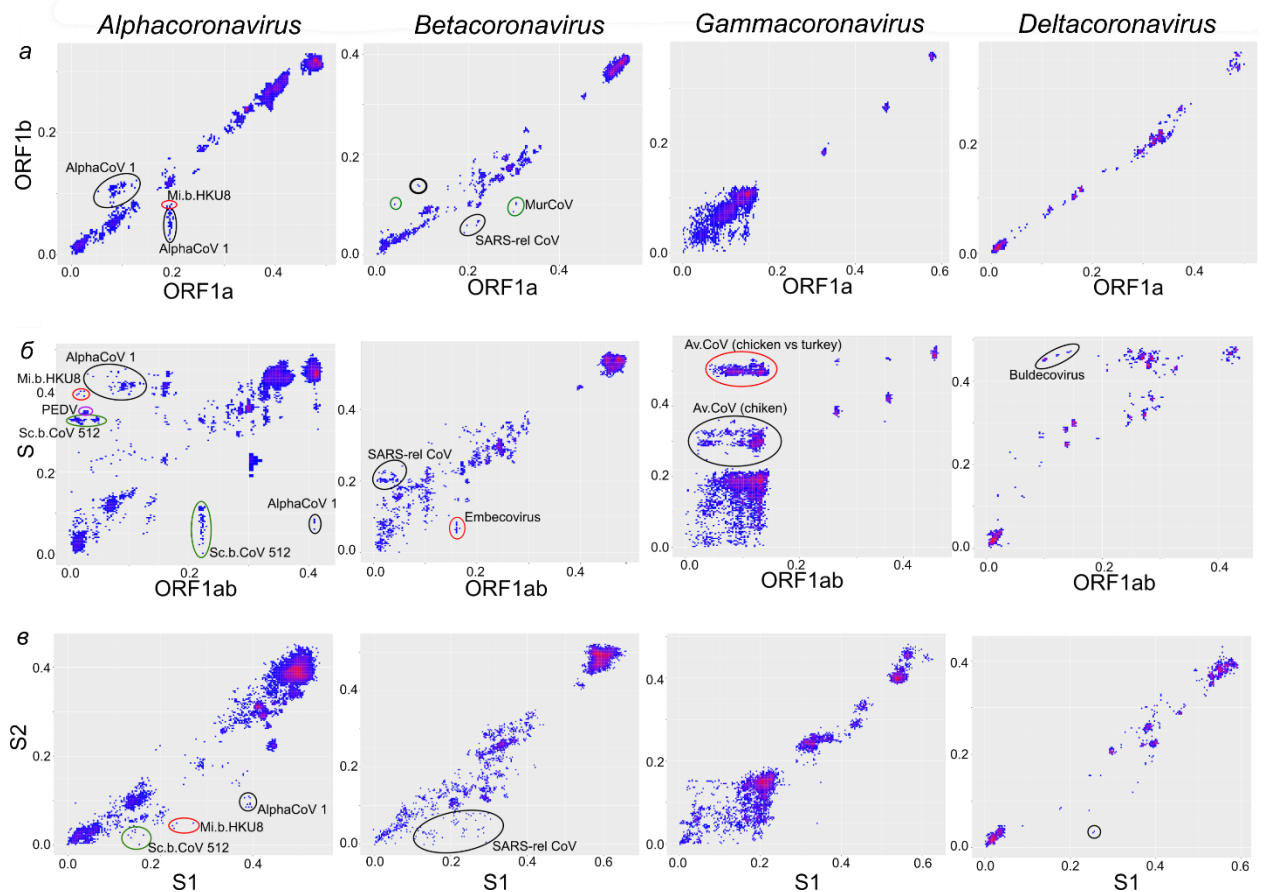


Рис. 35. Рекомбинация чаще встречается между ORF1 и S, чем внутри ORF1. Соответствие попарных нуклеотидных расстояний между ORF1a и ORF1b (а), ORF1ab и S (б), а также участками S1 и S2 белка S (в). Обозначения: AlphaCoV 1, *Alphacoronavirus 1*; Mi.b.HKU8, *Miniopterus bat coronavirus HKU8*; Sc.b.CoV 512, *Scotophilus bat coronavirus 512*; SARS-rel., SARS-related coronavirus; MurCoV, *Murine coronavirus*; Av.CoV, *Avian coronavirus*. Точки, которые соответствуют парам вирусов, подвергшимся рекомбинации, и обсуждаются в тексте, отмечены кружками.

Матрицы ОПП указывали на наличие «горячих точек» рекомбинации небольшого размера в области ORF1ab во всех родах коронавирусов. Детальный анализ подтвердил более высокую, чем в целом по ORF1ab, частоту рекомбинации в этих областях генома (данные не показаны). Важно отметить, что эти «горячие точки» были разными у четырех родов и могли быть обнаружены только с использованием матриц ОПП, но не матриц филогенетической совместимости. Внутри некоторых из этих «горячих точек» были очень высокие максимальные генетические расстояния (до 70%). Проверка вручную выравниваний геномов в этих областях выявила потенциально негосто-

логичные участки. Эти результаты теоретически могут быть объяснены приобретением фрагментов от вирусов других семейств, однако поиск BLAST не выявил предполагаемого источника этих геномных фрагментов.

ORF1ab кодирует неструктурный полипротеин, тогда как *S*, *E*, *M* и *N* являются структурными белками. Тем не менее, в отличие от рассмотренных ранее семейств, только ген *S* участвовал в рекомбинации чаще, чем неструктурные гены. Матрицы филогенетической совместимости и ОПП не предполагали заметно большей рекомбинации между генами ORF1ab, *E*, *M* и *N*, чем внутри ORF1ab (Рис. 33), а графики СПГР между генами ORF1ab, *E*, *M* и *N* были в принципе аналогичны графикам между ORF1a и ORF1b (данные не показаны). В плане рекомбинации не было четкого различия между ORF1ab и «минорными» структурными генами, поэтому последние далее не анализировались. Таким образом, с точки зрения общей для изученных в данной работе РНК-вирусов закономерности «разные профили рекомбинации в СО и НСО», структурные гены *E*, *M* и *N* фактически эквивалентны НСО, и только ген *S* соответствует структурным генам других семейств в плане иной рекомбинационной мобильности.

Наиболее заметное несоответствие попарных генетических расстояний, предполагающее частую рекомбинацию, наблюдалось между *S* и другими областями генома (Рис. 33). Отклонение большого количества попарных расстояний от линейного тренда может быть вызвано как наличием многочисленных недавних рекомбинационных событий, так и несколькими событиями рекомбинации, произошедшими давно между группами менее родственных вирусов. Поэтому были исследованы графики СПГР в *S* и других областях генома. Общей чертой всех родов коронавирусов было слабое соответствие генетических расстояний между умеренно родственными вирусами (менее 20–25% генетического расстояния), рассчитанных по ORF1ab и *S* (Рис. 35б). В этих границах генетических расстояний, примерно соответствующих под родам, было много отчетливых рекомбинационных событий во всех родах коронавирусов. Например, у  $\alpha$ -CoV была обнаружена рекомбинация между

представителями вида *Scotophilus bat coronavirus 512* (Рис. 35б, зеленые кружки). В  $\beta$ -CoV рекомбинация наблюдалась между вирусами вида *SARS-related coronavirus*, а также между вирусами подрода *Embecovirus*, который включает вид *Betacoronavirus 1* (Рис. 35б, черный кружок и красный кружок соответственно). В  $\gamma$ -CoV рекомбинация наблюдалась между представителями вида *Avian Coronavirus* и *Avian Coronavirus 9203*, выделенных из кур и уток соответственно. Некоторые из этих событий рекомбинации произошли совсем недавно, потому что вирусы отличались всего на несколько процентов нуклеотидной последовательности в одном из сравниваемых участков генома.

Общей чертой для разных родов коронавирусов было отсутствие корреляции между попарными генетическими расстояниями в ORF1 и S на генетических дистанциях ниже 10%, то есть наличие «облака» генетических расстояний, что подразумевает большое количество рекомбинационных событий. То, что речь идет именно о множественных рекомбинационных событиях, видно при сравнении филогенетических деревьев для этих участков генома (Рис. 34). В случае  $\gamma$ -CoV все пары вирусов в «облаке» генетических дистанций принадлежали к видам *Avian Coronavirus* и *Avian Coronavirus 9203*, которые включают все коронавирусы индеек, все генотипы вируса инфекционного бронхита IBV и некоторые коронавирусы уток. Подавляющее большинство полных геномов  $\gamma$ -CoV принадлежали IBV, так как этот вирус чаще секвенируют из-за его высокого ветеринарного значения. Представители вида *Avian Coronavirus*, выделенные от других хозяев, были недостаточно представлены в исследуемом наборе данных (Рис. 32). Таким образом, наличие большого числа последовательностей IBV и только одной последовательности коронавируса уток в «облаках» генетических расстояний может быть связано с неравномерностью выборки, а не с биологическими причинами, то есть не может быть сделан вывод о разной частоте рекомбинации у разных представителей рода  $\gamma$ -CoV. В случае  $\delta$ -CoV нельзя подтвердить наличие облака генетических расстояний из-за небольшого количества доступных



последовательностей и высокой представленности близкородственных вирусов свиней *Porcine coronavirus HKU15*.

В родах  $\alpha$ -CoV,  $\gamma$ -CoV и  $\delta$ -CoV наблюдалась относительно недавняя рекомбинация между сравнительно отдаленно родственными вирусами, у которых расстояние между нуклеотидными последовательностями одной из области генома было более 25%, а в другой – менее 10–15%. Наиболее ярким примером являются рекомбинационные события между вирусами видов *Miniopterus bat coronavirus HKU8*, PEDV и *Alphacoronavirus 1* (коронавирус кошек, коронавирус собак, вирус трансмиссивного гастроэнтерита TGEV), принадлежащих роду  $\alpha$ -CoV (Рис. 35б, красный, фиолетовый, зеленый круги соответственно). Еще одним примечательным случаем является облако некоррелированных генетических дистанций между представителями *Avian coronavirus*, выделенными из курицы и индейки, рода  $\gamma$ -CoV, а также рекомбинация между коронавирусом воробья HKU17 и другими неклассифицированными представителями подрода *Buldecovirus* в  $\delta$ -CoV, описанная ранее в литературе [250].

Для того чтобы получить более полное представление о закономерностях рекомбинации в гене *S*, были проверены соответствия попарных генетических расстояний, рассчитанных для двух доменов белка *S*, *S1* и *S2* (Рис. 35в). Несоответствие генетических расстояний между этими доменами было более выражено, чем внутри ORF1ab (Рис. 35а), но менее выражено, чем между ORF1ab и *S* (Рис. 35б). Таким образом, ген *S* чаще, но не всегда, обменивался между вирусами целиком. Закономерности рекомбинации внутри гена *S* различались у родов коронавирусов. У  $\gamma$ -CoV на графике СПГР в доменах *S1* и *S2* наблюдалось симметричное «облако» генетических дистанций до 20%, что может свидетельствовать о частой рекомбинации между доменами (Рис. 35в). «Облако» генетических расстояний было похоже на наблюдаемое в ORF1ab, но с более высокими расстояниями, что характерно для гена *S*. Наблюдались множественные недавние события рекомбинации между вирусами с генетическими расстояниями менее 2% в одной половине *S* и бо-

лее 15% в другой. В  $\alpha$ -CoV и  $\beta$ -CoV «облако» генетических дистанций, образующихся в результате рекомбинации между S1 и S2, было менее выраженным, что предполагает не настолько повсеместно распространенную рекомбинацию, как в  $\gamma$ -CoV, но это «облако» было асимметричным и затрагивало большие генетические расстояния. Так, например, последовательности с расстояниями до 40% в  $\alpha$ -CoV и 30% в  $\beta$ -CoV в доменах S1 могли сочетаться с близкородственным (до 10% различий) S2, но не наоборот (Рис. 35в, зеленый, красный и черный круги). В  $\beta$ -CoV большинство таких событий рекомбинации произошло между представителями *SARS-related coronavirus*, выделенными от летучих мышей, людей и панголинов (Рис. 35в, черный круг).

Количество известных последовательностей  $\delta$ -CoV значительно меньше, чем у остальных трех родов. Более того, на графиках СПГР в доменах S или в S и ORF1ab для  $\delta$ -CoV практически не было попарных дистанций в диапазоне 5–20%, отклоняющихся от линейного тренда, где наблюдалось большинство событий рекомбинации в других родах коронавирусов. Было несколько очевидных событий рекомбинации внутри S, но количество и разнообразие известных последовательностей  $\delta$ -CoV не позволяют провести подробный анализ закономерностей рекомбинации и сопоставить их с другими родами коронавирусов. По той же причине анализ рекомбинации между доменами S у  $\delta$ -CoV оказался малоинформативным, и дальнейший анализ проводился только для трех других родов.

Домен S1, N-концевой участок белка S, состоит из двух субдоменов: N-концевого субдомена S1-NTD и C-концевого субдомена S1-CTD. Оба субдомена могут содержать сайты связывания с рецептором. Так как субдомены S1-NTD и S1-CTD имеют меньшую длину (300–400 нт), чем S1 и S2 (около 1500 нт), количественно сравнивать частоту рекомбинации было затруднительно, и были отмечены только общие закономерности. Профиль рекомбинации между S1-NTD и S1-CTD (Рис. 36а) в целом был аналогичен таковому между S1 и S2 (Рис. 35в). Для  $\gamma$ -CoV была характерна рекомбинация между вирусами с генетическими дистанциями в доменах ниже 25%, при этом было

множество очень недавних переносов. Рекомбинация между субдоменами S1-CTD и S1-NTD у  $\beta$ -CoV включала ряд событий с участием вирусов, которые отличались до 40% в нуклеотидной последовательности S1-NTD (Рис. 36а). Как и в случае рекомбинации между доменами S1 и S2, рекомбинация между S1-NTD и S2 в  $\beta$ -CoV была асимметричной и затрагивала вирусы, сильнее отличающиеся в S1-NTD, чем в S1-CTD или S2 (Рис. 36а,в). Таким образом, большинство событий рекомбинации, обнаруженных у  $\beta$ -CoV при сравнении S1 и S2 (Рис. 35в), соответствовали переносу только S1-NTD, а не всего S1. Все события рекомбинации, наблюдаемые у  $\beta$ -CoV, происходили между представителями вида *SARS-related coronavirus*. Было выявлено несколько относительно недавних событий рекомбинации между S1-NTD и S1-CTD в пределах видов *Scotophilus bat coronavirus 512* и *Mink coronavirus 1* (Рис. 36а, кружки) у  $\alpha$ -CoV. Графики СПГР в S1-NTD и S2, S1-CTD и S2 были аналогичны графикам для S1 и S2, но позволили выявить приблизительные границы рекомбинационных событий. Все события рекомбинации происходили только на уровне вида и затрагивали вирусы видов *Alphacoronavirus 1*, *Miniopterus bat coronavirus HKU 8*, *PEDV* и *Scotophilus bat coronavirus 512* (Рис. 36в, г).

Матрицы ОПП указывали на то, что профиль рекомбинации в гене S  $\beta$ -CoV отличается от других трех родов (Рис. 33б). Это было подтверждено графиками СПГР (Рис. 35в, Рис. 36). Однако сравнение областей гена S вне связи с другими регионами может ввести в заблуждение относительно направления переносов генов, поэтому были проверены соответствия попарных генетических расстояний в доменах S и ORF1ab (Рис. 37).

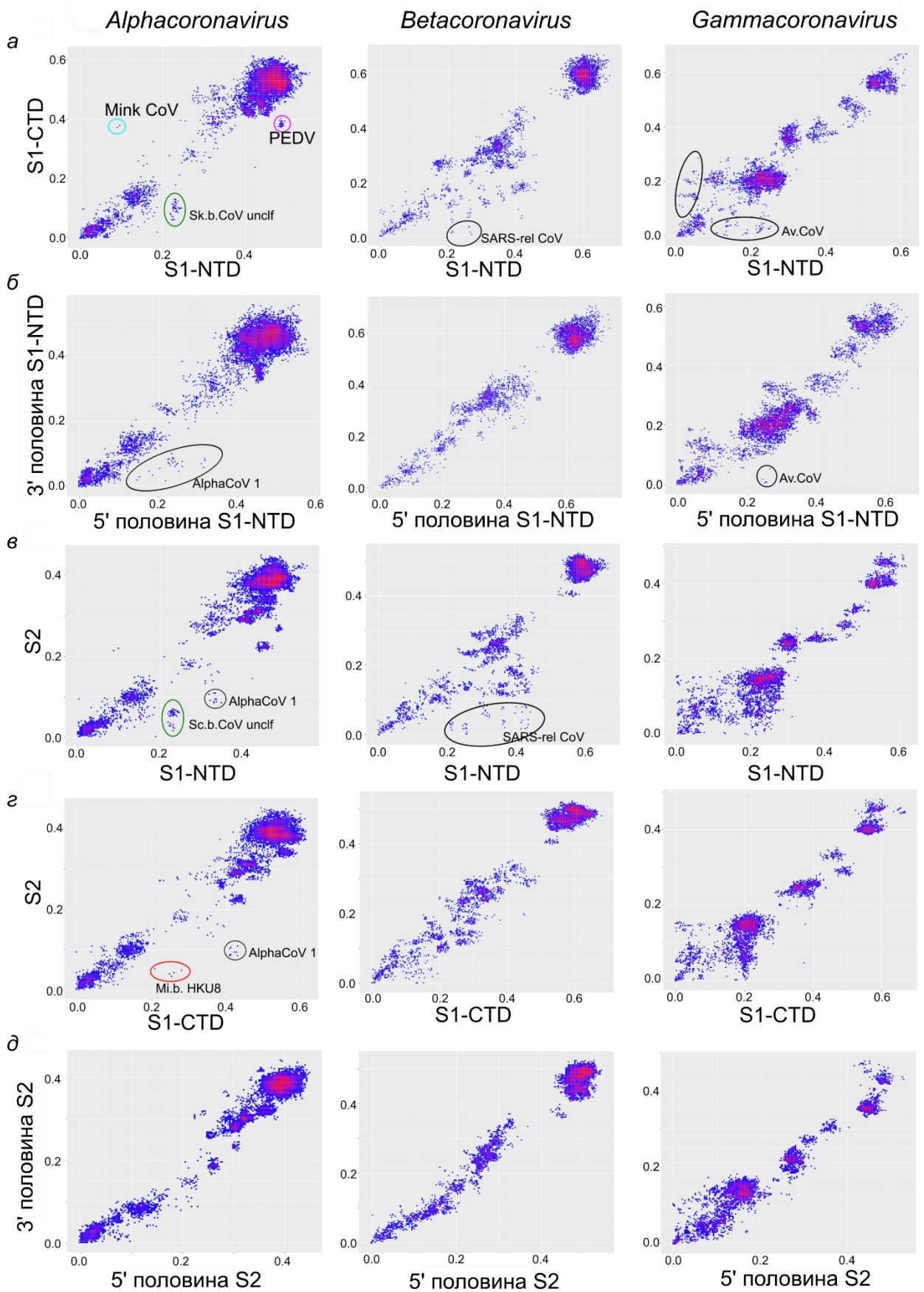


Рис. 36. Рекомбинация внутри доменов белка S менее выражена, чем между ними. Соответствие попарных нуклеотидных расстояний в разных участках

белка S: домены S1-NTD и S1-CTD (а), две половины домена S1-NTD(б), домен S1-NTD и участок S2 (в), домен S1-CTD и участок S2 (г), две половины участка S2 (д). AlphaCoV 1, *Alphacoronavirus 1*; Mi.b.HKU8, *Miniopterus bat coronavirus HKU 8*; Sc.b.CoV 512, *Scotophilus bat coronavirus 512*; SARS-rel., SARS-related coronavirus; Av.CoV, *Avian coronavirus*. Точки, которые соответствуют парам вирусов, подвергшимся рекомбинации, и обсуждаются в тексте, отмечены кружками.

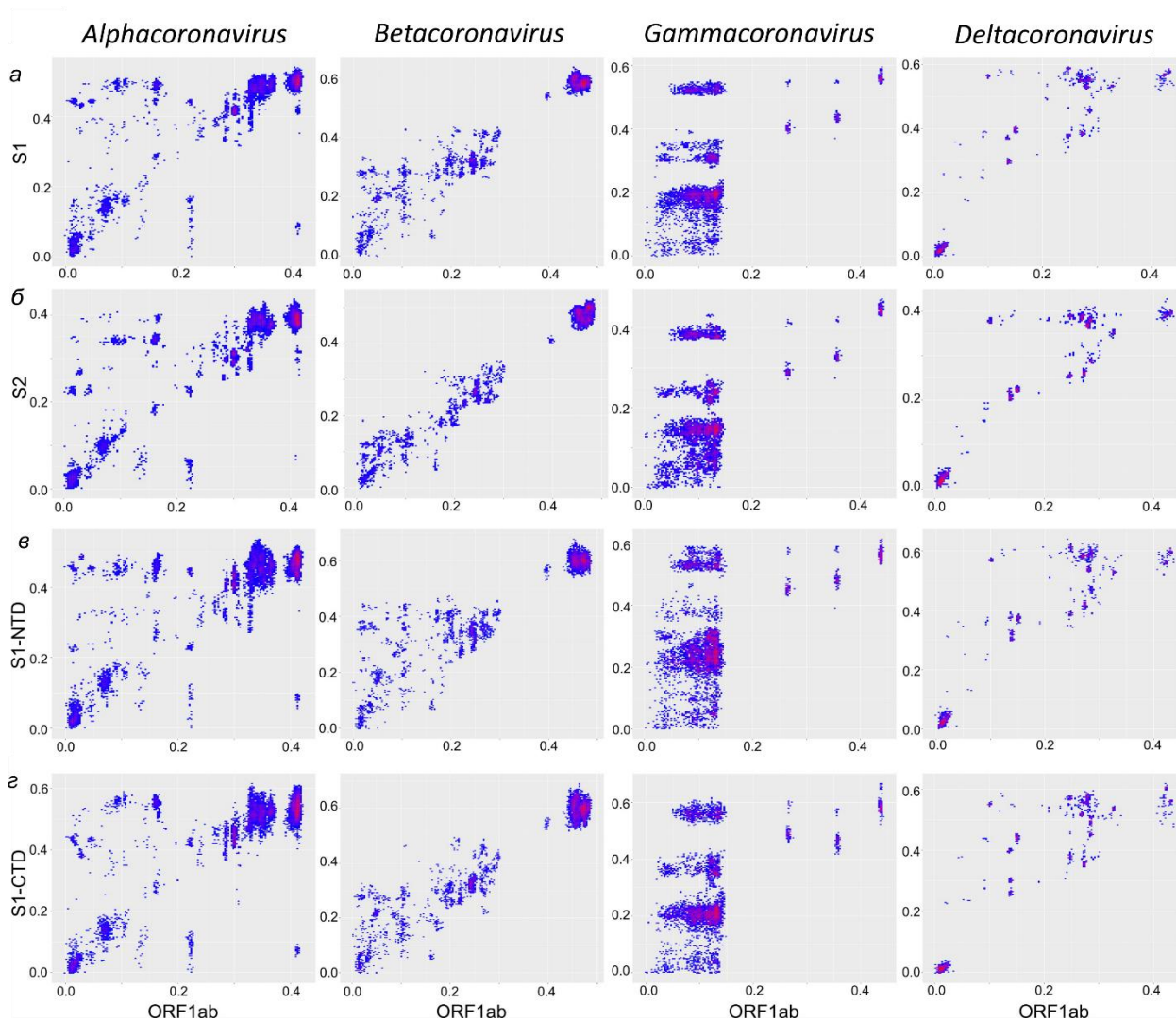


Рис. 37. У  $\beta$ -CoV домен S1-NTD является наиболее «подвижной» частью белка S, в то время как у других родов S обменивается целиком. Соответствие попарных нуклеотидных расстояний между ORF1ab и разными областями белка S: S1 (а), S2 (б), S1-NTD (в), S1-CTD (г).

Графики СПГР между ORF1ab и различными областями белка S подтвердили, что S1-NTD был наиболее подвижной частью S по отношению к остальным участкам генома у  $\beta$ -CoV. На это указывает большее отклонение попарных расстояний, особенно по сравнению с графиком СПГР ORF1ab и

S2 (Рис. 37). У других родов графики СПГР ORF1ab и отдельных доменов S были схожи, что подтверждает то, что S обычно рекомбинирует как единое целое.

### **3.5 Влияние способа формирования выборки нуклеотидных последовательностей и артефактов выборки на результат Байесовского филогенетического анализа**

В связи с широким использованием технологий геномного секвенирования, количество доступных последовательностей для вирусов с большим медицинским значением превышает десятки тысяч, в то время как для других вирусов доступно в сотни раз меньше геномных последовательностей. Для того чтобы понять, как неравномерное семплирование вирусов может влиять на результаты филогенетической реконструкции, был проведен филогенетический анализ искусственно уменьшенных выборок последовательностей полного участка VP1 (891 нт) генома энтеровируса типа EV-A71. Тип EV-A71 обладает самой высокой нейровирулентностью среди неполиомиелитных энтеровирусов. Он вызывал массовые вспышки ящуроподобного заболевания в Азии [321]. EV-A71 подразделяют на семь подтипов: A, B0–B5, C1–C5, D, E, F и G [181,322]. Самыми распространенными подтипами EV-A71 являются подтипы B и C [323,324]. Благодаря своему медицинскому значению и повсеместному распространению EV-A71 является самым широко секвенируемым типом и удобной моделью для симуляционных исследований. Всего по состоянию на июль 2019 г. было доступно 6902 последовательности VP1 EV-A71 с известными датами выделения. Эта выборка последовательностей («основная выборка») в дальнейшем использовалась для тестирования алгоритмов сокращения выборок нуклеотидных последовательностей.

Места выделения изолятов, геномные последовательности которых доступны в базе GenBank, неоднородны. Во многих случаях данные поступают из одного или нескольких крупных исследований, часто исследования проводятся в одном регионе. Чтобы смоделировать ситуацию, когда было проведе-

но всего несколько исследований молекулярной эпидемиологии вируса, был предложен следующий подход. Все последовательности в «основной выборке» были разделены на группы согласно первым пяти символам в их идентификаторе Genbank. Так как последовательностям из одного исследования обычно даются близкие идентификаторы, то такие группы соответствуют последовательностям, отсекуемым в одном исследовании. Последовательности из групп добавляются в выборку до тех пор, пока в выборке не будет определенное количество последовательностей. Таким образом было создано по 10 выравниваний, содержащих 50, 100, 200 или 400 последовательностей. Этот диапазон количества последовательностей в выравниваниях соответствует количеству последовательностей, доступных в GenBank для большинства типов энтеровирусов или видов пикорнавирусов.

Для каждого сгенерированного выравнивания был проведен Байесовский филогенетический анализ для вывода филогенетического дерева с разверткой по времени, высоты корня (возраст самого последнего общего предка) и скорости накопления замен. Байесовский филогенетический анализ не сошелся для четырех из десяти выравниваний, содержащих 50 последовательностей, и для двух из десяти выравниваний, содержащих по 100 последовательностей (Рис. 38а, б, красные кресты). Это может быть связано с тем, что в таких выборках содержалось большое количество похожих последовательностей, полученных в одном исследовании, и несколько сильно отличающихся последовательностей из других исследований с большим разбросом в датах выделения изолятов. Результаты иллюстрируют то, что даже значительное количество последовательностей, полученное в одном или нескольких исследованиях, может быть недостаточно для Байесовского филогенетического анализа. Различия в скоростях накопления замен и возрасте корней деревьев были существенными даже между сгенерированными выборками, содержащими 400 последовательностей (Рис. 38г). Скорости накопления замен варьировались от  $2,4 \times 10^{-3}$  до  $7,0 \times 10^{-3}$  з/с/г, а высота деревьев варьиро-

валась от 60 до 105 лет с неперекрывающимися 95% интервалами плотности апостериорной вероятности (HPD).

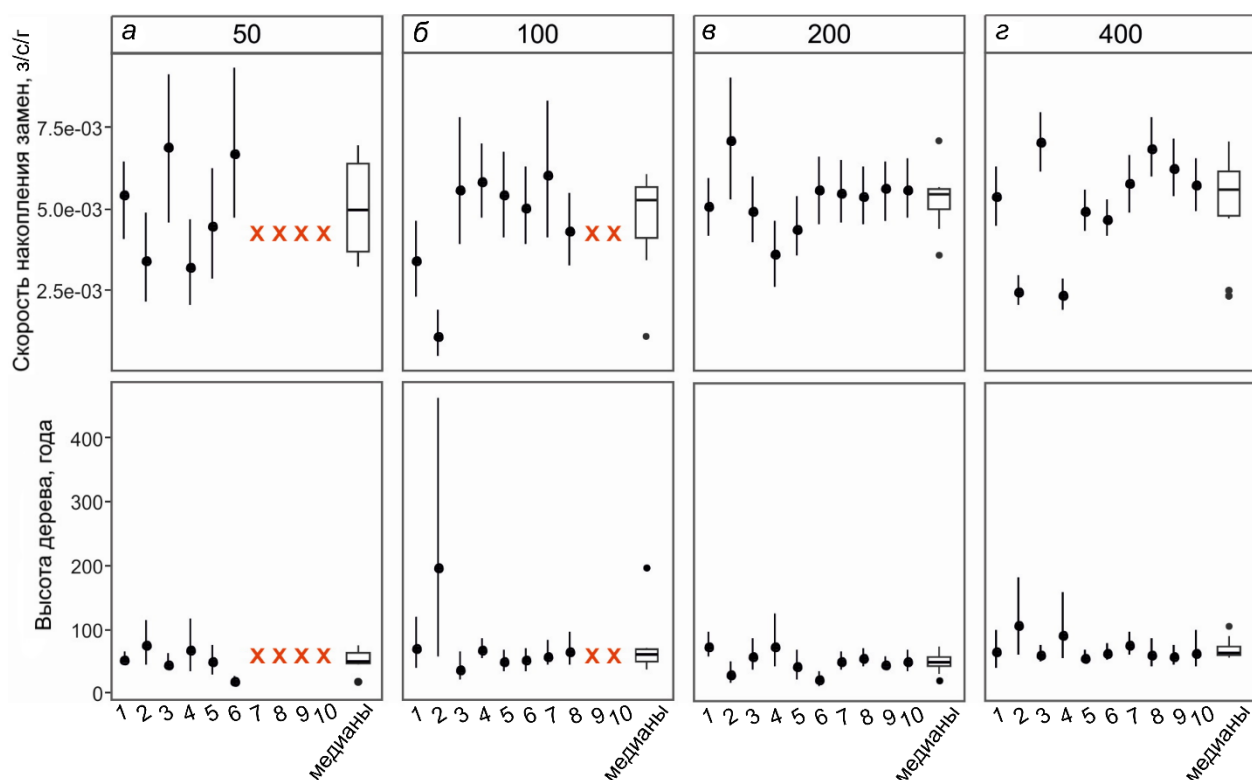


Рис. 38. Влияние количества последовательностей в выравнивании на скорость накопления замен и высоту корня дерева, полученного с помощью Байесовского филогенетического вывода. Медианные значения и 95% интервалы HPD скорости накопления замен и высоты деревьев показаны точками и линиями для каждого из десяти повторений выборки с указанным количеством последовательностей соответственно. Красные кресты обозначают прогоны, которые не сошлись за 200 миллионов поколений. В конце каждой панели даны диаграммы размаха для медианных значений параметров («медианы») (а-г).

Таким образом, проведенный опыт показывает, что неравномерность выборки, присущая большинству филогенетических исследований, оказывает критическое влияние на оценки скорости накопления замен и возраста самого недавнего общего предка – ключевых параметров эволюции пикорнавирусов.

### 3.5.1 Влияние размера набора данных и способа формирования выборки на оценки эволюционных параметров

Анализ тысяч последовательностей вычислительно сложен и не всегда информативен. Во многих случаях необходимо уменьшить набор данных,



при этом максимально сохранив его репрезентативность. Для решения этой задачи можно использовать следующие подходы:

1. Случайный выбор последовательностей.

В этом случае последовательности, отсекаемые в больших исследованиях, например, в расследованиях вспышек, будут чрезмерно представлены, в то время как редкие (но потенциально очень информативные) последовательности могут быть потеряны. Это приводит к значительной вариабельности оценок скорости накопления замен и высоты корня (Рис. 39а), а также оценки динамики численности популяции (Рис. 40а).

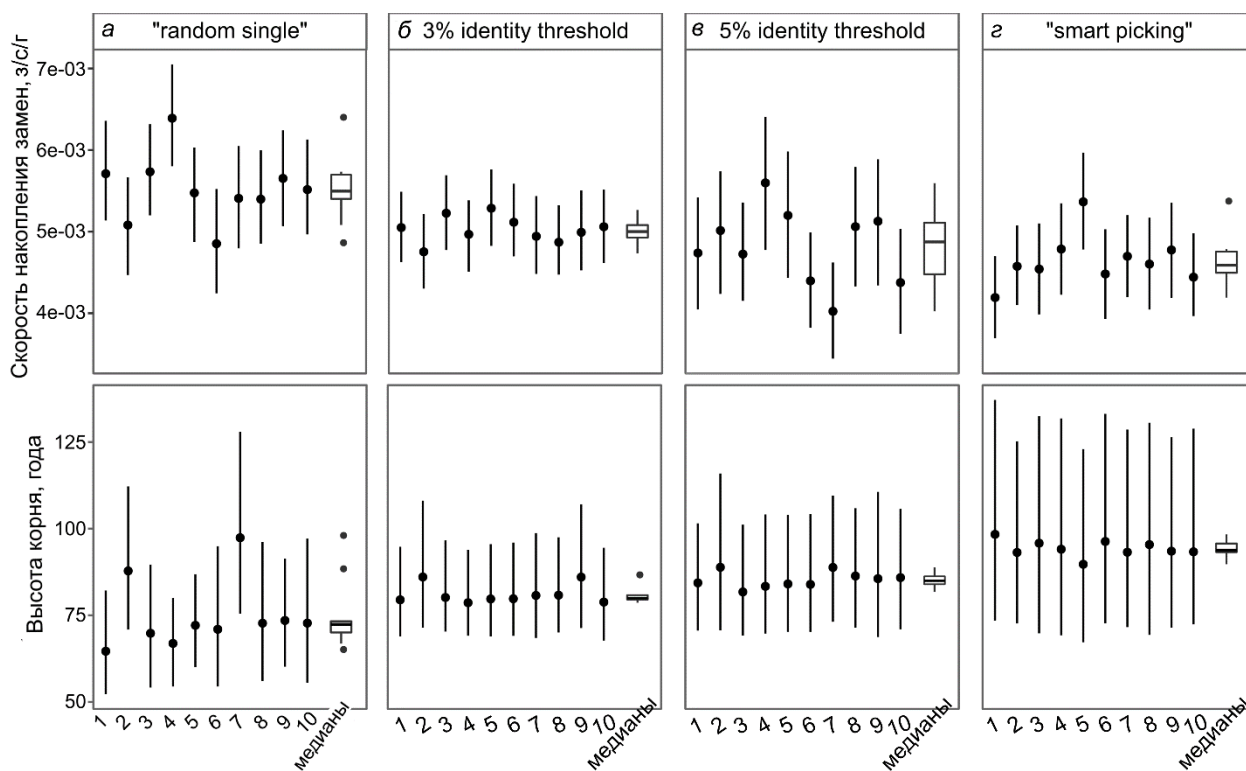


Рис. 39. Влияние алгоритма подготовки выборки нуклеотидных последовательностей на Байесовский вывод скорости накопления замен и высоту филогенетического дерева. Было подготовлено десять выравниваний из 6902 последовательностей EV-A71 VP1 с использованием следующих алгоритмов: выбор 400 случайных последовательностей (“random single”) (а); удаление последовательностей, отличающихся друг от друга менее чем на 3% (380 последовательностей) (б); и 5% (95 последовательностей) (в); и алгоритм «smart picking» (447 последовательностей) (г). Медианные значения и 95% интервалы HPD скорости накопления замен и высоты деревьев показаны точками и линиями. В конце каждой панели даны диаграммы размаха для медианных значений параметров («медианы»).

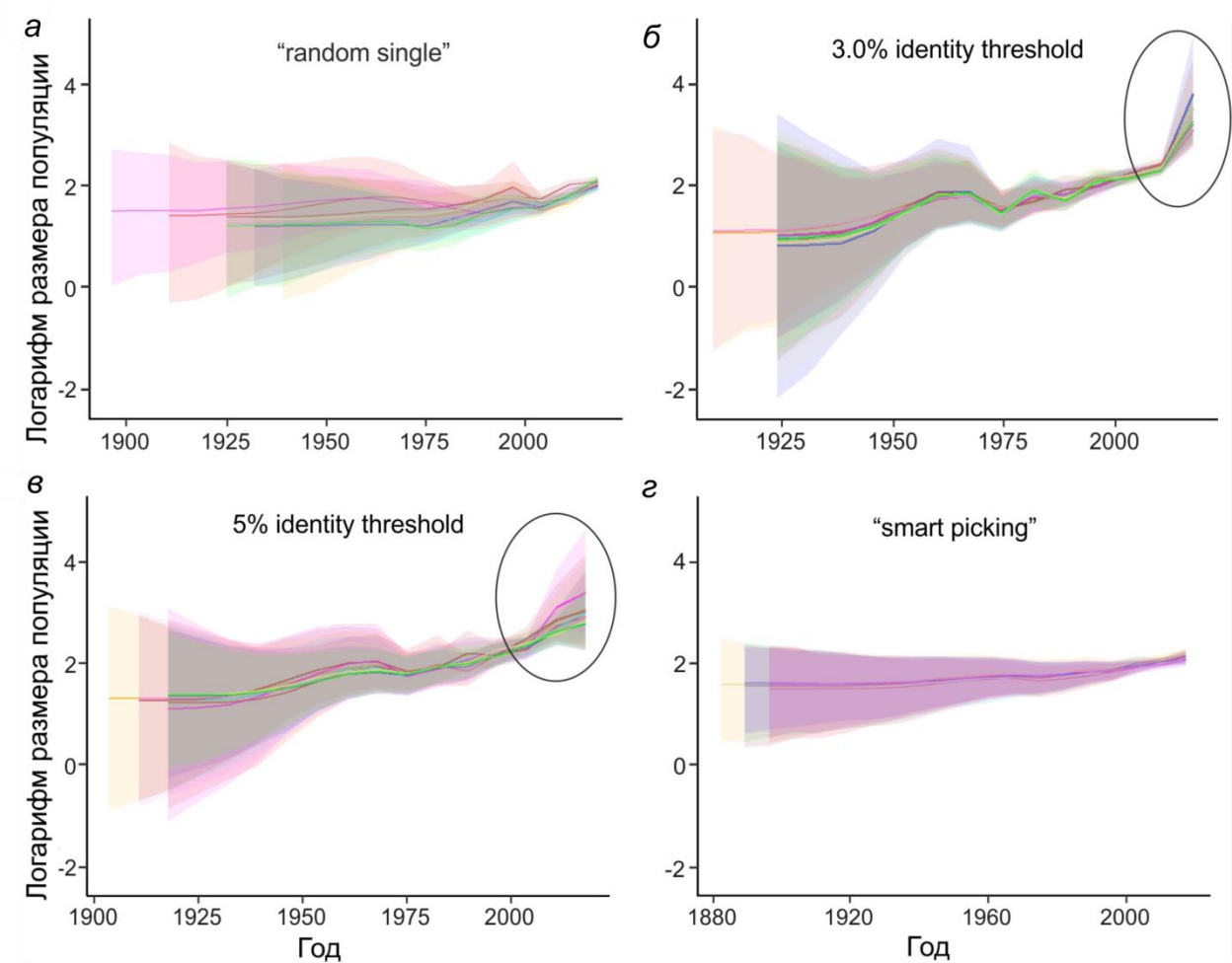


Рис. 40. Влияние алгоритма подготовки набора данных на результаты реконструкции динамики популяции вирусов с помощью модели Bayesian Skygrid. Использовались следующие алгоритмы построения выборок: выбор 400 случайных последовательностей (“random single”) (а); удаление последовательностей, отличающихся друг от друга менее чем на 3% (380 последовательностей) (б); и 5% (95 последовательностей) (в); и алгоритм «smart picking» (447 последовательностей) (г). 10 повторений для каждого алгоритма показаны разными цветами. Заштрихованные области указывают на 95% интервалы НРД оценок численности популяции. Удаление похожих последовательностей привело к ложному взрывному росту популяции в момент времени, соответствующий порогу минимального различия последовательностей (обведено).

## 2. Удаление схожих последовательностей.

Этот метод предполагает удаление из выборки похожих (до 3–5% различий) последовательностей. Подход реализован в программах Jalview [325], CD-HIT [326], UCLUST [327], инструменте skipredundant программного обеспечения EMBOSS [328]. Такой подход гарантирует, что сокращенный набор данных будет столь же информативным с точки зрения генетического

разнообразия, как и исходный, поскольку все редкие последовательности будут сохранены. Фильтрацию схожих последовательностей можно использовать в большинстве филогенетических исследований, потому что она проста и приводит к гораздо более ограниченной вариабельности оценок эволюционных параметров, чем случайный выбор последовательностей (Рис. 39б). Однако сокращение выборки с чрезмерно высоким порогом идентичности последовательностей может привести к увеличению вариабельности эволюционных параметров (Рис. 39в). В случае статистических филогенетических исследований фильтрация по идентичности сама по себе может внести систематическую ошибку. Прежде всего, это может привести к артефактам в результатах Байесовского вывода изменения численности вирусной популяции, потому что искусственное удаление схожих последовательностей приводит к удалению самых поздних узлов филогенетического дерева и, таким образом, моделирует взрывной рост популяции в то время, которое соответствует порогу сходства (Рис. 40б, в, обведено).

### 3. «Smart picking».

В качестве альтернативы методам сокращения выборки, описанным выше, в диссертационной работе был предложен метод «smart picking». Он заключается в следующем. Сначала все последовательности разделяются на группы согласно первым пяти символам (две буквы и три цифры) идентификатора GenBank. Такие группы последовательностей, как правило, соответствуют отдельным исследованиям, потому что при депонировании последовательностей в GenBank им часто назначают близкие идентификаторы. Отбор последовательностей осуществляется следующим образом. Сначала в выборку последовательностей включаются группы, содержащие небольшое количество последовательностей (порог задает пользователь). Затем в выборку добавляется определенный процент случайных последовательностей из каждого из более крупных исследований. Таким образом, снижается вероятность потери редких последовательностей. При таком методе сокращения размера выборки различия оценок эволюционных параметров были менее

выраженными в отличие от других методов (Рис. 39г), а оценки изменения численности вирусной популяции были воспроизводимы, и в них не были обнаружены артефакты (Рис. 40г).

Выбор между двумя последними подходами следует делать, исходя из исходных данных и задач исследования. В любом случае выравнивание последовательностей должно быть дополнительно проверено вручную на возможные ошибки секвенирования, такие как сдвиг рамки в кодирующей последовательности, участки неестественной гетерогенности в определенных последовательностях и т. д., которые встречаются в записях GenBank и могут повлиять на анализ.

### 3.5.2 Влияние ошибок секвенирования на оценки эволюционных параметров

Качество секвенирования может отличаться в разных лабораториях, и наличие последовательностей с ошибками секвенирования может повлиять на результат филогенетического анализа. Для того чтобы проверить чувствительность Байесовских филогенетических методов к ошибкам в последовательностях вирусов, были подготовлены выравнивания, которые включали последовательности с искусственно введенными заменами. В качестве исходного набора данных было использовано выравнивание полного участка VP1 генотипа B1 EV-A71 ( $n = 56$ ). Для того чтобы имитировать ошибки секвенирования, в одну последовательность (891 нт) выравнивания вводили 5, 10, 15 или 20 случайных замен, или 0.56%, 1.1%, 1.7% и 2.2% изменений последовательности соответственно. Затем проводили Байесовский филогенетический анализ. Анализ был проведен независимо для вируса с базальным положением на дереве (AB524092, изолят 1973 г. из Нидерландов, далее именуемый «базальным») и изолята из поверхностной клады с несколькими близкородственными вирусами (AF135866, изолят 1980 г. из США, далее именуемый «верхушечный»).

Регрессионный анализ расстояний от корня до листьев филогенетического дерева и дат выделения изолятов является классическим подходом для

обнаружения последовательностей вирусов с возможными ошибками. Также график соответствия расстояния от корня до листа и даты выделения вируса используется для того, чтобы продемонстрировать временную структуру популяции, или наличие временного сигнала. Наличие временного сигнала является необходимым условием для того, чтобы проводить статистический филогенетический анализ с моделями молекулярных часов [309]. Отклонение от линейного тренда было заметно только для «верхушечной» последовательности AF135866 с 20 введенными мутациями (Рис. 41а). Таким образом, классический метод оценки качества данных не выявлял последовательности количеством ошибок до 2% от общей длины. Был проведен Байесовский филогенетический анализ выравниваний с ошибками в «базальной» или «верхушечной» последовательностях. Скорости накопления замен на ветвях были извлечены из МСС деревьев с помощью скрипта на языке Python. Поскольку для филогенетического вывода использовались ослабленные логнормальные молекулярные часы (relaxed lognormal clock), скорости накопления замен на ветвях имели логнормальное распределение. Для преобразования скоростей к нормальному распределению, более подходящему для оценки экстремальных значений, в дальнейшем использовались десятичные логарифмы скоростей накопления замен. Эффект введения мутаций в одну последовательность из выравнивания на среднюю скорость накопления замен по веткам дерева, скорость накопления замен на ветке, ведущей к ошибочной последовательности и среднее квадратическое отклонение скоростей замен на ветках дерева представлен на Рисунке 41б. Введенные мутации оказали незначительное влияние на скорость накопления замен ветви, ведущей к «базальной последовательности», но сильно повлияли на среднюю скорость и скорость накопления замен ветви «верхушечной» последовательности. В случае введения мутаций в «верхушечную» последовательность, построение графика распределения скоростей накопления замен на ветках дерева позволило четко идентифицировать ошибочную последовательность всего с пятью ошибками секвенирования (0,56%) (Рис. 41в).

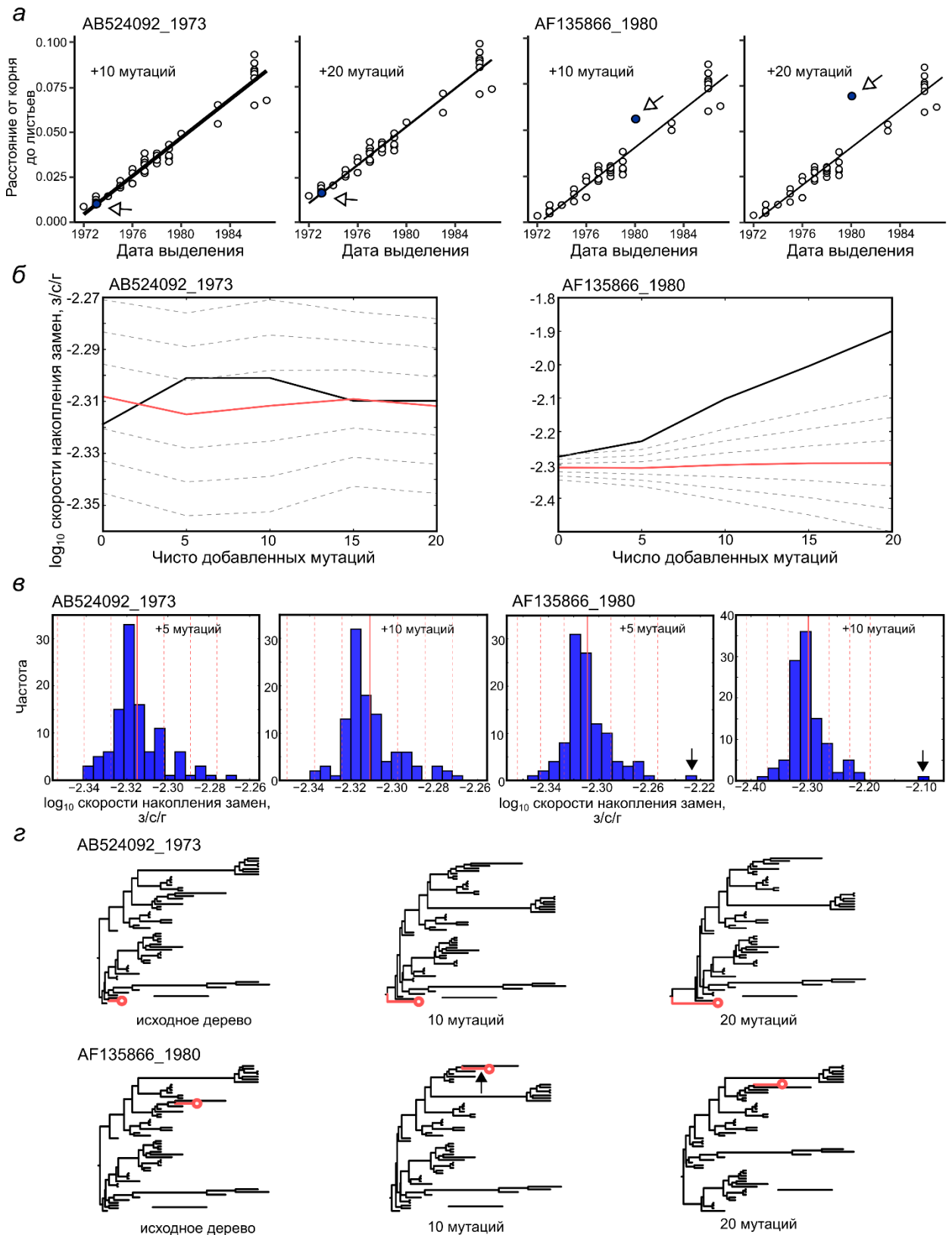


Рис. 41. а – Анализ временной структуры генотипа B1 EV-A71 после введения 10 и 20 случайных мутаций в нуклеотидные последовательности VP1 AB524092 и AF135866. Синие точки соответствуют изолятам, в последовательности которых были введены мутации. б – Влияние добавления случайных мутаций в последовательности AB524092 и AF135866 на скорости замен, полученные в Байесовском филогенетическом анализе генотипа B1 EV-A71.

Средняя скорость накопления замен на дереве показана жирной красной линией; скорость накопления замен на ветке, ведущей к измененной последовательности, черной линией; прерывистые линии отражают среднюю скорость  $\pm 1$ СКО, 2СКО и 3СКО. в – Частотное распределение десятичных логарифмов скоростей накопления замен на ветках деревьев генотипа В1 EV-A171, которые включали последовательности АВ524092 и АF135866 после добавления случайных мутаций. Сплошные и пунктирные красные линии обозначают средние скорости накопления замен и средние  $\pm 1$ СКО, 2СКО и 3СКО соответственно. Стрелка указывает на скорость накопления замен, ведущих к АВ524092 и АF135866 после добавления 5 и 10 случайных мутаций. г – Топология Байесовских филогенетических деревьев после добавления случайных мутаций в последовательности АВ524092 и АF135866. Под деревом указано количество добавленных мутаций. Ветка, ведущая к последовательности с мутациями, окрашена в красный цвет и заканчивается кружком. Черные стрелки указывают на ветку, ведущую к изоляту АF135866, который имел аномальную скорость накопления замен на рисунке (в). Длины ветвей деревьев соответствуют времени. Длина отрезка-индикатора масштаба соответствует 5 годам.

Таким образом, влияние ошибок секвенирования на результаты статистического филогенетического анализа может зависеть от позиции последовательности на дереве и наличия близкородственных последовательностей.

### 3.5.3 Влияние ошибок аннотации дат выделения вируса на оценки эволюционных параметров

При построении деревьев с разверткой по времени Байесовские филогенетические методы используют даты выделения изолятов. Для того чтобы оценить эффект ошибок в датах выделения были проанализированы выравнивания полного участка VP1 генотипа В1 EV-A71, в которых дата выделения одной последовательности изменена на 5, 10 или 20 лет. Как и в предыдущем анализе, изменения даты выделения проводилось независимо для двух последовательностей – «базальной» и «верхушечной». Ошибка аннотации в 5 лет от фактической даты привела к заметному отклонению расстояния от корня до листа от линии регрессии как для «базальной», так и для «верхушечной» последовательностей (Рис. 42а, синие точки). Поскольку записи с предположительно неправильной аннотацией даты выделения не являются редкостью в GenBank, перед выполнением статистического филогенетиче-

ского анализа рекомендуется провести проверку графика соответствия дат выделения и расстояний от корня до листьев на наличие выбросов.

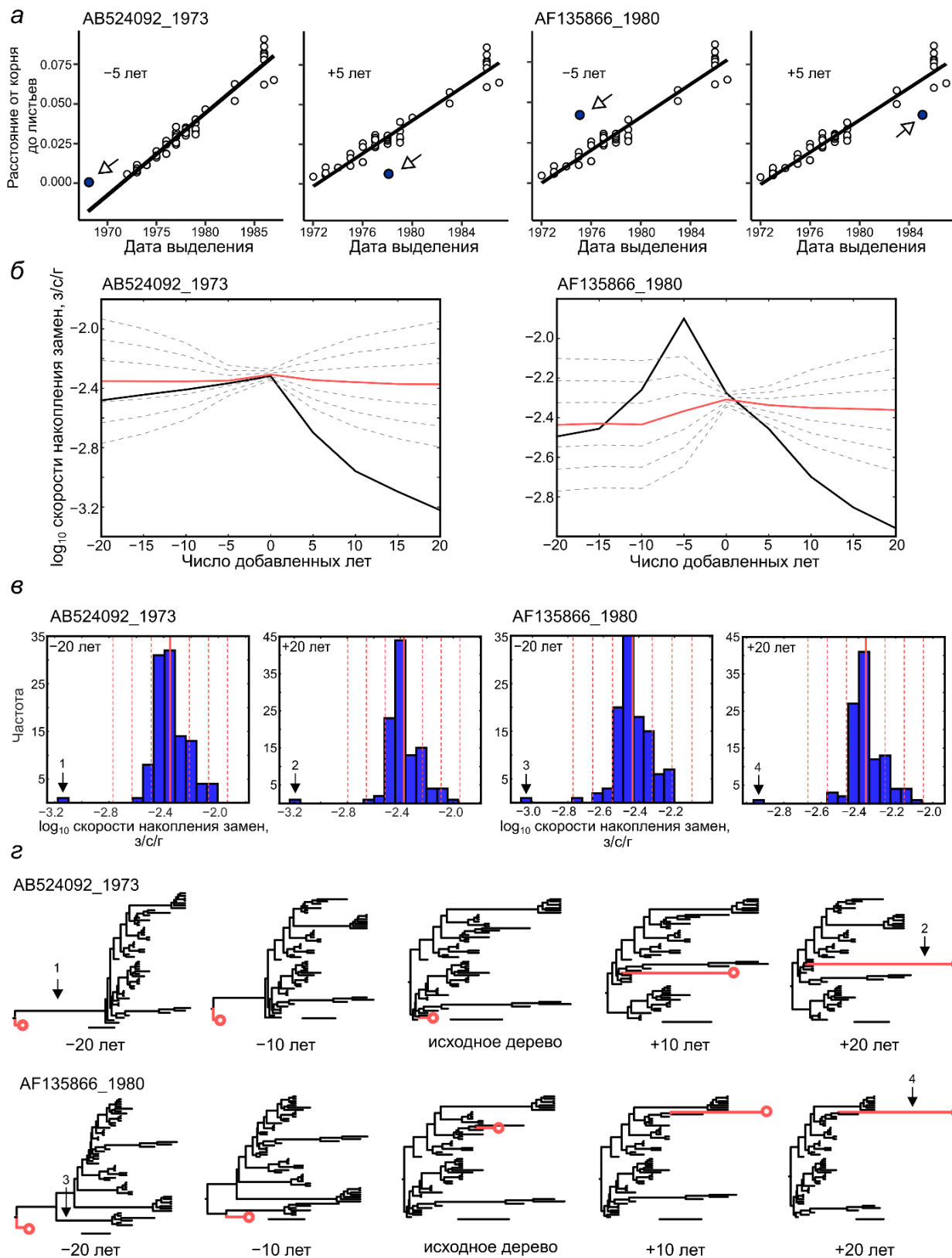


Рис. 42. а – Анализ временной структуры генотипа B1 EV-A71 после вычитания и прибавления 5 лет к дате выделения изолятов AV524092 и AF135866.



Синие точки соответствуют изолятам, в год выделения которых были внесены искусственные изменения. б – Влияние изменения года выделения изолятов АВ524092 и АF135866 на скорости замен, полученные в Байесовском филогенетическом анализе генотипа В1 EV-A71. Средняя скорость накопления замен на дереве показана жирной красной линией; скорость накопления замен на ветке, ведущей к измененной последовательности, черной линией; прерывистые линии отражают среднюю скорость  $\pm 1$ СКО, 2СКО и 3СКО. в - Частотное распределение десятичных логарифмов скоростей накопления замен на ветках деревьев генотипа В1 EV-A171, которые включали последовательности АВ524092 и АF135866 после изменения года выделения на 20 лет. Сплошные и пунктирные красные линии обозначают средние скорости накопления замен и средние  $\pm 1$ СКО, 2СКО и 3СКО соответственно. Стрелка указывает на скорости, отличие от средней для которых превышает 3СО. г - Топология Байесовских филогенетических деревьев после изменения года выделения изолятов АВ524092 и АF135866 на 10 и 20 лет. Ветка, ведущая к последовательности с мутациями, окрашена в красный цвет и заканчивается кружком. Пронумерованные стрелки указывают на скорости накопления замен, которые отклоняются от средней больше чем на 3 СО. Длины ветвей деревьев соответствуют времени. Длина отрезка-индикатора масштаба соответствует 5 годам.

Если не исключить последовательности с неправильными датами выделения, они могут существенно повлиять на дальнейший анализ. Байесовские методы направлены на вывод наиболее вероятной филогенетической истории всего набора последовательностей в соответствии с выбранной эволюционной моделью (модель замен, тип молекулярных часов, модель ветвления дерева) и могут сгладить аномалии, вызванные одной последовательностью. Следовательно, ошибки аннотации отдельных последовательностей могут как влиять, так и не влиять на скорость накопления замен ветки, ведущей к соответствующей последовательности (Рис. 42б, черные линии). Однако даже небольшая ошибка в дате одной последовательности может оказывать существенное влияние на СКО скорости накопления замен ветвей дерева (Рис. 42б, пунктирные линии). Например, ошибка в 5 лет увеличила СКО скорости накопления замен по дереву до шести раз для выравнивания с измененной «базальной» последовательностью и до семи раз для выравнивания с измененной «верхушечной» последовательностью. В то же время более значительная ошибка аннотации (–20 лет) не вызвала изменения средней

скорости накопления замен на ветвях дерева, но увеличила СКО скоростей в два раза. Даже если бы отдельные скорости замен на ветвях не имели бы большого значения для конкретного исследования, увеличение дисперсии параметров могло бы привести к увеличению времени расчетов, что может иметь решающее значение для больших наборов данных.

Несмотря на то, что изменение даты выделения не всегда влияло на скорость накопления замен ветки, идущей к «ошибочной» последовательности, ошибки аннотации приводили к появлению аберрантных ветвей в другом месте дерева (Рис. 42в, г, обозначены стрелками 1, 3). Ошибки в аннотации даты влияли и на общую топологию: «верхушечная» последовательность при уменьшении даты выделения смещалась к корню дерева, а «базальная» последовательность при увеличении даты выделения, напротив, смещалась к более поздним узлам (Рис. 42г).

Таким образом, тщательный отбор данных, проверка последовательностей и дат выделения изолятов, предварительный анализ временной структуры набора данных могут повысить точность статистических филогенетических исследований.

### **3.6 Молекулярная эволюция типов энтеровирусов человека**

3.6.1 Характеристика выборки последовательностей энтеровирусов. Оценка влияния неравномерности выборки на результаты филогенетического анализа

Результаты анализа рекомбинации у энтеровирусов, опубликованные ранее, указывали на независимую эволюционную историю структурных и неструктурных областей генома даже на короткой временной шкале [329]. Это наблюдение стало основанием для изучения молекулярной эволюции типов энтеровирусов человека на основе последовательности участка генома, кодирующего белок капсида VP1, отдельно от остальных частей генома. Для этого был проведен Байесовский филогенетический анализ 32 типов энтеровирусов (Таблица 9). Так как виды энтеровирусов имеют сложную генетическую структуру, и входящие в них типы могут иметь разные закономерности

эволюции и необязательно составляют одну популяцию, сначала каждый тип был проанализирован по-отдельности. Количество последовательностей в выравниваниях типизирующего фрагмента VP1 длиной около 300 нт было в 2–15 раз больше, чем в выравниваниях полного фрагмента VP1 длиной около 900 нт (Таблица 8). Географическое покрытие выравниваний типизирующих фрагментов также было более репрезентативным, поэтому они использовались для филогенетического анализа.

Выравнивания типизирующих фрагментов VP1 для отдельных типов отличались по количеству последовательностей, распределению последовательностей по годам выделения и по странам. Количество последовательностей в выравнивании типизирующих фрагментов VP1 отдельных типов составило от 35 до 629 последовательностей (Таблица 8). Прототипные последовательности, или самые первые выделенные последовательности каждого типа, датируются 1947–1970 годами. Таким образом, время между самым «старым» и самым «молодым» изолятом, или так называемая глубина выборки, составило в среднем 64 года.

Таблица 9. Результаты Байесовского филогенетического анализа

Серотип	Год выдел. прототип. штамма	Год выдел. самого старого непрототипного	Н корня, годы	Н корня, 95% ДИ, года	V замен $\times 10^{-2}$ , з/с/г	V накопления замен $\times 10^{-2}$ , 95% ДИ, з/с/г	Н корня (без прототипа), года	Н корня (без прототипа), 95% ДИ, года	V замен $\times 10^{-2}$ , з/с/г	V замен (без прототипа), 95% ДИ, з/с/г
CV-A2	1947	2000	77	[67, 93]	1,12	[0.74, 1.64]	37	[22, 56]	1.2	[0.74, 1.64]
CV-A4	1948	1998	71	[66, 81]	0,71	[0.52, 0.89]	45	[31, 60]	0.77	[0.57, 0.99]
CV-A5	1950	2000	77	[63, 118]	0,83	[0.44, 1.36]	54	[21, 87]	0.94	[0.48, 1.66]
CV-A6	1949	1992	116	[65, 161]	0,9	[0.72, 1.12]	97	[58, 141]	0.96	[0.78, 1.18]
CV-A9	1950	1962	102	[84, 123]	0,61	[0.50, 0.74]	92	[76, 114]	0.63	[0.49, 0.75]
CV-A10	1950	2001	116	[81, 156]	0,83	[0.63, 1.03]	115	[84, 156]	0.82	[0.65, 1.03]
CV-A13	1952	1989	65	[61, 72]	0.3	[0.2, 4.34]	51	[36, 69]	0.29	[0.18, 3.9]
CV-A16	1951	1981	200	[110, 304]	0,82	[0.66, 0.99]	37	[34, 40]	0.80	[0.67, 0.92]
CV-A24	1952	1985	127	[95, 167]	0,42	[0.29, 0.57]	99	[70, 129]	0.55	[0.37, 0.73]
CV-B1	1948	1984	84	[65, 106]	0,91	[0.70, 1.13]	73	[62, 88]	0.98	[0.81, 1.15]
CV-B2	1947	1975	76	[69, 85]	0,62	[0.51, 0.73]	73	[63, 86]	0.63	[0.51, 0.74]
CV-B3	1949	1970	129	[87, 176]	0,68	[0.55, 0.80]	127	[91, 168]	0.69	[0.56, 0.79]
CV-B4	1951	1972	102	[80, 128]	0,67	[0.55, 0.80]	73	[62, 90]	0.68	[0.54, 0.81]
CV-B5	1952	1954	130	[92, 173]	0,77	[0.66, 0.89]	121	[83, 164]	0.79	[0.67, 0.91]
E11	1953	1963	122	[90, 159]	0,65	[0.54, 0.77]	111	[81, 144]	0.68	[0.56, 0.79]
E13	1953	1982	100	[71, 135]	1,09	[0.83, 1.37]	89	[61, 123]	1.08	[0.82, 1.32]
E14	1954	1980	74	[58, 94]	1,21	[0.87, 1.57]	60	[39, 82]	1.15	[0.83, 1.5 ]
E18	1955	1987	82	[59, 109]	0,62	[0.42, 0.82]	48	[28, 77]	0.75	[0.46, 1.05]

Для последовательностей, которые использовались в филогенетическом анализе, годы выделения вирусов представлены неравномерно (Рис. 43). Большая часть последовательностей гена VP1 из базы данных GenBank была выделена после 2000 года. Самые первые последовательности прототипных штаммов относятся к 1950–1960 годам. Есть временные интервалы (1960–1970) для которых последовательности в GenBank отсутствуют. Это может быть связано с тем, что в те годы преобладали серологические способы типирования энтеровирусов, и после типирования изоляты не хранили в архивах. Большинство вирусов было выделено в странах, где действует надзор за неполиомиелитными энтеровирусами – США, страны Европы, Китай, Индия, Россия и др. Количество последовательностей из Африки и Центральной Америки крайне было мало.

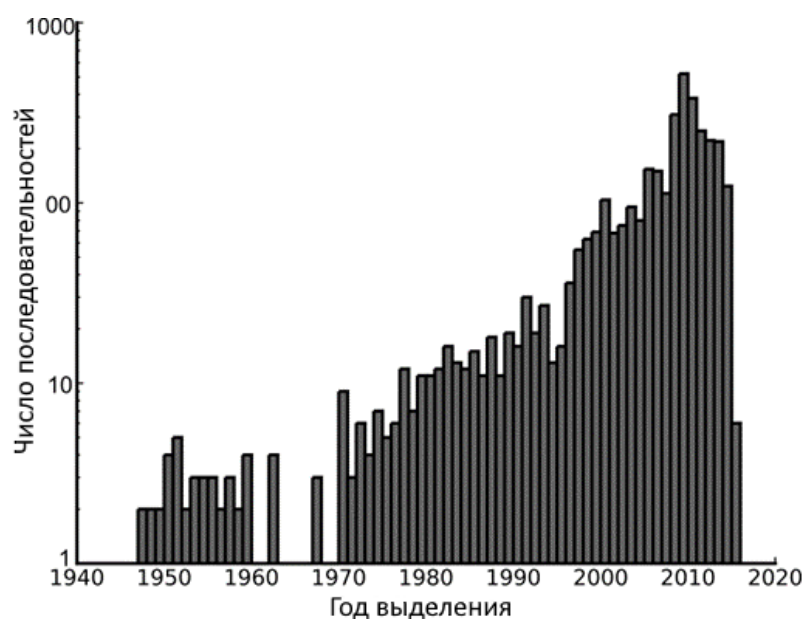


Рис. 43. Распределение количества последовательностей VP1, которые были использованы в филогенетическом анализе, по годам выделения изолятов.

В первую очередь было необходимо оценить, как неравномерность выборок последовательностей VP1 для отдельных типов энтеровирусов влияет на ключевые параметры, выводимые с помощью Байесовского филогенетического анализа: скорость накопления замен и высоту корня, или возраст последнего общего предка исследуемых вирусов. Результаты Байесовского филогенетического анализа по всем 32 вирусам приведены в Таблице 9. Из по-

лученных расчетов следует, что время существования последнего общего предка отдельных типов находится в пределах 55–200 лет назад, а скорость накопления замен у разных типов составляет  $0.4\text{--}1.21 \times 10^{-2}$  з/с/г. При этом наблюдаются следующие закономерности: количество последовательностей в выравниваниях оказывает влияние на высоту корня (Рис. 44а, Spearman  $r=0.56$ ,  $p\text{-value}=0.01$ ), но не влияет на скорость замен (Рис. 44б, Spearman  $r=-0.28$ ,  $p\text{-value}=0.06$ ). Для тех типов, которые представлены более чем 100 последовательностями (11 из 32 типов), статистически значимая корреляция между объемом выборки и высотой корня отсутствует (Spearman  $r=0.33$ ,  $p\text{-value}=0.24$ ).

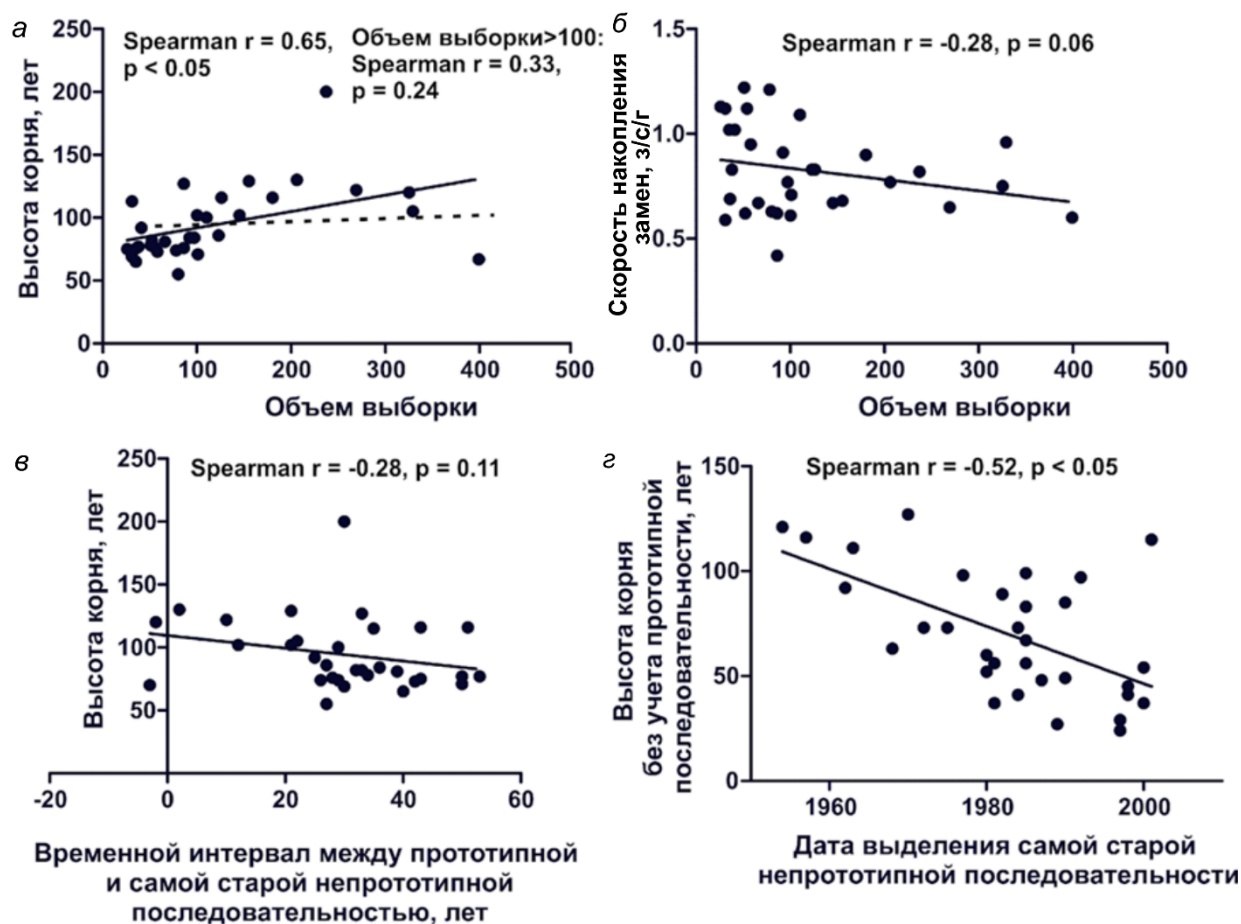


Рис. 44. Влияние размера выборки на высоту корня (а) и скорость замен (б). Влияние временного интервала между временем выделения прототипного штамма и временем выделения самого старого непрототипного изолята на высоту корня (в). Корреляция между высотой корня и глубиной выборки по времени после удаления прототипного штамма (г). Каждая точка соответствует измерениям для отдельного типа.

На основании проведенного исследования можно сделать вывод о том, что выборки, содержащие не менее 100 уникальных (различающихся более чем на 2%) последовательностей достаточны для проведения филогенетического анализа.

Анализ выборок последовательностей типов показал, что для некоторых типов распределение последовательностей по годам исторически более равномерное, чем для других, и промежутки времени между датой выделения прототипного штамма и следующего после него небольшой (EV-A71, E30, CV-B5). Напротив, для других типов такой временной интервал достаточно длинный. Так, например, прототипный штамм CVA2 был выделен в 1947 году, а следующий известный геном датируется 2000 годом. Интервал между временем выделения прототипной последовательности и самой старой не-прототипной составляет от -3 (если архивный изолят не был официально признан прототипным и был секвенирован позже прототипного) до 53 лет. Тем не менее, оказалось, что значения этих временных интервалов не влияли ни на высоту корня (Рис. 44в, Spearman  $r=-0.28$ ,  $p\text{-value}=0.11$ ), ни на скорость замен (Spearman  $r=-0.29$ ,  $p\text{-value}=0.09$ ).

Прототипные штаммы для большинства вирусов были выделены примерно в одно и то же время, между 1947–1955 годами, поэтому выравнивания, содержащие прототипные штаммы, плохо подходят для анализа влияния глубины выборки на высоту корня. Удаление прототипного штамма из выравниваний привело к уменьшению глубины выборки на 0–53 лет (в среднем, на 32 года). После удаления прототипных штаммов расчетная скорость замен незначительно увеличилась (в среднем на  $0.2 \times 10^{-3}$  з/с/г) и коррелировала с уменьшением глубины выборки (Spearman  $r=0.41$ ,  $p\text{-value}=0.02$ ), а время последнего общего предка типов уменьшилось на 2–163 года (в среднем 24.6 лет). Таким образом, при отсутствии в выравнивании прототипного штамма последний общий предок типа оказывается моложе примерно на 25 лет. При удалении из выборки прототипных штаммов наблюдалась корреляция между

глубиной выборки и высотой корня (Рис. 44г, Pearson  $r=-0.52$ ,  $p\text{-value}<0.05$ ) для типов без учета прототипных штаммов.

Таким образом, оценка влияния неравномерности некоторых параметров выборок последовательностей типизирующих фрагментов VP1 для исследуемых типов энтеровирусов показала следующее:

1. для анализа филогенетики типов энтеровирусов (участков генома, кодирующих типизирующий фрагмент VP1) требуется не менее 100 уникальных последовательностей;
2. неравномерная представленность последовательностей по годам практически не влияет на оценку возраста корня;
3. определяющим фактором для оценки корня является возраст самого старого изолята, т.е. общий предок энтеровирусов в среднем на  $N$  лет старше, чем их самый старый изолят типа.

Объяснить влияние величины возраста самого старого изолята только на высоту корня, но не на скорость замен, можно прохождением популяций энтеровирусов через «бутылочные горлышки» и вымиранием линий, что приводит к сбросу возраста корня наблюдаемого дерева.

### 3.6.2 Особенности топологии филогенетических деревьев для отдельных типов энтеровирусов

Если бы глобальная популяция энтеровирусов проходила через «бутылочные горлышки», то это отразилось бы на топологии филогенетических деревьев: все вирусы после гипотетического «бутылочного горлышка» группировались бы вместе на филогенетическом дереве. Действительно, в 16 из 32 типов (CV-A2, CV-A5, CV-A10, CV-A16, E1, E2, E7, E14, E18, E19, E20, E24, E25, E29, E33, EV-D68) прототипный штамм занимал базальное положение, причем в эховирусах чаще, чем в коксакивирусах (Рис. 45, 46). Общий предок (или несколько общих предков) современных вирусов для большинства типов существовал после выделения прототипного штамма в 1950-х годах. Такая топология филогенетических деревьев косвенно подтверждает ги-



потезу о прохождении глобальной популяции энтеровирусов через «бутылочные горлышки» и вымирании большинства вариантов вирусов, которые гипотетически существовали в 1950-х годах. Самые яркие примеры такой топологии можно наблюдать в филогенетических деревьях типов CVA2, CVA5, CVA10, CVA16, E19, E24 и E33 (Рис. 45, 46), у которых последний общий предок современных вирусов существовал 24–54 года назад.

Кроме того, можно выделить ещё несколько типов топологий деревьев. В 8 типах (CVA4, CVA6, CVA9, CVB1, CVB2, CVB3, CVB4, E6) вымирание генетических вариантов середины 20 века было неполным, и небольшая часть современных вирусов – это варианты, ответвившиеся от «мажорной» популяции более 50 лет назад (Рис. 45, 46). В 8 типах (CVA24, CVB5, E3, E11, E13, E21, E30, EV-A71) современные вирусы представляют собой несколько превалирующих разнообразных групп, которые разошлись после 1950-х годов. Эта группа включает наиболее распространенные и изученные типы. Неясно, наблюдаются у этих типов варианты вирусов, ответвившиеся после 1950-х годов, из-за того, что выборки для этих серотипов более репрезентативны, или из-за большей относительной приспособленности этих типов, которая предотвратила вымирание линий и обеспечила их более широкое распространение в человеческой популяции. Таким образом, характер группирования внутри разных типов представляет собой возможные эволюционные сценарии, которые варьируются от балансирования на грани вымирания единственной линии до разделения на несколько глобально распространенных линий.

Более того, в филогенетических деревьях некоторых типов можно найти сразу несколько паттернов группирования. Например, CVA4, CVA10, E19 и EV-D68, по-видимому, прошли через «бутылочные горлышки», о чем свидетельствуют длинные ветви, ведущие от корня к предку современных штаммов (около 1950-х – 1980-х годов), после чего последовало образование нескольких основных линий вирусов. Таким образом, эффект «бутылочного горлышка» и вымирания линий вирусов может происходить в глобальной

популяции энтеровирусов в любое время и с любой частотой, что ведет к преобладанию одной или нескольких линий современных вирусов и гипотетическому исчезновению «старых» линий.

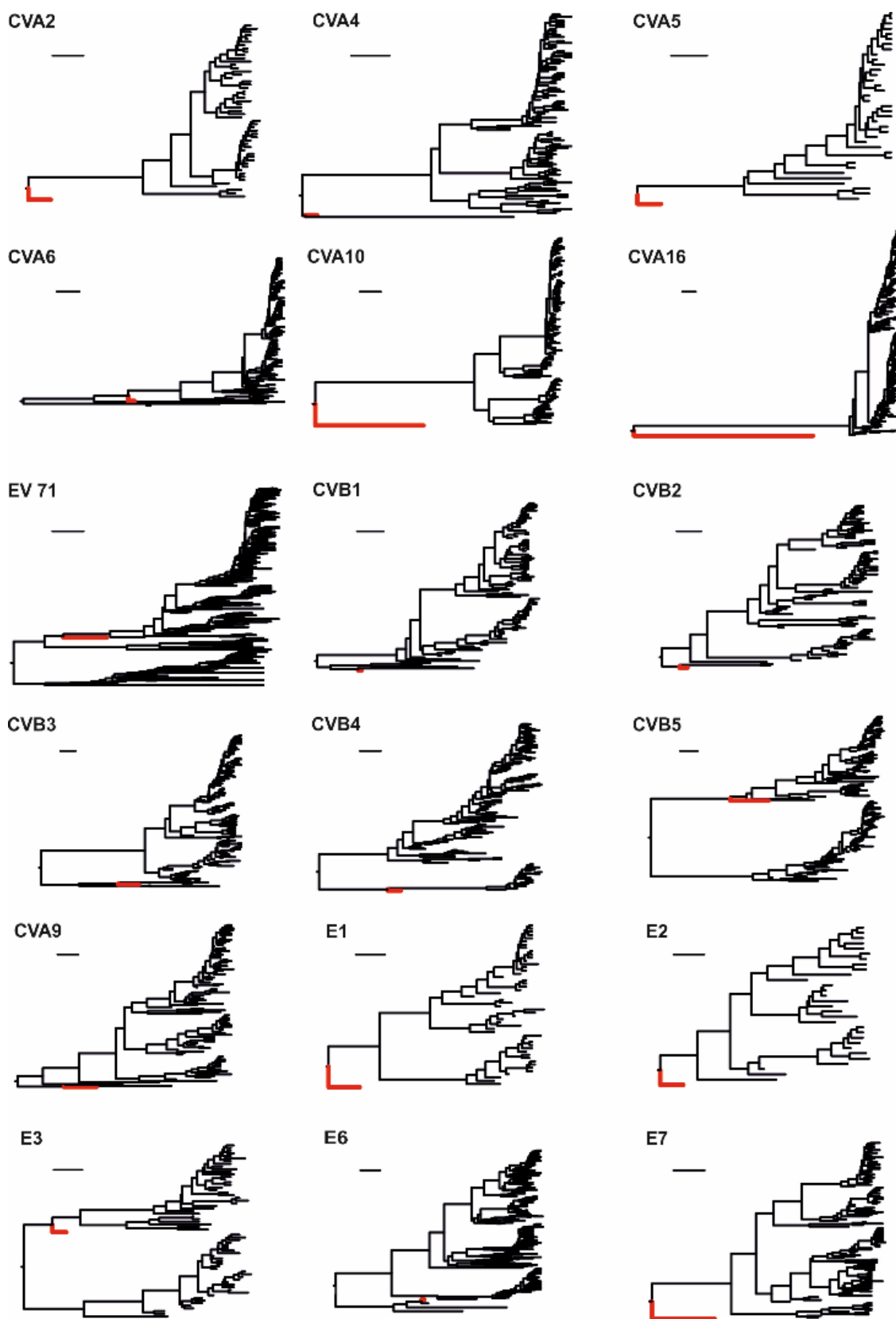


Рис. 45. Филогенетические деревья, построенные по типизирующему фрагменту гена VP1 (294–339 нт). Прототипные штаммы показаны красным цветом. Длины ветвей деревьев соответствуют времени. Длина отрезка-индикатора масштаба соответствует 10 годам.

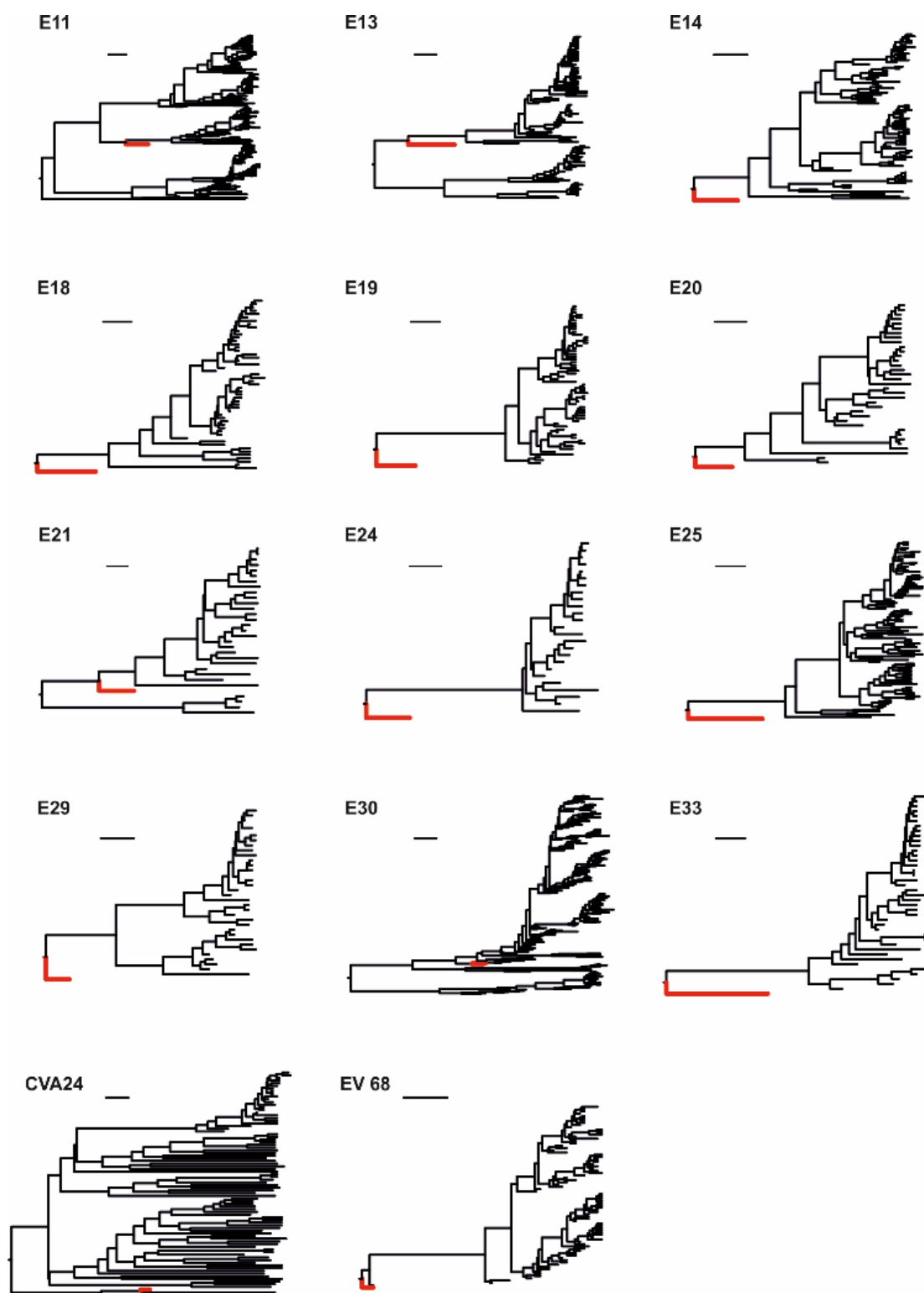


Рис. 46. Филогенетические деревья, построенные по типизирующему фрагменту гена VP1 (294–339 нт). Прототипные штаммы показаны красным цветом.

Длины ветвей деревьев соответствуют времени. Длина отрезка-индикатора масштаба соответствует 10 годам.

### 3.6.3 Филогенетический анализ видов *Enterovirus A* и *Enterovirus B*

Молекулярные часы при прохождении глобальной популяции энтеровирусов через «бутылочные горлышки» могут постоянно сбрасываться. Этот факт не позволяет использовать для определения времени возникновения типа время существования последнего общего предка известных на данный момент вирусов. Для того чтобы понять, действительно ли современные типы появились недавно, или же бесконечно повторяющиеся «бутылочные горлышки» регулярно сбрасывали молекулярные часы, был проведен Байесовский филогенетический анализ на уровне видов.

Важно отметить, что результаты такого анализа нужно интерпретировать с осторожностью. Во-первых, участок генома, кодирующий белок капсида, необязательно представляет целостную популяцию. Во-вторых, механизмы происхождения типов неизвестны, и, возможно, они включают интенсивное накопление мутаций за короткий промежуток времени [187].

Для филогенетического анализа использовались выравнивания полных последовательностей гена VP1 видов энтеровирусов А и В, причем последовательности с идентичностью более 95% (EV-A) и более 93% (EV-B) были исключены. Выравнивания *Enterovirus A* и *Enterovirus B* содержали 163 и 441 последовательность соответственно. Скорости замен составили  $4.0 \times 10^{-3}$  [95% интервал HPD  $3.44 \times 10^{-3}$ – $4.55 \times 10^{-3}$ ] и  $4.1 \times 10^{-3}$  [ $3.68 \times 10^{-3}$ – $4.52 \times 10^{-3}$ ] з/с/г, а высота корня – 687 [547–830] и 574 [473–685] лет для *Enterovirus A* и *Enterovirus B* соответственно (Рис. 47). Дерево имело хорошее разрешение, апостериорные вероятности большинства узлов на уровне выше индивидуальных типов были равны 1.

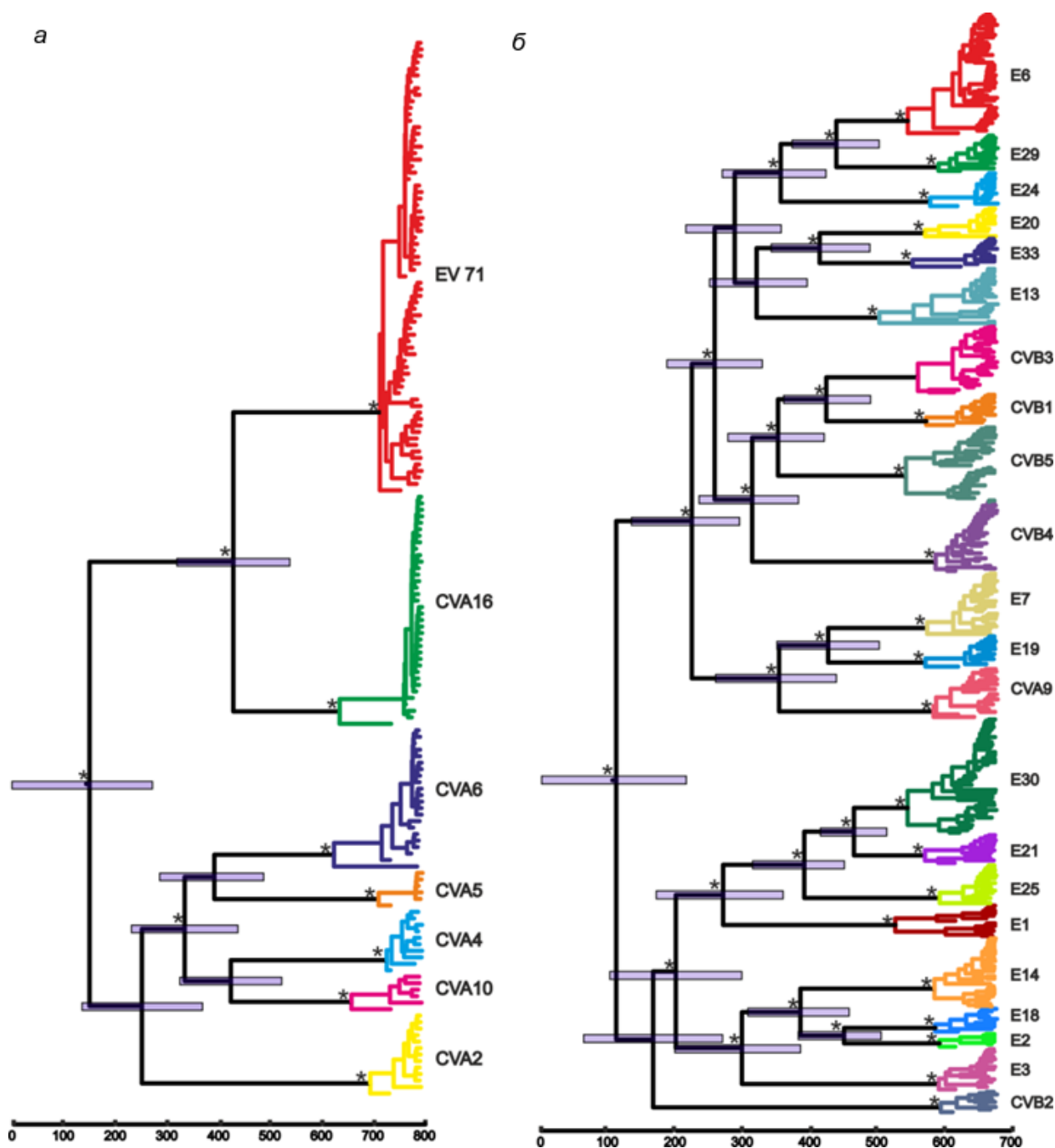


Рис. 47. Филогенетические деревья видов *Enterovirus A* (а) и *Enterovirus B* (б), построенные по полному гену VP1. Шкала показывает время в годах. Планки погрешностей на узлах выше уровня типов показывают 95% доверительный интервал. Типы покрашены разными цветами. Звездочками отмечены узлы с апостериорной вероятностью более 0.9.

Времена последних общих предков отдельных типов, выведенные при анализе полных VP1 последовательностей, попадали в 95% доверительный интервал времен, выведенных с помощью анализа выравнивания коротких последовательностей, но в целом были примерно на 30% больше. Вероятно,

это вызвано тем, что скорости замен, выведенные при анализе видов, находились на нижней границе распределения скоростей, наблюдаемых при анализе отдельных типов.

Ветви, ведущие к отдельным типам, были значительно длиннее (80–427 лет), чем ветви внутри типов. Такая топология деревьев совместима с гипотетическими «бутылочными горлышками», которые могли приводить к вымиранию большинства линий, предшествовавших современным вирусам. С другой стороны, самое недавнее разделение двух отдельных типов (общий предок E21 и E30) датируется всего лишь 214 [167–264] годами назад, а время последних общих предков ряда других пар типов было менее 300 лет назад. Таким образом, регулярное прохождение через «бутылочные горлышки» могло существенно влиять на эволюцию типов, но есть свидетельства, говорящие о том, что известное разнообразие типов энтеровирусов могло возникнуть в течение нескольких последних веков.

#### 3.6.4 Молекулярно-генетический анализ клинических изолятов коксакивируса A2, выделенных от детей с острым вялым параличом

В рамках надзора за неполиомиелитными энтеровирусами с 2001 по 2020 год в 5 из 7280 случаев ОВП у детей от 1 до 4 лет были выделены вирусы типа коксакивирус A2 (CVA2) в 2008 (два случая), 2015 (один случай) и 2019 (два случая) годах. Способность неполиомиелитных вирусов вызывать ОВП до сегодняшнего дня является предметом обсуждений и не доказана окончательно. Для проверки гипотезы о возможном возникновении варианта CVA2 с измененными патогенетическими свойствами в этой работе были определены нуклеотидные последовательности типизирующего фрагмента VP1 выделенных изолятов. В соответствии с практическими рекомендациями ВОЗ, три из пяти случаев можно было отнести к паралитическому полиомиелиту (больные Н.Т., К.О. и Д.А.). Для того чтобы понять филогенетические взаимоотношения этих изолятов в сопоставлении с другими последовательностями CVA2, выделенными в России и других странах мира, доступными в

базе GenBank, было построено Байесовское филогенетическое дерево с раз-  
верткой по времени (Рис. 48). В целом, последовательности CVA2 группиро-  
вались на филогенетическом дереве по географическим регионам, в которых  
они были выделены. На филогенетическом дереве можно выделить две  
большие клады. Одна клада включала преимущественно последовательности  
из Китая, а вторая – из России, Европы, стран Северной и Южной Америк и  
Азии. Среди всех последовательностей CVA2, доступных в Genbank, только  
7 были ассоциированы со случаями ОВП. Эти изоляты были равномерно  
распределены на дереве. Вирусы, выделенные в данном исследовании от  
больных с ОВП, были филогенетически близки с вирусами, выделенными в  
разных регионах России и ассоциированными с острыми респираторными  
заболеваниями, герпетической ангиной, ящуроподобным заболеванием и  
редко – с асептическим вирусным менингитом. Таким образом, филогенети-  
ческий анализ вирусов CVA2, выделенных от больных с ОВП, не выявил их  
связи с конкретной филогенетической группой, поэтому нельзя утверждать о  
возникновении новой линии CVA2, связанной со случаями ОВП.

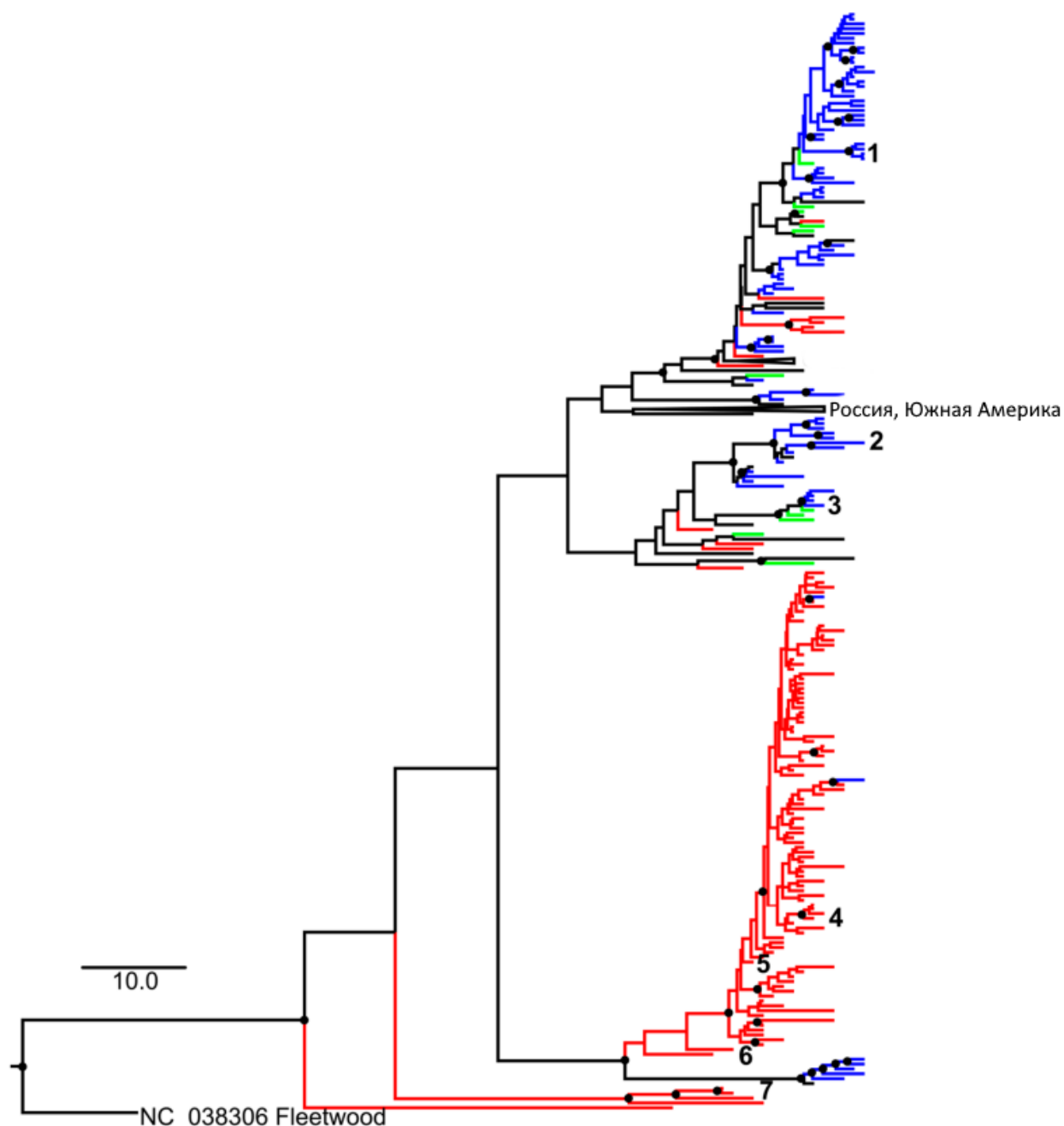


Рис. 48. Байесовское филогенетическое дерево, построенное по последовательностям типизирующего фрагмента VP1 типа CVA2, доступным в базе Genbank. Ветви, ведущие к последовательностям, собранным в России, Восточной Азии (Китай, Япония) и Европейском союзе, окрашены синим, красным и зеленым цветами соответственно. Последовательности изолятов, ассоциированных с ОВП, пронумерованы следующим образом: (1) – больной Н.Т.; 2 – больной К.О.; (3) – больной Д.А.; (4) – 2014 г., изолят из Тайваня [330]; (5) – 2008 г., изолят из Китая; (6), (7) – 2006 г., изолят из Китая. Узлы с апостериорной вероятностью выше 0.95 отмечены кругами. Длина отрезка индикатора масштаба соответствует 10 годам. Клады дерева, не содержащие изолятов, ассоциированных с ОВП, сжаты.



## 4 ОБСУЖДЕНИЕ

### 4.1 Общие закономерности естественной рекомбинации вирусов семейств *Caliciviridae*, *Astroviridae* и *Coronaviridae*

Рекомбинация является одним из механизмов образования генетического разнообразия РНК-вирусов. Для того чтобы рекомбинация между вирусами произошла, необходимо соблюдение ряда условий. Во-первых, рекомбинация требует коинфекции клетки. Во-вторых, необходимо, чтобы рекомбинирующие фрагменты генома были совместимы, и рекомбинация давала селективное преимущество новому вирусу, которое позволит ему закрепиться в популяции, или хотя бы не снижала его приспособленность. Экспериментально рекомбинацию внутри клеток широко изучали у коронавируса МНУ и полиовируса. У полиовируса и МНУ механическая частота рекомбинации в эксперименте оценивается в 10–20% и 25% за пассаж соответственно [270, 271]. Экспериментальные исследования МНУ показали, что рекомбинация происходит почти случайным образом по всему геному в течение относительно короткого промежутка времени, но после серийных пассажей рекомбинантных вирусов в культуре тканей выживают рекомбинанты с перекрестами в определенных областях [245], что скорее всего связано с давлением отбора. Таким образом, закономерности рекомбинации, которые наблюдаются в естественных условиях, скорее всего являются результатом экологических ограничений (возможность коинфекции) и отбора жизнеспособных рекомбинантов.

Число работ по изучению рекомбинации у РНК-вирусов исчисляется сотнями, но практически все они посвящены одному роду или, редко, семейству. В данной работе был проведен систематический анализ естественной рекомбинации (+)РНК-вирусов трех семейств – *Caliciviridae*, *Astroviridae* и *Coronaviridae* для того чтобы, используя единую методологию, выявить общие свойства и отличия в профиле рекомбинации у этих широко распространенных и социально значимых семейств вирусов.

Частая рекомбинация на стыке ORF1, кодирующей неструктурные белки, и ORF2, кодирующей главный белок капсида VP1, является характерной чертой норовирусов и учитывается при их типировании [23, 46]. Матрицы филогенетической совместимости и отклонения попарных расстояний (ОПР) в нашем исследовании характеризовались наличием выраженной границы на стыке ORF1/2, указывающей на то, что филогенетические деревья и попарные генетические расстояния, построенные по ORF1, были наименее совместимы с таковыми по областям ORF2-ORF3 (Рис. 18). Полномасштабный анализ выборки последовательностей норовирусов всеми алгоритмами детекции рекомбинации, реализованными в RDP4, также показал, что большинство рекомбинантных фрагментов составляли почти полные области, кодирующие VP1 и VP2 (Рис. 19).

События рекомбинации между НСО и СО генома были обнаружены во всех проанализированных геногруппах норовирусов, но наиболее часто в рекомбинации участвовали человеческие вирусы геногруппы GII. Человеческие норовирусы являются главной причиной острого гастроэнтерита во всех возрастных группах, и чаще всего вспышки заболевания вызывают различные генотипы геногруппы GII [65]. В литературе описаны десятки случаев межгенотипной рекомбинации между вирусами GII (систематический обзор [87]), и считается, что она является ключевым фактором для генерации новых вариантов, вытесняющих из циркуляции старые. В соответствии с этой концепцией, на графиках СПГР попарные генетические расстояния между вирусами GII, образовывали симметричное «облако» в пределах до 40% различий в нуклеотидной последовательности, что может свидетельствовать о многочисленных событиях рекомбинации (Рис. 20). Многие пары вирусов, расстояния между которыми отклонялись от линейного тренда, могли отличаться всего на несколько процентов в одной из областей генома, что говорит о том, что события рекомбинации произошли совсем недавно. Времена полужизни рекомбинантных форм человеческих норовирусов геногрупп GI и GII были впервые оценены в данном исследовании двумя способами и составляли 10.4 и 8.4–11.3 лет соответственно. Таким образом, время полужизни геногрупп

норовирусов, заражающих человека, в целом больше, чем у энтеровирусов (1.3, 9.8, 3.1 лет у типов E9, E11, E30 соответственно [147]), но находится в близком диапазоне.

Кроме геногруппы GII, рекомбинация между HCO и CO была детектирована во всех проанализированных геногруппах. Многие из выявленных в данной работе рекомбинационных событий были ранее описаны в литературе [87]. В рекомбинации участвовали и умеренно родственные вирусы с различиями по нуклеотидной последовательности менее 20%, и более дивергентные вирусы с различиями от 20% до 40% (Рис. 20б). Таким образом, структурные и неструктурные белки даже достаточно дивергентных норовирусов разных генотипов совместимы между собой, а ко-инфекция одной клетки такими вирусами происходила достаточно регулярно для того, чтобы обеспечить наблюдаемую высокую частоту рекомбинации. При этом не наблюдалось рекомбинации между норовирусами, отличающимися более чем на 30% нуклеотидной последовательности в ORF1 и более чем на 40% в ORF2-3. Таким образом, для естественных комбинаций CO и HCO норовирусов характерен полный промискуитет до определенного уровня генетических дистанций и отсутствие рекомбинации для более дивергентных вирусов.

Помимо рекомбинации на стыке ORF1/2, в литературе были описаны единичные события рекомбинации внутри CO генома между вирусами GII, GV [89,92,93,99,100]. В этом исследовании умеренно частая рекомбинация между разными генотипами GII, норовирусами мышей GV также была обнаружена, однако в отличие от рекомбинации между HCO и CO, недавних событий (т.е. с участием вирусов, очень сходных хотя бы в одном участке генома) не было (Рис. 20в). Сообщений о рекомбинации норовирусов внутри ORF1 опубликовано ещё меньше [94,95], что также подтвердилось в нашем исследовании (Рис. 20а).

Все события естественной рекомбинации саповирусов, обнаруженные в этой работе, происходили в ORF1 в месте соединения участков, кодирующих полимеразу и капсидный белок VP1 (Рис. 25), что согласуется с данными из литературы [102–111]. В отличие от норовирусов, у саповирусов и лаговире-

сов VP1 кодируется не отдельной ОРС, а граничит с неструктурным полипротеином в ORF1. Таким образом, место «горячей» точки рекомбинации определялось не границами ОРС, а функциональностью белков.

Практически все события рекомбинации у норовирусов и саповирусов происходили между вирусами одноименных геногрупп, за исключением нескольких событий межгеногрупповой рекомбинации между саповирусами человека [112], норовирусами человека, а также саповирусами свиней [113]. Случаи коинфекции разными геногруппами были показаны и у свиней [68,114], и у людей во время вспышек пищевого гастроэнтерита [115,116]. Тем не менее, межгеногрупповые рекомбинанты, детектированные в данном исследовании, отличались от своих «родителей» более чем на 35% во всех участках генома, что говорит о том, что рекомбинация могла произойти до того, как вирусы дивергировали и стали относиться к разным геногруппам. В данной работе такая возможность обсуждается впервые, поскольку при использовании стандартных инструментов для анализа рекомбинации не очевидно, насколько «давно» она могла иметь место.

В целом, на основании анализа профилей рекомбинации по геному можно предположить, что у саповирусов и норовирусов совместимость СО и НСО разных геногрупп была лимитирующим фактором для наблюдаемой (т.е. зафиксированной в популяции) естественной рекомбинации.

Лаговирусы отличаются от норовирусов и саповирусов меньшим генетическим разнообразием. Если у норовирусов отличия между аминокислотными последовательностями главного белка капсида VP1 вирусов разных геногрупп могут достигать 55% [62], то всё генетическое разнообразие лаговирусов укладывается примерно в 30% различий в нуклеотидной последовательности VP1 [75]. Как и у других калицивирусов, у лаговирусов место стыка полимеразы и VP1 является «горячей» точкой рекомбинации. На матрицах филогенетической совместимости и ОПП видна четкая граница между этими областями (Рис. 27). Многочисленные события рекомбинации между НСО и СО, описанные в литературе, происходили между разными генотипами и во-

влекали и патогенные, и непатогенные генотипы (Таблица 2). Это подтверждают график СПГР между НСО и СО (Рис 28), который характеризуется большим количеством отклоняющихся от линии регрессии точек, причем рекомбинация происходила практически во всем диапазоне генетических дистанций между вирусами разных генотипов.

На графиках СПГР внутри НСО наблюдалась линейная зависимость между генетическими дистанциями с немногочисленными группами точек, отклоняющимися от линейного тренда, принадлежащими сравнениями внутри генотипов GI.1 (RH DV) и GI.3 (RCV). Эти возможные события рекомбинации были подтверждены методами детекции рекомбинации RDP4, и ещё не были описаны в литературе. Внутри СО рекомбинационные события были также крайне редкими, и, помимо описанной ранее рекомбинации между GI.1(RH DV) [119–122], удалось детектировать недавнюю рекомбинацию у лаговировусов зайцев GII.1 (EBHSV).

Таким образом, для калицивирусов в целом характерен обмен целыми блоками неструктурных и структурных генов и крайне редкая рекомбинация внутри участков, кодирующих структурные и неструктурные белки.

У астровирусов, подобно калицивирусам, наименьшая согласованность наблюдалась между деревьями и попарными дистанциями, построенными по НСО (ORF1) и СО (ORF2) генома. При этом график СПГР в ORF1ab и ORF2 характеризовался отклонением большого количества попарных генетических расстояний, образующих «облако». Многие из отклоняющихся попарных расстояний соответствовали событиям рекомбинации, описанным ранее в литературе [145,146,156,166,273,147–152,154,155]. «Облако» генетических дистанций было несимметрично: среди астровирусов встречаются вирусы с очень сходными НСО и дивергировавшими белками капсида, но не наоборот. По этому признаку бóльшая вариабельность рекомбинационно-мобильных структурных генов у астровирусов напоминает таковую у гена шиповидного белка S у коронавирусов. Это могло быть связано как с более быстрой дивергенцией белков капсида, так и с элиминацией части вариантов неструктур-

ных белков из глобальной популяции. Так или иначе, для всех изученных вирусов можно говорить о «независимой» эволюции СО и НСО генома, по аналогии с феноменом, описанным ранее для пикорнавирусов [329].

Несмотря на общее сходство профиля рекомбинации у астровирусов с калицивирусами (более частая рекомбинация между СО и НСО), можно отметить несколько различий. Резким отличием астровирусов от калицивирусов было большое количество событий рекомбинации внутри СО и НСО, причем с вовлечением достаточно дивергентных вирусов, отличающихся на 20-30% нуклеотидной последовательности (Рис. 31). Ещё одним важным отличием является то, что у астровирусов не было четкой границы между блоками генов в плане рекомбинационной мобильности. Наблюдалось много событий рекомбинации рядом с границей СО и НСО, но не точно по ней, как у калицивирусов. Очевидно, что с одной стороны имеет место та же модульная динамика генома, но закономерности биологической совместимости генов отличаются – нет признаков строгой несовместимости фрагментов внутри СО или НСО между разными вирусами. Это наблюдение согласуется и со значительной частотой рекомбинации внутри СО и НСО у астровирусов.

В этом исследовании были обнаружены разные закономерности рекомбинации у коронавирусов млекопитающих и птиц в разных областях генома, которые, вероятно, являются следствием разной совместимости фрагментов генома. Согласно матрицам филогенетической совместимости и матрицам ОПП, ген белка S был наиболее часто рекомбинантным по сравнению с остальными участками генома. В рекомбинации между ORF1ab и S, особенно у  $\alpha$ -CoV и  $\gamma$ -CoV, могли участвовать значительно менее сходные вирусы (с нуклеотидной дистанцией до 40%) по сравнению с рекомбинацией внутри ORF1ab. Некоторые из таких событий были описаны ранее в литературе [193–195, 209, 220], и это исследование подтверждает, что рекомбинация между ORF1ab и S – это систематическая закономерность. Несмотря на то, что гены E, M и N являются структурными, они вели себя аналогично ORF1ab, а не S.

Барьеры совместимости генетических фрагментов, по-видимому, не действуют на генетических расстояниях ниже 3–10% для  $\alpha$ -CoV и  $\beta$ -CoV и 10–25% для  $\gamma$ -CoV, потому что ниже этих пределов рекомбинация была повсеместна даже в пределах консервативных доменов белка S. У  $\gamma$ -CoV генетические расстояния до 10–25% образуют симметричные «облака генов», что не наблюдается у других родов, вероятно, из-за того, что на малых генетических расстояниях отклоняющиеся точки неотличимы от статистического шума.

Пределы различий в нуклеотидной последовательности вирусов, участвующих в рекомбинации, могут зависеть от возможности и частоты коинфекции. Практически все  $\gamma$ -CoV, проанализированные в данной работе, были выделены из единственного хозяина – курицы (Рис. 32). Торговля птицей и сельскохозяйственными животными, а также иммунизация живыми вакцинами могут повысить вероятность коинфекции и частоту рекомбинации [253]. Известные  $\alpha$ -CoV и  $\beta$ -CoV были выделены из многих видов хозяев. Рекомбинация между вирусами, заражающими разных хозяев, требует пересечения видового барьера, который отсутствует у  $\gamma$ -CoV кур. Большинство случаев рекомбинации среди  $\alpha$ -CoV и  $\beta$ -CoV, выявленных в этом исследовании и описанных в литературе [191,192,195,196,208,274], происходили между вирусами, заражающими одного и того же хозяина. Как в нашем исследовании, так и в литературе есть примеры рекомбинации между вирусами разных хозяев [180,194,209–212,220,275–277]. Однако приведенный выше анализ показывает, что частота таких событий, по-видимому, была недостаточной для формирования «облака генов», как это наблюдается у  $\gamma$ -CoV во всех областях генома. Таким образом, специфичность к определенному виду хозяина (и, следовательно, вероятность коинфекции) определяет возможность формирования «облака генов» на близких генетических дистанциях у коронавирусов.

Дивергировавшие белки S обычно наблюдались у вирусов с близкородственными ORF1ab, но не наоборот. Лучше всего это иллюстрирует рекомбинация между коронавирусами курицы и индейки ( $\gamma$ -CoV), при которой ге-

ны *S* этих коронавирусов различались более чем на 40% при схожей ORF1ab, но не наоборот [249]. Неизвестно, действительно ли разнообразие генов *S* в биосфере больше, чем других генов коронавирусов, или же такой результат связан с тем, что нами пока что не изучено всё разнообразие коронавирусов в биосфере. Первое предположение может означать, что *S* и остальная часть генома имеют разные эволюционные закономерности внутри рода: коровые гены (ORF1ab-*E-M-N*) сцеплены вместе, а *S* может чаще и с меньшими ограничениями переходить к новым вирусам, в том числе заражающим другие виды, что приводит к его большому общему разнообразию в биосфере.

Закономерности рекомбинации внутри гена *S* также различались у родов коронавирусов. У  $\gamma$ -*CoV S* представляет собой гомогенное «облако генов», по-видимому, без барьеров совместимости между доменами *S* и субдоменами S1-NTD и S1-CTD. Такая закономерность сообщалась ранее [202,278] и наблюдалась в данном исследовании только между достаточно близкородственными вирусами (менее 20–25% расстояния нуклеотидной последовательности). Обмены между более дивергентными коронавирусами курицы и индейки затрагивали только полный ген *S* [249]. Точно так же в  $\alpha$ -*CoV* рекомбинация между далекими вирусами затрагивала полноразмерный *S* [192,195] (Рис. 37), а внутри *S* была менее распространенной и несистематической. У  $\beta$ -*CoV* между отдаленно родственными вирусами чаще передавался S1 или даже S1-NTD, а не полный ген *S* (Рис. 37) [336]. Это довольно неожиданно, поскольку, по крайней мере, у SARS-подобных коронавирусов рецептор-связывающий домен картирован в S1-CTD [279,280]. Это наблюдение предполагает, что у  $\beta$ -*CoV* существует ограничение совместимости между S2 и ORF1ab, и предпочтительнее рекомбинация внутри *S*, а не с участием полного *S* (как у  $\alpha$ -*CoV* и  $\gamma$ -*CoV*). Модульная структура *S* обсуждалась ранее на основании экспериментов по реконструкции рекомбинантных коронавирусов и заменой отдельных фрагментов шиповидного белка [271]. Данное исследование показывает, что обмен доменами *S* происходит систематически в масштабе рода во всех родах коронавирусов в естественных условиях.



Таким образом, у коронавирусов наиболее часто рекомбинация происходит между геном структурного белка S и другими участками генома. Гены остальных структурных белков (N, E, M) по профилю естественной рекомбинации соответствуют неструктурным генам. Рекомбинация внутри гена белка S происходит чаще, чем в неструктурной области, причем с большей частотой происходит обмен целыми доменами белка S.

#### **4.2 Влияние способа формирования выборки нуклеотидных последовательностей и артефактов выборки на результат Байесовского филогенетического анализа**

Изучение динамики естественной рекомбинации требует уверенного датирования событий в эволюционной истории вируса. Байесовские филогенетические методы вошли в практику около 15 лет назад и широко используются в исследованиях эволюции и молекулярной эпидемиологии вирусов. В отличие от других филогенетических методов, Байесовские методы предоставляют возможность применять более сложные модели молекулярных часов, динамики изменения популяции и распространения вирусов, а также позволяют использовать большой спектр данных о вирусных образцах, помимо нуклеотидных или аминокислотных последовательностей [340]. Вопрос выбора адекватной исследуемым данным математической модели стоит перед любым филогенетическим исследованием. Сложность моделей в Байесовском филогенетическом выводе накладывает особо строгие требования к процедуре выбора подходящей модели. Так модель с избыточным количеством параметров может привести к высокой дисперсии выводимых параметров, а выбор определенной модели ветвления дерева может приводить к систематическому уменьшению или увеличению высоты корня. Практические рекомендации для выбора моделей с описанием возможных трудностей хорошо описаны в литературе, а статистические методы выбора модели для Байесовского филогенетического анализа реализованы в виде удобного программного обеспечения [282,283].

Напротив, правила для подготовки выборки последовательностей для филогенетического анализа не формализованы. Число доступных геномных данных для разных таксонов вирусов может отличаться на несколько порядков, и неизвестно в какой степени неравномерное семплирование вирусов может влиять на результаты филогенетической реконструкции. Для того чтобы ответить на этот вопрос, в данном исследовании была смоделирована ситуация, типичная для мало секвенируемых вирусов, когда было проведено всего несколько исследований молекулярной эпидемиологии вируса. На основе выборки последовательностей EV-A71, одного из самых широко секвенируемых энтеровирусов, были составлены выравнивания, включавшие от 50 до 400 последовательностей, и был проведен Байесовский филогенетический анализ. Только в выравниваниях с количеством последовательностей более 100 удалось добиться сходимости филогенетического анализа во всех повторностях опыта (Рис. 38). Это свидетельствует о том, что даже значительное количество последовательностей, полученное в одном или нескольких исследованиях, может быть недостаточным для Байесовского филогенетического вывода. При этом 95%-интервалы НРД скоростей накопления замен и возрастов корня деревьев не всегда перекрывались даже для выравниваний, содержащих 400 последовательностей. Таким образом, к выводу эволюционных параметров мало изученных вирусов необходимо относиться с осторожностью.

Ещё одной важной проблемой является вопрос сокращения выборки вирусных последовательностей для филогенетического анализа. Для вирусов с высоким медицинским или ветеринарным значением количество доступных последовательностей превышает тысячи, однако в связи с более высокой вычислительной сложностью Байесовский филогенетический анализ таких больших выборок может занимать недели и иметь плохую сходимость. Нередко для того, чтобы не анализировать большое количество последовательностей, исследователи отбирают последовательности изолятов вручную [120,124]. Существуют и автоматические подходы, например, исключение из выборки похожих последовательностей [266,267,269,284]. Влияние способа

сокращения выборки на результаты филогенетического анализа неизвестно. В данном исследовании было рассмотрено три способа уменьшения выборки: случайный выбор последовательностей, удаление похожих последовательностей с разными порогами схожести, а также предложен новый подход, который позволяет снизить вероятность потери редких последовательностей и сократить число последовательностей из больших исследований. При последнем подходе сокращения выборки различия оценок скоростей накопления замен и высоты корня дерева были менее выраженными в отличие от других методов, а оценки изменения численности вирусной популяции были воспроизводимы (Рис. 39г, Рис. 40г). Удаление схожих последовательностей, как и предложенный нами подход, гарантирует сохранение редких последовательностей в выборке. Однако при максимальном пороге различия в 5% 95%-интервалы HPD скоростей накопления замен перекрывались не во всех повторностях (Рис. 39в). Также удаление схожих последовательностей приводило к систематической ошибке в результатах Байесовского вывода изменения численности вирусной популяции: на графиках наблюдался взрывной рост популяции в то время, которое соответствует порогу отсечки, разделенному на скорость накопления замен для данного вируса (Рис. 40б,в). Самые несогласованные результаты филогенетического вывода в разных повторностях наблюдались для первого подхода – случайного выбора последовательностей (Рис. 39а, Рис. 40а). Таким образом, способ сокращения последовательностей в выборке для филогенетического анализа может оказывать существенное влияние на результаты и даже приводить к систематическим ошибкам.

При анализе больших выборок геномных последовательностей увеличивается вероятность пропустить последовательности с ошибками в аннотации. Например, в базе данных нуклеотидных последовательностей GenBank, которую часто используют в качестве источника последовательностей для анализа, могут быть неправильно отмечены даты забора образцов. Также возможно случайное включение в выборку лабораторных или вакцинных штаммов, которые много лет хранились в замороженном виде. Для модели

молекулярных часов «правильной» датой для таких образцов является дата выделения от пациента, а не дата их секвенирования. Так как Байесовский вывод опирается на априорные данные о последовательностях и датах и выделения, проблемы с качеством данных могут влиять на результаты Байесовского филогенетического анализа порой без каких-либо указаний на то, что эти проблемы с качеством присутствуют [343,344]. Кроме того, качество секвенирования может отличаться в разных лабораториях. Одной из задач данной работы было оценить степень влияния возможных ошибок секвенирования и ошибок в датах выделения на филогенетический анализ. Для этого использовалось выравнивание последовательностей полного гена VP1 генотипа B1 EV-A71 и ошибки моделировались независимо в двух последовательностях – близкой к корню дерева и из «поверхностной» клады. Введение ошибок в данные увеличивало время сходимости филогенетического вывода до двух раз. Влияние на скорость замен на ветке дерева, ведущей к ошибочной последовательности, могло быть незначительным, однако даже небольшие ошибки в дате выделения (на 5 лет) или последовательности вируса (5 мутаций -1%) могли оказывать влияние на стандартное отклонение скоростей замен во всем филогенетическом дереве, увеличивая его в несколько раз. Также введенные ошибки в аннотацию даты выделения вирусов и их последовательности влияли на общую топологию дерева в разной степени в зависимости от исходного положения вируса на филогенетическом дереве. Таким образом, даже если ошибка в данных не оказывала существенного влияния на скорость накопления замен конкретной ветки филогенетического дерева, увеличение стандартного отклонения параметров может приводить к увеличению времени расчетов. При этом анализ скоростей замен на филогенетическом дереве не всегда выявляет aberrantные геномы, но выявление статистически значимых отклонений (например, скорости накопления замен) является достаточно надежным признаком неестественных событий в эволюции вируса. Тщательная проверка исследуемых вирусов на предмет ошибок в датах выделения образцов и последовательностях могут значительно повысить точность Байесовского филогенетического вывода.

### 4.3 Молекулярная эволюция типов энтеровирусов

Повсеместная рекомбинация между структурными и неструктурными областями генома обеспечивает практически независимую эволюционную траекторию этих частей генома в глобальной популяции вируса. Это позволяет (и требует, в некоторых случаях) анализировать филогенетику СО и НСО этих вирусов по-отдельности. Одной из наиболее актуальных и изученных групп (+)РНК-вирусов как источника возникающих заболеваний являются энтеровирусы человека. В задачи этой работы входило проведение систематического анализа филогенетики типов (СО генома) энтеровирусов для того, чтобы понять закономерности их возникновения.

Расчетный возраст общих предков наиболее распространенных типов энтеровирусов, выведенный в данной работе с помощью Байесовского филогенетического анализа, оказался сравнительно недавним – в пределах 55–200 лет назад. Однако пикорнавирусы или их предки почти наверняка существовали уже на заре появления жизни на Земле [345]. Для того чтобы разрешить это противоречие, в первую очередь была выполнена проверка влияния параметров выборки на скорость накопления замен и высоту корня – ключевые параметры, выводимые с помощью Байесовского филогенетического анализа. Число уникальных последовательностей для каждого типа составило от 38 до 629. Как правило, наиболее старый изолят датировался примерно 1950 годом, тогда как другие вирусы были выделены гораздо позже. Объем и глубина выборки (возраст самого старого изолята) не оказывали влияния на оценки скоростей накопления замен, но влияли на высоту корня (возраст последнего общего предка типа). Объем выборки более чем из 100 уникальных последовательностей оказался достаточным для того, чтобы исключить эффект влияния размера выборки на высоту корня. Уменьшение глубины выборки за счет удаления прототипного штамма из выравнивания типа приводило к соответствующему уменьшению высоты корня. Таким образом, возраст самой старой известной последовательности имел самое большое влияние на выведенный возраст корня. Отсутствие связи между глубиной выборки и скоростью накопления замен подтверждает отсутствие значимого эффекта насыщения

мутаций на оценки Байесовского филогенетического анализа. Размер и глубина выборки по-разному влияли на выведенные возраст корня и скорости накопления замен, поэтому можно утверждать, что уменьшение возраста корня при удалении самого старого изолята обусловлено объективными биологическими закономерностями, а не с особенностями реализации алгоритма или репрезентативностью выборки исходных данных.

Ранее было показано, что современные вирусы, принадлежащие одному типу, отличаются небольшим разнообразием нуклеотидной последовательности VP1 относительно старых изолятов и больше похожи друг на друга, чем на вирусы своего типа, выделенные в 1950–1980 годах [5,183,346]. Предполагалось, что популяция вирусов могла проходить через глобальные «бутылочные горлышки», которые вели к вымиранию большинства линий и уменьшали разнообразие вирусов [346]. Также было показано, что глобальное разнообразие доминирующих генотипов EV-A71 в каждый момент времени ограничено [5]. Анализ топологии деревьев, выведенных в данной работе, показал, что прототипные штаммы занимали прикорневое положение в 16 из 32 исследуемых типов, главным образом, в вирусах echo (11 из 16 типов). В деревьях для видов *Enterovirus A* и *Enterovirus B* большинство современных штаммов внутри различных типов формировали отдельные кластеры с совсем недавними общими предками. Такая топология филогенетических деревьев может быть связана с тем, что глобальные популяции этих типов проходили через «бутылочные горлышки» в последние десятилетия.

С точки зрения эпидемиологии, «бутылочные горлышки» совместимы с ежегодными циклами расширения и сокращения популяции вирусов в странах с умеренным климатом [140]. Эти циклы также могут соответствовать наблюдаемому возникновению и последующему исчезновению рекомбинантных форм энтеровирусов в Европе [177]. Другой механизм, который мог бы привести к прохождению популяций через «бутылочные горлышки» - «эпидемические» многолетние циклы многих типов энтеровирусов [182]. Увеличение размера популяции вирусов может происходить с периодом в несколько лет и объясняется колебаниями популяционного иммунитета.

Результаты филогенетического анализа, проведенного в данной работе, подтверждают гипотезу о том, что через глобальные «бутылочные горлышки» проходило большинство типов, и они были одной из главных движущих сил эволюции энтеровирусов. Остается открытым вопрос: как глобальные популяции вирусов отдельных типов могли регулярно претерпевать такие серьезные сокращения численности популяции, но не вымирать? Одним из объяснений может быть то, что приматы могут служить резервуарами для энтеровирусов и обеспечивать долгосрочное сохранение вариантов вирусов. С другой стороны, возможно, что такие циклы расширения популяции и вымирания являются характерной чертой эпидемиологии и эволюции энтеровирусов только в регионах с сезонным климатом.

Именно в регионах с сезонным климатом была выделена большая часть энтеровирусов (Северная Атлантика, Китай), представленных в Genbank, и, возможно, результаты анализа оказались бы иными, если бы в выборках последовательностей были представлены регионы с другим типом климата и эпидемиологией. Так, например, в данном исследовании среди индийских изолятов наблюдались последовательности, которые характеризовались наибольшим разнообразием и сильно отличались от других последовательностей своего типа [347]. 116 из 117 современных изолятов CVA4 отличались от прототипного штамма High Point, выделенного в 1948 году, тогда как единственный изолят из Кении, датированный 1999 годом, группировался на дереве с прототипным штаммом (Рис. 46). При исследовании энтеровирусов Мадагаскара был открыт новый ранее неизвестный генотип F EV-A71 [36]. Таким образом, представленность изолятов из разных географических зон имеет большое значение для исследования молекулярной эволюции энтеровирусов и отличий в закономерностях эволюции между странами и континентами.

Гипотетические «бутылочные горлышки», через которые могла проходить глобальная популяция энтеровирусов, мешают точному определению времени возникновения типов энтеровирусов. Несмотря на то, что выведенные времена последних общих предков известных изолятов могут быть до-

статочны надежными, они относятся только к известным на данный момент вирусам и могут не отражать время возникновения типов. К использованию коалесцентного Байесовского анализа для нескольких типов сразу нужно относиться осторожно, потому что отдельные типы необязательно представляют собой единую популяцию. Кроме того, нельзя исключать того, что при образовании нового типа возможны квантовые эволюционные события (множественные почти мгновенные несинонимичные мутации) [187].

Для того чтобы определить верхние границы времен возникновения типов, в данной работе был проведен Байесовский филогенетический анализ видов *Enterovirus A* и *Enterovirus B*. Для деревьев видов были характерны длинные ветви, ведущие к отдельным типам. Такая топология совместима с гипотезой о «бутылочных горлышках» и свидетельствует о том, что вероятность разделения одного типа на отдельные группы (новые типы) относительно мала. По-видимому, типы энтеровирусов соответствуют пикам приспособленности, и популяции «блуждают» рядом с этими пиками. Это согласуется с предположением о том, что появление нового типа происходит скорее за счет полной перестройки капсида, а не некоторого числа плавных изменений [184].

Относительно недавние времена узлов деревьев, которые соединяют несколько типов, подтверждают гипотезу о том, что современные типы могли возникнуть в течение последних нескольких сотен лет. Однако даты, полученные при построении деревьев для видов, нужно интерпретировать с осторожностью, так как эффект насыщения синонимичными мутациями может быть достаточно значительным для вирусов, дивергировавших более 100 лет назад. Датирование событий, которые произошли намного раньше, чем был выделен самый «старый» вирус, может быть неправильным и ошибочным на порядки [16].

Филогенетические отношения последовательностей внутри некоторых типов свидетельствуют о том, что процесс образования типов может происходить на наших глазах. Так, например, популяция E30 по всей видимости разделилась на две части (генотип A и другие генотипы), нуклеотидные ди-



станции между которыми превзошли порог типа, установленный ICTV, после дивергенции [348]. Филогенетические отношения внутри некоторых других типов (например, CBV5) тоже согласовывались с разделением популяции на несколько линий, которые достаточно приспособлены для циркуляции. Эти примеры показывают, что типы (по крайней мере, по определению ICTV) могут возникать не только при генетическом шифте, которому может способствовать изоляция популяции вирусов из-за смены клеточного рецептора или хозяина, но и при генетическом дрейфе.

## ВЫВОДЫ

1. Общим свойством эволюции вирусов с несегментированным РНК-геномом положительной полярности относительно мРНК является высокая частота рекомбинации между участками, кодирующими структурные и неструктурные белки. Профили рекомбинации внутри участков генома, кодирующих структурные и неструктурные белки, различаются между семействами, но сходны внутри одного семейства РНК-вирусов (*Caliciviridae*, *Astroviridae*, *Coronaviridae*).
2. Для калицивирусов характерен обмен целыми блоками неструктурных и структурных генов и крайне редкая рекомбинация внутри участков, кодирующих структурные и неструктурные белки. Для астровирусов характерен обмен целыми блоками неструктурных и структурных генов и умеренная рекомбинация внутри структурных и неструктурных генов.
3. У коронавирусов наиболее часто рекомбинация происходит между геном структурного белка S и другими участками генома. Рекомбинация внутри гена белка S происходит чаще, чем в неструктурной области, причем с большей частотой происходит обмен целыми доменами белка S. Гены остальных структурных белков (N, E, M) по профилю естественной рекомбинации соответствуют неструктурным генам.
4. Участки генома, кодирующие капсидные белки энтеровирусов, благодаря рекомбинации имеют эволюционную историю, практически оторванную от генов неструктурных белков, и существуют в биосфере как высоко динамичные пулы, в пределах которых отдельные типы могут регулярно подвергаться почти полному исчезновению практически в глобальном масштабе даже в течение десятков лет.
5. Филогенетический анализ вирусов CVA2, выделенных от больных с острым вялым параличом, не выявил их связи с конкретной филогенетической группой, что опровергает гипотезу о возникновении варианта вируса с измененными патогенетическими характеристиками.

## СПИСОК ЛИТЕРАТУРЫ

1. Simmonds P., Welch J. Frequency and Dynamics of Recombination within Different Species of Human Enteroviruses Frequency and Dynamics of Recombination within Different Species of Human Enteroviruses. 2006. Vol. 80, № 1. P. 483–493.
2. Nascimento F.F., Reis M. dos, Yang Z. A biologist's guide to Bayesian phylogenetic analysis // *Nat. Ecol. Evol.* 2017. Vol. 1, № 10. P. 1446–1454.
3. Russel P.M. et al. Model Selection and Parameter Inference in Phylogenetics Using Nested Sampling // *Syst. Biol.* / ed. Stadler T. 2019. Vol. 68, № 2. P. 219–233.
4. Baele G., Lemey P., Vansteelandt S. Make the most of your samples: Bayes factor estimators for high-dimensional models of sequence evolution // *BMC Bioinformatics.* 2013. Vol. 14, № 1. P. 85.
5. Tee K.K. et al. Evolutionary genetics of human enterovirus 71: origin, population dynamics, natural selection, and seasonal periodicity of the VP1 gene. // *J. Virol.* 2010. Vol. 84, № 7. P. 3339–3350.
6. Savolainen-Kopra C. et al. A large Finnish echovirus 30 outbreak was preceded by silent circulation of the same genotype // *Virus Genes.* 2011. Vol. 42, № 1. P. 28–36.
7. Wright S. Classification of the Factors of Evolution // *Cold Spring Harb. Symp. Quant. Biol.* 1955. Vol. 20. P. 16–24.
8. Drake J.W. Rates of spontaneous mutation among RNA viruses. // *Proc. Natl. Acad. Sci.* 1993. Vol. 90, № 9. P. 4171–4175.
9. Choi K.H. Viral Polymerases. 2012. P. 267–304.
10. Sanjuán R. et al. Viral Mutation Rates // *J. Virol.* 2010. Vol. 84, № 19. P. 9733–9748.
11. Amicone M. et al. Mutation rate of SARS-CoV-2 and emergence of mutators during experimental evolution // *Evol. Med. Public Heal.* 2022. Vol. 10, № 1. P. 142–155.
12. Ogando N.S. et al. The Curious Case of the Nidovirus Exoribonuclease: Its Role in RNA Synthesis and Replication Fidelity // *Front. Microbiol.* 2019. Vol. 10.
13. Eckerle L.D. et al. High Fidelity of Murine Hepatitis Virus Replication Is Decreased in nsp14 Exoribonuclease Mutants // *J. Virol.* 2007. Vol. 81, № 22. P. 12135–12144.
14. Jenkins G.M. et al. Rates of Molecular Evolution in RNA Viruses: A Quantitative Phylogenetic Analysis // *J. Mol. Evol.* 2002. Vol. 54, № 2. P. 156–165.
15. Duffy S., Shackelton L.A., Holmes E.C. Rates of evolutionary change in viruses: Patterns and determinants // *Nat. Rev. Genet.* 2008. Vol. 9, № 4. P. 267–276.
16. Duchêne S., Holmes E.C., Ho S.Y.W. Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. // *Proc. Biol.*

- Sci. 2014. Vol. 281, № 1786. P. 20140732-.
17. Eigen M., Schuster P. The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle // *Naturwissenschaften*. 1977. Vol. 64, № 11. P. 541–565.
  18. Domingo E. et al. Nucleotide sequence heterogeneity of an RNA phage population // *Cell*. 1978. Vol. 13, № 4. P. 735–744.
  19. Domingo E., Perales C. Viral quasispecies // *PLOS Genet*. 2019. Vol. 15, № 10. P. e1008271.
  20. Wilke C.O. Quasispecies theory in the context of population genetics // *BMC Evol. Biol*. 2005. Vol. 5, № 1. P. 44.
  21. Moya A., Holmes E.C., González-Candelas F. The population genetics and evolutionary epidemiology of RNA viruses // *Nat. Rev. Microbiol*. 2004. Vol. 2, № 4. P. 279–288.
  22. Bordería A. V. et al. Group Selection and Contribution of Minority Variants during Virus Adaptation Determines Virus Fitness and Phenotype // *PLOS Pathog.* / ed. Coyne C.B. 2015. Vol. 11, № 5. P. e1004838.
  23. Codoñer F.M. et al. The Fittest versus the Flattest: Experimental Confirmation of the Quasispecies Effect with Subviral Pathogens // *PLoS Pathog.* 2006. Vol. 2, № 12. P. e136.
  24. Sanjuán R. et al. Selection for Robustness in Mutagenized RNA Viruses // *PLoS Genet.* / ed. Singh Malik H. 2007. Vol. 3, № 6. P. e93.
  25. Burch C.L., Chao L. Evolvability of an RNA virus is determined by its mutational neighbourhood // *Nature*. 2000. Vol. 406, № 6796. P. 625–628.
  26. Geoghegan J.L., Holmes E.C. Evolutionary Virology at 40 // *Genetics*. 2018. Vol. 210, № 4. P. 1151–1162.
  27. Acevedo A., Brodsky L., Andino R. Mutational and fitness landscapes of an RNA virus revealed through population sequencing // *Nature*. 2014. Vol. 505, № 7485. P. 686–690.
  28. Sanjuán R., Moya A., Elena S.F. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus // *Proc. Natl. Acad. Sci*. 2004. Vol. 101, № 22. P. 8396–8401.
  29. Vale P.F. et al. The Distribution of Mutational Fitness Effects of Phage  $\phi$ X174 on Different Hosts // *Evolution (N. Y)*. 2012. Vol. 66, № 11. P. 3495–3507.
  30. Visher E. et al. The Mutational Robustness of Influenza A Virus // *PLOS Pathog.* / ed. Ferguson N.M. 2016. Vol. 12, № 8. P. e1005856.
  31. Charlesworth B. Effective population size and patterns of molecular evolution and variation // *Nat. Rev. Genet*. 2009. Vol. 10, № 3. P. 195–205.
  32. Sender R. et al. The total number and mass of SARS-CoV-2 virions // *Proc. Natl. Acad. Sci*. 2021. Vol. 118, № 25.
  33. Jin T., Yin J. Patterns of virus growth across the diversity of life // *Integr. Biol*. 2021. Vol. 13, № 2. P. 44–59.
  34. Gutiérrez S., Michalakis Y., Blanc S. Virus population bottlenecks during within-host progression and host-to-host transmission // *Curr. Opin. Virol*.

2012. Vol. 2, № 5. P. 546–555.
35. Bentley K., Evans D.J. Mechanisms and consequences of positive-strand RNA virus recombination // *J. Gen. Virol.* 2018. Vol. 99, № 10. P. 1345–1356.
  36. Han G.-Z., Worobey M. Homologous Recombination in Negative Sense RNA Viruses // *Viruses.* 2011. Vol. 3, № 8. P. 1358–1373.
  37. Patiño-Galindo J.Á., Filip I., Rabadan R. Global Patterns of Recombination across Human Viruses // *Mol. Biol. Evol.* / ed. Crandall K. 2021. Vol. 38, № 6. P. 2520–2531.
  38. Lai M.M. RNA recombination in animal and plant viruses // *Microbiol. Rev.* 1992. Vol. 56, № 1. P. 61–79.
  39. Cooper P.D. et al. On the Nature of Poliovirus Genetic Recombinants // *J. Gen. Virol.* 1974. Vol. 23, № 1. P. 41–49.
  40. Nagy P.D., Bujarski J.J. Efficient system of homologous RNA recombination in brome mosaic virus: sequence and structure requirements and accuracy of crossovers // *J. Virol.* 1995. Vol. 69, № 1. P. 131–140.
  41. Pilipenko E.V., Gmyl A.P., Agol V.I. A model for rearrangements in RNA genomes // *Nucleic Acids Res.* 1995. Vol. 23, № 1870–1875.
  42. Chetverin A.B. et al. Nonhomologous RNA Recombination in a Cell-Free System: Evidence for a Transesterification Mechanism Guided by Secondary Structure // *Cell.* 1997. Vol. 88, № 4. P. 503–513.
  43. Gmyl A.P. et al. Nonreplicative homologous RNA recombination: Promiscuous joining of RNA pieces? // *RNA.* 2003. Vol. 9. P. 1221–1231.
  44. Scheel T.K.H. et al. Productive Homologous and Non-homologous Recombination of Hepatitis C Virus in Cell Culture // *PLoS Pathog.* / ed. Lindenbach B.D. 2013. Vol. 9, № 3. P. e1003228.
  45. Austermann-Busch S., Becher P. RNA Structural Elements Determine Frequency and Sites of Nonhomologous Recombination in an Animal Plus-Strand RNA Virus // *J. Virol.* 2012. Vol. 86, № 13. P. 7393–7402.
  46. Kleine Büning M. et al. Nonreplicative RNA Recombination of an Animal Plus-Strand RNA Virus in the Absence of Efficient Translation of Viral Proteins // *Genome Biol. Evol.* 2017. Vol. 9, № 4. P. 817–829.
  47. Turner P.E., Elena S.F. Cost of Host Radiation in an RNA Virus // *Genetics.* 2000. Vol. 156, № 4. P. 1465–1470.
  48. Cooper L.A., Scott T.W. Differential Evolution of Eastern Equine Encephalitis Virus Populations in Response to Host Cell Type // *Genetics.* 2001. Vol. 157, № 4. P. 1403–1412.
  49. Novella I.S. et al. Lack of evolutionary stasis during alternating replication of an arbovirus in insect and mammalian cells // *J. Mol. Biol.* 1999. Vol. 287, № 3. P. 459–465.
  50. Miralles R., Moya A., Elena S.F. Effect of population patchiness and migration rates on the adaptation and divergence of vesicular stomatitis virus quasispecies populations // *J. Gen. Virol.* 1999. Vol. 80, № 8. P. 2051–2059.
  51. Cuevas J.M., Moya A., Elena S.F. Evolution of RNA virus in spatially

- structured heterogeneous environments // *J. Evol. Biol.* 2003. Vol. 16, № 3. P. 456–466.
52. Lemey P. et al. Bayesian Phylogeography Finds Its Roots // *PLoS Comput. Biol.* / ed. Fraser C. 2009. Vol. 5, № 9. P. e1000520.
53. Bouckaert R. Phylogeography by diffusion on a sphere: whole world phylogeography // *PeerJ*. 2016. Vol. 4. P. e2406.
54. Lemey P. et al. Phylogeography takes a relaxed random walk in continuous space and time // *Mol. Biol. Evol.* 2010. Vol. 27, № 8. P. 1877–1885.
55. Hassel C. et al. Phylogeography of Coxsackievirus A16 Reveals Global Transmission Pathways and Recent Emergence and Spread of a Recombinant Genogroup // *J. Virol.* 2017. Vol. 91, № 18. P. e00630-17.
56. Omondi G. et al. Phylogeographical and cross-species transmission dynamics of SAT1 and SAT2 foot-and-mouth disease virus in Eastern Africa. // *Mol. Ecol.* 2019. Vol. 28, № 11. P. 2903–2916.
57. Wille M. et al. Virus–virus interactions and host ecology are associated with RNA virome structure in wild birds // *Mol. Ecol.* 2018. Vol. 27, № 24. P. 5263–5278.
58. Wang F. et al. Genetic characterization of a novel calicivirus from a goose // *Arch. Virol.* 2017. Vol. 162, № 7. P. 2115–2118.
59. Shi M. et al. The evolutionary history of vertebrate RNA viruses // *Nature*. Springer US, 2018. Vol. 556, № 7700. P. 197–202.
60. Adler J.L., Zickl R. Winter Vomiting Disease // *J. Infect. Dis.* 1969. Vol. 119, № 6. P. 668–673.
61. Vinjé J. et al. ICTV Virus Taxonomy Profile: Caliciviridae // *J. Gen. Virol.* 2019. Vol. 100, № 11. P. 1469–1470.
62. Kroneman A. et al. Proposal for a unified norovirus nomenclature and genotyping // *Arch. Virol.* 2013. Vol. 158, № 10. P. 2059–2068.
63. Chhabra P. et al. Updated classification of norovirus genogroups and genotypes // *J. Gen. Virol.* 2019. Vol. 100, № 10. P. 1393–1406.
64. van Beek J. et al. Molecular surveillance of norovirus, 2005–16: an epidemiological analysis of data collected from the NoroNet network // *Lancet Infect. Dis.* 2018. Vol. 18, № 5. P. 545–553.
65. Ludwig-Begall L.F., Mauroy A., Thiry E. Noroviruses—The State of the Art, Nearly Fifty Years after Their Initial Discovery // *Viruses*. 2021. Vol. 13, № 8. P. 1541.
66. Madeley C.R., Cosgrove B.P. CALICIVIRUSES IN MAN // *Lancet*. 1976. Vol. 307, № 7952. P. 199–200.
67. Numata K. et al. Molecular characterization of morphologically typical human calicivirus Sapporo // *Arch. Virol.* 1997. Vol. 142, № 8. P. 1537–1552.
68. Oka T. et al. Genetic Characterization and Classification of Human and Animal Sapoviruses // *PLoS One* / ed. Jin D.-Y. 2016. Vol. 11, № 5. P. e0156373.
69. Oka T. et al. Comprehensive review of human sapoviruses // *Clin. Microbiol.*

- Rev. 2015. Vol. 28, № 1. P. 32–53.
70. Yinda C.K. et al. Novel highly divergent sapoviruses detected by metagenomics analysis in straw-colored fruit bats in Cameroon // *Emerg. Microbes Infect.* 2017. Vol. 6, № 1. P. 1–7.
  71. GAVIER-WIDEN D., MORNER T. Epidemiology and diagnosis of the european brown hare syndrome in Scandinavian countries : a review // *Rev. Sci. Tech. l'OIE.* 1991. Vol. 10, № 2. P. 453–458.
  72. Liu S.J. et al. A new viral disease in rabbits // *Anim. Husb. Vet. Med. (Xumu yu Shouyi).* 1984. Vol. 16, № 6. P. 253–255.
  73. Capucci L. et al. Detection and preliminary characterization of a new rabbit calicivirus related to rabbit hemorrhagic disease virus but nonpathogenic // *J. Virol.* 1996. Vol. 70, № 12. P. 8614–8623.
  74. Abrantes J. et al. Rabbit haemorrhagic disease (RHD) and rabbit haemorrhagic disease virus (RHDV): a review // *Vet. Res.* 2012. Vol. 43, № 1. P. 12.
  75. Le Pendu J. et al. Proposal for a unified classification system and nomenclature of lagoviruses // *J. Gen. Virol.* 2017. Vol. 98, № 7. P. 1658–1666.
  76. Green K. *Caliciviridae: the noroviruses* // *Fields virology*, 6th ed. 2013. P. 582–608.
  77. McFadden N. et al. Norovirus Regulation of the Innate Immune Response and Apoptosis Occurs via the Product of the Alternative Open Reading Frame 4 // *PLoS Pathog.* / ed. Basler C.F. 2011. Vol. 7, № 12. P. e1002413.
  78. Desselberger U. *Caliciviridae Other Than Noroviruses* // *Viruses.* 2019. Vol. 11, № 3. P. 286.
  79. Goodfellow I., Taube S. *Calicivirus Replication and Reverse Genetics* // *Viral Gastroenteritis.* Elsevier, 2016. P. 355–378.
  80. Fuentes C. et al. Identification of Human Astrovirus Genome-Linked Protein (VPg) Essential for Virus Infectivity // *J. Virol.* 2012. Vol. 86, № 18. P. 10070–10078.
  81. Smertina E. et al. Calicivirus Non-structural Proteins: Potential Functions in Replication and Host Cell Manipulation // *Front. Microbiol.* 2021. Vol. 12.
  82. McCormick C.J. et al. Translation Termination Reinitiation between Open Reading Frame 1 (ORF1) and ORF2 Enables Capsid Expression in a Bovine Norovirus without the Need for Production of Viral Subgenomic RNA // *J. Virol.* 2008. Vol. 82, № 17. P. 8917–8921.
  83. Hardy M.E. et al. Human calicivirus genogroup II capsid sequence diversity revealed by analyses of the prototype Snow Mountain agent // *Arch. Virol.* 1997. Vol. 142, № 7. P. 1469–1479.
  84. Jiang X. et al. Characterization of a novel human calicivirus that may be a naturally occurring recombinant // *Arch. Virol.* 1999. Vol. 144, № 12. P. 2377–2387.
  85. Mathijs E. et al. Experimental evidence of recombination in murine noroviruses // *J. Gen. Virol.* 2010. Vol. 91, № 11. P. 2723–2733.

86. Zhang H. et al. Isolation and Analysis of Rare Norovirus Recombinants from Coinfected Mice Using Drop-Based Microfluidics // *J. Virol.* 2015. Vol. 89, № 15. P. 7722–7734.
87. Begall L.F.L., Mauroy A., Thiry E. Norovirus recombinants: Recurrent in the field, recalcitrant in the lab - a scoping review of recombination and recombinant types of noroviruses // *J. Gen. Virol.* 2018. Vol. 99, № 8. P. 970–988.
88. Nayak M.K. et al. Detection of a novel intergenogroup recombinant Norovirus from Kolkata, India // *Virology.* 2008. Vol. 377, № 1. P. 117–123.
89. Phan T.G. et al. Genetic heterogeneity, evolution, and recombination in noroviruses // *J. Med. Virol.* 2007. Vol. 79, № 9. P. 1388–1400.
90. Takano T. et al. Molecular characterization and pathogenicity of a genogroup GVI feline norovirus // *Vet. Microbiol.* 2015. Vol. 178, № 3–4. P. 201–207.
91. Di Martino B. et al. A novel feline norovirus in diarrheic cats // *Infect. Genet. Evol.* 2016. Vol. 38. P. 132–137.
92. Rohayem J., Münch J., Rethwilm A. Evidence of Recombination in the Norovirus Capsid Gene // *J. Virol.* 2005. Vol. 79, № 8. P. 4977–4990.
93. Lam T.T.Y. et al. The recombinant origin of emerging human norovirus GII.4/2008: Intra-genotypic exchange of the capsid P2 domain // *J. Gen. Virol.* 2012. Vol. 93, № 4. P. 817–822.
94. Waters A., Coughlan S., Hall W.W. Characterisation of a novel recombination event in the norovirus polymerase gene // *Virology.* 2007. Vol. 363, № 1. P. 11–14.
95. Laconi A. et al. Identification of two divergent swine Noroviruses detected at the slaughterhouse in North East Italy // *Porc. Heal. Manag.* 2020. Vol. 6, № 1. P. 9.
96. Eden J.-S. et al. Recombination within the Pandemic Norovirus GII.4 Lineage // *J. Virol.* 2013. Vol. 87, № 11. P. 6270–6282.
97. Ford-Siltz L.A. et al. Genomics analyses of GIV and GVI noroviruses reveal the distinct clustering of human and animal viruses // *Viruses.* 2019. Vol. 11, № 3.
98. Chhabra P., Walimbe A.M., Chitambar S.D. Complete genome characterization of Genogroup II norovirus strains from India: Evidence of recombination in ORF2/3 overlap // *Infect. Genet. Evol. Elsevier B.V.*, 2010. Vol. 10, № 7. P. 1101–1109.
99. Thackray L.B. et al. Murine Noroviruses Comprising a Single Genogroup Exhibit Biological Diversity despite Limited Sequence Divergence // *J. Virol.* 2007. Vol. 81, № 19. P. 10460–10473.
100. Müller B. et al. Genetic diversity and recombination of murine noroviruses in immunocompromised mice // *Archives of Virology.* 2007. Vol. 152, № 9. P. 1709–1719.
101. Katayama K. et al. Novel Recombinant Sapovirus // *Emerg. Infect. Dis.* 2004. Vol. 10, № 10. P. 1874–1876.
102. Phan T.G. et al. Emergence of intragenotype recombinant sapovirus in Japan



- // Infect. Genet. Evol. 2007. Vol. 7, № 4. P. 542–546.
103. dos Anjos K. et al. The possible molecular evolution of sapoviruses by inter- and intra-genogroup recombination // Arch. Virol. 2011. Vol. 156, № 11. P. 1953–1959.
  104. Phan T.G. et al. Novel intragenotype recombination in sapovirus // Clin Lab. 2006. Vol. 52, № 7–8. P. 363–366.
  105. Nguyen T.A. et al. Norovirus and Sapovirus Infections among Children with Acute Gastroenteritis in Ho Chi Minh City during 2005-2006 // J. Trop. Pediatr. 2007. Vol. 54, № 2. P. 102–113.
  106. Liu X. et al. Molecular detection and characterization of sapovirus in hospitalized children with acute gastroenteritis in the Philippines // J. Clin. Virol. 2015. Vol. 68. P. 83–88.
  107. Lasure N., Gopalkrishna V. Epidemiological profile and genetic diversity of sapoviruses (SaVs) identified in children suffering from acute gastroenteritis in Pune, Maharashtra, Western India, 2007-2011 // Epidemiol. Infect. 2017. Vol. 145, № 1. P. 106–114.
  108. Kumthip K. et al. Genetic recombination and diversity of sapovirus in pediatric patients with acute gastroenteritis in Thailand, 2010-2018 // PeerJ. 2020. Vol. 2020, № 2. P. 2010–2018.
  109. Li J. et al. Genomic organization and recombination analysis of a porcine sapovirus identified from a piglet with diarrhea in China // Virol. J. 2017. Vol. 14, № 1. P. 57.
  110. Shan T. et al. The Fecal Virome of Pigs on a High-Density Farm // J. Virol. 2011. Vol. 85, № 22. P. 11697–11708.
  111. Jeong C. et al. Genetic diversity of porcine sapoviruses // Vet. Microbiol. 2007. Vol. 122, № 3–4. P. 246–257.
  112. Hansman G.S. et al. Intergenogroup recombination in sapoviruses // Emerg. Infect. Dis. 2005. Vol. 11, № 12. P. 1916–1920.
  113. Kuroda M. et al. Genetic diversity and intergenogroup recombination events of sapoviruses detected from feces of pigs in Japan // Infect. Genet. Evol. Elsevier, 2017. Vol. 55, № September. P. 209–217.
  114. Reuter G., Bíró H., Szűcs G. Enteric caliciviruses in domestic pigs in Hungary // Arch. Virol. 2007. Vol. 152, № 3. P. 611–614.
  115. Iizuka S. et al. Detection of sapoviruses and noroviruses in an outbreak of gastroenteritis linked genetically to shellfish // J. Med. Virol. 2010. Vol. 82, № 7. P. 1247–1254.
  116. Nakagawa-Okamoto R. et al. Detection of multiple sapovirus genotypes and genogroups in oyster-associated outbreaks. // Jpn. J. Infect. Dis. 2009. Vol. 62, № 1. P. 63–66.
  117. Abrantes J. et al. Recombination at the emergence of the pathogenic rabbit haemorrhagic disease virus Lagovirus europaeus/GI.2 // Sci. Rep. Nature Publishing Group UK, 2020. Vol. 10, № 1. P. 1–11.
  118. Szillat K.P. et al. Full-genome sequencing of German rabbit haemorrhagic disease virus uncovers recombination between RHDV (GI.2) and EBHSV

- (GII.1) // *Virus Evol.* 2020. Vol. 6, № 2. P. 1–11.
119. Abrantes J., Esteves P.J., Van Der Loo W. Evidence for recombination in the major capsid gene VP60 of the rabbit haemorrhagic disease virus (RHDV) // *Arch. Virol.* 2008. Vol. 153, № 2. P. 329–335.
  120. Kovaliski J. et al. Molecular epidemiology of rabbit haemorrhagic disease virus in australia: When one became many // *Mol. Ecol.* 2014. Vol. 23, № 2. P. 408–420.
  121. Forrester N.L. et al. Recombination in rabbit haemorrhagic disease virus: Possible impact on evolution and epidemiology // *Virology.* 2008. Vol. 376, № 2. P. 390–396.
  122. Hu B. et al. Recombination between G2 and G6 strains of rabbit hemorrhagic disease virus (RHDV) in China // *Arch. Virol.* 2017. Vol. 162, № 1. P. 269–272.
  123. Silvério D. et al. Insights into the evolution of the new variant rabbit haemorrhagic disease virus (GI.2) and the identification of novel recombinant strains // *Transbound. Emerg. Dis.* 2018. Vol. 65, № 4. P. 983–992.
  124. Lopes A.M. et al. GI.1b/GI.1b/GI.2 recombinant rabbit hemorrhagic disease virus 2 (*Lagovirus europaeus*/GI.2) in Morocco, Africa // *Arch. Virol.* Springer Vienna, 2019. Vol. 164, № 1. P. 279–283.
  125. Hall R.N. et al. A strain-specific multiplex RT-PCR for Australian rabbit haemorrhagic disease viruses uncovers a new recombinant virus variant in rabbits and hares // *Transbound. Emerg. Dis.* 2018. Vol. 65, № 2. P. e444–e456.
  126. Mahar J.E. et al. crossm Endemic Strains of RHDV in the Australian Landscape within 18 Months of Its Arrival. 2018. Vol. 2, № March. P. 1–15.
  127. Lopes A.M. et al. Full genomic analysis of new variant rabbit hemorrhagic disease virus revealed multiple recombination events // *J. Gen. Virol.* 2015. Vol. 96. P. 1309–1319.
  128. Mahar J.E. et al. Benign Rabbit Caliciviruses Exhibit Evolutionary Dynamics Similar to Those of Their Virulent Relatives // *J. Virol.* 2016. Vol. 90, № 20. P. 9317–9329.
  129. Mahar J.E. et al. The discovery of three new hare lagoviruses reveals unexplored viral diversity in this genus // *Virus Evol.* 2019. Vol. 5, № 1. P. 1–11.
  130. Zell R. et al. ICTV Virus Taxonomy Profile: Picornaviridae // *J. Gen. Virol.* 2017. Vol. 98, № 10. P. 2421–2422.
  131. Hyypia T. et al. Classification of Enteroviruses Based on Molecular and Biological Properties. 1997. Vol. 78. P. 1–11.
  132. Virus Taxonomy: 2021 Release [Electronic resource]. URL: <https://talk.ictvonline.org/taxonomy/>.
  133. Simmonds P. et al. Recommendations for the nomenclature of enteroviruses and rhinoviruses // *Arch. Virol.* 2020. Vol. 165, № 3. P. 793–797.
  134. Racaniello V.R. Picornaviridae: The viruses and their replication // *Fields*

- Virology. 2013. P. 453–581.
135. ViralZone, ExPASy [Electronic resource]. URL: [http://viralzone.expasy.org/viralzone/all\\_by\\_species/97.html](http://viralzone.expasy.org/viralzone/all_by_species/97.html).
  136. Lee Y. et al. A protein covalently linked to poliovirus genome RNA // Proc Natl Acad Sci U S A. 1977. Vol. 1, № 74. P. 59–63.
  137. Andino R. et al. Poliovirus RNA synthesis utilizes an rnp complex formed around the 5'-end of viral RNA // EMBO J. 1993. № 12. P. 3587–3598.
  138. Agol V.I. The 5' - untranslated region of picornaviral genomes // 5' - untranslated Reg. picornaviral genomes. 1991. Vol. 40. P. 103–180.
  139. Todd S. et al. Replication-competent picornaviruses with complete genomic RNA 3' noncoding region deletions // J. Virol. 1997. Vol. 71. P. 8868–8874.
  140. Pallansch M.A., Roos R.P. Enteroviruses: Polioviruses, coxsackieviruses, echoviruses and newer enteroviruses // Fields Virology, sixth edition. 2013. P. 723–775.
  141. Morosky S. et al. The neonatal Fc receptor is a pan-echovirus receptor // Proc. Natl. Acad. Sci. 2019. Vol. 116, № 9. P. 3758–3763.
  142. Andino R. et al. Intracellular determinants of picornavirus replication // Trends Microbiol. 1999. № 7. P. 76–82.
  143. Steil B.P., Barton D.J. Cis-active RNA elements (CREs) and picornavirus RNA replication // Virus Res. 2009. Vol. 139. P. 240–252.
  144. Lukashev A.N. et al. Recombination in circulating Human enterovirus B: Independent evolution of structural and non-structural genome regions // J. Gen. Virol. 2005. Vol. 86, № 12. P. 3281–3290.
  145. Bessaud M. et al. Genetic relationship between cocirculating human enteroviruses species C // PLoS One. 2011. Vol. 6, № 9.
  146. Lukashev A.N. et al. Recombination strategies and evolutionary dynamics of the Human enterovirus A global gene pool // J. Gen. Virol. 2014. Vol. 95, № Pt\_4. P. 868–873.
  147. McWilliam Leitch E.C. et al. Evolutionary dynamics and temporal/geographical correlates of recombination in the human enterovirus echovirus types 9, 11, and 30. // J. Virol. 2010. Vol. 84, № 18. P. 9292–9300.
  148. Lukashev A.N. Role of recombination in evolution of enteroviruses // Rev Med Virol. 2005. Vol. 15. P. 157–167.
  149. Blomquist S. et al. Characterization of a recombinant type 3/type 2 poliovirus isolated from a healthy vaccinee and containing a chimeric capsid protein VP1 // J. Gen. Virol. 2003. Vol. 84. P. 573–580.
  150. Rotbart H.A. et al. Enterovirus meningitis in adults // Clin. Infect. Dis. 1998. Vol. 27. P. 896–989.
  151. Huang C.C. et al. Neurologic complications in children with enterovirus 71 infection // N. Engl. J. Med. 1999. Vol. 341. P. 936–942.
  152. Andréoletti L. et al. Viral causes of human myocarditis // Arch. Cardiovasc. Dis. 2009. Vol. 102, № 6–7. P. 559–568.
  153. Clements G.B., Galbraith D.N., Taylor K.W. Coxsackie B virus infection and onset of childhood diabetes // Lancet. 1995. Vol. 346. P. 221–223.

154. Li W. et al. Large outbreak of herpangina in children caused by enterovirus in summer of 2015 in Hangzhou, China // *Sci. Rep.* 2016. Vol. 6, № 1. P. 35388.
155. Yang Q. et al. Two Genotypes of Coxsackievirus A2 Associated with Hand, Foot, and Mouth Disease Circulating in China since 2008 // *PLoS One* / ed. Blackard J. 2016. Vol. 11, № 12. P. e0169021.
156. Tapparel C. et al. Picornavirus and enterovirus diversity with associated human diseases // *Infect. Genet. Evol.* 2013. Vol. 14. P. 282–293.
157. Bendig J. et al. Enterovirus sequences resembling coxsackievirus A2 detected in stool and spleen from a girl with fatal myocarditis // *J. Med. Virol.* 2001. Vol. 64. P. 482–486.
158. Molet L. et al. Enterovirus infections in hospitals of Ile de France region over 2013 // *J. Clin. Virol.* 2016. Vol. 74. P. 37–42.
159. Ohara N. et al. Fulminant Type 1 Diabetes Mellitus Associated with Coxsackie Virus Type A2 Infection: A Case Report and Literature Review // *Intern. Med.* 2016. Vol. 55, № 6. P. 643–646.
160. Marx A., Glass J.D., Sutter R.W. Differential Diagnosis of Acute Flaccid Paralysis and its Role in Poliomyelitis Surveillance // *Epidemiol. Rev.* 2000. Vol. 22, № 2. P. 298–316.
161. Chouikha A. et al. Circulation and Molecular Epidemiology of Enteroviruses in Paralyzed, Immunodeficient and Healthy Individuals in Tunisia, a Country with a Polio-Free Status for Decades // *Viruses.* 2021. Vol. 13, № 3. P. 380.
162. Fifteen years of acute flaccid paralysis surveillance in Hong Kong: Findings from 1997 to 2011 // *J. Paediatr. Child Health.* 2014. Vol. 50, № 7. P. 545–552.
163. Kim H. et al. Clinical and enterovirus findings associated with acute flaccid paralysis in the republic of Korea during the recent decade // *J. Med. Virol.* 2014. Vol. 86, № 9. P. 1584–1589.
164. Sousa I.P. et al. Molecular characterization and epidemiological aspects of non-polio enteroviruses isolated from acute flaccid paralysis in Brazil: a historical series (2005–2017) // *Emerg. Microbes Infect.* 2020. Vol. 9, № 1. P. 2536–2546.
165. World Health Organization (WHO) Regional Office for Europe [Electronic resource]. URL: <http://data.euro.who.int/cisid/default.aspx?TabID=539511>.
166. LIM K.A., Benyesh-Melnick M. Typing of viruses by combinations of antiserum pools. Application to typing of enteroviruses (Coxsackie and ECHO) // *J. Immunol.* 1960. Vol. 84, № 0022-1767 (Print). P. 309–317.
167. Oberste M.S. et al. Typing of human enteroviruses by partial sequencing of VP1 // *J Clin Microbiol.* 1999. Vol. 37. P. 1288–1293.
168. Oberste M.S. et al. Molecular evolution of the human enteroviruses: correlation of serotype with VP1 sequence and application to picornavirus classification. // *J. Virol.* 1999. Vol. 73, № 3. P. 1941–1948.
169. Nix W.A., Oberste M.S., Pallansch M.A. Sensitive , Seminested PCR Amplification of VP1 Sequences for Direct Identification of All Enterovirus

- Serotypes from Original Clinical Specimens. 2006. Vol. 44, № 8. P. 2698–2704.
170. Knipe D.M. et al. Virus Evolution // Fields virology. 6th ed. 2013. P. 389–421.
  171. Puenpa J. et al. Molecular epidemiology and the evolution of human coxsackievirus A6 // J. Gen. Virol. 2016. Vol. 97, № 12. P. 3225–3231.
  172. Zhao G. et al. Characterization of VP1 sequence of Coxsackievirus A16 isolates by Bayesian evolutionary method // Virol. J. Virology Journal, 2016. Vol. 13, № 1. P. 1–11.
  173. Hicks A.L., Duffy S. Genus-Specific Substitution Rate Variability among Picornaviruses // J. Virol. 2011. Vol. 85, № 15. P. 7942–7947.
  174. Huang H.W. et al. Phylodynamic reconstruction of the spatiotemporal transmission and demographic history of coxsackievirus B2 // BMC Bioinformatics. BMC Bioinformatics, 2015. Vol. 16, № 1. P. 1–9.
  175. Yarmolskaya M.S. et al. Molecular epidemiology of echoviruses 11 and 30 in Russia: Different properties of genotypes within an enterovirus serotype // Infect. Genet. Evol. Elsevier B.V., 2015. Vol. 30. P. 244–248.
  176. McWilliam Leitch E.C. et al. The association of recombination events in the founding and emergence of subgenogroup evolutionary lineages of human enterovirus 71. // J. Virol. 2012. Vol. 86, № 5. P. 2676–2685.
  177. McWilliam Leitch E.C. et al. Transmission Networks and Population Turnover of Echovirus 30 // J. Virol. 2009. Vol. 83, № 5. P. 2109–2118.
  178. Cabrerizo M., Trallero G., Simmonds P. Recombination and Evolutionary Dynamics of Human Echovirus 6 // J. Med. Virol. 2014.
  179. Linsuwanon P. et al. Molecular epidemiology and evolution of human enterovirus serotype 68 in Thailand, 2006-2011 // PLoS One. 2012. Vol. 7, № 5. P. 2006–2011.
  180. Tan X. et al. The persistent circulation of enterovirus 71 in people's republic of China: Causing emerging nationwide epidemics since 2008 // PLoS One. 2011. Vol. 6, № 9.
  181. Bessaud M. et al. Molecular comparison and evolutionary analyses of VP1 nucleotide sequences of new African human enterovirus 71 isolates reveal a wide genetic diversity // PLoS One. 2014. Vol. 9, № 3. P. 1–9.
  182. Khetsuriani N. et al. Enterovirus surveillance-United States, 1970-2005 // MMWR Surveill Summ. 2006. Vol. 55. P. 1–20.
  183. Lukashev A.N. et al. Analysis of echovirus 30 isolates from Russia and new independent states revealing frequent recombination and reemergence of ancient lineages // J Clin Microbiol. 2008. Vol. 46, № 2. P. 665–670.
  184. Simmonds P. Recombination and selection in the evolution of picornaviruses and other Mammalian positive-stranded RNA viruses. // J. Virol. 2006. Vol. 80, № 22. P. 11124–11140.
  185. Jorba J. et al. Calibration of multiple poliovirus molecular clocks covering an extended evolutionary range. // J. Virol. 2008. Vol. 82, № 9. P. 4429–4440.
  186. Bailly J.L.J. et al. Phylogeography of circulating populations of human

- echovirus 30 over 50 years: nucleotide polymorphism and signature of purifying selection in the VP1 capsid protein gene // *Infect Genet Evol.* 2009. Vol. 9, № 4. P. 699–708.
187. Smura T., Savolainen-Kopra C., Roivainen M. Evolution of newly described enteroviruses // *Future Virol.* 2011. Vol. 6, № 1. P. 109–131.
  188. Vignuzzi M. et al. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population // *Nature.* 2006. Vol. 439, № 7074. P. 344–348.
  189. Jiang P. et al. Evidence for emergence of diverse polioviruses from C-cluster coxsackie A viruses and implications for global poliovirus eradication. // *Proc. Natl. Acad. Sci. U. S. A.* 2007. Vol. 104, № 22. P. 9457–9462.
  190. Oberste M.S. et al. Complete genome sequences of all members of the species Human enterovirus A // *J. Gen. Virol.* 2004. Vol. 85, № 6. P. 1597–1607.
  191. Oberste M.S. et al. Enteroviruses 76, 89, 90 and 91 represent a novel group within the species Human enterovirus A // *J. Gen. Virol.* 2005. Vol. 86, № 2. P. 445–451.
  192. Oberste M.S. et al. Characterizing the picornavirus landscape among synanthropic nonhuman primates in Bangladesh, 2007 to 2008. // *J. Virol.* 2013. Vol. 87, № 1. P. 558–571.
  193. Oberste M.S. et al. Naturally acquired picornavirus infections in primates at the Dhaka zoo. // *J. Virol.* 2013. Vol. 87, № 1. P. 572–580.
  194. Harvala H. et al. Detection and genetic characterization of enteroviruses circulating among wild populations of chimpanzees in Cameroon: relationship with human and simian enteroviruses. // *J Virol.* 2011. Vol. 85, № 9. P. 4480–4486.
  195. Madeley C.R., Cosgrove B.P. 28 nm PARTICLES IN FÆCES IN INFANTILE GASTROENTERITIS // *Lancet.* 1975. Vol. 306, № 7932. P. 451–452.
  196. Mendenhall I.H., Smith G.J.D., Vijaykrishna D. Ecological Drivers of Virus Evolution: Astrovirus as a Case Study // *J. Virol.* 2015. Vol. 89, № 14. P. 6978–6981.
  197. Bosch A., Pintó R.M., Guix S. Human astroviruses // *Clin. Microbiol. Rev.* 2014. Vol. 27, № 4. P. 1048–1074.
  198. ViralZone: Astroviridae [Electronic resource]. URL: [https://viralzone.expasy.org/by\\_species/27](https://viralzone.expasy.org/by_species/27) (accessed: 11.07.2022).
  199. Méndez E., Arias C.F. Astroviridae // *Fields virology*, 6th ed. 2013. P. 609–628.
  200. Lulla V., Firth A.E. A hidden gene in astroviruses encodes a viroporin // *Nat. Commun.* 2020. Vol. 11, № 1. P. 4070.
  201. Jonassen T.O. et al. A common RNA motif in the 3' end of the genomes of astroviruses, avian infectious bronchitis virus and an equine rhinovirus. // *J. Gen. Virol.* 1998. Vol. 79, № 4. P. 715–718.
  202. De Grazia S. et al. Genetic heterogeneity and recombination in human type 2

- astroviruses // *J. Clin. Microbiol.* 2012. Vol. 50, № 11. P. 3760–3764.
203. Medici M.C. et al. Genetic heterogeneity and recombination in type-3 human astroviruses // *Infect. Genet. Evol.* 2015. Vol. 32. P. 156–160.
204. Ha H.J. et al. Complete genome sequencing of a recombinant strain between human astrovirus antigen types 2 and 8 isolated from South Korea // *Infect. Genet. Evol.* Elsevier B.V., 2016. Vol. 39. P. 127–131.
205. Zaraket H. et al. Characterization of astrovirus-associated gastroenteritis in hospitalized children under five years of age // *Infect. Genet. Evol.* Elsevier B.V., 2017. Vol. 53, № 2017. P. 94–99.
206. Walter J.E. et al. Molecular characterization of a novel recombinant strain of human astrovirus associated with gastroenteritis in children // *Arch. Virol.* 2001. Vol. 146, № 12. P. 2357–2367.
207. Wolfaardt M. et al. Evidence of a Recombinant Wild-Type Human Astrovirus Strain from a Kenyan Child with Gastroenteritis // *J. Clin. Microbiol.* 2011. Vol. 49, № 2. P. 728–731.
208. Babkin I. V. et al. Recombination analysis based on the HAstV-2 and HAstV-4 complete genomes // *Infect. Genet. Evol.* Elsevier B.V., 2014. Vol. 22, № January. P. 94–102.
209. Hata A. et al. Next-generation amplicon sequencing identifies genetically diverse human astroviruses, including recombinant strains, in environmental waters // *Sci. Rep.* Springer US, 2018. Vol. 8, № 1. P. 1–9.
210. Ahmed S.F. et al. Novel Astroviruses in Children, Egypt // *Emerg. Infect. Dis.* 2011. Vol. 17, № 12. P. 2391–2393.
211. Amimo J.O. et al. Whole genome sequence analysis of porcine astroviruses reveals novel genetically diverse strains circulating in east african smallholder pig farms // *Viruses.* 2020. Vol. 12, № 11. P. 1–17.
212. Ito M. et al. Whole genome analysis of porcine astroviruses detected in Japanese pigs reveals genetic diversity and possible intra-genotypic recombination // *Infect. Genet. Evol.* Elsevier B.V., 2017. Vol. 50. P. 38–48.
213. Zhao C. et al. Genomic characterization of a novel recombinant porcine astrovirus isolated in northeastern China // *Arch. Virol.* 2019. Vol. 164, № 5. P. 1469–1473.
214. Li M. et al. Prevalence and genome characteristics of canine astrovirus in southwest China // *J. Gen. Virol.* 2018. Vol. 99, № 7. P. 880–889.
215. Wang Y. et al. Genetic characterization and phylogenetic analysis of feline astrovirus from Anhui province in eastern China // *3 Biotech.* Springer International Publishing, 2020. Vol. 10, № 8. P. 1–10.
216. Lan D. et al. Molecular characterization of a porcine astrovirus strain in China // *Arch. Virol.* 2011. Vol. 156, № 10. P. 1869–1875.
217. Rivera R. et al. Characterization of phylogenetically diverse astroviruses of marine mammals // *J. Gen. Virol.* 2010. Vol. 91, № 1. P. 166–173.
218. Chen X. et al. A novel astrovirus species in the gut of yaks with diarrhoea in the Qinghai-Tibetan plateau, 2013 // *J. Gen. Virol.* 2015. Vol. 96, № 12. P. 3672–3680.

219. Atkins A. et al. Characterization of an outbreak of astroviral diarrhea in a group of cheetahs (*Acinonyx jubatus*) // *Vet. Microbiol.* 2009. Vol. 136, № 1–2. P. 160–165.
220. De Benedictis P. et al. Astrovirus infections in humans and animals - Molecular biology, genetic diversity, and interspecies transmissions // *Infect. Genet. Evol.* Elsevier B.V., 2011. Vol. 11, № 7. P. 1529–1544.
221. Martella V. et al. Analysis of the ORF2 of human astroviruses reveals lineage diversification, recombination and rearrangement and provides the basis for a novel sub-classification system // *Arch. Virol.* 2014. Vol. 159, № 12. P. 3185–3196.
222. Zhang W. et al. Epidemiology, genetic diversity and evolution of canine astrovirus // *Transbound. Emerg. Dis.* 2020. Vol. 67, № 6. P. 2901–2910.
223. Woo P.C.Y. et al. A novel astrovirus from dromedaries in the Middle East // *J. Gen. Virol.* 2015. Vol. 96, № 9. P. 2697–2707.
224. Ulloa J.C., Gutiérrez M.F. Genomic analysis of two ORF2 segments of new porcine astrovirus isolates and their close relationship with human astroviruses. // *Can. J. Microbiol.* 2010. Vol. 56, № 7. P. 569–577.
225. Tse H. et al. Rediscovery and genomic characterization of bovine astroviruses // *J. Gen. Virol.* 2011. Vol. 92, № 8. P. 1888–1898.
226. Luo Z. et al. Multiple novel and prevalent astroviruses in pigs // *Vet. Microbiol.* 2011. Vol. 149, № 3–4. P. 316–323.
227. Lv S.L. et al. High genetic diversity and recombination events of porcine astrovirus strains identified from ill and asymptomatic pigs in 2017, Hunan Province, China // *Virus Genes.* Springer US, 2019. Vol. 55, № 5. P. 673–681.
228. Belliot G., Laveran H., Monroe S.S. Detection and genetic differentiation of human astroviruses: Phylogenetic grouping varies by coding region // *Arch. Virol.* 1997. Vol. 142, № 7. P. 1323–1334.
229. Kauer R. V. et al. Discovery of novel astrovirus genotype species in small ruminants // *PeerJ.* 2019. Vol. 2019, № 7. P. 1–20.
230. Pativada M., Bhattacharya R., Krishnan T. Novel human astrovirus strains showing multiple recombinations within highly conserved ORF1b detected from hospitalized acute watery diarrhea cases in Kolkata, India // *Infect. Genet. Evol.* Elsevier B.V., 2013. Vol. 20. P. 284–291.
231. Masters P.S., Perlman S. *Coronaviridae* // *Fields virology*, 6th ed. 2013. P. 825–858.
232. Walker P.J. et al. Changes to virus taxonomy and the Statutes ratified by the International Committee on Taxonomy of Viruses (2020). // *Arch. Virol.* 2020. Vol. 165, № 11. P. 2737–2748.
233. King A.M.Q. et al. *Virus Taxonomy. Ninth Report of the International Committee on Taxonomy of Viruses.* 2011.
234. The 2021 ICTV taxonomy release [Electronic resource]. 2021. URL: <https://ictv.global/taxonomy>.
235. Hon C.-C. et al. Evidence of the Recombinant Origin of a Bat Severe Acute



- Respiratory Syndrome (SARS)-Like Coronavirus and Its Implications on the Direct Ancestor of SARS Coronavirus // *J. Virol.* 2008. Vol. 82, № 4. P. 1819–1826.
236. Sabir J.S.M. et al. Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia // *Science* (80-. ). 2016. Vol. 351, № 6268. P. 81–84.
237. Li X. et al. Emergence of SARS-CoV-2 through recombination and strong purifying selection // *Sci. Adv.* 2020. Vol. 6, № 27. P. eabb9153.
238. Woo P.C.Y. et al. Discovery of Seven Novel Mammalian and Avian Coronaviruses in the Genus Deltacoronavirus Supports Bat Coronaviruses as the Gene Source of Alphacoronavirus and Betacoronavirus and Avian Coronaviruses as the Gene Source of Gammacoronavirus and Deltacoronavi // *J. Virol.* 2012. Vol. 86, № 7. P. 3995–4008.
239. Jo W.K. et al. Potential zoonotic sources of SARS-CoV-2 infections // *Transbound. Emerg. Dis.* 2020. № October. P. 1–11.
240. Wille M., Holmes E.C. Wild birds as reservoirs for diverse and abundant gamma- and deltacoronaviruses // *FEMS Microbiol. Rev.* 2020. Vol. 44, № 5. P. 631–644.
241. Fehr A.R., Perlman S. Coronaviruses: An Overview of Their Replication and Pathogenesis. 2015. P. 1–23.
242. Baker S.C. Coronaviruses: Molecular Biology // *Desk Encyclopedia of General Virology.* 2010. P. 445–452.
243. Makino S. et al. High-frequency RNA recombination of murine coronaviruses. // *J. Virol.* 1986. Vol. 57, № 3. P. 729–737.
244. Baric R.S. et al. Establishing a genetic recombination map for murine coronavirus strain A59 complementation groups // *Virology.* 1990. Vol. 177, № 2. P. 646–656.
245. Banner L.R., Mc Lai M. Random nature of coronavirus RNA recombination in the absence of selection pressure // *Virology.* 1991. Vol. 185, № 1. P. 441–445.
246. Fu K., Baric R.S. Map locations of mouse hepatitis virus temperature-sensitive mutants: confirmation of variable rates of recombination. // *J. Virol.* 1994. Vol. 68, № 11. P. 7458–7466.
247. Zhang X. et al. Identification of a natural recombinant transmissible gastroenteritis virus between Purdue and Miller clusters in China // *Emerg. Microbes Infect.* 2017. Vol. 6, № 1. P. 1–10.
248. Boniotti M.B. et al. Porcine Epidemic Diarrhea Virus and Discovery of a Recombinant Swine Enteric Coronavirus, Italy // *Emerg. Infect. Dis.* 2016. Vol. 22, № 1. P. 83–87.
249. Jackwood M.W. et al. Emergence of a group 3 coronavirus through recombination // *Virology.* 2010. Vol. 398, № 1. P. 98–108.
250. Chen Q. et al. The emergence of novel sparrow deltacoronaviruses in the United States more closely related to porcine deltacoronaviruses than sparrow deltacoronavirus HKU17 // *Emerg. Microbes Infect.* 2018. Vol. 7, №

1. P. 1–4.
251. Herrewegh A.A.P.M. et al. Feline Coronavirus Type II Strains 79-1683 and 79-1146 Originate from a Double Recombination between Feline Coronavirus Type I and Canine Coronavirus // *J. Virol.* 1998. Vol. 72, № 5. P. 4508–4514.
252. Lau S.K.P. et al. Molecular Epidemiology of Human Coronavirus OC43 Reveals Evolution of Different Genotypes over Time and Recent Emergence of a Novel Genotype due to Natural Recombination // *J. Virol.* 2011. Vol. 85, № 21. P. 11325–11337.
253. Quinteros J.A. et al. Full genome analysis of Australian infectious bronchitis viruses suggests frequent recombination events between vaccine strains and multiple phylogenetically distant avian coronaviruses of unknown origin // *Vet. Microbiol.* 2016. Vol. 197. P. 27–38.
254. Marandino A. et al. Whole-genome characterization of Uruguayan strains of avian infectious bronchitis virus reveals extensive recombination between the two major South American lineages // *Infect. Genet. Evol.* 2017. Vol. 54. P. 245–250.
255. Moharam I. et al. Emerging infectious bronchitis virus (IBV) in Egypt: Evidence for an evolutionary advantage of a new S1 variant with a unique gene 3ab constellation // *Infect. Genet. Evol.* 2020. Vol. 85. P. 104433.
256. Lee C.-W., Jackwood M.W. Origin and evolution of Georgia 98 (GA98), a new serotype of avian infectious bronchitis virus // *Virus Res.* 2001. Vol. 80, № 1–2. P. 33–39.
257. Lee C.W., Jackwood M.W. Spike gene analysis of the DE072 strain of infectious bronchitis virus: origin and evolution // *Virus Genes.* 2001. Vol. 22, № 1. P. 85–91.
258. Kusters J.G.G. et al. Sequence evidence for RNA recombination in field isolates of avian coronavirus infectious bronchitis virus // *Vaccine.* 1990. Vol. 8, № 6. P. 605–608.
259. Jia W. et al. A novel variant of avian infectious bronchitis virus resulting from recombination among three different strains // *Arch. Virol.* 1995. Vol. 140, № 2. P. 259–271.
260. Mardani K. et al. Naturally occurring recombination between distant strains of infectious bronchitis virus // *Arch. Virol.* 2010. Vol. 155, № 10. P. 1581–1586.
261. Mondal S.P., Cardona C.J. Genotypic and phenotypic characterization of the California 99 (Cal99) variant of infectious bronchitis virus // *Virus Genes.* 2007. Vol. 34, № 3. P. 327–341.
262. Munyahongse S. et al. Genetic characterization of infectious bronchitis viruses in Thailand, 2014–2016: identification of a novel recombinant variant // *Poult. Sci.* 2020. Vol. 99, № 4. P. 1888–1895.
263. Li D. et al. Molecular evolution of porcine epidemic diarrhea virus and porcine deltacoronavirus strains in Central China // *Res. Vet. Sci.* 2018. Vol. 120. P. 63–69.

264. Huang Y.-W. et al. Origin, Evolution, and Genotyping of Emergent Porcine Epidemic Diarrhea Virus Strains in the United States // *MBio* / ed. Griffin D.E. 2013. Vol. 4, № 5.
265. Wesley R.D. The S gene of canine coronavirus, strain UCD-1, is more closely related to the S gene of transmissible gastroenteritis virus than to that of feline infectious peritonitis virus // *Virus Res.* 1999. Vol. 61, № 2. P. 145–152.
266. Escutenaire S. et al. Characterization of divergent and atypical canine coronaviruses from Sweden // *Arch. Virol.* 2007. Vol. 152, № 8. P. 1507–1514.
267. Decaro N. et al. Recombinant Canine Coronaviruses Related to Transmissible Gastroenteritis Virus of Swine Are Circulating in Dogs // *J. Virol.* 2009. Vol. 83, № 3. P. 1532–1537.
268. Corman V.M. et al. Evidence for an Ancestral Association of Human Coronavirus 229E with Bats // *J. Virol.* / ed. Schultz-Cherry S. 2015. Vol. 89, № 23. P. 11858–11870.
269. Corman V.M. et al. Rooting the Phylogenetic Tree of Middle East Respiratory Syndrome Coronavirus by Characterization of a Conspecific Virus from an African Bat // *J. Virol.* 2014. Vol. 88, № 19. P. 11297–11303.
270. Lau S.K.P. et al. Discovery and Sequence Analysis of Four Deltacoronaviruses from Birds in the Middle East Reveal Interspecies Jumping with Recombination as a Potential Mechanism for Avian-to-Avian and Avian-to-Mammalian Transmission // *J. Virol.* / ed. Gallagher T. 2018. Vol. 92, № 15.
271. Graham R.L., Baric R.S. Recombination, Reservoirs, and the Modular Spike: Mechanisms of Coronavirus Cross-Species Transmission // *J. Virol.* 2010. Vol. 84, № 7. P. 3134–3146.
272. Becker M.M. et al. Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice // *Proc. Natl. Acad. Sci.* 2008. Vol. 105, № 50. P. 19944–19949.
273. He W.-T.W. et al. Genomic Epidemiology, Evolution, and Transmission Dynamics of Porcine Deltacoronavirus // *Mol. Biol. Evol.* / ed. Pond S.K. 2020. Vol. 37, № 9. P. 2641–2654.
274. Pyrc K. et al. Mosaic Structure of Human Coronavirus NL63, One Thousand Years of Evolution // *J. Mol. Biol.* 2006. Vol. 364, № 5. P. 964–973.
275. So R.T.Y. et al. Diversity of Dromedary Camel Coronavirus HKU23 in African Camels Revealed Multiple Recombination Events among Closely Related Betacoronaviruses of the Subgenus Embecovirus // *J. Virol.* / ed. Pfeiffer J.K. 2019. Vol. 93, № 23.
276. Lee C.-W., Jackwood M.W. Evidence of genetic diversity generated by recombination among avian coronavirus IBV // *Arch. Virol.* 2000. Vol. 145, № 10. P. 2135–2148.
277. Lukashev A.N. et al. Close genetic relatedness of picornaviruses from European and Asian bats // *J. Gen. Virol.* 2017. Vol. 98, № 5. P. 955–961.

278. Huang C. et al. A Bat-Derived Putative Cross-Family Recombinant Coronavirus with a Reovirus Gene // *PLOS Pathog.* / ed. Baker S. 2016. Vol. 12, № 9. P. e1005883.
279. Luytjes W. et al. Sequence of mouse hepatitis virus A59 mRNA 2: Indications for RNA recombination between coronaviruses and influenza C virus // *Virology*. 1988. Vol. 166, № 2. P. 415–422.
280. Yamada K.D., Tomii K., Katoh K. Application of the MAFFT sequence alignment program to large data—reexamination of the usefulness of chained guide trees // *Bioinformatics*. 2016. Vol. 32, № 21. P. 3246–3251.
281. Capella-Gutierrez S., Silla-Martinez J.M., Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses // *Bioinformatics*. 2009. Vol. 25, № 15. P. 1972–1973.
282. UniProt Consortium T. UniProt: the universal protein knowledgebase // *Nucleic Acids Res.* 2018. Vol. 46, № 5. P. 2699–2699.
283. Katoh K., Standley D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability // *Mol. Biol. Evol.* 2013. Vol. 30, № 4. P. 772–780.
284. Blom N. et al. Cleavage site analysis in picornaviral polyproteins: Discovering cellular targets by neural networks // *Protein Sci.* 1996. Vol. 5, № 11. P. 2203–2216.
285. Hall T.A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT // *Nucleic Acids Symp. Ser.* 1999. Vol. 42. P. 95–98.
286. Jakobsen I.B., Easteal S. A program for calculating and displaying compatibility matrices as an aid in determining reticulate evolution in molecular sequences // *Bioinformatics*. 1996. Vol. 12, № 4. P. 291–295.
287. Martin D.P. et al. RDP4: Detection and analysis of recombination patterns in virus genomes // *Virus Evol.* 2015. Vol. 1, № 1. P. 1–5.
288. Deviatkin A.A.A.A.A., Lukashev A.N.A.N. Recombination in the rabies virus and other lyssaviruses // *Infect. Genet. Evol. Elsevier*, 2018. Vol. 60, № January. P. 97–102.
289. Martin D., Rybicki E. RDP: Detection of recombination amongst aligned sequences // *Bioinformatics*. 2000. Vol. 16, № 6. P. 562–563.
290. Sawyer S. Statistical tests for detecting gene conversion // *Mol. Biol. Evol.* 1989.
291. Salminen M.O. et al. Identification of Breakpoints in Intergenotypic Recombinants of HIV Type 1 by Bootscanning // *AIDS Res. Hum. Retroviruses*. 1995. Vol. 11, № 11. P. 1423–1425.
292. Smith J.M. Analyzing the mosaic structure of genes // *J. Mol. Evol.* 1992. Vol. 34, № 2. P. 126–129.
293. Posada D., Crandall K.A. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations // *Proc. Natl. Acad. Sci. U. S. A.* 2001. Vol. 98, № 24. P. 13757–13762.
294. Gibbs M.J., Armstrong J.S., Gibbs A.J. Sister-scanning: A Monte Carlo

- procedure for assessing signals in recombining sequences // *Bioinformatics*. 2000. Vol. 16, № 7. P. 573–582.
295. Weiller G.F. Phylogenetic profiles: A graphical method for detecting genetic recombinations in homologous sequences // *Mol. Biol. Evol.* 1998. Vol. 15, № 3. P. 326–335.
  296. Holmes E.C., Worobey M., Rambaut A. Phylogenetic evidence for recombination in dengue virus // *Mol. Biol. Evol.* 1999. Vol. 16, № 3. P. 405–409.
  297. Boni M.F., Posada D., Feldman M.W. An exact nonparametric method for inferring mosaic structure in sequence triplets // *Genetics*. 2007. Vol. 176, № 2. P. 1035–1047.
  298. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. 2016.
  299. Nguyen L.-T. et al. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies // *Mol. Biol. Evol.* 2015. Vol. 32, № 1. P. 268–274.
  300. Minh B.Q., Nguyen M.A.T., von Haeseler A. Ultrafast Approximation for Phylogenetic Bootstrap // *Mol. Biol. Evol.* 2013. Vol. 30, № 5. P. 1188–1195.
  301. Kalyaanamoorthy S. et al. ModelFinder: fast model selection for accurate phylogenetic estimates // *Nat. Methods*. 2017. Vol. 14, № 6. P. 587–589.
  302. Rambaut A. FigTree 1.4.4 [Electronic resource]. URL: <https://github.com/rambaut/figtree/releases> (accessed: 01.06.2019).
  303. Yu G. Using ggtree to visualize data on tree-like structures // *Curr. Protoc. Bioinforma.* 2020.
  304. Hill V., Baele G. Bayesian Estimation of Past Population Dynamics in BEAST 1.10 Using the Skygrid Coalescent Model // *Mol. Biol. Evol.* / ed. Shapiro B. 2019.
  305. Lanfear R. et al. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses // *Mol. Biol. Evol.* 2016. P. msw260.
  306. Drummond A. et al. Bayesian phylogenetics with BEAUti and the BEAST 1.7 // *Mol. Biol. Evol.* 2012. Vol. 29, № 8. P. 1969–1973.
  307. Shapiro B., Rambaut A., Drummond A.J. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences // *Mol Biol Evol.* 2006. Vol. 23. P. 7–9.
  308. GraphPad Prism [Electronic resource]. URL: [www.graphpad.com](http://www.graphpad.com).
  309. Rambaut A. et al. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen) // *Virus Evol.* 2016. Vol. 2, № 1.
  310. Kumar S., Stecher G., Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets // *Mol Biol Evol.* 2016. Vol. 33. P. 1870–1874.
  311. Pedregosa F. et al. Scikit-learn: Machine Learning in Python // *J. Mach. Learn. Res.* 2011. Vol. 12. P. 2825–2830.

312. Simmonds P., Midgley S. Recombination in the Genesis and Evolution of Hepatitis B Virus Genotypes // *J. Virol.* 2005. Vol. 79, № 24. P. 15467–15476.
313. Su S. et al. Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses // *Trends Microbiol.* Elsevier Ltd, 2016. Vol. 24, № 6. P. 490–502.
314. Boni M.F. et al. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic // *Nat. Microbiol.* Springer US, 2020. Vol. 5, № 11. P. 1408–1417.
315. Wertheim J.O. et al. A Case for the Ancient Origin of Coronaviruses // *J. Virol.* 2013. Vol. 87, № 12. P. 7039–7045.
316. Kobayashi M. et al. Molecular Evolution of the Capsid Gene in Norovirus Genogroup I // *Sci. Rep.* Nature Publishing Group, 2015. Vol. 5, № 1. P. 1–9.
317. Parra G.I. et al. Static and Evolving Norovirus Genotypes: Implications for Epidemiology and Immunity // *PLoS Pathog.* 2017. Vol. 13, № 1. P. 1–22.
318. Kobayashi M. et al. Molecular evolution of the capsid gene in human norovirus genogroup II // *Sci. Rep.* 2016. Vol. 6, № 1. P. 29400.
319. Kerr P.J., Kitchen A., Holmes E.C. Origin and Phylodynamics of Rabbit Hemorrhagic Disease Virus // *J. Virol.* 2009. Vol. 83, № 23. P. 12129–12138.
320. Forni D. et al. Molecular Evolution of Human Coronavirus Genomes // *Trends Microbiol.* 2017. Vol. 25, № 1. P. 35–48.
321. Solomon T. et al. Virology, epidemiology, pathogenesis, and control of enterovirus 71 // *Lancet Infect Dis.* 2010. Vol. 10. P. 778–790.
322. Saxena V.K. et al. Genetic Diversity of Enterovirus A71, India // *Emerg. Infect. Dis.* 2015. Vol. 21, № 1. P. 123–126.
323. McMinn P.C. Recent advances in the molecular epidemiology and control of human enterovirus 71 infection // *Curr. Opin. Virol.* Elsevier B.V., 2012. Vol. 2, № 2. P. 199–205.
324. Yi E.-J. et al. Enterovirus 71 infection and vaccines // *Clin. Exp. Vaccine Res.* 2017. Vol. 6, № 1. P. 4.
325. Waterhouse A.M. et al. Jalview Version 2 - a multiple sequence alignment editor and analysis workbench // *Bioinformatics.* 2009. Vol. 25, № 9. P. 1189–1191.
326. Fu L. et al. CD-HIT: Accelerated for clustering the next-generation sequencing data // *Bioinformatics.* 2012. Vol. 28, № 23. P. 3150–3152.
327. Edgar R.C. Search and clustering orders of magnitude faster than BLAST // *Bioinformatics.* 2010. Vol. 26, № 19. P. 2460–2461.
328. Rice P., Longden I., Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite // *Trends Genet.* 2000. Vol. 6, № 6. P. 276-7.
329. Lukashev A. Recombination among picornaviruses // *Rev. Med. Virol.* 2010. Vol. 20, № 5. P. 327–337.
330. Yen T.-Y. et al. A case of recombinant coxsackievirus A2 infection with neurological complications in Taiwan // *J. Microbiol. Immunol. Infect.* 2017.

- Vol. 50, № 6. P. 928–930.
331. Kirkegaard K., Baltimore D. The mechanism of RNA recombination in poliovirus // *Cell*. 1986. Vol. 47, № 3. P. 433–443.
  332. Baric R.S. et al. Murine Coronavirus Temperature Sensitive Mutants. 1990. P. 349–356.
  333. Woo P.C.Y. et al. Comparative Analysis of 22 Coronavirus HKU1 Genomes Reveals a Novel Genotype and Evidence of Natural Recombination in Coronavirus HKU1 // *J. Virol.* 2006. Vol. 80, № 14. P. 7136–7145.
  334. Flores-Alanis A. et al. The receptor binding domain of SARS-CoV-2 spike protein is the result of an ancestral recombination between the bat-CoV RaTG13 and the pangolin-CoV MP789 // *BMC Res. Notes*. 2020. Vol. 13, № 1. P. 398.
  335. Zhang X.W., Yap Y.L., Danchin A. Testing the hypothesis of a recombinant origin of the SARS-associated coronavirus // *Arch. Virol.* 2005. Vol. 150, № 1. P. 1–20.
  336. Hu B. et al. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus // *PLOS Pathog.* / ed. Drosten C. 2017. Vol. 13, № 11. P. e1006698.
  337. Thor S.W. et al. Recombination in Avian Gamma-Coronavirus Infectious Bronchitis Virus // *Viruses*. 2011. Vol. 3, № 9. P. 1777–1799.
  338. Wang Q.Q. et al. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2 // *Cell*. 2020. Vol. 181, № 4. P. 894-904.e9.
  339. Li F. Receptor Recognition Mechanisms of Coronaviruses: a Decade of Structural Studies // *J. Virol.* 2015. Vol. 89, № 4. P. 1954–1964.
  340. Nascimento F.F., Reis M. Dos, Yang Z. A biologist’s guide to Bayesian phylogenetic analysis // *Nat. Ecol. Evol.* Springer US, 2017. Vol. 1, № 10. P. 1446–1454.
  341. Baele G. et al. Improving the Accuracy of Demographic and Molecular Clock Model Comparison While Accommodating Phylogenetic Uncertainty // *Mol. Biol. Evol.* 2012. Vol. 29, № 9. P. 2157–2167.
  342. Edgar R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput // *Nucleic Acids Res.* 2004. Vol. 32, № 5. P. 1792–1797.
  343. Graur D., Pupko T. The Permian Bacterium that Isn’t // *Mol. Biol. Evol.* 2001. Vol. 18, № 6. P. 1143–1146.
  344. Worobey M. Phylogenetic Evidence against Evolutionary Stasis and Natural Abiotic Reservoirs of Influenza A Virus // *J. Virol.* 2008. Vol. 82, № 7. P. 3769–3774.
  345. Koonin E. V et al. The Big Bang of picorna-like virus evolution antedates the radiation of eukaryotic supergroups. // *Nat. Rev. Microbiol.* 2008. Vol. 6, № 12. P. 925–939.
  346. Palacios G. et al. Molecular epidemiology of echovirus 30: temporal circulation and prevalence of single lineages // *J Virol.* 2002. Vol. 76, № 10. P. 4940–4949.
  347. Rao C.D., Yergolkar P., Shankarappa K.S. Antigenic diversity of

- enteroviruses associated with nonpolio acute flaccid paralysis, India, 2007-2009 // *Emerg Infect Dis.* 2012. Vol. 18. P. 1833–1840.
348. Lukashev A., Vakulenko Y. Molecular evolution of types in non-polio enteroviruses // *J Gen Virol. Microbiology Society*, 2017. Vol. 98, № 12. P. 2968–2981.