

УДК 599.113

## ГЕНЕТИЧЕСКАЯ КОНЦЕПЦИЯ ВИДА В ПРИЛОЖЕНИИ К МУЛЬТИЛОКУСНЫМ ЯДЕРНЫМ ДАННЫМ

© 2024 г. В. Л. Лебедев<sup>1, \*</sup>, С. А. Косушкин<sup>2, \*\*</sup>, А. А. Банникова<sup>3, \*\*\*</sup>

<sup>1</sup>Зоологический музей Московского государственного университета им. М.В. Ломоносова  
ул. Большая Никитская, 2, Москва, 125009 Россия

<sup>2</sup>Институт молекулярной биологии им. В.А. Энгельгардта РАН  
ул. Вавилова, 32, Москва, 119991 Россия

<sup>3</sup>Московский государственный университет им. М.В. Ломоносова,  
биологический факультет, кафедра зоологии позвоночных  
Ленинские горы, 1, стр. 12, Москва, 119234 Россия

\*E-mail: wslebedev@mail.ru

\*\*E-mail: toki@mail.ru

\*\*\*E-mail: hylomys@mail.ru

Поступила в редакцию 05.11.2023 г.

После доработки 28.11.2023 г.

Принята в печать 05.12.2023 г.

На примере целого ряда видовых групп и сложных видовых комплексов млекопитающих эмпирически показано, что для изучения их генетической дифференциации разрешающей способности набора из 10–20 ядерных локусов недостаточно. Для удовлетворительного анализа филогенетических отношений (включая события ретикуляций) и определения уровня дивергенции в пределах таких групп близкородственных видов необходимо умножение числа ядерных локусов и переход на уровень глубокого секвенирования нового поколения. Однако вопрос о том, какая часть исследуемого генома позволит достигнуть желаемого результата, иными словами, какое именно количество генов необходимо для того, чтобы получить генетическую дистанцию, разрешающую отношения ветвей филогенетического дерева комплекса близкородственных видов, остается открытым. В качестве стандартного набора маркеров мы выбрали белок-кодирующие участки генов. Дистанции для 160–180 генов общей длиной 270–300 тыс. п.н. использовали для оценки уровня дивергенции близких видов млекопитающих. Выяснено, что в большинстве случаев дистанции между видами находятся в диапазоне от 0.15 до 0.75% (медиана 0.33%). Диапазон дистанций между полувидами составляет от 0.12 до 0.28% (медиана 0.14%). Внутривидовые дистанции всегда меньше 0.11%. У грызунов наблюдаются более высокие значения межвидовых дистанций – от 0.25 до 2.3% (медиана 0.72%); дистанции между полувидами находятся в диапазоне от 0.1 до 0.35%. Для определения числа ядерных локусов и их совокупной длины, необходимых для расчета генетической дистанции, использовали симуляции на основании модели, включавшей следующие параметры: средняя скорость эволюции генов, ее дисперсия, уровень полиморфизма в современных и предковых популяциях. Предварительно был проведен анализ распределения скоростей локусов среди таксонов млекопитающих на основе ранее опубликованных данных по ~50 тыс. ядерных маркеров. Показано, что относительная ошибка в 10–15%, сопоставимая с аналогичной величиной для митохондриальных дистанций по отдельным генам (длиной примерно 1000 п.н.) между близкими видами, достигается при использовании около 100 локусов длиной 300 п.н. На основе этих данных выдвинуто следующее предположение: порог межвидовой/внутривидовой генетической дистанции, рассчитываемой по наиболее часто используемым в молекулярной филогенетике млекопитающих экзонам, составляет около 0.15%. Данная гипотеза предполагает признание за формами *in statu nascendi* видовой ранг. Следует заметить, что не все “хорошие” виды имеют значительные генетические дистанции в связи с тем, что репродуктивные барьеры иногда формируются быстрее, чем накапливаются замены по геному.

DOI: 10.31857/S0044459624040014, EDN: UTXNCJ

## ЭВОЛЮЦИЯ ГЕНЕТИЧЕСКОЙ КОНЦЕПЦИИ ВИДА

Генетическая концепция вида (ГКВ) прошла длительную эволюцию, и существует множество ее интерпретаций (Bateson, 1909; Taverner, 1920; Muller, 1939; Simpson, 1943; Dobzhansky, 1950; Mayr, 1969; Nei, 1976; Baker, Bickham, 1986; Masters, Spencer, 1989; Avise, Ball, 1990; Mayden, 1997; Schilthuizen, 2000; Butlin, 2005). Основа генетической концепции видов впервые изложена Бейтсоном более 100 лет назад (Bateson, 1909) и получила дальнейшее развитие у Добжанского (Dobzhansky, 1937, 1950) и Мюллера (Muller, 1939). По Добжанскому, вид – это система популяций, генный обмен между которыми ограничен или запрещен из-за репродуктивных изолирующих механизмов или комбинации нескольких таких механизмов. Таким образом, ГКВ Добжанского – это практически изоляционная концепция, так как подчеркивает репродуктивную изолированность видов, чем фактически напоминает концепцию Бейтсона (Bateson, 1909). Позже была предложена модель Бейтсона–Добжанского–Мюллера (Bateson–Dobzhansky–Muller, BDM), согласно которой к образованию двух видов приводит накопление генетических изменений в двух отдельных популяциях (Gavrilets, 2003). Здесь имеется в виду довольно сложная архитектура репродуктивной изоляции, а именно, указание на “несовместимость” аллелей в локусах (Gavrilets, 2003).

В такой интерпретации ГКВ частично совпадает с биологической концепцией вида (БКВ), поскольку скрещивающиеся популяции, которые репродуктивно и генетически не изолированы, считаются конспецифичными, а популяции, репродуктивно и генетически изолированные, – отдельными видами. Эти две концепции расходятся в том случае, когда генетическая изоляция популяций существует, но репродуктивная изоляция отсутствует. В этом случае БКВ в трактовке Добжанского (Dobzhansky, 1937) и Майра (Mayr, 1942), но не Койна и Орра (Coyn, Orr, 2004) должна считать такие популяции конспецифичными, а ГКВ – принадлежащими к разным видам.

Многочисленные примеры гибридизации между видами, известные на сегодня, сильно ослабляют ориентацию исключительно на репродуктивно изолированные популяции. С учетом этого соображения Койн и Орр (Coyn, Orr, 2004) предложили новое определение биологической концепции вида, дополнив ее возможностью

гибридизации. С точки зрения этих авторов, ГКВ эквивалентна БКВ или является подмножеством последней.

В приложении к млекопитающим, согласно современной ГКВ, две генетические линии являются видами, если целостность их генофондов сохраняется даже в присутствии потока генов, т.е. защищена (Baker, Bradley, 2006). Таким образом, вид – это совокупность природных популяций, которые генетически совместимы друг с другом, но изолированы от других подобных совокупностей, а видообразование есть аккумуляция генетических изменений между двумя филетическими линиями, что обеспечивает защиту и изоляцию их генофондов (Baker, Bradley, 2006).

Практически в любой интерпретации ГКВ для вывода о репродуктивной изоляции и эволюционной независимости генетических линий предполагается измерение генетических различий (Mayden, 1997). Однако главное критическое замечание в адрес ГКВ состоит в неопределенности величины и единиц измерения генетической изменчивости, необходимой для решения вопроса “вид/подвид”. Таким образом, проведение границы между видовым и внутривидовым уровнем дивергенции по-прежнему остается нерешенной задачей теоретического уровня.

Бэйкер и Брэдли (Baker, Bradley, 2006) предложили операциональный критерий, который состоит в количественной оценке уровня дивергенции геномов через (1) генетические дистанции, (2) анализ потока генов между таксонами, (3) оценку уровня гибридизации, (4) оценку времени дивергенции. Брэдли и Бэйкер (Bradley, Baker, 2001) разработали методику, позволяющую практически реализовать ГКВ в пределах класса млекопитающих. Анализируя генетические дистанции между сестринскими видами в разных группах млекопитающих по митохондриальному гену цитохрома *b* (*cyt b*), Брэдли и Бэйкер (Bradley, Baker, 2001) пришли к заключению, что большинство из них отличается на 4–11% нуклеотидных замен, дистанция  $\leq 2\%$  обычна для внутривидового уровня, дистанция  $\geq 10\%$  характеризует сложившиеся виды “...наиболее прагматичным первым шагом является использование порога генетической дистанции в 5% по гену цитохрома *b*, поскольку во многих случаях адекватная оценка внутривидовой изменчивости не доступна; при этом дистанция  $\geq 5\%$  – это очевидно приблизительная оценка, она не дает окончательного решения” (Baker, Bradley, 2006, p. 657). Хотя эту методологию, как и всю генетическую

концепцию, нельзя применять без информации об уровнях генетических дистанций между сестринскими видами в составе определенного таксона, она сделалась полезным инструментом практикующих систематиков.

### ПЕРЕХОД К МУЛЬТИЛОКУСНЫМ И ГЕНОМНЫМ ДАННЫМ

Итак, ранее ГКВ в интерпретации Брэдли и Бэйкера (Bradley, Baker, 2001; Baker, Bradley, 2006) предложила подход, который состоит в оценке генетических дистанций и выборе в качестве видового уровня таких различий, которые соответствуют дивергенции между ближайшими таксонами, для которых показано наличие генетической изоляции. Этот количественный, довольно удобный и по-настоящему операциональный критерий тем не менее на практике реализуется только на базе митохондриальных данных, что нередко приводит к ошибкам из-за интрогрессии чужеродной митохондриальной ДНК (мтДНК) вследствие исторической или современной гибридизации. Применение этого критерия к ядерным данным затрудняется меньшей изменчивостью ядерных генов и использованием в разных случаях разных локусов и разного их числа.

Между тем мультилокусный анализ генетического разнообразия — это ключевой подход к современному решению филогенетических, филогеографических и популяционно-генетических задач. В настоящее время общепринятой практикой становится использование данных высокопроизводительного секвенирования, что связано с постепенным исчерпанием возможностей секвенирования сотен отдельных генов по методу Сэнгера, с одной стороны, и накоплением материала по полногеномным или редуцированным геномным данным (например, транскриптомы или маркеры RADSeq) — с другой. Стремительный переход от однолокусных, в том числе митохондриальных, наборов данных к геномным ресурсам поднимает вопрос о том, как из этих новых огромных массивов информации получать представление о структуре филогенетических отношений и о временах дивергенций и как эти данные транслировать в стандартные таксономические ранги. Это означает и необходимость трансформации современной генетической концепции вида.

Адаптация ГКВ к мультилокусному анализу будет означать появление нового операционального критерия вида — мультигенной дистанции.

В последнем случае встает важная задача определения минимального объема генетической информации (числа генетических локусов и их информативности), необходимого для получения устойчивого филогенетического сигнала, восстановления порядка кладогенетических событий, верной оценки величины генного потока и в конечном итоге вывода о том, соответствуют ли изучаемые группы видам или внутривидовым категориям. В результате приложения операционального критерия ГКВ к мультигенным ядерным данным должно быть установлено соответствие между уровнем дивергенции по сумме ядерных локусов и уровнем репродуктивной изоляции. В свою очередь, уровень различия ядерных генов должен быть сопоставлен со временем дивергенции и приведен в соответствие с таксономическим рангом. Поскольку количественные критерии ГКВ могут различаться в разных таксономических группах, важно определить диапазон этих различий. Следовательно, мультигенная дистанция и время дивергенции как критерии вида должны представлять собой не точечные значения, а диапазоны дистанций и времен. В рамках практической систематики это должно облегчить принятие таксономических решений в сложных с точки зрения систематики группах и определить ревизию целого ряда проблемных таксонов.

### МУЛЬТИГЕННАЯ ДИСТАНЦИЯ И ВРЕМЯ ДИВЕРГЕНЦИИ

Перевод генетической концепции в плоскость времени обуславливает практическую применимость еще одного видового критерия — времени дивергенции.

Использование генетических дистанций для сопоставления ранга таксонов, предложенное Бэйкером и Брэдли (Baker, Bradley, 2006), часто вызывает неприятие. На самом деле, оценка генетических дистанций предлагалась ими только как первая и наиболее прагматическая ступень анализа. Как альтернативу генетическим дистанциям эти авторы рассматривают продолжительность видообразования (“We considered duration of speciation for sister species as an alternative to the 1st step value of genetic distance between sister species for application of the Genetic Species Concept”; Baker, Bradley, 2006, p. 658) и полагают, что настоящим показателем видообразования будет оценка времени, связанного

с накоплением генетических изменений, необходимых для формирования изолирующих механизмов.

Таким образом, современная ГКВ оперирует временем, которое необходимо для достижения изоляции. Это очень важный момент, потому что время дивергенции никогда не принимается во внимание биологической концепцией, иногда — морфологической, если имеются ископаемые остатки, иногда — филогенетической, но всегда — ГКВ (Baker, Bradley, 2006).

Вопрос нестандартизованных таксономических рангов не является новым (Hennig, 1966; Avise, Johns, 1999; Avise, Mitchell, 2007). Хенниг (Hennig, 1966) был первым, кто отметил, что абсолютное время эволюции могло бы быть стандартом для синхронизации таксономических рангов. Обзор и обобщение проблем синхронизации линнеевских рангов в далеких группах организмов на основании времени происхождения был выполнен Эвайсом и Лю (Avise, Liu, 2011). В настоящее время линнеевские ранги по-прежнему активно используются, но уже, очевидно, не соответствуют задачам таксономии и не отражают современный уровень знания о филогенетических связях организмов. Их объективные характеристики отсутствуют, они мало соотносимы в разных группах, их невозможно непосредственно наложить на шкалу времени (Avise, Liu, 2011), и потому временные несоответствия пронизывают нынешние классификации. В безранговой филогенетической систематике (филокодексе) система сводится к кодексу номенклатуры видов, для которых установлены филогенетические связи (Queiroz, Gauthier, 1992). Однако иерархическая система классификации представляется большинству систематиков более привычной и логичной, и научное сообщество пока не готово отказаться от нее. Можно ли не отказываться от рангов совсем, но сделать их объективными?

Разумеется, нет смысла напрямую сопоставлять внутривидовые, межвидовые или межродовые генетические и временные дистанции в далеких группах, например, у млекопитающих с такими у рептилий и амфибий. В последней группе они всегда будут выше по той простой причине, что любые таксоны рептилий и амфибий в основном древнее видов млекопитающих. Более того, таксоны рептилий и амфибий не только последовательно старше на каждом таксономическом уровне, чем аналогичные им таксоны птиц и млекопитающих, но похоже, что их эволюционный

возраст увеличивается быстрее с ростом таксономического ранга (Hedges, Kumar, 2003). Проблема согласованности рангов в разных группах на основе времен происхождения этих групп усугубляется тем, что клады каждого таксономического ранга имеют больший или меньший диапазон эволюционных возрастов. Например, эволюционный возраст принятых в современной таксономии семейств плацентарных млекопитающих по молекулярным данным изменяется в среднем от 12 до 60 млн лет (Meredith et al., 2011; Reis et al., 2012), а эволюционный возраст семейств амфибий варьирует от 55 до 180 млн лет (Zhang, Wake, 2009; Kamei et al., 2012; Feng et al., 2017). В диапазоне датировок семейств млекопитающих с некоторой натяжкой попадают времена дивергенции родов *Anura*: 65–20 млн лет назад (Feng et al., 2017).

Виды амфибий расходятся примерно от позднего миоцена до границы плиоцена и плейстоцена, при этом информация о времени дивергенции близких видов в составе родов амфибий для европейских и североамериканских таксонов (в основном плиоценовые и раннеплейстоценовые (Veith et al., 2003; Dufresnes et al., 2018) и позднемиоценовые (Hofman et al., 2007)) отличается от таковой для тропических групп: здесь миоценовая радиация отмечена даже для комплексов близких видов (Yang et al., 2017; Gorin et al., 2020). Возможно, впрочем, это артефакт меньшей изученности тропической фауны. Времена дивергенции близких видов рептилий также приходятся в основном на миоцен–плиоцен (Poyarkov et al., 2014; Schöneberg et al., 2023), иногда доходят до раннего плейстоцена (Salvi et al., 2021).

В любом случае, учитывая все известные сложности синхронизации рангов по времени, в рамках конкретных таксонов все же представляется вполне логичным использовать времена для размышлений о таксономическом ранге и статусе и для проведения границы между видовым и внутривидовым уровнем дифференциации.

## ЦЕЛЬ И ЗАДАЧИ ИССЛЕДОВАНИЯ

Вопрос о том, какая часть исследуемого генома позволит достигнуть желаемого результата, иными словами, какое именно количество генов необходимо для того, чтобы получить генетическую дистанцию, разрешающую отношения

ветвей филогенетического дерева комплекса близкородственных видов, остается открытым.

Учитывая, что разные функционально-структурные области генома эволюционируют с разной скоростью, задача определения уровня дивергенции по ядерным данным связана с задачей определения генетической дистанции по мультилокусным данным и оценкой ее ошибки. Поэтому на первом этапе исследования необходимо подобрать стандартный набор маркеров, который будет служить эталоном генетической дистанции. Наиболее рациональным выбором представляется экзом. Существующие в настоящее время данные в большинстве случаев представлены лишь относительно небольшим числом генов, последовательности которых получены с помощью таргетного секвенирования или из транскриптомов (десятки и сотни). Поэтому для получения сравнимых дистанций необходимо вводить поправочные коэффициенты, учитывающие отношение средней скорости по исследуемой выборке генов к средней скорости экзома в целом. Учитывая, что как отдельные локусы, так и целые геномы испытывают значительные изменения скоростей накопления мутаций в своей эволюции, очевидно, что сравнимость дистанций может быть достигнута только в пределах таксономически компактных групп животных.

Цель нашего исследования состоит в том, чтобы, сохраняя преемственность при переходе от классической систематики к геносистематике, обеспечить практикующих систематиков новым объективным инструментом, который облегчит принятие таксономического решения (вид/подвид). Во-первых, используя уже существующие мультилокусные данные, мы хотим определить уровень различий, характеризующий близкие виды и внутривидовые формы млекопитающих. На этом основании, возможно, удастся сформулировать критерий по ядерным генам, аналогичный предложенному ранее (Bradley, Baker, 2001; Baker, Bradley, 2006), и в конечном итоге разработать мультигенный и временной критерий вида для млекопитающих. Во-вторых, мы пытаемся оценить диапазон изменчивости мультигенной генетической дистанции и определить примерный объем данных, необходимый для получения информационно значимой генетической дистанции с необходимой точностью.

## МАТЕРИАЛ И МЕТОДЫ

### *Эмпирические дистанции между близкими видами*

Для эмпирического определения генетических дистанций между близкими видами (сестринские виды или члены единой видовой группы с неразрешенными взаимоотношениями) исходно было отобрано 188 белок-кодирующих генов длиной от 1000 до 4000 п.н. Для удобства работы использовались гены, кодирующая последовательность которых включала единственный экзон (согласно аннотации генома китайского хомячка *Cricetulus barabensis griseus* GCF\_000223135.1). Предварительно, с использованием ресурсов базы Orthomam (v.12a) (<https://orthomam.mbb.cnrs.fr/>; Allio et al., 2024), проверялось присутствие каждого гена кандидата в геномных последовательностях не менее 170 видов млекопитающих и их аутосомная принадлежность. Из списка видов, последовательности которых доступны в GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>), выбирали пары или группы близких видов и определяли ближайший к ним (референсный) вид, для генома которого существовала аннотация Refseq. Геномные аннотации в формате GTF из базы данных NCBI были использованы для извлечения координат CDS отобранных 188 генов. Сами последовательности CDS были извлечены из соответствующих референсных геномов при помощи программы bedtools getfasta. Поиск ортологов в остальных родственных геномах для каждой пары/группы близких видов проводили при помощи программы BLASTn с параметрами соответствия 95% сходства и 50% длины последовательностей. Наиболее подходящий результат поиска BLAST для каждого из геномов был сведен в множественное выравнивание ортологов определенного гена с помощью программы MAFFT (Katoh, Standley, 2013) с параметром  $-ep$  0.123. Все гены для конкретной группы сравниваемых геномов были сведены в конкатенат при помощи программы seqkit concat. Между конкатенированными последовательностями вычислялась p-дистанция в программе PAUP, version 4.0b10 (Swofford, 2003). В ряде случаев кроме пар близких видов в исследуемую группу включались более далекие виды из того же рода или внутривидовые формы. Список исследованных таксонов приведен в табл. 1.

*Симуляция генетических дистанций  
для разных значений параметров*

Помимо определения эмпирических дистанций, проводили симуляции генетических дистанций для разных значений параметров. Цель симуляций – оценить зависимость ошибки генетической дистанции от объема данных и определить, при каком объеме данных относительная ошибка дистанции будет сопоставима с аналогичной величиной для мтДНК. Предварительно был проведен анализ распределения скоростей локусов среди таксонов млекопитающих на основе данных по ~50 тыс. ядерных маркеров, использованных в работе Хекера и Хиллера (Hecker, Hiller, 2020). Параметры для симуляции включают, с одной стороны, характеристики эволюционного процесса, а с другой – параметры, касающиеся формирования объема данных в процессе исследования.

Важнейший параметр эволюционного процесса – скорость накопления замен, она может быть неодинаковой и отличаться как между генами, так и между линиями (концепция свободных часов). Мы предполагаем, что распределение скоростей разумно моделировать с помощью логнормального распределения со следующими параметрами:

$\mu_{exp}$  – средняя скорость накопления замен;  
 $\sigma(\mu)$  – дисперсия скорости, учитывает различия скоростей как в разных линиях, так и у разных генов;

$Ne$  – эффективная численность; в общем случае их три – для современных таксонов, между которыми считается дистанция, и для их общего предка; от этих величин будут зависеть времена коалесценции аллелей;

$T$  – время дивергенции.

В общем случае мы предполагаем, что сиквенсы соответствуют нефазированным генотипам, в которых гетерозиготные позиции кодируются с помощью стандартных обозначений для выродившихся позиций (ambiguity codes).

Подсчет дистанций производится между отдельными экземплярами, а не выборками.

Ожидаемая дистанция для гена  $k$  будет равна  $2\mu_k T = \mu_k(2T_{div} + 2t_{anc} - t_1 - t_2)$ , где  $T_{div}$  – время дивергенции, а  $t_{anc}$ ,  $t_1$ ,  $t_2$  – времена коалесценции аллелей у предка и современных форм.

Эти параметры моделируются с помощью экспоненциальных распределений с математическими ожиданиями, равными  $2Ne$ . Если уровень анцестрального полиморфизма соответствует

среднему значению для современных форм, то ожидаемая дистанция будет равна  $2\mu_k T_{div}$ . Это упрощение принято в наших симуляциях.

Наконец, последние два параметра характеризуют процесс формирования выборки – это число генов ( $n$ ) и длины их последовательностей ( $L$ ).

Учитывая, что при небольших временах дивергенции количество замен на ген длиной около 1000 п.н. обычно не превосходит несколько нуклеотидов, для моделирования числа замен в случайно выбранном гене использовалось распределение Пуассона с параметром  $L \times \mu_k(2T_{div} + 2t_{anc} - t_1 - t_2)$ , где  $\mu_k$  – случайная величина из логнормального распределения с параметрами  $(\mu_{exp}, \sigma(\mu))$ , а  $t_{anc}$ ,  $t_1$  и  $t_2$  – независимые случайные величины, извлеченные из экспоненциального распределения с математическим ожиданием  $2Ne$ . Процедуру повторяли по числу генов. Дистанция по совокупности генов равна сумме числа замен, деленной на  $nL$ . Симуляции повторяли 10000 раз для каждого набора параметров, включавшего  $\mu_{exp}$ ,  $\sigma(\mu)$ ,  $T_{div}$ ,  $Ne$ ,  $n$  и  $L$ . На основе эмпирических распределений вычисляли коэффициент вариации дистанции как показатель ее относительной ошибки. Эту величину сравнивали с коэффициентом вариации дистанции по митохондриальному гену *cyt b* для сходных уровней дивергенции (детали вычислений приведены в Приложении). На выходе определяли, какой объем данных ( $n$  и  $L$ ) достаточен для получения дистанций с ошибкой не выше, чем для митохондриальных данных. Все симуляции проводились в среде R с использованием встроенных функций `lgnorm`, `gexp`, `grois`.

## РЕЗУЛЬТАТЫ

### *Величины межвидовых и внутривидовых дистанций*

Априори было выделено три категории дистанций: 1) дистанции между видами в пределах рода, по возможности мы не включали туда виды из относительно далеких ветвей ранга подродов; 2) дистанции между конспецифичными формами, включая подвиды, а также домашними и дикими формами; 3) категория полувидов, статус которых по-разному принимается разными авторами (например, *Mus musculus* и *M. domesticus*), или крайне близкие виды, между которыми известны зоны гибридизации (например, косули *Capreolus capreolus* и *C. pygargus*). Учитывая, что скорость нуклеотидных замен в некоторых

группах, по крайней мере в некоторых таксонах, грызунов заметно отличается от таковой у других млекопитающих (Bergeron et al., 2023; наши данные), дистанции между видами Rodentia мы рассматриваем отдельно, выделяя все те же три категории. Результаты приведены в табл. 1 и на рис. 1. Как видим, дистанции

между видами для большинства млекопитающих находятся в диапазоне от 0.15 до 0.75% с медианой 0.33%. Дистанции между полувидами находятся в диапазоне от 0.12 до 0.28% с медианой 0.14%. Внутривидовые дистанции (межпопуляционные или внутривидовые) всегда меньше 0.11%.

**Таблица 1.** Мультигенные межвидовые и внутривидовые дистанции, рассчитанные по 188 ядерным локусам для грызунов, рукокрылых, парнокопытных и китообразных

Таксоны	Дистанции	Таксоны	Дистанции
<b>RODENTIA</b>			
<b>Mus</b>		<b>Acomys</b>	
<i>domesticus / castaneus / musculus</i>	0.00319** 0.00349**	<i>russatus/</i>	0.02166***
<i>domesticus</i>	0.00083+ 0.00115+	<i>percivali/</i>	0.02084***
<i>musculus</i>	0.00081+	<i>kempi/</i>	0.00929***
<i>castaneus</i>	0.0001+	<i>cahirinus / dimidiatus</i>	0.00078
<i>musculus + domesticus / spretus + spicilegus</i>	0.00759***	<b>Cavia</b>	
<i>spretus / spicilegus</i>	0.0069***	<i>aperea/</i>	0.00382***
<i>spretus</i>	6.65E+05+	<i>tschudii / porcellus</i>	0.0005*
<i>caroli/</i>	0.01704***	<b>Marmota</b>	
<i>pahari/</i>	0.03218	<i>marmota / himalayana</i>	0.00257***
<i>minutoides/</i>	0.03198	<i>monax/</i>	0.00396***
<b>Peromyscus</b>		<i>vancouverensis / flaviventris</i>	0.0037***
<i>maniculatus / polionotus</i>	0.00621***	<i>vancouverensis + flaviventris / marmot + himalayana + monax</i>	0.0043***
<i>maniculatus</i>	0.00037+	<b>Rhodomys</b>	
<i>nudipes / melanophrys</i>	0.00974***	<i>pumilio / dilectus</i>	0.0051***
<i>eremicus / californicus</i>	0.01259***	<b>Mastomys</b>	
<i>eremicus1 / eremicus2</i>	0.00567	<i>natalensis / coucha</i>	0.00583***
<i>nudipes + melanophrys / eremicus + californicus</i>	0.0195***		
<i>maniculatus + polionotus / eremicus + californicus + nudipes + melanophrys</i>	0.0233***		
<b>CHIROPTERA</b>			
<b>Rousettus</b>		<b>Pipistrellus</b>	
<i>aegyptiacus / madagascariensis</i>	0.0033***	<i>pygmaeus / pipistrellus</i>	0.0074***
<i>leshenaultii / aegyptiacus + madagascariensis</i>	0.00485***	<i>kuhlii / pygmaeus + pipistrellus</i>	0.0161***
<b>Pteropus</b>		<b>Hipposideros</b>	
<i>giganteus / vampyrus</i>	0.00283***	<i>armiger / pendlebury</i>	0.0037***
<i>pselaphon/</i>	0.00331***	<i>larvatus / pendlebury + armiger</i>	0.0065***
<i>alecto/</i>	0.00602***		
<b>CARNIVORA</b>			
<b>Mustela</b>		<b>Canis</b>	
<i>eversmannii / nigripes</i>	0.00118**	<i>dingo / familiaris</i>	0.0007*
<i>putorius / furo</i>	0.0004*	<i>lupus / familiaris+</i>	0.00095*
<i>eversmannii + nigripes / putorius</i>	0.0022***	<b>Neofelis</b>	
<i>lutreola/</i>	0.00348***	<i>diardi / nebulosa</i>	0.0022***

Таблица 1. Окончание

Таксоны	Дистанции	Таксоны	Дистанции
<b>Martes</b>		<b>Ursus</b>	
<i>martes / zibellina</i>	0.0026***	<i>maritimus / arctos</i>	0.0016***
<b>Felis</b>		<i>americanus / maritimus + arctos</i>	0.00265***
<i>catus / nigripes</i>	0.0035***	<b>Prionailurus</b>	
<i>chaus / catus + nigripes</i>	0.004***	<i>bengalensis / iriomotensis</i>	0.00094*
<b>Lynx</b>		<i>viverrinus / bengalensis+</i>	0.00254***
<i>canadensis / pardinus</i>	0.0013**	<b>Mirounga</b>	
<i>rufus / canadensis + pardinus</i>	0.00285***	<i>leonina / angustirostris</i>	0.0017***
CETARTIODACTYLA			
<b>Bubalus</b>		<b>Moschus</b>	
<i>bubalis / carabanensis</i>	0.0021**	<i>chryzogaster / berezovskii</i>	0.0017***
<i>depressicornis / bubalis</i>	0.0027***	<i>moschiferus / chryzogaster + berez</i>	0.0044***
<i>depressicornis / carabanensis</i>	0.0023***	<b>Ovis</b>	
<b>Capra</b>		<i>nivicola / canadensis</i>	0.00134**
<i>aegagrus / hircus</i>	0.0011	<i>aries / orientalis</i>	0.00155**
<i>aegagrus + hircus / sibirica</i>	0.0035***	<i>ammon / aries + ori</i>	0.00276**
<b>Cervus</b>		<i>pachyceros / ovis</i>	0.00357***
<i>canadensis / hanglu</i>	0.0013**	<b>Sus</b>	
<i>elaphus / canadensis + hanglu</i>	0.0016**	<i>scrofa / cebifrons</i>	0.0042***
<i>albirostris/</i>	0.0019***	<b>Camelus</b>	
<b>Capreolus</b>		<i>bactrianus / ferus</i>	0.0009*
<i>capreolus / pygargus</i>	0.0037***	<i>dromedarius / bactrianus+</i>	0.00245***
<b>Tursiops</b>		<b>Balaenoptera</b>	
<i>truncatus / aduncus</i>	0.002***	<i>acutorostrata / bonariensis</i>	0.0034***

Примечание. \*\*\*дистанции между видами; \*\*между полувидами; \*между внутривидовыми группами; + индивидуальные дистанции внутри группировок.

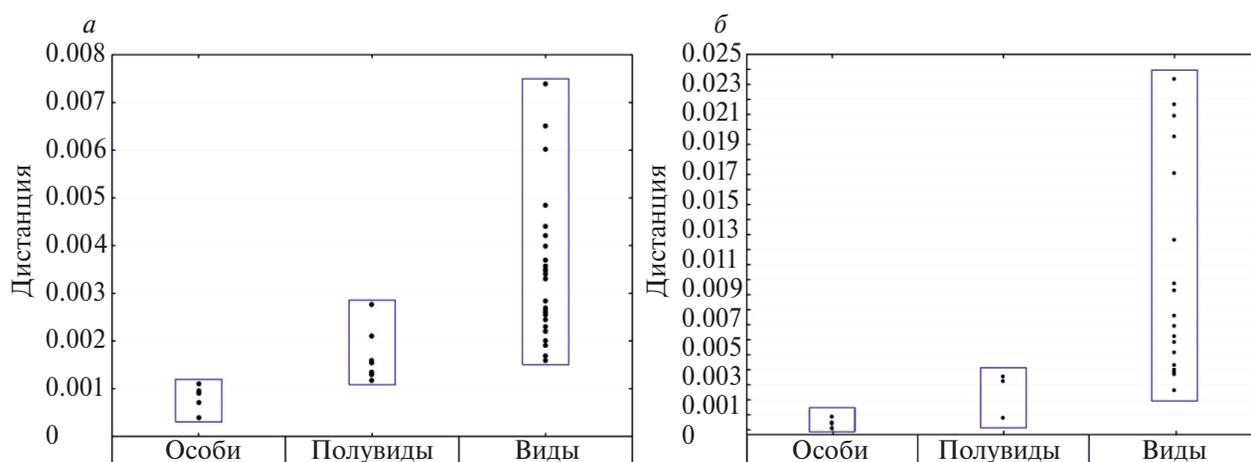


Рис. 1. Диапазон межвидовой и внутривидовой изменчивости: а – для разных отрядов млекопитающих, б – для отряда Rodentia.

У грызунов можно наблюдать более высокие значения межвидовых дистанций — от 0.25 до 2.3% с медианой 0.72%. Дистанции между полувидами находятся в диапазоне от 0.1 до 0.35%. Внутривидовые дистанции у грызунов меньше 0.12%.

*Определение необходимого объема выборки в зависимости от параметров модели*

Результаты симуляции эволюции *сут b* показали, что относительная ошибка (CV) для величин дистанций, характерных для близких видов или полувидов (4–6%), равна 0.11–0.14. Соответственно, нам требуется определить, при каких значениях параметров и объеме выборки ядерных генов будет достигаться пороговое значение, равное 0.11.

Симуляции показали, что относительная ошибка уменьшается с ростом величины самой дистанции и, соответственно, с ростом времени дивергенции (табл. 2, А). Пороговое значение CV при длине гена 300 п.н. для времени дивергенции в 500 тыс. лет (дистанция = 0.003) достигается при выборке в 100 генов, но требуется больше 200 генов для времени дивергенции в 200 тыс. лет (дистанция = 0.00125). Если длина гена увеличивается до 1000 п.н., то пороговое

значение для времени дивергенции 0.5 млн лет достигается уже при 50 генах (табл. 2, Б).

Вклад увеличения дисперсии скоростей генов (табл. 2, В) заключается в том, что увеличение  $\sigma(\mu)$  с  $1e-9$  до  $3e-9$  приводит к необходимости увеличения выборки с 100 до 200 генов.

При увеличении  $N_e$  объем выборки, необходимый для достижения порогового значения CV, существенно возрастает при величинах  $N_e$ , сопоставимых со временем дивергенции (табл. 2, Г).

*Примеры генетических различий между полувидами и дикими и домашними формами у млекопитающих*

Хотя в Кодексе зоологической номенклатуры ранг полувида не принят, тем не менее этот термин обычен и широко употребляется в литературе (Mallet, 2007), наряду с близким по значению понятием “*species in statu nascendi*” (Singh, 2021) для обозначения одного из начальных этапов формирования вида, когда ему не может еще быть присвоен таксономический ранг, однако мы наблюдаем процессы, которые могут привести к его полному становлению.

В качестве классического примера полувида можно привести надвидовой комплекс *Mus musculus / M. domesticus*. Эти географические

**Таблица 2.** Зависимость коэффициента вариации (CV) от параметров модели

Число генов	Параметры модели														
	А время дивергенции (тыс. лет)				Б длина гена (п.н.)			В $\sigma(\mu)$ – дисперсия скорости			Г $N_e$ – эффективная численность				
	200	300	500	800	300	600	1000	$1e-9$	$2e-9$	$3e-9$	5000	10000	50000	100000	300000
	Неварьирующие переменные														
	$\sigma(\mu) = 1e-9$ $\mu = 3e-9$ $N_e = 5e4$ длина гена = 300 п.н.				$\sigma(\mu) = 1e-9$ $\mu = 3e-9$ $N_e = 5e4$ $t = 500$ тыс. лет			$\mu = 3e-9$ $N_e = 5e4$ длина гена = 300 п.н.			$\sigma(\mu) = 1e-9$ $\mu = 3e-9$ $t = 500$ тыс. лет длина гена = 300 п.н.				
5	0.85	0.65	0.50	0.41	0.50	0.38	0.32	0.50	0.58	0.67	0.49	0.50	0.50	0.53	0.69
10	0.60	0.46	0.36	0.29	0.36	0.27	0.23	0.36	0.41	0.47	0.35	0.35	0.36	0.39	0.49
20	0.42	0.33	0.26	0.21	0.26	0.19	0.16	0.26	0.28	0.33	0.25	0.25	0.26	0.27	0.34
30	0.35	0.27	0.21	0.17	0.21	0.16	0.13	0.21	0.23	0.27	0.20	0.20	0.21	0.22	0.28
50	0.27	0.21	0.16	0.13	0.16	0.12	0.10	0.16	0.18	0.21	0.16	0.16	0.16	0.17	0.22
100	0.19	0.15	0.11	0.09	0.11	0.09	0.07	0.11	0.13	0.15	0.11	0.11	0.11	0.12	0.16
200	0.13	0.10	0.08	0.06	0.08	0.06	0.05	0.08	0.09	0.11	0.08	0.08	0.08	0.08	0.11
500	0.08	0.07	0.05	0.04				0.05	0.06	0.07					
1000	0.06	0.05	0.04	0.03				0.04	0.04	0.05					

формы с контактной зоной в Центральной Европе разными авторами рассматриваются как близкие виды или подвиды политипического вида *M. musculus*, при этом поток генов в гибридной зоне ограничен и асимметричен (Good et al., 2008; Baird, Macholan, 2012). Дистанция между ними по ядерным генам составляет 0.32%. Примерно тот же уровень различий характеризует *M. musculus* и *M. castaneus* (0.35%). Внутривидовые дистанции для *M. musculus* и *M. domesticus* составляют 0.08%, для *M. castaneus* — 0.01%. Дистанции от них до других близких видов, ранее тоже относимых к *M. musculus* s.l., существенно выше. *M. caroli* из Юго-Восточной Азии — это относительно далекий вид, однако принадлежащий к тому же подроду. *M. pahari* и *M. minotoides* относятся к другим подкладам и дистанции от них до *M. musculus* / *M. domesticus* значительно выше (~3%).

Другой пример близких викарирующих форм — хорьки р. *Mustela*. Степной и лесной хори, *M. eversmannii* и *M. putorius*, распространены парапатрично и скрещиваются в природе. Дистанция между ними 0.22%. К лесному хорю много ближе домашняя форма *M. furo* (0.04%). Ближайшая сестринская группа степного хоря — черноногий хорь *M. nigripes* (0.12%). Дистанция между хорями и относительно близкой к ним европейской норкой — 0.35%.

Дистанция между безусловно сестринскими видами бурый (*Ursus arctos*) и белым медведем (*U. maritimus*) — 0.16%. Дистанция до ближайшего к ним вида, *U. americanus*, 0.27%.

Ириомотейская кошка *Prionailurus iriomotensis* считается островным подвидом бенгальской кошки *P. bengalensis*, хотя исходно была описана как отдельный вид. По нашим данным, дистанция между ними 0.094%, что заметно меньше дистанции между *P. bengalensis* и *P. viverrinus* — 0.25%.

Некоторое время назад все разнообразие благородных оленей Голарктики по мтДНК было сведено к трем группам: европейский благородный олень *Cervus elaphus*, южно-азиатская форма *C. hanglu* и канадский олень *C. canadensis*, объединяющий оленей Сибири и Северной Америки (Mackiewicz et al., 2022). Данные по ядерным геномам подтверждают родство этих трех форм при диапазоне дистанций между ними 0.13–0.19%. К ним близок тибетский беломордый олень *C. albirostris* (0.19%), что тоже согласуется с митохондриальными данными.

Систематика рода горных баранов (*Ovis*) всегда трактовалась неоднозначно. Разные

исследователи включали сюда от двух до семи видов (Grubb, 2005). Данные по ядерным генам согласуются с делением на три группы. Азиатский муфлон *O. orientalis* оказался относительно близок к домашней овце *O. aries* (0.16%), чье происхождение связывают с подвидом *O. aries musimon*. С ними объединяется архар *O. ammon* с дистанцией 0.28%. Азиатские и американские снежные бараны очень близки (0.14%).

Среди диких и домашних форм млекопитающих наиболее показательны следующие примеры. Волк и домашняя собака в настоящее время считаются конспецифичными формами (Bergström et al., 2022). Дистанция между волком, домашней собакой и динго составляет 0.095%.

Дикий и домашний двугорбый верблюд, *Camelus ferus* и *C. bactrianus*, по митохондриальным данным относятся к разным, хотя и родственным,кладам. Геномные данные указывают на их близость при дистанции (0.09%). Дистанция между двугорбым и одногорбым верблюдами — 0.25%.

Между *Bos taurus* и *B. indicus* — родоначальниками домашнего рогатого скота — дистанция по ядерным генам составляет 0.15%. Несколько большая дистанция (0.2%) между речным и болотным буйволами, *Bubalus bubalis* и *B. carabanensis*, которые также участвовали в формировании домашнего скота. Несколько дальше от них целебеский буйвол *B. depressicornis* (0.23%).

Лабораторная морская свинка *Cavia porcellus* считается производной от дикого вида *C. tschudii*. Это полностью соответствует дистанциям по ядерным генам, которые между этими двумя формами очень малы (0.05%). Относительно далек от них другой дикий вид *C. aperea* (~0.4%).

## ОБСУЖДЕНИЕ

Ранее на примере видовых комплексов *Sorex araneus* (Rasporova et al., 2022), *Crocidura suaveolens* (Грицышин и др., 2023), *Dipus sagitta* (Лисенкова и др., 2023), а также некоторых видовых комплексов р. *Allactaga* (Lebedev et al., 2022) нами было показано, что разрешающей способности набора из 10–20 ядерных локусов недостаточно для изучения диверсификации филогенетических линий в сложных видовых комплексах и группах близкородственных видов. Для удовлетворительного анализа филогенетических отношений и событий гибридизации в пределах этих видовых групп необходим

переход к методикам глубокого секвенирования нового поколения.

### *Почему экзом?*

Довольно часто исследуют характеристики генома как целого, в частности такой показатель, как скорость мутирования на поколение или год по геному в среднем. Кроме того, довольно часто, рассматривая белок-кодирующие гены, исследователи фокусируются на интронных последовательностях, мотивируя свой выбор их большей скоростью эволюции, что приводит к большей информативности данных. Однако в качестве стандартного набора маркеров, с нашей точки зрения, больше подходит экзом, или как минимум большая выборка кодирующих последовательностей. Аргументом в пользу такого выбора может быть тот факт, что в этом случае проще установить гомологию. С другой стороны, эволюция интронов часто осложнена такими событиями, как инсерции диспергированных повторов и индели в tandemных повторах (Fedorova, Fedorov, 2003; Zhang, Hewitt, 2003), что может затруднять оценку уровня дивергенции и искажать результат. Кроме того, скорость изменения экзонов по сравнению с интронами ниже в 1.5–2 раза (Zhang, Hewitt, 2003), но проблем с выравниванием экзонных последовательностей значительно меньше.

### *Проблемы определения генетической дистанции и возможности их преодоления*

Результаты нашего исследования показывают, что у млекопитающих разница между видом и внутривидовой формой на уровне генома составляет 0.1–0.15%. Эти данные являются предварительными и подлежат уточнению после изучения большего числа групп, содержащих полувиды. Среди грызунов дистанции в видовом комплексе домовый мыши (см. ниже) указывают на больший уровень различий, чем в случае полувидов из других отрядов. В сочетании с более высоким уровнем различий между “хорошими” видами у грызунов это говорит в пользу предположения, что единого порога для всех млекопитающих не существует и его величина зависит, возможно, от скорости мутирования в тех или иных группах. Например, у грызунов скорость мутирования на единицу времени выше, чем у других млекопитающих (Bergeron et al., 2023).

Кроме того, следует проверить предположение, что разница между внутри- и межвидовой изменчивостью лучше видна не по генетическим

дистанциям, а по временам дивергенций. Если это окажется так, то будет подтверждена гипотеза, что процесс видообразования не напрямую связан с количеством различий по первичной структуре ДНК, а зависит от какого-то другого процесса, протекающего с более или менее постоянной скоростью относительно физического времени. С другой стороны, такой вывод противоречил бы существующим данным, которые указывают на положительную корреляцию между скоростью видообразования и скоростью молекулярной эволюции (Webster et al., 2003). Этот сложный вопрос требует дальнейших исследований.

Тем не менее, даже если ранг таксона лучше соответствует времени, чем дистанции, оценка уровня генетических различий является необходимой частью анализа. Соответственно, вопрос о необходимой точности оценки дистанции остается актуальным. Принимая величину коэффициента вариации в 0.1–0.15 за минимальный уровень относительной ошибки, мы приходим к выводу, что для получения надежной оценки дистанции по ядерным генам необходимый объем данных составляет 50–100 генов длиной примерно в 1000 п.н. или для генов длиной 300 п.н. требуется 100–200 генов. Необходимость большей длины совокупного выравнивания есть следствие существенно более медленной эволюции ядерных генов. Однако при высоких значениях анцестральной эффективной численности ( $N_e = 1e5$ ) существенным оказывается вклад ошибки, связанной с числом генов.

Тем самым результаты исследования показывают, что выборки ядерных генов числом менее 10 абсолютно недостаточны для проверки гипотезы по уровню дивергенции ядерного генома, потому что относительная ошибка очень велика (25–75%). В частности, невозможно проверить гипотезу о том, что уровень различия по митохондриальному геному соответствует уровню различий по ядерному геному, и, следовательно, отвергнуть гипотезу о заимствовании мтДНК. Этот вывод ставит предел возможностям сэнгеровского секвенирования в связи с временными и финансовыми затратами.

Итак, наши результаты по ядерным генетическим дистанциям в целом соответствуют ожиданиям в том, что дистанции между полувидами меньше дистанций между “хорошими” видами в тех же группах, при этом внутривидовые дистанции еще меньше. Основываясь на представленных данных, можно сделать следующее

закключение: порог межвидовой/внутривидовой генетической дистанции для млекопитающих, рассчитываемой по экзонам, которые наиболее часто используются в молекулярной филогенетике этой группы, составляет около 0.15%. Данная гипотеза предполагает признание за видами *in statu nascendi* видового ранга. Следует заметить, однако, что не все “хорошие” виды имеют значительные генетические дистанции в связи с тем, что репродуктивные барьеры иногда формируются быстрее, чем накапливаются замены по геному.

#### Потенциальные источники ошибок

Проведенная нами работа с последовательностями, получаемыми из депонированных в GenBank геномов, может содержать в себе потенциальные ошибки. Во-первых, при небольшом числе экзонов оценка дистанции может быть искажена давлением отбора на динамику появления мутаций в их последовательностях, однако мы предполагаем, что достаточно большая выборка локусов позволяет нивелировать это влияние. Во-вторых, геномная последовательность, представленная в базах данных, псевдогаплоидна, так как обычно не учитывается, что организм реально диплоидный. Кроме того, в самих геномных данных довольно часто встречаются ошибки. Оба последних фактора приводят к тому, что генетические дистанции между близкими геномами могут оказаться завышенными. Чтобы избавиться от искажения результата, желательно, вычисляя дистанцию, вводить поправку на эти источники ошибок, рассчитывая *net*-дистанцию с учетом дистанции между геномами внутри одного таксона. В этом случае мы можем оценить внутривидовую дистанцию между геномами. Однако, к сожалению, эти показатели в настоящее время практически недоступны для немодельных объектов.

#### БЛАГОДАРНОСТИ

Приносим благодарность В.А. Горину за помощь в обзоре литературы по временам дивергенций амфибий и рептилий.

**Таблица П1.** Некорректированные (*p*-)дистанции между симулированными последовательностями гена *cyt b*

Истинная дистанция	0.15	0.10	0.07	0.049	0.0343
Средняя <i>p</i> -дистанция	0.0914	0.0724	0.0551	0.0413	0.0303
CV	0.0888	0.1026	0.118	0.140	0.1628

#### ИСТОЧНИКИ ФИНАНСИРОВАНИЯ

Работа поддержана Российским научным фондом (грант РНФ № 21-14-00007).

#### КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют, что у них нет конфликта интересов.

#### СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Настоящая статья не содержит экспериментальных исследований с использованием животных в качестве объектов.

#### ПРИЛОЖЕНИЕ

В случае быстро эволюционирующих митохондриальных последовательностей использование для симуляции распределения Пуассона не подходит. Вместо этого мы симулировали последовательности для пары таксонов на основании реалистичных эволюционных моделей в программе Indelible v.1.03 (Fletcher, Yang, 2009). Истинная эволюционная дистанция между таксонами принимала одно из пяти значений (0.15–0.10–0.07–0.049–0.0343), для каждого варианта проводили 10000 симуляций. Три позиции кодонов имели каждая свою модель (соответственно, TrN + I + G, HKY + I + G, TrN + G), скорости позиций соотносились как 0.045 : 0.01 : 0.245. Параметры моделей приведены в контрольном файле ниже, вычислялись в PAUP\* 4.0b10 (Swofford, 2003). Результаты отражены в табл. П1.

Контрольный файл (часть):

```
[MODEL] m1
[submodel] TrN30.0 15.0
[statefreq] 0.25 0.25 0.25 0.25
[rates] 0.5 0.5 4
[MODEL] m2
[submodel] HKY15.0
[statefreq] 0.4 0.25 0.20 0.15
[rates] 0.8 1.0 4
[MODEL] m3
[submodel] TrN12.0 36.0
```

[statefreq] 0.20 0.40 0.35 0.05  
 [rates] 0.0 2.5 4  
 [TREE] t1 (A:1.0, B:1.0);  
 [treelength] 0.06300  
 [TREE] t2 (A:1.0, B:1.0);  
 [treelength] 0.01400  
 [TREE] t3 (A:1.0, B:1.0);  
 [treelength] 0.34300  
 [PARTITIONS] p0  
 [t1 m1 380]  
 [t2 m2 380]  
 [t3 m3 380]

## СПИСОК ЛИТЕРАТУРЫ

- Грицышин В.А., Лисенкова А.А., Сперанская А.С., Артюшин И.В., Шефтель Б.И. и др., 2023. Мультилокусный анализ филогенетических отношений в видовом комплексе *Crociodura suaveolens sensu lato*: сравнение с митохондриальными данными // Докл. РАН. Т. 509. № 3. С. 147–154.  
<https://doi.org/10.1134/S0012496623700308>
- Лисенкова А.А., Лебедев В.С., Ундрахбаяр Э., Богатырева В.Ю., Мельникова М.Н. и др., 2023. Филогения видового комплекса *Dipus sagitta* по результатам секвенирования ядерных генов // Докл. РАН. Т. 509. С. 155–160.  
<https://doi.org/10.1134/S0012496623700230>
- Allio R., Delsuc F., Belkhir K., Douzery E.J., Ranwez V., Scornavacca C., 2024. OrthoMaM v12: A database of curated single-copy ortholog alignments and trees to study mammalian evolutionary genomics // Nucleic Acids Res. V. 52. № D1. P. D529–D535.
- Avise J.C., Ball R.M., 1990. Principles of genealogical concordance in species concepts and biological taxonomy // Oxford Surveys in Evolutionary Biology. V. 7 / Eds Futuyma D.J., Antonovics J. Oxford: Oxford Univ. Press. P. 45–67.
- Avise J.C., Johns G.C., 1999. Proposal for a standardized temporal scheme of biological classification for extant species // Proc. Natl. Acad. Sci. USA. V. 96. P. 7358–7363.
- Avise J.C., Liu J.X., 2011. On the temporal inconsistencies of Linnean taxonomic ranks // Biol. J. Linn. Soc. V. 102. P. 707–714.  
<https://doi.org/10.1111/j.1095-8312.2011.01624.x>
- Avise J.C., Mitchell D., 2007. Time to standardize taxonomies // Syst. Biol. V. 56. № 1. P. 130–133.  
<https://doi.org/10.1080/10635150601145365>
- Baird S.J., Macholan M., 2012. What can the *Mus musculus musculus*/*M. m. domesticus* hybrid zone tell us about speciation? // Turkish J. Zool. V. 30. P. 459–464.
- Baker R.J., Bickham J.W., 1986. Speciation by monobrachial centric fusions // Proc. Natl. Acad. Sci. V. 83. P. 8245–8248.  
<https://doi.org/10.1073/pnas.83.21.8245>
- Baker R.J., Bradley R.D., 2006. Speciation in mammals and the genetic species concept // J. Mammal. V. 87. № 4. P. 643–662.  
<https://doi.org/10.1644/06-MAMM-F-038R2.1>
- Bateson W., 1909. Heredity and variation in modern lights // Darwin and Modern Science / Ed. Seward A.C. Cambridge: Cambridge Univ. Press. P. 85–101.
- Bergeron L.A., Besenbacher S., Zheng J., Li P., Bertelsen M.F., et al., 2023. Evolution of the germline mutation rate across vertebrates // Nature. V. 615. P. 285–290.  
<https://doi.org/10.1038/s41586-023-05752-y>
- Bergström A., Stanton D.W., Taron U.H., Frantz L., Sinding M.H.S., et al., 2022. Grey wolf genomic history reveals a dual ancestry of dogs // Nature. V. 607. № 7918. P. 313–320.  
<https://doi.org/10.1038/s41586-022-04824-9>
- Bradley B.D., Baker R.J., 2001. A test of the genetic species concept: cytochrome-*b* sequences and mammals // J. Mammal. V. 82. № 4. P. 960–973.  
[https://doi.org/10.1644/1545-1542\(2001\)082<0960:ATOTGS>2.0.CO;2](https://doi.org/10.1644/1545-1542(2001)082<0960:ATOTGS>2.0.CO;2)
- Butlin R.K., 2005. Recombination and speciation // Mol. Ecol. V. 14. № 9. P. 2621–2635.
- Coyne J.A., Orr H.A., 2004. Speciation. Sunderland: Sinauer Associates. 545 p.
- Dobzhansky T., 1937. Genetics and the Origin of Species. N.-Y.: Columbia Univ. Press. 364 p.
- Dobzhansky T., 1950. Mendelian populations and their evolution // Am. Nat. V. 84. № 819. P. 401–418.
- Dufresnes C., Lymberakis P., Kornilios P., Savary R., Perin N., Stöck M., 2018. Phylogeography of Aegean green toads (*Bufo viridis* subgroup): continental hybrid swarm vs. insular diversification with discovery of a new island endemic // BMC Evol. Biol. V. 18. P. 1–12.  
<https://doi.org/10.1186/s12862-018-1179-0>
- Fedorova L., Fedorov A., 2003. Introns in gene evolution // Origin and Evolution of New Gene Functions / Ed. Long M. Dordrecht: Springer. 199 p.
- Feng Y.-J., Blackburn D.C., Liang D., Hillis D.M., Wake D.B., et al., 2017. Phylogenomics reveals rapid, simultaneous diversification of three major clades of Gondwanan frogs at the Cretaceous–Paleogene boundary // Proc. Natl. Acad. Sci. V. 114. № 29. P. E5864–E5870.  
<https://doi.org/10.1073/pnas.1704632114>
- Fletcher W., Yang Z., 2009. INDELible: A flexible simulator of biological sequence evolution // Mol. Biol. Evol. V. 26. № 8. P. 1879–1888.  
<https://doi.org/10.1093/molbev/msp098>
- Gavrilets S., 2003. Perspective: Models of speciation: What have we learned in 40 years? // Evolution. V. 57. № 10. P. 2197–2215.  
<https://doi.org/10.1111/j.0014-3820.2003.tb00233.x>
- Good J.M., Dean M.D., Nachma M.W., 2008. A complex genetic basis to X-linked hybrid male sterility

- between two species of house mice // *Genetics*. V. 179. P. 2213–2228.  
<https://doi.org/10.1534/genetics.107.085340>
- Gorin V.A., Solovyeva E.N., Hasan M., Okamiya H., Karunathna D.M.S.S., et al., 2020. A little frog leaps a long way: compounded colonizations of the Indian Subcontinent discovered in the tiny Oriental frog genus *Microhyla* (Amphibia: Microhylidae) // *Peer J*. V. 8. Art. e9411.  
<https://doi.org/10.7717/peerj.9411>
- Grubb P., 2005. Order Artiodactyla // *Mammal Species of the World: A Taxonomic and Geographic Reference*. 3rd ed. / Eds Wilson D.E., Reeder D.M. Maryland: Johns Hopkins Univ. Press. P. 637–722.
- Hecker N., Hiller M., 2020. A genome alignment of 120 mammals highlights ultraconserved element variability and placenta-associated enhancers // *Gigascience*. V. 9. № 1. Art. giz159.  
<https://doi.org/10.1093/gigascience/giz159>
- Hedges S.B., Kumar S., 2003. Genomic clocks and evolutionary timescales // *Trends Genet.* V. 19. № 4. P. 200–206.  
[https://doi.org/10.1016/S0168-9525\(03\)00053-2](https://doi.org/10.1016/S0168-9525(03)00053-2)
- Hennig W., 1966. *Phylogenetic Systematic*. Urbana: Univ. of Illinois Press. 284 p.
- Hofman S., Spolsky C., Uzzell T., Cogălniceanu D., Babik W., Szymura J.M., 2007. Phylogeography of the fire-bellied toads *Bombina*: independent Pleistocene histories inferred from mitochondrial genomes // *Mol. Ecol.* V. 16. № 11. P. 2301–2316.  
<https://doi.org/10.1111/j.1365-294X.2007.03309.x>
- Kamei R.G., Mauro D.S., Gower D.J., Bocslaer I., van Sherratt E., et al., 2012. Discovery of a new family of amphibians from northeast India with ancient links to Africa // *Proc. Roy. Soc. Lond. B. Biol. Sci.* V. 279. № 1737. P. 2396–2401.  
<https://doi.org/10.1098/rspb.2012.0150>
- Katoh K., Standley D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability // *Mol. Biol. Evol.* V. 30. № 4. P. 772–780.  
<https://doi.org/10.1093/molbev/mst010>
- Lebedev V.S., Shenbrot G.I., Krystufek B., Mahmoudi A., Melnikova M.N., et al., 2022. Phylogenetic relations and range history of jerboas of the Allactaginae subfamily (Dipodidae, Rodentia) // *Sci. Rep.* V. 12. № 1. Art. 842.  
<https://doi.org/10.1038/s41598-022-04779-x>
- Mackiewicz P., Matosiuk M., Świsłocka M., Zachos F.E., Hajji G.M., et al., 2022. Phylogeny and evolution of the genus *Cervus* (Cervidae, Mammalia) as revealed by complete mitochondrial genomes // *Sci. Rep.* V. 12. Art. 16381.  
<https://doi.org/10.1038/s41598-022-20763-x>
- Mallet J., 2007. Subspecies, semispecies, superspecies // *Encyclop. Biodivers.* V. 5. P. 523–526.
- Masters J.C., Spencer H.G., 1989. Why we need a new Genetic Species Concept // *Syst. Biol.* V. 38. № 3. P. 270–279.  
<https://doi.org/10.2307/2992287>
- Mayden R.L., 1997. A hierarchy of species concepts: The denouement in the saga of the species problem // *Species: the Units of Biodiversity* / Eds Claridge M.F., Dawah H.A., Wilson M.R. L.: Chapman & Hall. P. 381–424.
- Mayr E., 1942. *Systematics and the Origin of Species*. N.-Y.: Columbia Univ. Press. 334 p.
- Mayr E., 1969. The biological meaning of species // *Biol. J. Linn. Soc.* V. 1. № 3. P. 311–320.
- Meredith R.W., Janecka J.E., Gatesy J., Ryder O.A., Fisher C.A., et al., 2011. Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification // *Science*. V. 334. № 6055. P. 521–524.  
<https://doi.org/10.1126/science.1211028>
- Muller H.J., 1939. Reversibility in evolution considered from the standpoint of genetics // *Biol. Rev.* V. 14. № 3. P. 261–280.
- Nei M., 1976. Mathematical models of speciation and genetic distance // *Population Genetics and Ecology* / Eds Karlin S., Nevo E. N.-Y.: Academic Press. P. 723–765.
- Poyarkov N.A., Orlova V.F., Chirikova M.A., 2014. The mitochondrial phylogeography and intraspecific taxonomy of the Steppe Racerunner, *Eremias arguta* (Pallas) (Lacertidae: Sauria, Reptilia), reflects biogeographic patterns in Middle Asia // *Zootaxa*. V. 3895. № 2. P. 208–224.
- Queiroz K., de Gauthier J., 1992. Phylogenetic taxonomy // *Annu. Rev. Ecol. Syst.* V. 23. P. 449–480.
- Raspopova A.A., Lebedev V.S., Searle J.B., Bannikova A.A., 2023. Discordant phylogenies in the *Sorex araneus* group (Soricidae, Mammalia): Footprints of past reticulations? // *Zool. Scr.* V. 52. № 4. P. 331–344.  
<https://doi.org/10.1111/zsc.12590>
- Reis M., dos Inoue J., Hasegawa M., Asher R.J., Donoghue P.C., Yang Z., 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny // *Proc. Roy. Soc. Lond. Ser. B. Biol. Sci.* V. 279. № 1742. P. 3491–3500.  
<https://doi.org/10.1098/rspb.2012.0683>
- Salvi D., Pinho C., Mendes J., Harris D.J., 2021. Fossil-calibrated time tree of *Podarcis* wall lizards provides limited support for biogeographic calibration models // *Mol. Phyl. Evol.* V. 161. Art. 107169.  
<https://doi.org/10.1016/j.ympev.2021.107169>
- Schilthuizen M., 2000. Dualism and conflicts in understanding speciation // *BioEssays*. V. 22. № 12. P. 1135–1141.
- Schöneberg Y., Winter S., Arribas O., Nicola M.R., Master M., et al., 2023. Genomics reveals broad hybridization in deeply divergent Palearctic grass and

- water snakes (*Natrix* spp.) // *Mol. Phyl. Evol.* V. 184. Art. 107787.  
<https://doi.org/10.1016/j.ympev.2023.107787>
- Simpson G.G., 1943. Criteria for genera, species, and subspecies in zoology and paleontology // *Ann. N.Y. Acad. Sci.* V. 44. № 2. P. 145–178.
- Singh B.N., 2021. Taxonomic status of *Drosophila pallidosa*: A species *in statu nascendi* // *J. Sci. Res.* V. 65. № 1. P. 85–87.
- Swofford D.L., 2003. PAUP\*. Phylogenetic analysis using parsimony (\*and other methods). v. 4.0b10. Sunderland: Sinauer Associates.
- Taverner P.A., 1920. A test of the subspecies // *J. Mammal.* V. 1. № 3. P. 124–127.
- Veith M., Kosuch J., Vences M., 2003. Climatic oscillations triggered post-Messinian speciation of Western Palearctic brown frogs (Amphibia, Ranidae) // *Mol. Phyl. Evol.* V. 26. № 2. P. 310–327.  
[https://doi.org/10.1016/S1055-7903\(02\)00324-X](https://doi.org/10.1016/S1055-7903(02)00324-X)
- Webster A.J., Payne R.J.H., Pagel M., 2003. Molecular phylogenies link rates of evolution and speciation // *Science.* V. 301. № 5632. P. 478.  
<https://doi.org/10.1126/science.1083202>
- Yang B.T., Zhou Yu., Min M.S., Matsui M., Dong B.J., et al., 2017. Diversity and phylogeography of Northeast Asian brown frogs allied to *Rana dybowskii* (Anura, Ranidae) // *Mol. Phyl. Evol.* V. 112. P. 148–157.  
<https://doi.org/10.1016/j.ympev.2017.04.026>
- Zhang D.X., Hewitt G.M., 2003. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects // *Mol. Ecol.* V. 12. № 3. P. 563–584.  
<https://doi.org/10.1046/j.1365-294X.2003.01773.x>
- Zhang P., Wake D.B., 2009. Higher-level salamander relationships and divergence dates inferred from complete mitochondrial genomes // *Mol. Phyl. Evol.* V. 53. № 2. P. 492–508.  
<https://doi.org/10.1016/j.ympev.2009.07.010>

## Genetic species concept as applied to multilocus nuclear data

V. S. Lebedev<sup>a, \*</sup>, S. A. Kosushkin<sup>b, \*\*</sup>, A. A. Bannikova<sup>c, \*\*\*</sup>

<sup>a</sup>Moscow State University, Zoological Museum  
 Bolshaya Nikitskaya, 2, Moscow, 125009 Russia

<sup>b</sup>Engelhardt Institute of Molecular Biology, RAS  
 Vavilova, 32, Moscow, 119991 Russia

<sup>c</sup>Moscow State University, Department of Vertebrate Zoology  
 Leninskiye Gory, 1, Bld. 12, Moscow, 119234 Russia

\*E-mail: wslebedev@mail.ru

\*\*E-mail: toki@mail.ru

\*\*\*E-mail: hylomys@mail.ru

Using the example of multiple species groups and species complexes, it is empirically shown that the resolving power of a set of 10–20 nuclear loci is insufficient for the study of their genetic differentiation. For the adequate analysis of phylogenetic relationships (including reticulation events) and assessment of divergence levels within such groups of closely related species, it is necessary to multiply the number of nuclear loci and transit to high-throughput next-generation deep sequencing. However, the question of the genomic coverage required for the purposes of such study remains open. In other words, exactly how many genes in a set are needed to measure the genetic distance resolving the relationships between branches on the phylogenetic tree reconstruction of a complex of closely related species? We chose protein-coding sequences as a standard set of markers. Distances for 160–180 genes with a combined length of 270 000–300 000 bp were used to estimate the divergence levels of closely related mammalian species. In most cases, the interspecific distances are within the 0.15–0.75% range (median of 0.33%). The range of distances between semi-species is 0.12–0.28% (median of 0.14%). Intraspecific distances are always lower than 0.11%. Rodents show higher values of interspecific distances, 0.25–2.3% (median of 0.72%); distances between half-species range from 0.1 to 0.35%. To determine the number of nuclear loci and their combined length sufficient for the calculation of a genetic distance, we use simulations based on a model that included the following parameters: the average rate of gene evolution, its dispersion, and the level of polymorphism in the modern and ancestral populations. We performed a preliminary analysis of the distribution of loci evolution rates among mammalian taxa based on the data on ~50 thousand nuclear markers. It is shown that a relative error of 10–15%, comparable to the same value for mitochondrial distances between close species based on individual genes (approximately 1000 bp in length), is achieved using approximately 100 loci of 300 bp in length. Based on these data, we propose the following working hypothesis: the threshold of interspecific/intraspecific genetic distance calculated on the basis of the exons most frequently used in mammalian molecular phylogenetics is approximately 0.15%. This hypothesis assumes a species rank for the forms *in statu nascendi*. It should be noted that not all “good” species have significant genetic distances, since in some cases reproductive barriers form faster than substitutions accumulate in the genome.