Вавиловский журнал генетики и селекции. 2016;20(6):816-822
DOI 10.18699/VJ16.153

Комбинирование и оценка конгруэнтности филогенетических сигналов от нескольких генов с помощью геометрического подхода

В.М. Ефимов[1, 2, 3] ✉, В.Ю. Ковалева[4], Ю.Н. Литвинов[4]

[1] Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия
[2] Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия
[3] Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский Томский государственный университет», Томск, Россия
[4] Федеральное государственное бюджетное учреждение науки Институт систематики и экологии животных Сибирского отделения Российской академии наук, Новосибирск, Россия

Рассмотрены возможности недавно предложенного нами нового алгоритма (DJ-метода) для анализа конгруэнтности и комбинирования молекулярно-генетических данных на основе матриц евклидовых расстояний. Этот подход назван геометрическим, поскольку евклидовы дистанции удовлетворяют аксиомам метрики, что обеспечивает возможность помещения множества точек, представляющих последовательности, в некоторое геометрическое пространство без искажения взаимных расстояний и наделения точек этого множества координатами в пространстве. Геометричность евклидовых расстояний позволяет применять к молекулярным данным весь арсенал методов многомерного анализа, что является актуальным для исследования соотношения внутри- и межвидовой изменчивости, вычисления центроидов таксонов и расстояний между ними, визуализации возможных направлений эволюции, комбинирования и оценки конгруэнтности филогенетических сигналов, относящихся к разным генам и даже к разным системам признаков. DJ-метод использован для оценки конгруэнтности и комбинирования филогенетических сигналов, получаемых от нескольких генов. Анализировались более 1500 нуклеотидных последовательностей двух ядерных (*apoB*, *brca1*) и двух митохондриальных (*co1*, *cytb*) генов 15 палеарктических и неарктических видов землероек-бурозубок рода *Sorex* (Soricidae, Eulipotyphla). Для каждого гена все его последовательности представлялись множеством точек в евклидовом пространстве. Для множества точек, относящихся к одному виду, вычислялся его центроид в этом же пространстве, а для каждого гена – матрица евклидовых расстояний между центроидами видов. Для оценки попарного сходства (конгруэнтности) матриц межвидовых расстояний применен тест Мантеля. Конгруэнтность генов яДНК составила 0.961, мтДНК – 0.748. Все матрицы межвидовых расстояний через взвешивание объединены в единую матрицу, по которой методом главных координат для всех видов построено единое пространство. В объединенном генетическом пространстве проявилось несколько направлений межвидовой изменчивости, отражающих разные по масштабу эволюционные события. По объединенной матрице межвидовых расстояний построено дерево, которое хорошо согласуется с принятой на сегодня зоологической систематикой. Это подтверждает работоспособность предложенного метода.

Ключевые слова: *Sorex*; мтДНК; яДНК; DJ-метод; филогенетика; евклидово пространство.

Combining and congruence evaluation of phylogenetic signals from different genes based on geometric approach

V.M. Efimov[1, 2, 3] ✉, V.Yu. Kovaleva[4], Yu.N. Litvinov[4]

[1] Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
[3] Tomsk State University, Tomsk, Russia
[4] Institute of Systematics and Ecology of Animals SB RAS, Novosibirsk, Russia

A new Euclidean distance based algorithm is used for analysis of congruence and combining molecular genetic data. This approach is called geometric, since Euclidean distance satisfies all metric axioms and the points representing the sequences can be placed in a geometric space without distorting the mutual distances and can be endowed with the coordinates in this space. Geometricness of Euclidean distances allows methods of multivariate analysis to be applied to molecular data, which is relevant to intra- and interspecies variability investigating, visualization of possible directions of evolution, combining data and evaluation of the congruence of phylogenetic signals. The algorithm is used for the analysis of more than 1500 nucleotide sequences of two nuclear (*apoB*, *brca1*) and two mitochondrial (*co1*, *cytb*) genes of 15 Palaearctic and Nearctic shrew species of the genus *Sorex* (Soricidae, Eulipotyphla). All sequences of each gene are represented as a set of points in Euclidean space. Centroids of a set of points belonging to the same species are calculated. The matrix of Euclidean distances between the species centroids is calculated for each gene. Mantel test is applied to estimate pairwise similarity (congruence) of interspecies distances matrices relating to different genes. nDNA gene congruence is 0.961, that of mtDNA is 0.748. All matrices of the interspecies distances are combined into a joint matrix by weighing. Joint genetic space for all species is built by principal coordinate method from the joint matrix. Several variability directions reflecting evolutionary events of different scale are visualized in a joint genetic space. In addition, the joint matrix of interspecies distances is used for building a

УДК 57.065
Поступила в редакцию 07.07.2015 г.
Принята к публикации 01.02.2016 г.
Опубликовано онлайн 09.06.2016 г.

✉ e-mail: efimov@bionet.nsc.ru

phylogenetic tree, which is consistent with the zoological systematics accepted for today. This confirms the efficiency of our proposed method.

Key words: *Sorex*; mtDNA; nuclear DNA; DJ-method; phylogenetics; Euclidean space.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Ефимов В.М., Ковалева В.Ю., Литвинов Ю.Н. Комбинирование и оценка конгруэнтности филогенетических сигналов от нескольких генов с помощью геометрического подхода. Вавиловский журнал генетики и селекции. 2016;20(6):816-822. DOI 10.18699/VJ16.153

HOW TO CITE THIS ARTICLE:

Efimov V.M., Kovaleva V.Yu., Litvinov Yu.N. Combining and congruence evaluation of phylogenetic signals from different genes based on geometric approach. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2016;20(6):816-822. DOI 10.18699/VJ16.153

Одна из целей молекулярной систематики – изучение филогенетических взаимоотношений некоторого набора таксонов по нуклеотидным последовательностям, относящимся к разным генам. Для достижения этой цели необходимо решить две самостоятельные и важные задачи: проанализировать конгруэнтность филогенетических сигналов от разных генов (Гречко и др., 2006; Гречко, 2013) и объединить информацию, относящуюся к разным генам.

В настоящее время для решения обеих задач применяются два подхода: конкатенация (метод суперматрицы) и консенсус (метод супердерева) (Delsuc et al., 2005; Gadagkar et al., 2005; Philippe et al., 2005; Jeffroy et al., 2006; Planet, 2006). При использовании первого подхода выровненные последовательности из нескольких генов стыкуются в одну, и филогенетический анализ выполняется на объединенных в общую матрицу последовательностях (Nylander, 2004; Wortley, Scotland, 2006; de Queiroz, Gatesy, 2007). При использовании второго отдельное дерево строится для каждого гена, а общее дерево оценивается на основе консенсуса (Bininda-Emonds et al., 2002; Burleigh et al., 2011). Оба подхода ориентированы на деревья как форму представления окончательного результата. Это обусловлено тем, что на сегодняшний день основными подходами в филогенетике являются методы максимальной парсимонии, максимального правдоподобия и другие статистические подходы, опирающиеся на модели эволюции непосредственно нуклеотидных последовательностей (Лукашов, 2009) и неизбежно приводящие к деревьям.

Однако, вопреки широко распространенному в молекулярной филогенетике мнению, деревья – не единственный способ представления филогенетических взаимоотношений. Они могут быть дополнены отображением взаимного расположения последовательностей в многомерном пространстве (геометрический подход) или даже объединены с ним (Cavalli-Sforza, Edwards, 1967; Klingenberg, Ekau, 1996; Scippa et al., 2008; Klingenberg, Gidaszewski, 2010; Ковалева и др., 2012; Polly et al., 2013).

Некоторые исследователи выделяют в отдельный подход дистанционные методы, основанные на матрицах расстояний, как первичных, посчитанных непосредственно по последовательностям, так и вторичных, полученных из деревьев для каждого гена (Criscuolo et al., 2006; Criscuolo, Michel, 2009). Это не вызывало бы возражений, если бы используемые дистанции были дистанциями в строгом математическом смысле этого слова. Внимательное рассмотрение показывает, что это не совсем так. По сути дела, дистанциями, расстояниями и даже метриками в молекулярной филогенетике называются произвольные меры различия между последовательностями. Для нуклеотидных и аминокислотных последовательностей уже предложено довольно много разнообразных генетических расстояний и постоянно предлагаются новые (Criscuolo, Michel, 2009), однако при этом не имеются в виду метрические свойства этих расстояний и, тем более, анализ взаимного расположения таксонов в каком-либо геометрическом пространстве. В частности, не являются евклидовыми такие известные генетические расстояния, как *p*-дистанция (см. Доп. материалы 1)[1], расстояние Нея (Beaumont et al., 1998), Джукса – Кантора, двупараметрическое расстояние Кимуры, LogDet (Ефимов и др., 2013) и, по-видимому, многие другие. Соответственно, анализ конгруэнтности и объединение расстояний производятся без учета их метрических свойств и достаточно произвольно. Объединение расстояний реализуется просто суммированием или усреднением (Criscuolo, Michel, 2009), анализ конгруэнтности – при помощи теста Мантеля (Mantel, 1967; Mantel, Valand, 1970). На практике дистанционные методы, используемые сегодня в молекулярной филогенетике, в том числе и как начальный этап при построении эволюционных моделей, ограничены тем, что по матрице объединенных расстояний между последовательностями тем или иным алгоритмом все равно строится филогенетическое дерево. Фактически большинство существующих дистанционных методов – это разновидности методов суперматрицы и супердерева.

Конечно, было бы гораздо удобнее, если бы дистанции между последовательностями сразу являлись метрическими расстояниями, т. е. удовлетворяли аксиомам метрики: неотрицательности, симметричности, неравенству треугольника. Еще лучше, если бы они были евклидовыми расстояниями. Выполнение аксиом метрики обеспечивает возможность помещения множества точек, представляющих последовательности, в некоторое геометрическое пространство без искажения взаимных расстояний и наделения точек этого множества координатами в этом пространстве (Havel et al., 1983; Gower, Legendre, 1986).

Для целей настоящей статьи будем называть геометрическим любое молекулярно-генетическое расстояние, удовлетворяющее аксиомам метрики, поскольку используемые в настоящее время в молекулярной филогенетике термины «расстояние», «дистанция», «метрика» не име-

[1] Дополнительные материалы см. в Приложении 2 по адресу: http://www.bionet.nsc.ru/vogis/download/pict-2016-20/appx2.pdf

ют достаточно определенного геометрического смысла. К геометрическим расстояниям относится известное генетическое расстояние Кавалли-Сфорца – Эдвардса (Cavalli-Sforza, Edwards, 1967). Оно выведено при предположении, что генетическая разница возникает главным образом за счет генетического дрейфа. Это расстояние является евклидовым и обладает хорошими статистическими свойствами. Авторы с самого начала имели в виду отображение эволюции в виде пучка расходящихся траекторий в некотором многомерном евклидовом пространстве и прекрасно понимали геометрический смысл предлагаемого ими расстояния. В нашей недавней работе введены два новых геометрических расстояния, которые служат евклидовыми аналогами расстояний Джукса – Кантора и Кимуры (Ефимов и др., 2013). Если генетические расстояния являются геометрическими, то это позволяет применять весь арсенал методов многомерного анализа для исследования соотношения внутри- и межвидовой изменчивости, вычисления центроидов таксонов и расстояний между ними, визуализации возможных направлений эволюции, комбинирования и оценки конгруэнтности филогенетических сигналов, относящихся к разным генам и даже к разным системам признаков (Ковалева и др., 2012, 2013; Ефимов и др., 2013). Обсуждение этих возможностей на примере ядерных и митохондриальных генов является целью настоящей работы.

Материалы и методы

Проанализированы взятые из GenBank нуклеотидные последовательности двух ядерных (*apoB*, *brca1*) и двух митохондриальных (*co1*, *cytb*) генов, относящихся к 15 палеарктическим и неарктическим видам землероек рода *Sorex* (Soricidae, Eulipotyphla): подрода *Sorex* – *Sorex araneus*, *S. bedfordiae*, *S. caecutiens*, *S. daphaenodon*, *S. isodon*, *S. minutissimus*, *S. minutus*, *S. roboratus*, *S. tundrensis* и подрода *Otisorex* – *S. cinereus*, *S. fumeus*, *S. haydeni*, *S. monticolus*, *S. palustris*, *S. trowbridgii*. В общей сложности обработано 1588 последовательностей (см. Доп. материалы 2).

Особенностью молекулярной филогенетики, осложняющей дальнейший анализ, является то, что ее цель – это установление филогенетических взаимоотношений между таксонами, а исходные нуклеотидные последовательности, как правило, относятся к особям, и их число для разных таксонов на практике сильно различается. По этой причине построение эволюционного дерева выполняется для последовательностей, и только потом по нему каким-либо образом строится таксономическое дерево. Из этого следует, что таксономическая принадлежность особей должна быть определена заранее. Исходной информацией служит множество нуклеотидных последовательностей, классифицированных по таксонам и генам. Для каждого сочетания «таксон – ген» имеется хотя бы одна последовательность. Последовательности по разным генам не обязательно соответствуют одним и тем же особям.

В работе (Ковалева и др., 2012) нами предложен новый алгоритм анализа соответствия и комбинирования молекулярно-генетических и морфологических данных на основе матриц расстояний между видами (DJ-метод). Для того чтобы получить межвидовые расстояния из нуклеотидных последовательностей, относящихся к особям, нам пришлось дополнительно для каждого вида конструировать модальную (консенсусную) последовательность из нуклеотидов, с максимальной частотой встречающихся в каждой позиции. Это решение оказалось не совсем удачным, и в следующей работе (Ковалева и др., 2013) мы попытались более последовательно применить геометрический подход.

Усовершенствование алгоритма сводится к следующему. Сначала для каждого гена строится матрица евклидовых расстояний между всеми последовательностями всех таксонов. Полученная матрица расстояний переводится в матрицу объект – признак методом главных координат (Torgerson, 1952; Gower, 1966). Это означает, что каждой последовательности ставится в соответствие некоторая точка в евклидовом пространстве, причем размерность этого пространства будет всегда меньше числа последовательностей. Для множества точек, относящихся к одному таксону, вычисляется его центроид, который и представляет данный таксон. Между центроидами таксонов вычисляется матрица евклидовых расстояний, которая переводится в матрицу объект – признак методом главных координат. Это значит, что теперь каждому таксону ставится в соответствие некоторая точка в некотором новом евклидовом пространстве, причем размерность этого пространства будет всегда меньше числа таксонов.

Дополнительным преимуществом такого подхода является возможность контроля над систематическим искажением результатов, возникающим вследствие разного числа анализируемых последовательностей для каждого таксона. Более того, легко решается и сопутствующая проблема, появившаяся из-за накопления в базах данных таксономически неправильно определенных или искаженных при секвенировании или выравнивании последовательностей (Dubey et al., 2009; Гречко, 2013). Поскольку при отображении последовательностей в евклидово пространство четко видна структура каждого таксона и соотношение внутри- и межтаксонной изменчивости, то возможно отфильтровать сомнительные последовательности от основного ядра каждого таксона.

Геометричность евклидовых расстояний между видами позволяет легко и однозначно решить задачу комбинирования филогенетических сигналов, относящихся к разным генам. Так как каждую матрицу евклидовых расстояний можно методом главных координат без потери информации перевести в матрицу объект – признак, то конкатенация этих матриц для разных генов позволяет получить единую координатную матрицу, по которой однозначно вычисляется объединенная матрица евклидовых расстояний. Очевидно, что та же самая объединенная матрица получится, если мы поэлементно просуммируем квадраты межвидовых расстояний в исходных матрицах и извлечем квадратный корень. При желании исходные матрицы можно брать с некоторыми неотрицательными весами (Cavalli-Sforza, Edwards, 1967).

В настоящей работе в качестве весов выбраны координаты первого собственного вектора (Abeysundera et al., 2014). Объединенная матрица расстояний между видами обработана методами Крускала (Kruscal, 1964) и UPGMA (Unweighted pair-group method using arithmetic averages).

Результаты

Сначала по каждому гену между всеми его последовательностями с помощью пакетов Jacobi 4 и Excel вычислены E_{PQ}-дистанции (при $\alpha = 1$) (Ефимов и др., 2013), являющиеся евклидовыми аналогами двупараметрического расстояния Кимуры (Kimura, 1980). Методом главных координат все последовательности представлены точками в многомерном евклидовом пространстве. На рис. 1 приведена конфигурация полных последовательностей гена *cytb* для трех видов бурозубок группы «*araneus*» на плоскости первых двух главных координат. Отчетливо видны как основные ядра видов, так и отклоняющиеся последовательности, которые были исключены из дальнейшей обработки (см. рис. 1).

Далее для каждого гена вычислены центроиды видов и матрица евклидовых расстояний между ними. Для оценки попарного сходства между полученными матрицами межвидовых расстояний применен тест Мантеля (см. таблицу).

Принято считать, что филогенетический сигнал всегда отображается в виде дерева. Однако нисколько не хуже отражает филогенетический сигнал и матрица межвидовых расстояний, вычисленная по молекулярно-генетическим последовательностям. Естественно полагать, что чем дальше на дереве отстоит гипотетический ближайший общий предок от данной пары видов, тем больше должно быть расстояние между этой парой в матрице межвидовых расстояний. А корреляция между полученными отдельно по каждому гену матрицами межвидовых расстояний (тест Мантеля) является оценкой сходства (конгруэнтности) филогенетических сигналов.

При оценке сходства филогенетических сигналов, полученных по каждому отдельному гену, оказалось, что ядерные гены *apoB* и *brca1* фактически дублируют друг друга, тогда как митохондриальные – *co1* и *cytb* – несут филогенетические сигналы, несколько отличающиеся как от ядерных генов, так и между собой, хотя корреляция между ними достаточно высока и достоверна ($p < 10^{-6}$) (см. таблицу).

В нашей работе в качестве весов выбраны координаты первого собственного вектора. Объединенная матрица расстояний между видами обработана методами Крускала и UPGMA (рис. 2 и 3).

Результаты обработки методом Крускала показали, что максимальная доля дисперсии (93.3 %) приходится на первую ось шкалирования. Вдоль этого направления четко обозначились различия между палеарктическими и неарктическими видами бурозубок, т. е. между подродами *Sorex* и *Otisorex* (см. рис. 2). Кроме того, по второй оси шкалирования видна обособленность *S. bedfordiae* от палеарктических видов и *S. trowbridgii* от неарктических видов. При этом на долю второй и третьей осей шкалирования приходится 5.14 и 1.53 % дисперсии соответственно. Понятно, что на фоне таких соотношений долей дисперсии сложно анализировать внутриродовую структуру таксона, поэтому матрицы квадратов межвидовых расстояний обработаны с помощью SVD-метода (Singular value decomposition method) по отдельности для палеарктических и неарктических видов.

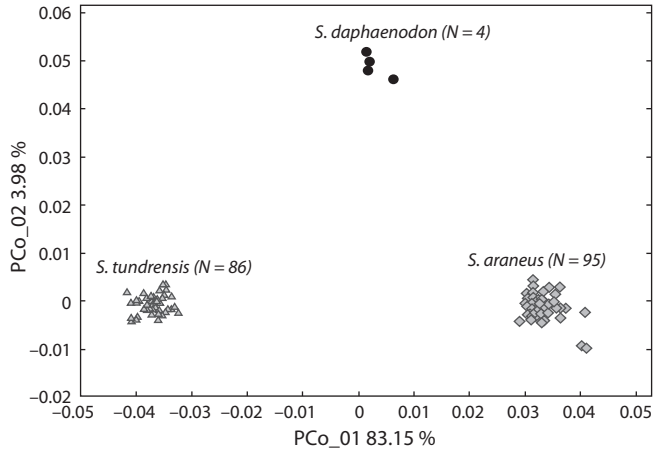

Рис. 1. Видовые облака группы «*araneus*», образованные нуклеотидными последовательностями гена *cytb* на плоскости первых двух главных координат.

Три отклоняющиеся последовательности *S. araneus* исключены из дальнейшей обработки.

В подроде *Sorex* заметной структурированности не наблюдается (рис. 4). Более или менее уверенно можно говорить о группе «*araneus*»: *S. araneus*, *S. tundrensis*, *S. daphaenodon*. На дендрограмме она образует обособленную ветвь (см. рис. 3). Ранее эта группа видов была выделена на основании кариологических данных (Dannelid, 1991; Ivanitskaya, 1994). В настоящее время монофилию группы поддерживают данные секвенирования гена *cytb* мтДНК (Ohdachi et al., 1997, 2006; Fumagalli et al., 1999). Остальные виды располагаются в трехмерном пространстве во всех направлениях, не образуя скоплений.

В частности, не выделилась группа 42 хромосомных бурозубок *«caecutiens»* (*S. caecutiens*, *S. roboratus*, *S. isodon*). В работе (Fumagalli et al., 1999) на основании гена *cytb* мтДНК она выделена в несколько ином составе (*S. cae-*

Коэффициенты корреляции между матрицами межвидовых расстоянии 15 видов бурозубок рода *Sorex*

Гены	*apoB*	*brca1*	*col*	*cytb*
apoB	1	0.961*	0.709*	0.759*
brca1	0.961*	1	0.684*	0.759*
col	0.709*	0.684*	1	0.748*
cytb	0.759*	0.759*	0.748*	1

Примечание. Применен тест Мантеля. Число пермутаций $Np \geq 10^6$.
* $p < 10^{-6}$.

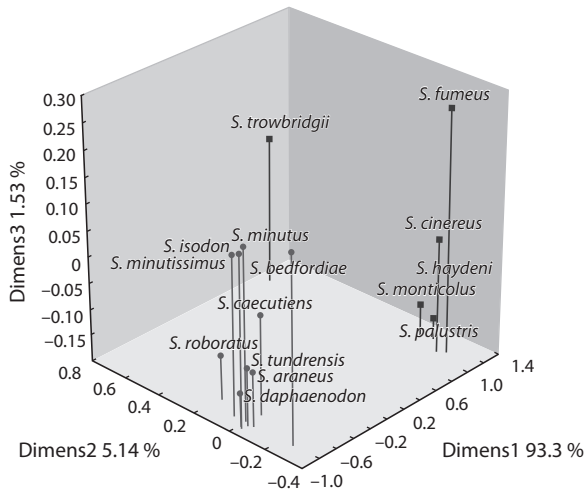

Рис. 2. Конфигурация видов бурозубок в трехмерном пространстве шкалирования методом Крускала объединенной матрицы расстояний между центроидами 15 видов рода *Sorex* по двум ядерным (*apoB*, *brca1*) и двум митохондриальным (*co1*, *cytb*) генам.

Dimens1, 2, 3 – оси многомерного шкалирования.

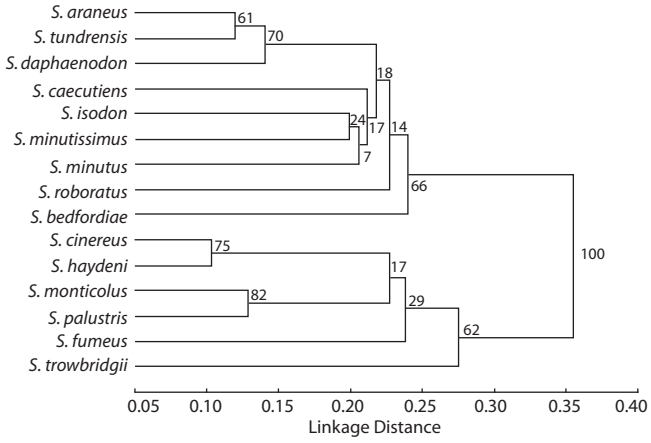

Рис. 3. Дендрограмма, построенная методом UPGMA по объединенной генетической матрице расстояний между центроидами 15 видов рода *Sorex*.

Коэффициент кофенетической корреляции 0.994; числа на ветвях – бутстреп-поддержка.

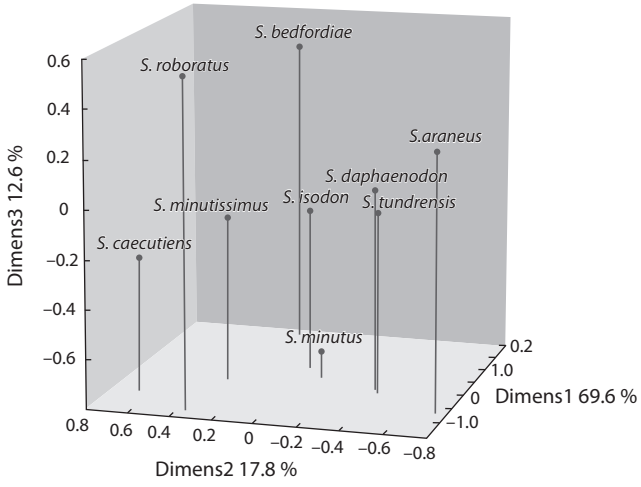

Рис. 4. Конфигурация видов бурозубок в трехмерном пространстве шкалирования методом Крускала объединенной матрицы расстояний между центроидами 9 видов подрода *Sorex* по двум ядерным (*apoB*, *brca1*) и двум митохондриальным (*co1*, *cytb*) генам.

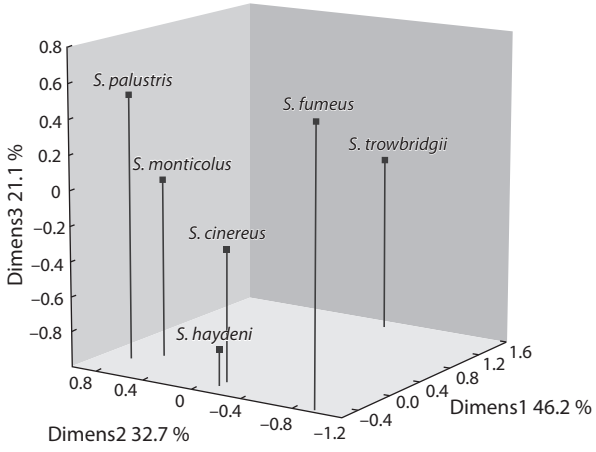

Рис. 5. Конфигурация видов бурозубок в трехмерном пространстве шкалирования методом Крускала объединенной матрицы расстояний между центроидами 6 видов подрода *Otisorex* по двум ядерным (*apoB*, *brca1*) и двум митохондриальным (*co1*, *cytb*) генам.

cutiens, *S. minutus*, *S. isodon*, *S. minutissimus*) с довольно слабой бутстреп-поддержкой.

В подроде *Otisorex* видны прежде всего высокая генетическая обособленность *S. trowbridgii* по первой оси шкалирования, *S. fumeus* – по второй, сестринские взаимоотношения *S. cinereus* и *S. haydeni* (группа «*cinereus*»), *S. monticolus* и *S. palustris* (группа «*vagrans*») (рис. 5).

Полученные данные согласуются как с результатами (Ohdachi et al., 2006), так и с хронологическими оценками времени дивергенции таксонов: расхождение *S. trowbridgii* с остальными видами – 8.91 Mya (млн лет назад), *S. fumeus* – 6.68 Mya, дивергенция группы «*vagrans*» от группы «*cinereus*» – 6.01 Mya, *S. cinereus* от *S. haydeni* – 3.71 Mya, *S. monticolus* от *S. palustris* – 1.96–2.41 Mya (Esteva et al., 2010).

Обсуждение

В филогенетике сложилась традиция воспринимать генетические «дистанции» как настоящие расстояния, хотя, строго говоря, они не всегда таковыми являются (Ефимов и др., 2013). Дистанционные методы опираются на матрицы расстояний между видами. Однако матрицы содержат значительно больше информации, чем получаемые из них филогенетические деревья (Ковалева и др., 2012). По меткому выражению (Pershina et al., 2011), за филогенетическими деревьями мы не видим леса. Лес в данном случае – многомерная эволюция, различные варианты которой можно увидеть, только если учитывать все расстояния. Существует достаточно развитая технология отображения с минимальной потерей информации матриц сходства/различия в удобное метрическое

пространство – многомерное шкалирование (Дэйвисон, 1988). В результате этого отображения мы получаем так называемое «таксономическое или эволюционное пространство», полнее представляющее «филогенетический сигнал» (Kitazoe et al., 2001, 2005; Huson, Bryant, 2006; Lee el al., 2006; Pershina et al., 2011; Дольник и др., 2012).

Из наших результатов следует, что достаточно конгруэнтные друг другу филогенетические сигналы выявились для всех четырех генов, как ядерных, так и митохондриальных (см. таблицу). В результате их объединения с помощью геометрического подхода получено дерево, практически совпадающее с принятой на сегодня зоологической систематикой (см. рис. 3) (Wilson, Reeder, 2005). Это подтверждает работоспособность предлагаемого нами метода. Кроме того, в объединенном генетическом пространстве выявляется несколько направлений изменчивости, отражающих эволюционные события, разные по масштабу и, по-видимому, относящиеся к разным эволюционным временам.

Следует отметить, что в последнее время наблюдается некоторое возрождение дистанционных методов. Вместо противопоставления филогенетических деревьев и многомерных геометрических представлений все чаще предпринимаются попытки построить филогенетическое дерево прямо в многомерном пространстве. С этой точки зрения особый интерес представляет евклидова задача Штейнера в *n*-мерном пространстве: для заданного множества точек найти составленную из отрезков кратчайшую сеть соединяющих путей (Lee et al., 2006; Brazil et al., 2009; Fonseca et al., 2014). Эта задача, известная не менее двух веков, имеет сейчас многочисленные применения в различных областях знания (Brazil et al., 2014).

Использование деревьев Штейнера как самых подходящих на роль филогенетических деревьев предложено еще в работе (Cavalli-Sforza, Edwards, 1967) и затем повторено в (Brazil et al., 2009). Прогресс в развитии вычислительных средств и алгоритмов делает подобную перспективу весьма актуальной. В настоящее время в Институте цитологии и генетики СО РАН разрабатывается пакет прикладных программ Jacobi 4, нацеленный на реализацию геометрического подхода для единообразного решения задач из различных областей биологии (Полунин и др., 2014). Задачу Штейнера предполагается включить в него в качестве одного из модулей.

Благодарности

Работа выполнена в рамках государственного задания по проекту № 0324-2015-0003 и гранта РФФИ № 14-04-00121-а.

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Список литературы

Гречко В.В. Проблемы молекулярной филогенетики на примере отряда чешуйчатых рептилий (отряд SQUAMATA): митохондриальные ДНК-маркеры. Мол. биология. 2013;47(1):61-82.

Гречко В.В., Федорова Л.В., Рябинин Д.М., Рябинина Н.Л., Чобану Д.Г., Косушкин С.А., Даревский И.С. Молекулярные маркеры ядерной ДНК в исследовании видообразования и систематики на примере ящериц комплекса «*Lacerta agilis*» (SAURIA: LACERTIDAE). Мол. биология. 2006;40(1):61-73.

Дольник А.С., Тамазян Г.С., Першина Е.В., Вяткина К.В., Порозов Ю.Б., Пинаев А.Г., Андронов Е.Е. Концепция универсальной таксономической системы бактерий: эволюционное пространство гена 16S-рРНК v. 1.0. С.-х. биология. 2012;5:111-120.

Дэйвисон М. Многомерное шкалирование. М.: Финансы и статистика, 1988.

Ефимов В.М., Мельчакова М.А., Ковалева В.Ю. Геометрические свойства эволюционных дистанций. Вавиловский журнал генетики и селекции. 2013;17(4/1):714-723.

Ковалева В.Ю., Абрамов С.А., Дупал Т.А., Ефимов В.М., Литвинов Ю.Н. Анализ соответствия и комбинирование молекулярно-генетических и морфологических данных в зоологической систематике. Изв. РАН. Сер. биол. 2012;4:404-414.

Ковалева В.Ю., Литвинов Ю.Н., Ефимов В.М. Землеройки (SORICIDAE, EULIPOTYPHLA) Сибири и Дальнего Востока: комбинирование и поиск конгруэнтности молекулярно-генетических и морфологических данных. Зоол. журн. 2013;92(11): 1383-1398.

Лукашов В.В. Молекулярная эволюция и филогенетический анализ. М.: Бином. Лаборатория знаний, 2009.

Полунин Д.А., Штайгер И.А., Ефимов В.М. Разработка программного комплекса JACOBI 4 для многомерного анализа микрочиповых данных. Вестн. Новосиб. ун-та. Серия: Информ. технологии. 2014;12(2):90-98.

Abeysundera M., Kenney T., Field C., Gu H. Combining distance matrices on identical taxon sets for multi-gene analysis with singular value decomposition. PLoS One. 2014;9(4):e94279. DOI 10.1371/journal.pone.0094279.

Beaumont M.A., Ibrahim K.M., Boursot P., Bruford M.W. Measuring genetic distance. Molecular tools for screening biodiversity: Plants and Animals (Eds. A. Karp, D.S. Ingram, P.G. Isaac). London: Chapman & Hall, 1998;315-325. DOI 10.1007/978-94-009-0019-6_58.

Bininda-Emonds O.R.P., Gittleman J.L., Steel M.A. The (super) tree of life: procedures, problems, and prospects. Annu. Rev. Ecol. Syst. 2002;33:265-289. DOI 10.1146/annurev.ecolsys.33.010802.150511.

Brazil M., Graham R.L., Thomas D.A., Zachariasen M. On the history of the euclidean Steiner tree problem. Archive History Exact Sciences. 2014;68:327-354. DOI 10.1007/s00407-013-0127-z.

Brazil M., Thomas D.A., Nielsen B.K., Winter P., Wulff-Nilsen C., Zachariasen M. A novel approach to phylogenetic trees: d-dimensional geometric Steiner trees. Networks. 2009;53(2):104-111.

Burleigh J.G., Bansal M.S., Eulenstein O., Hartmann S., Wehe A., Vision T.J. Genome-scale phylogenetics: inferring the plant tree of life from 18,896 gene trees. Syst. Biol. 2011;60(2):117-125. DOI 10.1093/sysbio/syq072.

Cavalli-Sforza L.L., Edwards A.W. Phylogenetic analysis. Models and estimation procedures. Am. J. Human Genet. 1967;19(3):233-257.

Criscuolo A., Berry V., Douzery E.J., Gascuel O. SDM: a fast distance & based approach for (super) tree building in phylogenomics. Syst. Biol. 2006;55(5):740-755. DOI 10.1080/10635150600969872.

Criscuolo A., Michel C.J. Phylogenetic inference with weighted codon evolutionary distances. J. Mol. Evol. 2009;68(4):377-392. DOI 10.1007/s00239-009-9212-y.

Dannelid E. The genus *Sorex* (Mammalia, Soricidae) – distribution and evolutionary aspects of Eurasian species. Mammal Rev. 1991;21(1): 1-20. DOI 10.1111/j.1365-2907.1991.tb00284.x.

de Queiroz A., Gatesy J. The supermatrix approach to systematic. Trends Ecol. Evol. 2007;22(1):34-41. DOI /10.1016/j.tree.2006.10.002.

Delsuc F., Brinkmann H., Philippe H. Phylogenomics and the reconstruction of the tree of life. Nature Rev. Genet. 2005;6(5):361-375. DOI 10.1038/nrg1603.

Dubey S., Michaux J., Brünner H., Hutterer R., Vogel P. False phylogenies on wood mice due to cryptic cytochrome-*b* pseudogene. Mol. Phylogen. Evol. 2009;50(3):633-641. DOI 10.1016/j.ympev.2008.12.008.

Esteva M., Cervantes F.A., Brant S.V., Cook J.A. Molecular phylogeny of long-tailed shrews (genus *Sorex*) from Mexico and Guatemala. Zootaxa. 2010;2615:47-65.

Fonseca R., Brazil M., Winter P., Zachariasen M. Faster exact algorithms for computing Steiner trees in higher dimensional euclidean spaces. 11th DIMACS Implementation challenge on Steiner tree problems. Providence, Rhode Island: Brown Univ., 2014.

Fumagalli L., Taberlet P., Stewart D.T., Gielly L., Hausser J., Vogel P. Molecular phylogeny and evolution of *Sorex* shrews (Soricidae: Insectivora) inferred from mitochondrial DNA sequence data. Mol. Phylogen. Evol. 1999;11(2):222-235. DOI 10.1006/mpev.1998.0568.

Gadagkar S.R., Rosenberg M.S., Kumar S. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. J. Experimental Zoology. Pt. B: Molecular and Developmental Evolution. 2005;304B(1):64-74. DOI 10.1002/jez.b. 21026.

Gower J.C. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika. 1966;53(3/4):325-338. DOI 10.2307/2333639.

Gower J.C., Legendre P. Metric and Euclidean properties of dissimilarity coefficients. J. Classification. 1986;3(1):5-48. DOI 10.1007/bf01896809.

Havel T.F., Kuntz I.D., Crippen G.M. The theory and practice of distance geometry. Bull. Mathem. Biol. 1983;45(5):665-720. DOI 10.1007/bf02460044.

Huson D.H., Bryant D. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 2006;23(2):254-267. DOI 10.1093/molbev/msj030.

Ivanitskaya E.Y. Comparative cytogenetics and systematics of *Sorex*: a cladistic approach. Advances in the biology of shrews (Eds. J.F. Merritt, G.L. Kirkland (Jr.), R.K. Rose). Pittsburgh: Carnegie Museum Nat. History, Spec. Publ. 1994;313-323.

Jeffroy O., Brinkmann H., Delsuc F., Philippe H. Phylogenomics: the beginning of incongruence? Trends Gen. 2006;22(4):225-231. DOI 10.1016/j.tig.2006.02.003.

Kimura M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 1980;16:111-120. DOI 10.1007/BF01731581.

Kitazoe Y., Kurihara Y., Narita Y., Okuhara Y., Tominaga A., Suzuki T. A new theory of phylogeny inference through construction of multidimensional vector space. Mol. Biol. Evol. 2001;18(5):812-828.

Kitazoe Y., Kishino H., Okabayashi T., Watabe T., Nakajima N., Okuhara Y., Kurihara Y. Multidimensional vector space representation for convergent evolution and molecular phylogeny. Mol. Biol. Evol. 2005;22(3):704-715. DOI 10.1093/molbev/msi051.

Klingenberg C.P., Ekau W. A combined morphometric and phylogenetic analysis of an ecomorphological trend: pelagization in Antarctic fishes (Perciformes: Nototheniidae). Biol. J. Linnean Soc. 1996;59(2):143-177. DOI 10.1111/j.1095-8312.1996.tb01459.x.

Klingenberg C.P., Gidaszewski N.A. Testing and quantifying phylogenetic signals and homoplasy in morphometric data. System. Biol. 2010;59(3):245-261. DOI 10.1093/sysbio/syp106.

Kruscal J.B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika. 1964;29(1):1-27. DOI 10.1007/BF02289565.

Lee S.H., Hwang K.S., Lee H.R., Kim S.S., Lee K.M., Lee C.H., Lee D. Embedding operational taxonomic units in three-dimensional space for evolutionary distance relationship in phylogenetic analysis. Proc. 5th WSEAS Intern. Conf. on Circuits, Systems, Electronics, Control and Signal Processing. USA. 2006;192-196.

Mantel N. The detection of disease clustering and a generalized regression approach. Cancer Research. 1967;27:209-220.

Mantel N., Valand R.S. A technique of nonparametric multivariate analysis. Biometrics. 1970;26:547-558. DOI 10.2307/2529108.

Nylander J.A.A. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University, 2. 2004.

Ohdachi S.D., Hasegawa M., Iwasa M.A., Vogel P., Oshida T., Lin L.-K., Abe H. Molecular phylogenetics of soricid shrews (Mammalia) based on mitochondrial cytochrome *b* gene sequences: with special reference to the Soricinae. J. Zool. 2006;270(1):177-191. DOI 10.1111/j.1469-7998.2006.00125.x.

Ohdachi S., Masuda R., Abe H., Adachi J., Dokuchaev N.E., Haukisalmi V., Yoshida M.C. Phylogeny of Eurasian soricine shrews (Insectivora, Mammalia) inferred from the mitochondrial cytochrome *b* gene sequences. Zool. Science. 1997;14(3):527-532.

Pershina E.V., Dolnik A.S., Tamazyan G., Ikonnikova E.V., Vyatkina K.V., Pinaev A.G., Andronov E.E. An evolutionary space for microbial evolution and community structure analysis. Department of Bioengineering and Bioinformatics of M.V. Lomonosov Moscow State Univ. 2011;54(3):40.

Philippe H., Delsuc F., Brinkmann H., Lartillot N. Phylogenomics. Ann. Rev. Ecol., Evol., System. 2005;541-562. DOI 10.1146/annurev.ecolsys.35.112202.130205.

Planet P.J. Tree disagreement: measuring and testing incongruence in phylogenies. J. Biomed. Inform. 2006;39(1):86-102. DOI 10.1016/j.jbi.2005.08.008.

Polly P.D., Lawing A.M., Fabre A.C., Goswami A. Phylogenetic principal components analysis and geometric morphometrics. Hystrix, Italian J. Mammal. 2013;24(1):33-41. DOI 10.4404/hystrix-24.1-6383.

Scippa G.S., Trupiano D., Rocco M., Viscosi V., Di Michele M., D'Andrea A., Chiatante D. An integrated approach to the characterization of two autochthonous lentil (Lens culinaris) landraces of Molise (south-central Italy). Heredity. 2008;101(2):136-144. DOI 10.1038/hdy.2008.39.

Torgerson W.S. Multidimensional scaling: I. Theory and method. Psychometrika. 1952;17(4):401-419. DOI 10.1007/BF02288916.

Wilson D.E., Reeder D.A.M. (Ed.). Mammal species of the world: a taxonomic and geographic reference. Baltimore: JHU Press, 2005;12:2142 p.

Wortley A.H., Scotland R.W. The effect of combining molecular and morphological data in published phylogenetic analyses. Syst. Biol. 2006;55(4):677-685. DOI 10.1080/10635150600899798.