

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПРОЦЕССА НАКОПЛЕНИЯ ГЕНЕТИЧЕСКИХ РАЗЛИЧИЙ В ЭВОЛЮЦИОННО-КОНСЕРВАТИВНЫХ НУКЛЕОТИДНЫХ ПОСЛЕДОВАТЕЛЬНОСТЯХ

Ю. В. Бондаренко, Н. В. Воронова

Белорусский государственный университет

Минск, Беларусь

e-mail: ybondarenko@buffergroup.com

Исследуется модель накопления генетических изменений в консервативных участках митохондриального генома животных. Проведено аналитическое и численное исследование модели. Показано, что предложенная модель хорошо согласуется с эмпирическими данными, полученными при анализе последовательностей COI в семи родственных таксонах насекомых.

Ключевые слова: математическое моделирование; нуклеотидные последовательности; скорость эволюции.

MATHEMATICAL MODELING OF THE ACCUMULATION OF GENETIC DIFFERENCES IN THE EVOLUTIONARILY CONSERVED NUCLEOTIDE SEQUENCE

Y. V. Bondarenko, N. V. Voronova

Belarusian State University

Minsk, Belarus

The model of accumulation of genetic changes in conservative genes of a mitochondrial genome of animals is studied. Analytical and numerical research of model is carried out. It is shown that the offered model well will be coordinated with the empirical data obtained in the analysis of sequences of COI in seven related taxa of insects.

Keywords: mathematical modeling; nucleotide sequences; rate of evolution.

ВВЕДЕНИЕ

Биологическая эволюция, выражающаяся в появлении новых видов живых существ, сопровождается накоплением генетических различий между вновь образующимися формами [1]. По сути, наличие этих различий является обязательным условием для того, чтобы «эволюционное событие» было зарегистрировано. Считается, что возникновение нового биологического вида из предковой формы должно сопровождаться изменением не менее чем 1 % генома [2]. Однако в действительности количество наблюдаемых изменений в отдельных эволюционных ветвях, приходящихся на один акт видообразования, может существенно различаться. Указанный факт лег в ос-

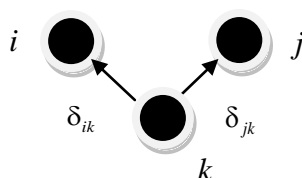
нову многочисленных поправок в существующие модели нуклеотидных замещений и алгоритмы построения филогенетических деревьев, а также потребовал введения нового понятия, формально противоречащего постулатам теории нейтральной эволюции, – скорости молекулярной эволюции, выражающейся в количестве нуклеотидных замещений в наследственном ряду последовательностей за млн лет.

Несмотря на то, что ситуация, когда в процессе эволюции изменение геномов происходит с варьирующей скоростью, широко распространена в природе, биологические основы этого явления не ясны. Математическое моделирование процесса накопления генетических различий в отдельных эволюционных линиях позволяет описать характер и скорость изменения геномов в разных таксонах, а в будущем, как считается, приблизиться к получению ответа на вопрос о биологических факторах, влияющих на скорость макроэволюции в разных линиях живых организмов. В нашей работе мы провели моделирование процесса накопления генетической дистанции по высоко консервативному митохондриальному гену субъединицы 1 цитохромоксидазы *c* (COI) в процессе эволюции реально существующих таксонов насекомых и рассчитали скорость молекулярной эволюции в каждой группе организмов.

МАТЕРИАЛЫ И МЕТОДЫ

Последовательности и таксоны. Последовательности COI семи родственных надсемейств насекомых (Aphidoidea, Coccoidea, Phylloxeroidea, Aleyrodoidea, Cicadoidea, Pentatomoidea, Psylloidea) получили из базы данных GenBank NCBI. Суммарная выборка составила 1414 последовательностей для которых было рассчитано 723 286 значений парных генетических дистанций. Компьютерный анализ последовательностей провели в программе MEGA6 [3]. Генетические дистанции рассчитали, используя TN93 модель.

Математическая модель. В процессе моделирования мы считали, что видообразование в каждой ветви эволюционного дерева происходит в соответствии с принципом дихотомического ветвления и с течением времени Δt_i будет происходить накопление генетических изменений в двух ветвях, исходящих из одного узла



где δ_{ik} – генетическая дистанция между предковым видом (k) и дочерним (i).

В дальнейшем событие, выражающееся в появлении двух дочерних видов из предковой формы, будем называть актом видообразования, а временной промежуток между этими двумя состояниями – тактом.

Учитывая возможность появления не только прямых, приводящих к росту генетической дистанции, но и обратных нуклеотидных замен, приводящих к ее уменьшению, генетическая дистанция между двумя дочерними видами будет задаваться как

$$\Delta G_{ij} = \delta_{ik} + \delta_{jk} - \delta_{ij},$$

где δ_{ij} – изменение генетической дистанции, определяемое обратными нуклеотидными заменами.

Поскольку размер анализируемой нуклеотидной последовательности значительно меньше длины генома, генетическая дистанция в анализируемом участке генома δ_i будет характеризоваться случайной величиной

$$\delta_i = \xi(\Delta t_i) - \widehat{\xi}(\Delta t_i),$$

где $\xi(\Delta t_i)$ – случайная величина, равная наблюдаемой генетической дистанции, $\widehat{\xi}(\Delta t_i)$ – случайная величина, равная вкладу обратных замен в значение генетической дистанции, накопленной за промежуток времени Δt_i ,

Поскольку считается, что каждый акт видообразования сопровождается появлением хотя бы одной нуклеотидной замены в последовательности гена COI [2], случайную величину $\xi(\Delta t_i)$ можно представить в следующем виде:

$$\xi(t) = \varepsilon + \delta_0 h(t),$$

где ε – «гарантированная» генетическая дистанция, сопровождающая акт видообразования; $\delta_0 h(t)$ – случайная функция, имеющая равномерное распределение на отрезке $[0, \delta_0]$.

Поскольку, как было сказано выше, генетическая дистанция между двумя последовательными узлами (видами) зависит от наличия не только прямых, но и обратных замен, она будет определяться выражением

$$\delta_i = \varepsilon + \delta_0 h(\Delta t_i) - \widehat{\delta}_0 \widehat{h}(\Delta t_i).$$

Генетическая дистанция, накопленная к некому моменту времени в каждой эволюционной ветви между последним общим предком всего таксона и рецентным видом, будет представлять собой сумму всех промежуточных генетических изменений:

$$\delta = \sum_{i=1}^N \delta_i(t).$$

При этом когда видообразование происходит по принципу дихотомии, количество этапов видообразования в таксоне будет выражаться как

$$N = \log_2 N_0,$$

где N – количество этапов видообразования; N_0 – количество существующих видов (в предлагаемой модели не учитываются вымершие виды).

В процессе эволюции таксона образуется множество значений генетических дистанций, каждая из которых определяется количеством изменений, накопленных между последним общим предком для всех видов таксона и существующим видом:

$$\{\delta^{(1)}, \delta^{(2)}, \dots, \delta^{(N_0)}\}.$$

Величину генетической дистанции между каждыми двумя видами в таксоне можно определить как:

$$\Delta G_{ij} = \delta^{(i)} + \delta^{(j)} - \theta(i, j),$$

где $\theta(i, j)$ – генетическая дистанция между последним общим предком для всех видов таксона и последним общим предком для видов (i) и (j) .

Так как изменения, накопленные в рамках анализируемого гена, не обязательно должны сопровождаться образованием нового вида, процесс видообразования также будет характеризоваться случайной величиной и соответствующим множеством

$$P = \{P_i | i \in [1, N]\},$$

где P_i – вероятность образования вида за определенный промежуток времени Δt_i .

Таким образом, генетическая дистанция между любыми двумя видами, обеспечиваемая появлением нуклеотидных замен в конкретном участке генома, будет определяться четырьмя величинами: P_i , ε , δ_0 , $\hat{\delta}_0$, где $\varepsilon = \text{const}$.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Основной проблемой при описании эволюционных процессов, протекающих на уровне нуклеотидных последовательностей, является отсутствие достаточно полных выборок, позволяющих исследовать масштаб эволюционных изменений на уровне целых таксонов. В частности, в нашей работе мы использовали последовательность гена, наиболее широко представленного в глобальных базах данных нуклеотидных последовательностей, однако даже в этом случае представленность варьировала от 3,5 % у крупных до 60 % у небольших таксонов. Для преодоления указанных ограничений предлагаем использовать разработанную нами модель.

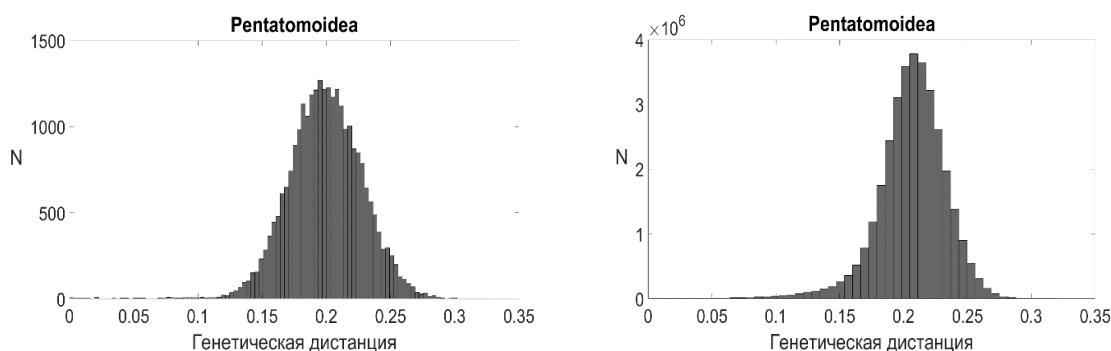


Рисунок. Распределение значений парных генетических дистанций для всех нуклеотидных последовательностей COI Pentatomoidea (слева) и распределение значений ΔG , рассчитанных с использованием предлагаемой модели (справа)

Для проверки применимости модели мы использовали нуклеотидные последовательности таксонов насекомых, для которых известны их эволюционный возраст и количество описанных видов. В процессе моделирования для каждого исследуемого таксона были рассчитаны значения $\Delta \delta$, при котором характер и свойства распределения значений парных генетических дистанций совпадал с эмпирическими данными, полученными при анализе распределения значений ΔG_{ij} , рассчитанных по выборке последовательностей COI (рисунок). При этом

$$\Delta \delta = \delta_0 - \hat{\delta}_0.$$

На основе анализа построенного с использованием предлагаемой модели распределения ΔG были также определены ΔG_{\max} , т. е. максимальное значение генетической дистанции между двумя видами, наиболее удаленными друг от друга в эволюционном отношении, и G для каждого таксона, равное

$$G = \sum_{i,j} \Delta G_{ij}.$$

Используя предлагаемую модель, была также определена скорость накопления генетической дистанции (V) в каждом таксоне с течением эволюционного времени:

$$V = \frac{N_0 \Delta \delta}{T},$$

где T – эволюционный возраст таксона по палеонтологическим данным (в млн лет), N_0 – количество ныне существующих (описанных) видов в таксоне.

В таблице представлены результаты вычислений для последовательности гена COI при следующих параметрах: вероятность прямой замены, увеличивающей значение генетической дистанции сверх $\varepsilon = 0,002$ (генетическая дистанция, определяемая одной транзицией в последовательности задаваемой длины) принята равной 0,02, поскольку известно, что близкие виды различаются в среднем на 2 % по последовательности используемого гена; отношение частоты прямых и обратных замен = 1/2, что обусловлено четырехбуквенным кодом ДНК; для участка генома заданной длины $\hat{\delta}_0 = 1,54 \cdot 10^{-5}$; вероятность видообразования за такт – 0,8.

Результаты моделирования для последовательностей COI исследуемых таксонов

*	Таксон						
	Aphidoidea	Coccoidea	Phylloxeroidea	Aleyrodoidae	Cicadoidea	Pentatomoidae	Psylloidea
N	5200	8000	76	1624	2500	5960	4000
T	200	135	75	125	115	135	90
$\Delta \delta$	0,0076	0,0198	0,0149	0,0251	0,0125	0,0115	0,0169
ΔG_{\max}	0,104	0,315	0,153	0,337	0,185	0,190	0,239
G	585,0	1560,0	11,626	402,75	481,50	940,16	598,06
V	0,143	1,173	0,0152	0,2949	0,272	0,508	0,751

* – расшифровка условных обозначений в тексте

Оставляя за рамками работы обсуждение биологического значения рассчитанных величин, а также особенностей эволюционной истории исследуемых таксонов, следует отметить, что данные, получаемые с использованием предлагаемой модели, хорошо согласуются с эмпирическими данными, полученными при анализе последовательностей COI. Принципиальной особенностью предлагаемой модели, как было сказано выше, является возможность предсказать события, происходящие в целом таксоне в отсутствие эмпирических данных. Использование предлагаемой модели позволяет рассчитывать такие показатели, как значения генетических дистанции между видами, не представленными в выборке нуклеотидных последовательностей, накопленная генетическая дистанция и скорость молекулярной эволюции в таксоне.

ЗАКЛЮЧЕНИЕ

Нами разработана математическая модель, описывающая эволюционные события, протекающие на уровне нуклеотидных последовательностей в процессе эволюции целых таксонов. Использование предлагаемой модели позволяет рассчитывать ряд показателей, которые не могут быть рассчитаны по экспериментальным данным.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Telford M. J., Budd G. E., Philippe H. Phylogenomic insights into animal evolution // *Current Biology*. 2015. Vol. 25(19). P. 876–887.
2. From barcodes to genomes: extending the concept of DNA barcoding / E. Coissac [et al.] // *Molecular Ecology*. 2016. Vol. 25. P. 1423–1428.
3. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. / K. Tamura [et al.] // *Mol. Biol. Evol.* 2013. Vol. 30. P. 2725–2729.