

УДК 575.852.113:595.773.4

## ОЦЕНКА РАВНОМЕРНОСТИ ХОДА МОЛЕКУЛЯРНЫХ ЧАСОВ В РОДОСЛОВНЫХ ВИДОВ ДРОЗОФИЛ ГРУППЫ *VIRILIS*

А.М. Куликов, О.Е. Лазебный, Е.Ю. Рыбакова

(Институт биологии развития имени Н.К. Кольцова РАН; e-mail: amkulikov@gmail.com)

Исследована филогения видов-двойников группы *Drosophila virilis* на основе изменчивости пяти ядерных и митохондриальных генов: *Adh*, *NonA*, *Fu*, *Ras1* и *16S-12S* рРНК. Предложена модель для оценки неравномерности накопления замен в филогенетических линиях группы видов по произвольному числу последовательностей ДНК с учетом множественных, параллельных и обратных замен. Реконструированы периоды неравномерного хода молекулярных часов в группе видов *virilis* и построено филогенетическое дерево данной группы на основе нейтральной изменчивости.

**Ключевые слова:** эволюция, молекулярные часы, нейтральная изменчивость, *Drosophila virilis*.

В работах по оценке скорости накопления изменчивости в последовательностях ядерных *Adh*, *NonA*, *Fu*, *Ras1* и *16S-12S* рРНК митохондриальных генов мы показали нарушение хода молекулярных часов в большинстве эволюционных линий группы видов-двойников *Drosophila virilis* [1, 2]. При этом неравномерность хода молекулярных часов была характерна для консервативных последовательностей генов *16S-12S* рРНК и *Ras1*, замены в которых преимущественно нейтральны [3–5]. Ранее с помощью математического моделирования была оценена роль позитивного отбора, действующего на locus [6]: интенсивный направленный отбор должен способствовать более равномерным темпам молекулярной эволюции. Полученные нами данные согласуются с выводами Катлера [6]. Именно локусы, подвергающиеся действию позитивного отбора, могут быть более предпочтительными для проведения филогенетических оценок. Однако направленный отбор по признакам, имеющим выраженный адаптивный характер, может приводить к накоплению конвергентных замен — гомоплазий в случае формирования независимых адаптаций к сходным условиям у видов, не входящих в монофилетическую группу. Очевидно, эта изменчивость связана в значительной степени с несинонимичными заменами, хотя нельзя полностью исключить и накопления гомоплазий в области промотора. Для того чтобы исключить влияние отбора при построении филогений, мы предлагаем использовать исключительно “молчащие” сайты, в том числе 4-кратно вырожденные позиции нуклеотидов и некодирующие последовательности ДНК.

Задачей работы была разработка алгоритма точной оценки равномерности накопления замен в эволюционных линиях сравниваемых видов по про-

извольному числу последовательностей, построение филогенетического древа на основе нейтральной изменчивости на примере дрозофил группы *virilis* и реконструкция периодов нарушения равномерного хода молекулярных часов в эволюционных линиях данной группы видов.

Используемый традиционно в качестве меры скорости накопления замен индекс дисперсии  $R(T)$ , равный частному дисперсии числа замен на их среднее число, не должен достоверно отличаться от единицы в случае нейтральной эволюции. Существуют формальные факторы, приводящие к смещению оценок  $R(T)$ . Это число узлов филогенетического древа или степень отличия конкретной филогении от филогении “звезды”, скорость смены поколений в разных родословных, недоучет параллельных и множественных замен в сравниваемых последовательностях и др. Методы оценки величины взвешенного индекса дисперсии были разработаны Гилеспи [7] и Балмером [8]. Эти методы приложимы к разным исходным данным и учитывают разные факторы. Гилеспи предложил использовать поправку на “эффекты происхождения”, общие для разных локусов одного генома. Балмер ввел поправку — коэффициент инфляции — на чрезмерное увеличение  $R(T)$ , возникающее при использовании сложных моделей, учитывающих множественные замены. Нами был разработан и применен новый метод, объединяющий обе поправки.

**Метод невзвешенного индекса дисперсии ( $R(T)$ ), с поправкой на коэффициент инфляции.** Оценивается отдельно для каждого из  $k$  локусов и каждой пары видов  $i-j$ ,  $R(T)_{i-j \text{ inf lat.}} = (s-2)S_a / (v + (s-4)c)(s-1)$ , где  $v$  и  $c$  — обобщенные оценочные показатели  $\text{Var}(d_{ij})$  и  $\text{Cov}(d_{ij}, d_{ik})$  для каждой  $i$ -й родословной (по Балмеру) соответственно,  $d$  — расстояние между соответствующими эволюционными линиями,

$s$  — число сравниваемых видов,  $S_a = \sum_{i=1}^i (a_i - \bar{a})^2$ , где  $a_i$  — число замен на один нуклеотид в данной родословной,  $\bar{a}$  — среднее число замен на один нуклеотид во всех рассматриваемых эволюционных линиях. Вариационно-ковариационную матрицу этих показателей, как показал Балмер, можно определить с высокой точностью по стандартным формулам сложения дисперсий и ковариаций независимых линейных функций, принимая значения дисперсий и ковариаций расстояний между видами за первичные данные:

$$\begin{aligned} \text{Var}(d_{ij}) &= \text{Var}\left(\sum w_i d_{ij}\right) = \\ &= \sum_{i=1}^i w_i^2 \text{Var}(d_{ij}) + \sum_{i \neq j} w_i w_j \text{Cov}(d_{ij}, d_{ik}), \\ \text{Cov}(d_{ij}, d_{ik}) &= \text{Cov}\left(\sum w_i d_{ij}, \sum u_i d_{ij}\right) = \\ &= \sum_{i=1} w_i u_i \text{Var}(d_{ij}) + \sum_{i \neq j} w_i u_j \text{Cov}(d_{ij}, d_{ik}), \end{aligned}$$

где  $w_i, w_j$  и  $u_i, u_j$  — веса расстояний  $d_{ij}, d_{ik}$  в  $i$ -й родословной.

Стандартные значения весов ( $w_i, w_j$  и  $u_i, u_j$ ) входят в формулы дисперсии и ковариации расстояний для данной эволюционной линии “ $i$ ” с положительным знаком и значением  $w_i = 1/(s - 1)$  для всех расстояний между видом “ $i$ ” и остальными видами и с отрицательным знаком и значением  $w_i = 1/(s - 1)(s - 2)$  для всех остальных расстояний.

Показатель индекса дисперсии для данной эволюционной линии  $R(T)_{i \text{ inflat}}$  определяется как среднее по всем парам, образуемым данным видом, как для одного локуса, так и по всем локусам. Статистика  $\chi^2 = (s - 2)S_a / (v + (s - 4)c)$  определяется для искомого показателя  $R(T)_{i-j \text{ inflat}}$  по индивидуальным значениям дисперсии и ковариации каждой эволюционной линии каждого анализируемого локуса. Пользуясь свойством аддитивности независимых распределений  $\chi^2$ , мы можем получить оценку для усредненных по всем локусам значений  $R(T)_{i-j \text{ inflat}}$ , для всех пар данного вида и внешних по отношению к нему, используя суммы соответствующих статистик  $\chi^2$  и суммы их степеней свободы, при этом необходимо вместо аргумента функции  $x$  использовать аргумент  $nx$ .

**Метод взвешенного  $R(T)$  с поправкой на коэффициент инфляции.** Данный показатель оценивается отдельно для каждого из  $k$  локусов и каждой пары видов  $i-j$ . Веса  $\omega_i$  определяются по средним  $a_i$  для всех  $k$  локусов:

$$\omega_i = s \sum_{k=1}^k a_i / \sum_{i=1}^i \sum_{k=1}^k a_i,$$

так что  $\sum \omega_i = s$ .

В соответствии со свойствами дисперсии и ковариации для произвольных случайных величин  $\xi_1$  и  $\xi_2$  верны равенства:  $\text{Var}(\omega_i^2 \xi) = \omega_i^2 \text{Var}(\xi)$  и  $\text{Cov}(\omega_i \xi_1, \omega_j \xi_2) = \omega_i \omega_j \text{Cov}(\xi_1, \xi_2)$ . Учитывая уже определенную вариационно-ковариационную матрицу, полученную для невзвешенных значений среднего числа замен на один нуклеотид  $a_i$ , воспользуемся свойствами дисперсии и ковариации и разделим значения матрицы на соответствующие им весовые коэффициенты. Соответственно получаем:

$$\begin{aligned} R(T)_{i-j \text{ inf lat., weight}} &= \\ &= (s - 2)S_a \text{ weight} / (v/\omega_i^2 + (s - 4)(c/\omega_i \omega_j))(s - 1), \end{aligned}$$

где  $S_a \text{ weight} = \sum_{i=1}^i (a_i/\omega_i - \bar{a}/\omega_i)$ ,

$$\chi^2 = (s - 2)S_a \text{ weight} / ((v/\omega_i^2) + (s - 4)(c/\omega_i \omega_j)).$$

Индекс дисперсии для  $i$ -й эволюционной линии  $R(T)_{i \text{ inflat., weight}}$  также определяется как среднее по всем парам, образуемым данным видом.

Результаты применения предложенных алгоритмов для оценки индекса дисперсии по последовательностям пяти генов в эволюционных линиях видов группы *virilis* подтвердил сделанные ранее выводы. Высокополиморфный локус *Adh*, связанный с пищевыми адаптациями к бродящим субстратам и находящийся под действием положительного отбора, имеет наименьший индекс дисперсии, а консервативная последовательность *12S-16S* рРНК генов мтДНК — наибольший. Среди ядерных генов наибольшее значение получено для высококонсервативного и обладающего исключительно низким популяционным полиморфизмом гена *Ras1*. Результаты анализа генетических расстояний, рассчитанных по заменам в нейтральных позициях общей последовательности пяти исследуемых генов (4-кратно вырожденные позиции нуклеотидов кодирующих последовательностей и некодирующие последовательности ДНК) для пар эволюционных линий представлены на рис. 1.

Виды *D. borealis*, *D. montana*, *D. lacicola* и *D. flavomontana* субфилады *montana* имеют наиболее низкие и сходные между собой показатели  $R(T)$ . Это свидетельствует о сходной равномерной скорости молекулярной эволюции в эволюционных линиях этих видов. При этом *D. flavomontana* имеет самые низкие значения  $R(T)$ . Значения  $R(T)$  в парах, включающих вид *D. virilis*, являются максимальными, что предполагает наиболее существенные нарушения равномерности хода молекулярных часов в эволюционной линии *D. virilis*. Остальные виды данной филады *virilis* также имеют более высокие индексы дисперсии по сравнению с видами из другой филады. Это свидетельствует о значительных изменениях скорости молекулярной эволюции на всех этапах расхождения видов филады *virilis*. При срав-

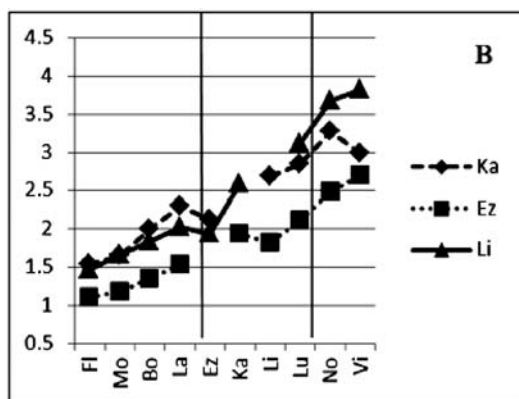
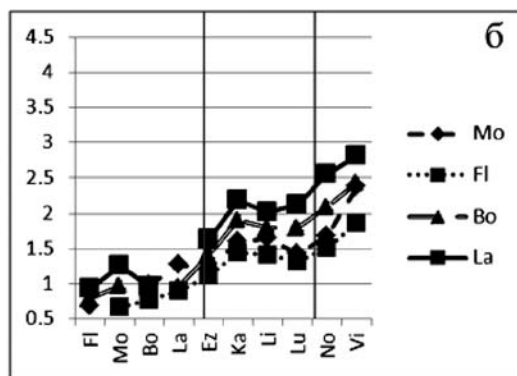
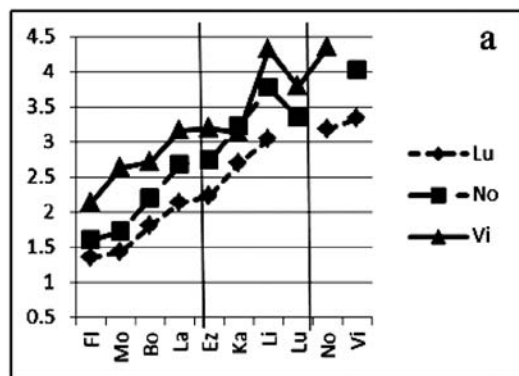


Рис. 1. Изменения  $R(T)_{i-j}$  *inflat.*, *weight* у видов, относящихся к одной филаде или субфиладе: а — эволюционные линии видов филады *virilis* (*D. virilis*, *D. lummei*, *D. novamexicana*); б — эволюционные линии видов субфилады *montana* филады *montana* (*D. montana*, *D. flavomontana*, *D. laticola*, *D. borealis*); в — эволюционные линии видов субфилады *kanekoi* филады *montana* (*D. kanekoi*, *D. ezoana*, *D. littoralis*). Виды обозначены двумя первыми буквами видового названия. Индексы  $R(T)$  рассчитаны по заменам в нейтральных сайтах (4-кратно вырожденные позиции нуклеотидов кодирующих последовательностей и некодирующие последовательности ДНК) общей последовательности пяти ядерных и мито-

нени представителей филады *virilis* и субфилады *montana* получены минимальные значения  $R(T)$ . Это означает, что на ранних этапах дивергенции от общего предка молекулярные часы шли более равномерно, чем на поздних этапах. Виды — представители субфилады *kanekoi* (*D. kanekoi*, *D. ezoana* и *D. littoralis*) — имеют более сложную картину распределения индекса дисперсии  $R(T)$ . Минимальные значения показателя характерны для *D. ezoana*. Виды *D. littoralis* и *D. kanekoi* при сравнении между собой демонстрируют промежуточное значение  $R(T) \approx 2,5$ , а при сравнении с представителями других субфилад — противоположные значения. Так, *D. littoralis* имеет показатели  $3,12 < R(T) < 3,82$  при сравнении с представителями филады *virilis*, что существенно выше аналогичных показателей *D. kanekoi* — от 2,8 до 3,3. Наоборот, при сравнении этих видов с представителями субфилады *montana* значения  $R(T)$  существенно ниже и близки для обоих видов (максимальные значения — 2,0 для *D. littoralis* и 2,3 для *D. kanekoi*).

Учитывая наиболее вероятную топологию дерева (рис. 2), построенного методом максимальной экономии по общей последовательности всех пяти генов с учетом изменчивости, близкой к нейтральной (4-кратно вырожденные позиции нуклеотидов кодирующих последовательностей и некодирующие последовательности ДНК), можно сделать следующие выводы. Действие отбора на изученные последовательности исключено в силу использованного близкой к нейтральной изменчивости и весовых

коэффициентов Гилеспи, поэтому неравномерная скорость накопления замен может быть связана с влиянием генетико-автоматических процессов. Одни и те же климатические условия могли по-разному влиять на виды-двойники дрозофил группы *virilis* в силу уже сложившихся адаптаций, времени распространения по ареалу и различной эффективно-

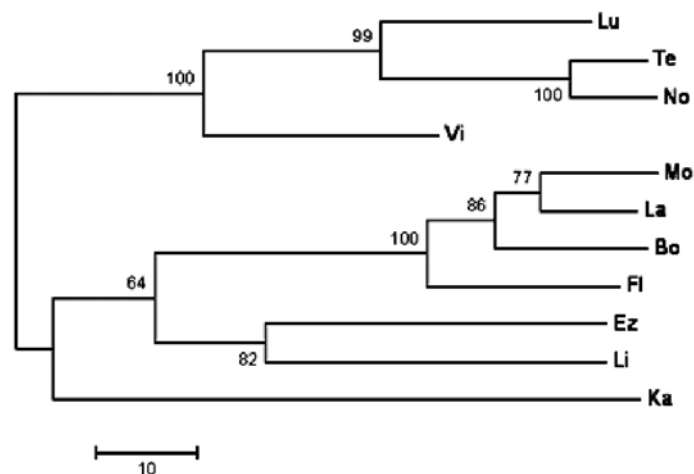


Рис. 2. Дендрограмма генетических расстояний, полученных по общей последовательности генов: *Adh*, *NonA*, *Fu*, *Ras1* и *16S-12S* рРНК с учетом изменчивости, близкой к нейтральной (4-кратно вырожденные позиции нуклеотидов кодирующих последовательностей и некодирующие последовательности ДНК), между видами-двойниками группы *virilis*, построенная методом максимальной экономии. Числа в узлах дерева обозначают индексы бутстреп-поддержки. Те — хромосомная форма *texana* вида *Drosophila americana* — одного из видов-двойников группы *virilis*, входящего в одноименную филаду

сти действия генетико-автоматических процессов в популяциях исследуемых видов. Например, разные виды группы *virilis* могли проходить через разное количество “бутылочных горлышек”.

\* \* \*

Работа поддержана Российским фондом фундаментальных исследований (грант № 09-04-01745).

## СПИСОК ЛИТЕРАТУРЫ

1. Куликов А.М., Лазебный О.Е., Горностаев Н.Г., Чекунова А.И., Митрофанов В.Г. Неравномерность скорости эволюции у дрозофил группы *virilis*. I. Применение филогенетически опосредованных тестов Такезаки // Изв. РАН. Сер. биол. 2010. Т. 37. № 1. С. 25—33.
2. Куликов А.М., Лазебный О.Е., Горностаев Н.Г., Чекунова А.И., Митрофанов В.Г. Неравномерность скорости эволюции у дрозофил группы *virilis*. II. Применение метода тестирования Таджми // Изв. РАН. Сер. биол. 2010. Т. 37. № 1. С. 206—209..
3. Gasperini R., Gibson G.J. Absence of protein polymorphism in the *Ras* genes of *Drosophila melanogaster* // Mol. Evol. 1999. Vol. 49. N 5. P. 583—590.
4. Сорокина С.Ю., Мюге Н.С., Андрианов Б.В., Митрофанов В.Г. Изменчивость 3'-концевого фрагмента гена

16S рРНК в группе бизкородственных видов дрозофил *virilis* // Генетика. 2005. Т. 41. № 8. С. 1049—1054.

5. Чекунова А.И., Куликов А.М., Михайловский С.С., Лазебный О.Е., Лазебная И.В., Митрофанов В.Г. Родственные отношения дрозофил группы *virilis*, рекомендованные на основе последовательностей гена *Drasl* // Генетика. 2008. Т. 44. № 3. С. 336—345.

6. Cutler D.J. Understanding the overdispersed molecular clock // Genetics. 2000. Vol. 154. P. 1403—1417.

7. Gillespie J.H. Lineage effects and the index of dispersion of molecular evolution // Mol. Biol. Evol. 1989. Vol. 6. N 5. P. 636—647.

8. Bulmer M. Estimating the variability of substitution rates // Genetics 1989. Vol. 123. P. 615—619.

Поступила в редакцию  
18.04.10.

#### ESTIMATION OF THE RATE OF MOLECULAR CLOCK IN LINEAGES OF THE *DROSOPHILA VIRILIS* SPECIES GROUP

A.M. Kulikov, O.E. Lazebny, E.Yu. Rybakova

Phylogeny of the *Drosophila virilis* species group is studied on the base of five nuclear and mitochondrial genes: *Adh*, *NonA*, *Fu*, *Ras1* и *16S-12S rRNA*. Mathematical simulation to estimate uneven accumulation of replacements in lineages of closely related species using arbitrary number of sequences was suggested. It accounts for multiple, parallel and reverse mutations. Periods of evolution with irregular rates in the *virilis* species group were reconstructed. Phylogeny tree for the group given on the base of neutral variability was constructed.

**Key words:** *evolution, molecular clock, neutral variability, Drosophila virilis.*

#### Сведения об авторах

Куликов Алексей Михайлович — канд. биол. наук, ст. науч. сотр., Учреждение Российской академии наук Институт биологии развития имени Н.К. Кольцова РАН. Тел. (499)135-87-81; e-mail: amkulikov@gmail.com

Лазебный Олег Евгеньевич — канд. биол. наук, ст. науч. сотр., Учреждение Российской академии наук Институт биологии развития имени Н.К. Кольцова РАН. Тел. (499)135-87-81; e-mail: oelazebny@gmail.com

Рыбакова Елена Юрьевна — аспирантка, Учреждение Российской академии наук Институт биологии развития имени Н.К. Кольцова РАН. Тел. (499)135-87-81.