

Эволюционный конструктор:
методика и комплекс программ для моделирования эволюции
трофически связанных популяций одноклеточных гаплоидных
организмов*

С. А. ЛАШИН, Н. А. КОЛЧАНОВ, Ю. Г. МАТУШКИН
Институт цитологии и генетики СО РАН, Новосибирск, Россия
e-mail: lashin@bionet.nsc.ru

Methodology for co-evolution's modelling of food web-like related populations of unicellular haploid organisms is presented. The corresponding software package "Evolutionary Constructor" is developed, that enables modelling of processes requiring immediate modifications of the model structure during calculations.

Введение

Математические модели биологических систем и, в частности, модели динамики популяций и генетической изменчивости, основанные на развитом математическом аппарате, представляют собой гибкое средство для исследования различных аспектов формирования биоразнообразия и процессов эволюции. Использование математических и компьютерных моделей позволяет привлечь во внимание различные формы трофических взаимодействий, пространственные формы распределения организмов, взаимодействие индивидуальных организмов, различные схемы размножения. Кроме того, процессы макро- и микроэволюции, протекающие на интервалах от тысяч до миллионов лет, принципиально не могут быть исследованы экспериментально, и роль моделирования здесь особенно велика. Поэтому разработка реалистичных многопараметрических моделей эволюции и видообразования, учитывающих влияние условий окружающей среды, трофических взаимодействий, географических факторов, генетической структуры организмов и прочих параметров, — одна из основных задач для биологии, математики и информатики XXI века.

Традиционные подходы к моделированию эволюционно-популяционных процессов включают в себя методы моделирования динамики популяций [1] и методы популяционной генетики [2, 3]. Методы моделирования динамики популяций описывают изменение численности популяций во времени с учетом условий окружающей среды, трофических взаимодействий популяций и других факторов. Большая часть этих методов основана на теории динамических систем [4, 5], а с их помощью, как правило, нельзя исследовать

*Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант № 06-04-49556), Программы президиума РАН "Происхождение и эволюция биосферы", проекта "Системная биология: компьютерно-экспериментальные подходы" Программы президиума РАН "Молекулярная и клеточная биология".

© Институт вычислительных технологий Сибирского отделения Российской академии наук, 2008.

изменение генетической структуры популяции. Что касается методов популяционной генетики, базирующихся в основном на теории вероятностей и математической статистике [6–8], то они, напротив, позволяют исследовать эволюцию генетической структуры популяции, но не дают возможности детально моделировать процессы популяционной динамики. Дальнейшее развитие методов эволюционно-популяционного моделирования привело к появлению “гибридных” методик моделирования (см., например, [9–11]), с которыми удастся одновременно исследовать изменение как численности, так и генетической структуры популяции. Однако подавляющее большинство этих методов (как и остальных методов генетики популяций) ориентировано на моделирование популяций диплоидных организмов (как правило, с половым размножением). Методы “портретного” индивидуально-ориентированного моделирования требовательны к размеру оперативной памяти, а также имеют большую вычислительную сложность. Это связано с использованием многочисленных объектов, так как для описания каждой особи необходим “свой” объект [12]. И если для моделирования популяций диплоидных организмов с половым размножением мощностей современных компьютеров достаточно (ввиду того, что, как правило, эффективная численность таких популяций не превышает 100–1000 особей), то для моделирования популяций гаплоидных организмов (в частности бактерий) имитационное моделирование “в лоб” зачастую проблематично (эффективная численность бактериальных популяций составляет 10^6 – 10^9 особей).

В данной статье мы предлагаем новую методику моделирования коэволюции популяций одноклеточных *гаплоидных организмов (клеток)* — “Эволюционный конструктор” (ЭК). Методика позволяет моделировать функционирование сетей популяций, трофически связанных между собой субстрат-продуктными отношениями, при влиянии окружающей среды. В процессе эволюции могут меняться размер популяции и ее генетическая структура — как под действием отбора, так и под действием мутаций. Отличительная особенность предложенной методики — возможность моделирования таких эволюционно-популяционных процессов, которые требуют перестройки структуры модели в процессе ее расчета. К таким процессам относятся, например, *горизонтальный перенос генетического материала* и, как следствие, *видообразование*, которые могут изменять в модели как количество переменных, так и количество уравнений. Также разработанная в рамках методики арифметика *генетических спектров* позволяет изменять степень подробности описания полиморфизма одного или нескольких генов в популяции, в том числе и непосредственно в процессе расчета модели.

1. Описание методики “Эволюционный конструктор”

Методика ЭК предназначена для моделирования сосуществования популяций трофически связанных одноклеточных гаплоидных организмов (клеток). На рис. 1 представлена схема основных объектов и процессов ЭК. Мы моделируем сети трофически связанных гаплоидных организмов, которые объединены в популяции по принципу генетической близости и которые находятся в едином объеме, называемом *окружающей средой*. Организмы могут потреблять (*утилизировать*) субстраты и синтезировать, а затем секретировать в окружающую среду продукты, которые, в свою очередь, могут потребляться другими организмами в качестве субстратов. Одни субстраты оказывают положительный эффект на рост популяции, другие, наоборот, могут оказывать ингибирующее действие. Эффективность утилизации субстратов и производства (*синтеза*) продуктов регулируется соответствующими генами. Кроме того, существуют так на-

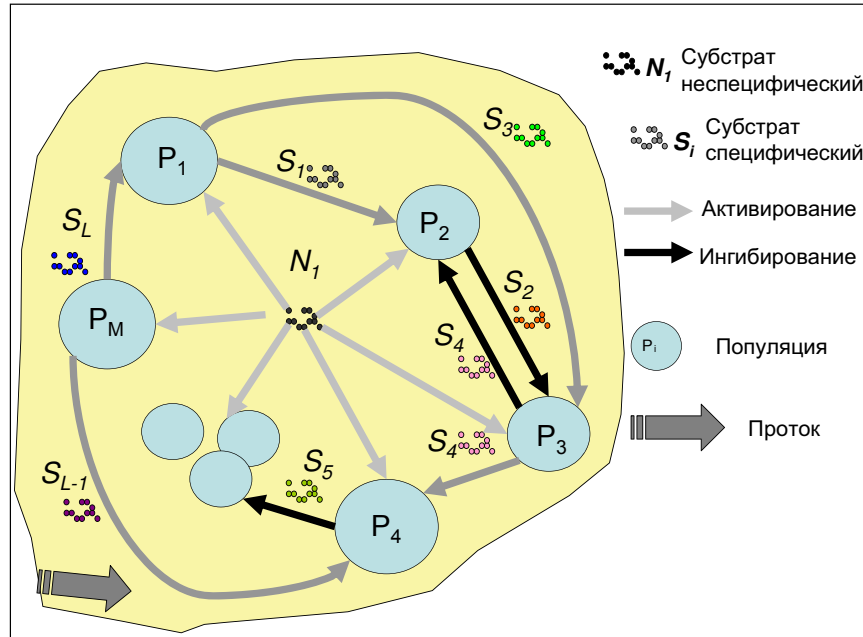


Рис. 1. Схема объектов и процессов эволюционного конструктора (в окружающей среде)

зываемые *неспецифическими* общие субстраты (на рис. 1 представлена схема с одним неспецифическим субстратом N_1), концентрация которых в окружающей среде регулируется протоком и присутствие которых необходимо для всех организмов трофической сети. В противоположность неспецифическим субстратам, субстраты, которые поступают в окружающую среду только благодаря деятельности клеток (синтезу и секреции), мы называем *специфическими* (на рис. 1 обозначены S_i).

Ниже приведено описание объектов и процессов ЭК.

1.1. Окружающая среда

Окружающая среда представляет собой замкнутую проточную систему фиксированного объема V_{total} . С окружающей средой связаны процессы *притока* субстратов в окружающую среду и *оттока* из окружающей среды как субстратов, так и клеток. Окружающая среда характеризуется следующими величинами:

V_{total} — объем окружающей среды, л;

k_{flow} — скорость протока, % от V_{total} за единицу времени;

$N_{env,i}$ — концентрация i -го неспецифического субстрата в окружающей среде;

$S_{env,i}$ — концентрация i -го специфического субстрата в окружающей среде;

$N_{flow,i}$ — концентрация i -го неспецифического субстрата во входящем протоке (в притоке).

В окружающей среде происходят следующие процессы:

— приток субстратов в окружающую среду — концентрация неспецифических субстратов в окружающей среде увеличивается в зависимости от скорости протока и концентрации соответствующих субстратов в протоке;

- отток субстратов из окружающей среды — концентрация как неспецифических, так и специфических субстратов уменьшается в зависимости от скорости потока;
- приток/отток неспецифических субстратов — описывается формулой

$$N_{env,i}(t+1) = N_{env,i}(t) + k_{flow}V_{total}(N_{flow,i} - N_{env,i}(t)); \quad (1)$$

- отток специфических субстратов — описывается следующей формулой (поступление специфических субстратов в окружающую среду связано с синтезом и секрецией этих субстратов клетками популяций и описывается далее):

$$S_{env,i}(t+1) = S_{env,i}(t)(1 - k_{flow}V_{total}). \quad (2)$$

1.2. Мономорфная популяция

Под *мономорфной* популяцией мы понимаем популяцию гаплоидных организмов:

- 1) которые потребляют одинаковое множество неспецифических и специфических субстратов;
- 2) которые продуцируют одинаковое множество специфических продуктов;
- 3) у которых совпадают как эффективности процессов утилизации, так и эффективности процессов синтеза.

Определим основные термины и понятия, которые будут использоваться далее в тексте.

Признак — константа скорости синтеза или утилизации какого-либо определенного субстрата; считается, что каждый признак однозначно определяется одним геном.

Генотип особи — представляет собой набор констант-признаков (генов), разделенных на три группы (вектора). Первая группа (c_i) характеризует эффективность утилизации специфических субстратов (s_i), вторая группа (d_i) — скорость выработки субстратов, третья группа (r_i) — эффективность утилизации неспецифических субстратов (n_i).

Мутация в данном случае — это изменение величины соответствующей константы признака.

В этих терминах можно считать особи мономорфной популяции генетически идентичными. И общий для всех особей такой популяции генотип назовем *генотипом мономорфной популяции*.

Мономорфная популяция однозначно определяется своим генотипом и размером (который может означать количество клеток в популяции или их концентрацию). Для того чтобы особи смогли *утилизировать* (использовать) субстраты для роста своей численности и производства продуктов, субстраты должны быть *поглощены* особями из окружающей среды. При недостатке в среде какого-либо субстрата может возникать конкуренция за этот субстрат, как внутрипопуляционная, так и межпопуляционная (когда особи сразу нескольких популяций могут потреблять один и тот же субстрат). При избытке субстрата его максимальное количество, поглощенное одной особью, определяется величиной *нормы потребления субстрата*. Данная величина видоспецифична и не может быть изменена вследствие мутации (в текущей версии методики). Таким образом, мономорфная популяция дополнительно характеризуется количеством *поглощенных* (потребленных) молекул субстратов. Примеры формул изменения численности популяции за одно поколение, в зависимости от численности популяции, количества

потребленных субстратов, скорости протока и коэффициента смертности, приведены ниже (легко могут быть использованы и другие формулы):

$$F_1(n_0, r_0, \mathbf{S}, \mathbf{C}, P) = \Delta P = \sqrt{r_0 n_0(P) \sum_{i=1}^N c_i s_i(P) - k_{flow} P - k_{death} P^2}; \quad (3)$$

$$F_2(n_0, r_0, \mathbf{S}, \mathbf{C}, P) = \Delta P = a_{basal}(n_0) P - \sqrt{\sum_{i=1}^N c_i s_i(P) - k_{flow} P - k_{death} P^2}, \quad (4)$$

где n_0 — количество единственного неспецифического субстрата, потребленного клетками популяции из среды (пропорционально размеру популяции); r_0 — скорость утилизации единственного неспецифического субстрата; \mathbf{S} — вектор значений специфических субстратов, потребленных клетками популяции из среды (пропорционально размеру популяции); \mathbf{C} — вектор скоростей утилизации соответствующих специфических субстратов; P — размер популяции; k_{flow} — скорость протока в окружающей среде (скорость вымывания); k_{death} — коэффициент смертности популяции; a_{basal} — “естественный прирост” популяции.

Формула (3) описывает простейший активаторный эффект специфических субстратов. Неспецифический субстрат при этом еще и необходим (при его отсутствии прирост популяции равен нулю). Формула (4) используется для моделирования ингибирования специфическими субстратами. Неспецифический субстрат при этом также необходим. Другие формулы изменения численности мономорфной популяции можно посмотреть, например, в [13].

1.3. Полиморфная популяция

Аналогично понятию мономорфной популяции мы определяем понятие *полиморфной популяции*. Полиморфную популяцию можно рассматривать как набор мономорфных субпопуляций (мономорфная субпопуляция может состоять из одной клетки или нескольких идентичных клеток). Полиморфная популяция характеризуется *обобщенным геномом* — набором *генетических спектров* популяции. Генетический спектр представляет собой распределение частот встречаемости признака в популяции (для одного гена) (рис. 2, а). Таким образом, генотипы особей полиморфной популяции представляются аналогом многомерного распределения, реализованного в виде вектора генетических спектров. *Мутация* в терминах генетического спектра означает изменение профиля генетического спектра (рис. 2, б).

Также мы вводим понятие *граничного значения признака* в популяции. Признак считается отсутствующим у той части популяции, которая соответствует части генетического спектра, меньшей граничного значения (например, на рис. 2, в, считается, что у субпопуляции, соответствующей части генетического спектра левее граничного значения, отсутствует данный признак). Понятие граничного значения признака дает удобный способ разделения единой сверхпопуляции на субпопуляции. Клетки считаются относящимися к одной популяции, если совпадают их множества присутствующих и отсутствующих признаков (с учетом граничных значений). В зависимости от ситуации, можно считать присутствие/отсутствие определенных признаков несущественным и рассматривать объединения соответствующих субпопуляций как единые популяции.

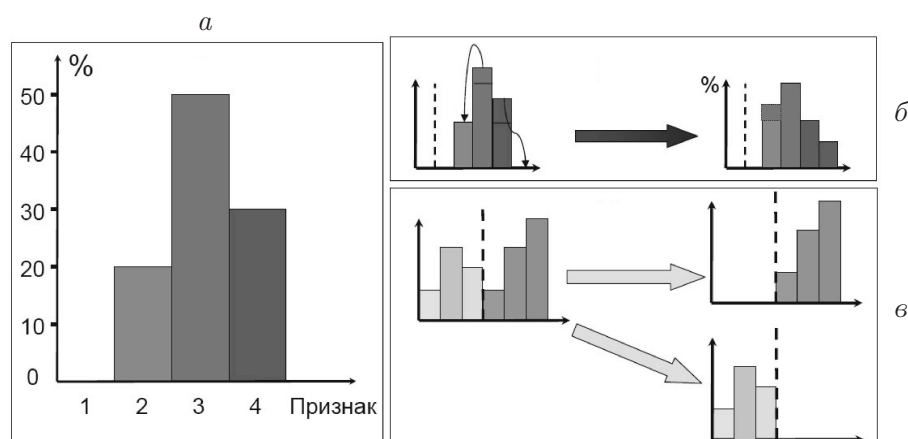


Рис. 2. Генетический спектр (а): у 20 % особей данной популяции значение признака равно 2, у 50 % — 3, у 30 % — 4; изменение генетического спектра популяции вследствие мутаций (б); расщепление популяции по граничному значению признака в генетическом спектре: образование двух спектров, однородных относительно граничного значения (в)

На рис. 2, в дан пример разделения популяции на две субпопуляции по граничному значению признака. В случае частичного перехода спектра за граничное значение возможен переход части клеток в другую популяцию — расщепление популяции. При этом, если до расщепления в системе уже была популяция с подобным спектром (присутствующих/отсутствующих признаков), то субпопуляция “присоединяется” к ней, если же такой популяции не было, то создается новая популяция. На данной особенности методики основывается возможность моделирования *горизонтального переноса генетического материала*. Преимущество такого подхода — в большой гибкости модели: в процессе расчета можно фактически менять как количество уравнений, так и количество переменных системы; также можно варьировать степень подробности описания полиморфизма генетическим спектром (количество аллелей).

1.4. Изменение численности полиморфной популяции

Прирост численности полиморфной популяции рассчитывается по следующей схеме. Полиморфная популяция “расщепляется” на множество мономорфных субпопуляций — доля каждого генотипа мономорфной субпопуляции зависит от вектора генетических спектров популяции. После этого для каждой мономорфной субпопуляции рассчитывается изменение численности по формулам, подобным приведенным выше (3) и (4). Затем происходит “слияние” измененных мономорфных субпопуляций в полиморфную популяцию (при слиянии вид генетических спектров популяции может измениться ввиду того, что эффективность потребления субстратов и, как следствие, эффективность размножения у клеток разных генотипов различается).

1.5. Синтез продуктов

Синтез продуктов клетками полиморфной популяции вычисляется по следующей интегральной формуле:

$$\Delta s_i = P \sum_{j \in \text{Spectr}_i} d_{ij} \left(\frac{P_j}{P} \right), \quad (5)$$

где Δs_i — количество синтезируемого субстрата типа i ; d_{ij} — значение признака в генетическом спектре Spectr_i ; P — размер популяции; P_j — доля особей со значением признака d_{ij} (в генетическом спектре).

1.6. Секреция продуктов популяциями в окружающую среду

Синтезированные продукты секретируются в окружающую среду и в дальнейшем могут быть использованы другими популяциями в качестве субстратов либо будут унесены из среды протоком.

1.7. Поглощение субстратов популяциями из окружающей среды

Популяции поглощают субстраты из окружающей среды. При избытке субстратов их потребление ограничивается *нормами потребления* соответствующих субстратов. Если же каких-то субстратов в окружающей среде недостаточно, то популяции конкурируют друг с другом за эти субстраты.

1.8. “Метаболизм”

Когда клетка синтезирует продукт, который может быть утилизирован ею самой, очевидно, что ей “нет смысла” секретировать этот продукт в окружающую среду и затем участвовать в конкуренции за него наравне с другими популяциями “на общих основаниях”. Поэтому при моделировании *внутренних* субстратов популяций мы рассматриваем две формы (состояния) субстратов. Первая форма — это субстраты, “готовые к утилизации”, вторая форма — это синтезированные субстраты. Субстраты первой группы восполняются за счет поглощения субстратов из окружающей среды и за счет перехода из синтезированной формы (если клетка может их синтезировать). При

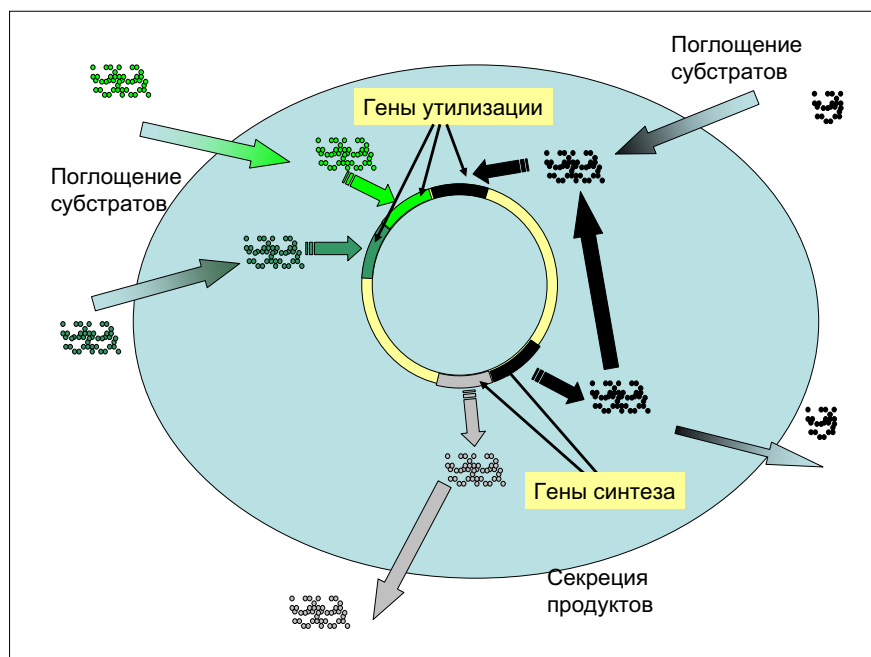


Рис. 3. Схема трофических процессов на примере одной клетки

размножении и синтезе продуктов субстраты первой группы утилизируются (расходуется). Субстраты второй группы восполняются только за счет синтеза. Если клетка может потреблять синтезированный субстрат, то часть этого субстрата (либо все количество — это зависит от потребностей клетки в данном субстрате) переходит в первую группу. Оставшаяся часть секретируется в окружающую среду. Принципиальная схема поглощения—утилизации—синтеза—секреции показана на рис. 3: клетка утилизирует три типа субстратов и синтезирует два типа субстратов, один из которых может утилизироваться самой клеткой. Синтезированный субстрат, который не может быть утилизирован клеткой (на рис. 3 светло-серый) в полном объеме секретируется в окружающую среду, тогда как второй субстрат (черный) — частично используется самой же клеткой (соответственно, в среду секретируется меньше либо вообще ничего не секретируется).

1.9. Итерационный процесс

Последовательность процессов, происходящих за один итерационный шаг, показана на рис. 4. Как видно, на каждой итерации поочередно происходят процессы поглощения

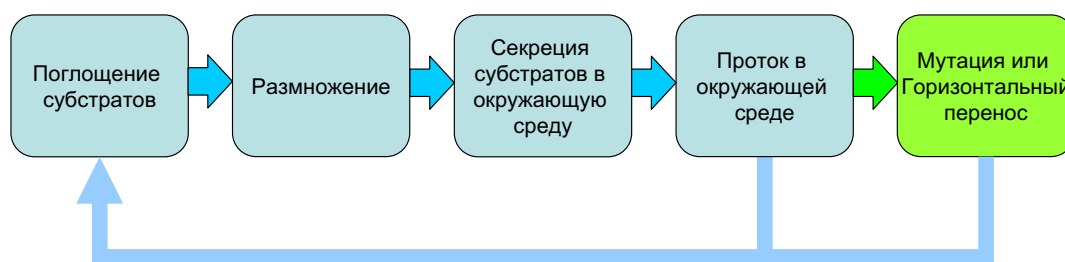


Рис. 4. Блок-схема одного итерационного шага

субстратов из окружающей среды (с учетом конкуренции), утилизации поглощенных субстратов в процессе размножения и синтеза продуктов, секреции синтезированных продуктов (или их части) в окружающую среду и действия протока в среде. Процессы мутации и горизонтального переноса генетического материала не обязательны на каждой итерации, а происходят лишь время от времени. Моменты этих событий могут задаваться изначально или генерироваться случайно.

2. Результаты моделирования

Описанная выше методика моделирования реализована в виде программного комплекса “Эволюционный конструктор” [13]. С помощью программы проверена адекватность методики ЭК. Рассчитаны модели популяции, питающейся одним субстратом в протоке (ср. с моделью Ферхюльста [1, 4]), модель конкуренции двух популяций, питающихся одним субстратом, а также другие модели динамики популяций. Кроме того, рассчитаны модели изменения генетического состава популяции при возникновении положительных мутаций в разных питательных условиях (рис. 5). Исследованные модели подтвердили адекватность методики моделирования ЭК.

Кроме базовых биологических моделей были исследованы более сложные модели функционирования трофических сетей одноклеточных организмов, получен ряд ориги-

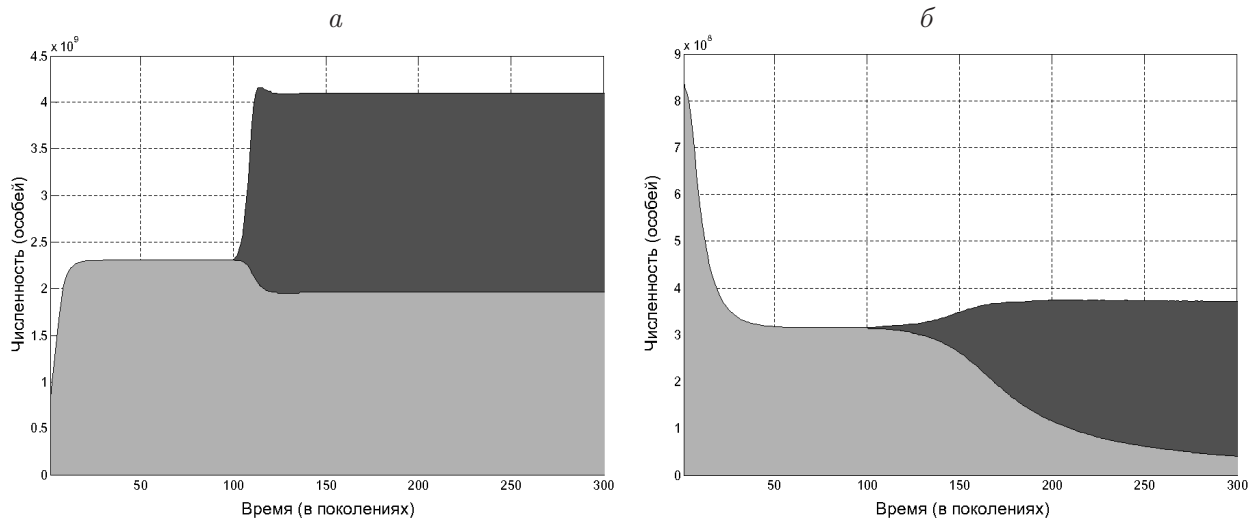


Рис. 5. Появление нового, более приспособленного, аллеля в популяции вследствие мутации в умеренных условиях окружающей среды (*a*), приводящее к незначительной внутривидовой конкуренции. Доля субпопуляции дикого типа (светлая область на рисунке) — того же порядка, что и доля мутантного типа (темная область). Появление нового, более приспособленного, аллеля в популяции вследствие мутации в жестких условиях окружающей среды (дефицит субстрата, *b*), приводящее к вытеснению дикого аллеля (светлая область на рисунке) мутантным (темная область)

нальных биологических результатов. Были исследованы трофические сети популяций-симбионтов, активирующих рост друг друга (производящих специфические субстраты, положительно влияющие на рост клеток другого вида). Также были исследованы сети популяций-антагонистов, ингибирующих друг друга. Была показана зависимость жизнеспособности ингибирующих сетей от типа субстратного ингибирования, существующего между участниками трофической сети — мутации, происходившие в этих сетях (пребывавших до мутации в состоянии равновесия), могли приводить к колебаниям численности популяций, как затухающим, так и незатухающим, и даже к хаотическим режимам, в зависимости от типа ингибирования. Было промоделировано влияние мутаций на выживаемость популяций в очень жестких условиях окружающей среды (в условиях дефицита большинства субстратов). Рассматривались два типа моделей, различающиеся стратегиями питания популяций. В одной случае питание было компенсаторным — отсутствие одного или нескольких специфических субстратов могло компенсироваться избыточным присутствием других субстратов. В другом случае питание было некомпенсаторным — отсутствие субстратов не компенсировалось другими и, соответственно, являлось критическим для популяций. Моделирование показало, что в “тяжелых” условиях внешней среды в долгосрочной перспективе эволюционное преимущество имеют трофические кольцевые системы популяций с компенсаторным типом питания. В таких системах “вовремя возникшие” мутации, усиливающие приспособленность особей (например, за счет увеличения эффективности потребления какого-либо субстрата), могут предотвращать как гибель самой популяции-мутанта, так и гибель остальных членов трофического кольца или, по крайней мере, значительно отдалить момент гибели, что в изменяющихся условиях окружающей среды может стать фактором выживания.

Была показана эволюционная значимость горизонтального переноса генетического

материала в трофических сетях активирующих популяций, особенно в изменяющихся условиях окружающей среды [14]. Был промоделирован процесс автономизации бактериальных популяций — образования в процессе эволюции новых популяций (видов), инкапсулирующих “внешний метаболизм” (в виде симбиотических трофических связей в рамках сети популяций) во “внутренний метаболизм” (метаболизм в истинном смысле значения). Показано, что этот процесс значительно оптимизирует общую пищевую приспособленность трофической системы.

Заключение

Разработанные методика моделирования и комплекс программ “Эволюционный конструктор” позволяют адекватно моделировать генетико-популяционные процессы с переменной численностью генов, аллелей, признаков и размеров популяций одноклеточных гаплоидных организмов. С помощью ЭК получены биологически значимые результаты.

Авторы благодарят Виталия Александровича Лихошвая и Валентина Владимировича Суслова за плодотворную дискуссию.

Список литературы

- [1] БАЗЫКИН А.Д. Нелинейная динамика взаимодействующих популяций. М.; Ижевск: Ин-т компьютерных исследований. 2003. 368 с.
- [2] ЛИ Ч. Введение в популяционную генетику. М.: Мир. 1978. 555 с.
- [3] ХЕДРИК Ф. Генетика популяций. М.: Техносфера. 2003. 592 с.
- [4] РИЗНИЧЕНКО Г.Ю. Математические модели в биофизике и экологии. М.; Ижевск: Ин-т компьютерных исследований. 2003. 184 с.
- [5] ФУРСОВА П.В., ЛЕВИЧ А.П. Математическое моделирование экологии сообществ // Проблемы окружающей среды (обзорная информация ВИНТИ). 2002. № 9.
- [6] BALLOUX F. EASYPOP (version 1.7). A computer program for the simulation of population genetics // J. Heredity. 2001. Vol. 92. P. 301–302.
- [7] ЕФРЕМОВ В.В. Равновесие между генетическим дрейфом и миграцией при разных величинах мутационного процесса: анализ при помощи имитационного моделирования // Генетика. 2005. Т. 41, № 9. С. 1283–1288.
- [8] ЖДАНОВА О.Л., ФРИСМАН Е.Я. Динамические режимы в модели однолокусного плотностно-зависимого отбора // Генетика. 2005. Т. 41, № 11. С. 1575–1584.
- [9] МОСКАЛЕЙЧИК Ф.Ф. Взаимосвязь воспроизводительной способности подразделенной популяции и ее генетической структуры // Генетика. 2005. Т. 41, № 10. С. 1419–1427.
- [10] DIECKMANN U., DOEBELI M. On the origin of species by sympatric speciation // Letters to Nature. 1999. Vol. 400. P. 354–357.
- [11] СЕМОВСКИЙ С.В., БУКИН Ю.С., ЩЕРБАКОВ Д.Ю. Видообразование в одномерной популяции: адаптивная динамика и нейтральная эволюция // Электронный журнал “Исследовано в России”. 2002. С. 1385–1396.
- [12] ШЕННОН Р. Имитационное моделирование систем — искусство и наука. М.: Мир. 1978. 420 с.

- [13] LASHIN S.A., SUSLOV V.V., KOLCHANOV N.A., MATUSHKIN YU.G. Simulation of coevolution in community by using the “Evolutionary Constructor” program // *Silico Biology*. 2007. Vol. 7, N 3. P. 261–275.
- [14] LASHIN S.A., LIKHOSHVAI V.A., KOLCHANOV N.A., MATUSHKIN YU.G. Modeling of horizontal gene transfer in prokaryotic populations with the “Evolutionary Constructor” program package // *Proc. of the Fifth Intern. Conf. on Bioinformatics of Genome Regulation and Structure (BGRS-2006)*. 2006. Vol. 3. P. 188–191.

Поступила в редакцию 8 апреля 2008 г.