

К АНАЛИЗУ СЛУЧАЙНЫХ ПРОЦЕССОВ ИЗОНИМИИ. I. СТРУКТУРА ИЗОНИМИИ

© 2021 г. В. П. Пасеков*

*Вычислительный центр им. А.А. Дородницына Федерального исследовательского центра
“Информатика и управление” Российской академии наук, Москва, 119991 Россия*

**e-mail: pass40@mail.ru*

Поступила в редакцию 03.02.2021 г.

После доработки 16.04.2021 г.

Принята к публикации 27.04.2021 г.

Анализируются закономерности случайной дивергенции фамильного состава популяции ограниченного размера, приближенно рассматриваемой в дискретном времени с неперекрывающимися поколениями. Предполагается, что фамилии передаются потомкам патрилинейно. На относительно малом промежутке времени $t/NE(t)$, где $NE(t)$ – средний гармонический эффективный размер популяции за t поколений, систематическими давлениями можно пренебречь. Фамильный состав очередного поколения t формально является случайной выборкой размера $Ne(t)/2$ из фамилий мужской составляющей родительской популяции, размера в 4 раза меньшего, чем количество гамет $2Ne(t)$, формирующих новое поколение, $Ne(t)$ – эффективный размер популяции в поколении t . В результате фамильная микроэволюция приближенно рассматривается как аналог процесса случайного генного дрейфа, одновременно протекающего в той же самой популяции с вчетверо меньшей, чем для фамилий, интенсивностью. Роль концентрации гомозигот при анализе фамильной структуры играет вероятность встречи индивидуумов с разными фамилиями. Результаты не предполагают монофилии фамилий и описывают умоглядную теоретическую совокупность популяций-реплик, как бы прошедших заново микроэволюционную историю рассматриваемой популяции в тех же условиях. Введено понятие фамильного аналога коэффициента инбридинга. Он практически превышает в 4 раза традиционный показатель инбридинга S . Райта. Оценивание коэффициента инбридинга реальных подразделенных популяций дает представление о статистической корреляции гомологичных генов объединяющихся гамет и не дифференцирует ситуации механической смеси субпопуляций или их общего происхождения. Лишь когда микроэволюция реальных субпопуляций с общим корнем независима, коэффициент инбридинга прямо связан с идентичностью генов по происхождению.

Ключевые слова: теоретическая популяционная генетика, популяции с неперекрывающимися поколениями, коэффициент инбридинга, дрейф фамилий, случайный генный дрейф, математические модели.

DOI: 10.31857/S001667582110009X

Интерес к популяционным фамильным данным у человека связан с задачами истории, географии заселения и миграций, происхождения, генеалогии населения [1, 2] и обусловлен просто вниманием людей к своим корням. Особенно важен также генетический аспект изучения фамильной структуры популяций [3–7]. Обширная библиография применения фамилий в популяционно-генетических исследованиях представлена в [8], критические моменты использования данных по изонимным бракам подчеркиваются в [9]. Применению моделей популяционной генетики в анализе фамильной структуры посвящены работы [10, 11] и др.

Во многом направления фамильных исследований объединяет ожидание кровнородственной связи (патрилинейная передача фамилий соответствует передаче Y-хромосомы) между людьми с одной фамилией (*изонимными*) и возможность получения информации о внутри- и межпопуляционном родстве [12]. С формализацией родства в терминах идентичности генов по происхождению можно познакомиться в [13, 14].

При изучении фамилий особый интерес вызывает связь коэффициента родства и инбридинга популяции с распространенностью в ней однофамильцев. Как известно, между передачей потомкам фамилии и передачей родительских генов имеются параллели. В определенной степени фа-

мии являются аналогами множественных аллелей аутосомного локуса. Одним из показателей родства служит коэффициент инбридинга популяции, важный, например, для понимания структуры наследственной отягощенности населения [15]. Семейные данные вносят свой вклад в оценивание инбридинга популяции и генетической структуры [4]. Здесь сделаем замечание относительно интерпретации коэффициента инбридинга. Данный показатель был введен С. Райтом в 1921, 1922 гг. для диаллельных локусов (см., например, [14]) как статистическая корреляция между гомологичными генами объединяющихся гамет в генотипе особи. Позднее широкое распространение получило определение коэффициента инбридинга как вероятности идентичности по происхождению аллелей (т.е. точных копий аллеля общего предка) индивидуума.

Понятно, что возникновению инбредности благоприятствует подбор родительских пар, связанных родством, например, при искусственном отборе. Может показаться неожиданным, но при случайном выборе супруги в популяции ограниченного размера вероятность идентичности аллелей по происхождению будет ненулевой. Объяснением возникающей инбредности в такой популяции служит *случайный генный дрейф*, приводящий к ненаправленным колебаниям концентраций аллелей от поколения к поколению. При относительно малом количестве поколений процесс дрейфа семейств отличается от генного дрейфа тем, что для первого из них время течет приблизительно в 4 раза быстрее. *В дальнейшем мы целиком фокусируемся на анализе возникновения инбредности популяции в результате случайного дрейфа как фактора увеличения вероятности идентичности аллелей по происхождению и дивергенции популяций.*

Цель настоящей серии работ в адаптации анализа генного дрейфа применительно к распределению семейств в популяциях, в систематическом изложении соответствующих результатов относительно семейной структуры и коэффициента инбридинга в статике, динамике и асимптотике на относительно малом времени дивергенции (малых коэффициентах инбридинга), а также в выявлении связи между характеристиками семейной и генетической структур. Показано, что при относительно малом времени доминирующим фактором динамики является случайный выборочный дрейф, приводящий к дивергенции популяции от начального состояния и дивергенции друг от друга популяций с общим происхождением на начальном этапе их микроэволюции.

В последующих частях данной работы планируется анализ зависимости от времени характеристик дивергенции возможных семейных состояний, получение точных в рамках модели зависимостей характеристик от времени и их аппроксимаций

при относительно малой длительности дивергенции. Использование понятия относительно времени и его связи с семейным аналогом коэффициента инбридинга показывает, что время для семейного дрейфа течет в 4 раза быстрее, чем для одновременно протекающего в популяции дрейфа генов.

В дальнейших публикациях для анализа семейного дрейфа будет применяться преобразование состояний для стабилизации темпа дивергенции. При использовании углового расстояния между начальным и текущим состояниями случайные изменения расстояний будут изотропны, и приблизительно у квадрата расстояния будет распределение хивадрат, что позволяет оценивать коэффициент семейного инбридинга (который вчетверо больше коэффициента инбридинга С. Райта, традиционно в популяционной генетике). Кроме того, детально анализируется связь вероятности случайной встречи двух изонимных индивидуумов с вероятностью идентичности по происхождению гомологичных аллелей объединяющихся гамет в диплоидных генотипах индивидуумов.

Кратко коснемся обозначений и терминологии. Под концентрациями семейств в популяции подразумеваются концентрации однофамильцев. Если используется информация о фамилиях мужчин и женщин, то у женщин имеются в виду девичьи фамилии. Термины “вероятность наблюдения пары однофамильцев, вероятность случайной встречи двух однофамильцев, появление изонимной пары в результате случайного выбора двух индивидуумов” являются синонимами. Векторы набраны полужирным шрифтом, к обозначениям семейных аналогов популяционно-генетических характеристик добавлено окончание *s*. Характеристики индивидуумов и их пар пишутся строчными буквами, а популяционные характеристики — заглавными (например, *f* и *F*), символ ◀ обозначает конец доказательства.

О СЛУЧАЙНОМ ВЫБОРОЧНОМ ДРЕЙФЕ СЕМЕЙСТВА

Чтобы проанализировать связь динамики распределения семейств в популяции с передачей потомкам наследственной информации, желательно хорошо представлять себе особенности этих процессов изолированно. В определенной степени фамилии являются аналогами множественных аллелей одного аутосомного локуса. Публикаций по однолокусной генетической структуре популяций конечного размера гораздо больше, чем по особенностям распределения семейств. Поэтому обратимся к анализу семейной структуры.

Проведем ее исследование в максимально простой ситуации с выполнением предположений, используемых в популяционной генетике. Пусть по-

колениа не перекрываются. Конечно, у человека существует перекрывание, но использование результатов, полученных для модели с неперекрывающимися поколениями (например, закона Харди–Вайнберга), обычно не приводит к противоречиям и эффективно при анализе генетической структуры популяций человека.

Положим, что фамилии передаются патрилинейно от отца к сыну без изменений, в браке супруга принимает фамилию мужа. Очевидно, после заключения браков *фамильная структура у супругов как мужчин, так и женщин одинакова*, и учет фамилий женщин после брака (дублей фамилий супругов) неинформативен. Все это свидетельствует о том, что можно отказаться от использования данных по фамилиям женщин и изучать только фамильную структуру мужчин — ведь даже если бы фамильный состав женщин изменялся произвольно это не оказывало бы никакого влияния на фамилии у мужчин в последующем.

Однако в некоторых отношениях использование девичьих фамилий женщин целесообразно, например при анализе изонимных браков. Добавим к этому дополнительные соображения о роли женщин в анализе фамильного состава. Рассмотрим популяцию в произвольном поколении τ . В нем фамильный состав совместно мужчин и девичьих фамилий у женщин в популяции размера $N(\tau)$ является случайной выборкой с возвращением объема $N(\tau)$ (включающей выборки фамилий по каждому полу) из совокупности фамилий у родителей-мужчин предыдущего поколения. Если наша задача состоит в оценивании фамильного состояния *в предыдущем поколении* $\tau - 1$, то источником информации будут как фамилии мужчин, так и *девичьи* фамилии женщин общим количеством $N(\tau)$ — размером всей случайной выборки из фамилий у родителей-мужчин.

Если же задача состоит в предсказании свойств фамильного состояния *в следующем поколении*, то они определяются размером $N(\tau + 1)/2$ случайной выборки (дающей фамильный состав мужчин в очередном поколении $\tau + 1$) из фамилий мужчин в момент τ . Таким образом, оценивание предыдущего состояния точнее, чем предсказание следующего, если не принимать во внимание динамику численности.

Итак, при изучении динамики фамильной структуры можно ограничиться анализом данных *только по мужчинам*, численность которых при образовании семей обозначим как $N/2$. Именно к стадии заключения браков будем относить фамильное состояние, или фамильную структуру популяции — концентрации фамилий среди муж-

чин. *Хотя далее будем говорить об N как общем размере популяции с учетом обоих полов, эта интерпретация условна и базовой величиной является $N/2$.*

Динамика распределения фамилий обусловлена различиями в количестве детей, оставляемых родительскими парами, точнее различиями в числе сыновей. В этом отношении носителей фамилий полагаем равнозначными (нейтральными), а различия в размере семей чисто случайными. Каждый индивидuum следующего поколения с равной вероятностью $1/\frac{N}{2}$ является потомком любого из $N/2$ родителей-отцов предыдущего поколения. Повторим, что при этом фамильный состав (фамильное состояние) мужчин следующего поколения удовлетворяет условиям случайной выборки с возвращением размера $N/2$ из совокупности фамилий родительского поколения. Понятно, что фамильный состав поколения потомков наследует родительский с выборочной ошибкой, накапливающейся в поколениях.

Эта схема известна в популяционной генетике как модель Райта–Фишера. В ней *генетический состав* нового поколения диплоидной популяции является результатом $2N$ случайных испытаний (N — размер популяции с учетом женщин и $2N$ — количество гамет, формирующих N диплоидных потомков). Вероятность при каждом из $2N$ испытаний получить i -ю гамету равна ее концентрации p_i в пуле гамет родителей. У нас аналогично *новый фамильный состав* мужской части популяции размера $N/2$ на следующем шагу является случайной выборкой $N/2$ фамилий родителей с присутствием выборкам отклонениями от состояния родительского поколения.

Здесь имеется полное сходство с генетической моделью гаплоидной популяции (или диплоидной со случайным скрещиванием), рассматриваемой в отношении одного (аутосомного) локуса с множественными аллелями. Роль фамилий играют аллели. Аналогом пар индивидуумов с разными фамилиями (следовательно, с разными родоначальниками) являются гетерозиготы, а аналогом пар однофамильцев будут гомозиготы. Согласно свойствам независимых биномиальных испытаний с вероятностью успеха p_i дисперсия $V(x_i)$ возможных значений доли x_i успехов в выборке размера K , т.е. дисперсия концентрации i -й фамилии находится как $V(x_i) = p_i(1 - p_i)/K$.

Отсюда для одной и той же популяции при одинаковых фамильном и генетическом состояниях интенсивность выборочных колебаний при смене поколения (дисперсия) концентраций фамилий в 4 раза больше, чем у одновременно протекающего в попу-

ляции генного дрейфа (выборка фамилий вчетверо меньше, чем выборка гамет).

В дальнейшем мы воспользуемся следующим свойством [16] класса моделей стохастических процессов со случайными и систематическими изменениями состояния, которому принадлежит модель Райта—Фишера.

Замечание 1. Пусть состояние популяции размера N с неперекрывающимися поколениями меняется за счет независимых случайных выборочных колебаний при смене поколений с дисперсией σ^2 порядка $1/N$ и неслучайного давления систематических факторов микроэволюции, также вызывающих изменения порядка $1/N$ за поколение.

Тогда, при малой величине отношения t/N , где t обозначает длительность процесса динамики состояния популяции в поколениях, давлением систематических факторов (отбора, миграций, мутаций) можно пренебречь.

Доказательство. Для конкретности рассмотрим модель Райта—Фишера. У нее изменения состояния популяции за поколение (шаг) не направлены и случайны. Поэтому среднее значение выборочного отклонения состояния за шаг не характеризует темп динамики, так как оно равно нулю. Скорость можно охарактеризовать, например, средним абсолютным отклонением за поколение, которое для отдельной координаты состояния (концентрации отдельной фамилии) примерно равно 0.8 от среднего квадратического отклонения σ в случае нормального и близких к нему распределений. Скорость изменений квадрата отклонения ненаправленного случайного процесса можно измерять дисперсией σ^2 состояний за шаг.

В случае некоррелирующих случайных изменений состояний при сменах поколений дисперсия суммарного отклонения через небольшое количество t поколений увеличится примерно в t раз, как и величина систематического отклонения. Значит среднее абсолютное отклонение будет порядка $\sqrt{t\sigma^2} = \sqrt{t1/N}$, а порядок систематического $t1/N$. Для их отношения имеем

$$\begin{aligned} \sqrt{t \times 1/N} / (t \times 1/N) &= \\ = \sqrt{N/t} \gg 1 \text{ при } N/t \gg 1 \text{ (при } t/N \ll 1). \end{aligned}$$

Следовательно, когда t мало по сравнению с N , систематическим отклонением можно пренебречь как много меньшим по сравнению с абсолютным. ◀

ФИНАЛЬНЫЕ ТЕОРЕТИЧЕСКИЕ ПОСЛЕДСТВИЯ ВЫБОРОЧНЫХ КОЛЕБАНИЙ ФАМИЛЬНОГО СОСТАВА ПОПУЛЯЦИИ

Рассмотрим гипотетическую исходную (родоначальную) популяцию (без предположения о монофилии фамилий, но на промежутке времени, когда мутациями и миграциями фамилий можно пренебречь). Пусть p_i обозначает концентрацию i -й фамилии среди родоначальников. Фамильный состав популяции в следующем поколении моделируется как состав случайной выборки с возвращением из фамилий родителей. Очевидно, что вероятность выбрать i -ю фамилию из совокупности, где ее концентрация равна p_i , совпадает с p_i . Изменчивость вклада родоначальников в концентрацию i -й фамилии в следующем поколении обусловлена случайностью результатов ряда биномиальных испытаний с вероятностью успеха p_i (появления потомка с i -й фамилией) и неудачи $1 - p_i$ (появления потомка с любой другой фамилией). Обозначим через x_i концентрацию i -й фамилии в следующем поколении. Напомним, что согласно свойствам биномиальных испытаний у выборки размера K будут следующие характеристики:

$$\begin{aligned} E\{x_i\} &= E\{x_i | p_i\} = p_i, \\ V(x_i) &\equiv E\{(x_i - E\{x_i\})^2\} = p_i(1 - p_i)/K, \end{aligned} \tag{1}$$

где $E\{x_i\}$ обозначает теоретическую среднюю величину (математическое ожидание) для случайных значений x_i ($E\{x_i | p_i\}$ — математическое ожидание x_i при условии вероятности успеха p_i), а $V(x_i)$ — дисперсия случайной величины x_i (у нас дисперсия возможных концентраций i -й фамилии x_i в выборке, т.е. новом поколении).

Ремарка 2. Термин “ожидаемая величина (математическое ожидание)” понимается следующим образом. Пусть в одних и тех же условиях из родительской популяции многократно и независимо производятся случайные выборки одного и того же размера. У каждой выборки может быть свой фамильный состав, т.е. у нас состав популяции нового поколения, отличающийся в общем случае от состава другой выборки. Такие популяции-реплики образуют умозрительную теоретическую совокупность. Ее характеризуют среднее (ожидаемое) значение x_i (или каких-либо характеристик популяции, выражающихся через $x \equiv \{x_i\}$), а также дисперсия значений x_i . Точно так же наблюдаемые результаты случайного процесса микроэволюции популяции при его повторении в тех же условиях не совпадут друг с другом. Здесь теоретическую совокупность образует множество возможных результатов-реплик популяции, повторя-

ющей заново микроэволюцию. Среднее значение x_i в теоретической совокупности является математическим ожиданием x_i . Ее характеристики обычно находятся математически.

Ожидаемую величину x , обозначаемую как $E\{x_i\}$, следует отличать от выборочной средней величины \bar{x} (это сумма наблюдаемых величин, деленная на число наблюдений). Значение $E\{x_i\}$ является *константой* (для биномиальных испытаний она равна вероятности успеха p), а \bar{x} является *случайной величиной*, варьирующей от выборки к выборке (у нас \bar{x} совпадает с концентрацией фамилии, равной суммарному количеству однофамильцев с данной фамилией, деленному на размер выборки, причем дисперсия \bar{x} равна $p(1-p)/K$).

Теоретически выборочные колебания концентраций x_i приводят в конце концов к такому фамильному состоянию популяции, когда она будет целиком состоять из однофамильцев (см. далее). Разумеется, данный теоретический предел не достигается в реальности, хотя бы потому, что предпосылки модели не могут строго выполняться неограниченно долго. Промежуточные стадии процесса случайного дрейфа многократно наблюдались в малых деревнях России, где резко преобладала одна или несколько фамилий. Такая ситуация описывалась не раз в художественной литературе упоминаниями типа “у нас в деревне все Смирновы”. Меня поражала картина, когда большинство жителей деревни оказываются однофамильцами. Например, при изучении популяций европейского севера России [17] встретила деревня с 91% однофамильцев при общем числе жителей 126 человек. Для городского человека такая ситуация парадоксальна. Так, когда я учился в школе в подмосковном городе с населением порядка 45 тыс. человек, в моем классе совсем не было однофамильцев.

Результат 3. Пусть фамильные (генетические) состояния популяции с неперекрывающимися поколениями формируются как последовательности независимых случайных выборок с возвращением (вложенных выборок) фамилий (аллелей) из фамилий (аллелей) родительского поколения. Тогда:

1) закономерности случайного процесса изменений фамильного состава популяции определяются ее мужской составляющей и соответствуют процессу генного дрейфа по одному аутосомному локусу с множественными аллелями (аналогами фамилий) в той же самой популяции;

2) хотя закономерности дрейфа фамилий такие же, как для синхронно идущего в популяции генного дрейфа, интенсивность дрейфа фамилий, фигурально говоря, вчетверо выше, чем дрейфа генов, при одинаковых фамильном и генетическом состояниях.

Доказательство. 1. Выборочная природа колебаний фамильного состава популяции воспроизводит свойства популяционно-генетической модели случайного генного дрейфа в популяции ограниченной численности при панмиксии. Это доказывает соответствие между процессами генного дрейфа и дрейфа фамилий, закономерности которых представляют собой закономерности последовательности вложенных выборок.

2. Как показано выше, в одной и той же популяции процесс генного дрейфа соответствует последовательности вложенных выборок в 4 раза большего размера (численности гамет), чем выборок в процессе дрейфа фамилий (численности мужчин). Тем самым интенсивность дрейфа фамилий, если измерять ее дисперсией возможных концентраций фамилии (аллеля) в выборке, будет вчетверо больше, чем для дрейфа генов при одинаковых фамильном и генетическом состояниях. ◀

Для вычисления математических ожиданий полезна следующая формула полного математического ожидания какой-либо числовой характеристики X :

$$E(X) = E\{E\{X|A\}\} = \sum E\{X|A_i\}Pr(A_i). \quad (2)$$

Здесь A – объединение событий $\{A_i\}$, представляющих собой полную систему несовместимых случайных событий, реализующихся с вероятностями $\{Pr(A_i)\}$ и таких, что обязательно происходит одно из событий системы; $E\{X|A_i\}$ обозначает условное (условие пишем после вертикальной черты) математическое ожидание для X при условии реализации соответствующего случайного события A_i . Например, в искусственной ситуации, когда полная система состоит из городских и сельских жителей, средний рост населения $E\{рост\}$ равен $E\{рост|житель города\} \cdot Pr(житель города) + E\{рост|житель села\} \cdot Pr(житель села)$.

Результат 4. Пусть в популяции с неперекрывающимися поколениями при заключении браков жена получает фамилию мужа. Положим, что фамилии наследуются патрилинейно, равноценны (нейтральны) в отношении их передачи потомкам, у каждого из которых с равной вероятностью отцом может быть любой из родительского поколения, а мутациями и миграциями фамилий можно пренебречь. Тогда:

1) в любом поколении (времени) вероятность случайно встретить индивидуума с i -й фамилией в наугад выбранной популяции из теоретической совокупности равна ее концентрации p_i в родоначальной популяции, $i = 1, 2, \dots$;

2) в предположении, что с течением времени в конкретной популяции останутся только однофа-

милльцы, вероятность фиксации i -й фамилии равна ее концентрации p_i среди родоначальников.

Доказательство. Разумеется, в конкретной популяции концентрация фамилии (вероятность встречи в ней индивидуума с данной фамилией) варьирует по поколениям, а постоянно лишь ее ожидаемое значение на возможных траекториях динамики. Динамика семейного состава начинается с формирования первого поколения, состав которого соответствует результату независимых испытаний, в каждом из которых потомок с i -й фамилией появляется с вероятностью p_i (вероятностью успеха), равной концентрации данной фамилии среди родоначальников. Пусть x_i обозначает концентрацию i -й фамилии в новом поколении. Согласно свойствам биномиальных испытаний (1) среднее (ожидаемое) значение для доли успехов $E\{x_i\}$ в испытаниях равно вероятности успеха, т.е. у нас среди популяций-реплик ожидаемое значение концентрации i -й фамилии в новом поколении равно p_i . Значит у концентраций и фамилий нет определенного тренда в отклонениях от своих начальных значений.

Возможные реализации биномиальных испытаний можно интерпретировать как мыслимые (теоретически возможные) семейные составы популяций-реплик, семейное разнообразие которых получено в результате повторных случайных выборок из той же самой родительской популяции. Тогда $E\{x_i\}$ имеет смысл ожидаемой для популяций-реплик концентрации i -й фамилии и является функцией начального состояния $x_i(0) = p_i$; $E\{x_i(1)\} = E\{x_i(1)|x_i(0) = p_i\} = p_i$.

1. Очевидно, в поколении τ значение концентрации i -й фамилии в популяции является случайной величиной. В следующем поколении $\tau + 1$ при условии $x_i(\tau)$ для распределения очередного значения концентрации $x_i(\tau + 1)$ имеем $E\{x_i(\tau + 1)|x_i(\tau)\} = x_i(\tau)$ согласно свойствам (1) биномиальных испытаний. По формуле полного математического ожидания (2)

$$E\{x_i(\tau + 1)\} = E\{E\{x_i(\tau + 1)|x_i(\tau)\}\} = E\{x_i(\tau)\},$$

$$E\{x_i(\tau)\} = E\{x_i(\tau - 1)\} = \dots = x_i(0) \equiv p_i.$$

2. Найдем вероятность фиксации i -й фамилии в предположении случайной утери со временем всех фамилий, кроме одной (см. далее). Поскольку $E(x_i(\tau)) = p_i$ в любом поколении, это равенство выполняется и когда в популяции осталась только одна из фамилий. Например, пусть i -я фамилия фиксировалась с неизвестной вероятностью U_i (тогда ее концентрация x_i в популяции равна еди-

нице) и утерялась с вероятностью $1 - U_i$ (тогда ее концентрация равна нулю). Отсюда согласно (2)

$$p_i = E\{x_i\} = E\{x_i|фиксация\}U_i + E\{x_i|утеря\}(1 - U_i) = 1U_i + 0(1 - U_i) = U_i. \blacktriangleleft$$

ОПИСАНИЕ ДИВЕРГЕНЦИИ ФАМИЛЬНОЙ СТРУКТУРЫ ПОДРАЗДЕЛЕННЫХ ПОПУЛЯЦИЙ

Напомним известные из биометрии свойства дисперсии случайных величин. Для любой случайной величины (например, для x с дисперсией $V(x)$) имеем

$$V(x) \equiv E\{(x - E\{x\})^2\} = E\{x^2\} - (E\{x\})^2,$$

$$E\{x^2\} = (E\{x\})^2 + V(x). \quad (3)$$

Перед анализом семейной структуры обратимся к использованию концепции *идентичности по происхождению* (ИПП) *гомологичных генов* (т.е. точных копий гена некоторого общего предка) применительно к определению коэффициента инбридинга. *Коэффициентом инбридинга индивидуума $f \geq 0$ называется вероятность ИПП его гомологичных генов*. Понятно, что согласно данному определению коэффициент инбридинга f является случайной величиной, принимающей (как вероятность) значения от 0 до 1 и варьирующей от индивидуума к индивидууму. *Коэффициентом инбридинга популяции $F \geq 0$ называется ожидаемый коэффициент инбридинга у случайно выбранного из популяции индивидуума ($F \equiv E\{f\} \geq 0$).*

В инбредной популяции с концентрациями $\{p_i\}$ аллелей аутосомного локуса концентрации $\{p_{ii}\}$ гомозиготных генотипов имеют вид (см., например, [18]):

$$p_{ii}(F) = Fp_i + (1 - F)p_i^2 = p_i^2 + Fp_i(1 - p_i) \geq p_i^2,$$

$$i = 1, 2, \dots, k; \quad 1 \geq p_i, \quad F \geq 0.$$

Соответственно концентрация $\sum p_{ii}(F)$ всех гомозигот находится как

$$\sum p_{ii}(F) = \sum (Fp_i + (1 - F)p_i^2) =$$

$$= \sum p_i^2 + \sum Fp_i(1 - p_i) \geq \sum p_i^2.$$

Концентрацией $H(p, F) \equiv 1 - \sum p_{ii}(F)$ всех гетерозигот будет

$$\begin{aligned}
 H(\mathbf{p}, F) &= 1 - \sum p_{ii}(F) = \\
 &= 1 - \left(\sum p_i^2 + F \sum p_i(1 - p_i) \right) = \\
 &= 1 - \sum p_i^2 - F \left(\sum p_i - \sum p_i^2 \right) = \left(1 - \sum p_i^2 \right) - \\
 &\quad - F \left(1 - \sum p_i^2 \right) = (1 - F) \left(1 - \sum p_i^2 \right) \equiv \\
 &\equiv (1 - F)H(\mathbf{p}, 0) \leq H(\mathbf{p}, 0).
 \end{aligned}$$

Таким образом, инбридинг проявляется в уменьшении концентрации гетерозигот и увеличении гомозиготности по сравнению со значениями, соответствующими соотношениям Харди–Вайнберга.

Коэффициент инбридинга можно выразить через дисперсии концентраций аллелей в популяциях-репликах или дисперсии по субпопуляциям в подразделенной популяции (см. далее).

Проиллюстрируем картину дивергенции фамилий в статике (в некоторый фиксированный момент) на примере умозрительного или реального множества из нескольких субпопуляций со своими фамильными составами из k фамилий. Такое множество часто называют *подразделенной популяцией*. Ее субпопуляции составляют теоретическую или реальную обследованную совокупность. Обозначим концентрации фамилий по всем субпопуляциям через $\mathbf{x}_i \equiv \{x_i^{(j)}\}$. Здесь верхний индекс у $x_i^{(j)}$ дает номер субпопуляции, а нижний – номер фамилии.

Положим, что у распределения концентрации i -й фамилии x_i по субпопуляциям среднее значение равно p_i , а дисперсия – $Vs(x_i)$. Статическую картину дивергенции фамильного состава в данном множестве субпопуляций будем описывать с помощью дисперсий $\{Vs(x_i)\}$ и вероятностей $\{Hs_i(\mathbf{x})\}$ встреч двух индивидуумов с i -й фамилией и с какой-либо другой, а также с использованием вероятности $Hs(\mathbf{x})$ случайной встречи двух индивидуумов с разными фамилиями. Повторим, что аналогом для $Hs_i(\mathbf{x})$ в популяционной генетике является концентрация гетерозигот с i -м аллелем, а для $Hs(\mathbf{x})$ – концентрация всех гетерозигот.

Сделаем небольшое отступление, касающееся смысла Hs и H . Hs является теоретической концепцией, относящейся к случайному событию (наблюдению разных фамилий у пары индивидуумов), которое могло произойти или нет, тогда как H относится к наблюдениям фактически реализовавшихся событий (и находится подсчетом гетерозигот). Вероятность Hs можно понимать двояко. С одной стороны, эту вероятность можно определить с учетом порядка фамилий как $x_i(1 - x_i)$, где x_i – концентрация i -й фамилии во всей сово-

купности популяций как едином целом. С другой стороны, ее можно понимать как вероятность случайно выбрать одну из (различающихся) популяций совокупности, скажем с концентрацией x_i , затем вычислить в ней значение $E\{Hs_i|\mathbf{x}\} = x_i(1 - x_i)$ согласно закону Харди–Вайнберга, а уже потом найти итоговую вероятность результата извлечений с учетом всевозможных выборов популяции по формуле (2) полного математического ожидания $E\{Hs_i\} = E\{E\{Hs_i|\mathbf{x}\}\}$.

Например, если представить себе, что у всех субпопуляций одна и та же независимая микроэволюционная история после отделения от общего корня, то состояние j -й субпопуляции соответствует состоянию на j -й из возможных траекторий динамики, играющих роль популяций-реплик. В фиксированный момент времени имеем как бы срез траекторий, и $E\{E\{Hs_i|\mathbf{x}\}\}$ дает ожидаемое значение Hs_i на траекториях в данный момент. В простейшем случае M субпопуляций и равновероятного выбора любой из них значение $E\{E\{Hs_i|\mathbf{x}\}\} = E\{x_i^{(j)}(1 - x_i^{(j)})\}$ находится как $\sum_{j=1}^M x_i^{(j)}(1 - x_i^{(j)})/M$. Далее вероятности $Hs(\mathbf{x})$ и $Hs_i(\mathbf{x})$ случайных встреч двух индивидуумов с разными фамилиями, например Hs_i , определим как $E\{E\{Hs_i|\mathbf{x}\}\}$ и будем для этого использовать более прозрачное обозначение $E\{Hs_i(\mathbf{x})\}$.

Стимулом для предлагаемой формы записи, приложимой также к широко распространенному обозначению H для концентрации гетерозигот по аутосомному локусу в генетике популяций, служит многозначность данного символа. Так, в популяционной генетике H может относиться к концентрации гетерозигот в одной свободно скрещивающейся популяции, в подразделенной популяции, состоящей из таких субъединиц, наконец и в умозрительной совокупности популяций с возможными генетическими состояниями.

Теоретически выводимые свойства статике умозрительной совокупности популяций сходны со свойствами реально существующих подразделенных популяций, состоящих из отдельных свободно скрещивающихся единиц. При первоначальном знакомстве с генетикой популяций мне потребовались усилия, чтобы разобраться можно ли теоретическую формулу $H = 2p_i(1 - p_i)(1 - F)$ для концентрации гетерозигот в популяции со свободным скрещиванием при накопленной инбредности прилагать к одной реальной популяции. Ведь в последней согласно закону Харди–Вайнберга $H = 2p_i(1 - p_i)$. Суть противоречия в том, что в первом случае H фактически подразумевает $E\{H\}$, т.е. ха-

рактическую характеристику распределения H по субпопуляциям, а не свойство одной из них.

В приведенных рассуждениях вероятности $Hs_i(\mathbf{x})$, $Hs(\mathbf{x})$ и $H_i(\mathbf{x})$, $H(\mathbf{x})$ рассматривались как случайные величины. Хотя в реальной подразделенной популяции состояния субпопуляций фиксированы, случайность вносит процедура выбора субпопуляций, например при вычислении математического ожидания, и можно говорить о вероятности $Hs_i(\mathbf{x})$ случайной встречи двух потомков разных родоначальников (с i -й фамилией и какой-либо иной). Вернемся к описанию фамильной дивергенции в заданном множестве субпопуляций с использованием вероятностного подхода. По формуле полного математического ожидания (2)

$$E\{Hs_j\} = E\{E\{Hs_j|\mathbf{x}\}\}, \quad E\{Hs_j|\mathbf{x}\} = x_j(1 - x_j),$$

где $E\{Hs_j|\mathbf{x}\}$ – условное математическое ожидание Hs_j . При условии концентрации x_i у i -й фамилии в популяции с учетом порядка фамилий в паре значение вероятности $Hs_j(\mathbf{x})$ находится однозначно как $x_i(1 - x_j)$ аналогично использованию закона Харди–Вайнберга в популяционной генетике. Отсюда согласно (3)

$$\begin{aligned} E\{Hs_j(\mathbf{x})\} &= E\{E\{Hs_j|x_i\}\} = E\{x_i(1 - x_j)\} = \\ &= E\{x_i\} - E\{x_i^2\} = p_i - p_i^2 - Vs(x_i) = \\ &= p_i(1 - p_i)(1 - Vs(x_i)/p_i(1 - p_i)) \equiv \\ &\equiv p_i(1 - p_i)(1 - Fs), \\ Fs &\equiv Vs(x_i)/p_i(1 - p_i), \end{aligned}$$

где Fs – фамильный аналог F . Когда концентрация i -й фамилии в j -й популяции равна $x_i^{(j)}$, то для нее рассматриваемая вероятность Hs_j равна $x_i^{(j)}(1 - x_i^{(j)})$. В то же время выбор популяции является случайным, и $Hs_j(\mathbf{x})$ находится с учетом выборов всевозможных популяций как $E\{x_i^{(j)}(1 - x_i^{(j)})\}$. Чтобы подчеркнуть, что подразумевается вычисление ожидаемого значения при случайном выборе популяции (при изменчивости по верхнему индексу, номеру популяции) этот индекс припишем к символу математического ожидания E как E_j .

Замечание 5. Пусть дана совокупность субпопуляций с концентрациями фамилий $\{x_i^{(j)}\}$, где i – номер фамилии, а j – номер субпопуляции. Обозначим вектор средних концентраций фамилий для распределений концентраций по субпопуляциям во всей совокупности через $\mathbf{p} = \{p_i\}$, а вектор дисперсий как $\{Vs(x_i)\}$.

Тогда в наугад выбранной субпопуляции вероятность случайной встречи пары индивидуумов с i -й

фамилией и с какой-либо другой $E_j\{Hs_i(\mathbf{x})\}$, вероятность встречи двух индивидуумов с любыми разными фамилиями $E_j\{Hs(\mathbf{x})\}$, вероятность встречи пары однофамильцев с i -й фамилией $E_j\{(x_i^{(j)})^2\}$ и вероятность встречи пары однофамильцев независимо от их фамилий $1 - E_j\{Hs(\mathbf{x})\}$ выражаются как

$$\begin{aligned} E_j\{Hs_i(\mathbf{x})\} &= p_i(1 - p_i) - Vs(x_i) \leq p_i(1 - p_i), \\ E_j\{Hs(\mathbf{x})\} &= Hs(\mathbf{p}) - Vs(\mathbf{x}) \leq Hs(\mathbf{p}), \end{aligned}$$

$$E_j\{(x_i^{(j)})^2\} = p_i^2 + Vs(x_i) \geq p_i^2,$$

$$\begin{aligned} 1 - E_j\{Hs(\mathbf{x})\} &= E_j\left\{\sum_{i=1}^k (x_i^{(j)})^2\right\} = \\ &= \sum_{i=1}^k p_i^2 + Vs(\mathbf{x}) \geq \sum_{i=1}^k p_i^2. \end{aligned}$$

Здесь $Vs(\mathbf{x}) \equiv \sum_{i=1}^k Vs(x_i)$, индекс j у E_j означает усреднение по субпопуляциям.

Доказательство. По условию $E_j\{x_i^{(j)}\} = p_i$. Согласно (3)

$$\begin{aligned} E_j\{Hs_i(\mathbf{x})\} &\equiv E_j\{x_i^{(j)}(1 - x_i^{(j)})\} = \\ &= E_j\{x_i^{(j)}\} - E_j\{(x_i^{(j)})^2\} = \\ &= p_i - (Vs(x_i) + E_j\{(x_i^{(j)})^2\}) = \\ &= p_i - Vs(x_i) - p_i^2 = p_i(1 - p_i) - Vs(x_i) \leq p_i(1 - p_i) \\ &\text{или } Vs(x_i) = p_i(1 - p_i) - E_j\{x_i^{(j)}(1 - x_i^{(j)})\}. \end{aligned}$$

Отсюда видно, что ожидаемая вероятность $Hs_i(\mathbf{x})$ встречи двух потомков разных родоначальников (с i -й фамилией и какой-либо другой) меньше соответствующего случайным встречам значения $p_i(1 - p_i)$ на величину межпопуляционной дисперсии $Vs(x_i)$ – аналог формулы Валунда для подразделенной популяции в популяционной генетике – см., например, [18, 19].

В фиксированной субпопуляции, скажем с номером j , вероятность встречи однофамильцев с i -й фамилией теоретически равна $(x_i^{(j)})^2$, а в случайно выбранной как ожидаемое значение этой вероятности при всевозможных результатах выбора:

$$\begin{aligned} E_j\{(x_i^{(j)})^2\} &= (E_j\{x_i^{(j)}\})^2 + \\ + Vs(x_i) &= p_i^2 + Vs(x_i) \geq p_i^2, \quad i = 1, 2, \dots, k, \end{aligned}$$

$$E_j\left\{\sum_{i=1}^k (x_i^{(j)})^2\right\} = \sum_{i=1}^k p_i^2 + Vs(\mathbf{x}) \geq \sum_{i=1}^k p_i^2,$$

$$Vs(\mathbf{x}) \equiv \sum_{i=1}^k Vs(x_i).$$

Вероятность встречи двух индивидуумов с разными фамилиями равна дополнению до единицы вероятности встречи однофамильцев. Пусть \mathbf{x} обозначает вектор концентраций всех фамилий во всех популяциях. Тогда вероятностью $Hs(\mathbf{x})$ случайной встречи в одной популяции двух индивидуумов с разными фамилиями с учетом формулы (3) будет

$$\begin{aligned} E_j \{Hs(\mathbf{x})\} &\equiv 1 - E_j \left\{ \sum_{i=1}^k (x_i^{(j)})^2 \right\} = \\ &= 1 - \sum_{i=1}^k p_i^2 - \sum_{i=1}^k Vs(x_i) = \\ &= 1 - \sum_{i=1}^k p_i^2 - Vs(\mathbf{x}) = Hs(\mathbf{p}) - Vs(\mathbf{x}) \leq Hs(\mathbf{p}). \blacktriangleleft \end{aligned}$$

Здесь $E_j \{Hs(\mathbf{x})\}$ и $Vs(\mathbf{x})$ дают представление о картине дивергенции субпопуляций внутри всей реальной или умозрительной совокупности. В подразделенной популяции (или в умозрительной совокупности популяций-реплик) вероятность $E_j \{Hs_i(\mathbf{x})\}$ встречи индивидуумов с i -й фамилией и с какой-то другой меньше соответствующего случайным встречам значения $p_i(1 - p_i)$ на величину межпопуляционной дисперсии $Vs(x_i)$. Одновременно вероятность $E_j \{Hs(\mathbf{x})\}$ встречи потомков разных родоначальников с учетом всех фамилий меньше значения $1 - \sum_i p_i^2 = Hs(\mathbf{p})$, соответствующего случайным встречам, на величину $Vs(\mathbf{x})$. В то же время вероятность случайно встретить своего однофамильца с i -й фамилией, напротив, больше значения p_i^2 , соответствующего случайным встречам во всем множестве как единой популяции, на величину межпопуляционной дисперсии $Vs(x_i)$.

ДИВЕРГЕНЦИЯ СУБПОПУЛЯЦИЙ И КОЭФФИЦИЕНТ ИНБРИДИНГА

Приведенные факты соответствуют эффектам инбридинга в генетике популяций. В генетическом контексте роль фамилий играют аллели, роль пар с разными фамилиями — гетерозиготы. При концентрации x_i у i -го аллеля по формуле С. Райта (см., например, [14, 18]) в инбредной популяции ожидаемой долей гетерозигот $E\{x_i(1 - x_i)\}$ будет $p_i(1 - p_i) \times (1 - F)$, где F — коэффициент инбридинга. Для фамилий $E\{x_i(1 - x_i)\}$ можно записать, как показано выше, в терминах $\{x_i^{(j)}\}$ в виде

$$\begin{aligned} E_j \{x_i^{(j)}(1 - x_i^{(j)})\} &= \\ &= p_i(1 - p_i)(1 - Vs(x_i)/(p_i(1 - p_i))) \equiv \\ &\equiv p_i(1 - p_i)(1 - Fs). \end{aligned}$$

Здесь видны параллели с дефицитом гетерозигот (нарушением закона Харди–Вайнберга) в популяционной генетике, а $Fs \equiv Vs(x_i)/p_i(1 - p_i)$ является фамильным аналогом случайного коэффициента инбридинга F . Таким образом,

$$Fs \equiv Vs(x_i)/p_i(1 - p_i), \quad Vs(x_i) = p_i(1 - p_i)Fs.$$

Отсюда

$$\begin{aligned} Vs(\mathbf{x}) &\equiv \sum_i Vs(x_i) = \\ &= \left(\sum_i p_i(1 - p_i) \right) Fs = \left(1 - \sum_i p_i^2 \right) Fs = Hs(\mathbf{p})Fs. \end{aligned}$$

Значит Fs можно представить как $Fs = Vs(\mathbf{x})/(1 - \sum_i p_i^2) = Vs(\mathbf{x})/Hs(\mathbf{p})$. В итоге для фамильной и популяционно-генетической структур имеем

$$\begin{aligned} Fs &= Vs(\mathbf{x}) / \left(1 - \sum_i p_i^2 \right) = Vs(\mathbf{x}) / Hs(\mathbf{p}), \\ F &= V(\mathbf{x}) / \left(1 - \sum_i p_i^2 \right) = V(\mathbf{x}) / H(\mathbf{p}), \end{aligned} \quad (4)$$

где $V(\mathbf{x}) \equiv \sum_i V(x_i)$ — сумма дисперсий концентраций аллелей.

ОБСУЖДЕНИЕ

Итак, существует значительное сходство между фамильной и генетической структурами, обязанное соответствию фамилий множественным аллелям аутосомного локуса и общей выборочной природе процессов случайного дрейфа фамилий и генов. Закономерности случайного дрейфа проявляются как свойства популяций-реплик, составляющих умозрительную теоретическую совокупность популяций с тождественной историей дивергенции от начального состояния и между собой. В каждый момент времени состояния реплик соответствуют срезу состояний на возможных траекториях динамики популяции, иначе говоря как бы множеству состояний подразделенной популяции, состоящей из таких реплик.

Для анализа реальных подразделенных популяций можно использовать приведенные в настоящей статье формулы, которые в равной степени приложимы как к совокупности субпопуляций с общим происхождением, так и к произвольному механическому набору случайно скрещивающихся субпопуляций. Например, популяционно-генетический аспект анализа подразделенной популяции (состоящей из удовлетворяющих соотношениям

Харди–Вайнберга субпопуляций) показывает нарушение этих соотношений в популяции как едином целом. В подразделенной популяции будет наблюдаться дефицит гетерозигот (эффект Валлунда, хотя внутри каждой отдельной субпопуляции дефицита нет).

Ситуация аналогична дефициту встреч индивидуумов с разными фамилиями по сравнению с ожидаемым при всеобщих чисто случайных встречах в популяции как едином целом. В статике дефициту соответствует коэффициент инбридинга, т.е. дивергенция субпопуляций, а динамика генотипической структуры подразделенной популяции рассматривалась, например, в [20]. Свойства генетических процессов в подразделенных популяциях анализировались также в [21] и [22].

Обратим внимание, что интерпретация коэффициента инбридинга зависит от характера субпопуляций в подразделенной популяции. В общем случае коэффициент инбридинга F в (4) численно равен статистической корреляции между гомологичными генами объединяющихся гамет как для произвольной механической смеси случайно скрещивающихся субпопуляций, так и для совокупности субпопуляций с общим происхождением. *Получение оценки F не позволяет определить с какой ситуацией имеет дело исследователь.*

Лишь в случае общего происхождения коэффициент инбридинга свидетельствует об идентичных по происхождению генах у субпопуляций, а статистическая корреляция связана с вероятностью идентичности по происхождению пары аллелей, образующих генотип индивидуума по рассматриваемому локусу. В общем случае динамика родственных популяций не будет независимой, и процедура оценивания коэффициента инбридинга должна учитывать характер их происхождения и условия микроэволюции.

Подчеркнем, что в любом случае существование корреляции между гомологичными генами объединяющихся гамет влияет на концентрации генотипов и тем самым определяет, например, ответную реакцию генетической структуры на действие отбора.

Настоящая статья не содержит каких-либо исследований с использованием в качестве объекта животных.

Настоящая статья не содержит каких-либо исследований с участием в качестве объекта людей.

СПИСОК ЛИТЕРАТУРЫ

1. *Бужилова А.П.* География русских фамилий // Восточные славяне. Антропология и этническая история. М.: Научный мир, 1999. С. 135–152.
2. *Балановская Е.В., Романов А.Г., Балановский О.П.* Однофамильцы или родственники. Подходы к изучению связи между гаплогруппами Y-хромосомы и фамилиями // Мол. биология. 2011. Т. 45. № 3. С. 473–485.
3. *Crow J.F., Mange A.P.* Measurement of inbreeding from the frequency of marriages between persons of the same surname // Soc. Biology. 1982. V. 29. № 1/2. P. 101–105.
4. *Lasker W.G.* Surnames and Genetic Structure. Cambridge: Cambr. University Press, 1985. 2005. 148 p.
5. *Ревазов А.А., Парадеева Г.М., Русакова Г.И.* Пригодность русских фамилий в качестве квазигенетического маркера // Генетика. 1986. Т. 22. № 4. С. 699–703.
6. *Tarskaia L., El'chinova G., Scapoli C. et al.* Surnames in Siberia: A study of the population of Yakutia through isonymy // Am. J. Phys. Anthropol. 2009. V. 138. P. 190–198.
7. *Сорокина И.Н., Чурносов М.И., Балтуцкая И.В. и др.* Антропогенетическое изучение населения центральной России. М.: Изд-во РАМН, 2014. 336 с.
8. *Сорокина И.Н., Рудых Н.А., Крикун Е.Н., Соколов С.Н.* Применение фамилий в популяционно-генетических исследованиях (на примере зарубежных популяций) // Научные ведомости. БелГУ. Сер. Медицина. Фармация. 2016. № 19(240). Вып. 35. С. 5–10.
9. *Rogers A.R.* Doubts about isonymy // Hum. Biol. 1991. V. 63. № 5. P. 663–668.
10. *Yasuda N., Cavalli-Sforza L.L., Skolnick M., Moroni A.* Evolution of surnames: An analysis of their distribution and extinction // Theor. Popul. Biol. 1974. V. 5. № 1. P. 123–142.
[https://doi.org/10.1016/0040-5809\(74\)90054-9](https://doi.org/10.1016/0040-5809(74)90054-9)
11. *Rossi P.* Self-similarity in population dynamics: surname distributions and genealogical trees // Entropy. 2015. V. 17. № 1. P. 425–437.
<https://doi.org/10.3390/e17010425>
12. *Lasker G.W.* A coefficient of relationship by isonymy: A method for estimating the genetic relationship between populations // Hum. Biol. 1977. V. 49. № 3. P. 489–493.
13. *Malecot G.* The Mathematics of Heredity. San Francisco; California: Freeman and Company, 1969. 88 p.
14. *Crow J.F., Kimura M.* An Introduction to Population Genetics Theory. The Blackburn Press, 2009. 591 p.
15. *Гинтер Е.К., Зинченко Р.А., Ельчинова Г.И. и др.* Роль факторов популяционной динамики в распространении наследственной патологии в российских популяциях // Мед. генетика. 2004. Т. 3. № 12. С. 548–555.
16. *Малютов М.Б., Пасеков В.П.* Об одной статистической задаче популяционной генетики // Теория вероятностей и ее применения. 1971. Т. 16. Вып. 3. С. 579–581.
17. *Пасеков В.П., Ревазов А.А.* К популяционной генетике населения европейского севера СССР. I. Данные по структуре шести деревень Архангельской области // Генетика. 1975. Т. 11. № 7. С. 145–155.
18. *Ли Ч.* Введение в популяционную генетику. М.: Мир, 1978. 555 с.

19. Вейр Б. Анализ генетических данных: дискретные генетические признаки. М.: Мир, 1995. 400 с.
20. Пасеков В.П. Влияние генного дрейфа на динамику генетической и фенотипической изменчивости в подразделенных популяциях // Матем. модели в экологии и генетике. М.: Наука, 1981. С. 148–173.
21. Maruyama T. Stochastic Problems in Population Genetics // Lecture Notes in Biomathematics. Berlin, N.Y.: Springer, 1977. V. 17. 245 p.
22. Свирежев Ю.М., Пасеков В.П. Основы математической генетики. М.: Наука, 1982. 511 с.

To Analysis of Random Processes of Isonymy. I. Isonymic Structure

V. P. Passekov*

*Dorodnicyn Computing Centre, Federal Research Center "Computer Science and Control"
of Russian Academy of Sciences, Moscow, 119991 Russia*

**e-mail: pass40@mail.ru*

The regularities of random dynamics of the surname composition in a population of limited size are analyzed. Population is approximately considered in discrete time with non-overlapping generations. Patrilinear surname inheritance is assumed. At a relatively small time interval $t/NE(t)$ where $NE(t)$ is an average harmonic effective population size over t generations, systematic pressures can be neglected. The surname composition of the next generation is a random sample of the size $Ne(\tau)/2$ from the surnames of the male component of the parent population, where $Ne(\tau)$ is effective size in the generation τ . As a result, surname microevolution is approximately considered to be an analogue of the process of random genetic drift which simultaneously proceeds in the same population with four times less intensity than for surnames. The role of homozygote concentrations in the analysis of surname structure is played by the probability of a random encounter of an individual with a namesake, heterozygote concentrations correspond to the probability of encounter of individuals with different surnames. The results do not suggest surname monophilia and refer to a speculative theoretical set of replica populations, as if having repeated the microevolutionary history of the population in question under the same conditions. The concept of a surname analogue of the inbreeding coefficient has been introduced. It is almost four times the traditional S. Wright's inbreeding coefficient. Estimating the inbreeding coefficient of real subdivided populations gives the statistical correlation of homologous genes of uniting gametes and does not differentiate situations of a mechanical mixture of subpopulations or their common origin. Only when the microevolution of real subpopulations with a common root is independent, the inbreeding coefficient is directly related to the identity of genes by descent.

Keywords: theoretical population genetics, populations with non-overlapping generations, inbreeding coefficient, random drift of surnames, random genetic drift, mathematical models.