

ОПИСАНИЕ ДИВЕРГЕНЦИИ СУБПОПУЛЯЦИЙ В ИЕРАРХИЧЕСКОЙ
СИСТЕМЕ ПРИ АНАЛИЗЕ ИЗОНИМИИ. III. КОЭФФИЦИЕНТ
ФАМИЛЬНОГО ИНБРИДИНГА

© 2023 г. В. П. Пасеков*

*Федеральный исследовательский центр “Информатика и управление”
Российской академии наук, Москва, 119991 Россия***e-mail: pass40@mail.ru*

Поступила в редакцию 03.05.2023 г.

После доработки 30.06.2023 г.

Принята к публикации 04.07.2023 г.

Рассматривается характерная для человека иерархическая структура метапопуляций с их подразделенностью на части (субпопуляции), обычно классифицируемые на основе административно-территориального деления, принципах биологической систематики и др. Целью настоящей работы является продолжение анализа общих свойств распределения концентрации фамилии по субпопуляциям при их иерархической подразделенности. Внимание концентрируется на описании фамильной дивергенции субпопуляций. Через показатели дивергенции выражается коэффициент фамильного инбридинга F_s , аналогичный традиционному коэффициенту инбридинга в популяционной генетике. Получено его разложение по уровням иерархии на сумму приращений коэффициента фамильного инбридинга, соответствующих отдельным уровням. Продемонстрирована связь найденного разложения с разложением дисперсии распределения концентрации фамилии по субпопуляциям. При дополнительном предположении о независимости концентрации фамилии и F_s найдена факторизация F_s по уровням иерархии в русле идей С. Райта. Она упрощается при малой фамильной дивергенции субпопуляций, когда коэффициент F_s всей метапопуляции равен сумме средних значений F_s на отдельных уровнях иерархии. Получена величина занижения фамильного инбридинга, когда в качестве единиц наблюдения вместо неподразделенных субпопуляций первого уровня иерархии служат субпопуляции более высокого ранга. Данные результаты являются статистическими характеристиками иерархической структуры, а не особенностью конкретной популяционной системы и не вытекают из той или иной модели микроэволюции. Они вычислительно не зависят от изучаемой иерархической системы, но позволяют характеризовать ее неоднородность количественно. Полученные результаты относятся к сельским и городским иерархическим метапопуляциям как отдельным компонентам всего населения. В Приложении, которое можно читать независимо, приведены без доказательств аналогии найденных свойств применительно к генетической структуре метапопуляции.

Ключевые слова: иерархическая структура популяций, метапопуляции, концентрации фамилии в субпопуляциях человека, характеристики неоднородности метапопуляций, разложение вероятности неизонимных встреч и коэффициента фамильного инбридинга по уровням иерархии.

DOI: 10.31857/S001667582312007X, **EDN:** PSWQUJ

Важность исследований фамильной структуры связана со сходством процессов генного и фамильного дрейфа, что позволяет рассматривать фамилии как квазигенетический маркер [1], а также косвенно судить о генетической структуре населения, важной с медико-биологической точки зрения [2]. Данные по фамилиям, в частности по частоте изонимных браков [3], дают возможность оценивать коэффициент инбридинга (см. критический анализ этого подхода в [4]). Связи фамилий с генетической структурой отражены в монографии [5], исследованиям фамильной структуры населения России посвящена, например,

монография [6]. В работе [7] приведена обширная библиография зарубежных исследований.

Параллельно и одновременно со случайным генным дрейфом в популяциях с общим происхождением протекает их фамильная дивергенция в результате дрейфа фамилий [8, 9], что приводит, в частности, к увеличению дисперсии концентрации фамилии и дисперсии концентрации гена в родственных субпопуляциях. В предыдущей части [10] данной работы внимание фокусировалось на описании свойств дисперсии распределения концентрации фамилии по субпопуляциям в иерархически подразделенной метапопуляции как

характеристике фамильной дивергенции. В следующей публикации [11] подробно рассмотрена общая (полная) вероятность неизоимной встречи (столкновения) индивидуума с фиксированной фамилией с носителем какой-либо другой в субпопуляции, случайно выбранной из рассматриваемой метапопуляции с иерархической структурой.

Целью настоящей работы является продолжение анализа в [11] общих свойств распределения концентрации фамилии по субпопуляциям при их иерархической подразделенности с использованием и обобщениями традиционных подходов популяционной генетики. Рассматривается задача выделения показателей дивергенции, свойственных собственно иерархической системе и независимых от истории ее формирования и взаимодействия между подсистемами, например от внутренних миграционных связей. В данной статье акцент лежит на анализе свойств коэффициента фамильного инбридинга, аналогичного коэффициенту инбридинга в популяционной генетике [12].

Особенности конкретного процесса микроэволюции приводят к своим значениям показателей дивергенции субпопуляций, но на них сказывается и популяционная структура. Поэтому важная задача состоит в том, чтобы выделить свойства показателей дивергенции, присущие самой иерархической структуре и независимые от характера микроэволюционного процесса, изолированности субпопуляций или существования миграций между ними.

Напомним, что под *иерархической системой* понимается система с многоуровневой структурой, элементы которой связаны отношениями подчинения, причем отдельный элемент, в свою очередь, является иерархической системой (более низкого ранга). У нас иерархической системой является метапопуляция, подразделенная на субпопуляции различного уровня иерархии, а подчинение означает вхождение субпопуляции в качестве составной части в субпопуляцию более высокого ранга.

Приведем пояснения относительно обозначений и терминологии. Под концентрациями фамилий в популяции подразумеваются концентрации однофамильцев (носителей данной фамилии). Векторы набраны полужирным шрифтом, к обозначениям фамильных аналогов популяционно-генетических характеристик (дисперсии, коэффициента фамильного инбридинга) добавлено окончание *s* (*Us* и *Fs* соответственно). Символ тождества ($\equiv$) у нас используется в смысле равенства по определению, а символ $\blacktriangleleft$ обозначает конец доказательства.

Дадим обозначения используемых далее характеристик метапопуляции с иерархической структурой и укажем на некоторые из их свойств:

s_i — субпопуляция i -го уровня иерархии (с единицами наблюдения $\{s_i\}$);

$s_{i,n}$ — субпопуляция s_i с единицами наблюдения $\{s_n\}$;

$x(s_i)$ — концентрация рассматриваемой фамилии в субпопуляции s_i ;

$x(s_i|s_j)$ — случайная величина, концентрация рассматриваемой фамилии в субпопуляции s_i , выбираемой наугад из s_j ;

$E_{s_j}\{x(s_i|s_j)\}$ — среднее значение случайной величины $x(s_i|s_j)$ (ее математическое ожидание, E — его символ, нижний индекс указывает на используемую при усреднении субпопуляцию); по концентрациям $\{x(s_i)\}$ находится среднее значение для $x(s_i|s_j)$, равное $x(s_j)$, или по рекуррентной формуле

$$x(s_j) \equiv E_{s_1}\{x(s_1)|s_j\}, \quad x(s_j) \equiv E_{s_{j-1}}\{x(s_{j-1})|s_j\}; \quad (1)$$

$\{x(s_1)|s_m\}$ — начальные данные для концентрации носителей интересующей фамилии в субпопуляциях первого уровня иерархии внутри метапопуляции s_m ;

$V(x(s_i|s_j))$ — дисперсия случайной величины $x(s_i|s_j)$;

$Hs(x(s_j))$, или $Hs(s_j)$ — вероятность (концентрация, доля) неизоимных встреч в s_j индивидуумов с рассматриваемой фамилией и с какой-нибудь иной с учетом их порядка, находящаяся по определению как среднее значение для субпопуляций $\{s_i\}$ в s_j , или по рекуррентной формуле

$$Hs(s_j) \equiv E_{s_1}\{Hs(s_1)|s_j\}, \\ Hs(s_j) \equiv E_{s_{j-1}}\{Hs(s_{j-1})|s_j\}; \quad (2)$$

$\{Hs(s_1)\}$ — начальные значения вероятностей рассматриваемых неизоимных встреч;

$Hs_0(s_i) \equiv x(s_i)(1 - x(s_i))$ — вероятность такой чисто случайной встречи в s_i с учетом порядка ее участников.

Концентрация $x(s_j)$ рассматриваемого признака в группе s_j (фамилии, аллеля или другого признака) согласно формуле полного математического ожидания находится при произвольном разбиении s_j на подгруппы $\{s_i\}$ как

$$x(s_j) \equiv E_{s_1}\{x(s_1)|s_j\} = \\ = E_{s_1}\{E_{s_i}\{x(s_1)|s_i\}|s_j\} = E_{s_1}\{x(s_i)|s_j\}, \quad (3)$$

т.е. как среднее значение для концентраций $\{x(s_i) = E_{s_i}\{x(s_1)|s_i\}\}$ в подгруппах $\{s_i\}$.

Для субпопуляции $s_j \equiv \{s_i\}$ эту формулу можно представить как

$$E_{s_i}\{E_{s_1}\{T(s_1)|s_i\}|s_j\} = E_{s_1}\{T(s_1)|s_j\}, \\ s_j \equiv \{s_i\}, \quad s_i \equiv \{s_1\}, \quad (4)$$

где T — обозначение рассматриваемого признака.

В популяции s , подразделенной на субпопуляции (группы) $\{s_i\}$ с межгрупповой дисперсией

$V_{s_{\text{betw}}}(x(s_1|s))$ и случайными встречами в s_1 с вероятностями $Hs_0(s_1) = x(s_1)(1 - x(s_1))$, согласно эффекту Валунда общая (полная) вероятность рассматриваемой неизонимной встречи $Hs(s)$ в выбранной наугад из s субпопуляции s_1 не совпадает с $Hs_0(s) = x(s)(1 - x(s))$, а находится как

$$Hs(s) = Hs_0(x(s)) - V_{s_{\text{betw}}}(x(s_1|s)). \quad (5)$$

Иначе эту формулу можно записать в виде

$$\begin{aligned} Hs(s) &= Hs_0(x(s)) \left(1 - \frac{V_{s_{\text{betw}}}(x(s_1|s))}{Hs_0(s)} \right) = \\ &= Hs_0(x(s))(1 - Fs(s)), \\ Fs(s) &\equiv V_{s_{\text{betw}}}(x(s_1|s))/Hs_0(s), \end{aligned} \quad (6)$$

где $Fs(s)$ — коэффициент фамильного инбридинга субпопуляции s .

Если субпопуляции $\{s_1\}$ дополнительно сгруппированы в $\{s_2\}$, т.е. s разбита на $\{s_2\}$, а каждая из s_2 на $\{s_1\}$, то по правилу сложения дисперсий полная дисперсия $V_{s_{\text{tot}}}(x(s_1|s))$ концентраций $\{x(s_1)\}$ равна сумме межгрупповой $V_{s_{\text{betw}}}(x(s_2|s))$ и средней внутригрупповой $E_{s_2}\{V_{s_{\text{in}}}(x(s_1|s_2))|s\}$ дисперсий:

$$\begin{aligned} V_{s_{\text{tot}}}(x(s_1|s)) &= \\ &= V_{s_{\text{betw}}}(x(s_2|s)) + E_{s_2}\{V_{s_{\text{in}}}(x(s_1|s_2))|s\}. \end{aligned} \quad (7)$$

Эта формула верна, если в многоуровневой субпопуляции s_m вместо второго уровня рассматривать k -й уровень иерархии:

$$\begin{aligned} V_{s_{\text{tot}}}(x(s_1|s_m)) &= V_{s_{\text{betw}}}(x(s_k|s_m)) + \\ &+ E_{s_k}\{V_{s_{\text{in}}}(x(s_1|s_k))|s_m\}, \quad 2 < k < m. \end{aligned} \quad (8)$$

Данное правило играет большую роль в дальнейшем.

Очевидно, при ходьбе пройденное расстояние складывается из приращений на каждом шагу. Аналогично любую числовую величину можно разбить на последовательность частей, границы которых играют роль своего рода делений измерительной линейки. Расстояния между соседними границами дают приращения измеряемой величины при движении вдоль такой “линейки”, а сумма приращений дает измеряемую величину. Если рассматривать полную дисперсию $V_{s_{\text{tot}}}(x(s_1|s_m))$ и в качестве формальных границ частей взять значения дисперсий $Vs(x(s_k|s_m))$ на уровнях иерархии, то сумма приращений дает разложение полной дисперсии. Согласно этому в предыдущей части [11] такое разложение полной дисперсии концентраций $\{x(s_1)\}$ в метапопуляции s_m с m уровнями иерархии представлено в виде

$$\begin{aligned} V_{s_{\text{tot}}}(x(s_1|s_m)) &= \sum_{i=1}^{m-1} \Delta_i Vs(s_m) \equiv \\ &\equiv \sum_{i=1}^{m-1} (Vs(x(s_i|s_m)) - Vs(x(s_{i+1}|s_m))) = \\ &= \sum_{i=1}^{m-1} E_{s_{i+1}}\{Vs(x(s_i|s_{i+1}))|s_m\}. \end{aligned} \quad (9)$$

Дальнейшее изложение придерживается следующего плана. На основании разложения полной дисперсии концентраций $\{x(s_1)\}$ в метапопуляции s_m с m уровнями иерархии получено разложение коэффициента фамильного инбридинга метапопуляции по уровням иерархии. Его изменения от одного уровня к соседнему определяются различиями между дисперсиями концентраций фамилии на смежных уровнях. Для субпопуляции каждого ранга введено понятие коэффициента фамильного инбридинга при условии заданного уровня у единиц наблюдения. Рассмотрены разложения коэффициента фамильного инбридинга метапопуляции в предположении случайных встреч в субпопуляциях, служащих единицами наблюдения, и при нарушении этого предположения. Исследовано понижение оценки дивергенции субпопуляций, когда не учитывается неоднородность (подрозделенность) субпопуляций, служащих единицами наблюдения.

Далее обобщен подход С. Райта к анализу иерархически подразделенной метапопуляции при дополнительном условии независимости концентрации аллеля и коэффициента фамильного инбридинга Fs в контексте анализа фамильной структуры. Получена факторизация Fs по уровням иерархии. Для нее найдена простая аппроксимация в случае малых коэффициентов. Полученные результаты относятся к сельским и городским иерархическим метапопуляциям как отдельным компонентам всего населения. В Приложении приведены без доказательств аналоги найденных в настоящей серии работ результатов по анализу фамильной структуры применительно к данным популяционной генетики. Оно предназначено для интересующихся популяционно-генетическими аспектами иерархически структурированных подразделенных популяций. Приложение можно читать независимо от предыдущего анализа фамильной структуры.

РАЗЛОЖЕНИЕ КОЭФФИЦИЕНТА ФАМИЛЬНОГО ИНБРИДИНГА ПО УРОВНЯМ ИЕРАРХИИ

При чисто случайных столкновениях индивидов в популяции s вероятность $Hs(s)$ встречи неизонимной пары лиц (одного с заданной фамилией с концентрацией $x(s)$ и другого с любой иной фамилией) с учетом порядка в паре имеет вид $Hs_0(s)$

$= x(s)(1 - x(s))$. Предположим, что столкновения неслучайны и вероятность $Hs(s)$ встреч описывается с помощью коэффициента фамильного инбридинга Fs как

$$Hs(s|Fs(s)) = x(s)(1 - x(s))(1 - Fs(s)) \equiv Hs_0(s)(1 - Fs(s)).$$

Отсюда коэффициент фамильного инбридинга выражается как

$$Fs(s) \equiv (Hs_0(s) - Hs(s))/Hs_0(s). \quad (10)$$

На разложении дисперсии и вероятности встреч основывается и разложение коэффициента фамильного инбридинга в соответствии с уровнями иерархии. Выбор для вероятности встреч в единице наблюдения s_1 (с концентрацией $x(s_1)$ интересующей фамилии) значения $Hs_0(s_1) = x(s_1)(1 - x(s_1))$, соответствующего чисто случайным столкновениям, приводит к следующему обобщению эффекта Валунда.

Результат 1. Пусть в иерархически подразделенной метапопуляции s_m с t уровнями иерархии в качестве единиц наблюдения служат субпопуляции $\{s_1\}$ первого уровня. Положим, что концентрации интересующей фамилии в $\{s_1\}$ равны $\{x(s_1)\}$, а вероятности встречи в них пары индивидуумов с этой фамилией и с какой-нибудь иной с учетом их порядка находятся как

$$\{Hs(s_1) = Hs_0(s_1) \equiv x(s_1)(1 - x(s_1))\}.$$

Тогда разложения коэффициента фамильного инбридинга $Fs(s_m)$ по компонентам отдельных уровней иерархии можно записать как

$$\begin{aligned} Fs(s_m) &= \sum_{i=1}^{m-1} \frac{Vs(x(s_i|s_m)) - Vs(x(s_{i+1}|s_m))}{Hs_0(s_m)} = \\ &= \sum_{i=1}^{m-1} \frac{E_{s_{i+1}} \{Vs_{in}(x(s_i|s_{i+1}))|s_m\}}{Hs_0(s_m)} = \\ &= \sum_{i=1}^{m-1} \Delta_i Fs(s_m), \quad (11) \\ \Delta_i Fs(s_m) &\equiv \frac{Vs(x(s_i|s_m)) - Vs(x(s_{i+1}|s_m))}{Hs_0(s_m)} = \\ &= \frac{E_{s_{i+1}} \{Vs_{in}(x(s_i|s_{i+1}))|s_m\}}{Hs_0(s_m)} \equiv \\ &\equiv Fs(s_{m,i}) - Fs(s_{m,i+1}) \geq 0. \end{aligned}$$

Вероятности $Hs(s_m)$ рассматриваемой встречи в наугад выбранной из s_m субпопуляции s_1 представляемы в виде

$$\begin{aligned} Hs(s_m) &= Hs_0(s_m) \times \\ &\times \left(1 - \sum_{i=1}^{m-1} \frac{Vs(x(s_i|s_m)) - Vs(x(s_{i+1}|s_m))}{Hs_0(s_m)} \right) = \\ &= Hs_0(s_m) \left(1 - \sum_{i=1}^{m-1} \frac{E_{s_{i+1}} \{Vs_{in}(x(s_i|s_{i+1}))|s_m\}}{Hs_0(s_m)} \right) = \\ &= Hs_0(s_m) \left(1 - \sum_{i=1}^{m-1} \Delta_i Fs(s_m) \right). \end{aligned} \quad (12)$$

Здесь

$$\begin{aligned} \Delta_i Fs(s_m) &\equiv Fs(s_{m,i}) - Fs(s_{m,i+1}) = \\ &= E_{s_{i+1}} \{Vs_{in}(x(s_i|s_{i+1}))|s_m\} / Hs_0(s_m), \\ Fs(s_m) &\equiv Vs(x(s_1|s_m)) / Hs_0(s_m), \\ Fs(s_{m,i}) &\equiv Vs_{in}(x(s_i|s_m)) / Hs_0(s_m), \\ Hs(s_m) &\equiv E_{s_1} \{Hs(s_1|s_m)\}, \\ Hs_0(s_m) &= x(s_m)(1 - x(s_m)), \end{aligned} \quad (13)$$

$Fs(s_m)$ и $Fs(s_{m,i})$ обозначают коэффициенты фамильного инбридинга в метапопуляции s_m и в $s_{m,i}$, когда в качестве единиц наблюдения служат субпопуляции $\{s_1\}$ и $\{s_i\}$ с уровнями иерархии 1 и i при условии, что вероятности рассматриваемых встреч в них равны $\{Hs_0(s_1)\}$ и $\{Hs_0(s_i)\}$ соответственно; $Hs(s_m)$ — общая вероятность рассматриваемых встреч в s_m .

Доказательство. В иерархически подразделенной метапопуляции s_m при $\{Hs(s_1) \equiv Hs_0(s_1) = x(s_1)(1 - x(s_1))\}$ согласно эффекту Валунда (5) и разложению (9) дисперсии $Vs(x(s_1|s_m))$

$$\begin{aligned} Hs(s_m) &= Hs_0(s_m) - Vs(x(s_1|s_m)) = \\ &= Hs_0(s_m) - \sum_{i=1}^{m-1} (Vs(x(s_i|s_m)) - Vs(x(s_{i+1}|s_m))) = \\ &= Hs_0(s_m) \left(1 - \sum_{i=1}^{m-1} \frac{Vs(x(s_i|s_m)) - Vs(x(s_{i+1}|s_m))}{Hs_0(s_m)} \right) = \\ &= Hs_0(s_m) \left(1 - \sum_{i=1}^{m-1} \frac{E_{s_{i+1}} \{Vs_{in}(x(s_i|s_{i+1}))|s_m\}}{Hs_0(s_m)} \right). \end{aligned}$$

В последнем равенстве использовано правило сложения дисперсий (8), когда в метапопуляции $s = s_m$ рассматриваются только три уровня иерархии (m , $i+1$, i). Тогда дисперсии $Vs_{tot}(x(s_i|s_m))$, $Vs_{betw}(x(s_{i+1}|s_m))$ и $E_{s_{i+1}} \{Vs_{in}(x(s_i|s_{i+1}))|s_m\}$ рассматриваются как полная, межгрупповая и средняя внутригрупповая дисперсии соответственно. В силу этого правила разность первых двух равна третьей, т.е. уменьшение межгрупповой дисперсии при переходе от i к смежному более высокому уровню равно средней внутригрупповой дисперсии для $\{s_{i+1}\}$:

$$Vs(x(s_i|s_m)) - Vs(x(s_{i+1}|s_m)) = E_{s_{i+1}} \{Vs_{in}(x(s_i|s_{i+1}))|s_m\}.$$

Согласно данному выше определению (13) коэффициента фамильного инбридинга $Fs(s_m)$ и (9) находим

$$\begin{aligned} Fs(s_m) &\equiv \frac{Vs(x(s_1|s_m))}{Hs_0(s_m)} = \\ &= \sum_{i=1}^{m-1} \frac{E_{s_{i+1}}\{Vs(x(s_i|s_{i+1})|s_m)\}}{Hs_0(s_m)} = \\ &= \sum_{i=1}^{m-1} \left(\frac{Vs(x(s_i|s_m))}{Hs_0(s_m)} - \frac{Vs(x(s_{i+1}|s_m))}{Hs_0(s_m)} \right). \end{aligned}$$

С учетом определения $\Delta_i Fs(s_m)$ подстановка данного разложения для Fs в $Hs_0(1 - Fs)$ дает (12). ◀

Рассмотрим компоненты i -го слагаемого в правой части последней формулы:

$$\begin{aligned} \frac{Vs(x(s_i|s_m))}{Hs_0(s_m)} &\equiv Fs(s_{m,i}), \\ \frac{Vs(x(s_{i+1}|s_m))}{Hs_0(s_m)} &\equiv Fs(s_{m,i+1}). \end{aligned}$$

Здесь отношение $Vs(x(s_i|s_m))/Hs_0(s_m)$ согласно определению (13) равно коэффициенту фамильного инбридинга $Fs(s_{m,i})$ метапопуляции s_m , где в качестве единиц наблюдения используются субпопуляции $\{s_i\}$ на i -м уровне иерархии и встречи (столкновения) внутри них чисто случайны. Данный коэффициент характеризует подразделенность метапопуляции на i -м уровне (интерпретируется как коэффициент фамильного инбридинга $Fs(s_m)$ у метапопуляции s_m при отсутствии в ней подразделенности на уровне ниже i). Аналогично $Fs(s_{m,i+1})$ характеризует подразделенность метапопуляции на уровне $i+1$, а разность $Fs(s_{m,i}) - Fs(s_{m,i+1})$ дает вклад $\Delta_i Fs(s_m)$ в коэффициент фамильного инбридинга $Fs(s_m)$.

Случайность встреч внутри любой из субпопуляций $\{s_i\}$ означает однородность каждой в смысле отсутствия препятствий для встреч. При этом формально s_i едина, а субпопуляции $\{s_{i-1}\}$ как бы условны (границы, при существовании которых столкновения происходят только между индивидами одной и той же субпопуляции, у них отсутствуют). Базовыми единицами являются такие субпопуляции, внутри которых столкновения случайны (значения концентраций в них служат начальными данными). Их группировка не оказывает влияния на общую (полную) вероятность встреч во всей метапопуляции подобно тому, как количество гетерозигот в метапопуляции складывается из количества в ее подразделениях независимо от их группировок при подсчетах (на сумму не влияет группировка слагаемых).

Рассмотрим погрешности выражений для вероятности рассматриваемых встреч и для коэффициента фамильного инбридинга, вносимые использованием вместо начальных данных (концентрации

интересующих встреч в единицах наблюдения $\{s_1\}$ со случайными столкновениями) агрегированных сведений для субпопуляций более высокого ранга $\{s_n\}$. В общем случае выбор значений $x(s_n)(1 - x(s_n))$ для вероятности встреч в субпопуляциях $\{s_n\}$, взятых в качестве единиц наблюдения уровня n , не обоснован. Использование для вероятности встреч в единице наблюдения s_n (с концентрацией $x(s_n)$ интересующей фамилии) значения $Hs_0(s_n) = x(s_n) \times (1 - x(s_n))$, соответствующего чисто случайным столкновениям, вместо неизвестного реального значения приводит к следующему утверждению.

Следствие 2. Пусть в иерархически подразделенной метапопуляции s_m с t уровнями иерархии в качестве единиц наблюдения вместо субпопуляций первого уровня $\{s_1\}$ с концентрациями фамилии $\{x(s_1)\}$ и случайными столкновениями рассматриваются субпопуляции уровня n ($1 < n < t - 1$) с концентрациями $\{x(s_n)\}$. Предположим, что при отсутствии или игнорировании сведений о наличии в них препятствий для встреч индивидов принимается решение о случайных столкновениях в каждой из субпопуляций $\{s_n\}$, т.е. вероятность $Hs(s_{m,n})$ встреч в s_m пары индивидов с интересующей фамилией и с какой-нибудь иной с учетом их порядка (в наугад выбранной из s_m субпопуляции s_n) находится как

$$\begin{aligned} Hs(s_{m,n}) &\equiv E_{s_n}\{Hs(s_n)|s_m\} = E_{s_n}\{Hs_0(s_n)|s_m\}, \\ Hs_0(s_n)(s_m) &\equiv x(s_n)(1 - x(s_n)), \end{aligned}$$

где $s_{m,n}$ обозначает метапопуляцию s_m с единицами наблюдения $\{s_n\}$.

Тогда коэффициент фамильного инбридинга $Fs(s_{m,n})$, определяемый из условия $Hs(s_{m,n}) = Hs_0(s_m) \times (1 - Fs(s_{m,n}))$, можно разложить по компонентам на уровнях иерархии как

$$\begin{aligned} Fs(s_{m,n}) &= \sum_{i=n}^{m-1} \frac{Vs(x(s_i|s_m)) - Vs(x(s_{i+1}|s_m))}{x(s_m)(1 - x(s_m))} = \\ &= \sum_{i=n}^{m-1} \frac{E_{s_{i+1}}\{Vs(x(s_i|s_{i+1})|s_m)\}}{x(s_m)(1 - x(s_m))} \equiv \sum_{i=n}^{m-1} \Delta_i Fs(s_m), \\ \Delta_i Fs(s_m) &\equiv \frac{Vs(x(s_i|s_m)) - Vs(x(s_{i+1}|s_m))}{x(s_m)(1 - x(s_m))} = \\ &= \frac{E_{s_{i+1}}\{Vs(x(s_i|s_{i+1})|s_m)\}}{x(s_m)(1 - x(s_m))} \geq 0. \end{aligned}$$

Коэффициент $Fs(s_{m,n})$ меньше значения $Fs(s_m)$, соответствующего единице наблюдения $\{s_1\}$, на сумму вкладов неучтенных уровней иерархии от 1 до $n - 1$:

$$\begin{aligned}
Fs(s_m) - Fs(s_{m,n}) &= \\
&= \sum_{i=1}^{m-1} \frac{Vs(x(s_i|s_m)) - Vs(x(s_{i+1}|s_m))}{x(s_m)(1-x(s_m))} - \\
&- \sum_{i=n}^{m-1} \frac{Vs(x(s_i|s_m)) - Vs(x(s_{i+1}|s_m))}{x(s_m)(1-x(s_m))} = \\
&= \sum_{i=1}^{n-1} \frac{E_{s_{i+1}}\{Vs(x(s_i|s_{i+1}))|s_m\}}{x(s_m)(1-x(s_m))} = \sum_{i=1}^{n-1} \Delta_i Fs(s_m) \geq 0.
\end{aligned}$$

Доказательство. Воспользуемся эффектом Валунда (5), согласно которому полная вероятность рассматриваемых встреч в метапопуляции s_m , где в качестве единиц наблюдения служат субпопуляции $\{s_1\}$, находится как

$$\begin{aligned}
Hs(s_m) &\equiv E_{s_1}\{Hs_0(s_1)|s_m\} = \\
&= Hs_0(s_m) \left(1 - \frac{Vs_{\text{betw}}(x(s_1|s_m))}{Hs_0(s_m)}\right) = \\
&= Hs_0(s_m)(1 - Fs(s_m)), \\
Fs(s_m) &\equiv \frac{Vs_{\text{betw}}(x(s_1|s_m))}{Hs_0(s_m)} \geq 0.
\end{aligned}$$

Пусть при анализе метапопуляции s в качестве единиц наблюдения выбраны субпопуляции $\{s_n\}$ на n -м уровне иерархии, например из-за отсутствия информации о существующей в них подразделенности или о наличии других препятствий для встреч индивидуумов, или же из-за игнорирования таких сведений. При упрощенном предположении о случайных столкновениях в каждой из субпопуляций $\{s_n\}$ получаем вероятность встреч $Hs(s_{m,n})$ в s_m заменой в приведенной формуле s_1 на s_n :

$$\begin{aligned}
Hs(s_{m,n}) &\equiv E_{s_n}\{Hs_0(s_n)|s_m\} = \\
&= Hs_0(s_m) \left(1 - \frac{Vs_{\text{betw}}(x(s_n|s_m))}{Hs_0(s_m)}\right) = \\
&= Hs_0(s_m)(1 - Fs(s_{m,n})), \\
Fs(s_{m,n}) &\equiv \frac{Vs_{\text{betw}}(x(s_n|s_m))}{Hs_0(s_m)} \geq 0.
\end{aligned}$$

В результате выбора единицы наблюдения $\{s_n\}$ вместо $\{s_1\}$ вероятность встреч увеличится согласно (5)–(6) на

$$\begin{aligned}
Hs(s_{m,n}) - Hs(s_m) &= \\
&= Vs_{\text{tot}}(x(s_1|s_m)) - Vs_{\text{betw}}(x(s_n|s_m)) = \\
&= Hs_0(s_m)(Fs(s_m) - Fs(s_{m,n})) = \\
&= E_{s_n}\{Vs_{\text{in}}(x(s_1|s_n))|s_m\} \geq 0,
\end{aligned}$$

где мы учли выражение для разности между общей и межгрупповой дисперсиями согласно (8).

С другой стороны, в соответствии с представлением $Hs = Hs_0(1 - Fs)$ коэффициент фамилного инбридинга уменьшится на

$$\begin{aligned}
Fs(s_m) - Fs(s_{m,n}) &= \\
&= E_{s_n}\{Vs_{\text{in}}(x(s_1|s_n))|s_m\} / Hs_0(s_m) \geq 0. \blacktriangleleft
\end{aligned}$$

Следствие 3. В условиях предыдущего следствия разложение вероятности встреч $Hs(s_{m,n}) \equiv E_{s_n}\{Hs(s_n)|s_m\}$ имеет вид

$$\begin{aligned}
Hs(s_{m,n}) &= x(s_m)(1-x(s_m))(1-Fs(s_{m,n})) = \\
&= Hs_0(s_m) \left(1 - \sum_{i=n}^{m-1} \frac{E_{s_{i+1}}\{Vs(x(s_i|s_{i+1}))|s_m\}}{x(s_m)(1-x(s_m))}\right) \equiv \\
&\equiv Hs_0(s_m) \left(1 - \sum_{i=n}^{m-1} \Delta_i Fs(s_m)\right).
\end{aligned}$$

Вероятность встреч $Hs(s_{m,n})$ больше значения $Hs(s_m)$ при единицах наблюдения $\{s_1\}$ первого уровня иерархии, где $Hs(s_1) = Hs_0(s_1) \equiv x(s_1)(1-x(s_1))$, на величину

$$\begin{aligned}
&\sum_{i=1}^{n-1} (Vs(x(s_i|s_m)) - Vs(x(s_{i+1}|s_m))) = \\
&= \sum_{i=1}^{n-1} E_{s_{i+1}}\{Vs(x(s_i|s_{i+1}))|s_m\} = \\
&= Vs(x(s_1|s_n)) \geq 0.
\end{aligned}$$

Отклонение $Hs(s_{m,n}) - Hs_0(s_m)$ возрастает на $E_{s_{n+1}}\{Vs(x(s_n|s_{n+1}))|s_m\} \geq 0$ при повышении ранга единицы наблюдения субпопуляции от n до следующего уровня.

Коэффициент фамилного инбридинга с повышением ранга единицы наблюдения от n до следующего уровня, напротив, уменьшается на $E_{s_{n+1}}\{Vs_{\text{in}}(x(s_n|s_{n+1}))|s_m\} / Hs_0(s_m)$.

Доказательство. Суммирование неотрицательных слагаемых $\Delta_i Fs(s_m)$ в разложении (11) коэффициента фамилного инбридинга $Fs(s_{m,n})$ начинается с номера иерархии у единицы наблюдения, т.е. не с единицы, как для $Fs(s_m)$, а с n . Следовательно, $Fs(s_{m,n})$ оказывается меньше значения $Fs(s_m)$ на величину суммы вкладов неучтенных уровней иерархии от 1 до $n-1$:

$$\begin{aligned}
Fs(s_{m,n}) &= \sum_{i=n}^{m-1} \Delta_i Fs(s_m), \\
Fs(s_m) - Fs(s_{m,n}) &= \sum_{i=1}^{n-1} \Delta_i Fs(s_m) = \\
&= \sum_{i=1}^{n-1} E_{s_{i+1}}\{Vs(x(s_i|s_{i+1}))|s_m\} / Hs_0(s_m) \geq 0.
\end{aligned}$$

Вероятность встреч $Hs(s_{m,n})$ оказывается больше значения $Hs(s_m)$ при единице наблюдения $\{s_1\}$:

$$\begin{aligned}
Hs(s_{m,n}) &= x(s_m)(1 - x(s_m))(1 - Fs(s_{m,n})) = \\
&= Hs_0(s_m) \left(1 - \sum_{i=n}^{m-1} \Delta_i Fs(s_m) \right), \\
Hs(s_{m,n}) - Hs_0(s_m) &= Hs_0(s_m)(1 - Fs(s_{m,n})) - \\
&\quad - Hs_0(s_m)(1 - Fs(s_m)) \quad (14) \\
&= Hs_0(s_m)(Fs(s_m) - Fs(s_{m,n})) = \\
&= E_{s_n} \{ Vs_{in}(x(s_i|s_n)) | s_m \} / Hs_0(s_m) = \\
&= \sum_{i=1}^{n-1} E_{s_{i+1}} \{ Vs(x(s_i|s_{i+1})) | s_m \} / Hs_0(s_m) \geq 0.
\end{aligned}$$

Очевидно расхождение $Hs(s_{m,n}) - Hs(s_m)$ при увеличении единицы наблюдения субпопуляции от n до $n + 1$ возрастает на величину $\Delta_n Hs(s_m) \equiv Hs_0(s_m) \Delta_n Fs(s_m) = E_{s_{n+1}} \{ Vs(x(s_n|s_{n+1})) | s_m \}$. Отсюда соответствующим изменением коэффициента фамильного инбридинга будет

$$\begin{aligned}
\Delta_n Fs(s_m) &= \Delta_n Hs(s_m) / Hs_0(s_m) = \\
&= E_{s_{n+1}} \{ Vs(x(s_n|s_{n+1})) | s_m \} / Hs_0(s_m). \blacktriangleleft
\end{aligned}$$

ФАКТОРИЗАЦИЯ КОЭФФИЦИЕНТА ФАМИЛЬНОГО ИНБРИДИНГА ПО УРОВНЯМ ИЕРАРХИИ ПРИ ЕГО НЕЗАВИСИМОСТИ ОТ КОНЦЕНТРАЦИИ ФАМИЛИИ

Другой вариант анализа иерархически подразделенной метапопуляции следует подходу С. Райта [12] и *требует дополнительного предположения о независимости концентрации x фамилии и коэффициента фамильного инбридинга Fs* . Отметим, что они разным образом описывают субпопуляции. Если концентрации характеризуют распространенность фамилий по отдельности, то коэффициент фамильного инбридинга характеризует распространенность пар фамилий при встречах носителей (как бы отталкивание или притяжение). Она может принимать разные значения даже при фиксированных концентрациях отдельных фамилий. Поэтому указанное предположение о независимости кажется естественным.

Согласно фамильному аналогу (5) эффекта Валунда в метапопуляции s_2 , подразделенной на субпопуляции первого уровня иерархии $\{s_1\}$ с концентрациями заданной фамилии $\{x(s_1)\}$ и случайными встречами индивидумов в них, вероятность $Hs(s_2)$ встречи в s_2 лиц с заданной фамилией и с любой иной находится как

$$\begin{aligned}
Hs(s_2) &= x(s_2)(1 - x(s_2)) - Vs(x(s_1|s_2)) = \\
&= x(s_2)(1 - x(s_2))(1 - Fs(s_2)), \\
Fs(s_2) &\equiv \frac{Vs(x(s_1|s_2))}{x(s_2)(1 - x(s_2))} \geq 0.
\end{aligned}$$

Далее для лучшей ориентации в соответствующих формулах используется обозначение вероятности рассматриваемых встреч как $Hs(s_{m,i})$, где $s_{m,i}$ относится к метапопуляции s_m , в которой единицами наблюдения являются субпопуляции $\{s_i\}$ на i -м уровне иерархии со случайными встречами индивидумов в них. Сохраним обозначение $Hs(s_m)$ за вероятностью встреч при единицах наблюдения первого уровня. Когда s_m разбита на $\{s_i\}$ и встречи в $\{s_i\}$ случайны, то $Hs(s_i) = Hs_0(s_i) = x(s_i)(1 - x(s_i))$, а

$$\begin{aligned}
Hs(s_{m,i}) &\equiv E_{s_i} \{ x(s_i)(1 - x(s_i)) | s_m \} = \\
&= x(s_m)(1 - x(s_m)) - Vs(x(s_i|s_m)) = \\
&= x(s_m)(1 - x(s_m))(1 - Fs(s_{m,i})), \quad (15) \\
Fs(s_{m,i}) &\equiv \frac{Vs(x(s_i|s_m))}{x(s_m)(1 - x(s_m))} = \frac{Vs(x(s_i|s_m))}{Hs_0(s_m)}.
\end{aligned}$$

Повторим, что коэффициент фамильного инбридинга $Fs(s_{m,i})$ субпопуляции s_m характеризует дивергенцию субпопуляций $\{s_i\}$ внутри s_m . Его нижний индекс i указывает на служащие единицами наблюдения субпопуляции $\{s_i\}$, встречи в которых предполагаются случайными. В обозначении $Fs(s_i)$ подразумевается, что единицами наблюдения являются субпопуляции $\{s_1\}$.

Рассмотренная выше подразделенная популяция s_2 может быть одной из частей популяции более высокого ранга s_3 . В случае трех уровней иерархии (т.е. при $m = 3$) по определению $Hs(s_3) \equiv E_{s_1} \{ Hs(s_1) | s_3 \}$ и в соответствии с (2) $Hs(s_3) = E_{s_2} \{ Hs(s_2) | s_3 \}$, т.е. концентрация $Hs(s_3)$ является средним значением как для концентраций Hs в $\{s_1\}$, так и (рекуррентно) для концентраций Hs в $\{s_2\}$. В терминах коэффициентов фамильного инбридинга Fs имеем

$$\begin{aligned}
Hs(s_3) &= E_{s_2} \{ Hs(s_2) | s_3 \} = \\
&= E_{s_2} \{ Hs_0(s_2)(1 - Fs(s_{2,1})) | s_3 \}.
\end{aligned}$$

При независимости Fs и концентрации фамилии x эта формула для $Hs(s_3)$ упрощается, поскольку тогда математическое ожидание произведения равно произведению математических ожиданий его сомножителей:

$$\begin{aligned}
Hs(s_3) &= E_{s_2} \{ Hs_0(s_2) | s_3 \} E_{s_2} \{ (1 - Fs(s_{2,1})) | s_3 \} = \\
&= Hs_0(s_3)(1 - Fs(s_{3,2}))(1 - E_{s_2} \{ Fs(s_{2,1}) | s_3 \}).
\end{aligned}$$

Рассмотрим последовательность значений Hs при малых уровнях иерархии:

$$\begin{aligned}
Hs(s_1) &= Hs_0(s_1), \quad Hs(s_2) = Hs_0(s_2)(1 - Fs(s_{2,1})), \\
Hs(s_3) &= Hs_0(s_3)(1 - Fs(s_{3,2}))(1 - E_{s_2} \{ Fs(s_{2,1}) | s_3 \}).
\end{aligned}$$

Из нее просматривается некоторая закономерность зависимости вида факторизации $Hs(s_m)$ от количества уровней иерархии. Первым сомножителем в факторизации является $Hs_0(s_m)$, который при рекуррентном переходе к $Hs(s_{m+1})$ дает согласно фа-

мильному аналогу (5) эффекта Валунда вместо себя два сомножителя: $Hs_0(s_{m+1})$ и $1 - Fs(s_{m+1,m})$. Вторым сомножителем является $1 - Fs(s_{m,m-1})$, заменяемый в результате усреднения по m при рекуррентном переходе к $Hs(s_{m+1})$ на $1 - E_{s_m}\{Fs(s_{m,m-1})|s_{m+1}\} = 1 - E_{s_{i+1}}\{Fs(s_{i+1,i})|s_{m+1}\}_{i=m-1}$. Прочие сомножители представимы как $1 - E_{s_{i+1}}\{Fs(s_{i+1,i})|s_m\}$ и заменяются на $1 - E_{s_{i+1}}\{Fs(s_{i+1,i})|s_{m+1}\}$, где $E_{s_{i+1}}\{Fs(s_{i+1,i})|s_{m+1}\}$ интерпретируется как средний коэффициент фамильного инбридинга субпопуляций $\{s_{i+1,i}\}$ на уровне $i+1$ с соответствующими единицами наблюдения уровня i в популяции s_{m+1} .

Таким образом, при очередном повышении уровня иерархии у s_m (включении s_m в качестве составной части в метапопуляцию s_{m+1}) повышается на единицу уровень иерархии аргумента у Hs_0 и количество сомножителей в факторизации $Hs(s_{m+1})$. При этом добавленный сомножитель равен разности между единицей и коэффициентом фамильного инбридинга $Fs(s_{m+1,m})$ у метапопуляции s_{m+1} (у остальных сомножителей того же типа, также включаемых в s_{m+1} , фигурирует средний коэффициент фамильного инбридинга для субпопуляций соответствующего уровня). Для метапопуляции s_3 разложение на множители можно представить в виде

$$\begin{aligned} Hs(s_3) &= Hs_0(s_3)(1 - Fs(s_{3,2})) \times \\ &\times \prod_{j=1}^{3-2} (1 - E_{s_{j+1}}\{Fs(s_{j+1,j})|s_3\}), \\ Fs(s_{3,2}) &\equiv Vs(x(s_2|s_3))/Hs_0(s_3). \end{aligned}$$

Здесь выражение с символом произведения $\prod$ используется как подсказка для обобщения формулы на произвольный уровень иерархической подразделенности.

Результат 4. Пусть рассматривается метапопуляция s_m с иерархической структурой подразделенности на субпопуляции. Положим, что концентрации заданной фамилии в субпопуляциях $\{s_i\}$ самого низкого уровня иерархии (единицы наблюдения) внутри s_m равны $\{x(s_i|s_m)\}$, вероятность $Hs(s_i|Fs_i)$ встречи в s_i пары индивидуумов с этой фамилией и с какой-нибудь другой (с учетом их порядка) равна

$$x(s_i)(1 - x(s_i))(1 - Fs(s_i)) = Hs_0(s_i)(1 - Fs(s_i)),$$

где $Fs(s_i)$ характеризует нарушение случайности встреч в $\{s_i\}$.

Тогда общая (полная) вероятность $Hs(s_m|Fs_1)$, т.е. вероятность такой встречи в наугад выбранной из s_m субпопуляции s_i при независимости в совокупности концентрации фамилии x и коэффициентов фамильного инбридинга Fs , факторизуется как

$$\begin{aligned} Hs(s_m|Fs_1) &= Hs_0(s_m)(1 - E_{s_1}\{Fs(s_1)|s_m\}) \times \\ &\times (1 - Fs(s_{m,m-1})) \prod_{i=1}^{m-2} (1 - E_{s_{i+1}}\{Fs(s_{i+1,i})|s_m\}), \\ Fs(s_{i+1,i}) &\equiv Vs(x(s_{i+1}|s_i))/Hs_0(s_{i+1}). \end{aligned} \quad (16)$$

Доказательство. Выясненные выше свойства рекуррентного преобразования для $Hs(s_m)$, согласно которым при увеличении m на единицу происходит увеличение количества специфического вида сомножителей в факторизации $Hs(s_m)$, подсказывают (16). Докажем по индукции предполагаемую формулу зависимости Hs от уровня иерархии m метапопуляции s_m . Согласно предыдущим результатам очевидно существует m , при котором $Hs(s_m)$ представимо в требуемом виде. Покажем, что данный вид сохраняется и для $Hs(s_{m+1})$. По рекуррентной формуле (2)

$$\begin{aligned} Hs(s_{m+1}|Fs_1) &\equiv E_{s_m}\{Hs(s_m)|s_{m+1}\} = \\ &= E_{s_m}\left\{Hs_0(s_m)(1 - E_{s_1}\{Fs(s_1)|s_m\}) \times \right. \\ &\times (1 - Fs(s_{m,m-1})) \prod_{i=1}^{m-2} (1 - E_{s_{i+1}}\{Fs(s_{i+1,i})|s_m\})|s_{m+1}\Big\}, \\ Hs_0(s_m) &\equiv x(s_m)(1 - x(s_m)). \end{aligned}$$

При независимости в совокупности x и Fs правая часть данного выражения равна произведению математических ожиданий соответствующих сомножителей по отдельности. Рассмотрим по очереди математические ожидания для сомножителей $Hs_0(s_m) = x(s_m)(1 - x(s_m))$, $1 - E_{s_1}\{Fs(s_1)|s_m\}$, $1 - Fs(s_{m,m-1})$ и для отдельного сомножителя $1 - E_{s_{i+1}}\{Fs(s_{i+1,i})|s_m\}$ произведения $\prod$. Согласно предыдущим результатам

$$\begin{aligned} 1. \quad &E_{s_m}\{x(s_m)(1 - x(s_m))|s_{m+1}\} = \\ &= x(s_{m+1})(1 - x(s_{m+1}))(1 - Fs(s_{m+1,m})). \\ 2. \quad &E_{s_m}\{1 - E_{s_1}\{Fs(s_1)|s_m\}|s_{m+1}\} = \\ &= 1 - E_{s_m}\{E_{s_1}\{Fs(s_1)|s_m\}|s_{m+1}\} = \\ &= 1 - E_{s_1}\{Fs(s_1)|s_{m+1}\}. \\ 3. \quad &E_{s_m}\{1 - Fs(s_{m,m-1})|s_{m+1}\} = \\ &= 1 - E_{s_m}\{Fs(s_{m,m-1})|s_{m+1}\} = \\ &= 1 - E_{s_{i+1}}\{Fs(s_{i+1,i})|s_{m+1}\}_{i=m-1}. \\ 4. \quad &E_{s_m}\{1 - E_{s_{i+1}}\{Fs(s_{i+1,i})|s_m\}|s_{m+1}\} = \\ &= 1 - E_{s_m}\{E_{s_{i+1}}\{Fs(s_{i+1,i})|s_m\}|s_{m+1}\} = \\ &= 1 - E_{s_{i+1}}\{Fs(s_{i+1,i})|s_{m+1}\}. \end{aligned}$$

Поясним эти равенства.

1. Правая часть п. 1 следует из формулы (6) эффекта Валунда с коэффициентом фамильного

инбридинга при соответствующих уровнях иерархии. В доказываемом выражении для $Hs(s_{m+1})$ первый сомножитель $x(s_{m+1})(1 - x(s_{m+1}))$ в п. 1 входит в $Hs(s_{m+1})$ как $Hs_0(s_{m+1}|Fs_1)$ до символа произведения $\prod$. Полученный в п. 1 второй сомножитель $1 - Fs(s_{m+1}, m)$ также входит в $Hs(s_{m+1})$ до символа произведения $\prod$.

2. В п. 2 сомножитель $1 - E_{s_m}\{Fs(s_1)|s_m\}|s_{m+1}$, очевидно, равен $1 - E_{s_1}\{Fs(s_1)|s_{m+1}\}$ согласно (4), что соответствует его значению для разложения Hs в метапопуляции s_{m+1} , а не s_m . Он сохраняет свое место до символа произведения $\prod$.

3. Сомножитель $E_{s_m}\{1 - Fs(s_{m,m-1})|s_{m+1}\}$ равен $1 - E_{s_{i+1}}\{Fs(s_{i+1}, i)|s_{m+1}\}|_{i=m-1}$, т.е. его можно переместить под знак произведения $\prod$, где он соответствует значению i , равному $m-1$. В результате верхний предел произведения увеличится на единицу и станет равным $m-1$ вместо $m-2$.

4. Каждый сомножитель произведения $\prod$ в $Hs(s_m)$ представим в виде $1 - E_{s_{i+1}}\{Fs(s_{i+1}, i)|s_m\}$ и при переходе к s_{m+1} преобразуется как

$$\begin{aligned} E_{s_m}\{1 - E_{s_{i+1}}\{Fs(s_{i+1}, i)|s_m\}|s_{m+1}\} &= \\ = 1 - E_{s_m}\{E_{s_{i+1}}\{Fs(s_{i+1}, i)|s_m\}|s_{m+1}\} &= \\ = E_{s_{i+1}}\{Fs(s_{i+1}, i)|s_{m+1}\}. \end{aligned}$$

Подстановка полученных выражений в $E_{s_m}\{Hs(s_m)|s_{m+1}\}$ при сделанном предположении относительно факторизации $Hs(s_m)$ дает

$$\begin{aligned} Hs(s_{m+1}|Fs_1) &\equiv E_{s_m}\{Hs(s_m)\} = \\ &= Hs_0(s_{m+1})(1 - E_{s_1}\{Fs(s_1)|s_{m+1}\}) \times \\ &\times (1 - Fs(s_{m,m-1})) \prod_{i=1}^{m-1} (1 - E_{s_{i+1}}\{Fs(s_{i+1}, i)|s_{m+1}\}), \end{aligned}$$

что совпадает с (15) для $Hs(s_{m+1})$ и доказывает предположение индукции. $\blacktriangleleft$

Доказанный результат в сущности является более детальной формой выводов С. Райта [12] при анализе иерархически подразделенных популяций.

Ремарка 5. В популяционной генетике известна [12] следующая формула С. Райта для концентрации гетерозигот $H(s)$ в подразделенной метапопуляции s :

$$H(s) = H_0(s)(1 - F_{IS}(s))(1 - F_{ST}(s)),$$

где $F_{IS}(s)$ и $F_{ST}(s)$ — неслучайный и случайный коэффициент инбридинга соответственно. Аналогично у нас отдельные составляющие выражения (15) для $Hs(s_m)$ имеют сходную интерпретацию. Среднюю фамилную инбредность субпопуляций первого уровня иерархии в s_m характеризует $E_{s_1}\{Fs(s_1)|s_m\}$, а

подразделенности метапопуляции отвечает произведение $(1 - Fs(s_{m,m-1})) \prod (...)$, отражающее фамилную дивергенцию субпопуляций в s_m . Компонент $E_{s_1}\{Fs(s_1)|s_m\}$ соответствует неслучайному коэффициенту инбридинга $F_{IS}(s_m)$ в популяционной генетике, а произведение $(1 - Fs(s_{m,m-1})) \prod (...)$ дает $1 - F_{ST}(s_m)$, где F_{ST} — аналог случайного коэффициента инбридинга. В терминах общего $F_{IT}(s_m)$, случайного и неслучайного коэффициентов фамилного инбридинга для выражения общей (полной) вероятности встреч $Hs(s_m)$ в метапопуляции s_m имеем

$$\begin{aligned} Hs(s_m) &= Hs_0(s_m)(1 - F_{IS}(s_m)) \times \\ &\times (1 - F_{ST}(s_m)) = Hs_0(s_m)(1 - F_{IT}(s_m)), \\ F_{IS}(s_m) &\equiv E_{s_1}\{Fs(s_1)|s_m\}, \\ 1 - F_{ST}(s_m) &\equiv (1 - Fs(s_{m,m-1})) \times \\ &\times \prod_{i=1}^{m-2} (1 - E_{s_{i+1}}\{Fs(s_{i+1}, i)|s_m\}). \end{aligned}$$

В анализируемых данных может отсутствовать информация о наличии преград для встреч в субпопуляциях, служащих единицами наблюдения. В связи с этим желательно проанализировать последствия реального существования препятствий встречам в единицах наблюдения, когда принято решение об их отсутствии. Рассмотрим данную ситуацию в предположении независимости коэффициентов фамилного инбридинга и концентраций фамилии.

Следствие 6. Пусть в условиях предыдущего результата в метапопуляции s_m с t уровнями иерархии и нулевым коэффициентом неслучайного фамилного инбридинга в качестве единицы наблюдения служат субпопуляции $\{s_n\}$ уровня n ($1 < n < t$), в которых концентрации интересующей фамилии равны $\{x(s_n)\}$, а вероятности $\{Hs(s_n)\}$ рассматриваемой неизонимной встречи неизвестны. Положим, что в качестве ее значения используется вероятность случайных столкновений индивидумов $Hs_0(s_n)$.

Тогда $Hs(s_m)$ будет меньше в $\prod_{i=1}^{n-1} (1 - E_{s_{i+1}}\{Fs(s_{i+1}, i)|s_m\})$ раз значения $Hs(s_{m,n})$, где $Fs(s_{i+1}, i)$ — неизвестные коэффициенты фамилного инбридинга в субпопуляциях.

Доказательство. Напомним, что согласно только что установленному результату полная вероятность встреч во всей метапопуляции s_m находится как

$$\begin{aligned} Hs(s_m) &= Hs_0(s_m)(1 - Fs(s_{m,m-1})) \times \\ &\times \prod_{i=1}^{m-2} (1 - E_{s_{i+1}}\{Fs(s_{i+1}, i)|s_m\}). \end{aligned}$$

Здесь подразумевается, что единицей наблюдения является субпопуляция s_1 первого уровня иерархии с

нулевым коэффициентом неслучайного фамильного инбридинга, и индекс i произведения начинается с единицы. Когда уровень единицы наблюдения равен n , то начальное значение i равно n . Соответственно при факторизации $Hs(s_{m,n})$ в отличие от $Hs(s_m)$ вместо $\prod_{i=1}^{m-1}$ фигурирует $\prod_{i=n}^{m-1}$ с другим начальным значением индекса i . Тем самым здесь опущено произведение $\prod_{i=1}^{n-1}$. По условию коэффициенты фамильного инбридинга у субпопуляций уровня иерархии n (у единиц наблюдения) приняты равными нулю, что возможно только когда такие коэффициенты равны нулю в каждой из субпопуляций с уровнями иерархии до n . В этом случае соответствующие им множители равны единице, и их можно вычеркнуть в формуле факторизации $Hs(s_m)$. Вычеркнутые множители при ненулевых в действительности коэффициентах фамильного инбридинга дают произведение $\prod_{i=1}^{n-1} (1 - E_{s_{i+1}}\{Fs(s_{i+1,i})|s_m\})$, показывающее во сколько раз изменилось значение $Hs(s_m)$ из-за пренебрежения возможным отклонением $Hs(s_n)$ от $Hs_0(s_n)$.

Таким образом получаем, что игнорирование инбредности субпопуляций первого уровня иерархии и подразделенности субпопуляций более высоких уровней до уровня единицы наблюдения приводит к завышению Hs согласно доказываемому выражению. ◀

Как известно, коэффициент инбридинга в популяциях человека обычно порядка сотых долей или тысячных. Если коэффициенты фамильного инбридинга также малы, то можно пренебречь их произведениями. В результате получим простую аппроксимацию для коэффициентов фамильного инбридинга в таких метапопуляциях.

Результат 7. В иерархически подразделенной метапопуляции s при малых коэффициентах фамильного инбридинга в составляющих ее субпопуляциях общая вероятность (15) случайной встречи $Hs(s_m|Fs)$ индивидуумов с данной фамилией и какой-нибудь другой с учетом их порядка в наугад выбранной из s_m единице наблюдения s_1 аппроксимируется следующим выражением:

$$\begin{aligned} Hs(s_m|Fs_1) &\sim x(s_m)(1 - x(s_m)) \times \\ &\times \left(1 - E_{s_1}\{Fs(s_1)|s_m\} - Fs(s_{m,m-1}) - \right. \\ &\quad \left. - \sum_{i=1}^{m-2} E_{s_{i+1}}\{Fs(s_{i+1,i})|s_m\} \right) \equiv \\ &\equiv x(s_m)(1 - x(s_m))(1 - Fs(s_m)), \\ Fs(s_m) &\equiv E_{s_1}\{Fs(s_1)|s_m\} + Fs(s_{m,m-1}) + \\ &\quad + \sum_{i=1}^{m-2} E_{s_{i+1}}\{Fs(s_{i+1,i})|s_m\} \geq 0, \end{aligned}$$

где $Fs(s_m)$ — коэффициент фамильного инбридинга субпопуляции уровня m , неотрицательные коэффициенты $\{Fs(s_{i+1,i})\}$ характеризуют подразделенность субпопуляций, а $E_{s_1}\{Fs(s_1)|s_m\}$ — среднюю инбредность субпопуляций, единиц наблюдения.

Таким образом, приближенно коэффициент инбридинга иерархически подразделенной метапопуляции равен сумме средних коэффициентов инбридинга субпопуляций на отдельных уровнях иерархии.

Доказательство. По определению коэффициента фамильного инбридинга $Hs(s_m|Fs_1) = Hs_0(s_m)(1 - Fs(s_m))$. С другой стороны, согласно (15)

$$\begin{aligned} Hs(s_m|Fs_1) &= Hs_0(s_m)(1 - E_{s_1}\{Fs(s_1)|s_m\}) \times \\ &\times (1 - Fs(s_{m,m-1})) \prod_{i=1}^{m-2} (1 - E_{s_{i+1}}\{Fs(s_{i+1,i})|s_m\}). \end{aligned}$$

Эти два выражения для $Hs(s_m)$ должны быть равны друг другу, т.е.

$$\begin{aligned} 1 - Fs(s_m) &= (1 - E_{s_1}\{Fs(s_1)|s_m\})(1 - Fs(s_{m,m-1})) \times \\ &\times \prod_{i=1}^{m-2} (1 - E_{s_{i+1}}\{Fs(s_{i+1,i})|s_m\}). \end{aligned}$$

При раскрытии скобок в правой части данного равенства в случае малых по условию коэффициентов фамильного инбридинга субпопуляций можно пренебречь их произведениями. Очевидно, в результате этого остающиеся члены дают требуемую аппроксимацию для $Fs(s_m)$, которая лаконично выражается с помощью единого суммарного коэффициента инбридинга. ◀

Из доказанного также ясно, что при $Fs(s_m) > 0$ вероятность Hs меньше значения, соответствующего случайным встречам. Отклонение увеличивается с каждым повышением уровня иерархии на величину разности единицы и Fs на соответствующем уровне иерархии.

ЗАКЛЮЧЕНИЕ

Проведенный анализ фамильной структуры мотивирован имеющимися параллелями в передаче фамилии и генов от родителя детям. На популяционном уровне передача фамилий и генов описывается одним и тем же процессом выборочного дрейфа, но отличающимся интенсивностью [8, 9]. На этом и на других аргументах основаны сотни статей в базе данных <https://pubmed.ncbi.nlm.nih.gov> по изучению фамильной структуры и ее связи с генетической. Указанные источники и приведенные ниже в списке литературы для первоначального ознакомления позволяют получить достаточно подробное представление о важности фамильных данных в популяционной генетике.

Существенной чертой обитания современного и первобытного человека является подразделен-

ность популяций, поддерживающаяся в течение длительного времени. При этом иерархическая подразделенность в силу территориального деления, происхождения, лингвистической классификации и т.д. является типичной. Более всего иерархическая структура характерна для семейных данных, часто представленных в соответствии с иерархическим административно-территориальным делением.

Важным показателем неоднородности иерархической системы является дивергенция составляющих ее субпопуляций. Настоящая серия статей автора [10, 11] посвящена анализу показателей дивергенции, присущих самой иерархической структуре и вычислительно независимых от характера микроэволюционного процесса, изолированности субпопуляций или существования миграций между ними.

Процесс выборочного дрейфа приводит к генетической и семейной дивергенции первоначально сходных популяций. Поэтому основное внимание мы уделяем показателям дивергенции субпопуляций в подразделенной метапопуляции. Таким показателем является нарушение случайного характера встреч (столкновений) индивидумов в метапопуляции и ее подразделениях. Формально вероятности встреч аналогичны концентрациям гетерозигот, поскольку в обоих случаях речь идет о вероятностях комбинаций пар (семей) при встрече и генов в диплоидных генотипах).

В данной статье рассматривается семейный аналог коэффициента инбридинга, важной характеристики в популяционной генетике. В изложении результатов сделан акцент на одной из сторон коэффициента семейного инбридинга, а именно на нем как характеристике нарушения случайности встреч индивидумов в метапопуляции. В контексте популяционной генетики это свойство аналогично отклонению концентрации гетерозигот в метапопуляции от значения, соответствующего закону Харди–Вайнберга. В указанном отношении получено специфическое для иерархической подразделенности обобщение эффекта Валунда в виде разложения коэффициента семейного инбридинга по компонентам, соответствующим отдельным уровням иерархии.

Конкретной целью всей серии работ автора по теоретическому изучению семейной структуры является адаптация и обобщение традиционных популяционно-генетических подходов к анализу и описанию свойств семейной структуры в иерархически подразделенной метапопуляции. Полученные результаты применимы и в соответствующей области популяционной генетики при минимальных коррекциях, и в Приложении, которое можно читать независимо, сформулирова-

ны популяционно-генетические аналоги найденных обобщений.

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.

ПРИЛОЖЕНИЕ

Как упоминалось ранее, сформулированные выше результаты для распределения концентрации фамилии дают представление о статистических свойствах числовых характеристик иерархически подразделенной на группы совокупности объектов любой природы и не зависят от закономерностей формирования групп. Например, они приложимы к биологическим популяциям, характеризующимся количественными признаками и иерархически подразделенным согласно территориальной принадлежности, происхождению, родственным связям и т.д. Близкой к изучаемой картине должна быть ситуация с концентрациями x аллеля аутосомного локуса в иерархически подразделенной метапопуляции, которые подобны данным по концентрации фамилии при анализе структуры изонимии у человека.

Настоящее приложение содержит приведенные без доказательств аналоги полученных в данной серии работ результатов по анализу семейной структуры применительно к материалам популяционной генетики. Объектом изучения в популяционной генетике является генетическая структура популяции (концентрации аллелей и генотипов). Типичные популяционно-генетические материалы охватывают ряд субпопуляций, образующих метапопуляцию. Ее генетическая структура определяется структурами субпопуляций, и важно исследовать связь между ними в типичных ситуациях.

Такой характерной для материалов по человеку ситуацией является иерархическая подразделенность метапопуляции. Формулируемые ниже свойства иерархически подразделенной метапопуляции в зависимости от генетической структуры субпопуляций относятся к ее текущему генетическому состоянию и вычислительно не зависят от существования миграций между субпопуляциями или их полной изолированности, они не определяются происхождением субпопуляций от общей родоначальной группы или их составом (вплоть до произвольной механической смеси) и моделями микроэволюции, которые влияют на результаты количественно. Наше внимание фокусируется на характеристиках дивергенции субпопуляций, которые определяют такую важную черту метапопуляции, как ее коэффициент инбридинга.

Читателям, чей интерес лежит только на генетических аспектах иерархической структуры ме-

тапопуляций, можно посоветовать сразу обратиться к данному приложению и просматривать соответствующие (близкие) доказательства в предшествующем тексте при желании глубже вникнуть в рассматриваемые проблемы. Чтобы приложение можно было читать независимо от предыдущего изложения, напомним используемые обозначения и приведем их модификации применительно к популяционно-генетическому контексту приложения.

Уровень иерархии m всей метапопуляции s_m считаем самым высоким, а субпопуляции первого уровня $\{s_i\}$ рассматриваем как неподразделенные с выполнением закона Харди–Вайнберга или с некоторым коэффициентом инбридинга, вызванным какими-либо причинами. Обозначения популяционно-генетических характеристик (дисперсии V , концентрации гетерозигот H , коэффициента инбридинга F) отличаются от использованных выше V_s , H_s , F_s для их фамильных аналогов отсутствием окончания s . Векторы набраны полужирным шрифтом, Символ s_i относится к субпопуляции на уровне (ранге) иерархии i , $\{x(s_i)\}$ дает концентрации носителей рассматриваемой характеристики, например, концентрации конкретного аллеля, гетерозиготы и т.д. в субпопуляции s_i (далее для краткости просто концентрации), $x(s_i|s_m)$, обозначает случайное значение рассматриваемой концентрации x в субпопуляции s_i , получаемое при выборе наугад s_i из содержащей ее (наряду с другими) субпопуляции s_m более высокого уровня m ($i < m$). Иначе говоря, $\{x(s_i|s_m)\}$ случайная величина, принимающая значения $\{x(s_i)\}$, равные концентрациям рассматриваемого далее какого-нибудь аллеля в субпопуляциях уровня i , входящих в s_m (при условии выбора наугад s_i из s_m). Ее средняя величина и дисперсия обозначаются как $x(s_m)$ и $V(x(s_i|s_m))$ соответственно.

Рассмотрим свойства распределения концентраций в подразделенной метапопуляции.

Замечание А1. Пусть анализируется подразделенная популяция s_m с размером $N(s_m)$ и с численностью $T_N(s_m)$ носителей рассматриваемой характеристики T (скажем, интересующего аллеля, гетерозиготы и т.п.) в s_m . Тогда концентрация $T(s_m) \equiv T_N(s_m)/N(s_m)$ данных носителей в популяции s_m , произвольно разбитой на группы $\{s_i\}$ с численностями $\{N(s_i)\}$, находится как

$$T(s_m) = \sum_{s_i \in s_j} T(s_i) \frac{N(s_i)}{N(s_m)} = \sum_{s_i \in s_j} T(s_i) Pr(s_i|s_m) = \\ = E_{s_i} \{T(s_i)|s_m\}, \quad Pr(s_i|s_m) \equiv \frac{N(s_i)}{N(s_m)},$$

т.е. концентрация $T(s_m)$ во всей популяции s_m равна среднему значению (математическому ожиданию) случайной величины $T(s_i|s_m)$, концентрации T в s_i

при выборе наугад s_i из s_m с вероятностью $Pr(s_i|s_m) \equiv N(s_i)/N(s_m)$. Здесь E является символом операции получения математического ожидания, нижний индекс у E указывает на переменную, служащую для усреднения.

Значение концентрации $T(s_m)$ не зависит от разбиения s_m на $\{s_i\}$, и $T(s_m)$ не меняется при изменении разбиения.

Типичным признаком в популяционной генетике служит концентрация x какого-либо аллеля. Как и для характеристики T , среднее значение $x(s_m)$ для концентрации аллеля во всей популяции s_m равно средней величине для его концентраций в составляющих s_m субпопуляциях. Скажем, если s_m разбита на $\{s_i\}$, то $x(s_m) = E_{s_i} \{x(s_i)|s_m\}$. То же самое можно сказать о разбиении s_m на $\{s_j\}$. Математическое ожидание случайной величины $x(s_i|s_m)$, обозначаемое как $x(s_m)$, равняется концентрации x в s_m : $x(s_m) = E_{s_i} \{x(s_i)|s_m\}$.

Данное свойство является частным случаем правила полного математического ожидания, согласно которому у нас концентрация $x(s)$ соответствующих носителей (аллеля, гетерозиготы и т.д.) в разбитой на субпопуляции $\{s_i\}$ метапопуляции s является средней величиной для концентраций в отдельных субпопуляциях (концентраций при условиях рассмотрения субпопуляций по одной), каково бы ни было данное разбиение. Для случайной величины x формула полного математического ожидания означает, что

$$x(s) \equiv E_{s_i} \{x(s_i)|s\} = E_{s_i} \{x(s_i)|s_i\}, \\ i = 1, 2, 3, \dots, \quad x(s_i) \equiv E_{s_i} \{x(s_i)|s_i\},$$

т.е. среднее значение концентрации x в некоторой иерархически подразделенной метапопуляции s является средней величиной для концентраций в разбивающих s подгруппах $\{s_i\}$. Приведенная формула показывает в явном виде равенство концентрации $x(s)$ рассматриваемого аллеля во всей метапопуляции s средней величине для его концентраций в отдельных частях $\{x(s_i)\}$ разбиения s одного и того же уровня иерархии i , какими бы ни были эти части и уровень.

Сформулированные результаты справедливы не только для концентрации аллеля, но и для концентрации H гетерозигот. Концентрация $H(s_i)$ в субпопуляции s_i является средним взвешенным значением $E_{s_i} \{H(s_i)|s_i\}$ концентраций гетерозигот в отдельных субпопуляциях-единицах наблюдения $\{s_1\}$, содержащихся в s_i , причем веса можно понимать как вероятности случайного выбора соответствующих субпопуляций $\{s_1\}$ из s_i . Другими словами, $H(s_i)$ находится как ожидаемое значение $H(s_1)$ при выборе наугад s_1 из s_i (т.е. средней вели-

чине для результатов, получаемых при выборе значений $H(s_1)$ с учетом вероятностей выбора s_1 .

Результат А2 (эффект Валунда). Пусть дана популяция s , разбитая на субпопуляции $\{s_1\}$, где концентрации рассматриваемого аллеля равны $\{x(s_1)\}$, а концентрации гетерозигот с этим аллелем с учетом порядка аллелей находятся как $\{x(s_1)(1 - x(s_1))\}$ по закону Харди–Вайнберга.

Тогда общая (полная) концентрация гетерозигот $H(s)$ с данным аллелем в наугад выбранной из s субпопуляции s_1 имеет вид

$$\begin{aligned} H(s) &\equiv E_{s_1} \{H(s_1)|s\} = \\ &= \sum_{s_1} x(s_1)(1 - x(s_1))Pr(s_1|s) = \\ &= x(s)(1 - x(s)) - V_{\text{betw}}(x(s_1|s)), \\ V_{\text{betw}}(x(s_1|s)) &= V(x(s_1|s)) = \\ &= \sum_{s_1} (x(s_1) - x(s))^2 Pr(s_1|s). \end{aligned}$$

Здесь $V_{\text{betw}}(x(s_1|s))$ обозначает межгрупповую дисперсию значений концентрации $\{x(s_1|s)\}$ интересующего аллеля в субпопуляциях (группах) $\{s_1\}$, составляющих популяцию s ; $Pr(s_1|s)$ вероятность выбора наугад субпопуляции s_1 из популяции s .

Иным образом общую (полную) концентрацию $H(s)$ можно представить как

$$\begin{aligned} H(s) &= x(s)(1 - x(s)) \times \\ &\times \left(1 - \frac{V_{\text{betw}}(x(s_1|s))}{x(s)(1 - x(s))}\right) \equiv H_0(s)(1 - F(s)), \\ F(s) &\equiv \frac{V_{\text{betw}}(x(s_1|s))}{x(s)(1 - x(s))} = \frac{V_{\text{betw}}(x(s_1|s))}{H_0(s)} \geq 0, \\ H_0(s) &\equiv x(s)(1 - x(s)), \end{aligned}$$

где $F(s)$ является коэффициентом инбридинга С. Райта, $H_0(s)$ обозначает концентрацию рассматриваемых гетерозигот в метапопуляции s при выполнении в ней закона Харди–Вайнберга.

Дисперсия значений $\{x(s_1|s)\}$ характеризует дивергенцию концентраций аллеля между субпопуляциями $\{s_1\}$. Чем она больше, тем больше дивергенция концентрации аллеля между субпопуляциями и тем меньше общая концентрация гетерозигот в подразделенной популяции. Коэффициент инбридинга также характеризует генетическую дивергенцию субпопуляций, и далее мы интерпретируем его как показатель отклонения концентрации гетерозигот от значения, соответствующего закону Харди–Вайнберга.

Обратим внимание, что результат Валунда обобщается на случай, когда в подразделенной популяции нарушаются соотношения Харди–Вайнберга в составляющих ее субпопуляциях $\{s_1\}$, где $H(s_1) \equiv x(s_1)(1 - x(s_1))(1 - F(s_1))$ с учетом порядка аллелей в гетерозиготах.

Следствие А3. Пусть дана популяция s_2 , разбитая на субпопуляции $\{s_1\}$ с концентрациями $\{x(s_1)\}$ рассматриваемого аллеля. Предположим, что в s_1 концентрацию $H(s_1)$ гетерозигот с данным аллелем с учетом порядка аллелей возможно описать с помощью коэффициента инбридинга $F(s_1)$ как

$$\begin{aligned} H(s_1) &= x(s_1)(1 - x(s_1))(1 - F(s_1)) = \\ &= H_0(s_1)(1 - F(s_1)), \quad s_2 = \{s_1\}, \end{aligned}$$

где $H_0(s_1) = x(s_1)(1 - x(s_1))$ обозначает концентрацию гетерозигот при выполнении закона Харди–Вайнберга в субпопуляции s_1 .

Тогда общая (полная) концентрация рассматриваемых гетерозигот $H(s_2|F_{s_1})$ в наугад выбранной из s_2 субпопуляции s_1 при независимости концентрации x рассматриваемого аллеля и коэффициента инбридинга $F(s_1)$ в субпопуляциях $\{s_1\}$ имеет вид

$$\begin{aligned} Hs(s_2|F_{s_1}) &= \\ &= (x(s_2)(1 - x(s_2)) - V_{s_{\text{betw}}}(x(s_1|s_2))) \times \\ &\times (1 - E_{s_1} \{F(s_1|s_2)\}). \end{aligned}$$

Здесь $V_{s_{\text{betw}}}(x(s_1|s_2))$ обозначает межгрупповую дисперсию концентраций $\{x(s_1|s_2)\}$ рассматриваемого аллеля в субпопуляциях (группах) $\{s_1\}$, составляющих популяцию s_2 .

Иным образом общая (полная) концентрация $H(s_2|F_{s_1})$ представима как

$$\begin{aligned} H(s_2|F_{s_1}) &= H_0(s_2) \left(1 - \frac{V_{s_{\text{betw}}}(x(s_1|s_2))}{x(s_2)(1 - x(s_2))}\right) \times \\ &\times (1 - E_{s_1} \{F(s_1|s_2)\}) = Hs_0(s_2) \times \\ &\times (1 - F(x(s_2)))(1 - E_{s_1} \{F(s_1|s_2)\}), \end{aligned}$$

где $F(x(s_2)) \equiv V_{s_{\text{betw}}}(x(s_1|s_2))/Hs_0(s_2) \geq 0$ является случайным коэффициентом инбридинга F_{ST} , а средний коэффициент инбридинга $E_{s_1} \{F(s_1|s_2)\}$ для субпопуляций $\{s_1\}$ в s_2 является коэффициентом неслучайного инбридинга F_{IS} согласно результатам С. Райта.

Мы видим, что дисперсия концентраций аллеля в субпопуляциях метапопуляции играет важную роль для величины коэффициента инбридинга. Поэтому представляет интерес выяснение структуры дисперсии концентрации аллеля в иерархически подразделенной метапопуляции. При этом для наших целей анализа структуры дисперсии существенную роль играет правило сложения дисперсии, согласно которому в совокупности из нескольких групп произвольных объектов с каким-либо числовым признаком x общая (полная) дисперсия x во всей совокупности равна сумме межгрупповой дисперсии (дисперсии распределения средних значений x в группах) и средней внутригрупповой дисперсии (средней дисперсии x внутри групп).

Понятно, что в соответствии с приведенной формулировкой данное статистическое правило

не зависит от природы объектов и принципов их объединения в группы, которые могут быть произвольными. Правило выполняется не только для дисперсий, но и при совместном изучении нескольких признаков для их матриц ковариаций. В случае популяционно-генетического анализа подразделенной популяции данное правило принимает следующий вид.

Замечание А4 (правило сложения дисперсий). Пусть метапопуляция s с тремя уровнями иерархии разбита на субпопуляции $\{s_2\}$, каждая из которых включает неподразделенные группы $\{s_1\}$, являющиеся единицами наблюдения с концентрациями $\{x(s_1)\}$ рассматриваемого аллеля в них.

Тогда полная дисперсия $V_{\text{tot}}(x(s_1|s))$ распределения концентрации аллеля по несгруппированным единицам наблюдения $\{s_1\}$ во всей метапопуляции s равна сумме межгрупповой дисперсии $V_{\text{betw}}(x(s_2|s))$, характеризующей дивергенцию концентраций аллеля $\{x(s_2)\}$ между субпопуляциями $\{s_2\}$, и средней внутригрупповой дисперсии $W(x(s_1|s_2)|s) \equiv E_{s_2}\{V_{\text{in}}(x(s_1|s_2))|s\}$, характеризующей среднюю дивергенцию концентраций $x(s_1)$ субпопуляций $\{s_1\}$ внутри $\{s_2\}$:

$$V_{\text{tot}}(x(s_1|s)) = V_{\text{betw}}(x(s_2|s)) + E_{s_2}\{V_{\text{in}}(x(s_1|s_2))|s\}.$$

Результат анализа свойств дисперсии для метапопуляции с тремя уровнями иерархии обобщается на ситуацию произвольного количества уровней.

Следствие А5. В иерархически подразделенной метапопуляции s_m с t уровнями иерархии полная дисперсия $V(x(s_1|s_m))$ распределения концентрации аллеля по субпопуляциям первого уровня допускает разложение по компонентам на уровнях иерархии в виде

$$\begin{aligned} V(x(s_1|s_m)) &= \sum_{i=1}^{m-1} E_{s_{i+1}}\{V(x(s_i|s_{i+1}))|s_m\} = \\ &= \sum_{i=1}^{m-1} (V(x(s_i|s_m)) - V(x(s_{i+1}|s_m))). \end{aligned}$$

Использование данного разложения дисперсии позволяет найти разложение ожидаемой концентрации гетерозигот по компонентам, соответствующим отдельным уровням иерархии.

Результат А6. Рассмотрим иерархически подразделенную метапопуляцию s_m с уровнями иерархии $i = 1, 2, \dots, t$. Положим, что в субпопуляциях $\{s_1\}$ концентрации рассматриваемого аллеля равны $\{x(s_1)\}$, а концентрации несущих его гетерозигот удовлетворяют закону Харди–Вайнберга и с учетом порядка аллелей находятся как $H(s_1) = H_0(s_1) \equiv x(s_1)(1 - x(s_1))$.

Если из s_m наугад выбрана субпопуляция s_1 первого уровня иерархии, то ожидаемая концентрация данных гетерозигот $H(s_m)$ в s_1 (общая, или полная концентрация) допускает следующее разложение:

$$\begin{aligned} H(s_m) &\equiv E_{s_1}\{H(s_1)|s_m\} = x(s_m)(1 - x(s_m)) - \\ &- V_{\text{betw}}(x(s_1|s_m)) = H_0(s_m) - \\ &- \sum_{i=1}^{m-1} E_{s_{i+1}}\{V_{\text{in}}(x(s_i|s_{i+1}))|s_m\}. \end{aligned}$$

Таким образом, ожидаемая концентрация гетерозигот $H(s_m)$ равна разности между значением $H_0(s_m) \equiv x(s_m)(1 - x(s_m))$, соответствующим выполнению закона Харди–Вайнберга во всей метапопуляции s_m , и разложением полной дисперсии $V(x(s_1|s_m))$ на средние внутригрупповые дисперсии $E_{s_{i+1}}\{V_{\text{in}}(x(s_i|s_{i+1}))|s_m\} = V(x(s_i|s_m)) - V(x(s_{i+1}|s_m))$, соответствующие отдельным уровням иерархии i .

На разложении дисперсии и концентрации гетерозигот основывается и разложение коэффициента инбридинга в соответствии с уровнями иерархии.

Результат А7. Пусть в иерархически подразделенной метапопуляции s_m с t уровнями иерархии в качестве единиц наблюдения служат субпопуляции $\{s_1\}$ первого уровня. Положим, что концентрации интересующего аллеля в $\{s_1\}$ равны $\{x(s_1)\}$, а концентрации его несущих гетерозигот с учетом порядка аллелей находятся в $\{s_1\}$ как $\{H(s_1) = H_0(s_1) \equiv x(s_1)(1 - x(s_1))\}$.

Тогда разложение коэффициента инбридинга $F(s_m)$ в наугад выбранной из s_m субпопуляции s_1 можно записать как

$$\begin{aligned} F(s_m) &= \sum_{i=1}^{m-1} \frac{V(x(s_i|s_m)) - V(x(s_{i+1}|s_m))}{H_0(s_m)} = \\ &= \sum_{i=1}^{m-1} \frac{E_{s_{i+1}}\{V_{\text{in}}(x(s_i|s_{i+1}))|s_m\}}{H_0(s_m)} = \sum_{i=1}^{m-1} \Delta_i F(s_m), \\ \Delta_i F(s_m) &\equiv \frac{V(x(s_i|s_m)) - V(x(s_{i+1}|s_m))}{H_0(s_m)} = \\ &= \frac{E_{s_{i+1}}\{V_{\text{in}}(x(s_i|s_{i+1}))|s_m\}}{H_0(s_m)} \equiv F(s_{m,i}) - F(s_{m,i+1}) \geq 0. \end{aligned}$$

Общая концентрация $H(s_m)$ рассматриваемых гетерозигот в наугад выбранной из s_m субпопуляции s_1 представима в виде

$$\begin{aligned} H(s_m) &= H_0(s_m) \times \\ &\times \left(1 - \sum_{i=1}^{m-1} \frac{V(x(s_i|s_m)) - V(x(s_{i+1}|s_m))}{H_0(s_m)} \right) = \\ &= H_0(s_m) \left(1 - \sum_{i=1}^{m-1} \frac{E_{s_{i+1}}\{V_{\text{in}}(x(s_i|s_{i+1}))|s_m\}}{H_0(s_m)} \right) = \\ &= H_0(s_m) \left(1 - \sum_{i=1}^{m-1} \Delta_i F(s_m) \right). \end{aligned}$$

Здесь

$$\begin{aligned}\Delta_i F(s_m) &\equiv F(s_{m,i}) - F(s_{m,i+1}) = \\ &= E_{s_{i+1}} \{V_{in}(x(s_i|s_{i+1}))|s_m\} / H_0(s_m), \\ F(s_m) &\equiv V(x(s_1|s_m)) / H_0(s_m), \\ F(s_{m,i}) &\equiv V_{in}(x(s_i|s_m)) / H_0(s_m), \\ H(s_m) &\equiv E_{s_1} \{H(s_1)|s_m\}, \\ H_0(s_m) &= x(s_m)(1 - x(s_m)),\end{aligned}$$

$F(s_m)$ и $F(s_{m,i})$ обозначают коэффициенты инбридинга в метапопуляции s_m и в $s_{m,i}$, когда в качестве единиц наблюдения в s_m служат субпопуляции $\{s_1\}$ и $\{s_i\}$ с уровнями иерархии 1 и i при условии, что концентрации рассматриваемых гетерозигот в них равны $\{H_0(s_1)\}$ и $\{H_0(s_i)\}$ соответственно; $H(s_m)$ общая концентрация рассматриваемых гетерозигот в s_m .

Обратим внимание на то, что когда в качестве единиц наблюдения используются субпопуляции $\{s_n\}$ с уровнем иерархии n , то в общем случае выбор значений $x(s_n)(1 - x(s_n))$ для концентрации гетерозигот в $\{s_n\}$ по закону Харди–Вайнберга не обоснован. Рассмотрим погрешности выражений для концентрации гетерозигот и коэффициента инбридинга, вносимые использованием вместо начальных данных (концентрации гетерозигот с интересующим аллелем в единицах наблюдения с нулевым коэффициентом инбридинга) агрегированных сведений для субпопуляций более высокого ранга.

Следствие А8. Пусть в иерархически подразделенной метапопуляции s_m с t уровнями иерархии в качестве единиц наблюдения вместо субпопуляций первого уровня $\{s_1\}$ с концентрациями интересующего аллеля $\{x(s_1)\}$ и случайным скрещиванием рассматриваются субпопуляции уровня n ($1 < n < t - 1$) с концентрациями $\{x(s_n)\}$. Предположим, что при отсутствии или игнорировании сведений о нарушении в них закона Харди–Вайнберга принимается решение об отсутствии нарушений в каждой из субпопуляций $\{s_n\}$, т.е. концентрация $H(s_{m,n})$ в s_m несущих интересующий аллель гетерозигот с учетом порядка аллелей (в наугад выбранной из s_m субпопуляции s_n) находится как

$$\begin{aligned}H(s_{m,n}) &\equiv E_{s_n} \{H(s_n)|s_m\} = E_{s_n} \{H_0(s_n)|s_m\}, \\ H_0(s_n)(s_m) &\equiv x(s_n)(1 - x(s_n)),\end{aligned}$$

где $s_{m,n}$ обозначает метапопуляцию s_m с единицами наблюдения $\{s_n\}$.

Тогда коэффициент инбридинга $F(s_{m,n})$, определяемый из условия $H(s_{m,n}) = H_0(s_m)(1 - F(s_{m,n}))$ можно разложить по компонентам на уровнях иерархии как

$$\begin{aligned}F(s_{m,n}) &= \sum_{i=n}^{m-1} \frac{V(x(s_i|s_m)) - V(x(s_{i+1}|s_m))}{x(s_m)(1 - x(s_m))} = \\ &= \sum_{i=n}^{m-1} \frac{E_{s_{i+1}} \{V(x(s_i|s_{i+1}))|s_m\}}{x(s_m)(1 - x(s_m))} \equiv \sum_{i=n}^{m-1} \Delta_i F(s_m), \\ \Delta_i F(s_m) &\equiv \frac{V(x(s_i|s_m)) - V(x(s_{i+1}|s_m))}{x(s_m)(1 - x(s_m))} = \\ &= \frac{E_{s_{i+1}} \{V(x(s_i|s_{i+1}))|s_m\}}{x(s_m)(1 - x(s_m))} \geq 0.\end{aligned}$$

Коэффициент $F(s_{m,n})$ меньше значения $F(s_m)$, соответствующего единице наблюдения $\{s_1\}$, на сумму вкладов неучтенных уровней иерархии от 1 до $n - 1$:

$$\begin{aligned}F(s_m) - F(s_{m,n}) &= \sum_{i=1}^{m-1} \frac{V(x(s_i|s_m)) - V(x(s_{i+1}|s_m))}{x(s_m)(1 - x(s_m))} - \\ &- \sum_{i=n}^{m-1} \frac{V(x(s_i|s_m)) - V(x(s_{i+1}|s_m))}{x(s_m)(1 - x(s_m))} = \\ &= \sum_{i=1}^{n-1} \frac{E_{s_{i+1}} \{V(x(s_i|s_{i+1}))|s_m\}}{x(s_m)(1 - x(s_m))} = \sum_{i=1}^{n-1} \Delta_i F(s_m) \geq 0.\end{aligned}$$

Описанные результаты базируются на разложении дисперсии концентрации интересующего аллеля в иерархически подразделенной метапопуляции.

Структуру показателей инбридинга такой метапопуляции можно также исследовать и с помощью подхода С. Райта, основанного на *дополнительном предположении о независимости коэффициента инбридинга и концентрации аллеля* в субпопуляциях. Рассмотрим данное предположение в условиях, допускающих нарушение закона Харди–Вайнберга в субпопуляциях уровня единицы наблюдения.

Результат А9. Пусть рассматривается метапопуляция s_m с иерархической структурой подразделения на субпопуляции. Положим, что концентрации интересующего аллеля в субпопуляциях $\{s_i\}$ самого низкого уровня иерархии (единицы наблюдения) внутри s_m равны $\{x(s_i|s_m)\}$, концентрация $H(s_1)$ гетерозигот с этим аллелем в s_1 (с учетом порядка) находится как $x(s_1)(1 - x(s_1))(1 - F(s_1)) = H_0(s_1)(1 - F(s_1))$.

Тогда общая (полная) концентрация $H(s_m)$ гетерозигот в наугад выбранной из s_m субпопуляции s_1 при независимости в совокупности концентрации аллелей и коэффициентов инбридинга F факторизуется как

$$\begin{aligned}H(s_m) &= H_0(s_m)(1 - E_{s_1} \{F(s_1)|s_m\})(1 - F(s_{m,m-1})) \times \\ &\times \prod_{i=1}^{m-2} (1 - E_{s_{i+1}} \{F(s_{i+1,i})|s_m\}), \\ F(s_{i+1,i}) &\equiv V(x(s_{i+1}|s_i)) / H_0(s_{i+1,i}), \\ H_0(s_i) &\equiv x(s_i)(1 - x(s_i)).\end{aligned}$$

Отдельные составляющие данного выражения имеют следующую интерпретацию. Инбредность

субпопуляций первого уровня иерархии в s_m характеризует $E_{s_1}\{F(s_1)|s_m\}$, а подразделенность метапопуляции отражает произведение $(1 - F(s_{m,m-1})) \prod (...)$, характеризующее дивергенцию субпопуляций в s_m . Компонент $E_{s_1}\{F(s_1)|s_m\}$ соответствует неслучайному коэффициенту инбридинга $F_{IS}(s_m)$, а $F_{ST}(s_m) \equiv (1 - F(s_{m,m-1}))(1 - \prod_{i=1}^{m-2} (1 - E_{s_{i+1}}\{F(s_{i+1,i})|s_m\}))$ случайному коэффициенту инбридинга F_{ST} .

В реальных данных может быть не представлена информация о нарушениях закона Харди–Вайнберга в субпопуляциях, служащих единицами наблюдения. В связи с этим желательно проанализировать последствия существования нарушений в единицах наблюдения (например, из-за скрытой подразделенности, либо недоступности наблюдению, либо простого игнорирования имеющихся сведений), когда принято решение об их отсутствии. Эта проблема рассматривалась выше при разложении концентрации гетерозигот и коэффициента инбридинга по уровням иерархии. Теперь обратимся к ней при факторизации данных показателей в предположении независимости концентраций аллеля и коэффициентов инбридинга.

Следствие A10. Пусть в условиях предыдущего результата в метапопуляции s_m с t уровнями иерархии в качестве единицы наблюдения служат субпопуляции $\{s_n\}$ уровня n ($1 < n < t$), в которых концентрации интересующего аллеля равны $\{x(s_n)\}$, а концентрации несущих его гетерозигот неизвестны. Пусть в качестве значения неизвестных концентраций гетерозигот используется $H_0(s_n)$.

Тогда $H(s_m)$ будет меньше значения $H(s_{m,n})$ в $\prod_{i=1}^{n-1} (1 - E_{s_{i+1}}\{F(s_{i+1,i})|s_m\})$ раз, где $F(s_{i+1,i})$ неизвестные коэффициенты инбридинга в субпопуляциях.

Поскольку обычно коэффициенты инбридинга в популяциях человека порядка тысячных, максимум сотых долей, то разумно аппроксимировать полученные формулы более простыми выражениями, пренебрегая произведениями коэффициентов.

Результат A11. В иерархически подразделенной метапопуляции s при малых коэффициентах инбридинга в ее субпопуляциях общая концентрация $H(s_m)$ гетерозигот с заданным аллелем в наугад выбранной из s_m единицы наблюдения s_1 с коэффициентом инбридинга $F(s_1)$ аппроксимируется следующим выражением:

$$\begin{aligned} H(s_1|s_m) &\sim x(s_m)(1 - x(s_m))(1 - E_{s_1}\{F(s_1)|s_m\} - \\ &- F(s_{m,m-1}) - \sum_{i=1}^{m-2} E_{s_{i+1}}\{F(s_{i+1,i})|s_m\}) \equiv \\ &\equiv x(s_m)(1 - x(s_m))(1 - F(s_m)) \equiv \\ &\equiv H_0(s_1|s_m)(1 - F(s_m)), \\ F(x(s_m)) &\equiv E_{s_1}\{F(s_1)|s_m\} + \\ &+ F(s_{m,m-1}) + \sum_{i=1}^{m-2} E_{s_{i+1}}\{F(s_{i+1,i})|s_m\}, \end{aligned}$$

где $F(s_m)$ — коэффициент инбридинга метапопуляции s_m , неотрицательные коэффициенты $\{F(s_{i+1,i})\}$ характеризуют подразделенность субпопуляций. Таким образом, приближенно коэффициент инбридинга иерархически подразделенной метапопуляции равен сумме средних коэффициентов инбридинга субпопуляций на отдельных уровнях иерархии.

СПИСОК ЛИТЕРАТУРЫ

1. Ревазов А.А., Парадеева Г.М., Русакова Г.И. Пригодность русских фамилий в качестве квазигенетического маркера // Генетика. 1986. Т. 22. № 4. С. 699–703.
2. 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.
3. Crow M.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. Rogers A.R. Doubts about isonymy // Hum. Biology. 1991. V. 63. № 5. P. 663–668.
5. Lasker W.G. Surnames and Genetic Structure. Cambridge: Cambr. Univ. Press, 1985. 148 p.
6. Сорокина И.Н., Чурносоев М.И., Балтуцкая И.В. и др. Антропогенетическое изучение населения центральной России. М.: Изд-во РАМН, 2014. 336 с.
7. Сорокина И.Н., Рудых Н.А., Крикун Е.Н., Сокорев С.Н. Применение фамилий в популяционно-генетических исследованиях (на примере зарубежных популяций) // Науч. ведомости БелГУ. Сер. Медицина. Фармация. 2016. № 19 (240). Вып. 35. С. 5–10.
8. Пасеков В.П. К анализу случайных процессов изонимии. I. Структура изонимии // Генетика. 2021. Т. 57. № 10. С. 1194–1204. <https://doi.org/10.31857/S001667582110009X>
9. Пасеков В.П. К анализу случайных процессов изонимии. II. Динамика дивергенции популяций // Генетика. 2021. Т. 57. № 11. С. 1318–1329. <https://doi.org/10.31857/S0016675821101114>

10. Пасеков В.П. Описание дивергенции субпопуляций в иерархической системе при анализе изонимии. I. Дисперсия как показатель дивергенции // Генетика. 2022. Т. 58. № 6. С. 713–727.
11. Пасеков В.П. Описание дивергенции субпопуляций в иерархической системе при анализе изонимии. II. Вероятности неизонимных встреч // Генетика. 2023. Т. 59. № 11. С. 1326–1340.
12. Wright S. The interpretation of population structure by F statistics with special regard to systems of mating // Evolution. 1965. V. 19. P. 395–420.

Description of the Divergence of Subpopulations in the Hierarchical System under the Analysis of Isonymy. III. Surname Inbreeding Coefficient

V. P. Passekov*

Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, Moscow, 119991 Russia

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

The hierarchical structure of metapopulations characteristic of humans with their subdivision into parts (subpopulations), usually classified on the basis of administrative-territorial division, principles of biological systematics, etc., is considered. Attention is focused on the description of surname divergence of subpopulations. The surname inbreeding coefficient F_s is expressed through the divergence indices, which is similar to the traditional inbreeding coefficient in population genetics. Its expansion by hierarchy levels into the sum of increments of the surname inbreeding coefficient corresponding to separate levels is obtained. The relationship between the found decomposition and the decomposition of the dispersion of the distribution of the surname concentration by subpopulations is demonstrated. With an additional assumption about the independence of the concentration of the surname and F_s , the factorization of F_s by hierarchy levels is found in line with the ideas of S. Wright. It is simplified with a small surname divergence of subpopulations when the F_s coefficient of the entire metapopulation is equal to the sum of the average F_s values at individual levels of the hierarchy. The underestimation of surname inbreeding was obtained, when subpopulations of a higher rank serve as units of observation instead of undivided subpopulations of the first level of the hierarchy. These results are statistical characteristics of the hierarchical structure, and not a feature of a particular population system, and do not follow from one or another model of microevolution. They are computationally independent of the hierarchical system under study, but allow characterizing their heterogeneity quantitatively. The results obtained refer to rural and urban hierarchical metapopulations as separate components of the entire population. In an appendix that can be read independently, genetic analogues of the properties found are given without proof in respect to the genetic structure of a metapopulation.

Keywords: hierarchical structure of populations, metapopulations, surname concentrations in human subpopulations, characteristics of heterogeneity of metapopulations, decomposition of the probability of non-isonimic encounters and the coefficient of surname inbreeding by hierarchy levels.