

УДК 578.087.1:575.1:599.9

А. С. ЗУРАБЯН, Ш. М. КОЧАРЯН, А. В. ПЕТРОСЯН

МАТЕМАТИЧЕСКИЕ КРИТЕРИИ ОПРЕДЕЛЕНИЯ НАСЛЕДСТВЕННОГО ХАРАКТЕРА ПЕРЕДАЧИ ПРИЗНАКА У ЧЕЛОВЕКА И ОПРЕДЕЛЕНИЕ ТИПА НАСЛЕДОВАНИЯ

Получены формулы, позволяющие при их последовательном приложении к собранному материалу определить, наследуется ли данный признак и если наследуется, то по какой схеме в том случае, если известна частота признака в популяции.

В исследованиях по генетике человека и медицинской генетике очень важно установление наследственного характера признака (или заболевания, или врожденного уродства), и в том случае, если наследственный характер передачи наблюдается, то установить схему наследования признака. Для моногенных признаков в настоящее время мы располагаем довольно хорошими количественными или полуколичественными критериями [2], которые, однако, в ряде случаев бывает трудно, а зачастую и невозможно приложить к собранному материалу в силу известных недостатков человека как объекта генетических исследований. Применение популяционных методов, основанных на законе равновесия Гарди, позволяет в значительной степени избежать этих трудностей. В случае двух и более генов, определяющих признак, мы не располагаем ничем более, кроме качественных методов, позволяющих в лучшем случае определить полигенный характер наследования.

Поэтому в настоящей работе нами приводятся выражения, выведенные на основании закона равновесия Гарди и теоремы, доказанной нами в предыдущей работе [1], позволяющие по частоте признака в популяции и у родственников I степени выявить, наследуется ли этот признак моногенно или полигенно рецессивно, и в случае полигенного типа наследования установить число генов, вовлеченных в определение признака.

В предыдущей работе [1] нами была доказана необходимость и достаточность условия независимости комбинирования генов в гаметах (зиготах) относительно h локусов ($h=1,2,\dots,n$) для достижения равновесия в большой панмиктической популяции, на которую наложены следующие ограничения: в ней отсутствуют давления отбора, мутаций и миграции, поколения не перекрываются; кроме того, предполагается, что вероятность для каждого из двух аллелей любого гена попасть в любую случайно выбранную гамету, продуцируемую определенным генотипом, равняется $1/2$, т. е. рассматриваемые гены фактически не сцеплены друг с другом.

Пусть каждая особь такой популяции обладает n локусами, которые обозначим через $\alpha_1, \alpha_2, \dots, \alpha_n$.

Предположим, что в h -том локусе имеется два аллеля: α_{ph} и α_{qh} ($h = 1, 2, \dots, n$).

Тогда модель k -той гаметы можно представить в виде формального произведения

$$\gamma_k = \prod_{h=1}^n \alpha_{ph}^{u_{ph}^k} \alpha_{qh}^{u_{qh}^k} \quad (k = 1, 2, \dots, 2^n),$$

где u_{ph}^k (u_{qh}^k) показывает, сколько раз присутствует аллель α_{ph} (α_{qh}) в k -той гамете.

$$u_{ph}^k = 0, 1; \quad u_{qh}^k = 0, 1; \quad u_{ph}^k + u_{qh}^k = 1.$$

Модель s -того генотипа, Γ_s , будет

$$\Gamma_s = \prod_{h=1}^n \alpha_{ph}^{w_{ph}^s} \alpha_{qh}^{w_{qh}^s}, \quad (s = 1, 2, \dots, 3^n),$$

$$\text{где } w_{ph}^s = 0, 1, 2; \quad w_{qh}^s = 0, 1, 2; \quad w_{ph}^s + w_{qh}^s = 2.$$

Вероятность гаметы γ_k обозначим через g_k , а вероятность продуцирования k -той гаметы s -тым генотипом через $C_k(s)$. Через z_s обозначим вероятность генотипа Γ_s .

Тогда вероятность гаметы γ_k в следующем поколении, g_k^* , по формуле полной вероятности будет выражаться

$$g_k^* = \sum_{s=1}^{3^n} C_k(s) \cdot z_s. \quad (1)$$

Будем говорить, что популяция находится в равновесии, если начиная с t -того поколения будет соблюдаться условие

$$z_s^{(t)} = z_s^{(t+1)}.$$

Для равновесной популяции вероятность генотипа будет

$$z_s = 2^{-n} \cdot \prod_{h=1}^n \frac{\alpha_{ph}^{w_{ph}^s} \cdot \alpha_{qh}^{w_{qh}^s}}{(w_{ph}^s)! (w_{qh}^s)!}, \quad (2)$$

а гаметы

$$g_k = \prod_{h=1}^n \alpha_{ph}^{u_{ph}^k} \alpha_{qh}^{u_{qh}^k}. \quad (3)$$

В такой равновесной популяции можно рассматривать какой-либо признак $Q\left(\prod_{h=1}^n \alpha_{qh}^2\right)$ (либо заболевание), который в общем случае может определяться (детерминироваться) n генами. Индивид, обладающий этим набором генов, будем называть пораженным.

Введем обозначения:

$P(M:Q)$ — безусловная вероятность пораженной матери;

$P(F:Q)$ — безусловная вероятность пораженного отца;

$P(C:Q)$ — безусловная вероятность пораженного ребенка;

Причем все эти безусловные вероятности равны между собой и равны k^2 — частоте признака в популяции.

$P(C:Q/M:Q)$ — условная вероятность пораженного ребенка у пораженной матери;

$P(M:Q/C:Q)$ и $P(F:Q/C:Q)$ — условные вероятности соответственно пораженной матери и отца у пораженного ребенка;

$P(P:Q/C:Q)$ — условная вероятность хотя бы одного пораженного родителя у пораженного ребенка;

$P(S:Q/C:Q)$ — условная вероятность пораженного сибса у пробанда.

1. Найдем $P(C:Q/M:Q)$, зная, что $P(M:Q) = k^2$

Исходя из (2), имеем:

$$P(M:Q) = \prod_{h=1}^n a_{qh}^2 = k^2.$$

Обозначим через X материнскую гамету, а через Y — отцовскую. Тогда

$$P(C:Q/M:Q) = P(X: \prod_{h=1}^n \alpha_{qh}/M:Q \times Y: \prod_{h=1}^n \alpha_{qh}/M:Q).$$

С учетом панмиксии последнее выражение равняется:

$$P(X: \prod_{h=1}^n \alpha_{qh}/M:Q) \cdot P(Y: \prod_{h=1}^n \alpha_{qh}).$$

А так как

$$P(X: \prod_{h=1}^n \alpha_{qh}/M:Q) = 1,$$

а по формуле (3)

$$P(Y: \prod_{h=1}^n \alpha_{qh}) = \prod_{h=1}^n \alpha_{qh},$$

то

$$P(C:Q/M:Q) = \prod_{h=1}^n \alpha_{qh} = k = \sqrt{P(M:Q)}. \quad (*)$$

Таким образом, вероятность пораженного ребенка у пораженного родителя равна корню квадратному из вероятности пораженного родителя или корню квадратному из частоты признака в популяции.

2. Найдем теперь вероятность пораженного родителя (хотя бы одного из двух) при условии, что ребенок у них поражен, $P(P:Q/C:Q)$.

Определим сначала $P(M:Q/C:Q)$

Теорема 1: Вероятность пораженной матери у пораженного ребенка равна:

$$P(M:Q/C:Q) = \frac{1}{\prod_{h=1}^n \left(1 + 2 \frac{a_{ph}}{a_{qh}}\right)}.$$

Доказательство:

Доказательство проводим методом математической индукции. *Базис индукции* для $n=1$, т. е. в случае одного локуса.

Выразим по Гарди вероятности генотипов через частоты генов:

$$P(a_{p1} a_{p1}) = a_{p1}^2$$

$$P(a_{p1} a_{q1}) = 2 a_{p1} a_{q1}$$

$$P(a_{q1} a_{q1}) = a_{q1}^2.$$

Пусть генотип $Q = a_{q1}a_{q1}$ будет „пораженным“, т. е. признак определяется рецессивным геном. Тогда вероятность пораженной матери у пораженного ребенка будет:

$$P(M:Q/C:Q) = \frac{a_{q1}^2}{2 a_{p1} a_{q1} + a_{q1}^2} = \frac{1}{1 + 2 \frac{a_{p1}}{a_{q1}}}. \quad (4)$$

Индукционный шаг. Предполагая, что для $k=n-1$ наше утверждение верно, докажем его для $k=n$.

Для равновесной популяции частоты генотипов выражаются формулой (2). Тогда для $n-1$ -го локуса

$$P_{n-1}(M:Q/C:Q) = \frac{\prod_{h=1}^{n-1} a_{qh}^2}{\sum_{w_{ph}^s \geq 2} 2^{n-1} \prod_{h=1}^{n-1} \frac{a_{ph}^{w_{ph}^s} a_{qh}^{w_{qh}^s}}{(w_{ph}^s)! (w_{qh}^s)!}}. \quad (5)$$

А по предположению индукции,

$$P_{n-1}(M:Q/C:Q) = \frac{1}{\prod_{h=1}^{n-1} \left(1 + 2 \frac{a_{ph}}{a_{qh}}\right)}.$$

Условная вероятность пораженной матери у пораженного ребенка для n пар генов будет иметь вид:

$$P_n(M:Q/C:Q) = \frac{\prod_{h=1}^n a_{qh}^2}{\sum_{w_{ph}^s \geq 2} 2^n \prod_{h=1}^n \frac{a_{ph}^{w_{ph}^s} a_{qh}^{w_{qh}^s}}{(w_{ph}^s)! (w_{qh}^s)!}}. \quad (6)$$

Сравним теперь выражения (5) и (6). При переходе от $n-1$ к n парам генов в числителе добавляется новый член произведения — a_{qn}^2 , а в знаменателе число членов удваивается за счет умножения на множитель $a_{qn}^2 + 2 a_{pn} a_{qn}$:

$$P_n(M : Q/C : Q) = \frac{\prod_{h=1}^{n-1} a_{qh}^2}{\sum_{w_{ph}^s \neq 2} 2^{n-1} \prod_{h=1}^{n-1} \frac{a_{ph}^{w_{ph}^s} a_{qh}^{w_{qh}^s}}{(w_{ph}^s)! (w_{qh}^s)!}} \cdot \frac{a_{qn}^2}{a_{qn}^2 + 2 a_{pn} a_{qn}}. \quad (7)$$

Применим к полученному выражению базис индукции и предположение индукции.

Тогда:

$$P_n(M : Q/C : Q) = \frac{1}{\prod_{h=1}^{n-1} \left(1 + 2 \frac{a_{ph}}{a_{qh}}\right)} \cdot \frac{1}{1 + 2 \frac{a_{ph}}{a_{qh}}} = \frac{1}{\prod_{h=1}^n \left(1 + 2 \frac{a_{ph}}{a_{qh}}\right)}. \quad (8)$$

Аналогично будет выражаться и $P(F : Q/C : Q)$.

Тогда условная вероятность того, что хотя бы один из родителей пораженного ребенка поражен, будет:

$$P(P : Q/C : Q) = \frac{2}{\prod_{h=1}^n \left(1 + 2 \frac{a_{ph}}{a_{qh}}\right)} - \frac{1}{\left[\prod_{h=1}^n \left(1 + 2 \frac{a_{ph}}{a_{qh}}\right)\right]^2}. \quad (9)$$

Исследуя выражение (9) на максимум при

$$P(C : Q) = \prod_{h=1}^n a_{qh}^2 = k^2,$$

получим, что он достигается при условии

$$a_{qi} = k \quad (1 \leq i \leq n), \text{ а остальные } a_{qh} = 1.$$

При этом

$$\max P(P : Q/C : Q) = \frac{2}{\frac{2}{k} - 1} - \frac{1}{\left(\frac{2}{k} - 1\right)^2}, \quad (10)$$

а минимум достигается при условии

$$a_{q1} = a_{q2} = \dots = a_{qn} = k^{\frac{1}{n}} \quad (11)$$

и равен

$$\min P(P : Q/C : Q) = \frac{2}{\left(\frac{2}{k^{\frac{1}{n}}} - 1\right)^n} - \frac{1}{\left(\frac{2}{k^{\frac{1}{n}}} - 1\right)^{2n}}. \quad (**)$$

3. Найдем условную вероятность пораженного sibса при условии наличия больного ребенка в sibстве.

$$P(S : Q/C : Q) = P(X : \prod_{h=1}^n a_{qh}/C : Q \times Y : \prod_{h=1}^n x_{qh}/C : Q).$$

С учетом панмиксии это выражение будет равняться

$$P(S : Q/C : Q) = P(X : \prod_{h=1}^n a_{qh}/C : Q) \cdot P(Y : \prod_{h=1}^n x_{qh}/C : Q). \quad (12)$$

Теорема 2. Вероятность того, что материнская гамета будет иметь вид $\prod_{h=1}^n a_{qh}$ при условии наличия больного ребенка у нее, равна

$$P(X : \prod_{h=1}^n a_{qh}/C : Q) = \frac{1}{\prod_{h=1}^n (a_{qh} + 2 a_{ph})}$$

Доказательство:

Используя формулу (1), получим:

$$P(X : \prod_{h=1}^n a_{qh}/C : Q) = \sum_{s=1}^{3^n} C_x(S) \cdot P(\Gamma_s/C : Q),$$

где $C_x(S)$ — вероятность продуцирования гаметы $\prod_{h=1}^n a_{qh}$ генотипом Γ_s .

$P(\Gamma_s/C : Q)$ — вероятность того, что мать обладает генотипом Γ_s при условии, что ребенок поражен.

$$\begin{aligned} P(X : \prod_{h=1}^n a_{qh}/C : Q) &= \sum_{C_x(S) \neq 0} C_x(S) \cdot P(\Gamma_s/C : Q) + \\ &+ \sum_{C_x(S)=0} C_x(S) \cdot P(\Gamma_s/C : Q) = \sum_{C_x(S) \neq 0} C_x(S) \cdot P(\Gamma_s/C : Q). \end{aligned}$$

Очевидно, что $P(\Gamma_s/C : Q) = 0$ для тех S , для которых $C_x(S) = 0$. Таким образом,

$$P(\Gamma_s/C : Q) = \frac{z_s}{\sum_{C_x(i) \neq 0} z_i}.$$

Тогда

$$P(X : \prod_{h=1}^n a_{qh}/C : Q) = \sum_{C_x(S) \neq 0} C_x(S) \cdot \frac{z_s}{\sum_{C_x(i) \neq 0} z_i} = \frac{\sum_{C_x(S) \neq 0} C_x(S) \cdot z_s}{\sum_{C_x(i) \neq 0} z_i}. \quad (13)$$

Согласно (1) и (3),

$$\sum_{C_X(S) \neq 0} z_s \cdot C_X(S) = \prod_{h=1}^n a_{qh}. \quad (14)$$

Рассмотрим теперь $\sum_{C_X(I) \neq 0} z_i$. Используя (2), получим

$$\sum_{C_X(I) \neq 0} z_i = \sum_{w_{ph}^s = 2} 2^n \prod_{h=1}^n \frac{a_{ph}^{w_{ph}^s} a_{qh}^{w_{qh}^s}}{(w_{ph}^s)! (w_{qh}^s)!}. \quad (15)$$

Подставим (14) и (15) в (13), а также умножим и разделим полученное выражение на $\prod_{h=1}^n a_{qh}$

$$P(X : \prod_{h=1}^n a_{qh} / C : Q) = \frac{\prod_{h=1}^n a_{qh}^2}{\prod_{h=1}^n a_{qh} \left(\sum_{w_{ph}^s = 2} 2^n \prod_{h=1}^n \frac{a_{ph}^{w_{ph}^s} a_{qh}^{w_{qh}^s}}{(w_{ph}^s)! (w_{qh}^s)!} \right)}.$$

Согласно формулам (7) и (8), это выражение равно

$$P(X : \prod_{h=1}^n a_{qh} / C : Q) = \frac{1}{\prod_{h=1}^n (a_{qh} + 2 a_{ph})}.$$

Аналогично для отца

$$P(Y : \prod_{h=1}^n a_{qh} / C : Q) = \frac{1}{\prod_{h=1}^n (a_{qh} + 2 a_{ph})}.$$

Тогда

$$P(S : Q / P : Q) = \frac{1}{\left[\prod_{h=1}^n (a_{qh} + 2 a_{ph}) \right]^2}.$$

Максимум выражения, как и в предыдущем случае достигается при условии (10), а минимум — при условии (11)

$$\max P(S : Q / C : Q) = \frac{1}{(2 - k)^2}. \quad (16)$$

$$\min P(S : Q / C : Q) = \frac{1}{\left(2 - k \frac{1}{n}\right)^{2n}}. \quad (***)$$

Формулы (10) и (16) малопригодны, так как сводят задачу к моногенному случаю, и они следуют из (**) и (***) для $n=1$. Интересной представляется формула (*), в которой показано, что вероятность по-

пораженного ребенка у пораженного родителя в популяционном обследовании не зависит от числа генов, вовлеченных в определение признака, и равна корню квадратному из частоты признака в популяции. Благодаря тому, что формулы (**) и (***) зависят от n , появляется возможность судить о наименьшем числе пар генов, вовлеченных в определение признака. Последовательное применение формул (*), (**) и (***) позволит не только судить о наследственной передаче признака, но и определить возможное минимальное число генов, ответственных за проявление исследуемого признака.

Таким образом, полученные вероятности пораженных родителей, детей и sibсов у пораженных и их максимальные и минимальные оценки при известной частоте признака в популяции могут, несомненно, быть полезны, наряду с другими методами, для выяснения наследственной природы изучаемого признака.

Институт кардиологии МЗ АрмССР,
Чаренцаванский филиал ВНИИ генетики,
ЕрНИИММ

Поступило 17.VII 1972 г.

Ա. Ս. ԶՈՒՐԱՐՅԱՆ, Շ. Մ. ՔՈԶԱՐՅԱՆ, Ա. Վ. ՊԵՏՐՈՍՅԱՆ

ՄԱՐԴՈՒ ՀԱՏԿԱՆԻՇՆԵՐԻ ԺԱՌԱՆԳՄԱՆ ԶԵՎԻ ՈՐՈՇԵԼՈՒ ՄԱԹԵՄԱՏԻԿԱԿԱՆ ԶԱՓԱՆԻՇՆԵՐԸ

Ա մ փ ո փ ո լ մ

Ստացված են բանաձևեր, որոնց օգնությամբ կարելի է որոշել հատկանիշների ժառանգման ձևը, եթե հայտնի է այդ հատկանիշի հաճախականությունը պոպուլյացիայում:

Л И Т Е Р А Т У Р А

1. Петросян А. В., Кочарян Ш. М., Зурабян А. С. Биологический журнал Армении, 2, 1973.
2. Ниль Д. Ж. и Шелл У. Наследственность человека. ИЛ, 1958.