

第七章 群体遗传学

Population genetics

研究基因在人群中的行为。

- 怎样估计群体的基因频率和基因型频率？
- 哪些因素可影响基因频率？
- 突变和选择对维持基因频率有什么作用？
- 怎样增进我们子孙后代的遗传素质？



AD: $Dd \times Dd \rightarrow$ 子代中3/4是患者

AR: $Dd \times Dd \rightarrow$ 子代中1/4是患者

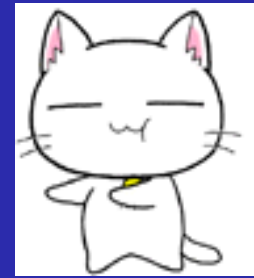
人群中显性基因 \uparrow ? 隐性基因 \downarrow ?

随着隐性纯合子患者死亡，隐性致病基因是否会在人群中消失掉？

一、Hardy-Weinberg平衡

条件：

- 无限大的群体
- 随机婚配
- 没有突变
- 没有选择
- 没有迁移
- 没有遗传漂变（小群体内基因频率随机波动）



结论：群体内一个位点上的基因型频率和基因频率将代代保持不变，处于遗传平衡状态。

Hardy-Weinberg平衡律的推导:

假设人群中某位点的等位基因A和a，有AA、Aa、aa三种基因型，A的频率为p，a的频率为q，

在平衡状态下， $p+q=1$

在随机婚配人群中A或a精子与A或a卵子随机结合，受精卵（子₁代）的基因型及其频率可表示如下：

		精子	
		A	a
卵	A	AA	Aa
	a	Aa	aa
子	p	p ²	pq
	q	pq	q ²

子₁代的基因型频率和基因频率分别是：

AA: p²; Aa: pq+pq=2pq; aa: q²

子₁代产生A、a基因配子的概率分别是：

$P_A = p^2 + 1/2(2pq) = p^2 + pq = p(p+q) = p$

$P_a = q^2 + 1/2(2pq) = q^2 + pq = q(p+q) = q$

子₁代群体在随机婚配的情况下，其婚配类型及频率：



		男性群体		
		AA p^2	Aa $2pq$	aa q^2
女性群体	AA p^2	AA x AA p^4	AA x Aa $2p^3q$	AA x aa p^2q^2
	Aa $2pq$	Aa x AA $2p^3q$	Aa x Aa $4p^2q^2$	Aa x aa $2pq^3$
	aa q^2	aa x AA p^2q^2	aa x Aa $2pq^3$	aa x aa q^4

男性群体与女性群体这三种基因型随机婚配共有六种婚配类型，由上表可写出这六种婚配类型的频率，见下表：

婚配类型	频率
AA x AA	p^4
AA x Aa	$2p^3q+2p^3q=4p^3q$
Aa x Aa	$4p^2q^2$
AA x aa	$p^2q^2+ p^2q^2= 2p^2q^2$
Aa x aa	$2pq^3+2pq^3=4pq^3$
aa x aa	q^4

子₁代每种婚配类型的频率及每种婚配类型所产生的子₂代各种基因型的频率列成下表：

子 ₁ 代 婚配类型	婚配频率	子 ₂ 代基因型及频率		
		AA	Aa	aa
AA x AA	p^4	p^4	—	—
AA x Aa	$4p^3q$	$2p^3q$	$2p^3q$	—
Aa x Aa	$4p^2q^2$	p^2q^2	$2p^2q^2$	p^2q^2
AA x aa	$2p^2q^2$	—	$2p^2q^2$	—
Aa x aa	$4pq^3$	—	$2pq^3$	$2pq^3$
aa x aa	q^4	—	—	q^4

子₂代基因型频率：

$$= p^2 (p^2 + 2pq + q^2) [AA] + 2pq (p^2 + 2pq + q^2) [Aa] + q^2 (p^2 + 2pq + q^2) [aa]$$

$$= p^2 (p+q)^2 [AA] + 2pq (p+q)^2 [Aa] + q^2 (p+q)^2 [aa]$$

$$= p^2 [AA] + 2pq [Aa] + q^2 [aa]$$

Hardy-Weinberg公式可由简单的数学推导得出。在双等位基因位点，**Hardy-Weinberg**公式即二项式平方展开后所得各项。

$$(p+q)^2=p^2+2pq+q^2$$

即纯合基因型频率是基因频率的自乘，其系数是1，杂合基因型频率是二个有关基因频率的乘积，其系数是2。

设群体中某位点有n个等位基因 A_1 ， A_2 ，…… A_n ，其基因频率分别为 p_1 ， p_2 ，…… p_n ，则基因型频率可以下式表示：

$$[p_1(A_1)+p_2(A_2)+\dots\dots+p_n(A_n)]^2$$

二、Hardy-Weinberg平衡的应用

(一) 估计基因频率和基因型频率

(二) 检验遗传假设



(一)估计基因频率和基因型频率

1、常染色体上的基因

(1) 等显性基因

基因计数法：基因频率通过表现型频率来推算。

例：MN血型系统(M、N为等显性的等位基因)
调查上海地区汉族1,788人，其中M血型397人，N血型530人，MN血型861人，试计算M、N基因频率。

每个人每个位点都有一对等位基因，所以，1788个人在某个位点上应该共有 1788×2 个基因。

397个M血型的人，每个人在MN位点上的两个基因都是M基因，所以，M基因数共有 397×2 个。

861个MN血型的人，每个人在MN位点上各有一个M基因和N基因，所以M基因数为861个，N基因数也为861个。

530个N血型的人，应该共有 530×2 个N基因。

$$P_M = \frac{397 \times 2 + 861}{1788 \times 2} = 0.4628$$

$$P_N = \frac{530 \times 2 + 861}{1788 \times 2} = 0.5372$$



检验MN血型群体调查是否符合Hardy-Weinberg平衡:

表7-3 MN血型调查资料的遗传平衡吻合度 χ^2 测验

	$L^M L^M$	$L^M L^N$	$L^N L^N$	合计
观察数(O)	397	861	530	1788
期望数(E)	382.96 (np^2)	889.05 ($n2pq$)	515.99 (nq^2)	1788 (n)
$\frac{ O-E }{E}^2$	0.51	0.88	0.38	$\chi^2_{(1)}=1.77$

计算期望数:

$$np^2 = 1788 \times (0.4628)^2 = 1788 \times 0.2142 = 382.96$$

$$n2pq = 1788 \times 2 \times 0.4628 \times 0.5372 = 889.05$$

$$nq^2 = 1788 \times (0.5372)^2 = 1788 \times 0.2886 = 515.99$$

(2) 基因有显、隐性之分 (方根法)

a) 若为AR遗传, 患者肯定是隐性纯合子,

则 $q^2=x$ (x 为群体发病率)

$$q = \sqrt{x}$$

由于 $p+q=1$, $p=1-q$, 因此显性基因频率是

$$p = 1 - \sqrt{x}$$

例：尿黑酸尿症（AR遗传）

调查得知，约1,000,000儿童中由1个是患儿，其发病率 $x = 0.000001$ 。按Hardy-Weinberg律计算：

尿黑酸尿症隐性基因频率为 $q = \sqrt{x} = 0.001$

正常显性基因频率为 $p = 1 - q = 0.999 \approx 1$

杂合子频率为 $2pq = 2 \times 1 \times 0.001 = 0.002$

杂合子频率和纯合子频率之比为：

$$2pq / q^2 = 2/q$$

b) 若为AD遗传，因为显性致病基因的纯合子个体都会死亡，所以患者都是杂合子。

则 $2pq=y$ (y 为群体发病率)

当 p 很小时，即致病基因在人群中分布频率比较低时， $q \approx 1$ ， $y=2pq \approx 2p$ ， $p=y/2$

(3)复等位基因

例：人类ABO血型是由三个复等位基因 I^A ， I^B ， i 所决定的，设它们的基因频率分别为 p 、 q 、 r 。

则 $p+q+r = 1$

基因型的频率为：

$$(p+q+r)^2=p^2+2pq+q^2+2qr+r^2+2pr$$

A型频率(\bar{A}) 为 p^2+2pr

B型频率(\bar{B}) 为 q^2+2qr

AB型频率(\overline{AB}) 为 $2pq$

O型频率(\bar{O}) 为 r^2

所以: $r = \sqrt{O}$

$q^2+2qr+r^2=(q+r)^2=(1-p)^2=B+O$

$1-p = \sqrt{B+O}$ $p = 1 - \sqrt{B+O}$

同理, $q = 1 - \sqrt{A+O}$ (或 $q = 1 - p - r$)



2、X染色体上的基因

男性是半合子，表现型频率=基因型频率
=基因频率。

例：Xg 血型基因频率的计算：

表7-4 Xg血型基因频率的计算

性别	血型	基因型	实得人数	表现型频率	基因频率
男性	Xg ^(a+)	Xg ^a Y	188	0.631	p=0.631
	Xg ^(a-)	XgY	110	0.369	q=0.369
女性	Xg ^(a+)	Xg ^a Xg ^a Xg ^a Xg	260	0.893	P=1-0.327 =0.673
	Xg ^(a-)	XgXg	31	0.107	q= $\sqrt{0.107}$ =0.327

表6-5 男女群体中Xg血型基因频率的期望值

性别	血型	观察值 (人数)	期望值 ₁ ^① (人数)	期望值 ₂ ^① (人数)
男性	Xg ^(a+)	188	298x0.631=188.04	298x0.673=200.55
	Xg ^(a-)	110	298x0.369=109.96	298x0.327=97.45
女性	Xg ^(a+)	260	291x(1-0.136)=251.37	291x(1-0.107)=259.86
	Xg ^(a-)	31	291x(0.369) ² =39.62	291x(0.327) ² =31.14

注：①期望值₁按男性群体的基因频率计算
期望人数

期望值₂按女性群体的基因频率计算
期望人数。

在随机婚配的人群中，**XR**遗传，设隐性致病基因的频率为 q ，显性基因的频率为 p ，

则男/女= $q/q^2=1/q$ ， $q<1$ ， $1/q>1$ ，

可见，**XR**遗传中，男性患者要多于女性患者。

而且，隐性致病基因越稀有，相对地男性患者的比例越高。

在**XD**遗传中,设显性致病基因频率为 p 隐性基因频率为 q ，则男/女= $p/(p^2+2pq)=1/(p+2q)$

$=1/(p+2-2p)=1/(2-p)$

$p<1$ ， $1/(2-p)<1$ ，可见，**XD**遗传中，女性患者要多于男性患者。

(二) 检验遗传假设

根据Hardy-Weinberg平衡可计算各种遗传假设的期望值，用来与观察值比较，以检验遗传假设是否符合于客观实际。

例：苯硫脲(PTC)尝味试验

对苯硫脲敏感-----尝味者

对苯硫脲不敏感-----非尝味者

假设受控于一对等位基因(T,t)

TT和Tt 为尝味者， tt为非尝味者

观察值:

- 群体数据: 已知在随机群体样本3,643人中, 尝味者是2,557人, 非尝味者是1,086人其表现型频率为0.298。按遗传假设非尝味者为隐性纯合子, 因此 $q = \sqrt{0.298} = 0.546$
- 家系数据:

婚配类型	家庭数	子代		非尝味者表现型频率
		尝味者	非尝味者	
尝味者x尝味者	425	929	130	0.123
尝味者x非尝味者	289	483	278	0.365

期望值:

- 尝味者x尝味者家庭子代的基因型及频率

婚配类型	婚配频率	子代	
		尝味者频率	非尝味者频率
TTxTT	p^4	p^4	——
TTxTt	$4p^3q$	$4p^3q$	——
TtxTt	$4p^2q^2$	$3p^2q^2$	p^2q^2

子代中非尝味者所占频率是:

$$\frac{p^2q^2}{p^4+4p^3q+4p^2q^2} = \frac{q^2}{p^2+4pq+4q^2} = \frac{q^2}{(p+2q)^2} = \left(\frac{q}{p+2q}\right)^2 = \left(\frac{q}{1-q+2q}\right)^2 = \left(\frac{q}{1+q}\right)^2$$

期望值:

- 尝味者 x 非尝味者家庭子代的基因型及频率

婚配类型	婚配频率	子代	
		尝味者频率	非尝味者频率
TT x tt	$2p^2q^2$	$2p^2q^2$	-
Tt x tt	$4pq^3$	$2pq^3$	$2pq^3$

子代中非尝味者的频率是:

$$\frac{2pq^3}{2p^2q^2 + 4pq^3} = \frac{q}{p + 2q} = \frac{q}{1 - q + 2q} = \frac{q}{1 + q}$$

根据群体调查数据，得出： $q = \sqrt{x} = 0.546$

代入公式，可以得出不同婚配类型，子代非尝味者的期望值：

尝味者 x 尝味者婚配的家庭中，子代非尝味者的期望值：

$$\left(\frac{q}{1+q}\right)^2 \times (929+130) = 0.125 \times 1059 = 132.375$$

尝味者 x 非尝味者婚配的家庭中，子代非尝味者的期望值：

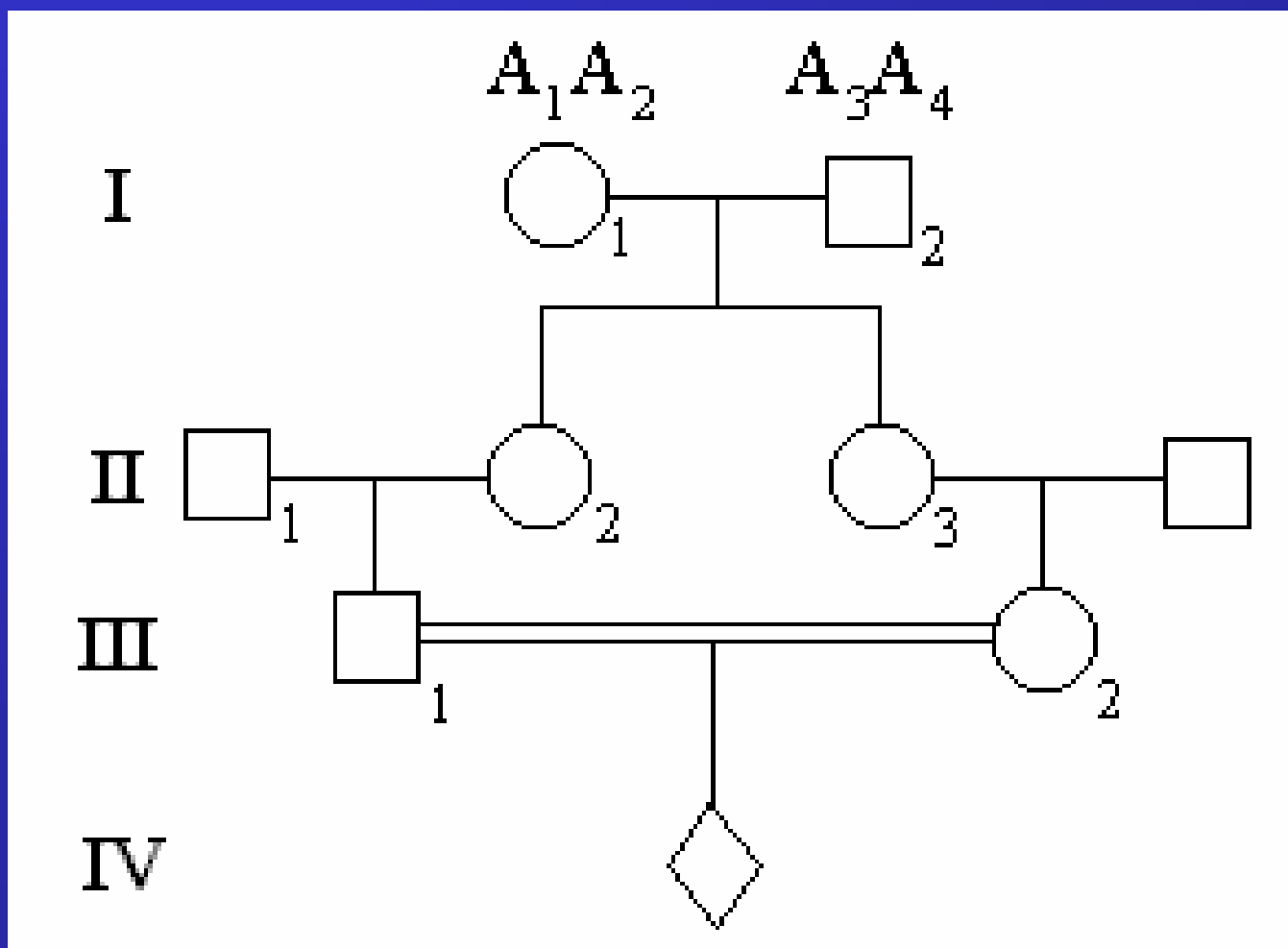
$$\frac{q}{1+q} \times (483+278) = 0.353 \times 761 = 268.633$$

- 在尝味者 \times 尝味者婚配类型中，子代非尝味者和尝味者的观察值分别是130和929，期望值分别是132.375和926.625。经卡方检验， P 大于0.80,二者无显著差异。
- 在尝味者 \times 非尝味者婚配类型中，子代非尝味者和尝味者的观察值分别是278和483，期望值分别是268.633和492.367。经卡方检验， P 大于0.30,二者无显著差异。
- 根据Hardy-Weinberg平衡，苯硫脲(PTC)尝味能力的遗传假设成立。

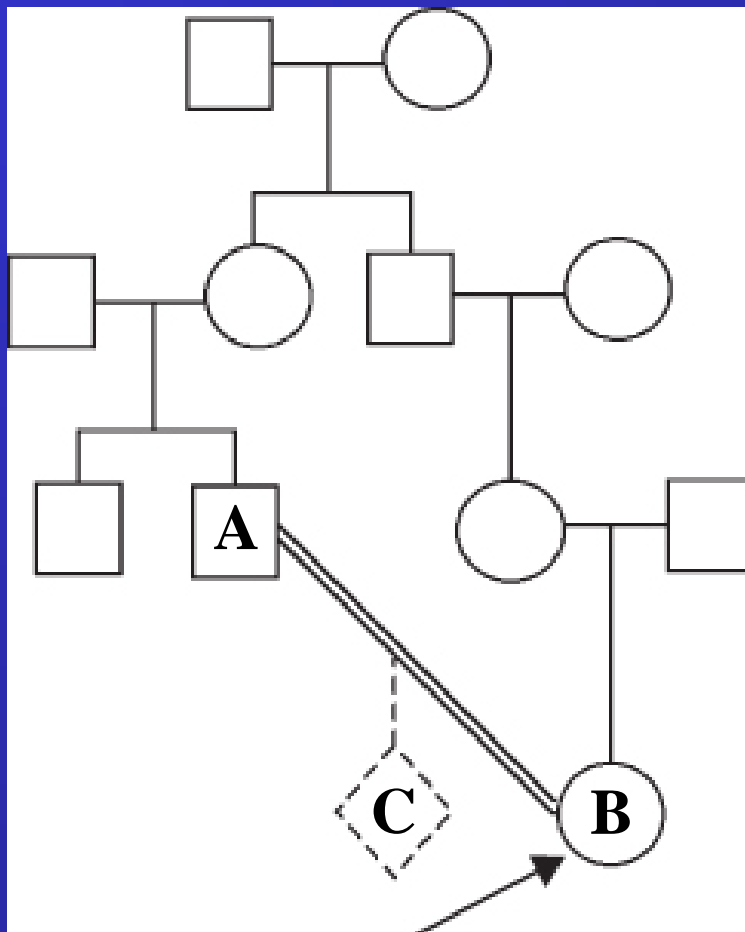
三、近婚系数(inbreeding coefficient, F)

- 近亲婚配：指两个配偶在几代之内曾有共同祖先。
- 近婚系数：有亲缘关系的配偶，他们从共同祖先得到同一基因又将这同一基因同时传递给他们的子女使之成为纯合子的概率。

1. 常染色体基因近婚系数的计算 (以姨表兄妹为例)



试计算A、B之间的亲缘系数 k 及C的近婚系数 F 。



first cousin once removed

2. 近亲婚配的有害效应:

$$\frac{\frac{q}{16} + \frac{15}{16} q^2}{q^2} = \frac{q+15q^2}{16q^2} = \frac{1+15q}{16q}$$

表7-10表亲婚配和随机群体婚配出生隐性纯合子的概率

基因频率 q	来自随机婚配双亲的隐性纯合子 q ²	pq/16	来自表亲婚配的隐性纯合子 q ² +pq/16	二者之比 (1+15q) /16q
0.20	0.0400	0.010	0.0500	1.25
0.10	0.0100	0.005625	0.015625	1.56
0.04	0.0016	0.0024	0.0040	2.50
0.02	0.0004	0.001225	0.001625	4.06
0.01	0.0001	0.000619	0.000719	7.19
0.001	0.000001	0.0000625	0.0000635	63.50

四、影响Hardy-Weinberg 平衡的因素

- 1、突变
- 2、选择
- 3、随机遗传漂变
- 4、迁移
- 5、遗传异质性



1. 突变 (mutation)

自然突变率： $u=n \times 10^{-6}$ /配子/位点/代

表7-11 几种遗传病的突变率

病名	遗传方式	突变率 (突变数 / 10^6 配子)
软骨发育不全症	AD	10
视网膜母细胞瘤	AD	24~46
强直性肌营养不良	AD	8~16
神经纤维瘤	AD	1 00
Marfan综合征 (蜘蛛指综合征)	AD	4.2~5.8
血友病 (甲型+乙型)	XR	22~57
Duchenne假肥大性肌营养不良	XR	43~95
白化病	AR	28
苯丙酮尿症	AR	25
多发性肠息肉	AD	13
Huntington舞蹈病	AD	5

2、选择 (selection)

(1) 生殖适合度 (f) :

为后代提供基因能力的一种量度

以正常的生育力为1作比较。

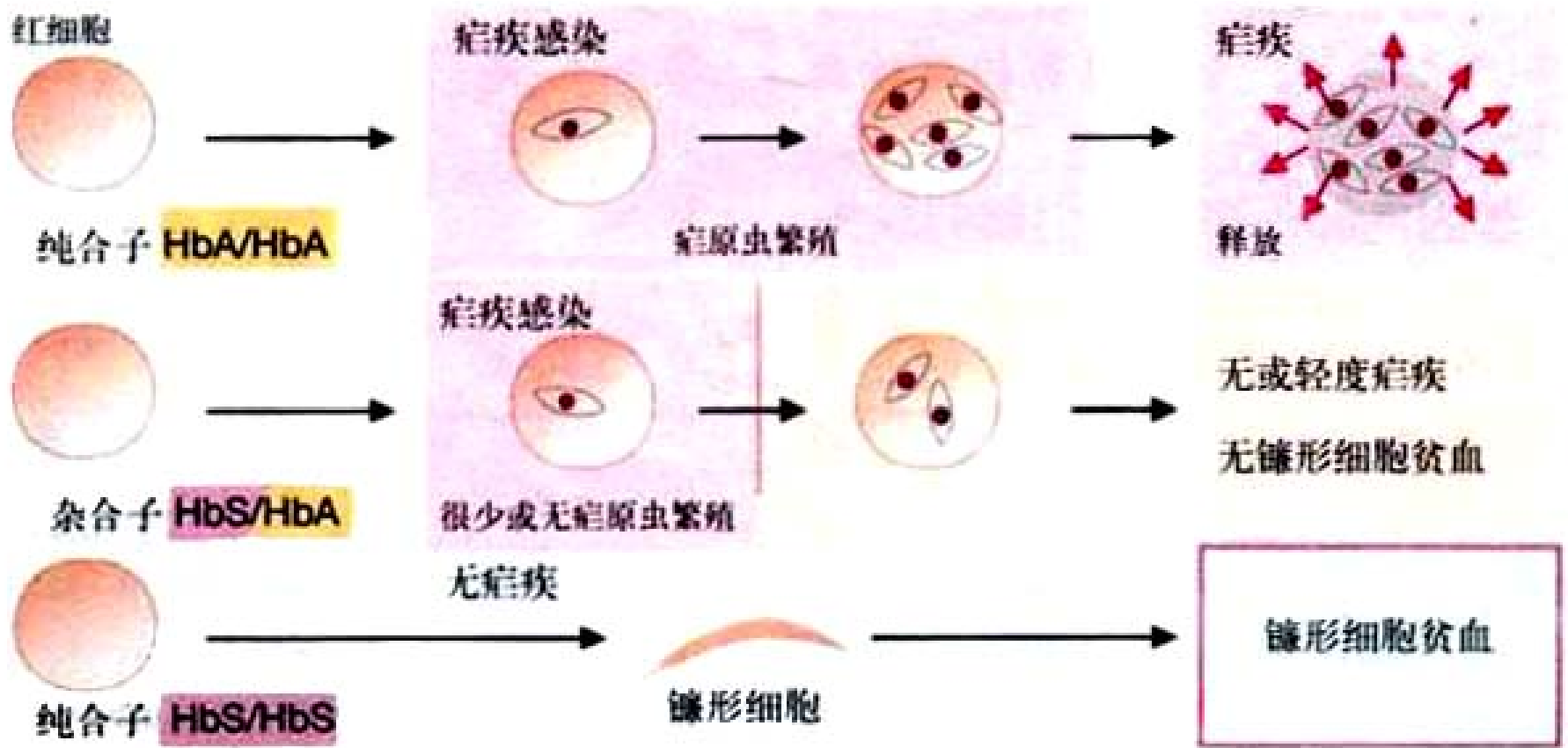
如调查软骨发育不全症患者108人，有孩子27人；患者未受累同胞457人，有孩子582人。如果软骨发育不全症患者生育能力与正常同胞相同，108人中应有多少孩子？

$$582 \times (108/457) = 137.5$$

$$f = 27/137.5 = 0.1963$$

(2) 杂合子优势(heterozygote advantage)

在某些隐性遗传病中，在特定的条件下杂合子可能比正常纯合子个体更有利与生存而繁殖后代。



HbS杂合子在疟区的选择性优势

3、随机遗传漂变 (random genetic drift)

小群体或隔离的人群中，基因频率的随机波动称为遗传漂变。

例如北美印第安人红细胞血型大部分是O型。



4、迁移 (migration)

不同种族和不同民族的基因频率可有差异。

迁移的结果使不同人群通婚，彼此掺入外来基因，导致基因流动，可改变原来群体的基因频率。

5、遗传异质性

由于相似的病理性状（如白化症）可受控于若干不同的基因位点，若不加以严格区分，往往会使Hardy-Weinberg平衡复杂化。

在实际应用中，符合理想群体的情况一般是不存在的，但Hardy-Weinberg平衡还是可以应用的，这是因为：

- 我们所调查的群体不可能大到足以显示出这些因素。
- 各种影响因素可相互抵消，如突变和选择。

五、历史上的优生运动不符合群体遗传学原理

表 7-12 随机婚配群体中去除 aa 后基因频率的变化

基因型	频率	去除 aa	新基因数	新基因频率
AA	p^2	p^2	$(A)p^2+pq=p$ $(a)0+pq=pq$
Aa	$2pq$	$2pq$		
aa	q^2	0		
合计	1.00	$p(p+2q)=p(1+q)$	$p+pq=p(1+q)$	1.00

新的隐性基因频率用 q' 表示:

$$q' = \frac{pq}{p+pq} = \frac{q}{1+q}$$

用 q_0 和 q_1 、 q_2分别表示原代和经一代、二代选择后的基因频率。

$q_1=q_0/(1+q_0)$; $q_2=q_1/(1+q_1)$ 等等, 则:

$$q_2 = \frac{q_0/(1+q_0)}{1 + q_0/(1+q_0)} = \frac{q_0}{1 + 2q_0}$$

这样, 对于经 n 代的选择来说, 就可以得到:

$$q_n=q_0/(1+nq_0);$$

想知道需要经过多少代的选择才能使 q_0 降到 q_n 简单的办法就是利用公式

$q_n = q_0 / (1 + nq_0)$, 求出 n 的值:

$$q_n (1 + nq_0) = q_0$$

$$nq_0q_n = q_0 - q_n$$

$$n = 1/q_n - 1/q_0$$

基因频率如果要降到原先的 $1/2$, 则:

$$n = 1/q_0$$