



遗传学

哈尔滨工业大学

HARBIN INSTITUTE OF TECHNOLOGY

主讲教师 张凤伟 副教授

第九章 群体遗传分析

主讲教师 张凤伟

□ 群体遗传学（population genetics）：是应用数学和统计学方法研究群体遗传结构及其变化规律的科学。

主要作用：

- 阐明生物进化的遗传机制
- 探讨遗传病的发病频率及规律
-

第一节 哈迪 - 温伯格平衡定律

基本概念



□ 孟德尔群体(Mendilian population): 指**一定地域内**一群可以相互交配且能够产生可育子代的所有生物个体;

□ 基因库(gene pool): 群体中所有个体的全部基因;

关系: 一个孟德尔群体, 享有一个共同的基因库;

群体的遗传结构：

两个主要参数：基因频率和基因型频率(相互推算)

- 基因频率(gene frequency)：指在一个群体内的特定基因座上，某一等位基因占该基因座等位基因总数的比率，也称等位基因频率(alleles frequency)；
- 基因型频率(genotype frequency)：群体中某特定基因型个体的数目，占个体总数的比率；

基因频率和基因型频率的相关符号

(两个等基因)

	等位基因		基因型		
基因符号	A	a	AA	Aa	aa
频率符号	p	q	D	H	R

基因频率和基因型频率的相关符号

(复等位基因)

	等位基因			基 因 型					
基因符号	A_1	A_2	A_3	A_1A_1	A_1A_2	A_1A_3	A_2A_2	A_2A_3	A_3A_3
频率符号	p_1	p_2	p_3	P_{11}	P_{12}	P_{13}	P_{22}	P_{23}	P_{33}

怎样由基因型频率计算基因频率?

设：群体总个体数 N $\left\{ \begin{array}{l} AA \cdots n_1 \\ Aa \cdots n_2 \\ aa \cdots n_3 \end{array} \right. \begin{array}{l} \longrightarrow N = n_1 + n_2 + n_3 \\ \longrightarrow \text{等位基因总数为 } 2N \end{array}$

则：3种基因型的频率为：

$$AA: D = \frac{n_1}{N}, \quad Aa: H = \frac{n_2}{N}, \quad aa: R = \frac{n_3}{N}; \quad D + H + R = 1$$

\therefore 每个AA个体有两个A，Aa有一个A和一个a，aa有两个a；

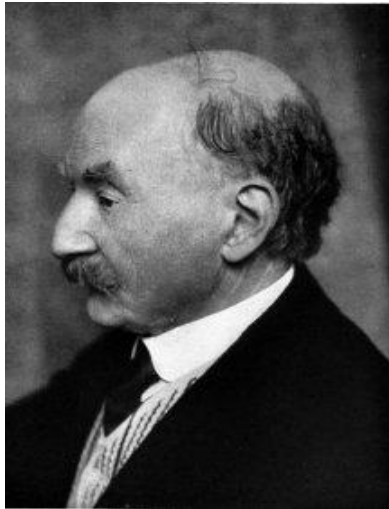
\therefore 等位基因A的频率： $p = \frac{2n_1 + n_2}{2N} = D + \frac{1}{2}H \cdots \textcircled{1}$

等位基因a的频率： $q = \frac{2n_3 + n_2}{2N} = R + \frac{1}{2}H \cdots \textcircled{2}$

Hardy – Weinberg 平衡定律



1908年，英国数学家**哈迪（Hardy）**和德国医生**温伯格（Weinberg）**分别独立推导出随机交配群体基因频率和基因型频率变化规律，称为Hardy-weinberg定律，又称**遗传平衡定律**。



Hardy



Weinberg

■ 定律内容

在随机交配的大群体中，若没有其它因素(选择、突变和迁移)的影响：

- ✓ 基因频率保持世代不变；
- ✓ 无论基因频率和基因型频率如何，只要经过一个世代的随机交配，一个基因座的基因型频率就达到平衡，即基因型频率也保持世代不变；
- ✓ 平衡群体中，基因型频率和基因频率的关系为：

$$D = p^2$$

$$R = q^2$$

$$H = 2pq$$

■ 定律推导

设：一个初始群体内三个基因型频率是：

$$AA = D_0, \quad Aa = H_0, \quad aa = R_0$$

$$\text{则： } p_0 = D_0 + \frac{1}{2}H_0$$

$$q_0 = R_0 + \frac{1}{2}H_0$$

亲代	基因型频率			基因频率	
	AA	Aa	aa	A	a
	D_0	H_0	R_0	p_0	q_0

亲代三种基因型产生A和a型配子的频率				
基因型	AA	Aa	aa	总和
基因型频率 (概率)	D_0	H_0	R_0	
A配子概率	D_0	$\frac{1}{2} H_0$	0	p_0
a配子概率	0	$\frac{1}{2} H_0$	R_0	q_0



一个世代产生的配子中含某一等位基因的概率与当代该等位基因的基因频率相等

若个体间随机交配，配子间的随机结合可得以下结果：

雌配子	雄配子	
	$A(p_0)$	$a(q_0)$
$A(p_0)$	$AA(p_0^2)$	$Aa(p_0q_0)$
$a(q_0)$	$Aa(p_0q_0)$	$aa(q_0^2)$

→

$$D_1 = p_0^2$$

$$H_1 = 2p_0q_0$$

$$R_1 = q_0^2$$

由上式及公式①和② →

$$\begin{cases} p_1 = D_1 + \frac{1}{2}H_1 = p_0^2 + \frac{1}{2} \cdot 2p_0q_0 = p_0 \\ q_1 = R_1 + \frac{1}{2}H_1 = q_0^2 + \frac{1}{2} \cdot 2p_0q_0 = q_0 \end{cases}$$

第一代	基因型频率			基因频率	
	AA	Aa	aa	A	a
	p_0^2	$2p_0q_0$	q_0^2	p_0	q_0

由配子间的随机结合可知：

雌配子	雄配子	
	$A (p_1=p_0)$	$a (q_1=q_0)$
$A (p_1=p_0)$	$AA (p_0^2)$	$Aa (p_0 q_0)$
$a (q_1=q_0)$	$Aa (p_0 q_0)$	$aa (q_0^2)$

→

$$D_2 = p_0^2$$

$$H_2 = 2 p_0 q_0$$

$$R_2 = q_0^2$$

由上式及公式①和②

→

$$\begin{cases} p_2 = D_2 + \frac{1}{2}H_2 = p_0^2 + \frac{1}{2} \cdot 2 p_0 q_0 = p_0 \\ q_2 = R_2 + \frac{1}{2}H_2 = q_0^2 + \frac{1}{2} \cdot 2 p_0 q_0 = q_0 \end{cases}$$

第二代	基因型频率			基因频率	
	AA	Aa	aa	A	a
	p_0^2	$2 p_0 q_0$	q_0^2	p_0	q_0

F_0

亲代	基因型频率			基因频率	
	AA	Aa	aa	A	a
	D_0	H_0	R_0	p_0	q_0

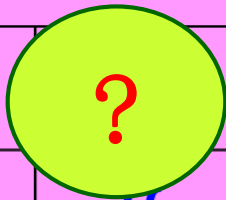
 F_1

第一代	基因型频率			基因频率	
	AA	Aa	aa	A	a
	p_0^2	$2 p_0 q_0$	q_0^2	p_0	q_0

 F_2

第二代	基因型频率			基因频率	
	AA	Aa	aa	A	a
	p_0^2	$2 p_0 q_0$	q_0^2	p_0	q_0

F_0

亲代	基因型频率			基因频率	
	AA		aa	A	a
	D_0	H_0	R_0	p_0	q_0

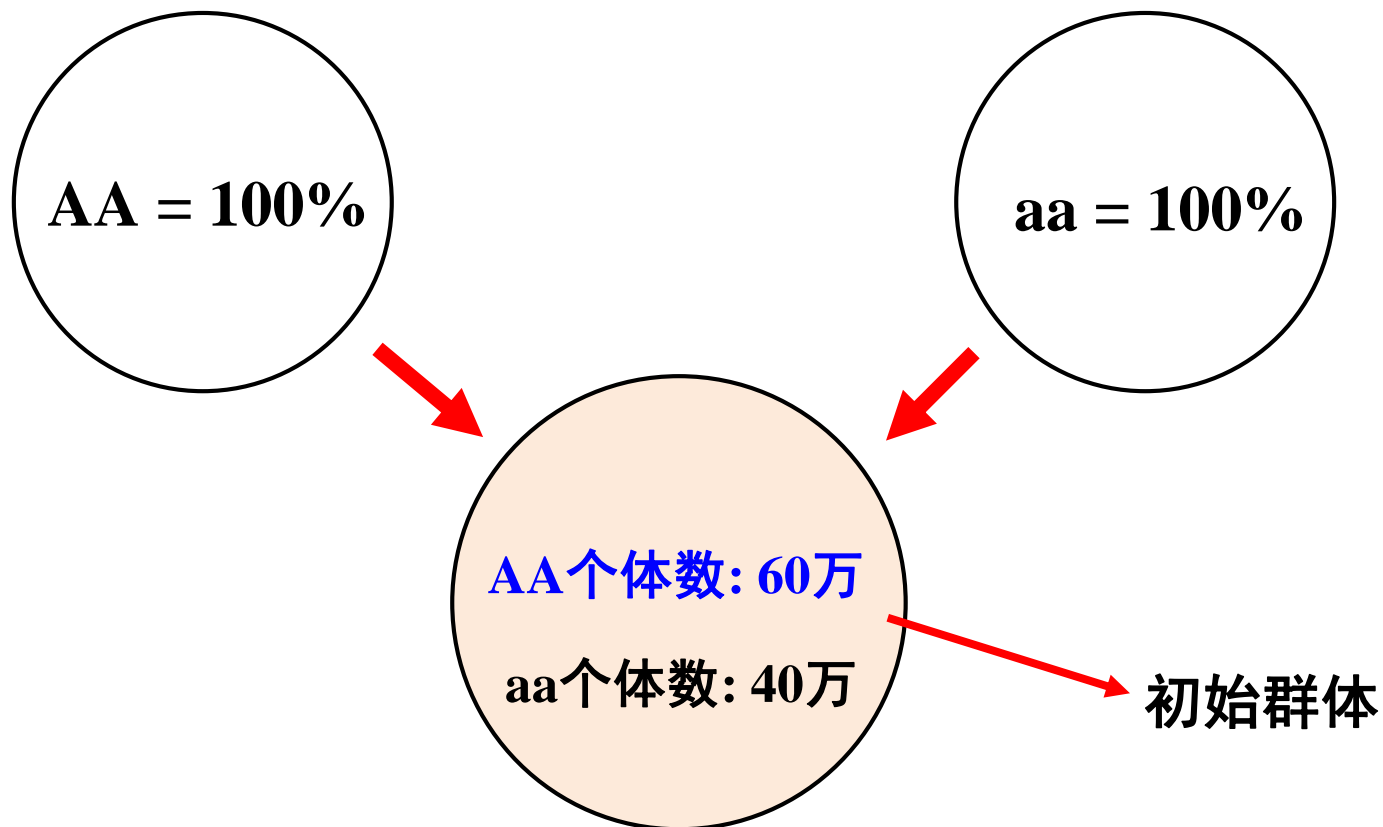
F_1

第一代	基因型频率			基因频率	
	AA	Aa	aa	A	a
	p_0^2	$2 p_0 q_0$	q_0^2	p_0	q_0

F_2

第二代	基因型频率			基因频率	
	AA	Aa	aa	A	a
	p_0^2	$2 p_0 q_0$	q_0^2	p_0	q_0

例如：



①. 初始群体(亲代):

基因型频率: $D_0 = 0.6$, $H_0 = 0$, $R_0 = 0.4$

基因频率: $p_0 = D_0 + (1/2) H_0 = 0.6 + (1/2) (0) = 0.6$

$q_0 = R_0 + (1/2) H_0 = 0.4 + (1/2) (0) = 0.4$

②. 随机交配后的F₁代:

基因型频率: $D_1 = p_0^2 = 0.6^2 = 0.36$

$$H_1 = 2 p_0 q_0 = 2 \times 0.6 \times 0.4 = 0.48$$

$$R_1 = q_0^2 = 0.4^2 = 0.16$$

基因频率: $p_1 = D_1 + (1/2) H_1 = 0.36 + (1/2) (0.48) = 0.6$

$$q_1 = R_1 + (1/2) H_1 = 0.16 + (1/2) (0.48) = 0.4$$

③. 随机交配后的F₂代:

基因型频率: $D_2 = p_1^2 = 0.6^2 = 0.36,$

$$H_2 = 2 p_1 q_1 = 2 \times 0.6 \times 0.4 = 0.48,$$

$$R_2 = q_1^2 = 0.4^2 = 0.16$$

基因频率: $p_2 = D_2 + (1/2) H_2 = 0.36 + (1/2) (0.48) = 0.6$

$$q_2 = R_2 + (1/2) H_2 = 0.16 + (1/2) (0.48) = 0.4$$

■ 定律内容

在随机交配的大群体中，若没有其它因素(选择、突变和迁移)的影响：

- ✓ 基因频率保持世代不变；
- ✓ 无论基因频率和基因型频率如何，只要经过一个世代的随机交配，一个基因座的基因型频率就达到平衡，即基因型频率也保持世代不变；
- ✓ 平衡群体中，基因型频率和基因频率的关系为：

$$D = p^2$$

$$R = q^2$$

$$H = 2pq$$

■ 复等位基因的平衡

➤ 已知基因型频率如何计算基因频率？

以三个等位基因 A_1 、 A_2 和 A_3 为例：

	等位基因			基 因 型					
基因符号	A_1	A_2	A_3	A_1A_1	A_1A_2	A_1A_3	A_2A_2	A_2A_3	A_3A_3
频率符号	p_1	p_2	p_3	P_{11}	P_{12}	P_{13}	P_{22}	P_{23}	P_{33}
个体数	-	-	-	n_{11}	n_{12}	n_{13}	n_{22}	n_{23}	n_{33}

N

➤ 已知基因型频率如何计算基因频率？

	等位基因			基 因 型					
基因符号	A ₁	A ₂	A ₃	A ₁ A ₁	A ₁ A ₂	A ₁ A ₃	A ₂ A ₂	A ₂ A ₃	A ₃ A ₃
频率符号	<i>p</i> ₁	<i>p</i> ₂	<i>p</i> ₃	<i>P</i> ₁₁	<i>P</i> ₁₂	<i>P</i> ₁₃	<i>P</i> ₂₂	<i>P</i> ₂₃	<i>P</i> ₃₃
个体数	—	—	—	n ₁₁	n ₁₂	n ₁₃	n ₂₂	n ₂₃	n ₃₃
基因型频率	—	—	—	$\frac{n_{11}}{N}$	$\frac{n_{12}}{N}$	$\frac{n_{13}}{N}$	$\frac{n_{22}}{N}$	$\frac{n_{23}}{N}$	$\frac{n_{33}}{N}$

已知

$$p_1 = \frac{2n_{11} + n_{12} + n_{13}}{2N} = P_{11} + \frac{1}{2}(P_{12} + P_{13})$$

$$p_2 = \frac{2n_{22} + n_{12} + n_{23}}{2N} = P_{22} + \frac{1}{2}(P_{12} + P_{23})$$

$$p_3 = \frac{2n_{33} + n_{13} + n_{23}}{2N} = P_{33} + \frac{1}{2}(P_{13} + P_{23})$$

以三个等位基因为例

由基因型频率推导基因频率：

$$p_1 = P_{11} + \frac{1}{2} (P_{12} + P_{13})$$

$$p_2 = P_{22} + \frac{1}{2} (P_{12} + P_{23})$$

$$p_3 = P_{33} + \frac{1}{2} (P_{13} + P_{23})$$

已知基因型频率计算基因频率

某个等位基因的基因频率等于该等位基因处于纯合状态时的基因型频率加上该等位基因处于杂合状态时基因型频率之和的二分之一；

平衡与非平衡群体

练习题1:

在某一基因座上共有3个等位基因 A_1 、 A_2 和 A_3 ，
若基因型频率 A_1A_1 为0.12、 A_2A_2 为0.3、 A_3A_3 为0.18、
 A_1A_2 为0.14、 A_1A_3 为0.1、 A_2A_3 为0.16，则 A_2 的基因
频率预期为（0.45）。

$$\begin{aligned} p_2 &= P_{22} + \frac{1}{2} (P_{12} + P_{23}) \\ &= 0.3 + \frac{1}{2} \cdot (0.14 + 0.16) \\ &= 0.45 \end{aligned}$$

➤ 已知基因频率如何计算基因型频率？

①. AA 16%, Aa 48%, aa 36%

← 平衡群体

②. AA 12%, Aa 56%, aa 32%

③. AA 5%, Aa 70%, aa 25%

④. AA 10%, Aa 60%, aa 30%

→ { A: $p = 0.4$
a: $q = 0.6$

•
•
•

以三个等位基因 A_1 、 A_2 和 A_3 为例：

	等位基因			基 因 型					
基因符号	A_1	A_2	A_3	A_1A_1	A_1A_2	A_1A_3	A_2A_2	A_2A_3	A_3A_3
频率符号	p_1	p_2	p_3	P_{11}	P_{12}	P_{13}	P_{22}	P_{23}	P_{33}

经历一次随机交配



	$A_1 (p_1)$	$A_2 (p_2)$	$A_3 (p_3)$
$A_1 (p_1)$	A_1A_1 (p_1^2)	A_1A_2 (p_1p_2)	A_1A_3 (p_1p_3)
$A_2 (p_2)$	A_1A_2 (p_1p_2)	A_2A_2 (p_2^2)	A_2A_3 (p_2p_3)
$A_3 (p_3)$	A_1A_3 (p_1p_3)	A_2A_3 (p_2p_3)	A_3A_3 (p_3^2)

$$A_1A_1 : P_{11} = p_1^2$$

$$A_1A_2 : P_{12} = 2p_1p_2$$

$$A_2A_2 : P_{22} = p_2^2$$

$$A_1A_3 : P_{13} = 2p_1p_3$$

$$A_3A_3 : P_{33} = p_3^2$$

$$A_2A_3 : P_{23} = 2p_2p_3$$

$$p_1' = p_1^2 + \frac{1}{2} \cdot (2p_1p_2 + 2p_1p_3) = p_1(p_1 + p_2 + p_3) = p_1$$

$$p_2' = p_2^2 + \frac{1}{2} \cdot (2p_1p_2 + 2p_2p_3) = p_2(p_1 + p_2 + p_3) = p_2$$

$$p_3' = p_3^2 + \frac{1}{2} \cdot (2p_1p_3 + 2p_2p_3) = p_3(p_1 + p_2 + p_3) = p_3$$

↓ 随机交配

	$A_1 (p_1)$	$A_2 (p_2)$	$A_3 (p_3)$
$A_1 (p_1)$	A_1A_1 (p_1^2)	A_1A_2 (p_1p_2)	A_1A_3 (p_1p_3)
$A_2 (p_2)$	A_1A_2 (p_1p_2)	A_2A_2 (p_2^2)	A_2A_3 (p_2p_3)
$A_3 (p_3)$	A_1A_3 (p_1p_3)	A_2A_3 (p_2p_3)	A_3A_3 (p_3^2)

$$A_1A_1 : P_{11} = p_1^2$$

$$A_2A_2 : P_{22} = p_2^2$$

$$A_3A_3 : P_{33} = p_3^2$$

$$A_1A_2 : P_{12} = 2p_1p_2$$

$$A_1A_3 : P_{13} = 2p_1p_3$$

$$A_2A_3 : P_{23} = 2p_2p_3$$

已知基因频率计算基因型频率

平衡群体

纯合基因型的基因型频率等于组成该基因型的等位基因的基因频率乘积，若为杂合基因型则再乘2；

练习题2:

有一个平衡群体，在某一基因座上共有4个等位基因 A_1 、 A_2 、 A_3 和 A_4 ，基因频率分别为0.1、0.2、0.3和0.4，则基因型为 A_1A_2 和 A_3A_3 的个体所占的比例预期分别为 0.04 和 0.09。

$$A_1A_2 : P_{12} = 2p_1p_2 = 2 \cdot 0.1 \cdot 0.2 = 0.04$$

$$A_3A_3 : P_{33} = p_3^2 = 0.3^2 = 0.09$$

■ ABO血型系统基因频率的计算

例如： 在1000人的群体中，ABO血型A型为450人、B型为130人、AB型为60人、O型为360人。若该群体处于平衡，试计算 I^A 、 I^B 和 i 的基因频率？

设： I^A 、 I^B 、 i 的基因频率分别为 p 、 q 和 r ，且群体平衡，则：

表型	实测表型频率	基因型	基因型频率	理论表型频率
A型	A	$I^A I^A$	p^2	$p^2 + 2pr$
		$I^A i$	$2pr$	
B型	B	$I^B I^B$	q^2	$q^2 + 2qr$
		$I^B i$	$2qr$	
AB型	AB	$I^A I^B$	$2pq$	$2pq$
O型	O	ii	r^2	r^2

$$r = \sqrt{O}$$

$$p = 1 - (q + r) = 1 - \sqrt{(q^2 + 2qr + r^2)} = 1 - \sqrt{B + O}$$

$$q = 1 - (p + r) = 1 - \sqrt{(p^2 + 2pr + r^2)} = 1 - \sqrt{A + O}$$

设： I^A 、 I^B 、 i 的基因频率分别为 p 、 q 和 r ，且群体平衡，则：

表型	实测表型频率	基因型	基因型频率	理论表型频率
A型	0.45	$I^A I^A$	p^2	$p^2 + 2pr$
		$I^A i$	$2pr$	
B型	0.13	$I^B I^B$	q^2	$q^2 + 2qr$
		$I^B i$	$2qr$	
AB型	0.06	$I^A I^B$	$2pq$	$2pq$
O型	0.36	ii	r^2	r^2

$$r = \sqrt{O} = \sqrt{0.36} = 0.6$$

$$p = 1 - \sqrt{B + O} = 1 - \sqrt{0.13 + 0.36} = 0.3$$

$$q = 1 - \sqrt{A + O} = 1 - \sqrt{0.45 + 0.36} = 0.1$$

■ 伴性基因的平衡

方式：不平衡群体经过多个世代的随机交配，
以振荡的方式快速地接近平衡；

推导：

设：将亲代群体按性别分为雌性群体和雄性群体，雌性群体A和a基因频率分别为 p_{xx} 和 q_{xx} ($p_{xx} + q_{xx} = 1$)，雄性群体A和a基因频率分别为 p_x 和 q_x ($p_x + q_x = 1$)。

亲代 群体	基因	雌性	雄性	雌雄频率差异
	A	p_{xx}	p_x	$p_x - p_{xx} = d$
	a	q_{xx}	q_x	$q_x - q_{xx} = d'$

则： 下一代基因型频率和基因频率推导如下：

子代 雌性 群体	雄配子	雌配子	
		$X^A (p_{xx})$	$X^a (q_{xx})$
	$X^A (p_x)$	$X^A X^A (p_x p_{xx})$	$X^A X^a (p_x q_{xx})$
	$X^a (q_x)$	$X^A X^a (q_x p_{xx})$	$X^a X^a (q_x q_{xx})$

子代 雄性 群体	雄配子	雌配子	
		$X^A (p_{xx})$	$X^a (q_{xx})$
	$Y(1)$	$X^A Y (p_{xx})$	$X^a Y (q_{xx})$

子代 雌性 群体	雄配子	雌配子	
		$X^A (p_{xx})$	$X^a (q_{xx})$
	$X^A (p_x)$	$X^A X^A (p_x p_{xx})$	$X^A X^a (p_x q_{xx})$
	$X^a (q_x)$	$X^A X^a (q_x p_{xx})$	$X^a X^a (q_x q_{xx})$

$$\begin{aligned}
 \text{A: } p_{1xx} &= p_x p_{xx} + \frac{1}{2} p_x q_{xx} + \frac{1}{2} p_{xx} q_x \\
 &= p_x p_{xx} + \frac{1}{2} [p_x (1 - p_{xx}) + p_{xx} (1 - p_x)] \\
 &= \frac{1}{2} (p_x + p_{xx})
 \end{aligned}$$

$$\begin{aligned}
 \text{a: } q_{1xx} &= q_x q_{xx} + \frac{1}{2} p_x q_{xx} + \frac{1}{2} p_{xx} q_x \\
 &= q_x q_{xx} + \frac{1}{2} [q_{xx} (1 - q_x) + q_x (1 - p_{xx})] \\
 &= \frac{1}{2} (q_x + q_{xx})
 \end{aligned}$$

子代 雄性 群体	雄配子	雌配子	
		$X^A (p_{xx})$	$X^a (q_{xx})$
	$Y(1)$	$X^A Y (p_{xx})$	$X^a Y (q_{xx})$

$$A: p_{1x} = p_{xx}$$

$$a: q_{1x} = q_{xx}$$



某一代雄性群体的基因(型)频率等于上一代雌性群体的基因频率。

经历一代的随机交配后



子代 群体	基因	雌性	雄性	雌雄频率差异
	A	$\frac{1}{2} (p_x + p_{xx})$	p_{xx}	$\frac{1}{2} (p_{xx} - p_x) = -\frac{1}{2}d$
	a	$\frac{1}{2} (q_x + q_{xx})$	q_{xx}	$\frac{1}{2} (q_{xx} - q_x) = -\frac{1}{2}d'$

例如：X染色体上A等位基因在雌雄群体的基因频率分别为 0.1和0.7，则：

	♀	♂	$d (\text{♂} - \text{♀})$
F_0	0.1	0.7	0.6
F_1	0.4	0.1	-0.3
F_2	0.25	0.4	0.15
F_3	0.325	0.25	-0.075
F_4	0.2875	0.325	0.0375
F_5	0.30625	0.2875	-0.01875
	⋮		

	个体数	等位基因数	A 频率	a 频率	A 数	a 数
雌性群体	N	2N	p_{xx}	q_{xx}	$2Np_{xx}$	$2Nq_{xx}$
雄性群体	N	N	p_x	q_x	Np_x	Nq_x
整个群体	2N	3N	\bar{p}	\bar{q}	$2Np_{xx} + Np_x$	$2Nq_{xx} + Nq_x$

亲代: A: $\bar{p} = \frac{1}{3}(2p_{xx} + p_x)$ a: $\bar{q} = \frac{1}{3}(2q_{xx} + q_x)$

下一代群体：

	基因	雌性	雄性
子代群体	A	$\frac{1}{2}(p_x + p_{xx})$	p_{xx}
	a	$\frac{1}{2}(q_x + q_{xx})$	q_{xx}

F_1 代: $A: \bar{p} = \frac{1}{3}(2p_{xx} + p_x)$ $a: \bar{q} = \frac{1}{3}(2q_{xx} + q_x)$

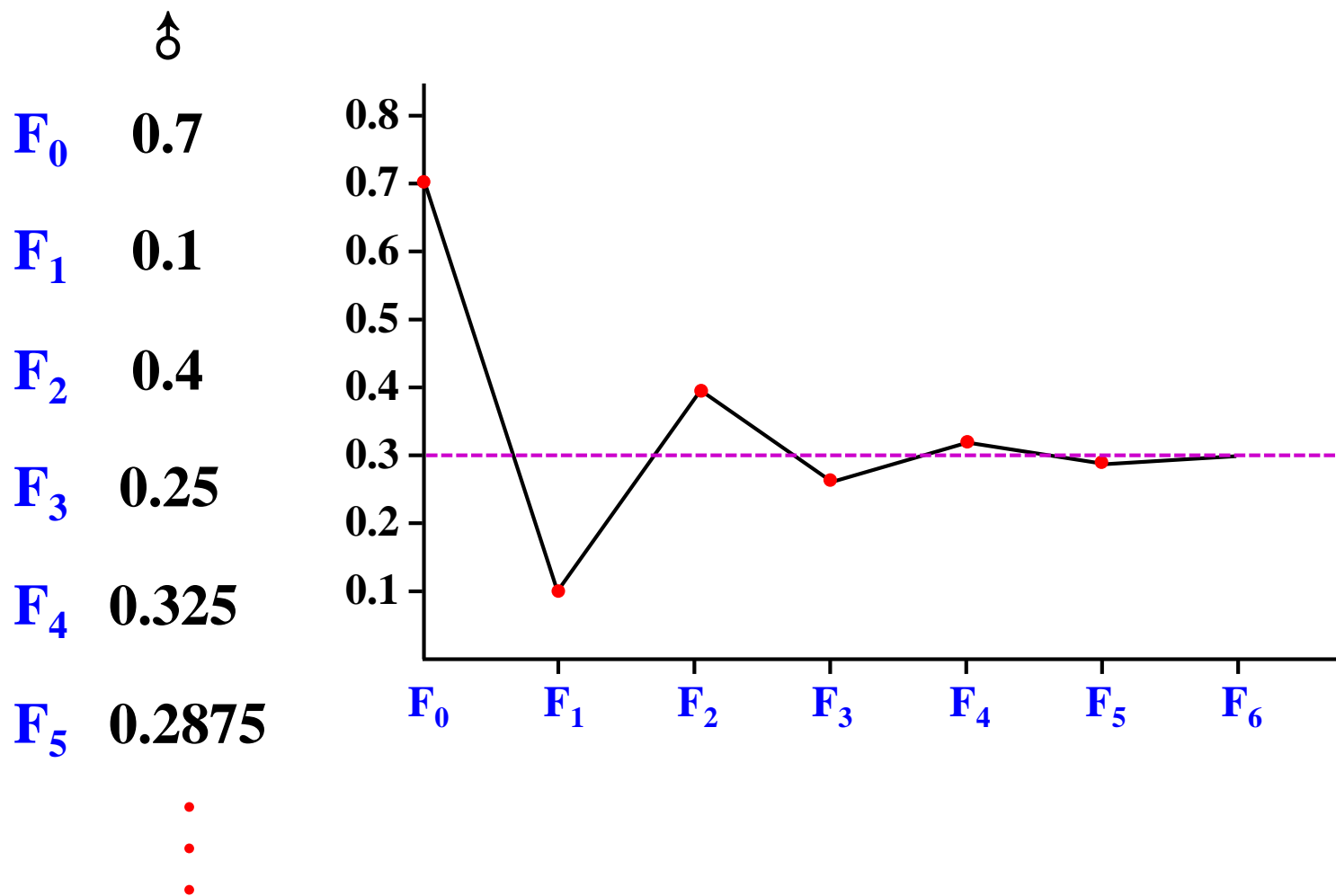
例如：A等位基因在雌雄群体的基因频率分别为0.1和0.7，则：

	♀	♂	$d (\text{♂} - \text{♀})$	$\bar{p} = \frac{1}{3} (2p_{xx} + p_x)$
F_0	0.1	0.7	0.6	0.3
F_1	0.4	0.1	-0.3	0.3
F_2	0.25	0.4	0.15	0.3
F_3	0.325	0.25	-0.075	0.3
F_4	0.2875	0.325	0.0375	0.3
F_5	0.30625	0.2875	-0.01875	0.3
	⋮			⋮

例如：A等位基因在雌雄群体的基因频率分别为0.1和0.7，则：

	♀	♂	$d (\text{♂} - \text{♀})$	$\bar{p} = \frac{1}{3} (2p_{xx} + p_x)$
F_0	0.1	0.7	0.6	0.3
F_1	0.4	0.1	-0.3	0.3
F_2	0.25	0.4	0.15	0.3
F_3	0.325	0.25	-0.075	0.3
F_4	0.2875	0.325	0.0375	0.3
F_5	0.30625	0.2875	-0.01875	0.3
	⋮	⋮		⋮

例如：A等位基因在雌雄群体的基因频率分别为0.1和0.7，则：



总结

X染色体上等位基因由不平衡到平衡的方式和规律：

方式：不平衡群体经过多个世代的随机交配，以振荡的方式快速地接近平衡；

规律：

1. 每随机交配一次，雌雄群体基因频率的差值减少一半；
2. 在建立平衡的过程中，雌雄两性群体中的基因频率随着随机交配世代的增加而交互递减；

总结

判定在X染色体上等位基因达到平衡的标准：

在随机交配条件下，下列情况达到了H-W平衡：

①雌雄群体基因频率相等；

②	雄性群体		雌性群体		
	X^A	X^a	$X^A X^A$	$X^A X^a$	$X^a X^a$
	p	q	p^2	$2pq$	q^2

对于遗传病：

伴X隐性遗传：

$$\text{男性发病率:女性发病率} = q : q^2 = 1 : q$$

伴X显性遗传：

$$\text{男性发病率:女性发病率} = p : (p^2 + 2pq) = 1 : (1 + q)$$

■ 如何判断群体是否达到H-W平衡

思路：

实际个体数 → 实际基因型频率 → 实际基因频率 → 假设群体平衡 → 理论基因型频率 → 理论个体数 → 卡方检验

例如：某血型调查报道，在420人中，M型(MM)137人，MN型(MN)196人、N型(NN)87人。在这个群体中的血型分布是否达到了H-W平衡？

令M和N的基因频率分别为 p 和 q ，基因频率如下：

$$\text{M: } p = 137/420 + \frac{1}{2} \times 196/420 = 0.56$$

$$\text{N: } q = 87/420 + \frac{1}{2} \times 196/420 = 0.44$$

假设群体平衡，在这样的基因频率下，3种基因型预期频率与血型的预期人数为：

M型： $p^2 = 0.56^2 = 0.3136$ ， $420 \times 0.314 = 132$ 人

MN型： $2pq = 2 \times 0.56 \times 0.44 = 0.4928$ ， $420 \times 0.493 = 207$ 人

N型： $q^2 = 0.44^2 = 0.1936$ ， $420 \times 0.194 = 81$ 人

进行 χ^2 测验时可利用以下公式（ O 是实测值， E 是理论值， Σ 是总和），即：

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

χ^2 检验： $\chi^2 = \sum (O-E)^2/E = (137-132)^2/132 + (196-207)^2/207$
 $+ (87-81)^2/81 = 1.215$

$df = 3-1-1 = 1$ ，查表得： $p > 0.05$ ，差异不显著，可以认为该血型分布符合H-W平衡。

χ^2 表

$P \backslash df$	0.99	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01
1	0.00016	0.004	0.016	0.064	0.148	0.455	1.074	1.642	2.706	3.841	6.635
2	0.0201	0.103	0.211	0.446	0.713	1.386	2.408	3.219	4.605	5.991	9.210
3	0.115	0.352	0.584	1.005	1.424	2.366	3.665	4.642	6.251	7.815	11.345
4	0.297	0.711	1.064	1.649	2.195	3.357	4.878	5.989	7.779	9.488	13.277
5	0.554	1.145	1.610	2.343	3.000	4.351	6.064	7.269	9.236	11.070	15.086
6	0.872	1.635	2.204	3.070	3.828	5.345	7.231	8.588	10.645	12.592	16.812
7	1.239	2.167	2.833	3.822	4.671	6.346	8.783	9.803	12.017	14.067	18.475
8	1.646	2.733	3.490	4.594	5.527	7.344	9.524	11.030	13.362	15.507	20.090
9	2.088	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	21.666
10	2.558	3.940	4.865	6.179	7.627	9.342	11.781	13.442	15.987	18.307	23.209

表内数字是各种 χ^2 值， df 为自由度， P 是在一定自由度下 χ^2 大于表中数值的概率。

第二节

影响哈迪-温伯格平衡的因素



■ 若仅存在正向突变

设：基因 $A \rightarrow a$ 为正向突变，每代突变率为 u

($u=x/n$, x 是一代中 A 突变为 a 基因数目, n 为 A 基因总数)

突变前： A 频率： p_0

突变后： A 频率： $p_0 (1 - u)$

$$p_1 = p_0 (1 - u)$$

$$p_2 = p_1 (1 - u) = p_0 (1 - u) (1 - u) = p_0 (1 - u)^2$$

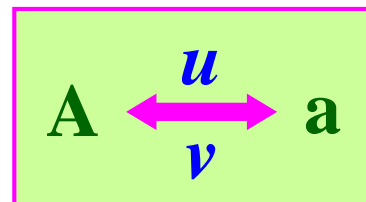
$$p_n = p_0 (1 - u)^n \quad (n \rightarrow \infty, p_n \rightarrow 0)$$

□ 突变压（mutation pressure）：假设在常染色体上有一对等位基因A和a，若A在长时间内不断突变为a，没有其它因素的干扰，最终这个群体中的A将全部消失，成为a的纯合群体，这就是突变对遗传结构产生的压力，称为突变压。

■ 若同时存在正向突变和回复突变

设：正向突变 ($A \rightarrow a$) 的频率为 u

回复突变 ($a \rightarrow A$) 的频率为 v



则： p 每代变化量 $qv - pu$

q 每代变化量 $pu - qv$

如果： $pu < qv$ 则群体中 A 增加

$pu > qv$ 则群体中 a 增加

$pu = qv$ 群体中基因频率保持不变(达到平衡)

平衡时: $pu = qv$

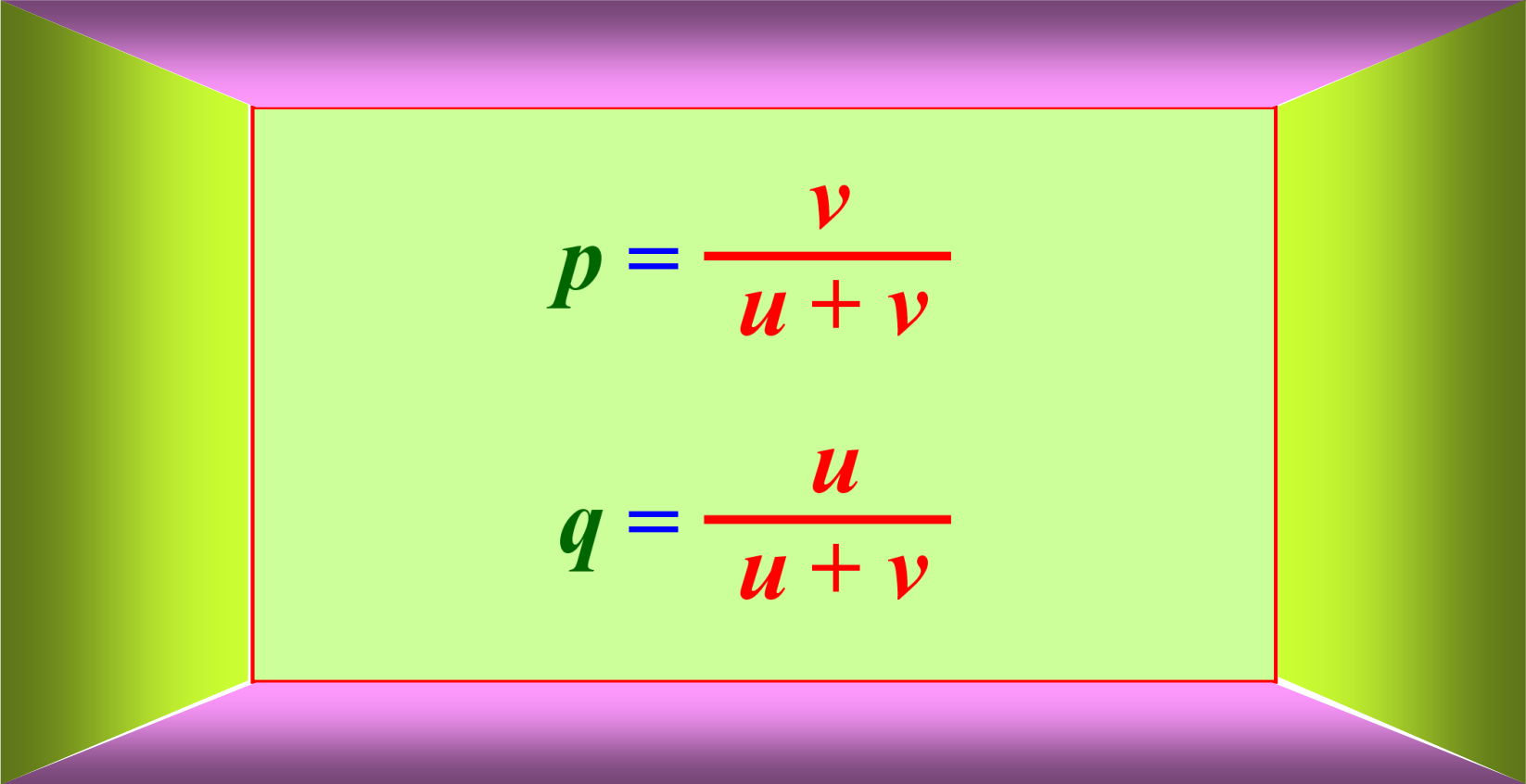
$$pu = (1 - p)v$$

$$pu = v - pv$$

$$v = pu + pv = p(u + v)$$

因此 $p = \frac{v}{u + v}$

同理 $q = \frac{u}{u + v}$


$$p = \frac{v}{u + v}$$

$$q = \frac{u}{u + v}$$



■ 适应值与选择系数

□ 适应值(adaptive value): 也称达尔文适合度, 指一种已知基因型的个体, 将它的基因传递给后代的**相对能力**, 用 ω 表示;

① ω 是一个相对值，将具有最高生殖效能的基因型个体的适应值定为1，其它基因型个体适应值是与之相比的比值；

基因型	AA	Aa	aa	总数
当代个体数	40	50	10	100
产生的可育子代数	80	90	10	180
平均可育子代数	$80/40=2$	$90/50=1.8$	$10/10=1$	
适应值 ω	1	0.9	0.5	

②适应值是一个相对概念，与环境因素有关；

例如：椒花蛾



②适应值是一个相对概念，与环境因素有关；

例如：椒花蛾

椒花蛾在污染区时：

	黑色	浅色
基因型	DD, Dd	dd
释放量a	154	64
回收量b	82	16
生存率b/a	0.53	0.25
适应值 ω	$0.53/0.53 = 1$	$0.25/0.53 = 0.47$



②适应值是一个相对概念，与环境因素有关；

例如：椒花蛾

椒花蛾在非污染区时：

	黑色	浅色
基因型	DD, Dd	dd
释放量a	406	393
回收量b	19	54
生存率b/a	0.047	0.137
适应值 ω	$0.047/0.137 = 0.343$	$0.137/0.137 = 1$



□ **选择系数(selective coefficient):** 一种基因型的个体在群体中不利于生存的程度, 用S表示且有 $S=1-\omega$, 显然 $\omega=1$ 时, $S=0$; $\omega=0$ 时, $S=1$.

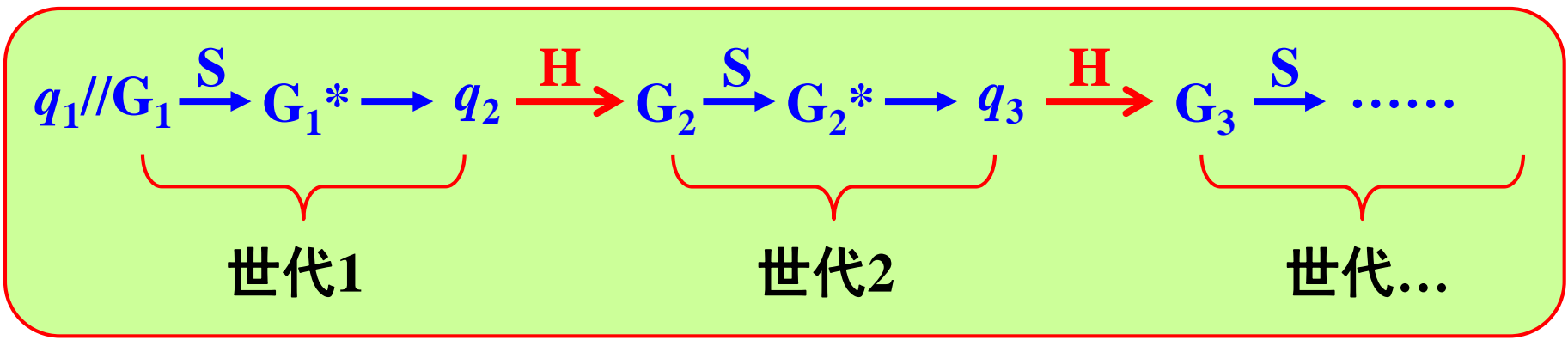
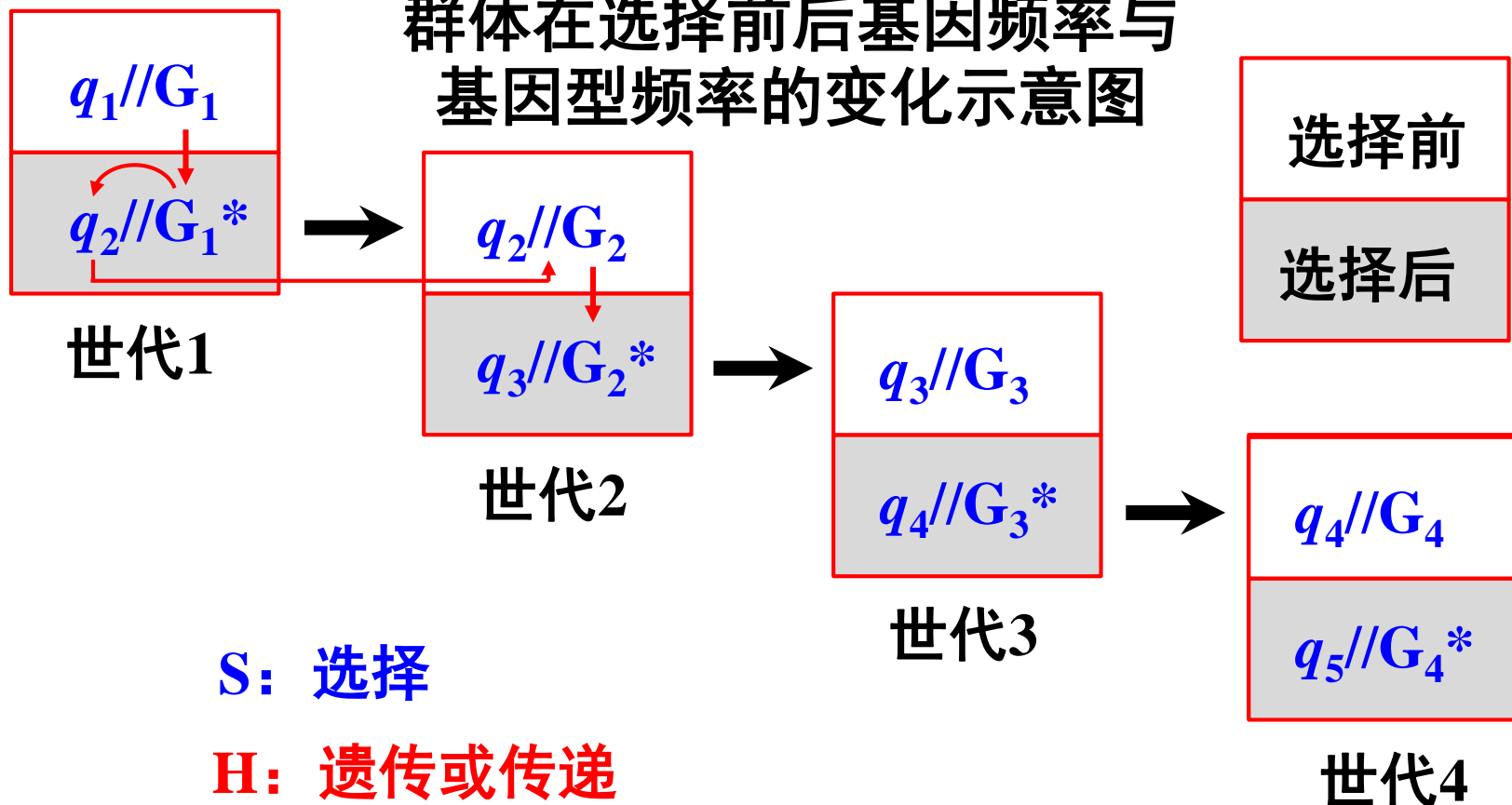
在污染区时

	黑色	浅色
基因型	DD, Dd	dd
释放量a	154	64
回收量b	82	16
生存率b/a	0.53	0.25
适应值 ω	$0.53/0.53 = 1$	$0.25/0.53 = 0.47$
选择系数S	0	0.53

在非污染区时

	黑色	浅色
基因型	DD, Dd	dd
释放量a	406	393
回收量b	19	54
生存率b/a	0.047	0.137
适应值 ω	$0.047/0.137 = 0.343$	$0.137/0.137 = 1$
选择系数S	0.657	0

群体在选择前后基因频率与基因型频率的变化示意图



■ 选择对H-W平衡的影响

1. 对隐性纯合体的不完全选择，即 $0 < S < 1$.

设：在一个平衡群体中，一对等位基因A和a，频率

分别为 p 、 q ，隐性纯合体的选择系数为 S ；

则：基因型频率为：

$$AA: D = p^2, \quad Aa: H = 2pq, \quad aa: R = q^2$$

且：而aa的 $\omega=1-S$ ，AA 和 Aa 的 $\omega=1$ ，则选择前后基因型的频率分别为：

基因型	AA	Aa	aa	全群体
选择前	$N \cdot p^2$	$N \cdot 2pq$	$N \cdot q^2$	N
适应值	1	1	1-S	-
选择后	$N \cdot p^2$	$N \cdot 2pq$	$N \cdot q^2(1-S)$	$N \cdot (1-Sq^2)$
相对频率	$\frac{N \cdot p^2}{N \cdot (1-Sq^2)}$	$\frac{N \cdot 2pq}{N \cdot (1-Sq^2)}$	$\frac{N \cdot q^2(1-S)}{N \cdot (1-Sq^2)}$	1

基因型	AA	Aa	aa	全群体
起始频率	p^2	$2pq$	q^2	1
适应值	1	1	1-S	-
选择后	p^2	$2pq$	$q^2(1-S)$	$1-Sq^2$
相对频率	$\frac{p^2}{1-Sq^2}$	$\frac{2pq}{1-Sq^2}$	$\frac{q^2(1-S)}{1-Sq^2}$	1

D

H

R

经过选择后的基因频率和下一代选择前的基因频率为：

$$\text{A: } p_1 = \frac{p^2 + pq}{1 - Sq^2} = \frac{p}{1 - Sq^2}$$

$$\text{a: } q_1 = \frac{[q^2(1 - S) + pq]}{1 - Sq^2} = \frac{q(1 - Sq)}{1 - Sq^2}$$

a基因频率的改变量：

$$\Delta q = q_1 - q = \frac{q(1 - Sq)}{1 - Sq^2} - q = \frac{-Sq^2(1 - q)}{1 - Sq^2}$$

q 很小时， $1 - Sq^2 \approx 1$ ，则 $\Delta q \approx -Sq^2(1 - q)$

不同 Δq 和不同S的选择效率

Δq		S			
q_0	q_n	1	0.5	0.1	0.01
0.99	0.75	1	8	38	382
0.75	0.50	1	3	18	176
0.50	0.25	2	6	31	310
0.25	0.10	6	14	71	710
0.10	0.01	90	185	924	9240
0.01	0.001	900	1805	9023	90231
0.001	0.0001	9000	18005	90023	900230

$\Delta q \approx -Sq^2(1-q)$, 相同的 $q_0 - q_n$ 区间, 当 $q = 2/3$ 时 Δq 有最大值

2. 对隐性纯合体的完全选择，即S=1:

基因型	AA	Aa	aa	全群体
起始频率	p^2	$2pq$	q^2	1
适应值	1	1	0	-
选择后	p^2	$2pq$	0	$p^2 + 2pq$
相对频率	$\frac{p^2}{p^2 + 2pq}$	$\frac{2pq}{p^2 + 2pq}$	0	1

经过选择后的基因频率和下一代选择前的基因频率为：

$$\text{A: } p_1 = \frac{p^2 + pq}{p^2 + 2pq} = \frac{1}{1 + q}$$

$$\text{a: } q_1 = \frac{pq}{p^2 + 2pq} = \frac{q}{1 + q}$$

对隐生纯合体的不完全选择：

$$\text{A: } p_1 = \frac{p^2 + pq}{1 - Sq^2} = \frac{p}{1 - Sq^2}$$

$$\text{a: } q_1 = \frac{[q^2(1 - S) + pq]}{1 - Sq^2} = \frac{q(1 - Sq)}{1 - Sq^2}$$

若S=1

经过选择后的基因频率和下一代选择前的基因频率为：

$$\text{A: } p_1 = \frac{p^2 + pq}{p^2 + 2pq} = \frac{1}{1 + q}$$

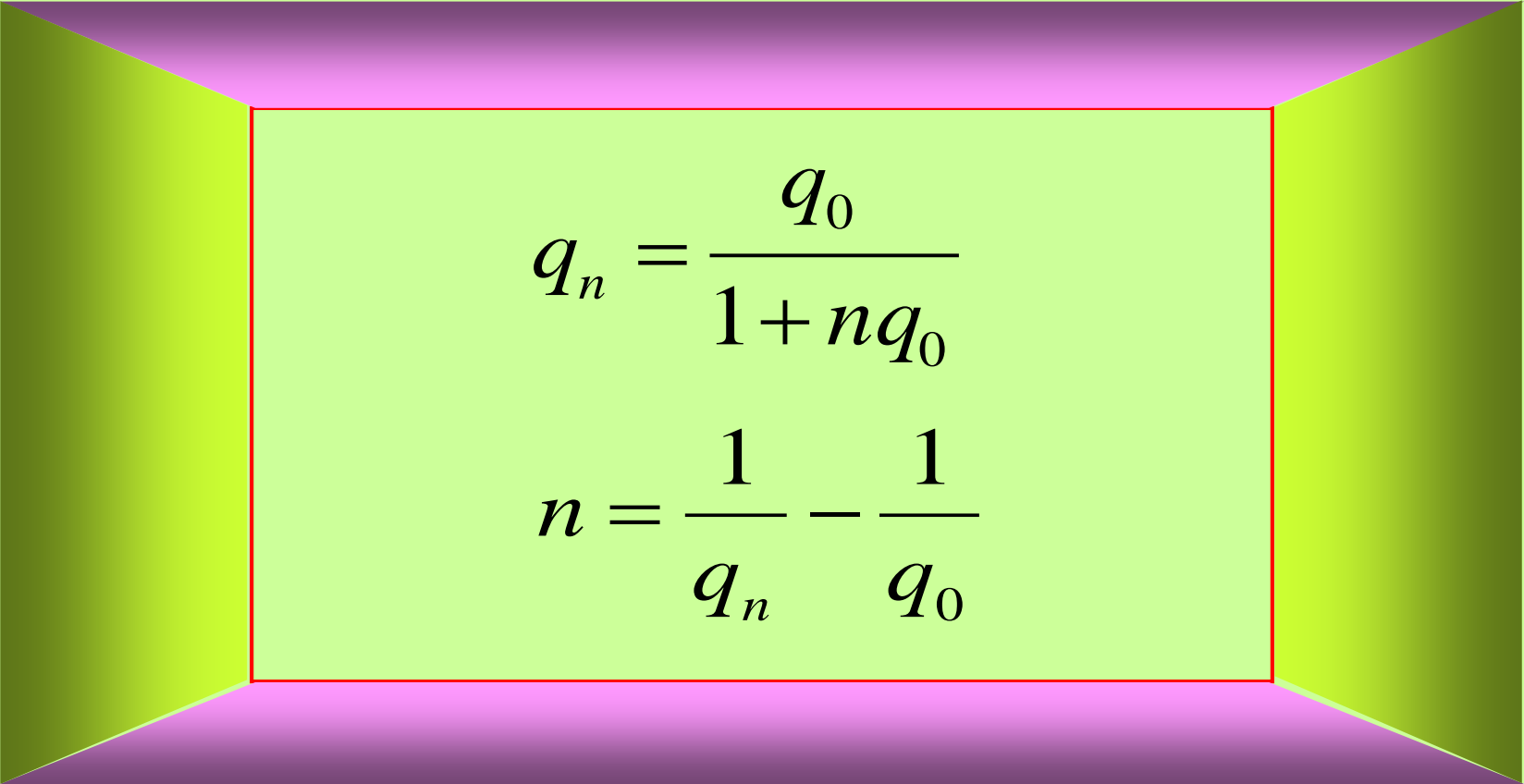
$$\text{a: } q_1 = \frac{pq}{p^2 + 2pq} = \frac{q}{1 + q}$$

$$\mathbf{F}_1: \quad q_1 = \frac{q_0}{1 + q_0}$$

$$\mathbf{F}_2: \quad q_2 = \frac{q_1}{1 + q_1} = \frac{\frac{q_0}{1 + q_0}}{1 + \frac{q_0}{1 + q_0}} = \frac{q_0}{1 + 2q_0}$$

$$\mathbf{F}_3: \quad q_3 = \frac{q_2}{1 + q_2} = \frac{\frac{q_0}{1 + 2q_0}}{1 + \frac{q_0}{1 + 2q_0}} = \frac{q_0}{1 + 3q_0}$$

$$\mathbf{F}_n: \quad q_n = \frac{q_0}{1 + nq_0} \longrightarrow \mathbf{n} = \frac{1}{q_n} - \frac{1}{q_0}$$


$$q_n = \frac{q_0}{1 + nq_0}$$

$$n = \frac{1}{q_n} - \frac{1}{q_0}$$

举例：已知人类白化病等位基因的频率为0.01，若白化纯合体不育（ $S=1$ ），要将此基因的频率分别降至0.001和0.0001所需的世代数。

利用 $n = \frac{1}{q_n} - \frac{1}{q_0}$ 计算：

$$n_1 = \frac{1}{0.001} - \frac{1}{0.01} = 900$$

$$n_2 = \frac{1}{0.0001} - \frac{1}{0.01} = 9900$$

3. 对显性体的选择

基因型	AA	Aa	aa	全群体
选择前	p^2	$2pq$	q^2	1
适应值	$1 - S$	$1 - S$	1	-
选择后	$p^2 (1 - S)$	$2pq (1 - S)$	q^2	$1 - Sp(2 - p)$
相对频率	$\frac{p^2 (1 - S)}{1 - Sp(2 - p)}$	$\frac{2pq (1 - S)}{1 - Sp(2 - p)}$	$\frac{q^2}{1 - Sp(2 - p)}$	1

D

H

R

$$p_1 = D + \frac{1}{2}H = \frac{p - Sp}{1 - Sp(2 - p)}$$

$$\Delta p = p_1 - p = \frac{p - Sp}{1 - Sp(2 - p)} - p = \frac{-Sp(1 - p)^2}{1 - Sp(2 - p)}$$

当 p 或 S 很小时，分母约为1， $\Delta p \approx -Sp(1-p)^2$



■ 选择隐性纯合体

设：一对等位基因A和a(频率分别为 p , q),
正向突变率 u , 回复突变率 v , 同时选择系数
S作用于aa

则：突变使a的改变量为： $\Delta q_m = pu - qv = u(1-q) - qv$

选择使a的改变量为： $\Delta q_s = \frac{-Sq^2(1-q)}{1-Sq^2} \approx -Sq^2(1-q)$

(注： q 很小时作近似处理，即 $1-Sq^2 \approx 1$)

平衡时： $\Delta q_m + \Delta q_s = 0$ ，即 $u(1-q) - qv = Sq^2(1-q)$

若 v 也很小， $vq \approx 0$ ，则 $u \approx Sq^2$ ， $q = \sqrt{\frac{u}{s}}$

可用于推算自发突变的频率 u 或平衡时 q

q 和 v 很小时: $q = \sqrt{\frac{u}{s}}$

可用于推算

自发突变的频率 u 或平衡时 q

举例：

人类全色盲为常染色体隐性遗传，纯合体的比例为8万分之1，且患者的子女为正常人的一半，求正向突变率。

解： $q^2 = 1/80000$ ， $S = 0.5$ ， 则 $u = Sq^2 = 6.25 \times 10^{-6}$

2. 选择的是显性纯合体：同样 $v = Sp^2$ ；

3. 选择显性体（AA和Aa）

选择 $\Delta p_s = -Sp(1-p)^2$ （当 p 或 S 很小时）；

突变 $\Delta p_m = qv - pu = v(1-p) - pu$ ；

平衡时： $\Delta q_m + \Delta q_s = 0$ ，即 $Sp(1-p)^2 = v(1-p) - pu$

当 p 和 u 很小时， $1-p \approx 1$ ， $S+u \approx S$ ，所以 $v = Sp$

例如：人类侏儒即软骨发育不全(AA, Aa)，适应值 $\omega=0.2$ ，正常a突变为A的频率 $\nu = 5 \times 10^{-5}$ ，求A基因的频率。

解： $S = 1 - \omega = 0.8$ ， $\nu = 5 \times 10^{-5}$ ，

则： $p = \nu/S = 6.25 \times 10^{-5}$



□ **迁移(migration)**: 指个体从一个群体迁入另一个群体或从一个群体迁出，然后参与交配繁殖，导致群体间的基因流动。

设: 一个群体中，一对等位基因A和a的基因频率为 p_0 和 q_0



若迁入者A和a基因频率为 p_0' 和 q_0' ，那么在当代迁移后或下一代未发生迁移之前，a的基因频率为：

设：群体个体数为N，则迁入和迁出个体数为mN

	目标群体	迁出	迁入
个体数	N	mN	mN
等位基因数	2N	2mN	2mN
a等位基因数	$2Nq_0$	$2mNq_0$	$2mNq_0'$
迁入迁出后a等位基因数	$2Nq_0 - 2mNq_0 + 2mNq_0'$		

$$q_1 = q_0 - mq_0 + mq_0' = mq_0' + (1 - m) q_0$$

$$q_1 = q_0 - mq_0 + mq_0' = \boxed{mq_0' + (1 - m) q_0}$$

$$q_2 = mq_0' + (1 - m) q_1 = mq_0' + (1 - m) [mq_0' + (1 - m) q_0]$$

$$= \boxed{mq_0' + (1 - m) mq_0' + (1 - m)^2 q_0}$$

$$q_3 = mq_0' + (1 - m) q_2$$

$$= mq_0' + (1 - m) [mq_0' + m(1 - m) q_0' + (1 - m)^2 q_0]$$

$$= \boxed{mq_0' + (1 - m) mq_0' + (1 - m)^2 mq_0' + (1 - m)^3 q_0}$$

$$q_4 = mq_0' + (1 - m) q_3$$

$$= mq_0' + (1 - m) [mq_0' + m(1 - m) q_0' + m(1 - m)^2 q_0' + (1 - m)^3 q_0]$$

$$= \boxed{mq_0' + (1 - m) mq_0' + (1 - m)^2 mq_0' + (1 - m)^3 mq_0' + (1 - m)^4 q_0}$$

可以类推：

$$q_n = mq_0' + (1-m)mq_0' + (1-m)^2 mq_0' + \cdots + (1-m)^{n-1} mq_0' + (1-m)^n q_0$$

$$= mq_0' [1 + (1-m) + (1-m)^2 + \cdots + (1-m)^{n-1}] + (1-m)^n q_0$$

$$= mq_0' \frac{1-(1-m)^n}{1-(1-m)} + (1-m)^n q_0$$

$$= q_0' - q_0'(1-m)^n + (1-m)^n q_0$$

$$= q_0' + (1-m)^n(q_0 - q_0')$$

连续n代迁移且m不变:

$$q_n = q_0' + (1 - m)^n (q_0 - q_0')$$

$$(1 - m)^n = (q_n - q_0') / (q_0 - q_0')$$

例：两个群体X、Y，群体X中的A基因频率 $p_0' = 0.028$ ，群体Y中A基因频率 $p_0 = 0.630$ ，X群体向Y群体发生基因迁移10代后，Y群体的 $p_n = 0.446$ 。求A基因迁移的频率。

解： $(1-m)^n = (p_n - p') / (p_0 - p')$

$$(1-m)^{10} = (0.446 - 0.028) / (0.630 - 0.028)$$

$$\text{可得 } m = 1 - \sqrt[10]{0.694}$$

$$= 0.036$$



许多中性或无利性状的存在不一定能用自然选择来解释，可能是遗传漂变的作用；

例如：人类不同种族所具有的血型频率差异在实际中并没有适应上的意义，可能同类人将血型差异一直传下来，可能是遗传漂变的结果；

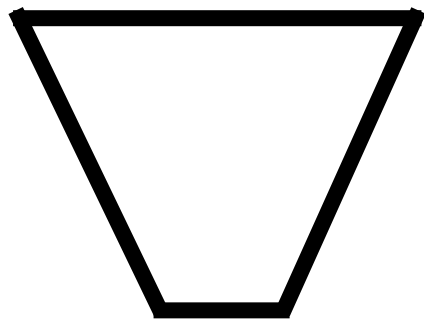
□ **遗传漂变(random genetic drift):** 一个**小群体**中，由于偶然事件导致群体中基因频率随机波动的现象。

特点：

1. 遗传漂变与抽取的样本数有关，样本数越小基因频率的波动越大，样本数越大基因频率的波动越小；
2. 环境条件的改变可能造成遗传漂变。

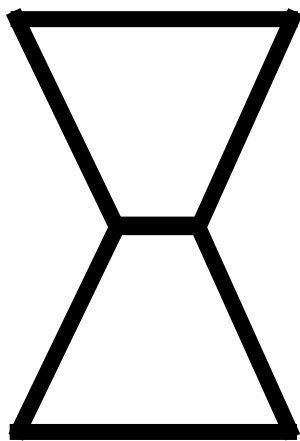
■ 奠基者效应和瓶颈效应

□ 奠基者效应(founder effect): 由少数几个个体的基因频率, 决定后代的基因频率;



■ 奠基者效应和瓶颈效应

□ 瓶颈效应(bottle effect): 一个大的群体通过“瓶颈”后，由少数几个个体再扩展成原来规模的群体，群体数量消长的过程对遗传造成的影响；



计算题

红绿色盲来源于性连锁的隐性等位基因， $1/10$ 的男性是色盲的，则：

- 1、女性中有多少是色盲的？
- 2、有多少婚配的家庭，他们的子、女各有一半是色盲？
- 3、有多少婚配的家庭，他们的孩子都是正常的？
- 4、在不平衡群体中，红绿色盲基因在女性中的频率是 0.2 ，在男性中的频率是 0.6 。经过一个世代的随机婚配，后代中多少女性是色盲？以及多少男性是色盲？
- 5、在4中，男性、女性后代的基因频率是多少？