

第一章

育种群体的遗传组成

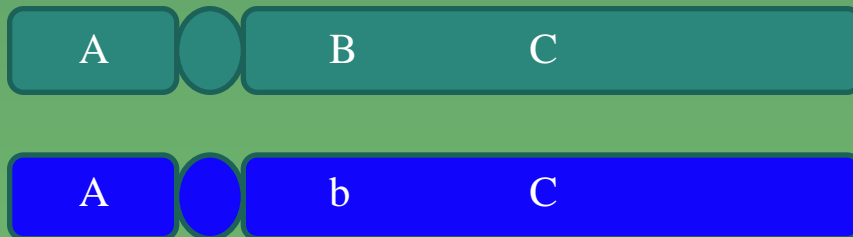
提 纲

- 群体的基因频率和基因型频率
- 不同交配系统下基因频率和基因型频率的变化
- 平衡的建立和平衡群体的性质
- 连锁对Hardy-Weinberg平衡的影响

群体的基因频率和基因型频率

基本概念

- 基因座 (locus)、基因 (gene)、等位基因 (allele)
- 基因型 (genotype)
 - 纯合子 (homozygous, homozygosity)
 - 杂合子 (heterozygous, heterozygosity)
- 表型 (phenotype)



♀

♂

姊妹染色体

群体的定义

- 群体遗传学的研究对象是遗传群体（population），群体是指具有共同特征的一些个体所组成的集合。生物群体可能包含所有生物个体如人、动物、植物、微生物等的种群，而对于群体遗传学中的群体，其定义比较狭窄，往往指生物的一个种、一个包含变异的品种或一个品种群，甚至一些个体杂交后的特定世代。

群体遗传学的主要任务

- 某一世代中，基因和基因型频率的大小、不同世代中基因和基因型频率的变化情况，即群体结构的变化情况
- 引起基因和基因型频率变化的条件和原因以及由此变化而产生的后果
- 生物群体上下代之间的遗传及基因变化规律
- 从量的角度研究群体遗传变异规律，诸如研究群体的基因频率、基因型频率以及研究基因型频率受各种因素影响而随世代或年代变化和群体基因型的适应性问题等

群体结构

- 基因频率
- 基因型频率
- 基因频率和基因型频率都相同的群体，则它们有相同的群体结构
- 如F₂和F₃群体，它们有着相同的基因频率，但基因型频率不相同，因此它们是两个不同的群体

群体的基因型频率

- 一对等位基因，为A和a，二倍体
- 三种基因型AA、Aa和aa
- (AA, Aa, aa) → (D, H, R) 表示不同基因型的个体数， $N=D+H+R$
- 三种基因型AA，Aa和aa的频率依次分别为：

$$p_{AA} = \frac{D}{N}$$

$$p_{Aa} = \frac{H}{N}$$

$$p_{aa} = \frac{R}{N}$$

群体的基因频率

$(AA, Aa, aa) \rightarrow (D, H, R)$

$N=D+H+R$ 为群体大小

等位基因A的频率

$$p = \frac{2 \times D + H}{2N} = \frac{D + \frac{1}{2}H}{N}$$

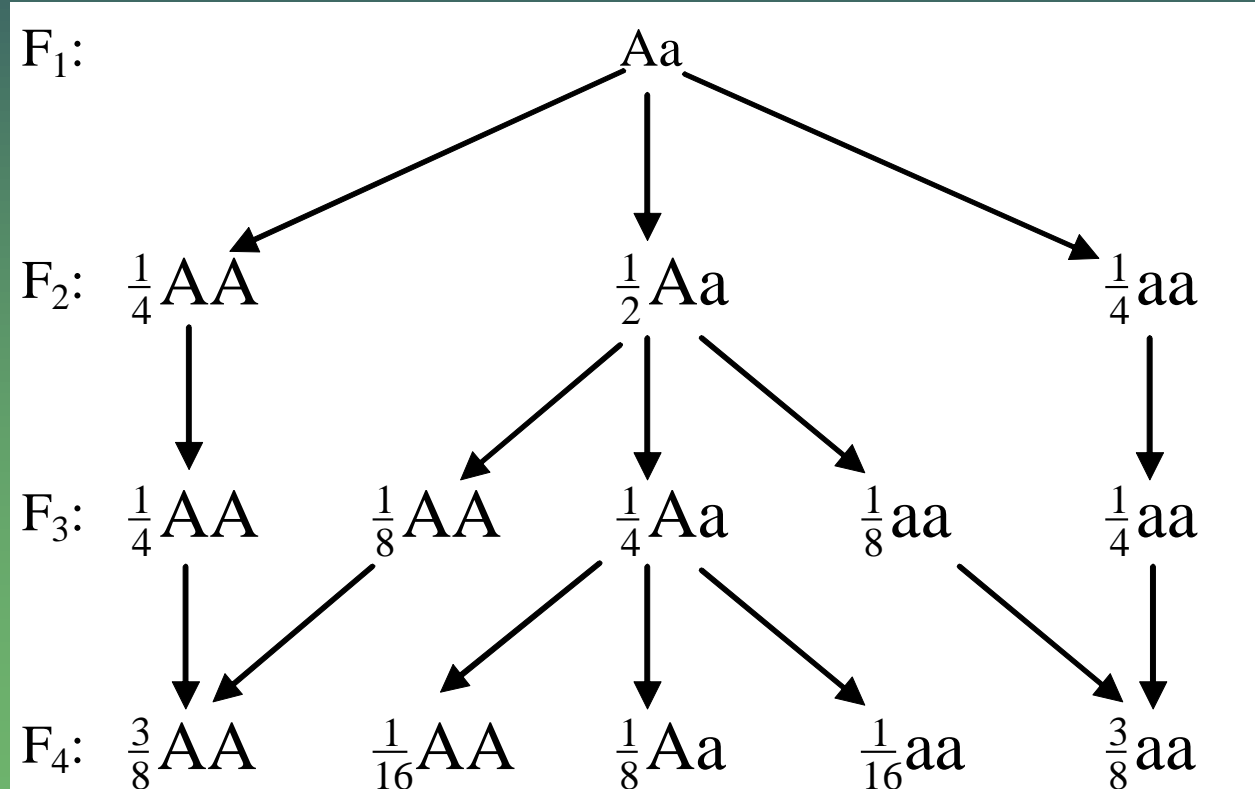
等位基因a的频率

$$q = \frac{H + 2 \times R}{2N} = \frac{\frac{1}{2}H + R}{N}$$

不同交配系统下基因频率和 基因型频率的变化

自交交配系统下基因频率和基因型频率的变化

一对等位基因A和a的情况下，其自交分离如图



杂合子Aa自交后, 各世代基因和基因型频率的变化

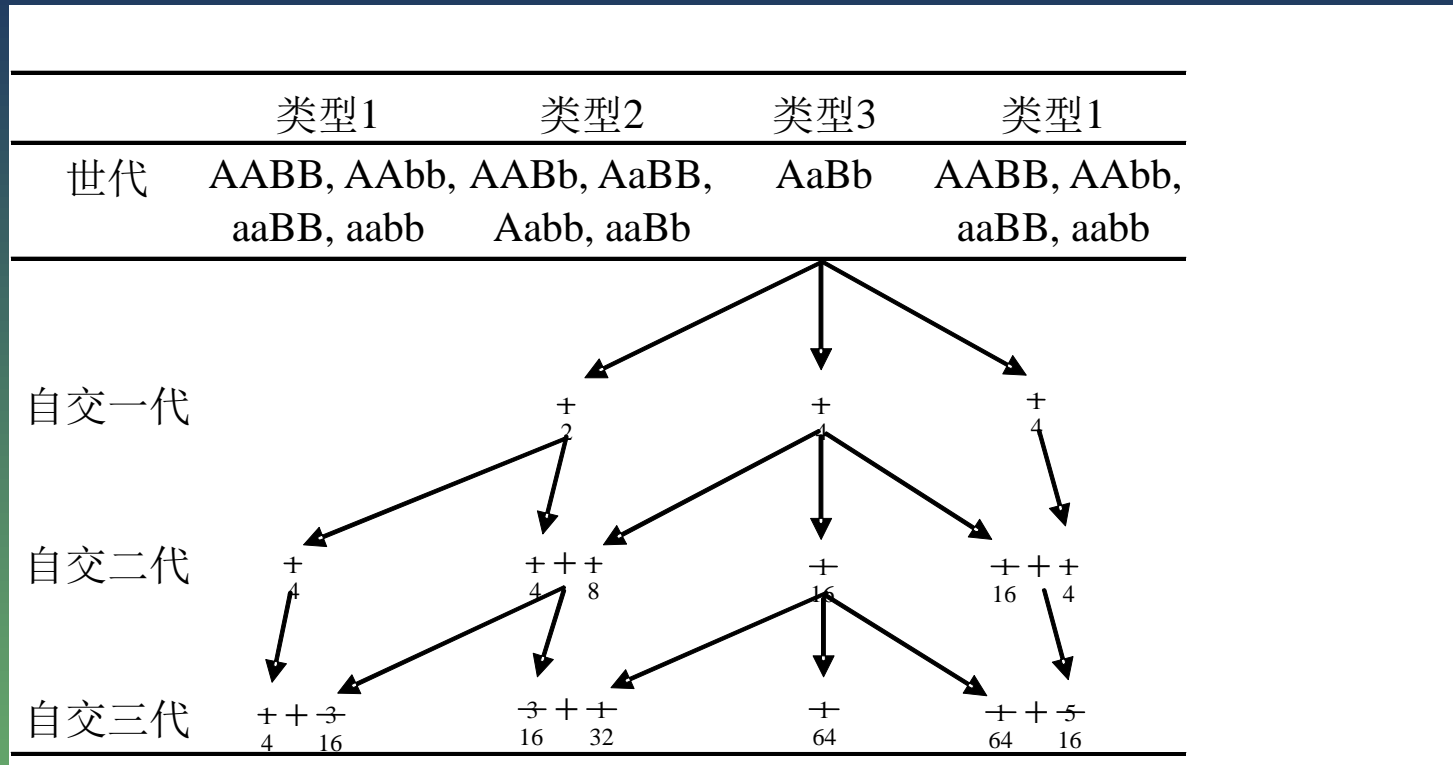
自交世代	基因型			基因	
	AA	Aa	aa	A	a
F1	0	1	0	0.5	0.5
F2	1/4	1/2	1/4	0.5	0.5
F3	3/8	1/4	3/8	0.5	0.5
F4	7/16	1/8	7/16	0.5	0.5
F5	15/32	1/16	15/32	0.5	0.5

多个位点，每个位点有一对等位基因的情况（以两对等位基因A/a和B/b为例）

基因型分为3种类

- 类型1：纯合型，即两对基因都是完全纯合的，如 AABB、AAbb、aaBB和aabb
- 类型2：部分杂合型，即一个位点是纯合的，另一个位点是杂合的，如AABb、Aabb、AaBB和aaBb
- 类型3：纯杂合型，即两对基因完全杂合，基因型只有一种：AaBb

两个独立遗传位点的自交后代分离



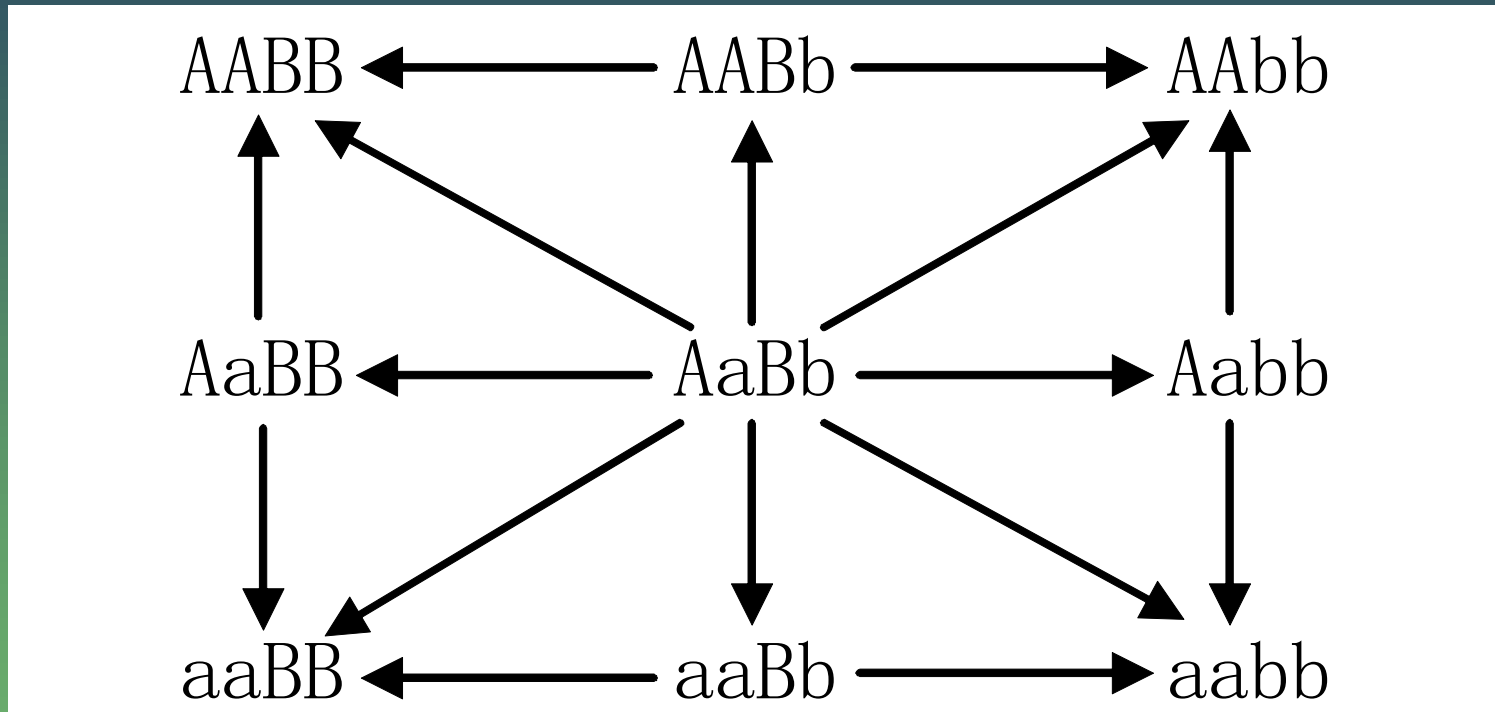
自交 n 个世代以后，三个类型的基因型频率用下式的展开表示

$$\frac{1}{2^{2n}} [1 + (2^n - 1)]^2$$

- 若个体为 m 对基因的杂合体 $AaBbCc\dots$, 经过几代自交后, 后代出现 $m+1$ 种类型的基因型, 各种类型的频率用此展开项来表示

$$\frac{1}{2^{mn}} [1 + (2^n - 1)]^m$$

自交系统基因型演变的直观形式



回交交配系统

$$P_1 : AA \times P_2 : aa$$



$$B_{11} : AA \times Aa = \frac{1}{2} AA + \frac{1}{2} Aa$$



$$B_{12} : AA \times \left(\frac{1}{2} AA + \frac{1}{2} Aa\right) = \frac{3}{4} AA + \frac{1}{4} Aa$$

回交育种

- 回交是植物育种中常采用的一种方法，特别是用来转移某一个特定基因所控制的性状，通常所说的回交是指 $P_1 \times P_2$ 的 F_1 个体和某一亲本的杂交。
- 如果 F_1 和 P_1 回交，这时亲本 P_1 称为轮回亲本，而另一亲本 P_2 称为非轮回亲本或供体亲本。
- 回交群体一般用 B_{ij} 表示， $i=1$ 或 2 表示轮回亲本， $j=1, 2, \dots$ 表示回交代数（也可用 P_iBC_j 更明确地表示不同轮回的不同回交时代）。
- 一次或两次回交有时又称有称为有限回交（Limited backcross），多次回交有时又称高代回交（advanced backcross）。有限回交CIMMYT育种家也称之为简单回交（simple backcross）。

锈病持久抗性研究： Ravi Singh



- 广适应品种
× 锈病持久
抗性材料
- 与广适应品
种回交1-2次
- 大分离群体
- 选择与广适
应品种类似、
同时又具有
持久抗性的
后代

随机交配

- 如果用 D 、 H 、 R 表示群体中相应于 AA 、 Aa 、 aa 的基因型频率，则群体的随机交配如下表

Male parents	Female parents		
	AA, D	Aa, H	aa, R
AA, D	D^2	DH	DR
Aa, H	HD	H^2	HR
aa, R	RD	RH	R^2

Hardy-Weinberg平衡 (HWE) 群体

- 在一个群体中，一个座位上两个等位基因A和a的频率为p和q，三种基因型AA、Aa和aa的频率为D、H和R，如果基因型频率和基因频率满足：

$$D=p^2、H=2pq、R=q^2，或者说$$

D、H、R对应于 $(p+q)^2$ 的展开项

- 则称该群体处于Hardy-Weinberg平衡状态

HWE群体的随机交配后代仍然处于HWE状态

AA: $D=p^2$ 、Aa: $H=2pq$ 、aa: $R=q^2$

交配类型	频率	后代基因型及其频率		
		AA	Aa	aa
AA × AA	p^4	1	0	0
AA × Aa	$4p^3q$	0.5	0.5	0
AA × aa	$2p^2q^2$	0	1	0
Aa × Aa	$4p^2q^2$	0.25	0.5	0.5
Aa × aa	$2pq^3$	0	0.5	0.5
aa × aa	q^4	0	0	1
合计	1	p^2	$2pq$	q^2

不管群体的起始频率如何, 只要经过一代随机交配, 群体就达到平衡

AA: D, Aa: H, aa: R; A: $p=D+0.5H$, a: $q=0.5H+R$

交配类型	频率	后代基因型及其频率		
		AA	Aa	aa
AA × AA	D^2	1	0	0
AA × Aa	$2DH$	0.5	0.5	0
AA × aa	$2DR$	0	1	0
Aa × Aa	H^2	0.25	0.5	0.5
Aa × aa	$2HR$	0	0.5	0.5
aa × aa	R^2	0	0	1
合计	1	$(D+0.5H)^2$	$(D+0.5H)(0.5H+R)$	$(0.5H+R)^2$

Hardy-Weinberg平衡定律

- In a large random-mating population with no selection, mutation, or migration, the gene frequencies and the genotype frequencies are constant from generation to generation; and furthermore, there is a simple relationship between the gene frequencies and the genotype frequencies (Falconer's text book, Page 5)
- 这个定律是由Hardy和Weinberg在1908年同时发现的，所以一般称为Hardy-Weinberg平衡定律

Hardy-Weinberg平衡定律

- The population with constant gene frequencies and the genotype frequencies is said to be in HWE
- HWE can be achieved in one generation of random mating
- 到达HWE的条件
 - **In a large random-mating population**
 - **With no selection, mutation, or migration**

平衡的建立和平衡群体的性质

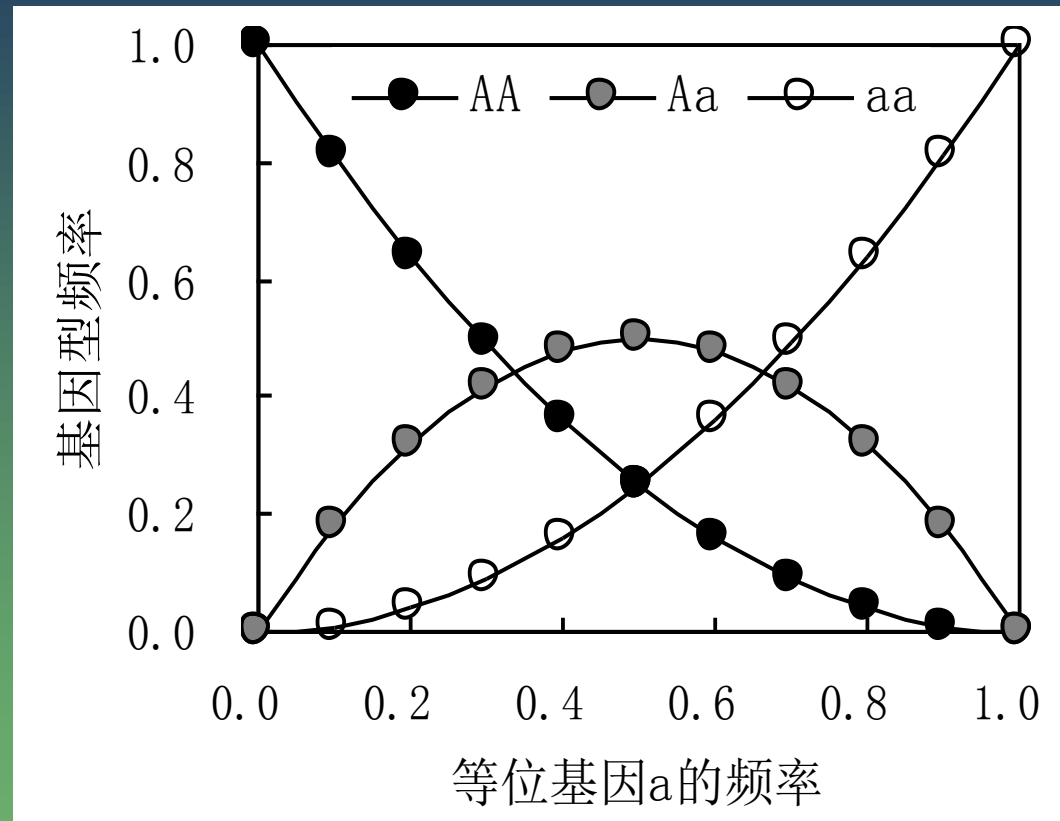
Hardy-Weinberg平衡群体的性质

- 在一个平衡群体中，杂合子的频率不会超过0.5，也不可能大于纯合子的频率之和

$$H = 2p(1-p) = 2p - 2p^2$$

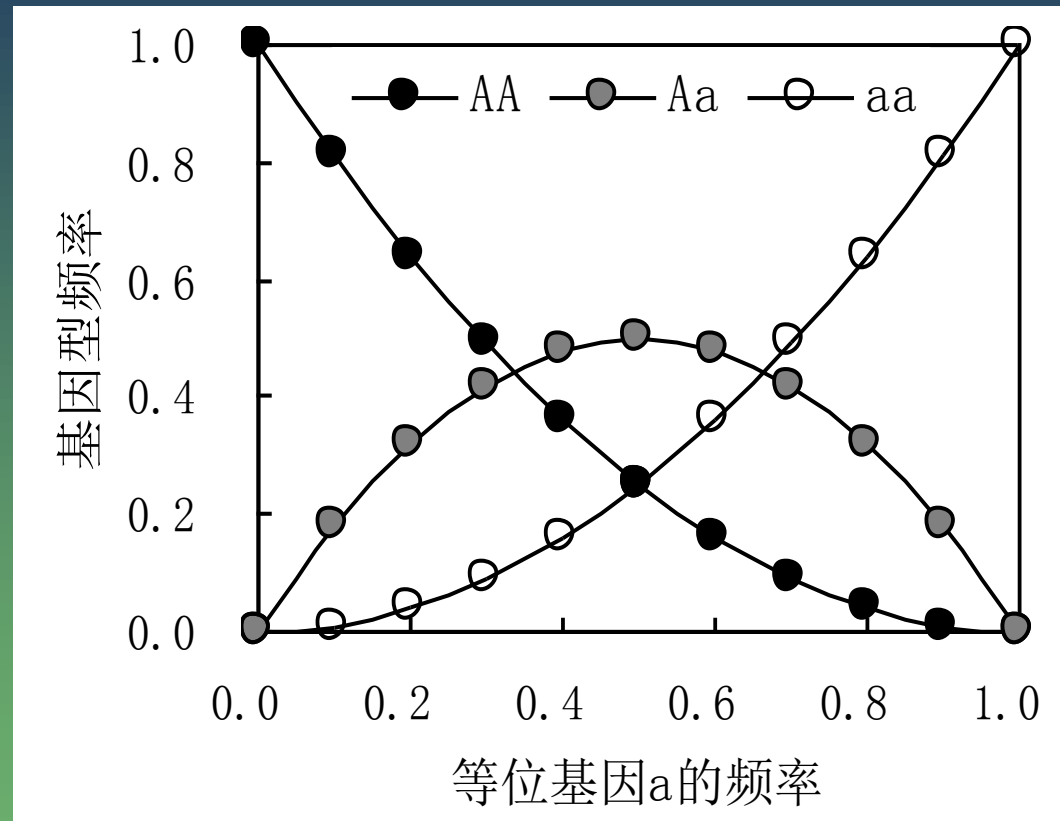
$$\frac{dH}{dp} = 2 - 4p$$

- H在 $p=0.5$ 处达到极大值0.5



Hardy-Weinberg平衡群体的性质

- 当一个等位基因的频率是另一个的2倍以上时，杂合体频率介于两种纯合体之间
- $D=H: p^2=2pq,$
 $p=2q=2-2p,$ 因此
 $p=2/3, q=1/3$



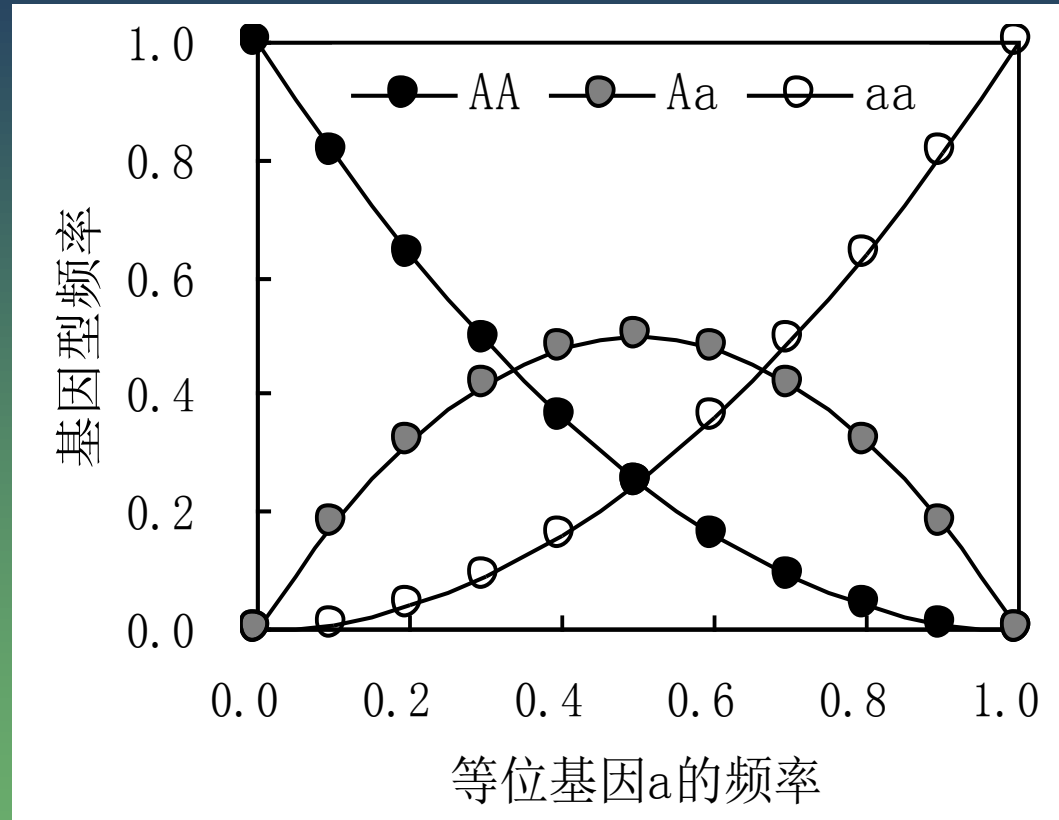
Hardy-Weinberg平衡群体的性质

- 当一个a基因有很小的频率 (q) 时, 该基因大多存在于杂合体中

- 杂合型Aa与隐性纯合型aa携带有害基因a的比例为

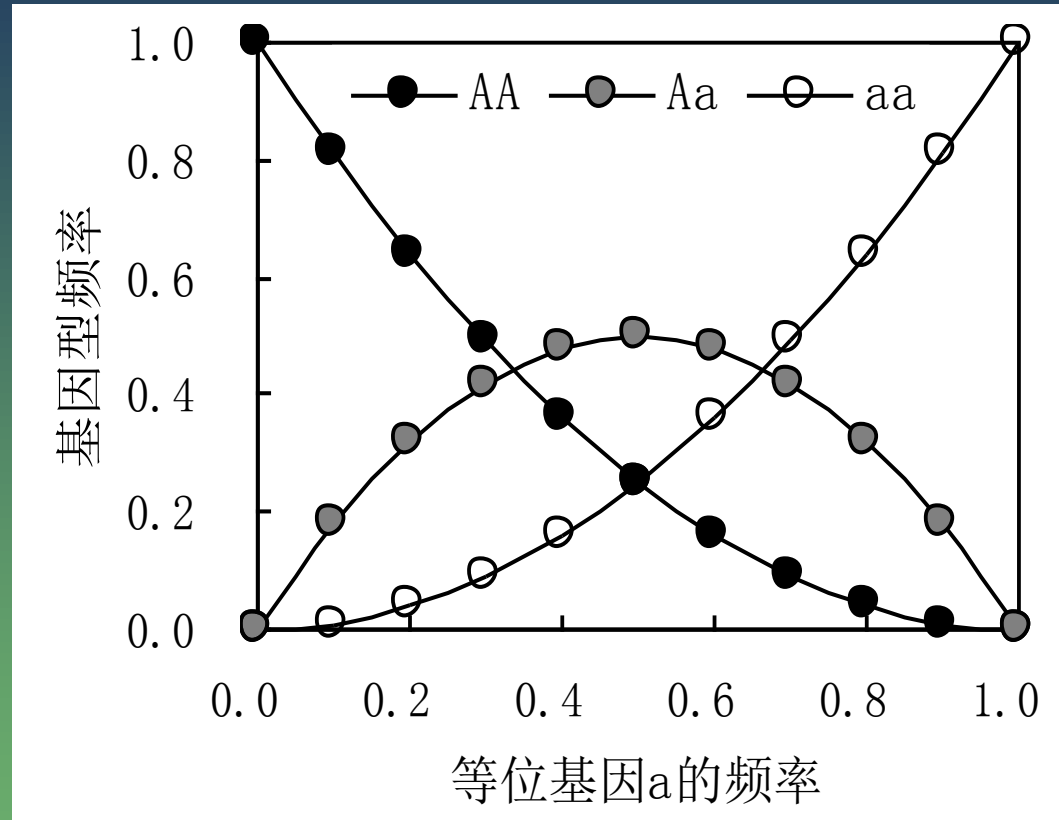
$$\frac{H}{2 \times R} = \frac{2pq}{2q^2} = \frac{p}{q}$$

- 因此, 对于很小的q, p/q会非常大



Hardy-Weinberg平衡定律的作用

- 测定基因和基因型频率
- 任一平衡群体的3种基因型频率在平衡图上必然形成一条直线，且该直线垂直交于X轴，交点将X轴分为两部分，从原点到交点为一基因频率，从交点到1.0为另一基因频率。



基因型随机交配与配子型随机交配等价

Male parents	Female parents		
	AA, D	Aa, H	aa, R
AA, D	D ²	DH	DR
Aa, H	HD	H ²	HR
aa, R	RD	RH	R ²

Male gametes	Female gametes	
	A, p	A, q
A, p	p ²	pq
a, q	qp	q ²

Hardy-Weinberg平衡定律的作用

- 计算杂合体的频率：有时需计算杂合体的频率，如某种疾病为隐性基因控制，纯合隐性表现症状，杂合体并不表现症状，那么往往需要了解**致病基因携带者在正常人群中所占的比重**，这便是杂合体所占的频率。

$$\frac{H}{D+H} = \frac{2q(1-q)}{[(1-q)^2 + 2q(1-q)]} = \frac{2q}{1+q}$$

- 当q很小时，

$$\frac{H}{D+H} = \frac{2q}{1+q} \approx 2q$$

Hardy-Weinberg平衡定律的作用

- 例如，苯酮尿是由一隐性基因控制的人类代谢方面疾病，婴儿出生后即可鉴定，在英国伯明翰地区连续三年对出生婴儿的检查发现，在55,715个婴儿中有5例为该病患者，现试求正常人群中此病携带者的比例。
- 在HWE的假定下，隐性基因的频率

$$q = \sqrt{\frac{5}{55,715}} = 0.0095$$

- 发病比例不到万分之一，但正常人中携带致病基因的比例接近2%

Hardy-Weinberg平衡定律的检验

$$\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} \sim \chi^2(df)$$

样本容量为747的冰岛人的血型数据

We want to test if the population is in HW equilibrium:

If $(M, N) = (p, q)$, then $(MM, MN, NN) = (p^2, 2pq, q^2)$

Genotype	MM	MN	NN	Total
Sample size	$N_1=233$	$N_2=385$	$N_3=129$	$N=747$
Predicted by HWE	242.70	366.18	138.38	747

- $(p, q) = (0.57, 0.43)$
- $(p^2, 2pq, q^2) = (0.3249, 0.4902, 0.1849)$
- $\chi^2=1.96$ (**df=1, P=0.16**), i.e. the population is in HW equilibrium

F1自交与F1随机交配是等价的， 因此F2是一个HWE群体

F1群体中, AA: $D=0$; Aa: $H=1$; aa: $R=0$; $p=q=0.5$

Male parents	Female parents		
	AA, $D=0$	Aa, $H=0$	aa, $R=0$
AA, $D=0$	0	0	0
Aa, $H=0$	0	1	0
aa, $R=0$	0	0	0

- F2群体中,
 - AA: $D=0.25=0.5^2$;
 - Aa: $H=0.5=2 \times 0.5 \times 0.5$;
 - aa: $R=0.25=0.5^2$.

复等位基因的HWE

- 在一个群体中, 一个座位上三个等位基因 A_1 , A_2 和 A_3 的频率为 q_1 , q_2 和 q_3 .
- 在HWE时, 六种基因型的频率为
 - $A_1A_1: q_1^2$
 - $A_1A_2: 2q_1q_2$
 - $A_1A_3: 2q_1q_3$
 - $A_2A_2: q_2^2$
 - $A_2A_3: 2q_2q_3$
 - $A_3A_3: q_3^2$

An example on ABO blood type

when we don't have the expected frequencies

Type	A	B	AB	O	Total
N_i	2162	738	228	2876	6004

- H_0 : this is a randomly mated population, i.e. the population is in HW equilibrium.
- Need to estimate $p(A)$, $p(B)$, $p(O)$ first.
- By HWE, we mean:
 - $p(AA)=p(A)^2$, $p(AO)=2*p(A)*P(O)$
 - $p(BB)=p(A)^2$, $p(BO)=2*p(B)*P(O)$
 - $p(AB)=2*p(A)*P(B)$
 - $p(OO)=p(O)^2$

性连锁基因

(XX为女性, XY为男性)

- HWGE时的基因型频率

	女性			男性	
基因型	AA	Aa	aa	A	a
频率	p^2	$2pq$	q^2	p	q

- 对于隐性致病基因a来说, 男性患病的比例远高于女性。并非说明男性不抗病, 这是遗传学在起作用。
- 隔代遗传: AA (♀) 与a (♂) 的后代正常, 但女儿中有一半的携带者, 即Aa (♀), 即使与正常男性A (♂) 结婚, 其儿子有一半的基因型为a (♂)

The background of the slide features a pattern of stylized green leaves and branches, rendered in various shades of green, creating a naturalistic and textured backdrop.

连锁对Hardy-Weinberg 平衡的影响

配子别不平衡

- 平衡群体中，配子基因型的频率等于基因频率的乘积，因此群体偏离Hardy-Weinberg平衡的程度常用下面的公式去计算

$$D_{ij} = p(A_i B_j) - p(A_i) p(B_j)$$

连锁不平衡的度量(D)

Gene, frequency	A ₁ , p _A	A ₂ , q _A	B ₁ , p _B	B ₂ , q _B
Gamete (haplotype)	A ₁ B ₁	A ₁ B ₂	A ₂ B ₁	A ₂ B ₂
Frequency, equilibrium	p _A p _B	p _A q _B	q _A p _B	q _A q _B
Frequency, actual	u	s	t	v
Difference from equilibrium	+D	-D	-D	+D

- 基因频率: $p_A = u + s$; $q_A = t + v$; $p_B = u + t$; $q_B = s + v$.
- 对于A₁B₁来说, $p(A_1B_1) - p(A_1)p(B_1) = u - (u+s)(u+t) = u(1-u-s-t) - st = uv - st$

连锁不平衡的度量(D)

Gene, frequency	A ₁ , p _A	A ₂ , q _A	B ₁ , p _B	B ₂ , q _B
Gamete (haplotype)	A ₁ B ₁	A ₁ B ₂	A ₂ B ₁	A ₂ B ₂
Frequency, equilibrium	p _A p _B	p _A q _B	q _A p _B	q _A q _B
Frequency, actual	u	s	t	v
Difference from equilibrium	+D	-D	-D	+D

$$D = uv - st$$

F2群体中的连锁不平衡

$A_1A_1B_1B_1 \times A_2A_2B_2B_2$

Gamete (haplotype)	A_1B_1	A_1B_2	A_2B_1	A_2B_2
Frequency	$0.5(1-r)$	$0.5r$	$0.5r$	$0.5(1-r)$

$$D = \left[\frac{1}{2} (1 - r) \right]^2 - \left(\frac{1}{2} r \right)^2 = \frac{1}{4} (1 - 2r)$$

- 当 $r=0$ 时, $D_{\max}=0.25$

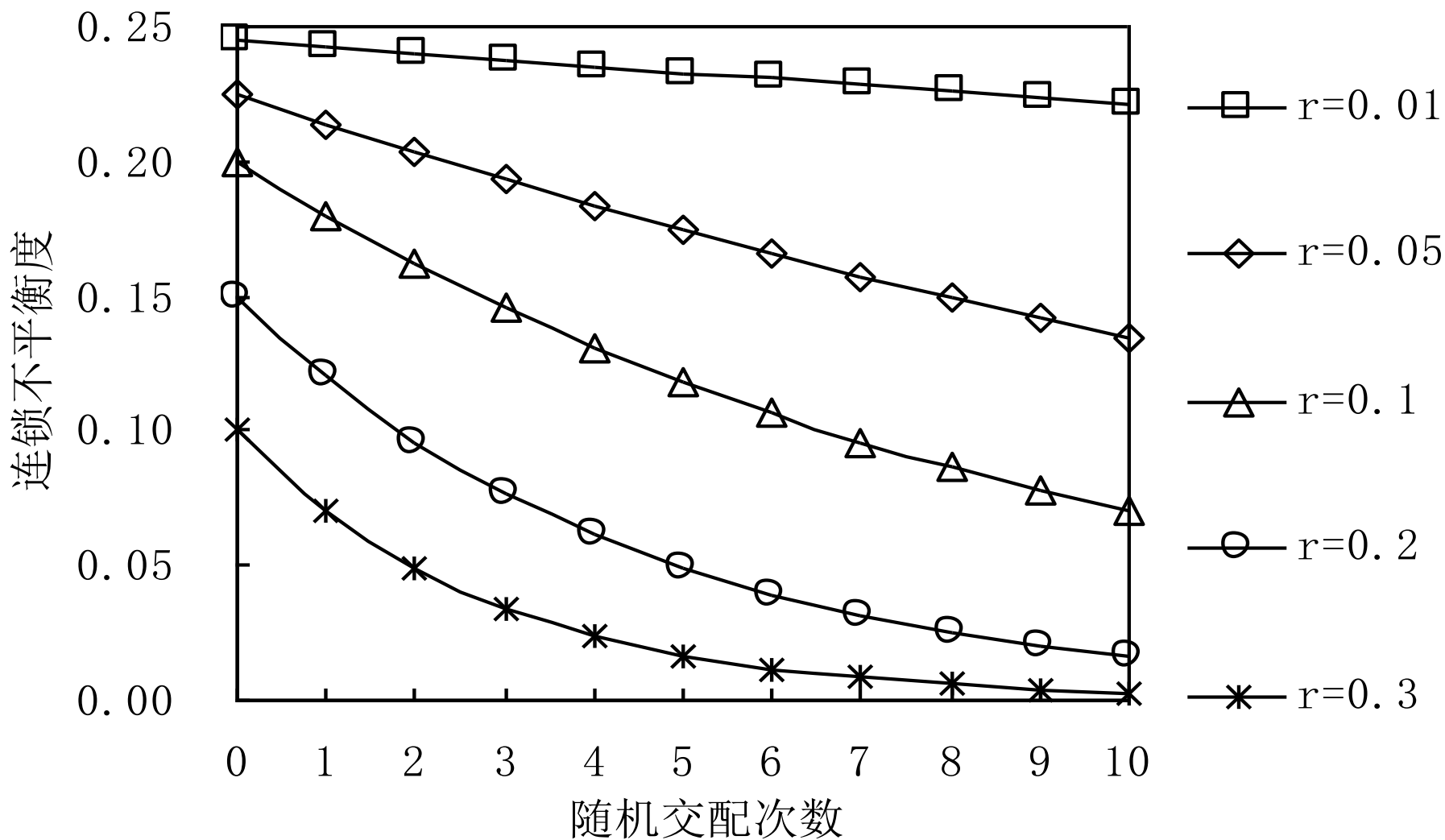
随机交配过程中的连锁不平衡

Gene, frequency	A_1, p_A	A_2, q_A	B_1, p_B	B_2, q_B
Gamete (haplotype)	A_1B_1	A_1B_2	A_2B_1	A_2B_2
Frequency, equilibrium	$p_A p_B$	$p_A q_B$	$q_A p_B$	$q_A q_B$
Frequency, actual	u	s	t	v

- 以配子 A_1B_1 为例, 有两种产生配子 A_1B_1 的途径
 - 由基因型 A_1B_1/A_xB_x ($x=1$ 或 2) 不发生重组产生, 频率为 $u(1-r)$
 - 由基因型 A_1B_x/A_xB_1 ($x=1$ 或 2) 发生重组产生, 频率为 $p_A p_B r$
- 因此, $D_1 = [u(1-r) + p_A p_B r] - p_A p_B = [u - p_A p_B](1-r) = D_0(1-r)$

多代随机交配后的 连锁不平衡

$$D_t = D_0 (1-r)^t$$



Two loci A and B in fruit fly, 17.8kb apart on one chromosome

Haplotype		Obs. samples	Obs. Frequency	Exp. Frequency	Exp. samples
A	B				
+	+	4	0.085	0.06	2.8
+	-	4	0.085	0.11	5.2
-	+	13	0.277	0.30	14.1
-	-	26	0.533	0.53	24.9

- $D=0.023$, Chi-square=0.93, $df=1$, $P > 0.5$
- LD is not significant

群体结构对LD的影响

- 不同频率的随机交配群体的混合也会导致不平衡
 - 群体1和群体2 为随机交配群体，位点A和B独立遗传
 - 混和群体一半来自群体1，另一半来自群体2

群 体	基因频率				基因型频率
	A1	A2	B1	B2	A1A1B1B1
群体1	0.7	0.3	0.7	0.3	0.2401
群体2	0.3	0.7	0.3	0.7	0.0081
混合群体	0.5	0.5	0.5	0.5	观测频率：0.1241 期望频率：0.0625

群体结构对 LD 的影响

群 体	Gene frequency			
	A ₁	A ₂	B ₁	B ₂
群体1	0.7	0.3	0.7	0.3
群体2	0.3	0.7	0.3	0.7
1:1混合群体	0.5	0.5	0.5	0.5

群 体	Haplotype frequency				LD
	A ₁ B ₁	A ₁ B ₂	A ₂ B ₁	A ₂ B ₂	
群体1	0.49	0.21	0.21	0.09	0
群体2	0.09	0.21	0.21	0.49	0
混合群体实际频率	0.29	0.21	0.21	0.29	0.04
混合群体平衡频率	0.25	0.25	0.25	0.25	

连锁在育种中的作用

- 如果两个优良基因连锁在一起，我们希望连锁越紧密越好，两个基因能一起传递到下一代，这种连锁最好不要被打破
- 如果一个优良基因和一个不利基因连锁在一起，则我们希望能打破这种连锁

Breeding is a tedious and long procedure when considering the nature of complex inheritance: linkage, epistasis, pleiotropy, GbyE, etc.

Dr. Borlaug spent more than 20 years from

this tall wheat to this semi-dwarf wheat,
which triggered the Green Revolution!

