

第1章

群体结构与交配系统

王建康

中国农业科学院作物科学研究所

wangjiankang@caas.cn

<http://www.isbreeding.net>

主要内容

- § 1.1 遗传学基本概念和理论
- § 1.2 群体的遗传结构
- § 1.3 自交和回交对群体结构的影响
- § 1.4 随机交配与随机交配群体
- § 1.5 连锁对群体结构的影响

§ 1.1 遗传学基本概念和理论

- § 1.1.1 基因座位、等位基因和基因型
- § 1.1.2 基因（基因型）到性状（表型）
- § 1.1.3 分离定律
- § 1.1.4 自由组合定律
- § 1.1.5 连锁和交换定律

基因的定义

- 基因一词泛指在生物繁殖过程中，能够由亲代传递给子代的实体遗传物质，这些遗传物质和环境一起决定着生物的各种表型性状。
- 人类大约有两万至两万五千个基因，基因影响发色、眼色、肤色、身高、体重等性状以及各种行为性状，这些性状同时又或多或少地受到环境的影响。
- 基因是携带遗传信息的一段DNA序列。DNA分子是由A、T、G、C 4种碱基，按照A与T配对、G与C配对原则，形成具有方向性的双螺旋结构。
- DNA序列经常被看作是控制性状的基本遗传单位。DNA分子有大有小。

基因座位

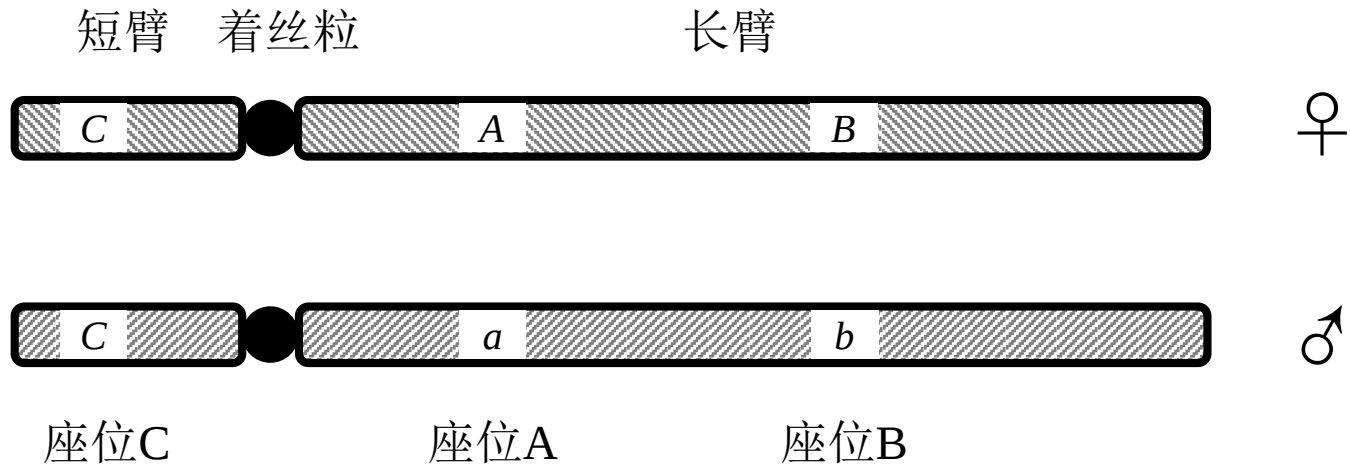
- 不同物种组织细胞的细胞核中，包含数目相对固定的一组染色体，基因按照一定的顺序线性排列在这些染色体上，每条染色体上排列的基因数以千计。
- 基因在染色体上所处的物理位置称为基因座位，或简称座位（locus）。

等位基因

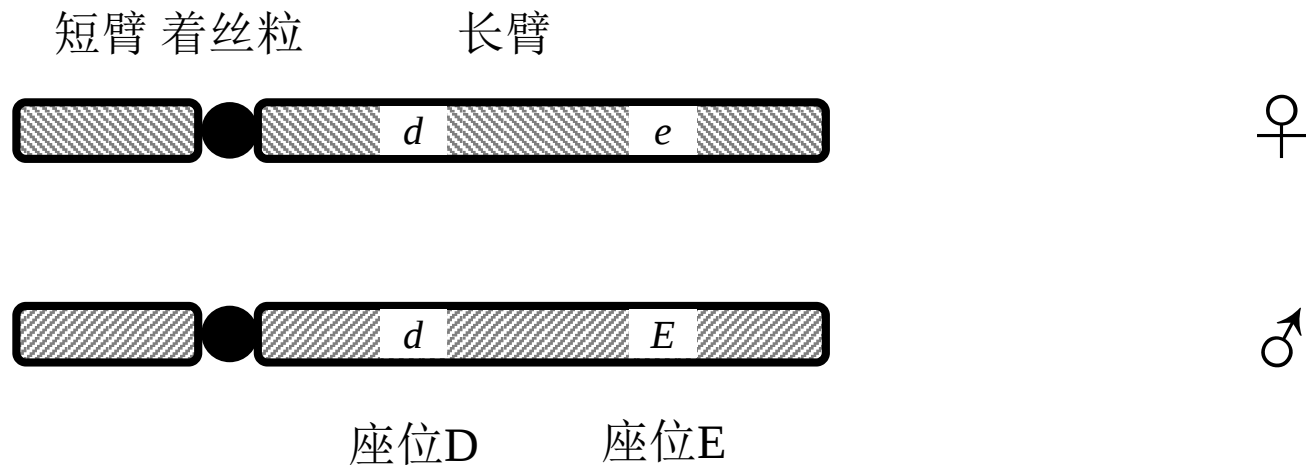
- 同一个座位上的基因称为等位基因（allele）。
- 在大多数高等生物的组织细胞中，染色体是成对出现的，即每条染色体存在有两份拷贝。
- 这两份拷贝互为同源性（homologous，即绝大部分位置上有相同的DNA序列，携带着相同的基因，也称为同源染色体），一份由母本产生的雌配子提供，一份由父本产生的雄配子提供。
- 这样的物种又称为二倍体（diploid），常用 $2n=x$ 表示二倍体物种的 n 对、 x 条染色单体。

一个二倍体生物个体组织细胞中的两对常染色体

第1对同源染色体



第2对同源染色体



纯合型、杂合型、复等位基因

- 一个个体可能携带两个完全相同的等位基因，这时称个体在这个座位上纯合的（homozygous），或称为纯合基因型（homozygote），简称纯合型；
- 也可能携带两个不同的等位基因，这时称个体在这个基因座位上杂合的（heterozygous），或称为杂合基因型（heterozygote），简称杂合型。
- 一个二倍体个体携带两个等位基因，但从群体的角度来看，一个座位上的等位基因可以只有一个，即无多态性座位；也可以有两个，还有可能多于两个。如等位基因数目多于两个，则这样的等位基因又称为复等位基因（multiple allele）

基因（基因型）到性状（表型）

- 与基因型相对应，一个个体表现出来的、外在的体征称为表现型（phenotype），简称表型。
- 基因遵循中心法则（Central Dogma），经过转录和翻译等一系列分子和生化过程，合成生长发育所需的蛋白质，基因通过基因的表达决定生物的表型。
- 以孟德尔杂交试验中的豌豆籽粒形状为例，控制圆鼓和皱缩两种表型的两个等位基因分别用W和w表示，对于它们在DNA序列上的差异、控制籽粒形状的生化代谢途径已经研究清楚。

基因的显隐性

- 对于圆鼓和皱缩两种表型来说，杂合型 Ww 与纯合型 WW 有相同的表型，因此称 W 是显性等位基因（dominant allele），简称显性基因（dominant gene）。
- 杂合型 Ww 中，等位基因 W 的存在抑制了纯合型 ww 表型的出现，因此称 w 是隐性等位基因（recessive allele），简称隐性基因（recessive gene）。
- 基因显隐性的确定需要基因型和表现型两方面的信息，离开性状谈基因的显隐性是没有意义的。同时，对于同样的两个等位基因，当考察的性状发生改变后，基因的显隐性也可能发生改变。

性状的显隐性

- 对于单基因控制的性状，也可类似于基因的显隐性来定义表型的显隐性。
- 对于圆鼓和皱缩两种籽粒形状表型来说，杂合型 Ww 为圆鼓籽粒，皱缩表型被隐藏起来了。因此，圆鼓相对于皱缩表现为显性，或圆鼓是一种显性表型（dominant phenotype）；
- 皱缩相对于圆鼓表现为隐性，或皱缩是一种隐性表型（recessive phenotype）。

基因与环境共同决定表型

- 表型是基因型和环境共同作用的结果，大部分情况下，个体的基因型和表型之间不存在明确的一一对应关系。两个相同的基因型，由于环境的差异也可以表现出不完全相同的表型；同样的表型，也可能来自于不同的基因型。
- 遗传研究中，很多时候只有个体的基因型和表型数据是不够的，还需要知道这些个体的亲代或者这些个体杂交繁殖后产生子代的基因型和表型。
- 一个性状还可能受多个座位上基因的影响。反过来，一个基因也有可能同时影响多个性状，这一现象称为基因的多效性（pleiotropy）。
- 遗传学研究中往往要区分个体的基因型和表现型。

孟德尔豌豆单交杂交试验结果

性状	杂交亲本	F ₁ 杂种表现	F ₂ 性状分离观测值	F ₂ 性状观测分离比
籽粒形状	圆鼓×皱缩	圆鼓	5474圆鼓， 1850皱缩	圆鼓:皱缩=2.96:1
种皮颜色	黄色×绿色	黄色	6022黄色， 2001绿色	黄色:绿色=3.01:1
花色	粉红×白色	粉红	705粉红， 224白色	粉红:白色=3.15:1
荚的形状	饱满×收缩	饱满	882饱满， 299收缩	饱满:收缩=2.95:1
未成熟荚颜色	绿色×黄色	绿色	428绿色， 152黄色	绿色:黄色=2.82:1
结荚习性	无限×有限	无限	651无限， 207有限	无限:有限=3.14:1
主茎高度	长型×短型	长型	787长型， 277短型	长型:短型=2.84:1

遗传学分离定律

- 分离定律（principle of segregation），又称为孟德尔第一定律（Mendel's first law），解释了一个座位上基因在亲代和子代之间的传递规律。
- 即一个座位上，个体如何产生配子、雌雄配子如何结合产生下一代个体的遗传学规律。

$$Ww \rightarrow \left(\frac{1}{2}W + \frac{1}{2}w\right)$$

$$\left(\frac{1}{2}W + \frac{1}{2}w\right) \times \left(\frac{1}{2}W + \frac{1}{2}w\right) = \frac{1}{4}WW + \frac{1}{2}Ww + \frac{1}{4}ww = \frac{3}{4}W_* + \frac{1}{4}ww$$

籽粒形状（圆粒对皱粒）和种皮颜色（绿皮对黄皮）两个性状的杂交试验

	亲本1	亲本2	F ₁ 杂种	F ₂ 世代（W _* 代表基因型WW和Ww，G _* 代表基因型GG和Gg）			
基因型	WWGG	wwgg	WwGg	W _* G _*	W _* gg	wwG _*	wwgg
表型	圆粒黄皮	皱粒绿皮	圆粒黄皮	圆粒黄皮	圆粒绿皮	皱粒黄皮	皱粒绿皮
观测值	全部	全部	全部	315	108	101	32
观测分离比	1	1	1	圆黄：圆绿：皱黄：皱绿 = 9.84 : 3.38 : 3.16 : 1; 圆粒：皱粒 = (315+108) : (101+32) = 3.18 : 1; 黄皮：绿皮 = (315+101) : (108+32) = 2.97 : 1			

遗传学自由组合定律

- 自由组合定律，又称为孟德尔第二定律（Mendel's second law），解释了两个或多个独立遗传座位上，基因在亲子之间的传递规律。

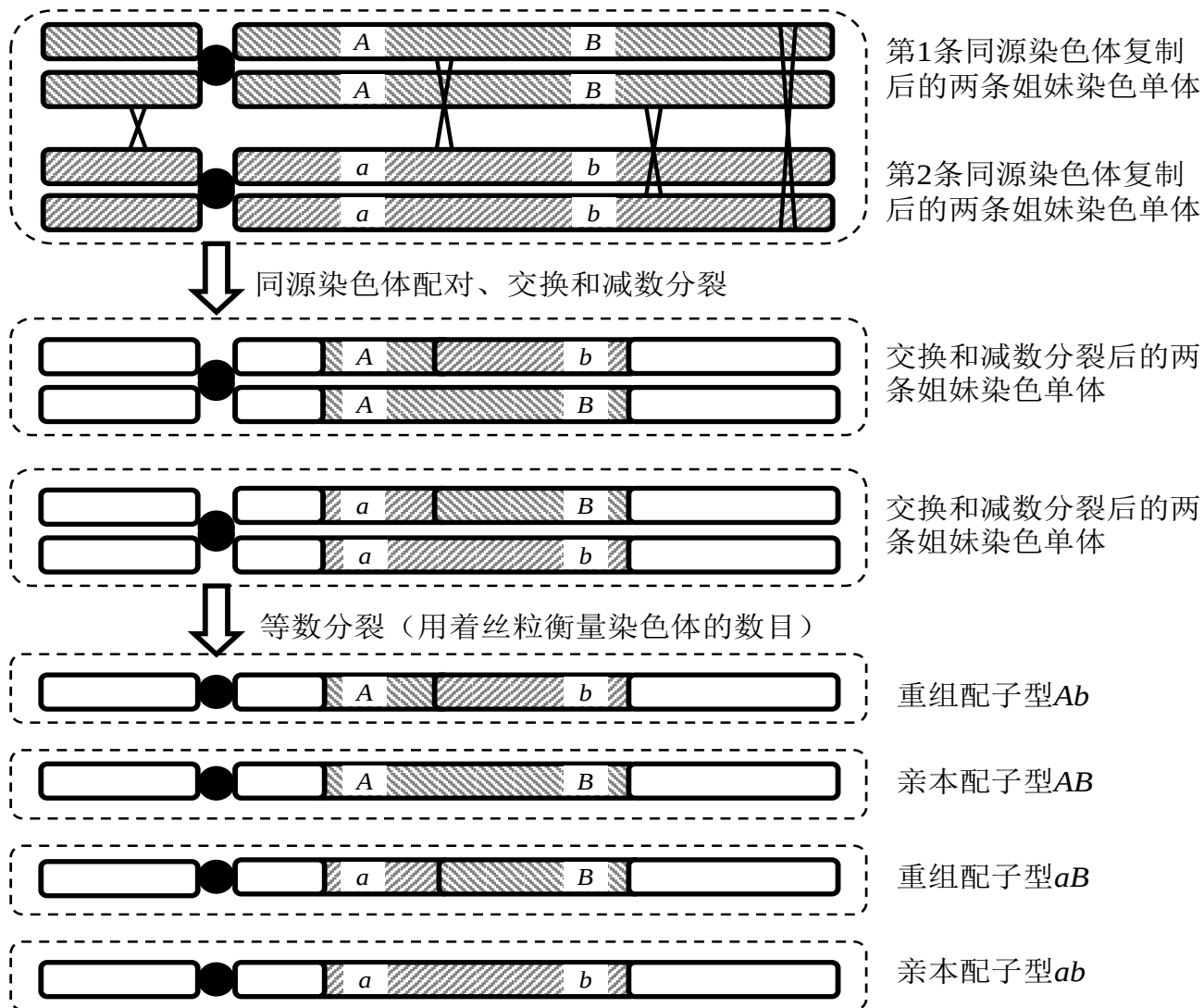
$$WwGg \rightarrow \left(\frac{1}{2}W + \frac{1}{2}w\right) \times \left(\frac{1}{2}G + \frac{1}{2}g\right) = \left(\frac{1}{4}WG + \frac{1}{4}Wg + \frac{1}{4}wG + \frac{1}{4}wg\right)$$

$$\left(\frac{1}{4}WG + \frac{1}{4}Wg + \frac{1}{4}wG + \frac{1}{4}wg\right) \times \left(\frac{1}{4}WG + \frac{1}{4}Wg + \frac{1}{4}wG + \frac{1}{4}wg\right)$$

$$= \left(\frac{1}{4}WW + \frac{1}{2}Ww + \frac{1}{4}ww\right) \times \left(\frac{1}{4}GG + \frac{1}{2}Gg + \frac{1}{4}gg\right)$$

$$= \left(\frac{9}{16}W_*G_* + \frac{3}{16}W_*gg + \frac{3}{16}wwG_* + \frac{1}{16}wwgg\right)$$

一对同源染色体上，双杂合基因型 AB/ab 产生配子过程的示意图



遗传学连锁和交换定律

- 同源染色体的非姐妹染色单体交换往往是随机发生的，染色体越长，发生交换的概率也越大；一对同源染色体上，也可同时发生多次交换。
- 对于两个连锁座位之间的一段染色体来说，如座位A和B，发生交换的概率可以利用重组型配子 Ab 和 aB 所占比例进行计算，重组配子的比例称为重组率（recombination frequency）。

$$AB / ab \rightarrow \left[\frac{1}{2}(1-r)AB + \left(\frac{1}{2}r\right)Ab + \left(\frac{1}{2}r\right)aB + \frac{1}{2}(1-r)ab \right]$$

重组率 r 的作用

- 重组率越小，说明两个座位上的等位基因结合在一起、传递到下一代的倾向越强。
- 当重组率等于0时， A 与 B 总是结合在一起， a 与 b 总是结合在一起，这样的连锁称为完全连锁（complete linkage）。
- 当重组率等于0.5时，配子的频率等于基因频率的乘积，表现为自由结合。
- 重组率直观地反映了两个座位之间的遗传距离或者连锁关系。通过重组率的估计，可以了解基因在染色体上的排列顺序，即构建遗传连锁图谱、定位控制性状的基因。

§ 1.2 群体的遗传结构

- § 1.2.1 群体的基因频率和基因型频率
- § 1.2.2 群体的杂合度和多样性
- § 1.2.3 双亲群体中基因型的期望分离比及其检验

基因频率和基因型频率

- 群体的遗传组成可以从两方面来考虑，一方面是群体中等位基因的频率，另一方面是群体中各种基因型的频率。
- 假设某一座位只有一对等位基因，用 A 和 a 表示，该群体由 n 个具有二倍体遗传特性的个体组成。群体中可能的基因型有 AA 、 Aa 和 aa 三种，这三种基因型的个体数分别用 n_{AA} 、 n_{Aa} 和 n_{aa} 表示，总样本量用 n 表示。

等位基因频率的估计

- n 个个体共携带有 $2n$ 个基因，基因型为 AA 的个体携带2个 A 基因，基因型为 aa 的个体携带2个 a 基因，基因型为 Aa 的个体携带1个 A 基因和1个 a 基因。因此， A 基因的频率和 a 基因的频率分别为：

$$p_A = \frac{2n_{AA} + n_{Aa}}{2n} = \frac{n_{AA} + \frac{1}{2}n_{Aa}}{n}$$

$$p_a = \frac{n_{Aa} + 2n_{aa}}{2n} = \frac{\frac{1}{2}n_{Aa} + n_{aa}}{n}$$

基因型频率的估计

- 三种基因型 AA 、 Aa 和 aa 的观测频率分别为：

$$p_{AA} = \frac{n_{AA}}{n} \quad p_{Aa} = \frac{n_{Aa}}{n} \quad p_{aa} = \frac{n_{aa}}{n}$$

基因频率和基因型频率的关系

- 根据它们的估计方法，不难看出，等位基因的频率还可以用三种基因型的频率表示为：

$$p_A = p_{AA} + \frac{1}{2} p_{Aa} \quad p_a = p_{aa} + \frac{1}{2} p_{Aa}$$

群体的杂合度

- 对于单个座位来说，杂合基因型的频率又称为群体的杂合度（heterozygosity）。
- 对于只包含两个等位基因 A 和 a 的座位，杂合度等于杂合基因型 Aa 的频率。
- 对于具有复等位基因 A_1 、 A_2 、 A_3 、...的座位 l ，杂合度等于所有可能杂合基因型频率之和。

平均杂合度的计算

- 座位 l 上的杂合度等于所有可能杂合基因型频率之和。

$$p_{lu} = \frac{1}{2} \sum_{v < u} p_{lvu} + p_{luu} + \frac{1}{2} \sum_{v > u} p_{luv} \quad , \quad \text{其中} \quad p_{luv} = \frac{n_{uv}}{n}$$

- 对于 m 个座位来说，平均杂合度（average heterozygosity）定义为：

$$H = \frac{1}{m} \sum_{l=1, \dots, m} H_l$$

基因的多样性

- 杂合度是从杂合基因型频率计算而来。在近交群体中，杂合基因型的频率很低，常用基因多样性（gene diversity）度量群体的异质性，基因多样性由等位基因频率计算而来。

基因多样性的计算

- 假定座位 l 上等位基因的频率用 p_{lu} 表示，基因多样性定义为：

$$D_l = 1 - \sum_u p_{lu}^2$$

- 对于 m 个座位来说，平均基因多样性定义为：

$$D = \frac{1}{m} \sum_{l=1, \dots, m} D_l = 1 - \frac{1}{m} \sum_{l=1, \dots, m} \sum_u p_{lu}^2$$

杂合度与多样性的关系

- 群体杂合度由杂合基因型的频率计算而来，杂合基因型频率越大，群体杂合度就越高。
- 基因多样性由等位基因的频率计算而来，等位基因越多、频率越一致，多样性就越高。
- 一般群体中，不能从等位基因的频率推测出基因型的频率。但是，在随机交配的大群体中，基因型频率可以由等位基因频率计算出来，也只有在这种情况下，群体的杂合度才等于群体的基因多样性。

双亲群体中基因型的期望分离比及其检验

- 如果一个群体有清楚的来源，则基因和基因型都有一个期望频率。如果不存在奇异分离，基因和基因型频率应该与期望频率之间不存在显著性差异。
- 一个群体的基因型频率是否符合一个已知的期望分离比，可以采用 χ^2 统计量进行适合性检验。

$$\chi^2 = \sum \frac{(O - E)^2}{E} \quad \chi^2 = \sum \frac{(|O - E| - 0.5)^2}{E}$$

§ 1.3 自交和回交对群体结构的影响

- § 1.3.1 交配系统及其分类
- § 1.3.2 自交对群体结构的影响
- § 1.3.3 回交对群体结构的影响

交配系统及其分类

- 动物以随机交配（random mating）为主，通过有性生殖（sexual reproduction）的方式繁殖后代；
- 植物可通过无性生殖（asexual reproduction）和有性生殖两种交配系统（mating system）繁殖后代。
- 无性生殖植物通过营养生长或无融合（apomixis）的方式进行繁殖，如香蕉、木薯、马铃薯、甘薯和甘蔗。

有性生殖及其分类

- 有性生殖又可分为自花授粉（self-pollination）和异花授粉（cross-pollination）两种类型。
- 自然条件下，有些植物以自花授粉为主繁殖后代，如水稻、小麦和大豆；有些植物以异花授粉为主繁殖后代，如玉米、高粱和黑麦；有些则同时采用自花和异花两种方式繁殖后代，如棉花和蚕豆。
- 借助人工手段，可以对无性生殖物种的不同无性系进行杂交甚至自交，对自花授粉物种的不同个体进行杂交，对异花授粉物种的不同个体进行自交。

交配制度与群体结构和育种的关系

- 物种的交配制度决定了物种的群体结构。对植物来说，也在很大程度上决定着农业生产的品种类型、以及改良该植物的适宜育种方法。
- 无性生殖植物的品种是克隆家系，来自同一个亲本的克隆家系与它们的亲本有着一模一样、但高度杂合的基因型。在任何育种群体的任何时期，任何表现优良的个体都能通过无性繁殖的方式成为一个新的无性系品种。
- 自花授粉和异花授粉的区别，主要体现在近交（inbreeding）群体和异交（outbreeding）群体的遗传结构上。

异交与群体结构和育种的关系

- 异交群体中，个体在很多座位上携带不同的等位基因，异交或自交会产生具有广泛分离的杂合后代。
- 异花授粉物种的自交后代一般会出现近交衰退（inbreeding depression）现象。
- 但是，异交群体中产生出来的两个纯合自交系之间的杂交，又能够恢复异交群体的杂合度。
- 有时，两个自交系的杂交后代会有更优异的表现，这一现象称为杂种优势（hybrid vigor或heterosis），是杂种品种选育的理论基础。

自交与群体结构和育种的关系

- 长期自交的群体中，个体的基因型基本上是纯合的，自交后代的基因型与亲代基本一致，不存在近交衰退现象。一个纯合基因型个体的自交后代，又称为一个纯系（pure line）。两个纯系间的杂交将产生出基因型高度杂合的杂种 F_1 ，对杂种 F_1 的连续自交可以产生众多不同于亲本基因型的纯系。
- 例如，两个纯系亲本在10个座位上存在差异，它们的重组近交家系后代群体中则存在 $2^{10}=1024$ 种不同的基因型，其中只有两种与亲本基因型相同，1022种都是亲本中没有的重组型。有些后代纯系会比它们的亲本有更优良的表现，这一现象称为超亲分离（transgressive segregation），是纯系品种选育的理论基础。

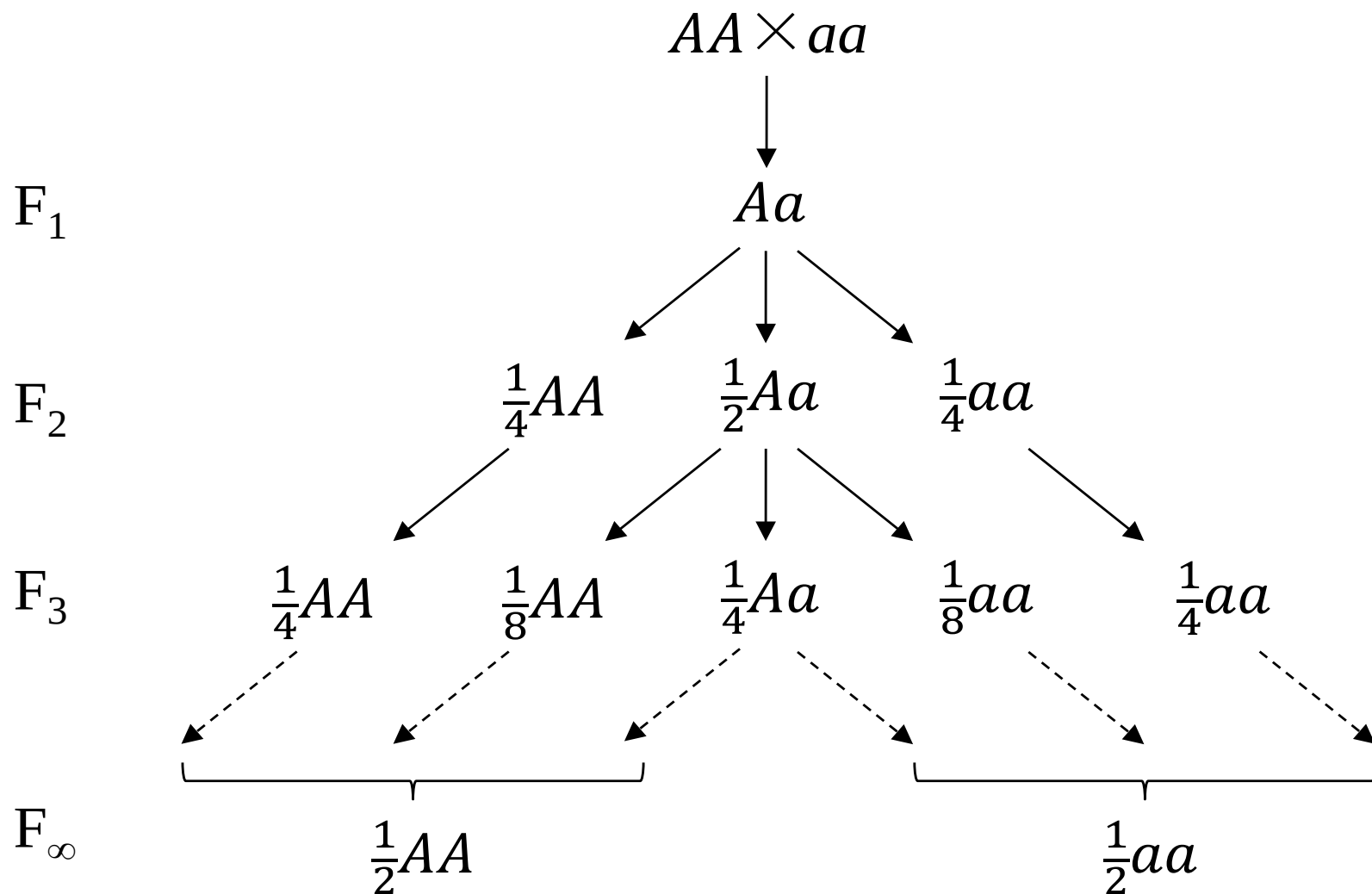
其它交配系统

- 自交和异交是两种极端类型的交配系统，异交又称随机交配。以自交系统为主的植物也可能存在一定比例的异交，以异交系统为主的植物也可能存在一定比例的自交。自交和异交的比例因物种而异，同时又受授粉时期、温度、湿度、风力等环境条件，以及传媒昆虫种群数量的影响。动物以异交为主，但有时也存在一定程度的近交。
- Wright（1921）将动物的交配系统分为随机交配、遗传同型交配、遗传非同型交配、表型同型交配和表型非同型交配5类。在第3章和第4章中我们将看到，利用近交系数（inbreeding coefficient）这一概念，可以把介于自交和随机交配之间的众多交配系统统一起来。

自交对群体结构的影响

- 就一对等位基因 A 和 a 来说，基因型 Aa 自交将产生 AA 、 Aa 和 aa 三种基因型。根据遗传学分离定律，它们在群体中所占的比例将分别为 0.25、0.5 和 0.25。
- 如继续自交，自交交配类型只有三种，即 $AA \times AA$ 、 $Aa \times Aa$ 和 $aa \times aa$ 。
 - 类型 $AA \times AA$ 占整个群体的0.25，子代基因型全部为 AA
 - 类型 $Aa \times Aa$ 占整个群体的0.5，其子代又分离出三种基因型，比例分别为 0.25、0.5 和0.25
 - 类型 $aa \times aa$ 占整个群体的0.25，子代基因型全部为 aa
- 综合起来考虑，自交二代的群体中，三种基因型 AA 、 Aa 和 aa 的比例分别为 0.375、0.25 和0.375。

一个座位上、两个等位基因 自F₁代的自交分离示意图



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
F_1	0	1	0	0.5	0.5
F_2	1/4	1/2	1/4	0.5	0.5
F_3	3/8	1/4	3/8	0.5	0.5
F_4	7/16	1/8	7/16	0.5	0.5
F_5	15/32	1/16	15/32	0.5	0.5

重复自交过程中，一对和多对独立遗传基因座位纯合基因型的百分数

世代	独立遗传基因座位个数							
	1	2	3	4	5	10	20	30
F ₂	50.00	25.00	12.50	6.25	3.13	0.10	0.00	0.00
F ₃	75.00	56.25	42.19	31.64	23.73	5.63	0.32	0.02
F ₄	87.50	76.56	66.99	58.62	51.29	26.31	6.92	1.82
F ₅	93.75	87.89	82.40	77.25	72.42	52.45	27.51	14.43
F ₆	96.88	93.85	90.91	88.07	85.32	72.80	52.99	38.58
F ₇	98.44	96.90	95.39	93.89	92.43	85.43	72.98	62.35
F ₈	99.22	98.44	97.67	96.91	96.15	92.46	85.48	79.03
F ₉	99.61	99.22	98.83	98.45	98.06	96.16	92.47	88.92
F ₁₀	99.80	99.61	99.42	99.22	99.03	98.06	96.17	94.30

回交、供体亲本、轮回亲本

- 回交是指两个亲本 P_1 和 P_2 的杂交 F_1 个体，与其中的某一亲本再进行杂交。
- 回交是植物育种的一种常用方法，尤其适用于转育由少数基因控制的一些简单性状。
- 如果 F_1 和 P_1 回交，这时亲本 P_1 称为轮回亲本，另一亲本 P_2 称为非轮回亲本或供体亲本。

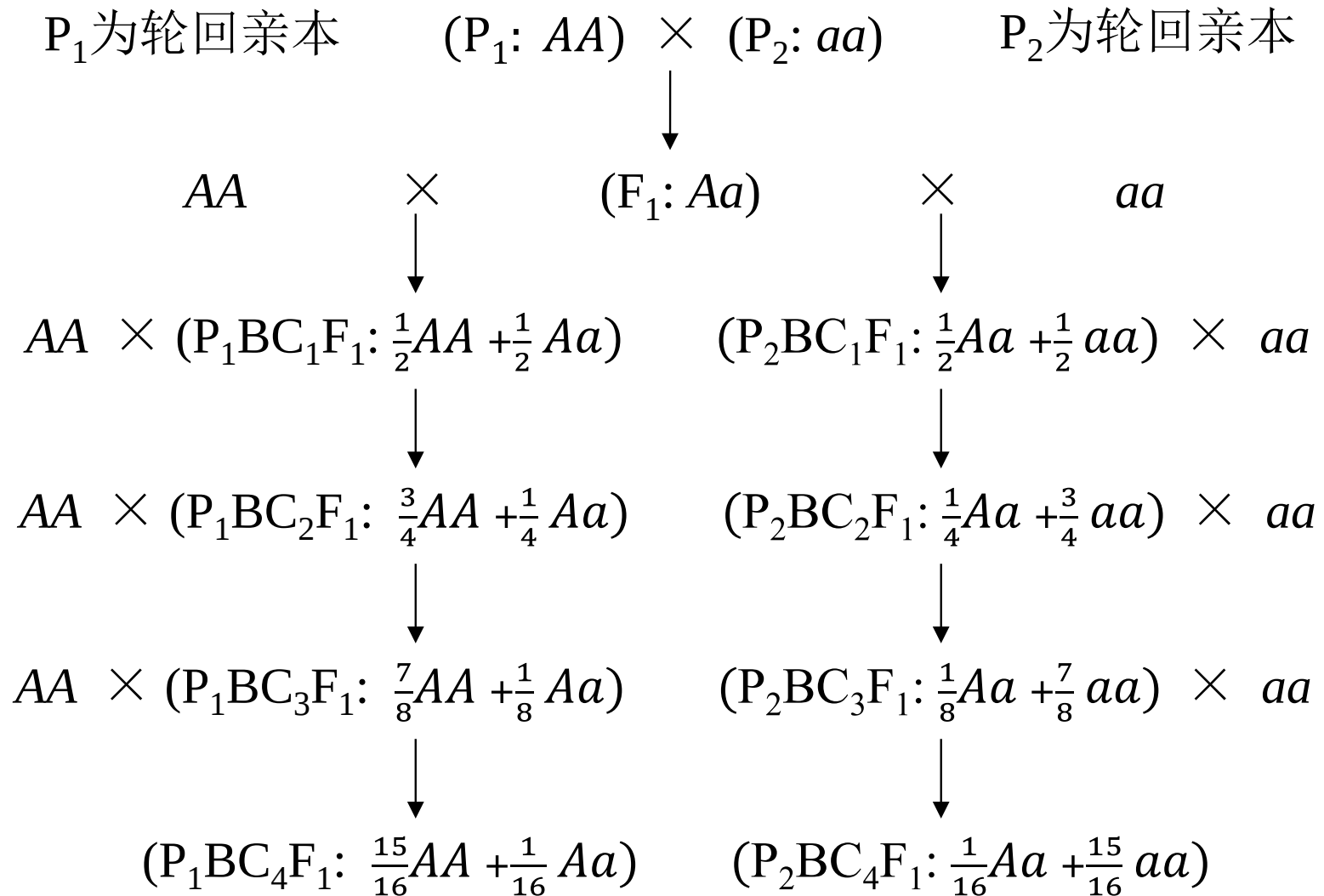
回交群体的表示和分类

- 回交群体一般用符号 BC_{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亲本的不同回交世代。在不至于混淆的情况下，也可用 BC_j 或 B_j 表示不同的回交世代。
- 如回交若干次后还需要自交，这时的回交世代可用 BC_jF_1 表示，回交之后的重复自交世代用 $BC_jF_2、BC_jF_3\dots$ 表示。
- 一次或少数几次回交有时又称为有限回交（limited backcross）或简单回交（simple backcross），多次回交有时又称高代回交（advanced backcross）或多代回交。

回交对群体结构的影响

- 以一对等位基因为例，每回交一代，杂合体频率减少一半，这一点与重复自交完全一样。
- 但是，自交群体中有两种纯合体 AA 和 aa ，而在回交世代中只有一种纯合体。与亲本 AA 的回交群体中只有纯合体 AA ，不会出现纯合体 aa ；与亲本 aa 的回交群体中只有纯合体 aa ，不会出现纯合体 AA 。
- 因此，在回交系统中，回交后代恢复成轮回亲本的纯合速度要比在自交系统中快。

回交系统中基因型频率的变化



杂合子 Aa 与亲本 AA 回交后, 各世代基因和基因型频率的变化

回交世代	基因型			基因	
	AA	Aa	aa	A	a
F1	0	1	0	0.5	0.5
BC1	1/2	1/2	0	0.75	0.25
BC2	3/4	1/4	0	0.875	0.125
BC3	7/8	1/8	0	0.9375	0.0625
BC4	15/16	1/16	0	0.96875	0.03125

无选择情况下，排除不良等位基因***b*** 连锁于优良等位基因***A***的概率

重组率	一次回交	二次回交	三次回交	4次回交	5次回交	自交
0.5	0.75	0.875	0.9375	0.9688	0.9844	0.5
0.2	0.36	0.488	0.5904	0.6723	0.7379	0.2
0.1	0.19	0.271	0.3439	0.4095	0.4686	0.1
0.02	0.0396	0.0588	0.0776	0.0961	0.1142	0.02
0.01	0.0199	0.0297	0.0394	0.0490	0.0585	0.01

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

§ 1.4 随机交配与随机交配群体

- § 1.4.1 随机交配的定义
- § 1.4.2 Hardy-Weinberg平衡定律
- § 1.4.3 Hardy-Weinberg平衡群体的应用
- § 1.4.4 Hardy-Weinberg平衡的检验
- § 1.4.5 复等位基因

随机交配的定义

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

父本基因型和频率	母本基因型和频率		
	AA, D	Aa, H	aa, R
AA, D	D^2	DH	DR
Aa, H	HD	H^2	HR
aa, R	RD	RH	R^2

两个群体间的随机交配

群体1的基因型和频率	群体2的基因型和频率		
	$AA: D_2$	$Aa: H_2$	$aa: R_2$
	$AA \times AA: D_1D_2$	$AA \times Aa: D_1H_2$	$AA \times aa: D_1R_2$
	$Aa \times AA: H_1D_2$	$Aa \times Aa: H_1H_2$	$Aa \times aa: H_1R_2$
$aa: R_1$	$aa \times AA: R_1D_2$	$aa \times Aa: R_1H_2$	$aa \times aa: R_1R_2$

结合类型	$AA \times AA$	$AA \times Aa$	$AA \times aa$	$Aa \times Aa$	$Aa \times aa$	$aa \times aa$
概率	D_1D_2	$D_1H_2 + H_1D_2$	$D_1R_2 + R_1D_2$	H_1H_2	$H_1R_2 + R_1H_2$	R_1R_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 对应于二项式 $(p+q)^2$ 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	0.25
Aa × aa	$4pq^3$	0	0.5	0.5
aa × aa	q^4	0	0	1
合计 (加权)	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	0.25
Aa × aa	$2HR$	0	0.5	0.5
aa × aa	R^2	0	0	1
合计	1	$(D+0.5H)^2$	$2(D+0.5H)(0.5H+R)$	$(0.5H+R)^2$

Hardy-Weinberg平衡定律

- 在随机交配群体中，两个等位基因A和a的频率分别为 p 和 q ，如果没有干扰因素，三种基因型AA、Aa、aa的频率总是 p^2 、 $2pq$ 、 q^2 ，也就是说基因频率和基因型频率将永远保持不变，群体的遗传结构处于一种平衡状态。
- 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。

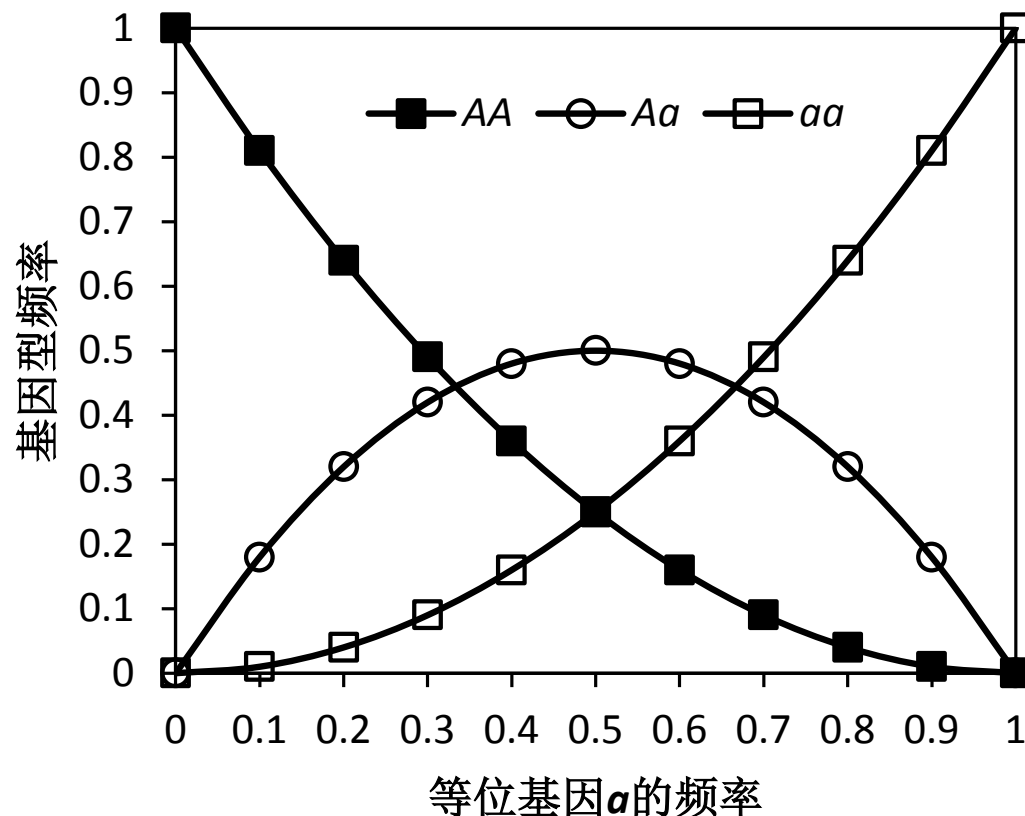
Hardy-Weinberg平衡群体的性质1

- 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

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

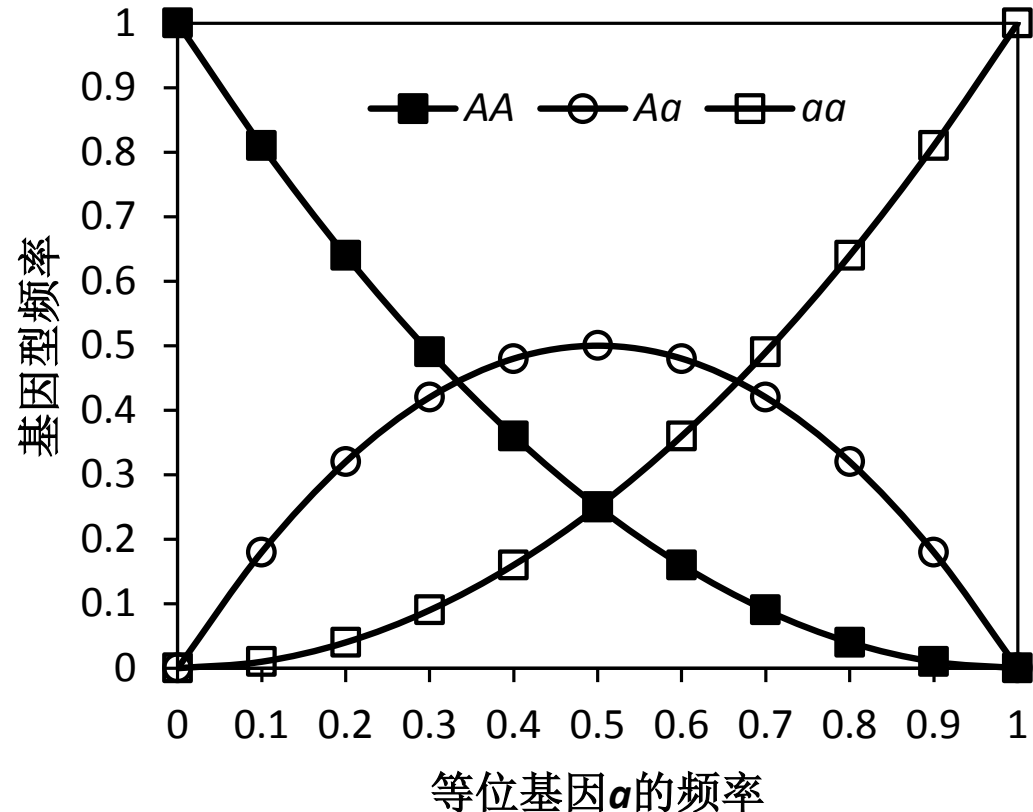
$$\frac{dH}{dp} = 2 - 4p$$

- H在 $p=0.5$ 处达到极大值0.5



Hardy-Weinberg平衡群体的性质2

- 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
 $p=2/3$, $q=1/3$

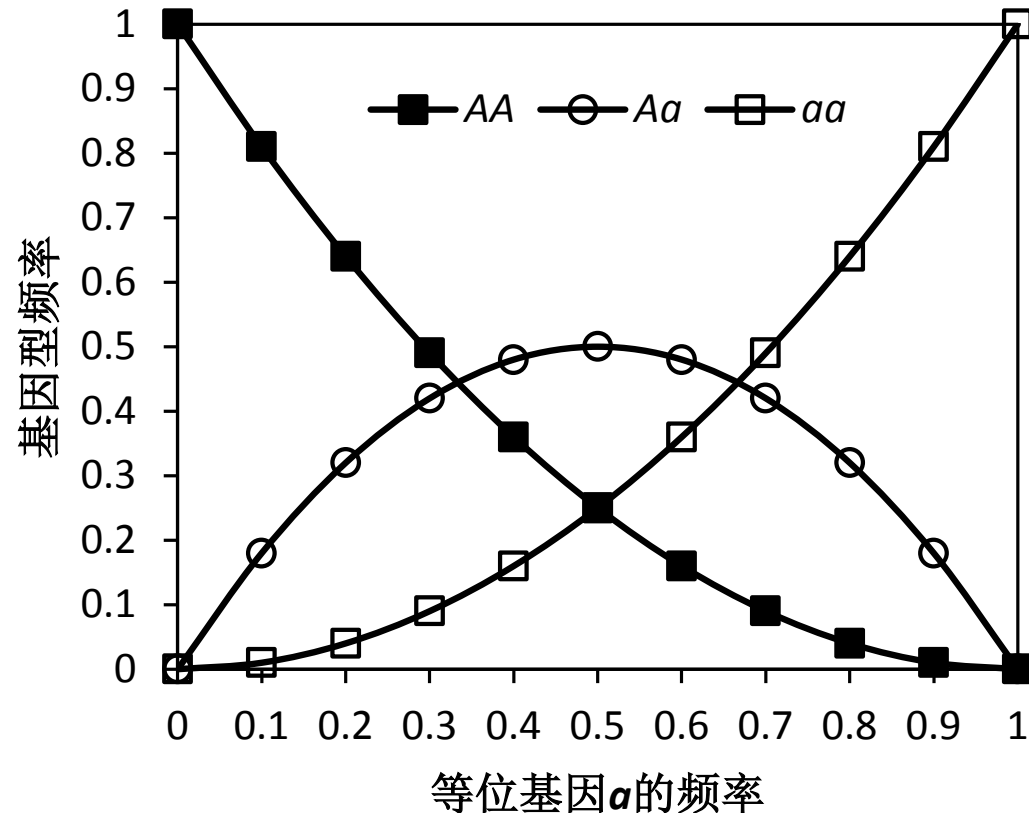


Hardy-Weinberg平衡群体的性质3

- 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

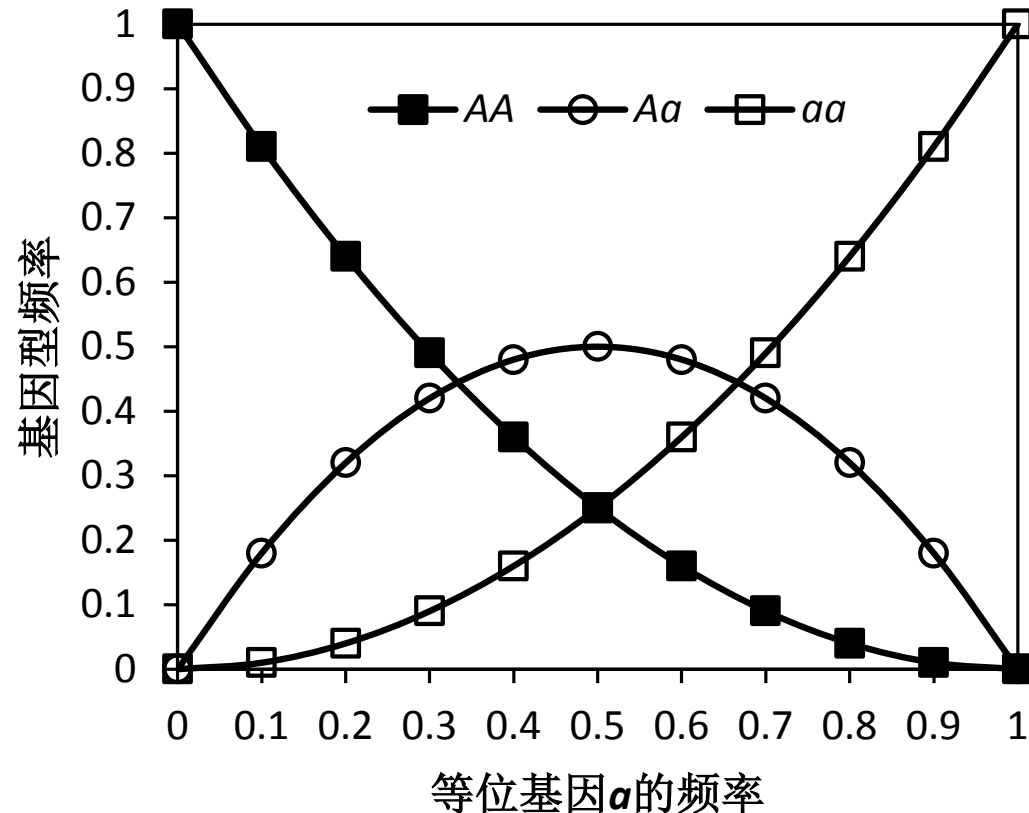
$$\frac{H}{2 \times R} = \frac{2pq}{2q^2} = \frac{p}{q}$$

- 因此, 对于很小的q, p/q会非常大



Hardy-Weinberg平衡定律的作用1

- 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。



Hardy-Weinberg平衡定律的作用2

- 计算隐性基因控制的性状上，正常表型中杂合体（或携带者）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2

- 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%

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 $(MM, MN, NN) = (p^2, 2pq, q^2)$

基因型	MM	MN	NN	Total
观测值	$N_1=233$	$N_2=385$	$N_3=129$	$N=747$
HW平衡频率	0.3245	0.4903	0.1852	1
期望值	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

F_1 自交与 F_1 随机交配是等价的, 因此 F_2 是一个HWE群体

F_1 群体中, AA: $D=0$; Aa: $H=1$; aa: $R=0$; $p=q=0.5$

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

- F_2 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

$$(p_1A_1 + p_2A_2 + p_3A_3 + \dots)^2 = p_1^2A_1A_1 + 2p_1p_2A_1A_2 + 2p_1p_3A_1A_3 + \dots$$

- 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, Y染色体十分保守)

- 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 (♂)

§ 1.5 连锁对群体结构的影响

- § 1.5.1 杂种F1产生配子的连锁不平衡
- § 1.5.2 随机交配群体中的连锁不平衡
- § 1.5.3 两个座位间不平衡的显著性检验
- § 1.5.4 多代随机交配与连锁不平衡
- § 1.5.5 群体结构与连锁不平衡
- § 1.5.6 多代随机交配后的累积重组率

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)

等位基因和频率	A, p_A	a, p_a	B, p_B	b, p_b
配子型(单倍型)	AB	Ab	aB	ab
平衡频率	$p_A p_B$	$p_A p_b$	$p_a p_B$	$p_a p_b$
观测频率	u	s	t	v
观测频率与平衡频率之差	+D	-D	-D	+D

- 基因频率: $p_A = u + s$; $p_a = t + v$; $p_B = u + t$; $p_b = s + v$.
- 对于AB来说, $p(AB) - p(A) p(B) = u - (u + s)(u + t) = u(1 - u - s - t) - st = uv - st$

单倍体观测频率的另一种表示

等位基因和频率	A, p_A	a, p_a	B, p_B	b, p_b
配子型 (单倍型)	AB	Ab	aB	ab
平衡频率	$p_A p_B$	$p_A p_b$	$p_a p_B$	$p_a p_b$
观测频率与平衡频率之差	+D	-D	-D	+D
观测频率	$p_A p_B + D$	$p_A p_b - D$	$p_a p_B - D$	$p_a p_b + D$

连锁不平衡的度量(D)

等位基因和频率	A, p_A	a, p_a	B, p_B	b, p_b
配子型(单倍型)	AB	Ab	aB	ab
平衡频率	$p_A p_B$	$p_A p_b$	$p_a p_B$	$p_a p_b$
观测频率	u	s	t	v
观测频率与平衡频率之差	$+D$	$-D$	$-D$	$+D$

$$D = uv - st$$

F₂群体中的连锁不平衡度

(AABB × aabb) F₁: AB/ab

配子型(单倍型)	AB	Ab	aB	ab
平衡频率	0.25	0.25	0.25	0.25
期望频率	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, $D_{\max}=0.25$

连锁不平衡的其它度量方法

- 当 $uv-st>0$ 时, $D_{\max} = \min\{p_A p_b, p_a p_B\}$
- 当 $uv-st<0$ 时, $D_{\max} = \min\{p_A p_B, p_a p_b\}$
- 标准化不平衡度 D' 定义为:

$$D' = \frac{D}{D_{\max}}$$

连锁不平衡的其它度量方法

- 另一种不平衡度的度量方法是将不平衡度平方后除以四种等位基因频率的乘积，即

$$r^2 = \frac{D^2}{p_A p_a p_B p_b}$$

- 从练习1.17可以知道，上面不平衡度的平方根正好等于两个座位间的相关系数，上面的不平衡度乘以总样本量正好等于不平衡检验的 χ^2 统计量。
- 因此，不平衡度参数 r^2 在自然群体的遗传分析中有更广泛的应用。

不同度量方法的异同

- 前面的三种不平衡度参数，各自从不同侧面反映座位间的不平衡程度，难以相互取代。
- 后两者参数的目的是消除基因频率对不平衡度的影响。但是，不平衡度和基因频率都是从配子观测频率计算出来的，这度量方法有时也难以完全排除等位基因频率的影响。
- 此外，这些参数都不能用来判断不平衡度的显著性，不平衡度的显著程度仍然要借助于统计假设检验。

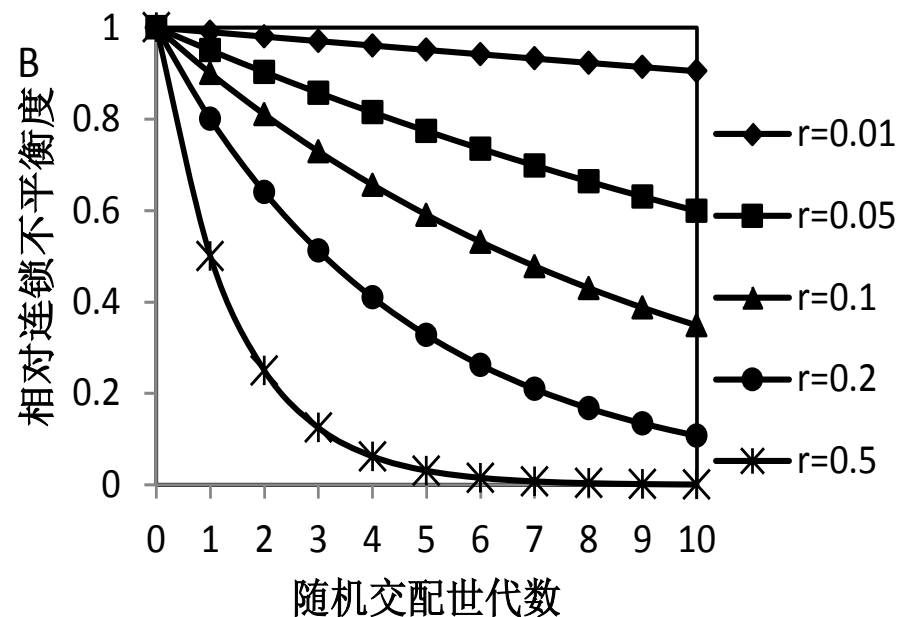
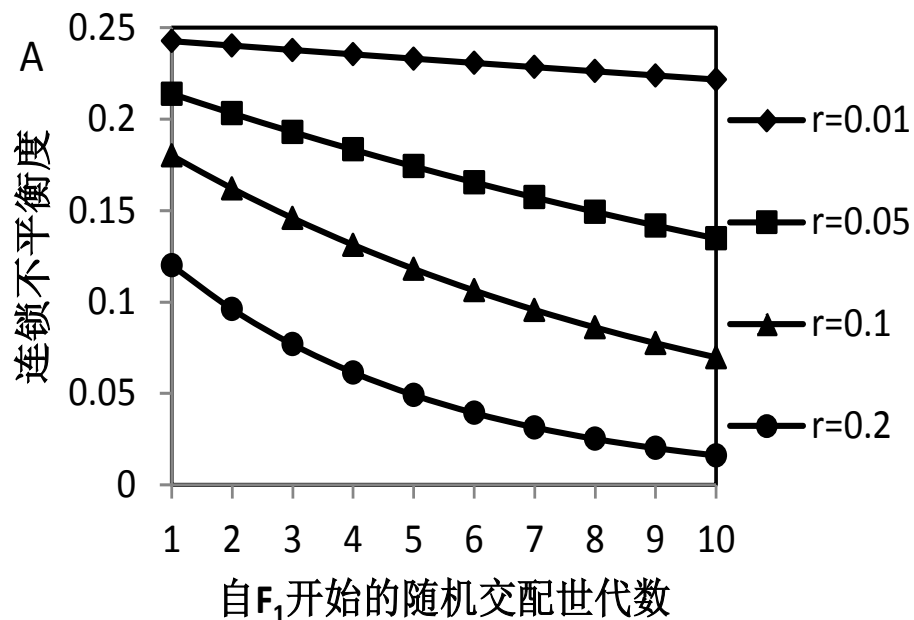
多代随机交配过程中的连锁不平衡

等位基因和频率	A, p_A	a, p_a	B, p_B	b, p_b
配子型 (单倍型)	AB	Ab	aB	ab
平衡频率	$p_A p_B$	$p_A p_b$	$p_a p_B$	$p_a p_b$
观测频率	u	s	t	v

- 以配子AB为例, 有两种产生配子AB的途径
 - 由基因型AB/XY ($X=A$ 或 a , $Y=B$ 或 b) 不发生重组产生, 这一事件发生的频率为 $u(1-r)$
 - 由基因型AY/XB ($X=A$ 或 a , $Y=B$ 或 b)发生重组产生, 这一事件发生的频率为 $p_A p_B r$
- 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

多代随机交配后的连锁不平衡度

- 随机交配过程中，交换和重组的不断发生，使得座位间的不平衡程度逐渐下降。对于连锁非常紧密、一次随机交配难以看到重组的座位来说，有可能通过多代随机交配检测出重组的存在，这样就不至于把紧密连锁的座位当作一个座位。
- F_1 : AB/ab 连续随机交配， t 个世代后的连锁不平衡度为：

$$D_t = D_0(1-r)^t = \frac{1}{4}(1-2r)(1-r)^t$$

多代随机交配后的累积重组率

- F_1 : AB/ab 连续随机交配, t 个世代后的累积重组率用 R 表示。令:

$$D_t = \frac{1}{4}(1-2r)(1-r)^t = \frac{1}{4}(1-2R)$$

- 得到:

$$R = \frac{1}{2} - 2D_t = \frac{1}{2} - \frac{1}{2}(1-2r)(1-r)^t$$

随机交配多代的累积遗传图距

随机交配代数	F ₂ 群体（等价于随机交配一代）的遗传图距/cM					
	0.1	0.2	1	2	5	10
2	0.15	0.30	1.50	2.99	7.44	14.75
3	0.20	0.40	2.00	3.98	9.88	19.50
4	0.25	0.50	2.49	4.97	12.31	24.25
5	0.30	0.60	2.99	5.96	14.75	29.00
6	0.35	0.70	3.49	6.95	17.19	33.75
7	0.40	0.80	3.99	7.94	19.63	38.50
8	0.45	0.90	4.48	8.93	22.06	43.25
9	0.50	1.00	4.98	9.92	24.50	48.00
10	0.55	1.10	5.48	10.91	26.94	52.75
多代重复自交	0.20	0.40	1.98	3.92	9.55	18.33

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!**

