



# 遗传学 (第3版)

## 第6章 数量性状遗传分析

1. 数量性状及其多基因学说
2. 数量性状遗传分析的统计学基础
3. 数量性状基因座及其作图
4. 数量性状遗传率及计算方法
5. 近亲繁殖与杂种优势





# 6.1 数量性状及其多基因学说

## 6.1.1 数量性状的概念

生物的性状分为两类：

质量性状 (qualitative character)

数量性状 (quantitative character 或 quantitative trait, QT)

所有能够度量的性状都可称为数量性状

数量遗传学 (Quantitative Genetics)

生物统计遗传学 (Biometrical Genetics)

统计遗传学 (Statistical Genetics)





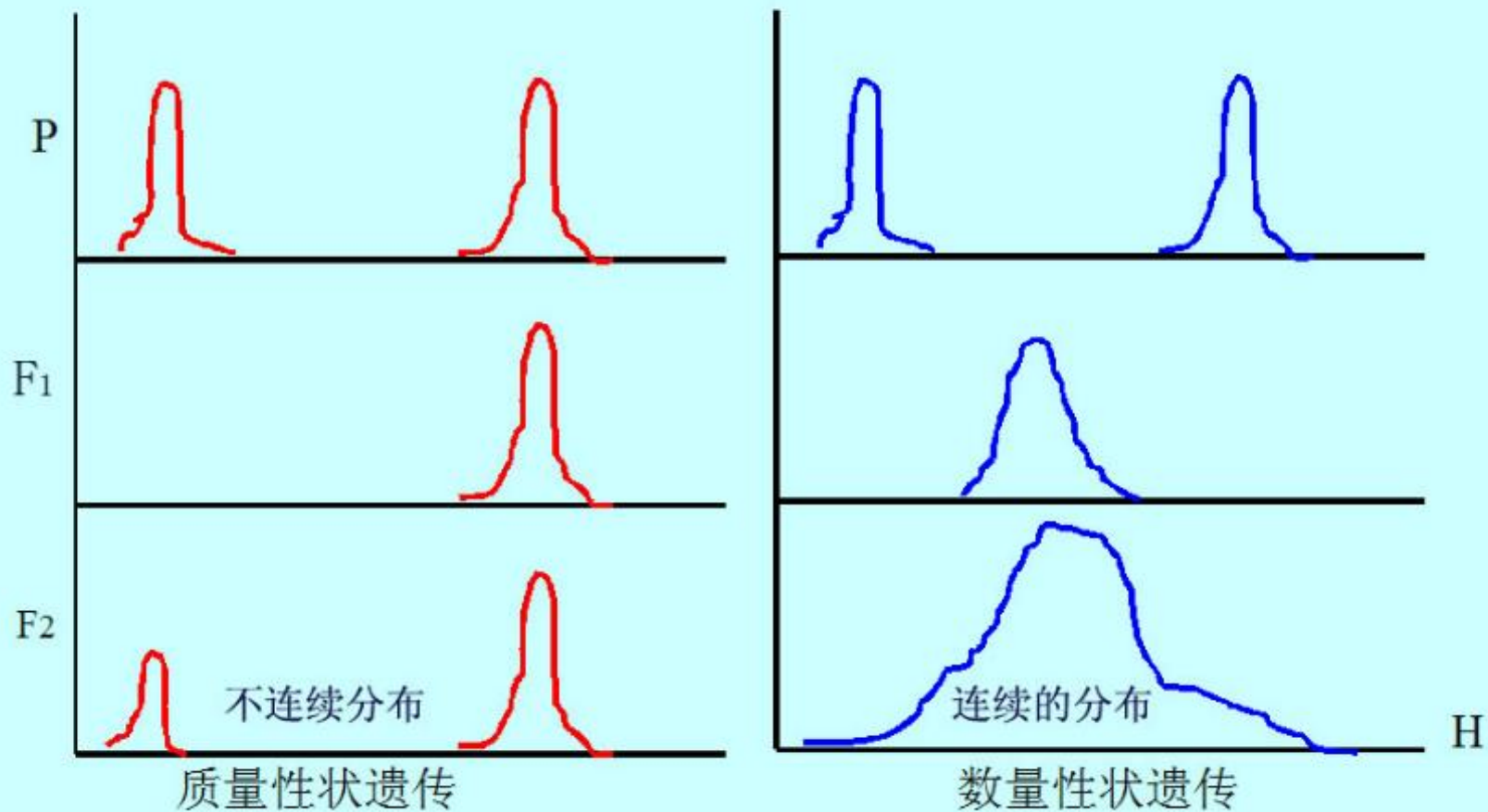
## (1) 质量性状与数量性状

- **质量性状**：表型之间截然不同，具有质的差别，用文字描述的性状。如水稻的糯与粳，人的A、B、O血型，豌豆的圆皱等性状。
- **数量性状**：性状之间呈连续变异状态，界限不清楚，不易分类，用数字描述的性状。如作物的株高，奶牛的泌乳量，棉花的纤维长度等。



## 质量性状与数量性状的比较

质量性状		数量性状
性状主要类型	品种特征、外貌特征	生产、生长性状
遗传基础	少数主基因控制	微效多基因系统控制
	遗传关系较简单	遗传关系复杂
变异表现方式	间断型	连续型
考察方式	描 述	度 量
环境影响	不敏感	敏 感
研究水平	家 庭	群 体
研究方法	系谱分析、概率论	生物统计



质量性状遗传与数量性状遗传的区别



数量性状包括两大类：

一是表现**连续变异的性状**，如牛的泌乳量、农作物的产量、棉花纤维、羊毛的长度等等；

二是表型**呈非连续变异**，而遗传物质的数量呈潜在的**连续变异的性状**，即只有超越某一遗传阈值时才出现的性状，如动、植物甚至包括人类的抗病力、死亡率以及单胎动物的产仔数等性状，称为**阈性状**（threshold character或threshold trait）。

数量性状表型的连续性特点：

第一，**一种基因型影响一组表型**的表现。其结果模糊了基因型所决定的不同表型之间的差异，因而**不能将一个特定的表型归属于一个特定的基因型**。

第二，许多不同基因座的等位基因都能使某一种被观察的表型发生改变。

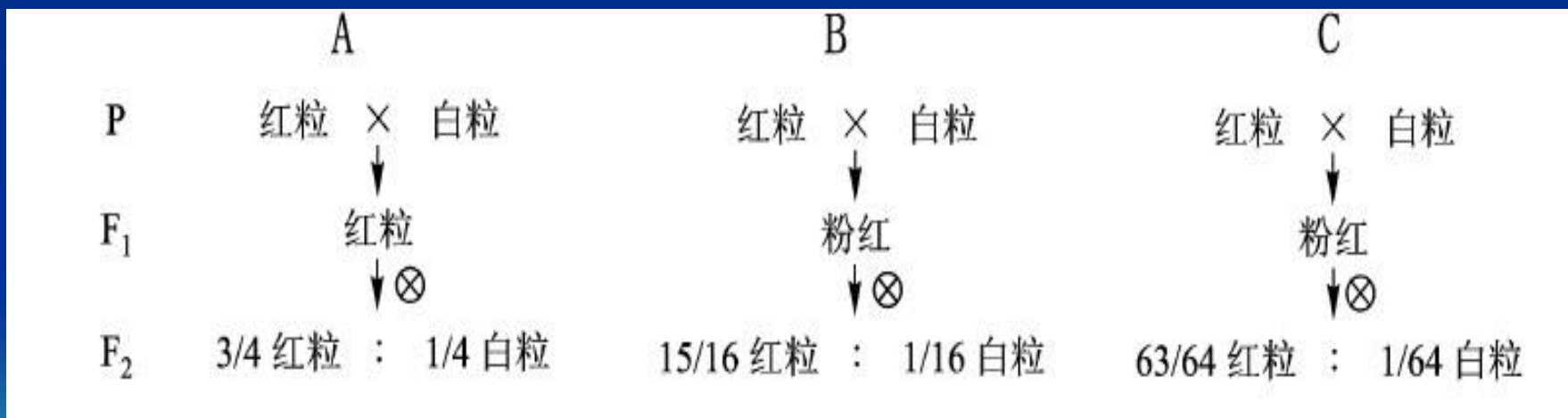




# 6.1.2 数量性状的多基因学说

## (1) 实验依据

1909年，瑞典遗传学家Nilsson-Ehle对小麦和燕麦中籽粒颜色的遗传进行了研究，他发现在若干个红粒与白粒的杂交组合中有如下A、B、C 3种情况：





## 他研究后进一步发现：

① 在小麦和燕麦中，有3对与种皮颜色有关的、种类不同但作用相同的基因，这3对基因中的任何一对在单独分离时都出现 $3/4:1/4$ 的比率，而3对基因同时分离时，则产生 $63/64:1/64$ 的比率。

② 上述的杂交在 $F_2$ 的红色籽粒中又呈现各种程度的差异，按红色的程度又可人为地分为：

在A中： $1/4$  红粒： $2/4$  中等红： $1/4$  白色；

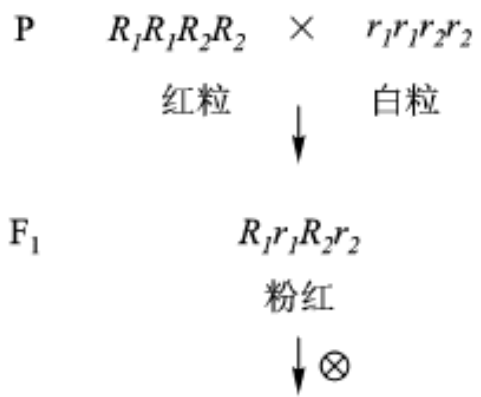
在B中： $1/16$  深红： $4/16$  红： $6/16$  中等红： $4/16$  淡红： $1/16$  白色

在C中： $1/64$  极深红： $6/64$  深红： $15/64$  次深红：

$20/64$  中等红： $15/64$  中淡红： $6/64$  淡红： $1/64$  白色

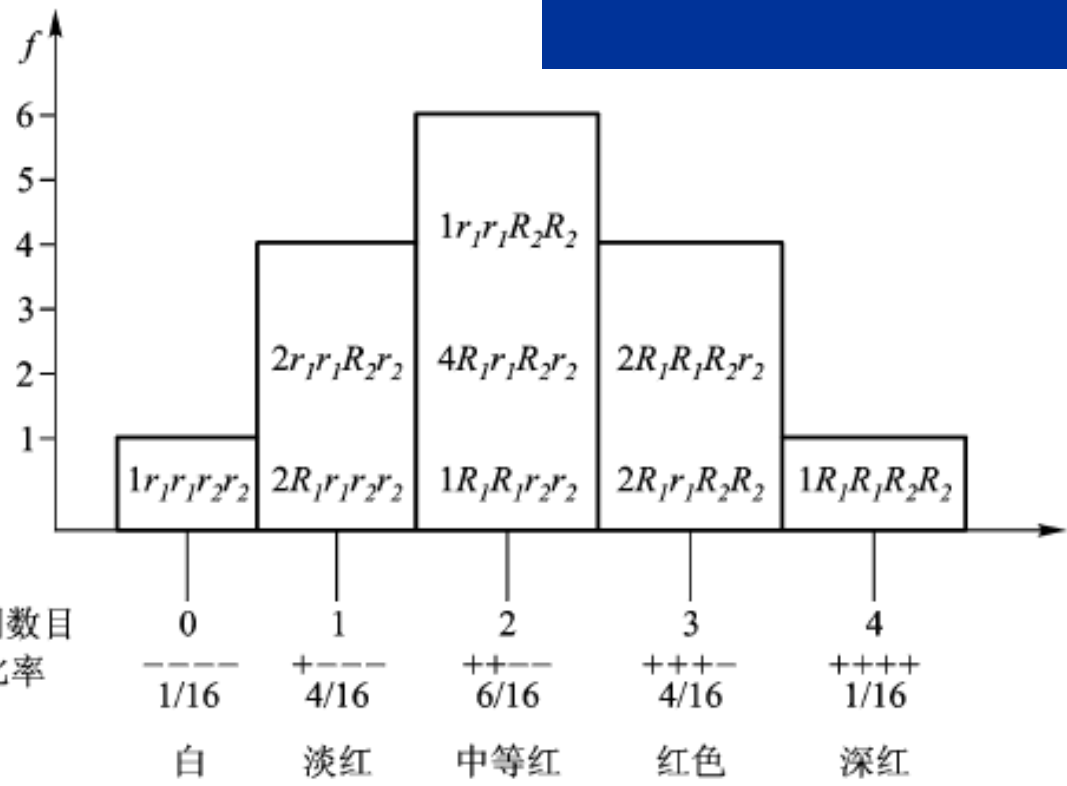
③ 红色籽粒深浅程度的差异与所具有的决定“红色”的基因数目有关，而与基因的种类无关。现以B组实验为例，说明种皮颜色的深浅程度与基因数目的关系。





设： $R_1 r_1$  及  $R_2 r_2$  为两对决定种皮颜色的基因，大写字母表示“增加”红色，小写字母表示“不增加”红色， $R$ 与 $r$ 不存在显隐性关系。

F<sub>2</sub> 基因型及其比率



假设每一列（直行）个体的表型相同，得到表型分布结果为1: 4: 6: 4: 1，该分布的各项系数是根据数学中二项式分布的基本原理推演的“杨辉三角”中遗漏单行，取其双数行得到的





随后美国学者Edward Murray East

进行烟草 (*Nicotiana longiflora*) 花冠长度的遗传学研究。将花冠的平均长度为 40.5 mm 和 93.3 mm 的纯系亲本进行杂交 → F<sub>1</sub> 呈中等长度，如所预期的一致，但长度稍有变异，这是由环境的变化所引起的。

花冠长度的遗传若由 4 对基因控制，则预期 F<sub>2</sub> 中落在每一亲本类型中的植株的表型频率为  $(1/2)^8 = 1/256$

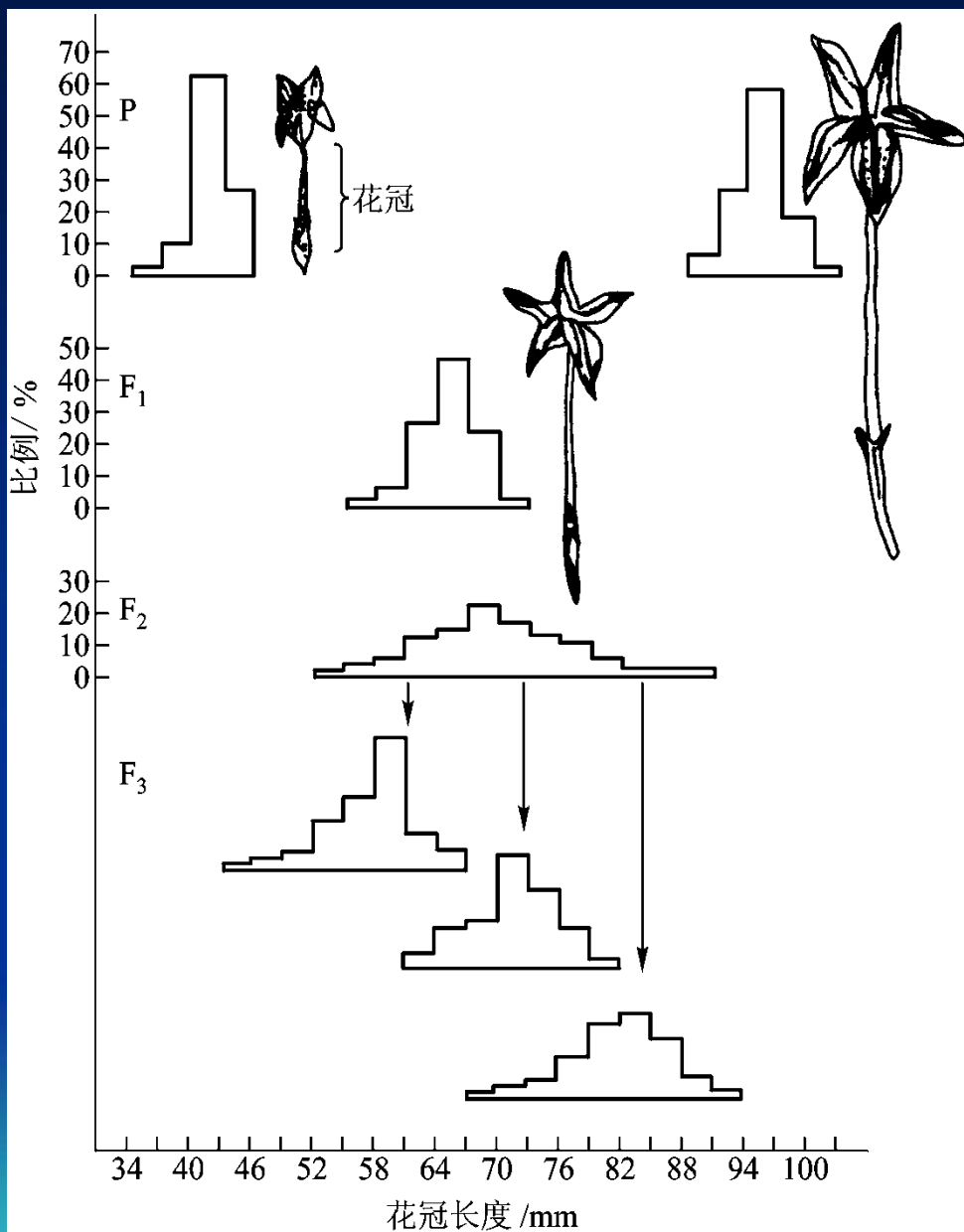
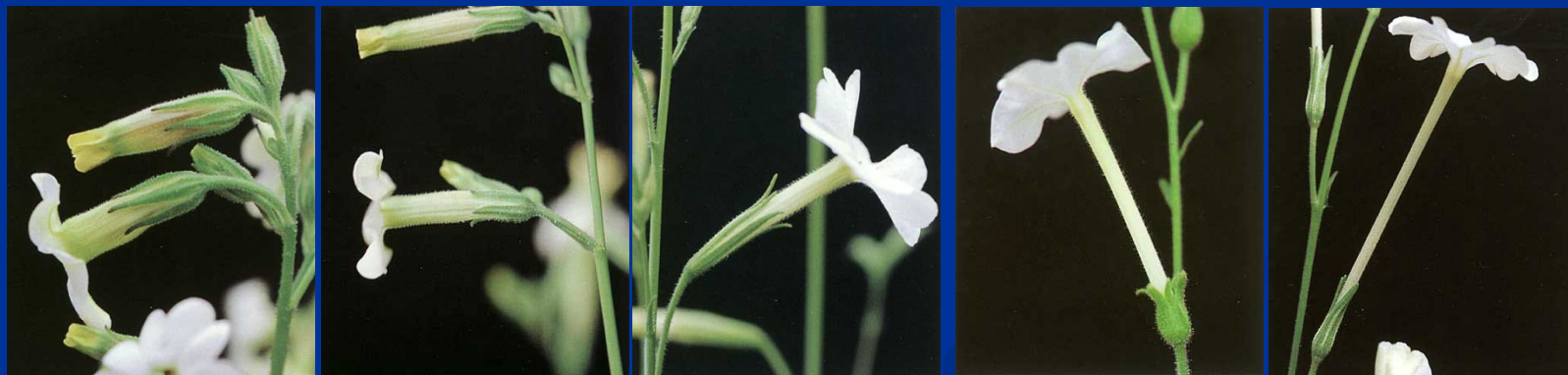


图6-1 烟草花冠长度的遗传分析

(1) East's Experiment, 1916

长花烟草: *Nicotiana longiflora*

Trait: 花冠 (corolla) 长度



Strain A: 37-43 mm

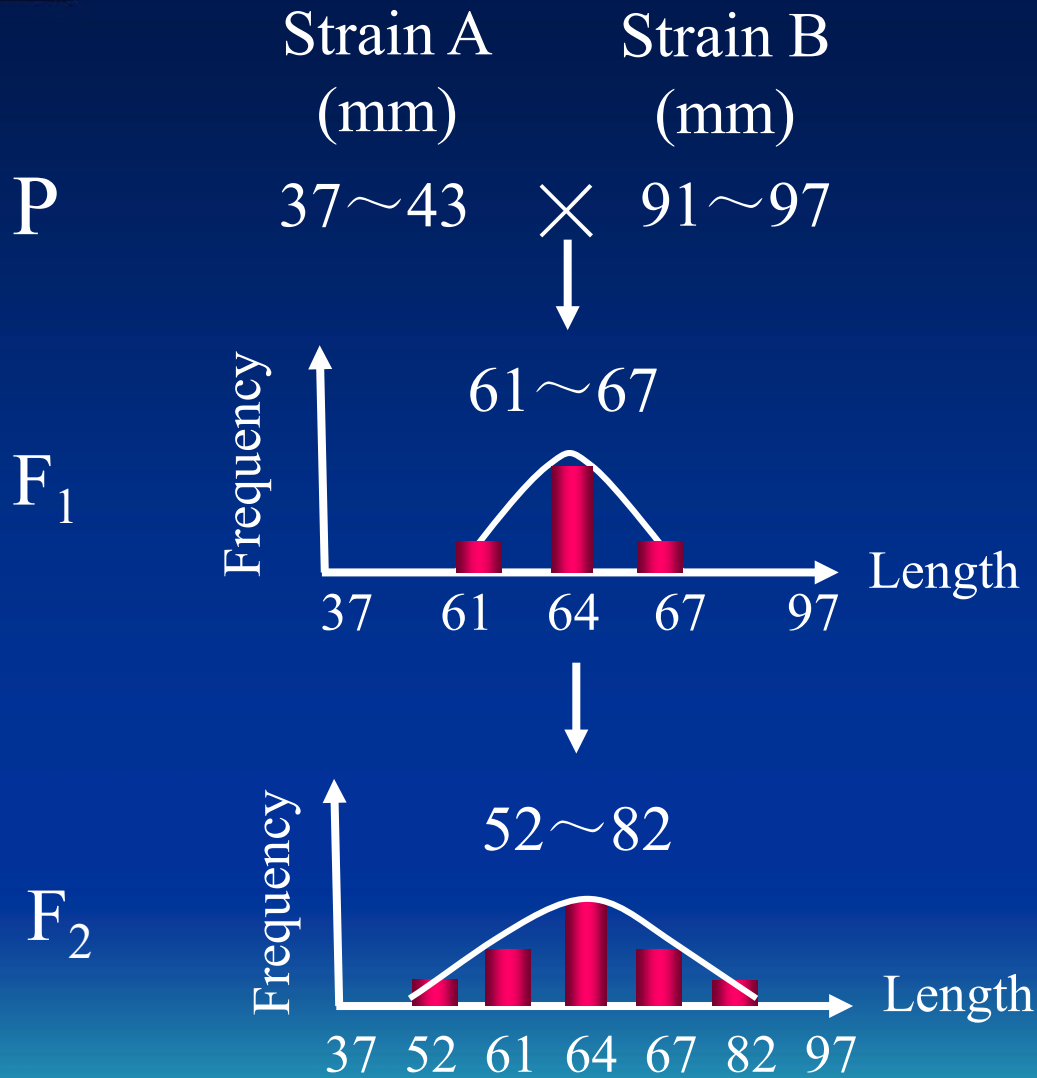
Strain B: 91-97 mm



武汉大学

Wuhan University

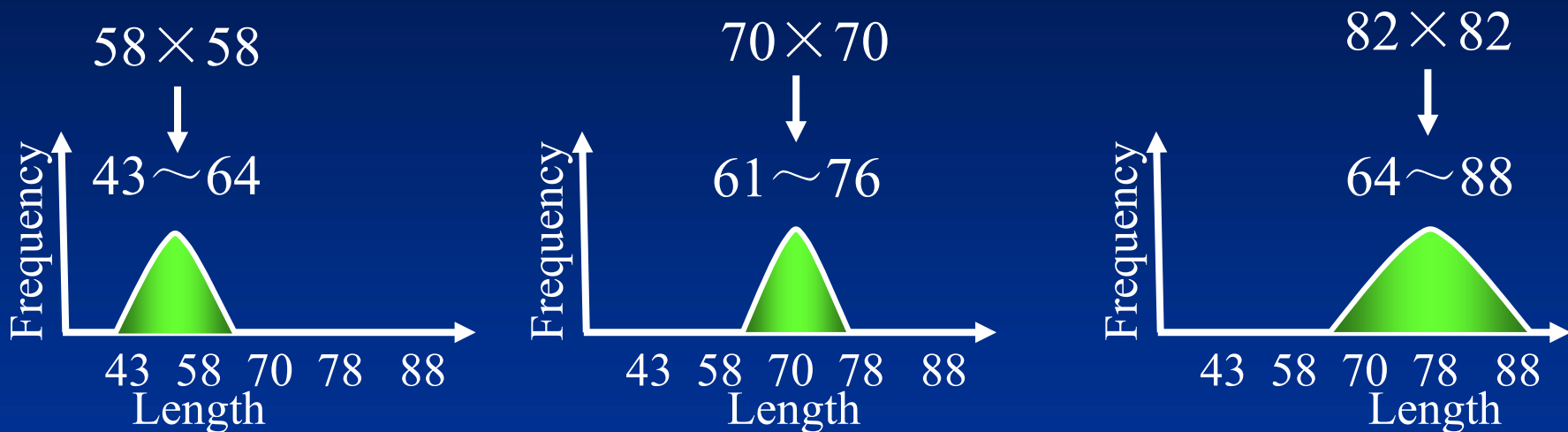
# Diagram



F<sub>2</sub> 比 F<sub>1</sub> 变异范围大 (由于基因型和环境影响)



# 分组 $F_2 \times F_2$ 杂交

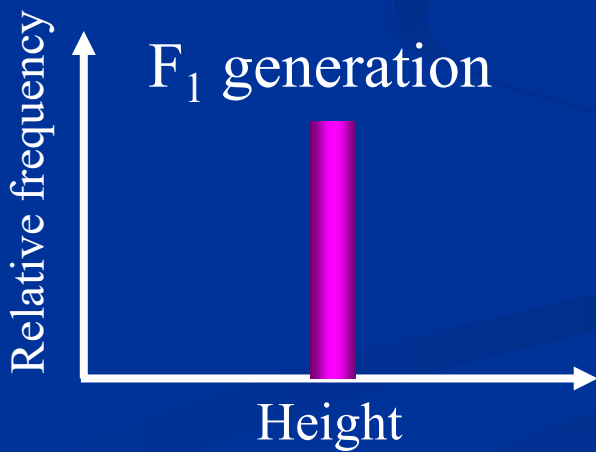
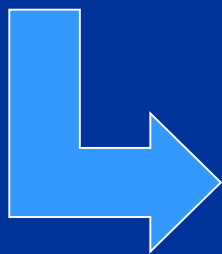
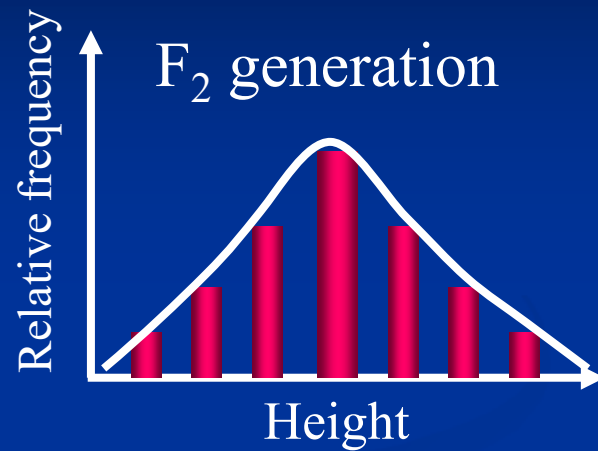
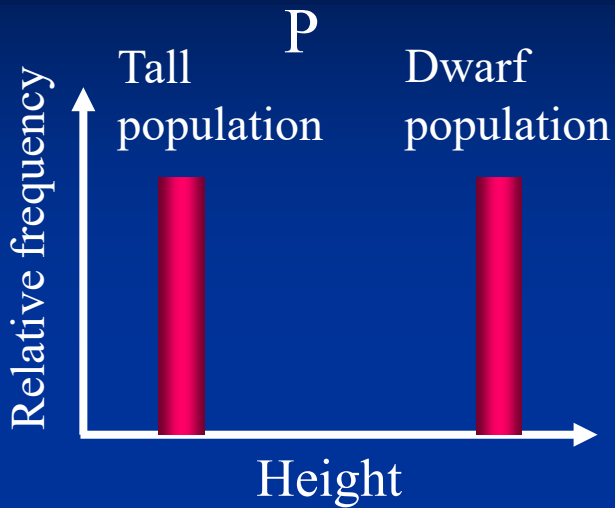


三组独立的 $F_3$  花冠长度

$F_3$  仍然是 $F_2$  亲本表型的中间类型，变异大



# Diagram







# East 杂交实验结果的特点

## 1) $F_1$ phenotype

中等长度

变异少

无亲本型

## 2) $F_2$ phenotype

大部分还是中等长度

变异多

有亲本型，但很少

整体符合正态分布 (Normal distribution)



## (2) 多基因假说的要点

Nilsson-Ehle总结了其实验分析的结果，提出了数量性状遗传的多基因学说。后又经统计学家Fisher及East等在玉米、烟草等植物的数量性状遗传的研究中进一步证明和完善，最终形成了解释和分析数量性状遗传的基因理论。其要点是：

① 数量性状是许多对微效基因（minor gene）或多基因（polygene）的联合效应共同作用的结果。

② 多基因中的每一对基因对性状表现的表现所产生的效应是微小的。多基因不能予以个别辨认，只能按性状的表现一并研究。

③ 微效基因的效应是相等，而且相互累加。

（可称多基因为加性基因——additive gene）

④ 微效基因之间一般不存在显隐性关系。

（通常用大写拉丁字母表示增效，小写字母表示减效）



⑤ **微效基因对环境敏感**，数量性状的表现容易受环境因素的影响而发生较大变化，难以识别个别基因的作用。

⑥ **多基因往往有多效性**。多基因一方面对于某种数量性状起微效基因的作用，同时在其他性状上可以作为修饰基因（改变其他基因效果的基因）而起作用，使之成为其他基因表现的遗传背景。

⑦ 多基因与主效基因（major gene）一样都由染色体所携带，并同样具有分离、重组、连锁等性质。在个别情况下，多基因的效应不是累加而是累积的（如果实的体积），有时，某几对基因间也表现显性，以致表型分布呈现偏态。

**发展：**数量性状的深入研究**进一步丰富**了多基因假说，**如存在主效基因**、基因效应大小可以不同、基因间存在上位性效应等。

多基因假说阐明了数量性状遗传的基本原因，为数量性状的遗传分析奠定了理论基础



数量性状不同于质量性状，在研究方法上有下列特点：

- ① 在杂交后代中，个别或少数后裔所能提供的信息量很少。  
研究的单位必须扩大到群体和许多世系才可能获得对其遗传规律和动态变化的认识。
- ② 对个体的性状进行测量或称重，在阈性状方面则计数。
- ③ 利用生物统计学的方法，计算性状的表型参数：平均数、方差、协方差、相关系数、回归系数等等。在此基础上进而计算遗传参数：遗传率、遗传相关系数等。



以假想的玉米穗长的遗传模式来直观地说明这一假说：

(1) 如果两亲本相差一对基因

P:	aa (6cm)	×	AA (18cm)
		↓	
F <sub>1</sub>	Aa (12cm)		
		↓	
F <sub>2</sub>	1aa	:	2Aa
		:	:
		:	1AA
频率	1/4	:	2/4
		:	:
		:	1/4
穗长	6cm	:	12cm
		:	:
		:	18cm

增加一个A，就相当于在短穗亲本的基础上增加6cm

(2) 假设该性状由三对等位基因 (A<sub>1</sub>a<sub>1</sub>, A<sub>2</sub>a<sub>2</sub>和A<sub>3</sub>a<sub>3</sub>) 控制，依据多基因假说，等位基因间无显性效应，非等位基因间无上位效应，基因的效应相同且可加

如果A<sub>1</sub>A<sub>1</sub>A<sub>2</sub>A<sub>2</sub>A<sub>3</sub>A<sub>3</sub>=18cm,     a<sub>1</sub>a<sub>1</sub>a<sub>2</sub>a<sub>2</sub>a<sub>3</sub>a<sub>3</sub>=12cm,

可知一个A基因的效应值 (A<sub>1</sub>=A<sub>2</sub>=A<sub>3</sub>) 为3cm,     (6A=18)

一个a基因的效应值 (a<sub>1</sub>=a<sub>2</sub>=a<sub>3</sub>) 为2cm。     (6a=12)

因此，每用一个a基因替换一个A基因，穗长将减少1cm



忽略环境效应的影响，如下杂交试验的结果将是：

$$P: \quad A_1A_1A_2A_2A_3A_3 \quad \times \quad a_1a_1a_2a_2a_3a_3$$

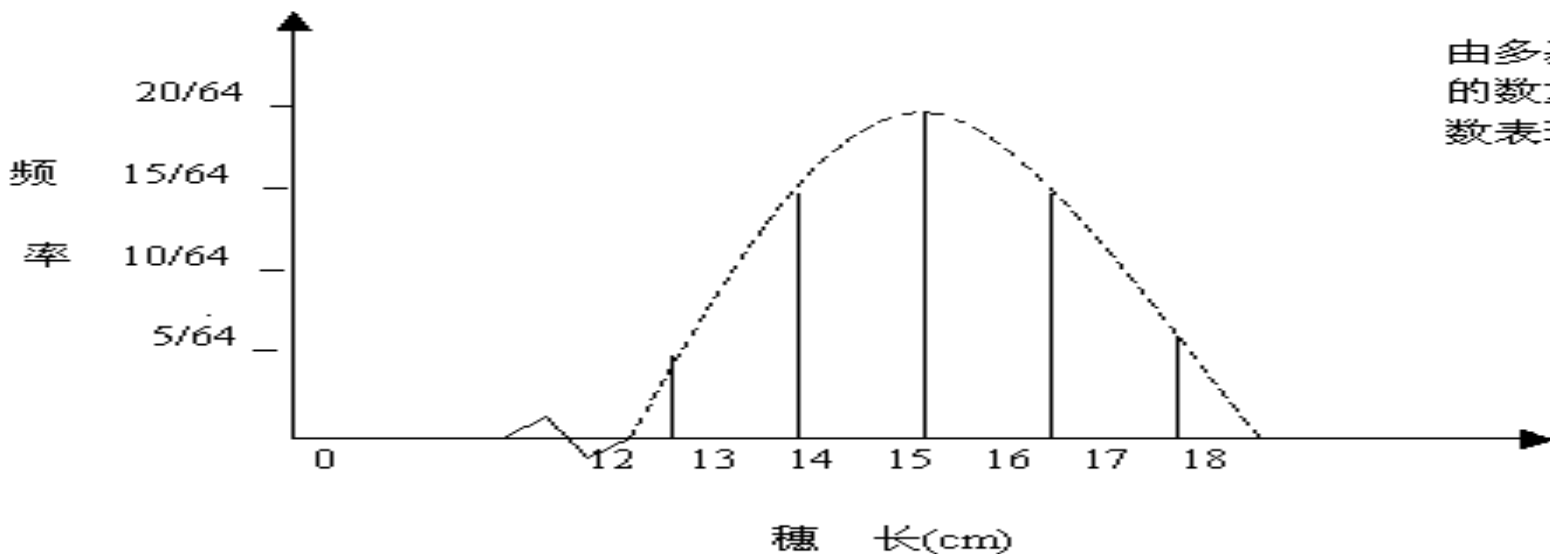
$$F_1: \quad A_1a_1A_2a_2A_3a_3$$

$$F_2: \quad 6A : 5A1a : 4A2a : 3A3a : 2A4a : 1A5a : 6a$$

$$\text{频率:} \quad 1/64 : 6/64 : 15/64 : 20/64 : 15/64 : 6/64 : 1/64$$

$$\text{穗长:} \quad 18 : 17 : 16 : 15 : 14 : 13 : 12$$

将F<sub>2</sub>的各基因频率作一曲线图：



由多基因系统控制的数量性状绝大多数表现为正态分布





F<sub>2</sub>基因类别数

等位基因数+1

一对等位基因

$$2+1=3$$

三对等位基因

$$6+1=7$$

在以上例子中：小写字母仅仅保持一个基数，叫做**无效等位基因 (null alleles)**；大写字母基因具有使数量增加的效应，每增加一个，效应加1份，基因数目越多，每份的效应越小。

无效等位基因：不能产生野生型表现，完全失去活性的突变基因

大写字母基因叫做**有效等位基因 (active alleles)**



## (3) 数量性状基因数的估计

根据 $F_2$ 代极端类型出现的频率估计时:

当性状受1对基因支配:  $AA \times aa \rightarrow F_2$  极端类型 1/4

2  $AABB \times aabb \rightarrow F_2$  极端类型 1/16

3  $AABBCC \times aabbcc \rightarrow F_2$  极端类型 1/64

n  $\rightarrow F_2$  极端类型  $(1/2)^{2n}$

若实际测得极端类型出现的频率为:

$1/a$ , 即  $a = 2^{2n} \rightarrow n = \lg a / (2 \lg 2)$ ,  $n$  就是所估计的基因对数



可以通过计算 $F_2$ 代的表型中具有某一极端类型的个体数在 $F_2$ 总群体中所占比值，推算出几对微效基因控制了该性状

$$4^n = \frac{F_2 \text{代个体总数}}{F_2 \text{代中极端类型个体数}}$$

其中 $n$ 为微效基因对的数目。

例如，当两个纯种亲本杂交后得到 $F_1$ ， $F_1$ 自交得到 $F_2$ ， $F_2$ 总共个体数为22016，其中与某一亲本的数量性状相同的个体数为86，则：

$$4^n = 22016 / 86 = 256, \text{ 因为 } a^n = b, n = \lg b / \lg a$$

$$n = \lg 256 / \lg 4 = 4 \text{ 估计有4对微效基因控制该性状}$$



多对微效基因作用于某一性状时，每个微效基因在总效应中占多大比值的推算：

① 微效基因的累加作用

$$\text{纯显性亲本表型值}(\bar{x}) = \text{每个显性基因表型值} \times \text{纯显性亲本基因数} + \text{纯隐性亲本表型值}(\bar{x})$$

例：短穗玉米 穗长  $\bar{x} = 6.6 \text{ cm}$

长穗玉米 穗长  $\bar{x} = 16.8 \text{ cm}$

两个亲本杂交得到 $F_1$ 为中间性状。 $F_1$ 自交得 $F_2$ 代，其中长穗和短穗各占群体的 $1/16$ 。问每个显性微效基因的表型值是多少？

解：首先求纯显性亲本基因数： $4^n = 16, n = 2$

即控制长穗玉米穗长的基因为2对（4个）

则：





$$\begin{aligned} \text{每个显性基因表型值} &= \frac{\text{纯显性亲本表型值}(\bar{x}) - \text{纯隐性亲本表型值}(\bar{x})}{\text{纯显性亲本基因数}} \\ &= \frac{16.8\text{cm} - 6.6\text{cm}}{4} = 2.55\text{cm} \end{aligned}$$

已知纯隐性亲本的穗长为6.6cm, 因此:

含1个显性基因的玉米穗长:  $6.6 + 2.55 = 9.15$  cm

含2个显性基因的玉米穗长:  $6.6 + (2 \times 2.55) = 11.7$  cm

依次类推。



## ②倍加作用的微效基因的表型值的估算：

$$\text{纯显性亲本表型值 } (\bar{x}) = (\text{每个显性基因表型值})^n \times \text{纯隐性亲本表型值 } (\bar{x})$$

其中n是纯显性亲本的微效基因数目。

**例如：**已知一亲本株高74cm，由2对（4个）显性基因控制，另一亲本株高2cm。每个显性基因的效应是倍加作用。请估算每个显性基因的表型值，并估算出每种基因型植株的高度。

解：代入上式得：

$$x^n \times 2 = 74 \quad n = 4 \quad x^4 = 74/2 \quad x = 2.47$$

纯隐性亲本植株株高2cm, 则有：

有1个显性基因，植株高应是  $2.47 \times 2 = 4.94\text{cm}$

有2个显性基因，植株高应是  $2.47^2 \times 2 = 12.20\text{cm}$

有3个显性基因，植株高应是  $2.47^3 \times 2 = 30.14\text{cm}$

有4个显性基因，植株高应是  $2.47^4 \times 2 = 74.44\text{cm}$

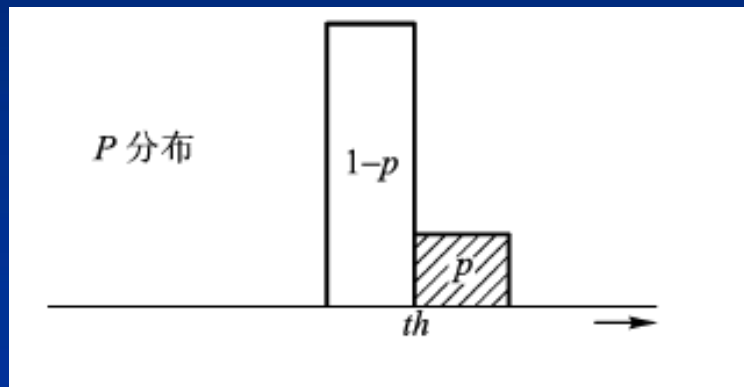
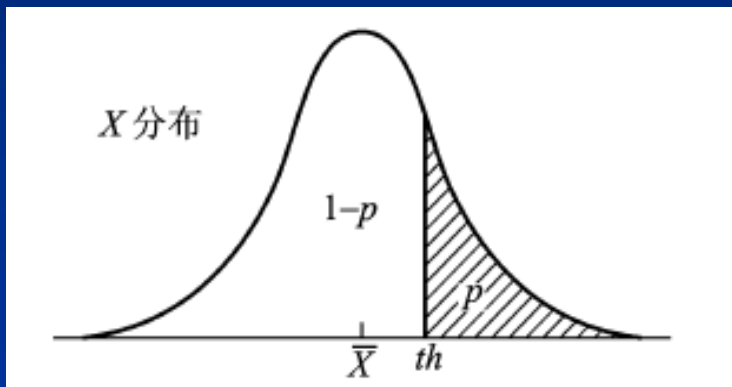




# 6.1.3 阈性状及其特性

**阈性状**(threshold character/trait)：遗传基础是微效多基因、表型是非连续变异的一类性状，是一类重要的数量性状。

阈性状的两种分布 { 呈正态分布（连续分布以X表示）  
可以计数的间断分布（以P表示）



表示发生率为20%的阈性状的两种分布的特征

## 表现特点

存在一个“阈”。阈的一侧表现一类性状，阈的另一侧表现另一类性状。如死亡与存活，中间只有一个临界点（阈值，threshold，th），在这种情况下个体在表型分布中只有两个值，0或1。所以，可以认为阈性状是一种超越某一遗传阈值时才表现的性状。



## 阈性状是一类重要的数量性状：

动、植物包括人类在内的抗病能力如“患病”或“正常”；“存活”或“死亡”；又如某些哺乳动物的前后肢的指（趾）数，多数个体有正常数目的指（趾），但少数个体可以出现多指（趾）等等，均属于阈性状。只含有一个阈值的阈性状又称为二者居一性状，或称全或无（all or none）性状。阈性状与非阈性状的数量遗传学分析的原理和方法基本相同，但在处理上有所差别。



人类多基因遗传病有唇裂±腭裂、腭裂、脊柱裂、无脑儿、先天性心脏病、精神分裂症、原发性高血压、冠心病、糖尿病、哮喘等。

一般认为是由遗传因素与环境效应共同决定个体是否容易患病→医学遗传学中称为易患（感）性（liability）。**易患性的变异是呈连续变异的，它表示人体内由基因决定的某种抗体物质的浓度差异。易患性高的个体，抗病力低，当一个个体的易患性超过一定限度——阈值时，该个体即表现为“患病”，性状就表达。连续分布的易患性（X）就被阈值区分出不连续的“发病”与“正常”两类，未越过阈值者属于“正常”，越过阈值者则为“患病”。在一定的环境条件下，阈值标志着患病所必需的最低的相关基因的数目。**



## 6.2 数量性状遗传分析的统计学基础

1、平均数：
$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

2、方差：
$$s^2 = \frac{\sum (x - \bar{x})^2}{n-1}$$

3、标准差：
$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

$$S = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}}$$



4、直线相关：

$$r_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

5、协方差：两个相关变量（ $x$ 、 $y$ ），共同变异量的度量：

$$\text{COV}_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{N} = \frac{1}{N} \sum x_i y_i - \bar{x} \bar{y}$$

6、回归系数：

$$b_{yx} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$

统计学常用参数的概念及其计算方法详见@6-1



## 6.3 数量性状基因座及其作图

### 6.3.1 数量性状基因座作图原理与步骤

经典的数量遗传分析方法→ 只能分析控制数量性状表现的众多基因的综合遗传效应，无法准确鉴别基因的数目、单个基因在染色体上的位置和遗传效应。

#### (1)、QTL的概念

Quantitative trait loci: **QTL** 数量性状位点(基因座)

在数量遗传中所分析的某个QTL只是一个统计的参数，它代表染色体（或连锁群）上影响数量性状表现的某个区段，它的范围可以超过10cM，在该区段可能会有1个甚至多个基因







## (2) QTL作图原理和步骤

➤ 一个数量性状往往受多个QTL影响，这些QTL分布于整个基因组的不同位置。利用特定的遗传标记可以确定影响某一性状的QTL在染色体上的数目、位置及其遗传效应，这就是QTL作图(QTL mapping) 也称作QTL定位。

➤ 1) 用于QTL定位的遗传标记

包括形态标记、细胞学标记、生化标记、DNA标记



## 2) 用于QTL定位的分子标记连锁图谱

如果分子标记覆盖整个基因组，控制数量性状的基因座( $Q_i$ ，即QTL)两侧会有相连锁的分子标记( $M_{i-}$ 和 $M_{i+}$ )。这些与数量性状基因紧密连锁的分子标记将表现不同程度的遗传效应。分析这些表现遗传效应的分子标记，就可以推断与分子标记相连锁的QTL的位置和效应。

➤ **QTL作图的基本原理** 是利用特定遗传分离群体中的遗传标记及相应的数量性状观测值，**分析遗传标记和性状之间的连锁关系**。如果分析结果证明某个遗传标记与性状连锁，则可认定在该标记附近存在一个或几个QTL。分析一个性状与已知连锁图的一系列标记之间的连锁关系，即可确定存在多少个QTL及这些QTL在标记图谱上的位置。**需要注意的是QTL作图中的连锁分析与质量性状不同，不能直接计算遗传标记和QTL之间的重组率，而是采用统计学方法计算它们之间连锁的可能性(LOD值)**，依据这种可能性是否达到某个阈值来判断遗传标记和QTL是否连锁，并进而确定其位置和效应。



## 以单标记和单基因为例（如图）

若标记M与性状Q无连锁(左图), 则不同的M标记基因型(MM、Mm、mm)所对应的Q基因型(QQ、Qq、qq)比例分布相同(均遵循孟德尔规律), 因此3种M标记基因型所对应的Q性状平均值会相等; 若标记M与Q存在连锁(右图), 不同的M标记基因型所对应的Q基因型比率分布会受M标记连锁影响发生改变, 因此3种M标记基因型所对应的Q性状在平均数上有差异, 这是数量性状定位的最基本原理。



## 6.3.2 QTL定位的基本步骤

### (1) 选择遗传标记

理想的作图标记应具有4个方面的特征：

- **数量丰富**：标记覆盖整个基因组；
- **多态性好**：个体或亲代与子代之间有不同的基因型；
- **中性**：同一基因位点的各种基因型都有相同的适应性，以避免不同基因型间的生存能力差异引起的试验误差；
- **共显性**：以保证直接区分同一基因位点的各种基因型。形态标记数量有限，通常不表现中性和共显性；

蛋白质标记可以满足中性和共显性，但它们又有数量不足或多态性不好的缺点，而DNA分子标记容易具备上述4个特征，已成为目前应用最广泛的作图标记。

常用的DNA分子标记有 RFLP, AFLP, VNTR (variable number of tandem repeat可变数目串联重复), SSR等。



## (2) 人工构建作图群体

适于QTL定位的群体应该是待测数量性状存在广泛变异，多个标记位点处于分离状态的群体，这样的群体一般是由亲缘关系较远的亲本间杂交，再经自交回交等方法进行人工构建的。如用高株×矮株或早熟期×晚熟期等，常用的群体有：

**F<sub>2</sub>群体**

**回交(BC)群体**

**双单倍体(doubled haploids, DH, 即加倍的单倍体群体)群体**

**重组近交系(recombinant inbred lines, RIL, 由F<sub>1</sub>连续多代自交产生)群体等。**

其中DH群体和RIL群体的分离单位是品系，品系间存在遗传差异而品系内个体间基因型相同，自交不分离，可以永久使用。



## (3) 获得某数量性状的表型值与遗传标记基因型的相关数据

从作图群体中**抽样（每株）提取DNA**做分子标记检测，记录每个被测个体的**标记基因型**。若标记的遗传图谱未知，还需要先依据各标记基因型分离资料制作标记的连锁图。

由于各种分子标记最后显示的都是电泳分离的带谱所以个体的标记基因型需要将每个标记的带纹与亲本比较并赋值来记录，例如在**共显性情况下**，两个纯合亲本各显示1条带，杂合体同时显示双亲的2条带。作图群体中应含有 $P_1$ 、 $P_2$ 和杂合型3种带型，这**3种带型即代表某一分子标记的3种基因型**。如果将含有 $P_1$ 带型的个体赋值为1， $P_2$ 带型赋值为3，杂合体赋值为2，即可得到**数据化的分子标记基因型**。在此基础上才能进行分子标记遗传图谱的制作和QTL图谱制作。





## (4) 测量数量性状

表 8-6 水稻窄叶青 8×京系 17F<sub>2</sub> 群体部分数量性状值与 RFLP 标记基因型值

株号	性 状			RFLP 标记*					
	生育期/d	株高/cm	颖花数/个	RG573	RG13	RZ70	RG697	RG474	RG435
1	126	115	129	1	1	3	1	1	1
2	107	127	356	2	2	2	3	3	3
3	112	123	341	2	1	1	2	1	1
4	95	117	202	1	1	3	1	2	1
5	115	121	285	3	3	1	3	2	3
6	110	127	311	2	2	1	2	1	2
7	101	103	126	3	3	2	2	1	2
8	119	116	237	2	2	2	0	3	3
9	118	112	217	1	2	3	0	2	2
10	115	111	186	3	3	3	0	3	3
11	104	115	330	2	3	3	0	2	2
12	109	106	162	2	2	3	2	2	2
13	107	115	270	2	2	2	0	2	2
14	114	123	314	3	3	2	3	3	3
15	123	97	146	1	1	3	0	1	1
.....				.....	.....				

\* 标记基因型值中的 0 表示带纹不清或其他原因造成的数据缺失。

在检测作图群体的每个个体的标记基因型值的同时，测定其数量性状值。将每个个体的数量性状表现型值和分子标记基因型值按顺序列表（表8-6），就形成了后续分析的基本数据。





## (5) 统计分析

- 用统计方法分析数量性状与标记基因型值之间是否存在关联，判断QTL与标记之间是否存在连锁（用Mapmaker / Exp3.0软件进行连锁遗传分析），确定QTL在标记遗传图谱上的数目、位置，估计 QTL的效应。



武汉大学

Wuhan University

简而言之，QTL定位一般包括以下步骤：

确定和筛选遗传标记

构建作图群体

检测和分析标记（检测分离世代群体中每一个体的标记基因型值和数量性状值）

构建标记的遗传图谱

测量数量性状

统计分析（数量性状位点定位）等步骤





现以玉米的两个近交系：

B73 × Mo17的F<sub>1</sub> × Mo17回交群体

(BC) 的试验为例说明影响其产量性状的QTL区间定位结果。

采用同工酶 (PGM2及AMP3) 和

RFLP标记 (C409, C282, C256, C449, C458, C228) 共计8个

标记位点；各标记间的距离为

厘摩 (cM)；*LOD*值为2的水平

线是该试验的阈值 (图6-5)

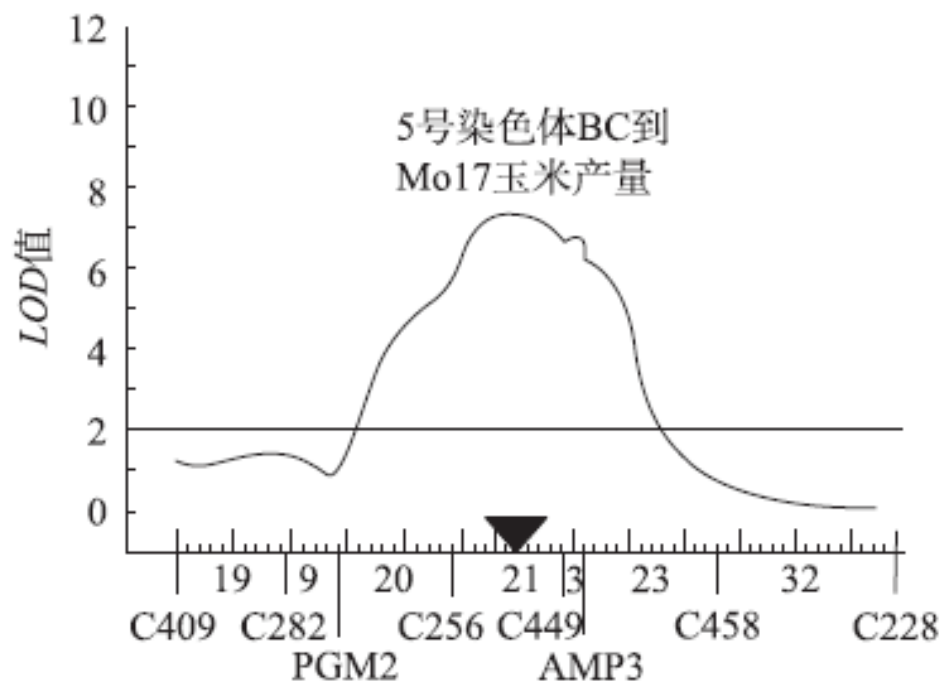


图 6-5 玉米第 5 号染色体上影响产量的 QTL 定位及 *LOD* 值(引自盛志廉等,1999)

图中 ▼ 所在位置表示具有最大 *LOD* 值的 QTL 所处位置,即表明影响玉米产量性状的 QTL 存在于标记位点 C256 与 C449 之间



## QTL分析的应用前景

QTL分析的应用主要有3个方面：

- ①由QTL定位得到的遗传图谱可以进一步转换成物理图谱，对QTL进行克隆和序列分析，在DNA分子水平上研究决定数量性状基因的结构和功能，进而应用基因工程的手段来操纵QTL。
- ②用于标记辅助选择（Marker-assisted selection, MAS）。在动植物育种上，利用标记与QTL的连锁，在实验室内对数量性状变异提早进行识别与选择，可以提高选择效率和精度。与传统方法相比，标记辅助选择对回交育种引入隐性有利基因、剔除非轮回亲本不利连锁基因更为快速有效，可以较少世代数完成目的基因的转育。
- ③利用标记与QTL连锁分析可以提供与杂种优势有关的信息，鉴定与杂种优势有关的标记位点，确定亲本在QTL上的差异，可以有效地预测。



## 6.4 数量性状遗传率及计算方法

### 6.4.1 数量性状的遗传率/遗传力 (heritability)

#### (1) 表型值及其方差的分量

① **表型值及其剖分** 某数量性状的表型值就是实际所度量或观察到的数值。表型值受许多外界因素如土壤、肥力、水分、光照、温度等的改变而发生变异，这种变异归因于环境因素。任何一个数量性状的表现都是遗传和环境共同作用的结果。所以**性状的表型值首先可以剖分为遗传和环境两个组成部分**：

$$P = G + E$$

其中： $P$  = 表型值； $G$  = 基因型值； $E$  = 环境效应





人的“身高”是一种典型的数量性状，是由多基因的遗传与环境效应共同作用的结果。人类正常身高表型呈正态分布，最高个体及最矮个体出现的频率很低，处在正态分布的两端。我国篮球名将**姚明**（图6-6）罕见的**身高为2.29 m**，比他的家乡上海男性同期平均身高（1.68 m）高出接近0.61 m，高出36.3%



图 6 - 6 姚明(图片引自新华网)



## ② 表型方差及其分量

在群体中，某数量性状的遗传变异属于总的表型变异的一部分，表型变异的其余部分是环境变异。在统计学上用方差作为变异的度量。总的表型方差可剖分为遗传方差和环境方差，于是：

总的表型方差 = 遗传方差 + 环境方差

所以： $V_P$ （表型方差）=  $V_G$ （遗传方差/基因型方差）+  $V_E$ （环境方差）

如果环境与基因型有互作

则： $V_P = V_G + V_E + 2COV_{GE}$

若G与E不相关

则： $V_P = V_G + V_E$

因为：基因型方差的组分有：加性方差、显性方差、互作方差





其中  $V_G = V_A + V_D$

所以:  $V_P = V_A + V_D + V_I + V_E$

其中:

$V_A$ =加性方差（育种值方差）。由于多基因的累加效应造成的遗传变异，能遗传且固定的组分。

$V_D$ =显性方差。由于等位基因间的显隐性关系而造成的一部分非加性的遗传变异，随着自交代数的增加而逐渐消失，所以能遗传但不能固定，是一种杂种优势现象。

$V_I$ =互作方差。由于非等位基因之间上下位关系所产生的非加性的遗传变异，此分量也被认为能遗传而不能固定。

## (2) 基因型值的尺度

一对等位基因 $A_1$ 和 $A_2$ ，其频率分别为 $p$ 和 $q$ ，当群体平衡时，其基因型频率为：

$$p^2 (A_1A_1) + 2pq (A_1A_2) + q^2 (A_2A_2)$$

设 $A_1A_1$ ， $A_1A_2$ ， $A_2A_2$ 的基因型值分别为 $a$ ， $d$ ， $-a$ 。

基因型值的坐标：

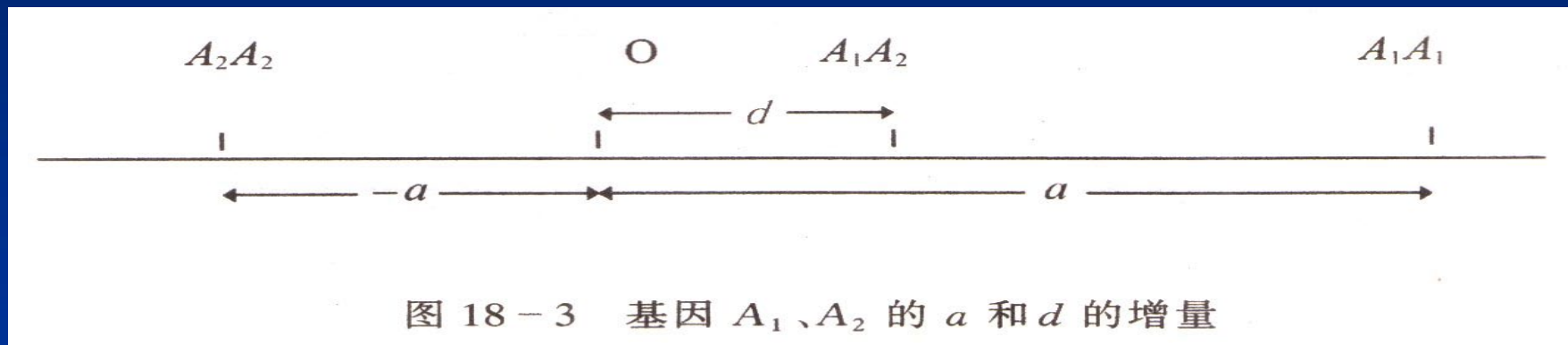


图 18-3 基因  $A_1$ 、 $A_2$  的  $a$  和  $d$  的增量

离差从纯合体的中点 $O$ 量起， $d$ 可正可负，因而 $A_1A_2$ 可以在 $O$ 的任何一边。

$O$ 点：两纯合体基因型值的算术平均值，作为量度 $a$ 和 $d$ 的原点。

“ $a$ ”：表示从 $O$ 点量起朝着坐标正的方向的一个增量，属于  $A_1A_1$ 的基因型值的增量（ $-a$ 正好相反）。

“ $d$ ”：表示杂合体的数值，取决于显性度，可正可负。

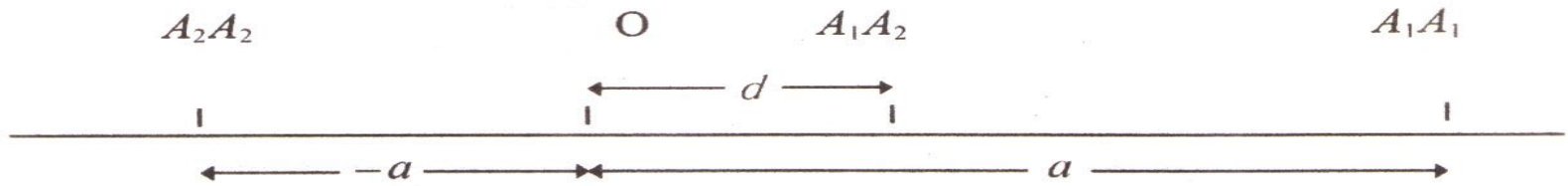


图 18-3 基因  $A_1$ 、 $A_2$  的  $a$  和  $d$  的增量

- (1) 无显性时  $d=0$ ，表示杂合体的累加效应完全等于纯合体的平均效应。
  - (2) 部分显性时， $a > d > 0$ （或  $-a < d < 0$ ）表示杂合体的累加效应与  $A_1A_1$  相似（或与  $A_2A_2$  相似）， $d$  偏于  $a$  的一边，说明  $A_1$  对  $A_2$  有部分显性； $d$  偏于  $-a$  一边，则说明  $A_2$  对  $A_1$  有部分显性。
  - (3) 完全显性时， $d=+a$  或  $-a$ 。当  $d=+a$  时， $A_1A_2$  的表型不能与  $A_1A_1$  相区别； $d=-a$  时，则  $A_1A_2$  的表型完全与  $A_2A_2$  相同。
  - (4) 超显性时， $d > +a$  或  $d < -a$  表明杂合体的表现超过了纯合体的表现。
- 所以， $d$  完全是显性效应的度量。一般用  $d/a$  来表示显性度。完全显性时，显性度  $d/a=1$ 。



若设： $A_1=5$ ， $A_2=3$

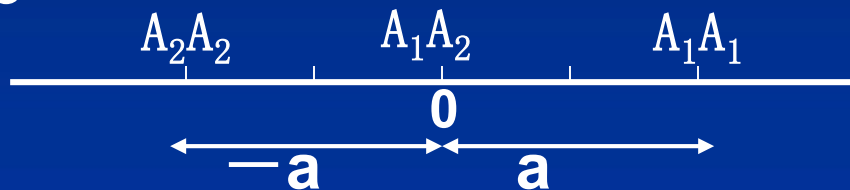
则  $A_1A_1=5+5=10$      $A_1A_2=5+3=8$      $A_2A_2=3+3=6$

基因型值的坐标尺度应为：O点  $(10+6) / 2=8$

所以： $a$ 值= $10-8=2$

$-a$ 值= $6-8=-2$

$d$ 值= $8-8=0$



在此， $A_1$  与  $A_2$  没有显性 ( $d=0$ ), 杂合体的累加效应完全等于纯和体的平均效应



### (3) 遗传力 (遗传率, heritability)

所谓遗传力就是遗传变量在总的表现变量所占的比值, 通常它是用百分率 (%) 来表示的。

#### ① 广义遗传力 (heritability in the broad sense, $H^2$ )

指数量性状基因型方差占表型方差的比例, 用公式表示为:

$$H^2 = \text{基因型方差} / \text{表型方差} \times 100 = V_G / V_P \times 100$$

因为:  $V_P = V_G + V_E$

所以:  $H^2 = V_G / V_P \times 100 = V_G / (V_G + V_E) \times 100$

通过广义遗传力的估计, 可以了解一个性状受遗传效应影响有多大, 受环境效应影响多大。



如果环境变量小，遗传力就高，表示表现型变异大都是可遗传的，也说明这个性状的遗传力传递力比较强；相反，当环境变量较大时，遗传力就小，表示表现型变异大都是不遗传的，说明了这个性状的遗传传递力比较弱。因此，遗传力的大小，就变成成为亲本和后代之间遗传关系的一个度量；或者说，遗传力的大小也可作为估算不同性状的遗传传递强弱的一个指标。

## ② 狭义遗传力 (heritability in the narrow sense, $h^2$ )

是指数量性状育种值方差（加性方差）占表型方差的比例：

$$h^2 = \text{育种值方差（加性方差）} / \text{表型方差} \times 100$$
$$= V_A / V_P \times 100$$

由于育种值是从基因型效应中已剔除显性效应和上位效应后的加性效应部分，在世代传递中是可以稳定遗传的，因此它在育种上具有重要意义。



表 6-2 某些数量性状的遗传率

种类	性状	遗传率
乳牛	泌乳量	0.30
	体重	0.37
猪	每窝仔数	0.15
	体长	0.55
来航鸡	产卵量	0.05
	成活率	0.10
小麦	粒重	0.10
黑腹果蝇	翅长	0.45
人类	产卵量	0.18
	身高	0.88
	胸围	0.61
	腰围	0.25
	收缩压	0.64
	体质指数	0.50
	口语能力	0.70
	数学能力	0.30
	智商	0.69
	总胆固醇	0.60
	指纹总峰数	0.97
	精神分裂症	0.80
	抑郁症	0.37

某数量性状的遗传率大，说明在该数量性状的表现中，由遗传所决定的比率较大，环境对它的影响较小。通常，与生物适应性无关的性状往往比与适应性有关的性状的遗传率要高一些。现列举几种动、植物包括人类的某些数量性状的遗传率，如表 6-2 所示。





## 6.4.2 估计遗传率的方法

在自花授粉作物中根据不同来源的资料，估计农作物数量性状遗传率的经典方法大致可以分为下列4种：

- ① 利用基因型一致的不分离群体（纯合亲本和其 $F_1$ 群体），估计环境方差，**求广义遗传率**；
- ② 用不同世代杂种群体消去环境方差和遗传方差中属于显性作用的方差，从而估计遗传方差中纯属于基因累加作用的方差，**求狭义遗传率**；
- ③ 利用上下代，亲代—子代的回归或相关关系估计狭义遗传率；
- ④ 用方差分析法分别估计总方差中各种方差组分，求遗传率。



## (1) 通过杂种第二代 ( $F_2$ ) 来估测遗传力 (广义遗传力)

首先以一对等位基因为例，然后推广到多个基因座。

设：群体中有  $A_1$ 、 $A_2$  基因。亲代的基因型为  $A_1A_1$ 、 $A_2A_2$ ，其  $F_1$  的基因型为  $A_1A_2$ 、一种表型，基因型平均值为  $d$ 。 $F_1$  自交后得到  $F_2$ ，在  $F_2$  群体中有 3 种基因型，其在群体中有不同的频率，计算平均效应时，应将各基因型按各自的频率作加权平均。得下表：

$F_2$ 基因型	基因型值 ( $x$ )	频率 ( $f$ )	$fx$	$fx^2$
$AA$	$a$	$1/4$	$1/4a$	$1/4a^2$
$Aa$	$d$	$2/4$	$2/4d$	$2/4d^2$
$aa$	$-a$	$1/4$	$1/4 - a$	$1/4(-a)^2$





F<sub>2</sub>群体中:

基因型值平均数:  $\mu = \frac{\sum fx}{\sum f} = \sum fx = \frac{1}{4}a + \frac{2}{4}d + \left(-\frac{1}{4}\right)a = \frac{1}{2}d$

在此  $\sum f = 1$

基因型方差: 根据方差定义,

$$\sigma^2 = \frac{\sum fx^2 - \frac{(\sum fx)^2}{\sum f}}{\sum f} = \sum fx^2 - \frac{(\sum fx)^2}{\sum f} = \sum fx^2 - (\sum fx)^2$$

∴  $\sum f = 1$

$$\begin{aligned} \therefore \text{基因型方差 } \sigma_G^2 &= \sum fx^2 - \frac{(\sum fx)^2}{\sum f} = \sum fx^2 - (\sum fx)^2 \\ &= \frac{1}{4}a^2 + \frac{2}{4}d^2 + \frac{1}{4}a^2 - \left(\frac{1}{2}d\right)^2 \\ &= \frac{2}{4}a^2 + \frac{2}{4}d^2 - \frac{1}{4}d^2 \\ &= \frac{2}{4}a^2 + \frac{2}{4}d^2 - \frac{1}{4}d^2 \end{aligned}$$





如果控制同一性状的基因有n对，这些基因又相互连锁，它们的作用是相等而相加的，并假定它们之间不存在相互作用。则F<sub>2</sub>的基因型方差为：

$$\begin{aligned}
 V_G(F_2) &= n \left( \frac{1}{2} a^2 + \frac{1}{4} d^2 \right) \\
 &= \left( \frac{1}{2} a_1^2 + \frac{1}{2} a_2^2 + \dots + \frac{1}{2} a_n^2 \right) + \left( \frac{1}{4} d_1^2 + \frac{1}{4} d_2^2 + \dots + \frac{1}{4} d_n^2 \right) \\
 &= \frac{1}{2} \sum_1^n a^2 + \frac{1}{4} \sum_1^n d^2
 \end{aligned}$$

设

$$A = \sum_1^n a^2 \quad D = \sum_1^n d^2$$

则： $V_G(F_2) = 1/2 A + 1/4 D$

其中：**1/2 A**和**1/4D**，分别定义为**加性方差**和**显性方差**，若考虑环境方差V<sub>E</sub>，

则 F<sub>2</sub>的表型方差： $V_P(F_2) = V_G + V_E = 1/2 A + 1/4 D + V_E$

由此可见：F<sub>2</sub>表型方差可分为3部分：即**加性（育种值）方差分量**、**显性方差分量**和**环境方差分量**。



在 $F_2$ 表型方差中， $V_G$ 不能直接估算，但可以通过间接的方法得到，这就是由混合基因型的群体（ $F_2$ 群体）的方差减去基因型一致的群体（亲代群体以及它们所产生的 $F_1$ 群体）的方差，可得 $V_G$ 的估计：

混合群体  
一致性群体

$$V_P (F_2) = V_E + V_G$$
$$V_P (F_1) = V_E + 0 \quad (-$$

---

$$\text{差异} = V_P (F_2) - V_P (F_1) = V_G$$



一个重要原理：如果用来作为亲本的（父本和母本）基因型一致，则它们的表型方差全部由环境因素所造成的，因为群体中 $V_G=0$ ，而由它们所产生的 $F_1$ 群体中基因型在每一个个体中都相同，因而 $V_G$ 亦等于0。于是我们称这一类群体为**基因型一致的群体**，它们的表型方差就完全等于环境方差。



较好环境方差估计应根据 $F_2$ 基因型分离比例求其加权均数即：

$$V_E = \frac{1}{4} V_{P_1} + \frac{1}{2} V_{F_1} + \frac{1}{4} V_{P_2}$$

那么：

$$H^2 = \frac{V_G}{V_P} \times 100\% = \frac{V_{F_2} - V_E}{V_{F_2}} \times 100\% = \frac{V_{F_2} - \frac{1}{4}(V_{P_1} + V_{P_2} + 2V_{F_1})}{V_{F_2}}$$

上述环境方差估计方法是最常用的一种

具体计算 $V_E$ 时可用下列方法之一：

- ①  $V_{F_1} = V_E$  （适用于异花授粉作物）
- ②  $1/2 (V_{P_1} + V_{P_2}) = V_E$  （适用于无 $F_1$ 数据）
- ③  $1/3(V_{F_1} + V_{P_1} + V_{P_2}) = V_E$





例：在玉米穗长试验的结果中：

➤  $V_{P1} = 2.401, V_{P2} = 2.177, V_{F1} = 2.349, V_{F2} = 5.072$

$$V_E = 1/3 (2.401+2.177+2.343) = 2.307$$

广义遗传率为：

$$H^2 = V_G / V_P \times 100$$

$$V_P (F_2) - V_P (F_1) = V_G$$

➤  $H^2 = (5.072-2.307) / 5.072 \times 100\% = 54\%$

在该杂交组合中，F<sub>2</sub>穗长的变异大约有54%是由于遗传差异造成的，46%是环境差异造成的。



(2) 通过杂种一代 ( $F_1$ ) 分别与亲本回交获得的回交的子代 (即  $B_1$ 、 $B_2$ ) 来估测遗传力

$B_1$ :  $F_1$  个体  $A_1A_2 \times P_1 (A_1A_1)$  的回交子代

$B_2$ :  $F_1$  个体  $A_1A_2 \times P_2 (A_2A_2)$  的回交子代

$B_1$  的遗传方差的估算:

$A_1A_2 \times A_1A_1$  回交的后代  $B_1$  是  $1/2 A_1A_1$ ,  $1/2 A_2A_2$ ,

$$n = 1/2 + 1/2 = 1$$

$A_1A_1$  的平均效应是  $-a$ ,  $A_1A_2$  的平均效应是  $d$ ,

$B_1$  的平均值  $= 1/2 (-a+d) = 1/2 (d-a)$

根据  $\bar{X}$  求  $B_1$  遗传方差

$$V = \frac{\sum (x - \bar{x})^2}{n}$$

$$B_1 V_G = \frac{1}{2} \left[ \left( d - \frac{d-a}{2} \right)^2 + \left( -a - \frac{d-a}{2} \right)^2 \right] = \frac{1}{4} (a - d)^2$$



$B_2$ 的遗传方差 ( $B_2V_G$ ) 的估算:

$A_1A_2 \times A_2A_2$ 回交的后代 $B_2$ , 是 $1/2 A_1A_2$ 、 $1/2 A_2A_2$ ,  
合计是 $n=1/2 + 1/2 = 1$

$A_2A_2$ 的效座是 $a$ ,  $A_1A_2$ 的效座是 $d$ ,  $B_2$ 的平均值

$$\bar{X} = 1/2 (A_1A_2 + A_2A_2) = 1/2 (d+a)$$

$$B_2V_G = \frac{1}{2} \left[ \left( a - \frac{a+d}{2} \right)^2 + \left( d - \frac{a+d}{2} \right)^2 \right] = \frac{1}{4} (a-d)^2$$

使 $a$ 与 $d$ 分开, 则有:  $B_1V_G + B_2V_G = 1/4 (a-d)^2 + 1/4 (a+d)^2$   
 $= 1/2a^2 + 1/2d^2 = 1/2A + 1/2D$

若推广到 $n$ 对基因, 这种回交子代遗传方差之和为:

$$B_1V_G + B_2V_G = 1/2 \sum a^2 + 1/2 \sum d^2 = 1/2A + 1/2D$$

再加上环境方差,  $B_1$ 和 $B_2$ 的总方差:  $V_{B1} + V_{B2} = 1/2A + 1/2D + 2V_E$



通过方差计算遗传度的方法

$$h^2 = (V_{F2} - V_{F1}) / V_{F2} \times 100\%$$

$$\text{方差 } V = \sum f (X - \bar{X})^2 / (n - 1)$$

F1—子一代； F2—子二代； X—变数；  $\bar{X}$ —平均数； f—频数； n—样本总数

$$\because V_p(F_2) = V_E + V_G$$

$$V_p(F_1) = V_E$$

最少控制基因对数

$$n = \frac{(\bar{X}_{P1} - \bar{X}_{P2})^2}{8(V_{F2} - V_{F1})}$$



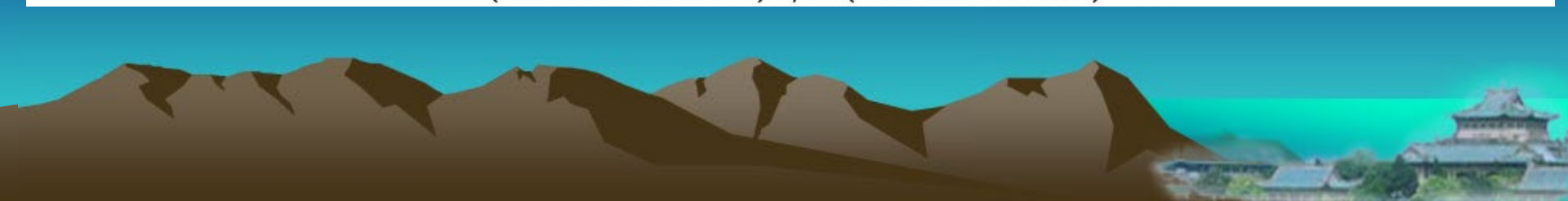
例：如下表，已知玉米穗长的频率分布，请利用方差计算其遗传率和最少控制基因对数

### 玉米穗长的频数分布表

	穗长 (x) (厘米)																	总数 n	统计数			
	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21		平均数 X	方差 V		
P <sub>1</sub>	4	21	24	8														57	6.632	0.665		
P <sub>2</sub>							3	11	12	15	26	15	10	7	2					101	16.802	3.561
F <sub>1</sub>					1	12	12	14	17	9	4								69	12.116	2.309	
F <sub>2</sub>		1	10	19	26	47	73	68	68	39	25	15	9	1						401	12.888	5.074

$$\text{遗传率 } h^2 = (V_{F_2} - V_{F_1}) / V_{F_2} \times 100\% = (5.074 - 2.309) / 5.074 \times 100\% \approx 54\%$$

$$\begin{aligned} \text{最少控制基因对数 } n &= (\bar{X}_{P_1} - \bar{X}_{P_2})^2 / 8(V_{F_2} - V_{F_1}) \\ &= (16.802 - 6.632)^2 / 8(5.074 - 2.309) \approx 5 \end{aligned}$$



### (3) 人类疾病遗传率的估计

计算人类多基因遗传病遗传度的高低在临床实践上有重要意义，传统的计算方法主要有两种，即Falconer公式和Holzinger公式。

#### ① Falconer公式——从群体和患者亲属发病率估计遗传率

Falconer1965年提出Falconer公式（Falconer method），它是根据先证者亲属的患病率与遗传率有关而建立的。亲属患病率越高，遗传率越大，所以可通过调查先证者亲属患病率和一般人群的患病率，算出遗传率( $h^2$ 或 $H$ )。





$$h^2 = \frac{b}{r} \quad (1)$$

(1) 式中， $h^2$ 为遗传率； $b$ 为亲属易患性对先证者易患性的回归系数； $r$ 为亲属系数。

当已知一般人群的患病率时，用下式计算回归系数：

$$b = \frac{X_g - X_r}{a_g} \quad (2)$$

当缺乏一般人群的患病率时，可设立对照组，调查对照组亲属的患病率，用下式计算回归系数：

$$b = \frac{p_c (X_c - X_r)}{a_r} \quad (3)$$

## Falconer公式中各符号的意义：

$$h^2 = \frac{b}{r}$$

b : 回归系数；

r : 亲属系数；

$$b = \frac{X_g - X_r}{a_g}$$

$X_g$  : 一般群体易患性平均值与阈值之间的标准差

$X_r$  : 先证者亲属易患性平均值与阈值之间的标准差

$a_g$  : 一般群体易患性平均值与一般群体中患者易患性平均值之间的标准差

$X, a$ 均可查Falconer表得出

$$b = \frac{p_c (X_c - X_r)}{a_r}$$

$q_g$ 为一般群体患病率； $q_c$ 为对照亲属患病率

$p_c = 1 - q_c$ ； $q_r$ 为先证者亲属患病率。

$X_c$  : 为对照组亲属中的易患性平均值与阈值之间的标准差数；

$X_r$ 为先证者亲属易患性平均值与阈值之间的标准差数；

$a_r$ 为先证者亲属易患性平均值与先证者亲属中患者易患性平均值之间的标准差数；

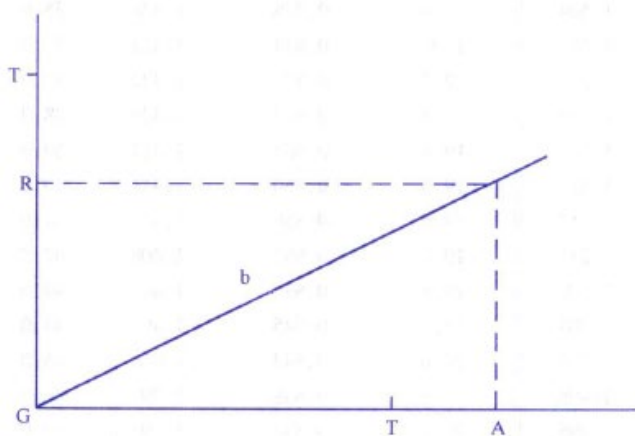


图 6-8 多基因遗传病患者亲属对患者的回归图解

注: T: 阈值; G: 一般群体易患性平均值; A: 患者易患性平均值; R: 患者亲属易患性平均值; b: 患者亲属对患者的回归系数。

从图 6-8 中可见斜率的式子为:

$$b = \frac{(R - G)}{(A - G)}, \quad R - G = (T - G) - (T - R)$$

$$b = \frac{(T - G) - (T - R)}{(A - G)}$$

$$b = \frac{(FG) - (T - R)}{(A - G)}$$

**回归系数:** 回归方程式  $Y=bX+a$  中的斜率  $b$ , 称为回归系数, 表  $X$  每变动一单位,  $Y$  亦变动  $b$  单位。  
或: 回归方程式中与自变量相乘的参数。

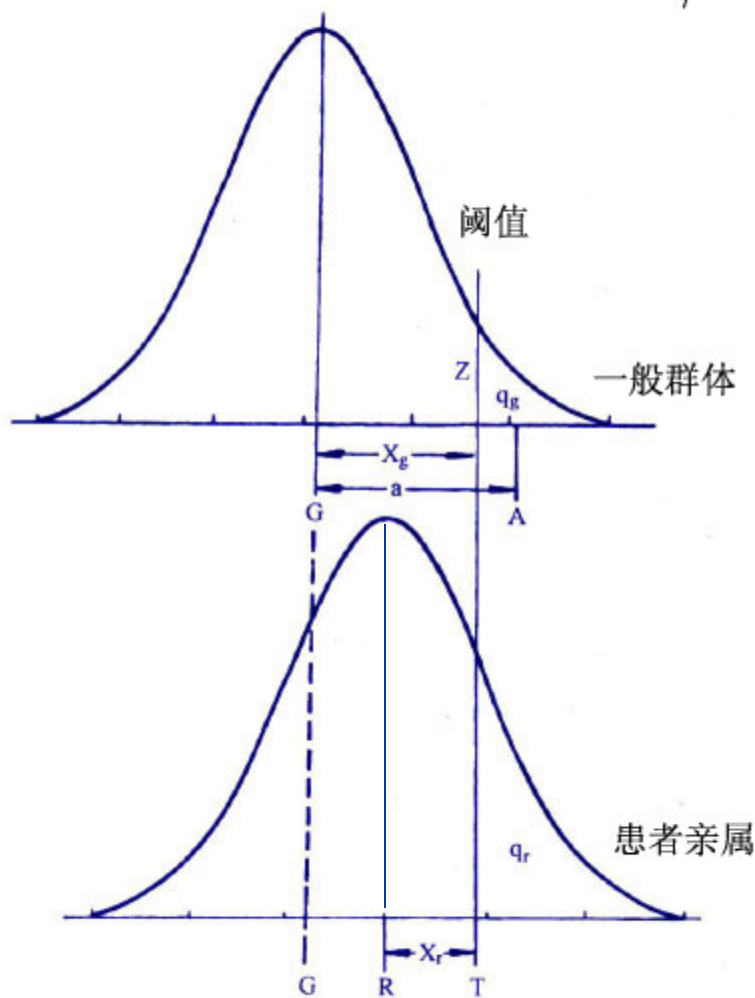
例如在简单直线回归方程式  $Y_0=a+bx$  中,  $b$  就是回归系数, 它代表因变量  $Y$  对自变量  $X$  的平均变动..

如果将患者易患性和患者亲属易患性的正态分布图与上述回归图相对应(图 6-9), 则  $T-G$  为  $X_g$ ,  $T-R$  为  $X_r$ ,  $A-G$  为  $a_g$ ,  $R-G$  为  $a_r$ 。因此, 回归的斜率为:

$$b = \frac{X_g - X_r}{a_g}$$

Falconer 提出, 从回归系数 ( $b$ ) 和亲缘系数 ( $r$ ) 可计算遗传度, 即:

$$h^2 = b/r$$



- 注: G 总人群平均易患性  
 A 总人群中患者(先证者)的平均易患性  
 R 患者一级亲属的平均易患性  
 T 易患性阈值  
 $q_g$  总人群发病率  
 $q_r$  患者一级亲属发病率  
 $X_g$  总人群平均易患性距离阈值的标准差数  
 $X_r$  患者一级亲属平均易患性距离阈值的标准差数  
 $a$  患者平均易患性距离总人群平均易患性的标准差数  
 Z 阈值 T 处的纵坐标高度

图 6-9 多基因遗传病总人群和患者一级亲属易患性分布的比较

在亲属系数中  
有1/2的可能性是  
父母和外祖父母之  
与其表兄妹、堂兄

$X_g$ : 一般群体易患性平均值与阈值之间的标准差

$X_r$ : 先证者亲属易患性平均值与阈值之间的标准差

$a_g$ : 一般群体易患性平均值与一般群体中患者易患性平均值之间的标准差

$X, a$ 均可查Falconer表得出

例如，有人调查先天性房间隔缺损在一般群体中的患病率为1/1000 (0.1%)，在100个先证者的家系中调查，先证者的一级亲属共有 669人(双亲200人，同胞279人，子女190人)，其中有22人发病，依次求得先证者一级亲属的患病率为  $22 / 669 \times 100\% = 3.3\%$  ( $q_r$ )，然后查Falconer表。按群体患病率查得 $X_g$ 和 $a_g$ ，再根据亲属患病率查得 $X_r$ 和 $a_r$ ，然后代入公式(2)求出b值。

$$b = \frac{X_g - X_r}{a_g} = \frac{3.090 - 1.838}{3.367} = 0.372$$

所以：

$$h^2 = \frac{b}{r} = \frac{0.37}{0.5} = 0.74 = 74\%$$

群体患病率为  
0.1%时，  
 $X_g=3.090$   
 $a_g=3.367$

$$q_r = 22/669 = 3.3\%$$

$q_r=3.3\%$ 时  
 $X_r=1.838$   
 $a_r=2.231$

以上计算结果表明，遗传因素对先天性房间隔缺损发生的贡献为74%，经显著性检验该遗传率有统计学意义。





## ② Holzinger公式——从双生子的发病一致率估计遗传率

Holzinger公式（Holzinger formula）（1929）是根据遗传率越高的疾病，一卵双生的患病一致率与二卵双生患病一致率相差越大而建立的。

**一卵双生**（monozygotic twin, MZ）：由一个受精卵形成的两个双生子，他们的遗传基础理论上完全相同，其个体差异主要由环境决定；

**二卵双生**（dizygotic twin, DZ）：由两个受精卵形成的两个双生子，相当于同胞，因此他们的个体差异由遗传基础和环境因素共同决定。

所谓**患病一致率**是指双生子中一个患某种疾病，另一个也患同样疾病的频率。





$$h^2 = \frac{C_{MZ} - C_{DZ}}{100 - C_{DZ}} \quad \text{或} \quad h^2 = \frac{C_{MZ} - C_{DZ}}{1 - C_{DZ}}$$

其中， $C_{MZ}$ 为一卵双生子的同病率； $C_{DZ}$ 为二卵双生子的同病率。

例如，对躁狂抑郁性精神病的调查表明，在15对单卵双生子中，共同患病的有10对；在40对双卵双生子中，共同患病的有2对。依此来计算单卵双生子的同病率为67%，双卵双生子的同病率为5%。代入上式

:

$$h^2 = \frac{C_{MZ} - C_{DZ}}{100 - C_{DZ}} = \frac{67 - 5}{100 - 5} = 0.65 = 65\%$$

以上结果表明，在躁狂抑郁性精神病中，遗传因素的贡献为65%。

表 9-3 多基因遗传病的发病率及遗传率(摘引自陈竺,2001)

疾病	群体发病率/%	先证者一级亲属发病率/%	遗传率/%
唇裂±腭裂	0.17	4	76
腭裂	0.04	2	76
脊柱裂	0.3	4	60
无脑儿	0.5	4	60
各型先天性心脏病	0.5	2.8	35
精神分裂症	0.5~1.0	10~15	80
原发性癫痫	0.36	3~9	55
原发性高血压	4~10	15~30	62
冠心病	2.5	7	65
青少年型糖尿病	0.2	2~5	75
哮喘	1~2	12	80
消化性溃疡	4	8	37
原发性肝癌	0.05	5.45	52



## 6.5 近亲繁殖与杂种优势

### 6.5.1 近交与杂交的遗传学效应

**异型交配** (nonassortative mating) : 基因型不同的纯合子之间的交配。

**同型交配** (assortative mating) : 相同基因型之间的交配

**近交** (inbreeding) : 也称近亲繁殖或近亲婚配, 是有亲缘关系的个体相互交配, 繁殖后代。近亲繁殖按亲缘关系的近、远程度一般可分为: 全同胞 (full-sib) (同父、母的兄妹)、半同胞 (half-sib) [同父 (或母) 异母 (或父) 的兄妹] 和表兄妹 (first cousins) 之间的交 (婚) 配。

**自交** (selfing) : 植物的自花授粉、动物的自体受精 (self-fertilization), 自交是近亲繁殖中最极端的方式。



## (1) 近交使基因纯合，杂交使基因杂合

以一对等位基因为例有3种交配类型：

$$AA \times AA, aa \times aa, Aa \times Aa$$

在 $AA \times AA$ ， $aa \times aa$ 交配类型→产生的子代的基因型（或表型）全部都是与亲本相同的**纯合体**；

第三种交配类型  $Aa \times Aa$  属于杂合体间的同型交配，其后代为 $1 / 4AA + 1 / 2Aa + 1 / 4aa$ ，经过n代连续的同型交配的过程和最终结果显示于表6-5：

表 6-5  $Aa \times Aa$  连续  $n$  代同型交配的结果

世代	基因型频率			基因频率	
	AA	Aa	aa	A	a
0	0	1	0	0.5	0.5
1	1/4	2/4	1/4	0.5	0.5
2	3/8	2/8	3/8	0.5	0.5
3	7/16	2/16	7/16	0.5	0.5
⋮	⋮	⋮	⋮	⋮	⋮
$n$	$(2^n - 1)/2^{n+1}$	$2/2^{n+1}$	$(2^n - 1)/2^{n+1}$	0.5	0.5
⋮	⋮	⋮	⋮	⋮	⋮
$\infty$	1/2	0	1/2	0.5	0.5

注:表 6-5 及其以下 3 个自然段中所涉及的基因型频率、基因频率等概念详见第 19 章。

表6-5显示:

- ① 群体中Aa的基因型频率按 $H_n = 1 / 2H_{n-1}$ 的规律每代减少前一代的1 / 2而迅速递减, 其极限值为0
- ② 同型交配群体中纯合体基因型频率每代都有增加,其频率各占1 / 2
- ③ 完全同型交配的群体中的基因频率并没有改变



在家养动物近交过程中，由于近交个体有限，加上严格的选择。因此，实际上基因频率会发生显著变化，这并非近交自身的遗传效应，而是遗传漂变和选择作用的结果。

杂交的遗传效应则使基因杂合，增加杂合体的频率。杂交实质上是纯合子间的异型交配，子代必然都是杂合体。

## 近交效应

### ① 增加后代群体的纯合度，使遗传性状稳定

亲缘关系越近的个体近交，其后代纯合体比率就上升越快，如每代进行同一类型的近交，自交6代后纯合子的比率可达成100%。

### ② 衰退现象的出现

由于纯合子比率增加，使原来处于杂合状态被掩盖的隐性性状得以表现，表现为生活力、繁殖力下降。抗病力下降等，如植物白化的出现。

对于人类，近亲结婚导致基因纯合化，使隐性遗传病发病率上升。例如：半乳糖血症，表兄妹结婚后代发病率为非近亲结婚时的19倍。





## (2) 近交系数与亲缘系数

**近交系数** (coefficient of inbreeding,  $F$ 或 $f$ ) : 个体在一个特定的基因座位上接受两个遗传上或者说血缘上相同的等位基因的概率。

或: 个体从其双亲共同祖先得到一对遗传上是等同的纯合基因的概率  
它反应了亲缘关系相近程度, 是度量个体近交程度的重要遗传参数, 用 $F$ 表示。

若某个体的 $F$ 值大, 说明其父母亲缘关系近;

若 $F=0$ , 说明其双亲无亲缘关系

**亲缘系数** (coefficient of relationship,  $R$ ) : 有亲缘关系的两个个体的基因组中携带相同而且同源基因的比例, 称之为亲缘系数。

它是两个个体亲缘程度的度量。亲缘系数愈大, 亲缘关系愈近, 亲缘系数为0, 则可认为个体 ( $X$ 、 $Y$ ) 间在近期几个世代内没有共同祖先

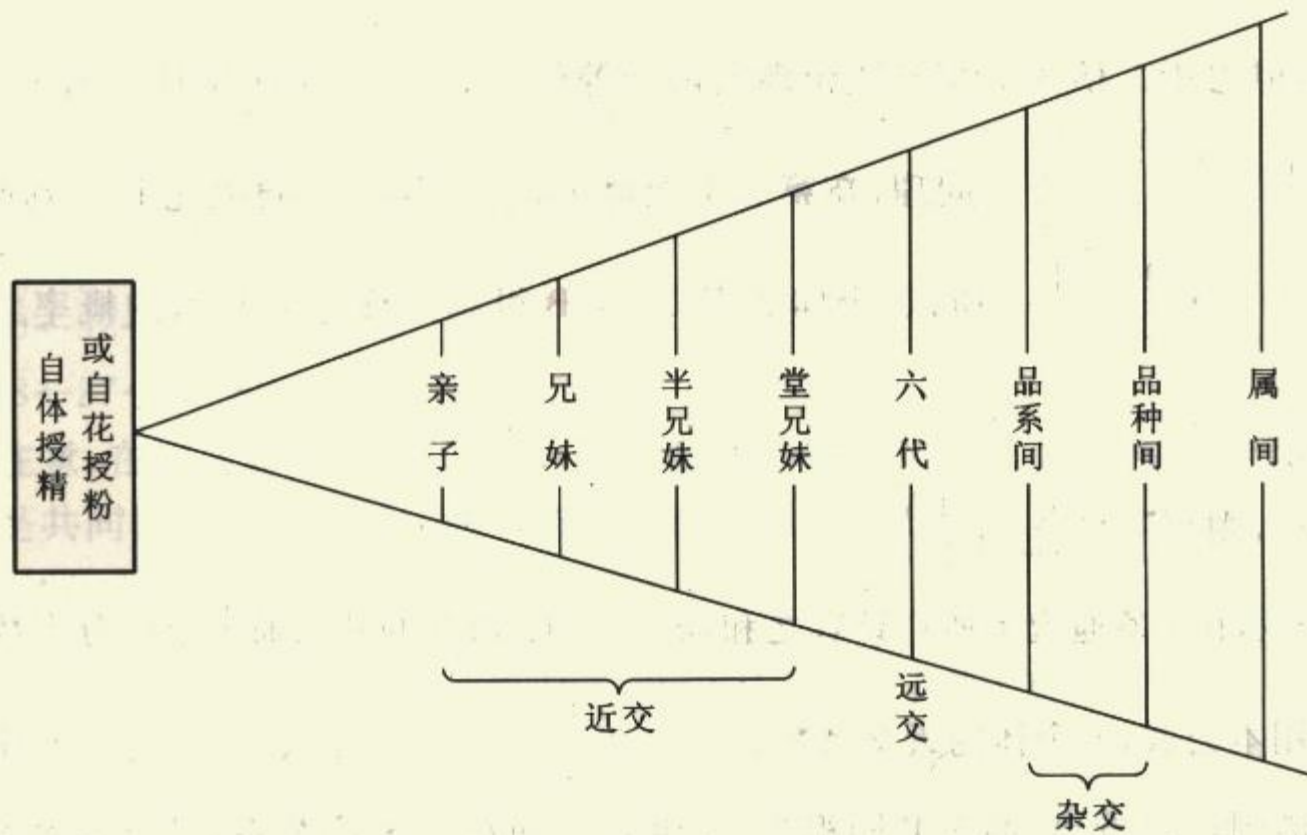


图 14-1 亲缘关系远近示意图



### (3) 运用通径分析方法计算近交系数和亲缘系数

Wright(1921)提出通径分析 ( path analysis) 原理与方法

#### ① 通径与通径链

在一个相关变量的网络系统中，连接“原因”与“结果”的每一个单箭头线条称为一条通径 ( path)

因果关系: 单箭头线，方向由“因”到“果”，称为通径线

通径线的系数称为**通径系数**：每一条通径中的“原因”对于某一个“结果”所起作用的相对大小，用通径系数 ( path coefficient ) 表示。即度量各“原因变量”对“结果变量”的直接影响的系数称为通径系数。

由一条或一条以上的通径所组成的完整的通道称为**通径链**。

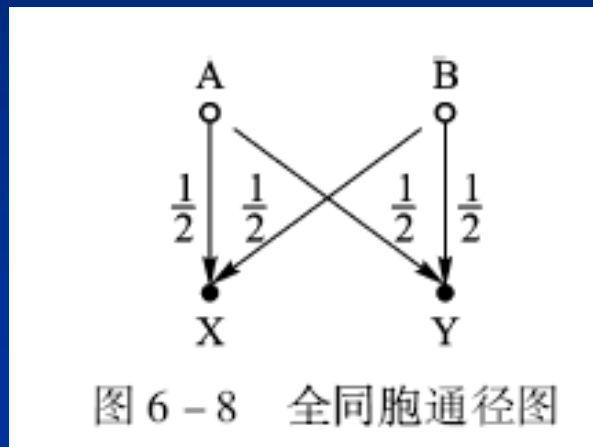


图 6 - 8 全同胞通径图



例如，在图6-7中，个体A、B为“原因变量”（以 $\circ$ 表示），个体X、Y为“结果变量”（以 $\bullet$ 表示）。由“原因变量”A、B指向“结果变量”X、Y的单箭头分别为4条途径，（ $A \rightarrow X$ ； $A \rightarrow Y$ ； $B \rightarrow X$ ； $B \rightarrow Y$ ），而连接个体A、B与个体X、Y的有两条途径链即： $X \leftarrow A \rightarrow Y$ 及 $X \leftarrow B \rightarrow Y$ 。由具有同父、同母的全同胞系谱图（图6-7）转换而成为全同胞途径图（图6-8）。

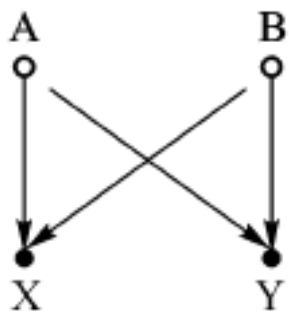


图6-7 全同胞系谱图

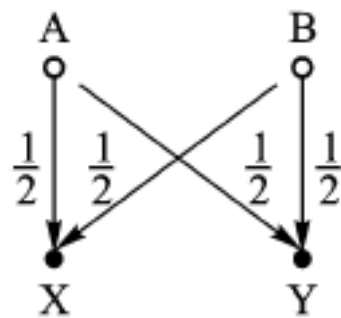
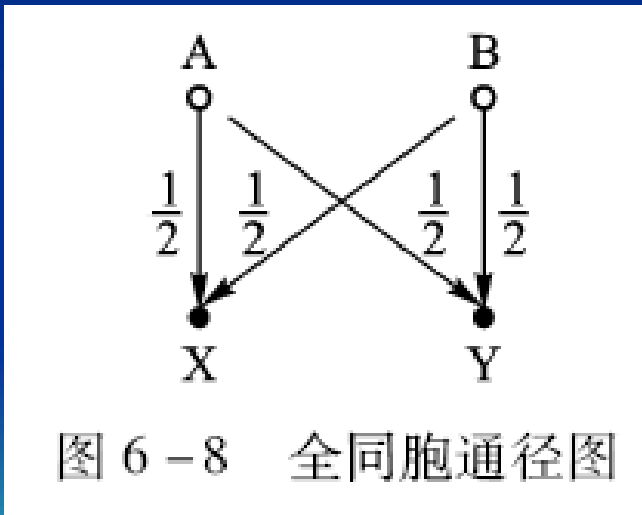


图6-8 全同胞途径图

② **途径分析的理论及其应用** 途径分析证明，在随机交配群体中个体世代的每一条途径的途径系数 =  $1/2$ 。而各类亲属关系的特定个体X、Y间的亲缘系数：

$$R_{XY} = \sum \left(\frac{1}{2}\right)^L \dots\dots\dots (1)$$

其中 $R_{XY}$ 表示X、Y个体的**亲缘系数**，L表示沿着某两个特定亲属间（X、Y间）的连接途径链条中的**箭头数**， $\Sigma$ 表示所有这类链条之和。



在图6—8途径图中由 $X \leftarrow A \rightarrow Y$ 及由 $X \leftarrow B \rightarrow Y$ 两条途径链连接X、Y两个特定个体，一条经过共同祖先A，另一条经过共同祖先B，而A、B两个个体不存在共同祖先，属随机婚配。每条途径链的系数由各途径的系数相乘，然后对每条途径链的系数求和，则有：

$$R_{XY} = (1/2)^2 + (1/2)^2 = 2(1/2)^2 = 1/2$$







# 同理可求表（堂）兄妹间的亲缘系数（图6-9）

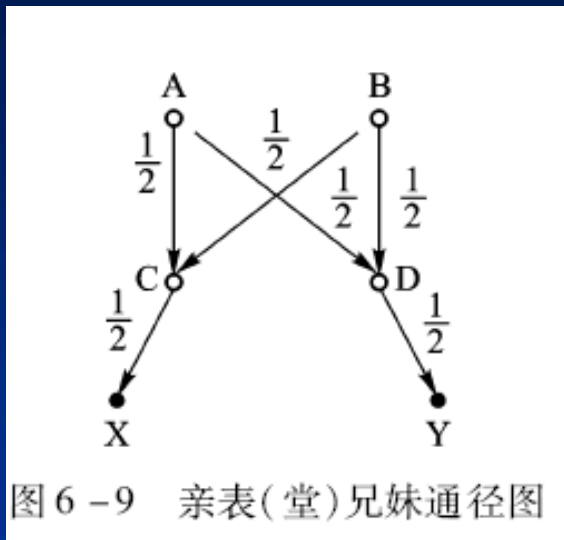


图 6 - 9 亲表(堂)兄妹通径图

连接两个特定亲属间的两条通径链条中各有4个箭头： $E \leftarrow C \leftarrow A \rightarrow D \rightarrow F$ 及 $E \leftarrow C \leftarrow B \rightarrow D \rightarrow F$ 。所以表（堂）兄妹的亲缘系数为：

$$R_{EF} = (1/2)^4 + (1/2)^4 = 2(1/2)^4 = 1/8$$

依通径分析原理所推导的近交系数（ $F_X$ ）的通式：

$$F_X = R_{SD} \times 1/2 \cdots \cdots (2)$$

该式表示在系谱中可追溯的最近共同祖先是随机婚配的（如图6-7至图6-12中的A、B个体），所以共同祖先的亲缘系数为零，个体X的近交系数是其双亲亲缘系数的1/2

$R_{SD}$ 表示S个体与D个体的亲缘系数，即X、Y个体的亲缘系数

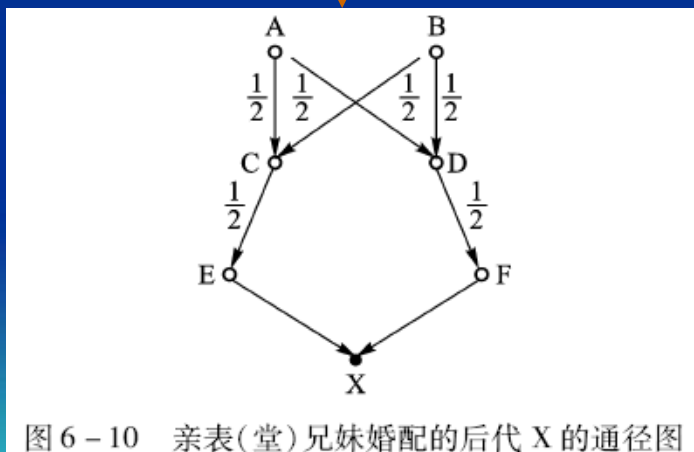
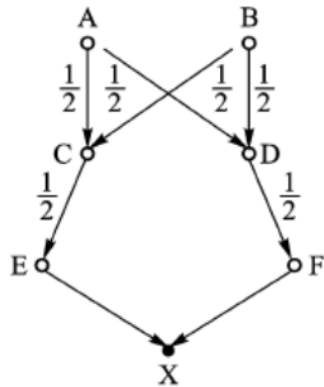


图 6 - 10 亲表(堂)兄妹婚配的后代 X 的通径图





当我们计算图6-10中表（堂）兄妹婚配的子女（X）的近婚（交）系数时：

$$F_X = R_{EF} \times 1/2 = 1/8 \times 1/2 = 1/16$$

图6-10 亲表(堂)兄妹婚配的后代X的通径图

通径分析方法计算近交系数的另一公式为：

$$F_X = \Sigma (1/2)^{n_1 + n_2 + 1}$$

令  $n_1 + n_2 + 1 = N$  时，则有： $F_X = \Sigma (1/2)^N \dots\dots (3)$

其中  $\Sigma$  表示各条通径链的数值求和， $N$  表示一条连接通径中包括X的双亲在内的个体数，运用上式时，只需从个体X的一个亲本出发逐代向上追溯，经过共同祖先再逐代往下寻找，一直回到另一个亲本的链条中总共所经历的个体数。（所求个体的双亲到一个共同祖先之间所有的长辈数）



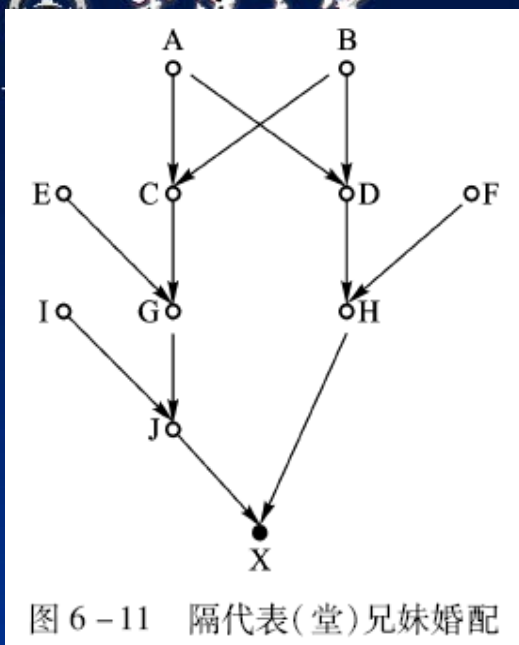


图 6-11 隔代表(堂)兄妹婚配

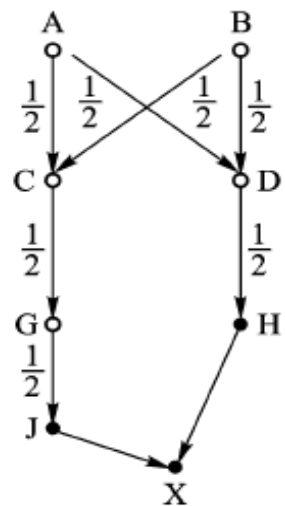


图 6-12 隔代表(堂)亲通径图  
J与H为个体X的双亲

由通径图6-12可见：由  $J \leftarrow G \leftarrow C \leftarrow A \rightarrow D \rightarrow H$  的通径链中包括6个个体，由  $J \leftarrow G \leftarrow C \leftarrow B \rightarrow D \rightarrow H$  的通径链中也包括了6个个体，由X的一个亲本（J）出发逐代向上追溯，经过共同祖先（A，B）再逐代往下沿**箭头**寻回到另一个亲本（H），总共经历了两条通径链

根据（3）式，得：

$$F_X = (1/2)^6 + (1/2)^6 = 1/32$$

或根据（1）式先求：

$$R_{JH} = (1/2)^5 + (1/2)^5 = 1/16$$

然后依（2）式可求：

$$F_X = R_{JH} \times 1/2 = 1/16 \times 1/2 = 1/32$$

（注意（3）式为个体数所求得得到的 $F_X$ ；而（1）式 $R_{JH}$ 为通径链中的箭头数计算而来。在计算时，个体数也好，箭头数也好都不能包括X个体在内）

各类亲属的亲缘系数及婚配后代的近交系数列于表 6-6。

表 6-6 几类亲属间亲缘系数及婚配后代的近交系数(仿自杜若甫,2005)

亲属关系	亲缘系数	近交(婚)系数
亲子(图 6-13)	$1/2(0.5)$	$1/4(0.25)$ (限于动、植物)
全同胞(图 6-7 及图 6-8)	$1/2(0.5)$	$1/4(0.25)$ (限于动、植物)
舅甥女、叔侄女、姨甥、姑侄(图 6-14)	$1/4(0.25)$	$1/8(0.125)$ (限于动、植物)
同父(母)异母(父)同胞(即半同胞)(图 6-15)	$1/4(0.25)$	$1/8(0.125)$ (限于动、植物)
亲表(堂)亲(图 6-9,图 6-10)	$1/8(0.125)$	$1/16(0.0625)$
隔山亲表(堂)亲(图 6-16)	$1/16(0.0625)$	$1/32(0.03125)$
表舅甥女、表(堂)叔侄女、表(堂)姑侄、表姨甥[又称隔代表(堂)亲](图 6-11,图 6-12)	$1/16(0.0625)$	$1/32(0.03125)$

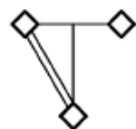


图 6-13 亲子交配(限于动、植物)

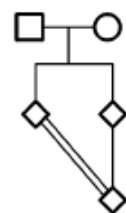


图 6-14 舅甥女等 4 种交配类型(限于动、植物)

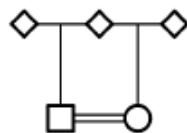


图 6-15 半同胞交配(限于动、植物)

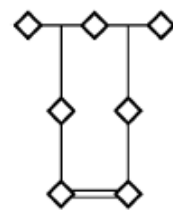


图 6-16 隔山亲表(堂)兄妹婚配



#### (4) 近交降低群体基因型值的平均值，杂交提高群体均值

一个数量性状的基因型值是由基因的累加效应值和非累加效应值组成。非累加效应值中的显性效应和大部分上位效应都存在于杂合体中，因此也可大致称非加性效应值为杂合效应值。

#### (5) 近交使群体分化，杂交使群体一致

在一对等位基因 (A, a) 遗传的群体中，同型交配的最终结果是使群体分化成两个纯系，AA系与aa系。在两对基因 (A, a及B, b) 的情况下，分化成AABB、AAbb、aaBB和aabb 4 个不同的纯系。杂交则相反，它能使个体的基因型杂合化，因而使群体趋向一致。最为典型的情况是两个纯系AA与aa间的杂交， $F_1$ 群体中100%为Aa的杂合体，群体达到了完全的一致性。

#### (6) 近交与人工选择相结合是提高杂种优势的重要手段之一

通过人工选择保留理想的纯合体，近交与人工选择结合能较快地加大群体间基因频率的差异，因而也成为提高杂种优势的有力手段。



## 6.5.2 杂种优势及其遗传理论

### (1) 杂种优势

杂种优势 (heterosis) 是指两个遗传组成不同的亲本杂交产生的杂种一代 ( $F_1$ )，在生长势、生活力、繁殖力、适应性以及产量和品质等性状上比双亲优越的现象。

利用杂种第一代这种超亲现象，以获得更大的经济效益，称为杂种优势利用。

$$\bar{x}_{F_1} > \frac{\bar{x}_{p1}}{2} + \frac{\bar{x}_{p2}}{2}$$



## (2) 杂种优势表现的特点:

### ① 杂种优势的复杂多样性。

杂种优势的表现因组合不同、性状不同、环境条件不同而呈复杂多样性。

② 杂种优势强弱和亲本的遗传差异及纯度密切相关。杂交亲本间的遗传差异愈大，杂种优势愈明显；在双亲亲缘关系和性状有一定的差异的前提下，亲本的纯度愈高，则杂种优势愈强。

### ③ $F_2$ 及以后世代杂种优势衰退

例： $aa \times AA \rightarrow F_1 (Aa)$

$Aa \times Aa \rightarrow F_2 (1AA: 2Aa: 1aa)$

只有一半的杂合基因型个体表现为杂种优势，另一半纯合个体基因型的性状趋向双亲，不表现杂种优势。因此杂交种只能种植一代，必须年年制种。

### ④ 细胞质对 $F_1$ 杂种优势的影响。





### (3) 杂种优势形成的机制有两种学说：

#### ① 显性说 (dominance hypothesis, Bruce, 1910)

异花授粉的植物和异体受精的动物，杂合性程度很高，隐性的有害基因，近亲繁殖或自交，后代逐渐纯合化，分离出形态和生理状况不良的个体，使它们的生活力减退，适应度降低。所以近亲繁殖带来的不良效应和衰退是由于原来处于杂合性状的基因发生分离的结果。

若双亲对很多座位上的不同等位基因是纯合体，形成杂种后，显性的有利基因的效应集积起来，而隐性有害基因的作用被遮盖起来，出现了明显的优势。

即:杂种优势是双亲的显性基因集中在杂种中所引起的互补作用

P AAbbCCDDee × aaBBccddEE



F<sub>1</sub> AaBbCcDdEe (出现杂种优势)

野生型基因一般是显性的，显性基因多编码具有生物学活性的蛋白质，突变基因一般是隐性的，隐性基因多编码失去或降低活性的蛋白质。因此杂合体AaBb的生活力高于纯合体AAbb或aaBB。这便是显性假说所说的杂种优势的生化基础。





② 超显性说 (overdominance hypothesis) :

East和Shull 1918年提出基因处于杂合态时比两个纯合态都好

$$P \quad a_1a_1b_1b_1c_1c_1d_1d_1e_1e_1 \times a_2a_2b_2b_2c_2c_2d_2d_2e_2e_2$$

$$F_1 \quad a_1a_2b_1b_2c_1c_2d_1d_2e_1e_2 \quad (\text{出现杂种优势})$$

该假说认为同基因座杂基因间及不同基因座之间，是复杂的互作关系，而不是显隐性关系，杂种优势的形成来源于基因的互作，其中包括同基因座在杂合状态下基因之间的互作和不同基因之间的基因互作（亦称上位性互作）。在杂合状态下，基因表达水平可大大超过纯合状态的基因表达水平。由于基因型在杂合状态下的性状表现超过基因型在显性情况下的表现，故称为超显性现象。



## 超显性假说所说的杂种优势的生化基础两种可能情况：

①两个等位基因各自编码一种蛋白质，这两种蛋白质的相互作用的结果比各自独立存在更有利于个体的生存。例如人类的镰刀形血红蛋白杂合体（HbA / HbS）的红细胞中同时存在着两种血红蛋白：成人血红蛋白（HbA）和镰刀形细胞血红蛋白（HbS）。杂合体既不是贫血症患者，又较不易为疟原虫感染，因而在疟疾流行的地区更有利于生存。

②两个杂合等位基因所编码的多肽结合成为活力高于相同亚基所形成的蛋白质。等位基因的这一相互作用形式至少曾经在粗糙脉孢菌的谷氨酸脱氢酶基因中发现。



上述两种假说一个强调显性基因作用，另一个强调基因间相互作用，它们既非互相排斥又不能概括一切。根据数量性状的遗传分析，杂种优势的遗传实质在于**显性效应**、**累加效应**以及**异位显性**、**互补作用**和**超显性**等各种基因互作效应。在某一材料中可能只有某一种基因互作决定某种程度的杂种优势

谢谢!

习题

P146-4、6、8





## 补： 人类遗传学图谱构建中连锁与否的判断

人类遗传连锁图的构建是以系谱分析为基础，通过重组率的计算排列出不同多态位标的遗传距离而构建的。

在人类遗传学研究中，由于通常不知道父母的基因型和父母中标记基因的连锁相是相斥还是相引，因而无法简单地通过计算重组体出现的频率来进行连锁分析，而必须通过适当的统计模型来估算重组率，并采用**或然比**（似然比 likelihood ratio）检验的方法来推断连锁是否存在，





即比较假设两个座位间存在连锁 ( $r < 0.5$ ) 的概率与假设没有连锁 ( $r = 0.5$ ) 的概率。这两种概率之比可以用似然比统计量来表示:

即  $L(r) / L(0.5)$ , 其中  $L()$  为似然函数, 为方便起见, 常将  $L(r) / L(0.5)$  取以10为底的对数, 称为LOD值(Lod score)或对数优势比 (logarithms of the odds)。为了确定两对基因之间是否存在连锁, 一般要求似然比大于1000: 1, 即LOD > 3; 而要否定连锁存在, 则要求似然比小于100: 1, 即LOD < 2.

[返回](#)

[返回 \(93\)](#)



武汉大学

Wuhan University

在其他生物遗传图谱的构建中，似然比的概念也用来反映重组值的可靠程度或作为连锁是否真实存在的一种判断尺度。





## 1. 重组率与非重组率的计算

例：以下4个家系显示多态DNA标记与由一个显性等位基因D决定的性状的分离情况。

M和m分别代表分子标记的等位基因。

假设： 在每个家系中双杂合子亲本为MD/md

每个子代分为重组和非重组类型

在4个家系中子代只有家系（B）mD/md为重组

其余9人都是非重组

重组频率为： $r = 1/(1+9) = 0.1$

问题是数值太小以至于chi-square test 统计意义都不适合





我们如何确定这些家系中MD是否连锁？

第一步：使用二项式展开式

计算出每个家系中子代的概率（重组）

仅仅考虑是重组子还是非重组子，

不考虑性别及其他的问题。

在家系（B）中，出现1个重组子和3个非重组子，

所以其出现的概率是：

$$(4! / 1!3!) r^1(1-r)^3 = 4r(1-r)^3$$

每个家系中子代出现的概率都表示在方框图中第1行

第二步：对每个家系计算子代中重组子和非重组子分布

的概率，假设 $r=0.5$ （独立分配），基因位于

不同的染色体上或在相同的染色体上被分离

（非重组）。这些概率表示在方框图中第2行





## 2. 或然率的计算

$$\begin{aligned} \text{(A) Likelihood ratio} &= L(r) / L(0.5) \\ &= (1-r) / 1/2 = 2(1-r) \end{aligned}$$

$$\text{Lod} = \log_{10}(2) + \log_{10}(1-r)$$

$$\begin{aligned} \text{(B) Likelihood ratio} &= L(r) / L(0.5) \\ &= 4r(1-r)^3 / 4(1/2)^4 = 16r(1-r)^3 \end{aligned}$$

$$\text{Lod} = \log_{10}(16) + \log_{10}(r) + 3\log_{10}(1-r)$$

.....

