

# 第7章

## 双亲杂交后代的遗传分析

王建康

中国农业科学院作物科学研究所

[wangjiankang@caas.cn](mailto:wangjiankang@caas.cn)

<http://www.isbreeding.net>

# 本章的主要内容

- § 7.1 单环境多基因型表型数据的方差分析
- § 7.2 六个基本世代均值和方差的构成
- § 7.3 自交后代均值和方差的分解
- § 7.4 基因间的上位性互作

# § 7.1 单环境多基因型表型数据的方差分析

- § 7.1.1 表型观测值的线性分解
- § 7.1.2 表型离差平方和的分解
- § 7.1.3 遗传效应的差异显著性检验
- § 7.1.4 基因型值预测和广义遗传力估计

# 遗传群体及其表型鉴定

- 假定一个遗传群体由 $g$ 个不同的基因型（可以看作是个体或家系）组成，这些基因型可以来自两个或多个亲本的某个杂交世代，也可以来自一个自然遗传群体，还可以是一定时期推广的栽培品种或者是种质资源库中保存的一批遗传材料等等。
- 在特定环境条件下，每个基因型获得 $r$ 次重复表型观测数据。不同基因型的平均表现或均值也不同，每个基因型都可被看作是一个抽样总体，一个基因型的 $r$ 次重复观测数据是来自该基因型总体的一组样本。尽管存在基因型的差异，但可以认为它们的方差是相同的或同质的，均等于随机误差方差。

# 表型观测值的分布

- 用 $\mu_i$ 表示第 $i$ 个基因型的均值或平均表现，是一个待估计的未知参数。在观测误差服从均值是0、方差是 $\sigma_\varepsilon^2$ 的正态分布，且相互独立的假定下，第 $i$ 个基因型的第 $k$ 个表型值 $y_{ik}$ 则服从下面的正态分布。

$$y_{ik} \sim N(\mu_i, \sigma_\varepsilon^2)$$

其中， $i=1, 2, \dots, g$ ， $k=1, 2, \dots, r$

# 表型观测值的线性分解

- 用 $\mu$ 表示 $g$ 个基因型的总平均，每个基因型的平均表现 $\mu_i$ 与总平均 $\mu$ 之间的差异定义为该基因型的遗传效应，用 $G_i$ 表示。表型值 $y_{ik}$ 还可以用线性模型表示出来。

$$\mu = \frac{1}{g} \sum_{i=1}^g \mu_i \quad G_i \hat{=} (\mu_i - \mu)$$

$$y_{ik} = \mu + G_i + \varepsilon_{ik} \quad \varepsilon_{ik} \sim N(0, \sigma_{\varepsilon}^2) \text{ iid}$$

其中 $i=1, 2, \dots, g, k=1, 2, \dots, r$

# 表型方差的分解和广义遗传力

- 遗传方差  $\sigma_G^2 \hat{=} \frac{1}{g-1} \sum_{i=1}^g G_i^2$

其中， $g-1$ 称为遗传效应或遗传方差的自由度

- 单环境表型方差的分解  $\sigma_P^2 = \sigma_G^2 + \sigma_\varepsilon^2$
- 广义遗传力 $H^2$ 定义为遗传方差 $\sigma_G^2$ 占表型方差 $\sigma_P^2$ 的比例。遗传方差和遗传力 $H^2$ 是重要的群体遗传参数，从它们的估计值，可以了解一个群体由基因型所决定变异的大小，以及占表型变异的比例。

# 样本均值的分布和观测值的离差

- 总平均数的分布


$$\bar{y} = \frac{1}{gr} \sum_{i,k} y_{ik} \sim N\left(\mu, \frac{1}{gr} \sigma_{\varepsilon}^2\right)$$

- 单个基因型的平均数分布

$$\bar{y}_{i\cdot} = \frac{1}{r} \sum_k y_{ik} \sim N\left(\mu_i, \frac{1}{r} \sigma_{\varepsilon}^2\right)$$

- 样本观测值离差的构成

$$y_{ik} - \bar{y} = (\bar{y}_{i\cdot} - \bar{y}) + (y_{ik} - \bar{y}_{i\cdot})$$

$$y_{ik} = \mu + G_i + \varepsilon_{ik}$$




# 离差平方和的分解 $SS_T = SS_G + SS_\varepsilon$

- 总平方和（或总效应平方和）

$$SS_T = \sum_{i,k} (y_{ik} - \bar{y})^2 = \sum_{i,k} (y_{ik} - \bar{y}_{i\cdot})^2 + r \sum_i (\bar{y}_{i\cdot} - \bar{y})^2$$

- 误差平方和（或误差效应平方和），包含着误差方差 $\sigma_\varepsilon^2$ 的信息

$$SS_\varepsilon = \sum_{i,k} (y_{ik} - \bar{y}_{i\cdot})^2$$

- 基因型平方和（或基因型效应平方和），包含着遗传方差 $\sigma_G^2$ 的信息

$$SS_G = r \sum_i (\bar{y}_{i\cdot} - \bar{y})^2$$

# 效应平方和的期望

- 误差平方和的期望

$$\begin{aligned} E(SS_{\varepsilon}) &= E \sum_{i,k} (y_{ik} - \bar{y}_{i\cdot})^2 = E \sum_{i,k} [(y_{ik} - \mu_i) - (\bar{y}_{i\cdot} - \mu_i)]^2 \\ &= g(r-1)\sigma_{\varepsilon}^2 \end{aligned}$$

- 基因型平方和的期望

$$\begin{aligned} E(SS_G) &= rE \sum_i (\bar{y}_{i\cdot} - \bar{y})^2 = rE \sum_i [(\bar{y}_{i\cdot} - \mu_i) - (\bar{y} - \bar{\mu}) + (\mu_i - \bar{\mu})]^2 \\ &= (g-1)\sigma_{\varepsilon}^2 + (g-1)r\sigma_G^2 \end{aligned}$$

# 均平方（简称均方）及其期望

- 误差均方及其期望， $g(r-1)$ 为其自由度

$$MS_{\varepsilon} = \frac{SS_{\varepsilon}}{g(r-1)} \quad E(MS_{\varepsilon}) = \sigma_{\varepsilon}^2$$

- 基因型均方及其期望， $(g-1)$ 为其自由度

$$MS_G = \frac{SS_G}{g-1} \quad E(MS_G) = \sigma_{\varepsilon}^2 + r\sigma_G^2$$

# 方差的无偏估计

- 误差方差的无偏估计

$$\hat{\sigma}_{\varepsilon}^2 = MS_{\varepsilon}$$

- 基因型方差（遗传方差）的无偏估计

$$\hat{\sigma}_G^2 = \frac{1}{r} (MS_G - MS_{\varepsilon})$$

# 群体遗传效应的差异显著性检验

- 很多时候，我们首先关心的是，群体的遗传方差是否显著，或者说 $g$ 个基因型是否有显著不同的均值或平均表现。如果在平均表现上无显著差异，说明它们的遗传效应均与0无显著差异，也就是对下面的统计假设进行检验。

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_g \text{ 或}$$

$$H_0 : G_1 = G_2 = \cdots = G_g = 0$$

# 差异显著性的检验过程

- 当零假设成立时，效应方差 $\sigma_G^2=0$ 。这时从 $MS_G$ 的期望公式看出，遗传效应均方应该与误差均方相差不大；当零假设不成立时，效应方差 $\sigma_G^2>0$ ，遗传效应均方应该明显大于误差均方。

- 统计上通常利用效应均方与误差均方的比值，来检验基因型间是否有显著性差异，这个比值服从自由度是 $(g-1)$ 和 $g(r-1)$ 的 $F$ 分布，即

$$F = \frac{MS_G}{MS_\varepsilon} \sim F[g-1, g(r-1)]$$

- 如果 $F$ 值的显著性概率没有超过一定的显著性水平，一般取0.05或0.01，说明基因型间无显著差异，该遗传群体可能不适合这一观测性状的遗传研究。但要注意的是，这并不是说该群体对其它性状也不适合。同时还要考虑表型测量误差是否太大，群体种植的环境是否有代表性等因素。

# 方差分析表

变异来源	自由度	平方和	均方	F统计量	期望均方	方差估计值
Source	Degree of freedom	Sum of squares	Mean square	F-value	Expected MS	Estimate of variance
基因型	$g-1$	$SS_G$	$MS_G$	$MS_G/MS_\varepsilon$	$\sigma_\varepsilon^2 + r\sigma_G^2$	$(MS_G - MS_\varepsilon)/r$
随机误差	$g(r-1)$	$SS_\varepsilon$	$MS_\varepsilon$		$\sigma_\varepsilon^2$	$MS_\varepsilon$
总离差	$gr-1$	$SS_T$				

# 完全随机区组设计的观测数据

- 如田间试验采用的是完全随机区组设计，将每个重复安排在差异相对较小的一个区组内，重复的个数就等于区组的个数，区组间的差异有时也可能是显著的。这时的方差分析模型中，还应包含区组的效应。

$$y_{ik} - \bar{y} = (\bar{y}_{i\cdot} - \bar{y}) + (\bar{y}_{\cdot k} - \bar{y}) + (y_{ik} - \bar{y}_{i\cdot} - \bar{y}_{\cdot k} + \bar{y})$$

$$y_{ik} = \mu + G_i + B_k + \varepsilon_{ik}$$



# 完全随机区组设计的方差分析表

变异来源	自由度	平方和	均方	F统计量	期望均方	方差估计值
区组或重复	$r-1$	$SS_B$	$MS_B$	$MS_B/MS_\varepsilon$	$\sigma_\varepsilon^2 + g\sigma_R^2$	$(MS_G - MS_\varepsilon)/g$
基因型	$g-1$	$SS_G$	$MS_G$	$MS_G/MS_\varepsilon$	$\sigma_\varepsilon^2 + r\sigma_G^2$	$(MS_G - MS_\varepsilon)/r$
随机误差	$(g-1) \times (r-1)$	$SS_\varepsilon$	$MS_\varepsilon$		$\sigma_\varepsilon^2$	$MS_\varepsilon$
总离差	$gr-1$	$SS_T$				

# 基因型值与表型平均数

- 每个遗传研究群体，都会包含多种基因型。遗传和育种研究中，首先关心的是这些基因型在一种或多种环境下的平均表现是什么，或基因型值的大小，在此基础上才能开展进一步的遗传研究（如基因定位和克隆）和育种应用（如根据基因型的构成来选择优良个体）。
- 在频率派统计学看来，基因型的平均表现是一个确定的数值，但事先是未知的，表型鉴定的首要目的就是估计这些未知参数。在 § 6.3 中我们已经看到，大多数情况下，基于表型观测数据的样本均值就是基因型值很好的估计，有时也把基因型值称为基因型的平均表现。

# 基因型值的预测（或估计）

- 单环境试验中，从分布模型公式7.1可以看出观测值 $y_{ik}$ 包含了第 $i$ 个基因型值的信息。对基因型 $i$ 来说，在误差效应服从独立正态分布的假定下，重复间平均数是基因型值 $\mu_i$ 的最优线性无偏估计（best linear unbiased estimate, 简称BLUE），估计值的方差等于误差方差除以重复数，即

$$\hat{\mu}_i = \frac{1}{r} \sum_k y_{ik} = \bar{y}_i. \quad V(\hat{\mu}_i) = \frac{1}{r} \sigma_\varepsilon^2$$

# 遗传力及其估计

- 遗传力是一个群体概念，分广义遗传力和狭义遗传力两种。广义遗传力（heritability in the broad sense）是群体中所有遗传因素产生的方差占表型方差的比例，狭义遗传力（heritability in the narrow sense）特指育种值产生的加性遗传方差占表型方差的比例。
- 方差分析是估计这两种遗传力最直接、最常用的方法。狭义遗传力的计算要用到加性方差，估计时需要用到多个遗传群体，有时还要知道不同群体中个体之间的亲缘关系，这方面的内容将在第8章详细介绍。

# 广义遗传力的估计

- 广义遗传力的估计，一般不需要知道个体间的亲缘关系，只要能够估计出群体的总遗传方差和误差方差就可以。广义遗传力用符号 $H^2$ 表示，把方差分析表中得到的方差估计值代入下面的公式，就得到 $H^2$ 的估计值。

$$H^2 = \frac{\sigma_G^2}{\sigma_P^2} = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_\varepsilon^2}$$

- 上面公式可视为单个表型观测值的遗传力。这里的单个观测单位可以是个体水平，也可能是小区水平，依具体试验而定。

# 重复平均数的遗传力

- 对于重复观测数据，一般用重复平均数（即样本均值）进行基因定位和育种选择，其中的遗传方差与单个观测数据中的遗传方差相同，但是误差方差只有单个观测数据误差方差的 $1/r$ 。重复平均数的表型方差和重复平均数的遗传力公式为

$$\sigma_{\bar{P}}^2 = \sigma_G^2 + \frac{1}{r} \sigma_{\varepsilon}^2 \quad H^2 = \frac{\sigma_G^2}{\sigma_{\bar{P}}^2} = \frac{\sigma_G^2}{\sigma_G^2 + \frac{1}{r} \sigma_{\varepsilon}^2}$$

- 由于降低了随机误差方差，重复平均数有较高的遗传力。因此，在报道遗传力时，还要指明表型方差是来自单个观察值，还是来自重复平均数。

# 水稻双亲衍生的10个RIL家系在三个环境下的直链淀粉含量（%）

基因型	环境I		环境II		环境III	
	重复1	重复2	重复1	重复2	重复1	重复2
RIL1	15.3	15.1	14.4	14.6	14.5	14.8
RIL2	14.5	15.0	15.8	15.7	16.3	15.7
RIL3	14.0	14.9	15.9	15.8	15.2	16.1
RIL4	13.2	14.0	16.0	16.8	15.0	15.5
RIL5	15.4	15.9	16.7	16.6	15.4	15.6
RIL6	15.5	15.6	16.1	16.7	16.0	17.0
RIL7	13.2	14.1	14.3	14.9	14.1	14.5
RIL8	11.9	12.6	14.2	14.4	15.2	16.1
RIL9	12.8	13.5	14.5	14.6	15.3	15.5
RIL10	12.8	13.6	14.6	15.5	14.2	14.2

# 单环境方差分析（不考虑区组效应）

环境	变异来源	自由度	平方和	均方	F值	显著概 率	方差估 计值
I	基因型间	9	23.655	2.628	11.866	0.0003	1.203
	随机误差	10	2.215	0.222			0.222
	总离差	19	25.870				
II	基因型间	9	14.785	1.643	14.347	0.0001	0.764
	随机误差	10	1.145	0.115			0.115
	总离差	19	15.930				
III	基因型间	9	9.718	1.080	6.066	0.0047	0.451
	随机误差	10	1.780	0.178			0.178
	总离差	19	11.498				



# 单环境方差分析（考虑区组效应）

环境	自由度			均方			方差估计值		
	机误	区组	基因型	机误	区组	基因型	机误	区组	基因型
I	9	1	9	0.066	1.625	2.628	0.066	0.156	1.281
II	9	1	9	0.074	0.481	1.643	0.074	0.041	0.784
III	9	1	9	0.118	0.722	1.080	0.118	0.060	0.481

# 方差分析全模型和简化模型

- 方差分析模型中效应类型的增加，一定会降低误差的平方和，同时也降低误差的自由度。均方是否会下降，要看所增加效应的平方和是否足够大。
- 严格地讲，方差分析应该从包含尽可能多的效应模型开始，这样的模型又称全模型（full model）。在全模型下， $F$ 检验不显著的效应应该从方差分析模型中剔除出去，然后对剔除不显著效应之后的模型重新进行分析，剔除不显著效应之后的模型称为简化模型（reduced model）。
- 前面的数据，区组效应在三个环境中均达到0.05的显著性水平，方差分析中考虑区组是恰当的，不用考虑简化模型的方差分析。

# § 7.2 六个基本世代均值和方差的构成

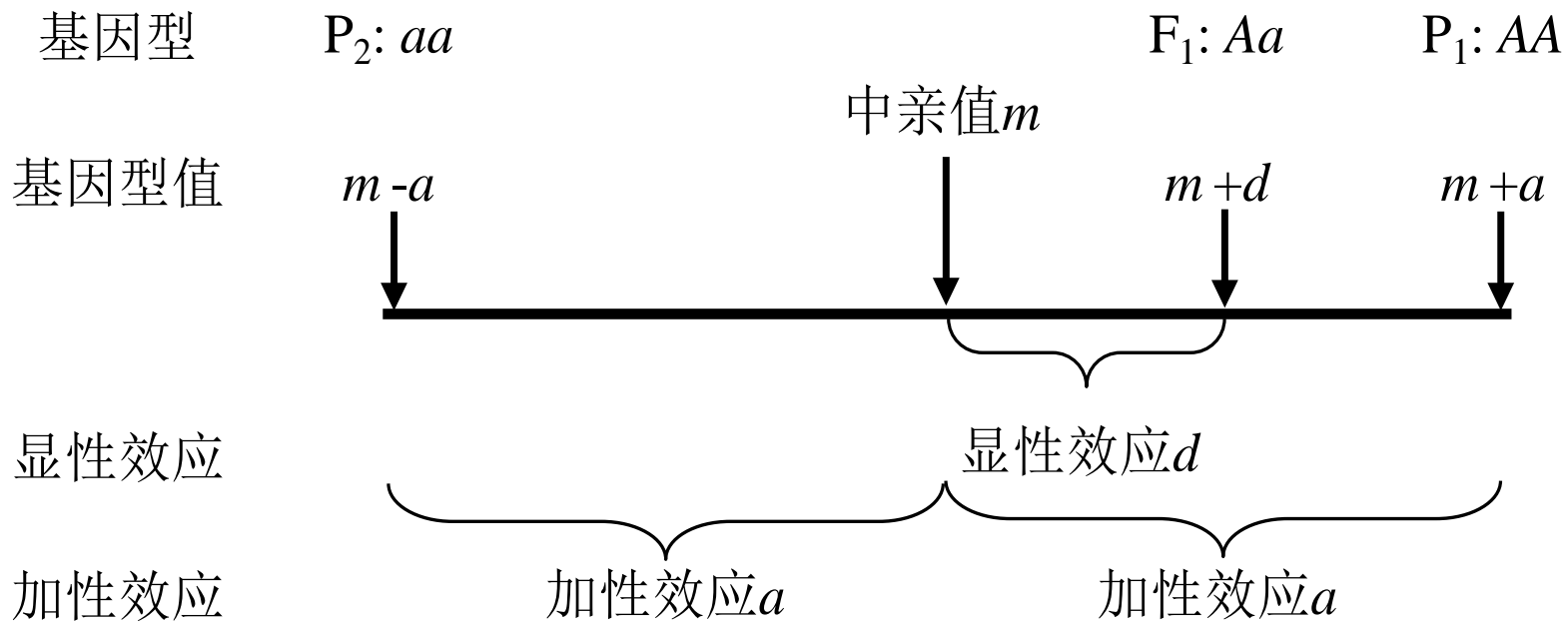
- § 7.2.1 单基因座位的加显性遗传模型
- § 7.2.2 多基因座位的加显性遗传模型
- § 7.2.3 分离世代的均值与遗传方差
- § 7.2.4 遗传方差和遗传力的估计
- § 7.2.5 有效因子个数的估计
- § 7.2.6 加显性模型的检验与实例分析

# 加显性遗传模型

- 数量遗传中，常假定不同座位上的遗传效应是可加的。每个座位上，纯合基因型的效应称为加性效应，杂合基因型的效应称为显性效应。显性效应是同一个座位上、不同等位基因结合在一起、产生出有别于纯合基因型平均表现的效应，有时也称座位内基因互作（intra-genic interaction）。
- 个体未必在所有座位上都是纯合或是杂合的。这时，个体在遗传构成上，会同时包含加性效应和显性效应。
- 将基因型值分解为加性效应和显性效应的模型称为加显性模型（additive and dominance model），是数量遗传中最简单、最基本的模型。

# 单座位（或单基因）加显性模型

- 先讨论一对等位基因（用 $A$ 和 $a$ 表示）的模型，称单座位或单基因模型（single locus model）。假定基因型 $Aa$ 是两个纯合基因型亲本 $AA$ 和 $aa$ 的杂交后代，三种基因型的平均表现由下图的箭头指明。



# 中亲、加性效应和显性效应

- $m$ 表示中亲值，即两个亲本表现的平均数。
- 亲本离中亲值的距离称为加性效应（additive effect），用 $a$ 表示，也可看作高值亲本（AA）与低值亲本（aa）之差的一半。
- $F_1$ 平均表现与中亲值 $m$ 之间的离差称为显性效应（dominant effect），用 $d$ 表示。

$$m = \frac{1}{2}(\bar{P}_1 + \bar{P}_2) \quad a = \frac{1}{2}(\bar{P}_1 - \bar{P}_2) \quad d = \bar{F}_1 - \frac{1}{2}(\bar{P}_1 + \bar{P}_2) = \bar{F}_1 - m$$

$$\bar{P}_1 = m + a \quad \bar{P}_2 = m - a \quad \bar{F}_1 = m + d$$

# 显性度

- 显性效应与加性效应的比值 $d/a$ ，称为基因的显性度（degree of dominance）。
- 狭义地看， $a$ 为正值；广义地看， $a$ 只是加性效应的代数符号，可以为正值，也可以为负值；显性度既可以是正，也可以是负，其绝对值可以小于1，也可以大于1。

# 根据显性度的划分遗传模型

- 加性模型（additive model）的显性度为0， $F_1$ 的表现正好等于中亲值；部分显性（partial dominance model）的显性度介于0和1之间， $F_1$ 的表现介于中亲和亲本AA之间；部分隐性（partial recessive model）的显性度介于-1和0之间， $F_1$ 的表现介于亲本aa和中亲之间，也称负向部分显性；
- 显性或完全显性（dominance model or complete dominance model）的显性度为1， $F_1$ 的表现正好与亲本AA完全相同；隐性或完全隐性（recessive model or complete recessive model）的显性度为-1， $F_1$ 的表现正好与亲本aa完全相同，也称负向显性；
- 超显性（over-dominance model）的显性度大于1， $F_1$ 的表现高于亲本AA；超隐性（over-recessive model）的显性度小于-1， $F_1$ 的表现低于亲本aa，也称负向超显性。

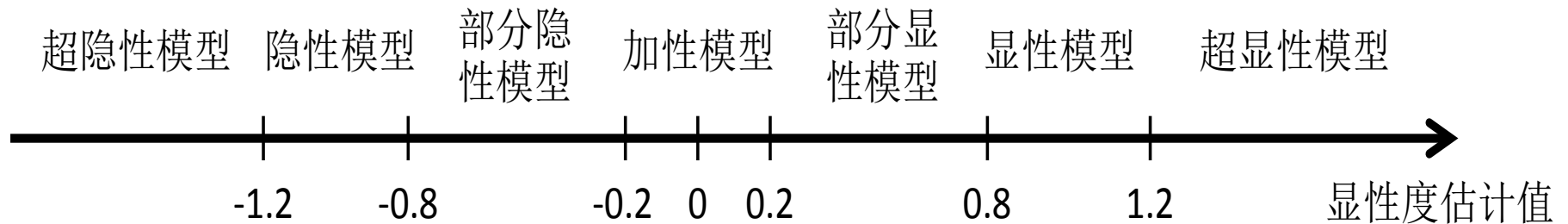


# 显隐性的相对性

- 从绝对值上看，部分显性和部分隐性显性度的绝对值低于1，显性和隐性显性度的绝对值等于1，超显性和超隐性显性度的绝对值大于1。
- 在第1章中曾说过，性状的显隐性与等位基因的显隐性都是相对概念。上面按照加显性效应定义的各种模型也是相对的，关键是要看两个亲本中把谁的基因型视为 $AA$ 、谁的视为 $aa$ 。

# 根据估计值划分遗传模型

- 实际数据估计出的加显性效应都存在一定的误差，加性座位上显性度的估计值也不会正好就等于0，显性座位上的估计值也不会正好就等于1。这时，可参照下图的标准来划分前面的7种遗传模型。



- 以-1、0、1左右0.2来划分遗传模型带有人为因素，如果遗传群体很大，遗传效应的估计误差很小，使用0.1为标准也未尝不可。如果控制性状的基因数目很多，遗传效应不能准确估计，还可考虑使用0.25的标准。

# 例子

- 利用20个单株得到两个纯系亲本的株高平均数分别为116.30cm和98.45cm，60个F<sub>1</sub>单株得到的株高平均数为117.68cm。
- 假定株高差异只受一个座位上两个等位基因的控制，根据单基因加显性模型得到中亲值 $m$ 、加性效应 $a$ 、显性效应 $d$ 的估计值分别为107.38、8.93、10.31。显性度的估计值为1.15。可以认为高秆等位基因相对于矮秆等位基因表现为显性。
- 当然，如认为株高的测量误差很小，三个群体的样本平均数都非常接近真实的群体均值，也可以认为高秆等位基因表现出一定的超显性。

# 多基因控制的性状

- 如果性状受多个座位上基因的控制，基因型平均数可按单基因效应累加的方法表示。
- 如考虑三个座位A、B、C，大写字母表示增效等位基因，小写字母表示减效等位基因。
- 亲本P<sub>1</sub>的基因型为*AAbbcc*，
- 亲本P<sub>2</sub>的基因型为*aaBBCC*，
- 它们杂交F<sub>1</sub>的基因型为*AaBbCc*。

# 多基因的加显性效应

- 不同座位的遗传效应用座位名称作为下标来区分。对于亲本 $P_1$ 来说，座位A上是纯合基因型 $AA$ ，基因型值中应该加上加性效应 $a_A$ ；座位B上是纯合基因型 $bb$ ，基因型值中应减去加性效应 $a_B$ ；座位C上是纯合基因型 $cc$ ，基因型值中应减去加性效应 $a_C$ 。因此，亲本 $P_1$ 的基因型值等于 $m+a_A-a_B-a_C$ 。
- 与亲本 $P_1$ 类似，得到亲本 $P_2$ 的基因型值为 $m-a_A+a_B+a_C$ 。
- 两个亲本杂交得到的 $F_1$ 在三个座位上都是杂合基因型，基因型值中应该加上这三个座位上的显性效应 $d_A$ 、 $d_B$ 、 $d_C$ 。

# 多基因的群体均值

- 下面的等式分别给出这三种单一基因型群体的均值表达式。

$$\bar{P}_1 = m + a_A - a_B - a_C$$

$$\bar{P}_2 = m - a_A + a_B + a_C$$

$$\bar{F}_1 = m + d_A + d_B + d_C$$

# 多基因的加显性效应

- 对于多基因控制的性状，亲本间差异是多个座位上等位基因差异引起的。数量性状的基因型往往难以鉴定，单个座位的加显性效应不易计算，不能从单个座位的效应来计算不同基因型的均值。但可以沿用单座位模型的效应定义方法，

$$m = \frac{1}{2}(\bar{P}_1 + \bar{P}_2) \quad \Sigma a = \frac{1}{2}(\bar{P}_1 - \bar{P}_2) = a_A - a_B - a_C$$

$$\Sigma d = \bar{F}_1 - \frac{1}{2}(\bar{P}_1 + \bar{P}_2) = d_A + d_B + d_C$$

- 因此有，

$$\bar{P}_1 = m + \Sigma a \quad \bar{P}_2 = m - \Sigma a \quad \bar{F}_1 = m + \Sigma d$$

# 多基因的加显性遗传模型

- 与单座位模型相对照可以看到，如果把那里的加显性效应用多基因加显性效应代替的话，得到的就是多座位模型。
- 因此，在不同座位之间基因型值可加的假定下，多基因座位加显性模型就是单座位模型的推广，它们统称为加显性效应模型，或简称加显性模型。
- 后续章节里只侧重于单基因座位模型的推导。



# 分离世代的均值与遗传方差

- 分离世代的群体均值和方差，除了依赖于各种基因型的平均表现外，还依赖于各种基因型的频率。频率越高的基因型对群体均值的贡献当然也就越大。
- 对于单基因加显性模型来说，基因型 $AA$ 、 $Aa$ 、 $aa$ 的均值用 $X_2$ 、 $X_1$ 、 $X_0$ 表示，分别等于 $m+a$ 、 $m+d$ 、 $m-a$ 。三种基因型的频率分别用 $f_2$ 、 $f_1$ 、 $f_0$ 表示，均值和频率中的下标可以理解成等位基因 $A$ 的个数。
- 群体均值和遗传方差的计算公式如下，与上一章的公式6.37和公式6.38在分组数 $k=3$ 的情况下完全一致。

$$\bar{X} = f_2 X_2 + f_1 X_1 + f_0 X_0$$

$$V_X = f_2 X_2^2 + f_1 X_1^2 + f_0 X_0^2 - \bar{X}^2$$

# 回交世代的均值与遗传方差

- $B_1$ 表示 $F_1$ 与亲本 $P_1$ 的回交，它的均值和遗传方差如下：

$$\bar{B}_1 = m + \frac{1}{2}a + \frac{1}{2}d \quad V_{B_1} = \frac{1}{4}(a - d)^2$$

- $B_2$ 表示 $F_1$ 与亲本 $P_2$ 的回交，它的均值和遗传方差如下：

$$\bar{B}_2 = m - \frac{1}{2}a + \frac{1}{2}d \quad V_{B_2} = \frac{1}{4}(a + d)^2$$

# F<sub>2</sub>世代的均值与遗传方差

- F<sub>2</sub>群体中三种基因型的频率分别为1/4、1/2、1/4，均值分别为 $m+a$ 、 $m+d$ 、 $m-a$ ，F<sub>2</sub>的均值是这三种基因型值的加权平均。

$$\begin{aligned}\bar{F}_2 &= \frac{1}{4}(m+a) + \frac{1}{2}(m+d) + \frac{1}{4}(m-a) \\ &= m + \frac{1}{4}(a) + \frac{1}{2}(d) + \frac{1}{4}(-a) = m + \frac{1}{2}d\end{aligned}$$

- F<sub>2</sub>世代的遗传方差计算为：

$$V_{F_2} = \frac{1}{4}(a)^2 + \frac{1}{2}(d)^2 + \frac{1}{4}(-a)^2 - \left(\frac{1}{2}d\right)^2 = \frac{1}{2}a^2 + \frac{1}{4}d^2$$

# 多基因的回交和F<sub>2</sub>世代均值与遗传方差

- 如果性状受多基因控制,

$$\bar{B}_1 = m + \frac{1}{2}\Sigma a + \frac{1}{2}\Sigma d \quad V_{B_1} = \frac{1}{4}\Sigma(a-d)^2 = \frac{1}{4}\Sigma a^2 + \frac{1}{4}\Sigma d^2 - \frac{1}{2}\Sigma ad$$

$$\bar{B}_2 = m - \frac{1}{2}\Sigma a + \frac{1}{2}\Sigma d \quad V_{B_2} = \frac{1}{4}\Sigma(a+d)^2 = \frac{1}{4}\Sigma a^2 + \frac{1}{4}\Sigma d^2 + \frac{1}{2}\Sigma ad$$

$$\bar{F}_2 = m + \frac{1}{2}\Sigma d \quad V_{F_2} = \frac{1}{2}\Sigma a^2 + \frac{1}{4}\Sigma d^2$$

# F<sub>2</sub>群体的遗传方差分解

- 纯系双亲衍生的后代群体中，F<sub>2</sub>既可以被看作F<sub>1</sub>的自交后代，也可以被看作F<sub>1</sub>的随机交配后代，每个座位上的基因频率与基因型频率满足HW平衡定律。因此，双亲遗传研究中常把F<sub>2</sub>视为参照群体或基础群体。
- 在F<sub>2</sub>遗传方差的计算公式中，通常用V<sub>A</sub>表示F<sub>2</sub>群体的加性方差，V<sub>D</sub>表示F<sub>2</sub>群体的显性方差。于是，在加显性效应模型下，F<sub>2</sub>群体的遗传方差可分解为加性方差和显性方差两部分。

$$V_A = \frac{1}{2} \sum a^2 \quad V_D = \frac{1}{4} \sum d^2 \quad V_{F_2} = V_A + V_D$$

# 多基因的平均显性度 $V_{F_2} = V_A + V_D$

- $F_2$ 群体中的加性方差是由加性效应决定的，此效应在随机交配和自交过程中都能遗传到后代群体中，同时又可以通过连续自交的方式固定下来。
- 显性方差是由显性效应所决定，此效应在随机交配过程中是不能遗传给后代的；自交过程中，显性效应也会逐渐消失，即不能通过连续自交的方式固定下来。
- 如果每个座位的基因效应相等，那么  $\sqrt{\frac{2V_D}{V_A}} = \frac{d}{a}$
- 因此，有时也利用上面的公式来估计多基因的平均显性度（average degree of dominance）。

# 回交群体遗传方差的和与差

- 两个回交群体遗传方差之和包含着1倍的加性方差和2倍的显性方差；两个遗传方差之差等于加显性效应的积和。

$$V_{B_1} + V_{B_2} = \frac{1}{2} \Sigma a^2 + \frac{1}{2} \Sigma d^2 = V_A + 2V_D$$

$$V_{B_2} - V_{B_1} = \Sigma ad$$

- 通常用 $\Sigma ad$ 的正负粗略判断亲本中增效基因的分布。如果 $\Sigma ad > 0$ ，说明亲本 $P_1$ 具有较多的增效基因；反之则说明亲本 $P_2$ 具有较多的增效基因。

# 加显性方差的估计

- 后面还会看到， $F_2$ 衍生的诸多家系群体，包括自交家系、全同胞家系、半同胞家系等，这些群体的遗传方差都是两种遗传方差的线性组合。因此，通过创造这些家系群体，就能估计 $F_2$ 基础群体中的加性方差 $V_A$ 和显性方差 $V_D$ 。
- 加性方差 $V_A$ 对于评价群体的育种价值、确定适宜的选择方法、遗传进度的预测等方面，都具有重要应用。这里先给出一种加性方差和显性方差的计算方法。

$$V_A = 2V_{F_2} - (V_{B_1} + V_{B_2}) \quad V_D = (V_{B_1} + V_{B_2}) - V_{F_2}$$



# 分离群体的遗传方差和遗传力估计

- 传统数量性状的遗传研究中，一般情况下都不知道单个基因的效应，不能从等式7.49和7.50来计算不同群体的遗传方差，并对遗传方差进行分解。
- 遗传分析可供利用的数据是个体或家系的表型数据，§ 7.1已经给出利用重复表型观测数据的方差分析估计遗传方差和遗传力的方法。这里介绍利用多个双亲群体的表型方差估计遗传方差的原理和方法。

# 随机误差方差的估计

- 当基因型与环境之间无互作时，纯系理论告诉我们，表型值等于基因型值加上随机误差效应，表型方差也就等于遗传方差加上误差方差，即

$$P = G + \varepsilon \quad W_P = V_G + V_\varepsilon$$

- 为与上一节的遗传方差区分，这里的表型方差和之后的样本方差一律用 $W$ 带下标表示。对于两个纯系 $P_1$ 和 $P_2$ 以及它们的 $F_1$ 而言，群体中的不同个体具有相同基因型，表型变异中无遗传变异，表型上的差异全部归于环境变异。因此，这三个群体的表型方差均可以用来估计误差方差。

$$W_{P_1} = W_{P_2} = W_{F_1} = V_\varepsilon$$

# 自由度不等时误差方差的估计

$$V_{\varepsilon} = \frac{df_{P_1} W_{P_1} + df_{F_1} W_{F_1} + df_{P_2} W_{P_2}}{df_{P_1} + df_{F_1} + df_{P_2}}$$

- 如果对三个不分离群体进行方差分析，则分子中的积和其实就是方差分析的总离差平方和。每个自由度等于样本大小减1，减去的1可以看作是计算样本均值时损失了一个自由度。总自由度等于总样本量减3，3可以看作在计算离差平方和时需要估计3个群体均值。
- 当三个自由度相差不大时，也可以用 $P_1$ 、 $F_1$ 、 $P_2$ 表型方差的加权平均数估计误差方差。

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

# 分离群体的遗传方差和遗传力

- 遗传方差的估计

$$V_{B_1} = W_{B_1} - V_{\varepsilon} \quad V_{B_2} = W_{B_2} - V_{\varepsilon} \quad V_{F_2} = W_{F_2} - V_{\varepsilon}$$

- 广义遗传力的估计

$$H_{B_1}^2 = \frac{V_{B_1}}{V_{B_1} + V_{\varepsilon}} \quad H_{B_2}^2 = \frac{V_{B_2}}{V_{B_2} + V_{\varepsilon}} \quad H_{F_2}^2 = \frac{V_{F_2}}{V_{F_2} + V_{\varepsilon}}$$

- 狭义遗传力的估计

$$h_{B_1}^2 = \frac{\frac{1}{2}V_A}{V_{B_1} + V_{\varepsilon}} \quad h_{B_2}^2 = \frac{\frac{1}{2}V_A}{V_{B_2} + V_{\varepsilon}} \quad h_{F_2}^2 = \frac{V_A}{V_{F_2} + V_{\varepsilon}}$$

# 有效因子个数的估计

- 关于控制数量性状基因数的估计，最早是1921年提出的Castle-Wright公式。
- 假定某一数量性状受 $k$ 个座位上基因的控制，每个座位有相同的加性效应 $a$ ，无显性效应，即 $d=0$ 。假定有两个亲本 $P_1$ 和 $P_2$ ， $P_1$ 集中了所有的增效基因， $P_2$ 集中了所有的减效基因， $P_1$ 与 $P_2$ 相差 $2k$ 个加性效应 $a$ 。

$$\bar{P}_1 = m + ka \quad \bar{P}_2 = m - ka \quad V_{F_2} = V_A = \frac{1}{2}ka^2$$
$$k = \frac{(\bar{P}_1 - \bar{P}_2)^2}{8V_{F_2}} = \frac{(\bar{P}_1 - \bar{P}_2)^2}{8(W_{F_2} - V_\varepsilon)}$$

# 有效因子个数的估计

- 如果存在显性效应，Castle-Wright公式就会低估基因个数。如能利用多个分离世代群体估计出加性方差，就能利用公式7.63估计基因个数，这时就可以去掉公式7.62的无显性假定。

$$k = \frac{(\bar{P}_1 - \bar{P}_2)^2}{8V_A}$$

- 这两个公式都是特定假设条件下对基因个数的一种估计。如果假设条件不成立，如基因效应不相等，增减效等位基因在两个亲本中分散分布，估计出的基因个数与真实基因个数之间就会有较大出入。为与真实基因个数相区分，这两个公式给出的估计，也称为有效因子数（number of effective factors）。

# 黄花烟草品系 ‘V22’ 和 ‘V73’ 及其杂交后代株高（cm）的样本均值和样本方差

世代	调查单株数	样本均值	样本方差	样本均值的方差
$P_1$	20	116.30	20.07	1.0034
$P_2$	20	98.45	29.05	1.4525
$F_1$	60	117.68	58.19	0.9699
$B_1$	120	116.00	58.66	0.4888
$B_2$	120	109.16	73.62	0.6135
$F_2$	160	111.78	78.66	0.4916

# 加显性模型的尺度检验

- 加显性模型下，6个基本世代平均数满足三个条件

$$A = 2\bar{B}_1 - \bar{P}_1 - \bar{F}_1 = 0$$

$$B = 2\bar{B}_2 - \bar{P}_2 - \bar{F}_1 = 0$$

$$C = 4\bar{F}_2 - 2\bar{F}_1 - \bar{P}_1 - \bar{P}_2 = 0$$

- 利用表7.7的数据，得到统计量 $A$ 、 $B$ 、 $C$ 分别为-1.98、2.19、-2.99，方差分别为3.93、4.88、14.20， $t$ 值分别为-1.00、0.99、-0.79，显著性概率分别为0.3186、0.3221、0.4278。因此，统计量 $A$ 、 $B$ 、 $C$ 均与0无显著差异，可以认为表7.7中的株高性状符合加显性遗传模型。



# 加显性模型的回归检验

$$\mathbf{y} = \begin{bmatrix} \bar{P}_1 = 116.30 \\ \bar{P}_2 = 98.45 \\ \bar{F}_1 = 117.68 \\ \bar{B}_1 = 116.00 \\ \bar{B}_2 = 109.16 \\ \bar{F}_2 = 111.78 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & 1 \\ 1 & 0.5 & 0.5 \\ 1 & -0.5 & 0.5 \\ 1 & 0 & 0.5 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} m \\ \Sigma a \\ \Sigma d \end{bmatrix}$$

估计方法	估计值			回归模型的矫正决定系数 $R^2$	回归模型显著性概率
	中亲值	加性效应	显性效应		
简单最小二乘	107.30	8.51	10.23	0.9856	<0.001
加权最小二乘	107.07	7.66	10.46	0.6666	<0.001

# 6个基本世代的遗传方差和遗传力

世代	表型方差 $W_x$	遗传方差 $V_x$	广义遗传力 $H^2$	狭义遗传力 $h^2$
$P_1$ 、 $P_2$ 和 $F_1$	45.02	0	0	0
$B_1$	58.66	13.64	0.2325	0.2134
$B_2$	73.62	28.60	0.3885	0.1700
$F_2$	78.66	33.64	0.4277	0.3183
平均显性度	0.83			
有效因子数 $k$	1.18 (公式7.62), 1.59 (公式7.63)			

$$V_A = 2V_{F_2} - (V_{B_1} + V_{B_2}) = 2 \times 33.64 - (13.64 + 28.60) = 25.04$$

$$V_D = (V_{B_1} + V_{B_2}) - V_{F_2} = (13.64 + 28.60) - 33.64 = 8.60$$

# 结果分析

- 从遗传效应估计值看，显性效应超过了加性效应。但是，由于加性方差是效应平方的0.5倍，显性方差是效应平方的0.25倍，因此加性方差仍有较高的估计值。
- $B_1$ 和 $B_2$ 的遗传方差中只包含0.5倍的加性方差 $V_A$ ， $F_2$ 的遗传方差中包含1倍的加性方差 $V_A$ 。利用公式7.59计算出的狭义遗传力见最后1列。
- 根据加显性方差估计得到的平均显性度估计值为0.83，按照图7.2的划分标准，属于完全显性。
- 根据公式7.62和7.63得到的有效因子数 $k$ 列于表7.9最后1行。由于显性效应的存在，公式7.62得到较低的估计值。但总的来说，在表7.7的群体中，控制株高性状的基因可能不会很多。

# § 7.3 自交后代均值和方差的分解

- § 7.3.1 F3世代均值和方差的构成
- § 7.3.2 F3世代中的环境方差
- § 7.3.3 任意自交世代衍生的后代家系群体

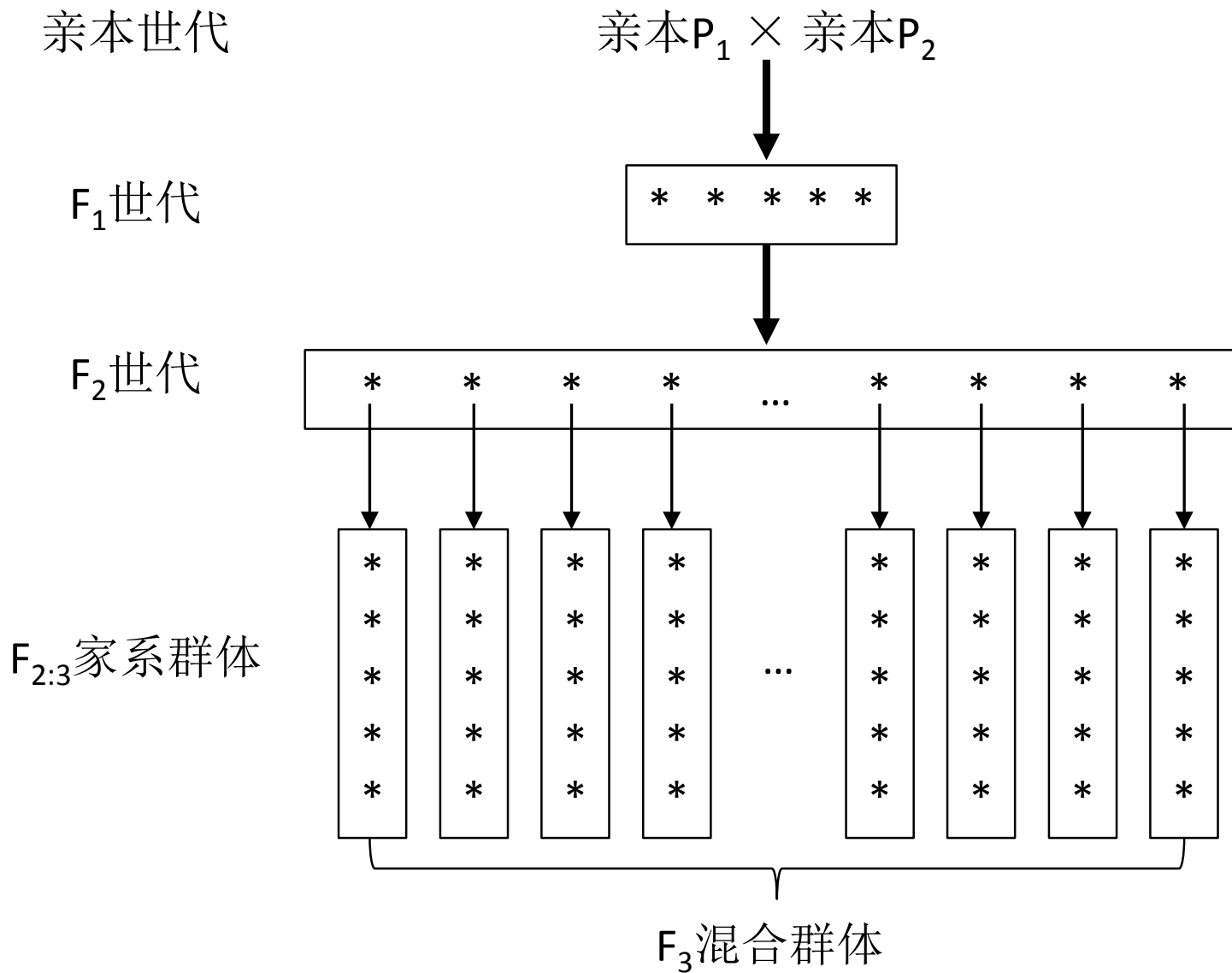
# 纯系亲本的杂交后代

- 以两个纯系为亲本，通过杂交、回交、自交及连续自交衍生出的各种世代，是植物遗传和育种研究中极为常见的群体。在纯系亲本的假定下，亲本和 $F_1$ 都是单一基因型构成的群体，种植规模一般为几个到几十个单株。如果两个亲本的共祖先系数为0，则 $F_1$ 的近交系数也为0。
- $F_2$ 世代有最大的基因型分离，根据物种的特点，种植规模一般为几十到几百个；植株较小的物种，如小麦和水稻等作物，一个 $F_2$ 群体甚至可达几千或上万个单株。当两个亲本的基因型在 $k$ 个座位上存在差异时， $F_2$ 群体中可能基因型的个数为 $3^k$ 。当 $k=5$ 时， $3^k=243$ ； $k=8$ 时， $3^k=6561$ ； $k=10$ 时， $3^k=59049$ 。因此，即使两个亲本在数个座位上存在等位基因差异， $F_2$ 群体中每个个体的基因型也可以被看作是唯一的。

# 双亲衍生的F<sub>3</sub>世代

- 收获F<sub>2</sub>植株上的自交种子，种植得到的群体就是F<sub>3</sub>世代。在遗传育种中，F<sub>3</sub>世代有两种种植方式。一是将F<sub>2</sub>植株上收获的种子混合起来进行种植，形成一个F<sub>3</sub>混合群体；二是将每个F<sub>2</sub>植株上收获的自交种子种植成一个家系，形成一个F<sub>2:3</sub>家系群体。
- 由于带有家系结构，每个F<sub>2:3</sub>家系都能够追踪到唯一一个F<sub>2</sub>单株，即每个F<sub>2:3</sub>家系是F<sub>2</sub>群体中同一个个体的自交后代。每个家系的平均表现及家系内的变异，是由F<sub>2</sub>亲本基因型决定的。因此，F<sub>2:3</sub>家系的表现一定程度上反映了F<sub>2</sub>个体的表现。
- 家系内的个体具有共同祖先，单个家系的种植规模不用太大，一般在几个到几十个单株即可。如果把所有F<sub>2:3</sub>家系混合起来，这其实就是F<sub>3</sub>混合群体。也就是说，去掉家系结构后的F<sub>2:3</sub>群体，等同于F<sub>3</sub>混合群体。

# 双亲衍生的F<sub>2:3</sub>家系群体示意图



# F<sub>3</sub>世代的群体均值和遗传方差

- 在单基因加显性模型下，如不考虑家系结构，基因型AA、Aa和aa在F<sub>3</sub>世代中的频率分别为3/8、1/4和3/8。群体均值和遗传方差为

$$\bar{F}_3 = m + \frac{3}{8}a + \frac{1}{4}d + \frac{3}{8}(-a) = m + \frac{1}{4}d$$

$$V_{F_3} = \frac{3}{8}a^2 + \frac{1}{4}d^2 + \frac{3}{8}(-a)^2 - \left(\frac{1}{4}d\right)^2 = \frac{3}{4}a^2 + \frac{3}{16}d^2$$



# F<sub>3</sub>世代的群体均值和遗传方差

F <sub>2</sub> 世代 基因型	频率 $f$	基因型值 $X$	F <sub>2:3</sub> 家系世代			F <sub>2:3</sub> 家系均值 $Y$	家系内方差
			AA	Aa	aa		
AA	$\frac{1}{4}$	$a$	1	0	0	$a$	0
Aa	$\frac{1}{2}$	$d$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}d$	$\frac{1}{2}a^2 + \frac{1}{4}d^2$
aa	$\frac{1}{4}$	$-a$	0	0	1	$-a$	0
均值		$\bar{F}_2 = \frac{1}{2}d$	$\bar{F}_3 = \frac{1}{4}d$			$\bar{F}_3 = \frac{1}{4}d$	$V_{wF_{2:3}} = \frac{1}{4}a^2 + \frac{1}{8}d^2$
方差		$V_{F_2} = \frac{1}{2}a^2 + \frac{1}{4}d^2$	$V_{F_3} = \frac{3}{4}a^2 + \frac{3}{16}d^2$			$V_{bF_{2:3}} = \frac{1}{2}a^2 + \frac{1}{16}d^2$	

F<sub>2</sub> 与 F<sub>3</sub> 家系间的协方差

$$Cov_{F_{2:3}} = \frac{1}{2}a^2 + \frac{1}{8}d^2$$

表中基因型值、以及均值和方差的计算时，都没有考虑中亲值  $m$ 。因此，表中的基因型值和所有均值需要加上中亲值  $m$ 。是否考虑中亲值  $m$ ，对表中所有方差的计算没有影响。

# F<sub>2:3</sub>世代的家系间和家系内遗传方差

- 利用F<sub>2:3</sub>家系群体，可以计算出两个方差。一个利用三种家系的均值计算，这个方差反映了家系均值（或家系平均数）之间的变异，称为家系间方差（between family variance），计算公式是

$$V_{bF_{2:3}} = \frac{1}{4}a^2 + \frac{1}{2}\left(\frac{1}{2}d\right)^2 + \frac{1}{4}(-a)^2 - \left(\frac{1}{4}d\right)^2 = \frac{1}{2}a^2 + \frac{1}{16}d^2$$

- 另一个是利用三种家系的方差计算，这个方差反映了家系内的平均变异，称为家系内方差（within family variance），计算公式是

$$V_{wF_{2:3}} = \frac{1}{4} \times 0 + \frac{1}{2}V_{F_2} + \frac{1}{4} \times 0 = \frac{1}{4}a^2 + \frac{1}{8}d^2$$

# F<sub>2</sub>个体与F<sub>2:3</sub>家系间的协方差

- 此外，还可以把F<sub>2</sub>的基因型值作为X、家系均值作为Y计算出一个协方差。这个协方差反映了F<sub>2</sub>个体与其F<sub>3</sub>后代之间的相关关系，计算公式是

$$\begin{aligned} Cov_{F_{2:3}} &= \frac{1}{4} \times a \times a + \frac{1}{2} \times d \times \left(\frac{1}{2}d\right) + \frac{1}{4} \times (-a) \times (-a) - \left(\frac{1}{2}d\right) \times \left(\frac{1}{4}d\right) \\ &= \frac{1}{2}a^2 + \frac{1}{8}d^2 \end{aligned}$$

# 多基因模型下的遗传方差分解

- 推广到多基因的情形， $F_2$ 世代的加显性方差用 $V_A$ 和 $V_D$ 表示，

$$V_{F_3} = \frac{3}{2}V_A + \frac{3}{4}V_D \quad V_{bF_{2:3}} = V_A + \frac{1}{4}V_D \quad V_{wF_{2:3}} = \frac{1}{2}V_A + \frac{1}{2}V_D$$

$$Cov_{F_{2:3}} = V_A + \frac{1}{2}V_D$$

- 这些遗传方差计算公式，可以从理论上指导育种过程中的选择工作。例如，与 $F_2$ 世代相比， $F_3$ 世代具有1.5倍的 $V_A$ ，表明 $F_3$ 世代的选择要比 $F_2$ 更加有效。 $F_3$ 世代的1.5倍 $V_A$ 中，家系间占了1倍、家系内仅占0.5倍，表明在 $F_3$ 家系之间的选择要比家系内的选择更为有效。
- $F_2$ 与 $F_3$ 家系之间存在协方差，协方差中包含了1倍的 $V_A$ 。这一方面说明，通过 $F_2$ 个体的选择来提高 $F_3$ 家系的平均表现是可行的；另一方面还说明，如难以对 $F_2$ 单株进行选择，通过 $F_3$ 家系来间接选择 $F_2$ 单株也是可行的。

# F<sub>3</sub>世代中的环境方差

- 利用表型观测数据计算的F<sub>3</sub>家系间和家系内这两个方差，都包含非遗传的成分。用 $V_{Ew}$ 表示家系内共同的环境方差（within-family environmental variance），用 $V_{Eb}$ 表示家系间的环境方差（between-family environmental variance）。如用 $y_{ij}$ 表示第 $i$ 个家系中的第 $j$ 个观测值， $G_{ij}$ 表示个体的基因型值， $Eb_i$ 表示第 $i$ 个家系的环境效应， $Ew_{ij}$ 表示家系内个体的随机环境效应，则观测值可用下面的模型表示

$$y_{ij} = G_{ij} + Eb_i + Ew_{ij}$$

$$Eb_i \sim N(0, V_{Eb}) \quad Ew_{ij} \sim N(0, V_{Ew}) \quad \text{且相互独立}$$

# F<sub>2:3</sub>家系内的表型方差和遗传方差

- 在计算第*i*个家系的家系内表型方差 $V_i$ 时，家系内的环境效应 $Eb_i$ 为一固定值。因此，家系内的方差等于家系内的遗传方差加上家系内的环境方差，即

$$V_i = V(G_{ij}) + V(Eb_i) + V(Ew_{ij}) = V(G_{ij}) + V(Ew_{ij})$$

- 从中可以看出，第*i*个家系内的方差包含两部分，即家系内不同基因型引起的遗传方差和家系内个体的随机环境效应引起的误差方差。家系内的环境方差只包含一份 $V_{Ew}$ 。因此，F<sub>3</sub>家系内的平均方差也只包含一份 $V_{Ew}$ 。下面的公式给出利用表型值计算的家系内平均方差。

$$W_{wF_{2:3}} = V_{wF_{2:3}} + V_{Ew} = \frac{1}{2}V_A + \frac{1}{2}V_D + V_{Ew}$$

# F<sub>2:3</sub>家系间的表型方差和遗传方差

- 假定 $n$ 为家系大小，第 $i$ 个家系的平均数为

$$\bar{y}_{i\bullet} = \bar{G}_{i\bullet} + Eb_i + \bar{E}w_{i\bullet}$$

- 家系平均数中包含的环境方差有两部分，一部分是家系间的方差 $V_{Ec}$ ，另一部分是家系内的抽样误差，即

$$V(\bar{y}_{i\bullet}) = V(\bar{G}_{i\bullet}) + V(Eb_i) + V(\bar{E}w_{i\bullet})$$

- 利用表型值 $y_{ij}$ 计算的家系平均数方差为

$$W_{bF_{2:3}} = V_{bF_{2:3}} + \frac{1}{n}V_{wF_{2:3}} + V_{Eb} = V_A + \frac{1}{4}V_D + \frac{1}{n}V_{wF_{2:3}} + V_{Eb}$$

# 任意自交世代的群体均值和遗传方差

- 对任意 $F_t$ 世代，杂合基因型 $Aa$ 的频率表示为

$$H_t = \left(\frac{1}{2}\right)^{t-1}$$

- 单基因模型下，世代均值和遗传方差为

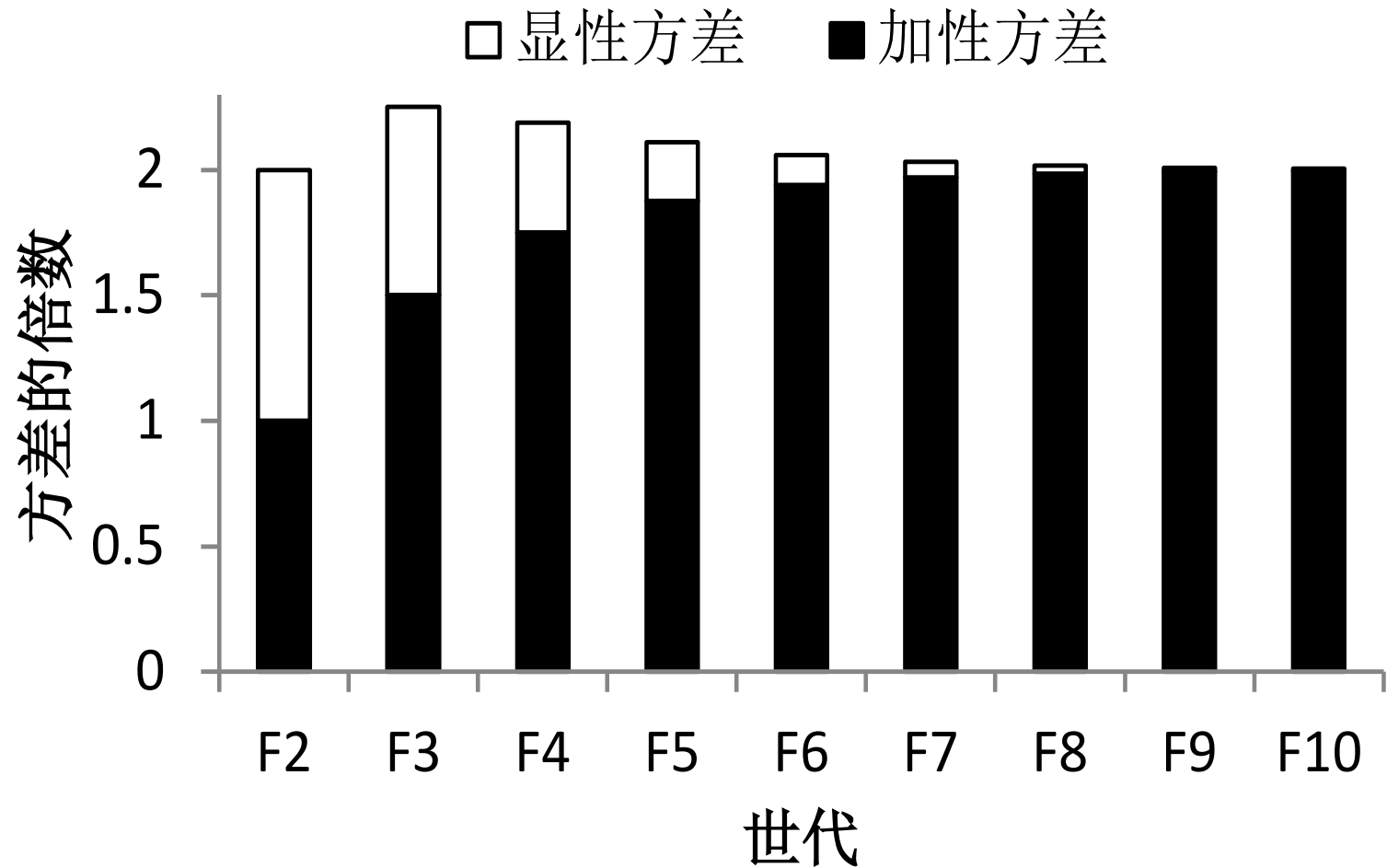
$$\bar{F}_t = m + H_t d \quad V_{F_t} = (1 - H_t)a^2 + H_t(1 - H_t)d^2$$

- 推广到多基因模型时，

$$V_{F_t} = 2(1 - H_t)V_A + 4H_t(1 - H_t)V_D$$



# 以F<sub>2</sub>世代的遗传方差为基础，不同世代中加性方差和显性方差的倍数



# 任意世代的家系间和家系内遗传方差

- 从 $F_t$ 世代开始，经过 $m-t$ 代自交衍生的家系世代用 $F_{t:m}$ 表示，其中 $m=t+1$ 、 $t+2$ 、 $t+3\dots$ （如把 $m-t=0$ 看作是 $F_t$ 世代自身，则 $m$ 也可以等于 $t$ ）。

$$V_{F_m} = (1 - H_m)a^2 + H_m(1 - H_m)d^2$$

$$V_{bF_{t:m}} = (1 - H_t)a^2 + H_m(H_t - H_m)d^2$$

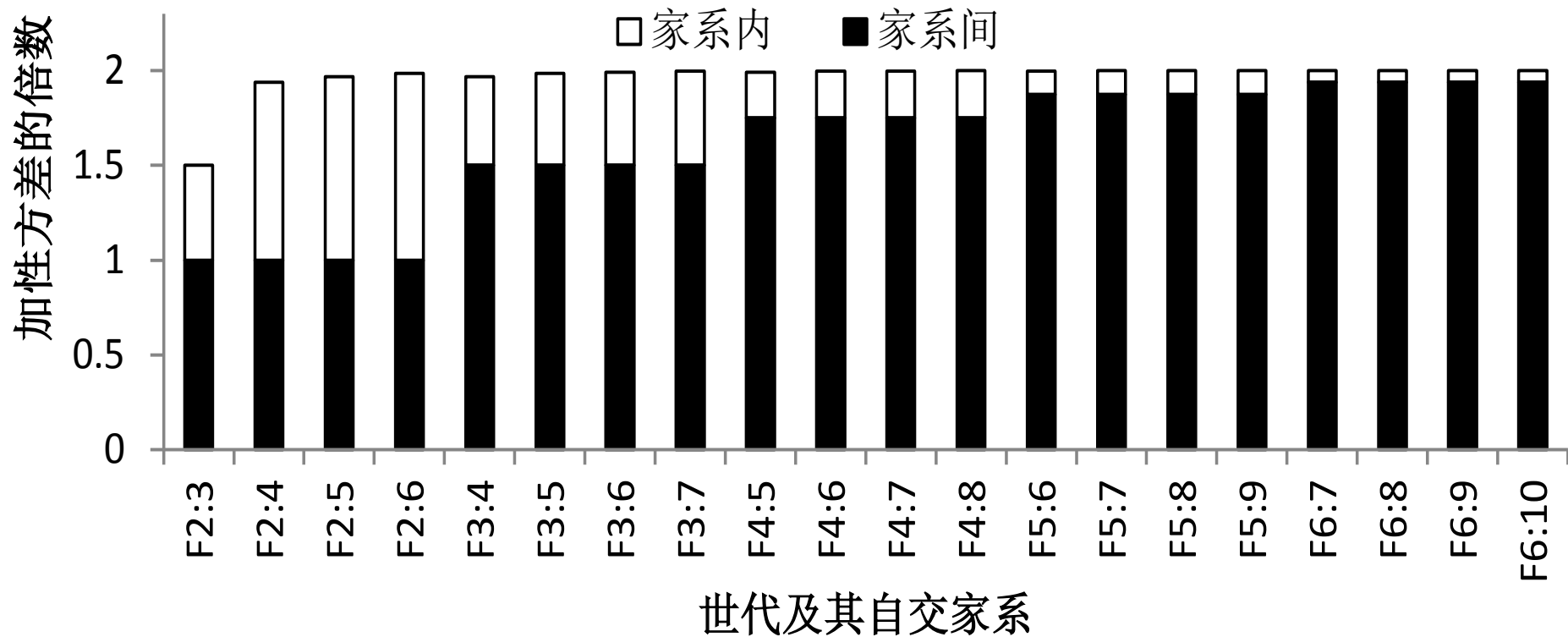
$$V_{wF_{t:m}} = V_{F_m} - V_{bF_{t:m}} = (H_t - H_m)a^2 + H_m(1 - H_t)d^2$$

$$Cov_{F_{t:m}} = (1 - H_t)a^2 + H_m(1 - H_t)d^2$$

# 任意世代的家系间和家系内遗传方差

- 从上面的公式可以看出， $F_{t:m}$ 家系均值之间的方差，以及 $F_t$ 世代与 $F_{t:m}$ 家系之间的协方差中，加性方差的大小只依赖于 $F_t$ 世代中杂合基因型的频率。
- $F_t$ 世代杂合基因型频率越低（即对应于较大的 $t$ ）， $F_{t:m}$ 家系间的加性方差就越大， $F_t$ 与 $F_{t:m}$ 家系之间的加性方差就越大。

# 以F<sub>2</sub>世代遗传方差为基础，不同世代中家系间和家系内加性方差的倍数



# 家系间和家系内加性方差与选择方法

- $F_{2:3}$ 家系之间的加性方差与其后代 $F_{2:4}$ 、 $F_{2:5}$ 等的加性方差相等，说明在这些家系之间进行选择的效果也是类似的。这其实就是纯系育种过程中，早代选择（early generation selection）的理论基础。也就是说，对于难以在个体水平上进行选择的性状，自交一代家系间选择与自交多代家系间是等效的，育种中应该尽早利用自交家系进行选择。
- 加性方差在 $F_2$ 和 $F_3$ 世代间的变化幅度最大，即从1倍的 $V_A$ 增加到1.5倍的 $V_A$ 。因此， $F_3$ 世代的选择效率要高于 $F_2$ 世代，尤其是遗传力较低的复杂性状。

# $F_{t:m}$ 家系与 $F_{t:n}$ 家系之间的协方差

$$\text{Cov}(F_{t:m}, F_{t:n}) = (1 + F_t) \left[ V_A + \frac{(1 - F_m)(1 - F_n)}{1 - F_t} V_D \right]$$

其中， $F_t = 1 - \left(\frac{1}{2}\right)^{t-2}$  是自交世代 $F_t$ 的近交系数

- $F_{t:m}$ 家系与 $F_{t:n}$ 家系的协方差中，加性方差所占的比例只与 $F_t$ 世代的近交系数有关，而与 $F_m$ 世代和 $F_n$ 世代的近交系数没有关系，再次说明了早代选择的重要性。
- 当 $m=n=t$ 时，上面公式给出的协方差就是公式7.80给出的 $F_t$ 世代遗传方差。当 $n=t$ 时，上面公式给出的协方差就是公式7.85给出的 $F_t$ 世代与 $F_{t:m}$ 家系之间的协方差。因此，这一公式可以看作自交过程中任意世代遗传方差、任意两个世代间协方差的一般表达式。

# § 7.4 基因间的上位性互作

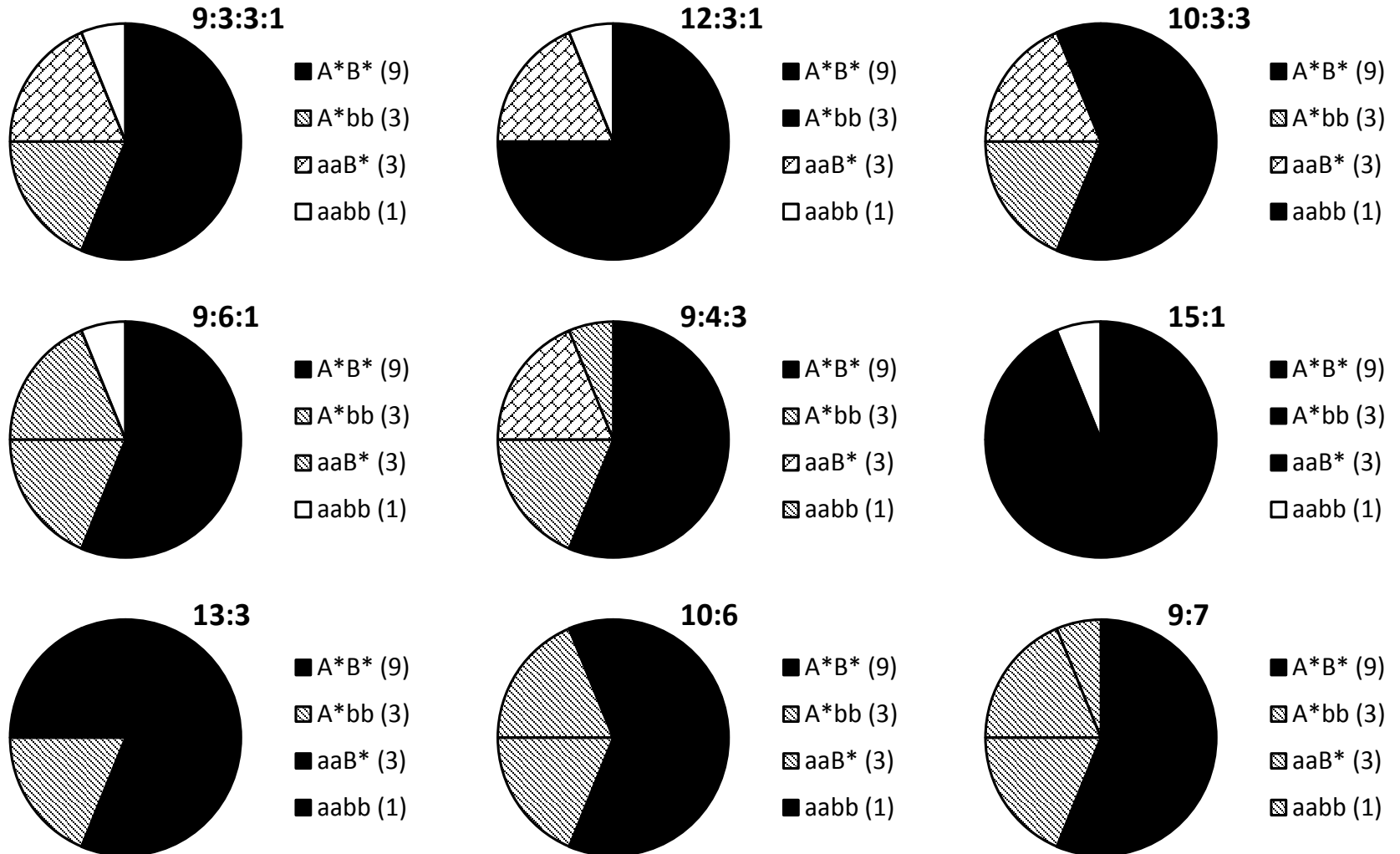
- § 7.4.1 两个座位间的上位性互作
- § 7.4.2 上位性模型的遗传方差分解
- § 7.4.3 上位性互作的重要性

# 两个座位间的上位性互作

- 同一个座位内不同等位基因之间互作引起的显性效应外，有时不同座位上的基因之间也会存在相互作用，这样的互作称上位性（epistasis）或基因间互作（inter-genic interaction）。
- 基因间的上位性互作，首先是从两个基因控制的质量性状在 $F_2$ 群体中的表型分离比观测到的。如一个性状受两对独立遗传显性基因控制，用 $A/a$ 和 $B/b$ 表示这两对等位基因。没有上位性互作存在时， $F_2$ 世代的4种表型将呈现出9:3:3:1的分离比。这4种表型对应的基因型分别用 $A_*B_*$ 、 $A_*bb$ 、 $aaB_*$ 、 $aabb$ 表示，等位基因 $A$ 后面的星号表示这个位置既可以是 $A$ ，也可以是 $a$ ，等位基因 $B$ 后面的星号表示这个位置既可以是 $B$ ，也可以是 $b$ 。当这两对独立显性基因之间存在互作时， $F_2$ 世代的表型就可能只有三种、甚至两种，并呈现出各种有规律的分離比。



# 两对显性独立遗传基因控制的质量性状 在F<sub>2</sub>世代中各种可能的表型分离比



# 上位性互作的4种类型

- 根据基因型是否纯合或是杂合，两个座位（用A和B表示）上的上位性互作可以划分成4种类型，即纯合基因型之间的上位性（additive  $\times$  additive epistasis，用 $aa$ 表示），座位A纯合基因型和座位B杂合基因型之间的上位性（additive  $\times$  dominance epistasis，用 $ad$ 表示），座位A杂合基因型和座位B纯合基因型之间的上位性（dominance  $\times$  additive epistasis，用 $da$ 表示），以及两个座位上杂合基因型之间的上位性（dominance  $\times$  dominance epistasis，用 $dd$ 表示）。
- 考虑上位性效应时的遗传模型，称为加性-显性-上位性效应遗传模型，简称上位性模型。

# 两对独立遗传基因在加显性模型和上位性模型下的9种基因型值

基因型	频率	基因型值符号	基因型值	
			加显性模型	上位性互作模型
<i>AABB</i>	$\frac{1}{16}$	$\mu_{AABB}$	$m + a_A + a_B$	$m + a_A + a_B + aa$
<i>AABb</i>	$\frac{2}{16}$	$\mu_{AABb}$	$m + a_A + d_B$	$m + a_A + d_B + ad$
<i>AAbb</i>	$\frac{1}{16}$	$\mu_{AAbb}$	$m + a_A - a_B$	$m + a_A - a_B - aa$
<i>AaBB</i>	$\frac{2}{16}$	$\mu_{AaBB}$	$m + d_A + a_B$	$m + d_A + a_B + da$
<i>AaBb</i>	$\frac{4}{16}$	$\mu_{AaBb}$	$m + d_A + d_B$	$m + d_A + d_B + dd$
<i>Aabb</i>	$\frac{2}{16}$	$\mu_{Aabb}$	$m + d_A - a_B$	$m + d_A - a_B - da$
<i>aaBB</i>	$\frac{1}{16}$	$\mu_{aaBB}$	$m - a_A + a_B$	$m - a_A + a_B - aa$
<i>aaBb</i>	$\frac{2}{16}$	$\mu_{aaBb}$	$m - a_A + d_B$	$m - a_A + d_B - ad$
<i>aabb</i>	$\frac{1}{16}$	$\mu_{aabb}$	$m - a_A - a_B$	$m - a_A - a_B + aa$

# 利用基因型值计算各种遗传效应

$$\begin{array}{l}
 m \\
 a_A \\
 d_A \\
 a_B \\
 d_B \\
 aa \\
 ad \\
 da \\
 dd
 \end{array}
 =
 \begin{bmatrix}
 0.25 & 0 & 0.25 & 0 & 0 & 0 & 0.25 & 0 & 0.25 \\
 0.25 & 0 & 0.25 & 0 & 0 & 0 & -0.25 & 0 & -0.25 \\
 -0.25 & 0 & -0.25 & 0.5 & 0 & 0.5 & -0.25 & 0 & -0.25 \\
 0.25 & 0 & -0.25 & 0 & 0 & 0 & 0.25 & 0 & -0.25 \\
 -0.25 & 0.5 & -0.25 & 0 & 0 & 0 & -0.25 & 0.5 & -0.25 \\
 0.25 & 0 & -0.25 & 0 & 0 & 0 & -0.25 & 0 & 0.25 \\
 -0.25 & 0.5 & -0.25 & 0 & 0 & 0 & 0.25 & -0.5 & 0.25 \\
 -0.25 & 0 & 0.25 & 0.5 & 0 & -0.5 & -0.25 & 0 & 0.25 \\
 0.25 & -0.5 & 0.25 & -0.5 & 1 & -0.5 & 0.25 & -0.5 & 0.25
 \end{bmatrix}
 \times
 \begin{array}{l}
 \mu_{AABB} \\
 \mu_{AABb} \\
 \mu_{AAbb} \\
 \mu_{AaBB} \\
 \mu_{AaBb} \\
 \mu_{Aabb} \\
 \mu_{aaBB} \\
 \mu_{aaBb} \\
 \mu_{aabb}
 \end{array}$$

# 纯合基因型的基因间互作

- 两个座位上的纯合基因型有4种，即AABB、AAbb、aaBB、aabb，它们的基因型值中只包含加性效应和加加互作效应。纯系群体中，不存在显性及与显性有关的上位性效应。根据表7.12的上位性模型，4种基因型值与遗传效应的关系用下面的4个等式表示。

$$\mu_{AABB} = m + a_A + a_B + aa$$

$$\mu_{AAbb} = m + a_A - a_B - aa$$

$$\mu_{aaBB} = m - a_A + a_B - aa$$

$$\mu_{aabb} = m - a_A - a_B + aa$$

# 纯合基因型间的效应和方差

- 反过来还可以利用下面的4个公式计算各种遗传效应的大小，同时还能得到群体遗传方差。

$$m = \frac{1}{4}(\mu_{AABB} + \mu_{AAbb} + \mu_{aaBB} + \mu_{aabb}) \quad a_A = \frac{1}{4}(\mu_{AABB} + \mu_{AAbb} - \mu_{aaBB} - \mu_{aabb})$$

$$a_B = \frac{1}{4}(\mu_{AABB} - \mu_{AAbb} + \mu_{aaBB} - \mu_{aabb}) \quad aa = \frac{1}{4}(\mu_{AABB} - \mu_{AAbb} - \mu_{aaBB} + \mu_{aabb})$$

$$V_G = a_A^2 + a_B^2 + aa^2 = 2V_A$$

- 从中可以看出，加加上上位性效应也包含在纯系群体的遗传方差中，加加上上位性是可以过选择被固定下来的，因此可以被纯系育种所利用。

# 4个近等基因系在一个品质性状上的观测值

重复	近等基因系的基因型			
	AABB	AAbb	aaBB	aabb
重复1	237	241	210	198
重复2	235	236	207	201
重复3	230	238	202	208
重复4	240	254	236	195
重复5	258	247	225	207
重复6	258	246	235	207
重复7	232	224	219	151
重复8	233	220	211	152
重复9	242	200	209	151
基因型均值	240.56	234.00	217.11	185.56
总平均	219.31			
基因型效应估计	21.25	14.69	-2.19	-33.75

# 方差分析的线性模型

- 4个近等基因系之间的差异显著性，可以通过方差分析方法进行检验。但是，这里除了关心近等基因系之间是否存在显著的遗传差异外，更关心的是A和B这两个座位上的两个等位基因间是否存在显著差异。如果这两个座位都存在显著差异，哪个座位的遗传贡献更大？它们之间是否存在上位性互作？
- 要回答这些问题，可以把表7.13的数据视为一个双因素试验。用*i*=1、2表示基因型AA、aa，*j*=1、2表示基因型BB、bb， $\alpha_i$ 和 $\beta_j$ 分别表示两个因素的主效应， $\tau_{ij}$ 表示它们之间的互作效应，那么观测值 $y_{ijk}$  ( $k=1, 2, \dots, 9$ ) 的方差分析线性模型为

$$y_{ijk} = \mu + \alpha_i + \beta_j + \tau_{ij} + \varepsilon_{ijk}$$



# 两个独立遗传座位4种基因型的方差分析表

变异来源	自由度	平方和	均方	F值	P值	方差估计
基因型	3	16302.31	5434.10	17.79	$5.66 \times 10^{-07}$	569.84
座位A	1	11628.03	11628.03	38.06	$6.66 \times 10^{-07}$	629.03
座位B	1	3268.03	3268.03	10.70	0.0026	164.58
互作	1	1406.25	1406.25	4.60	0.0396	122.30
随机误差	32	9777.33	305.54			305.54
总和	35	26079.64				

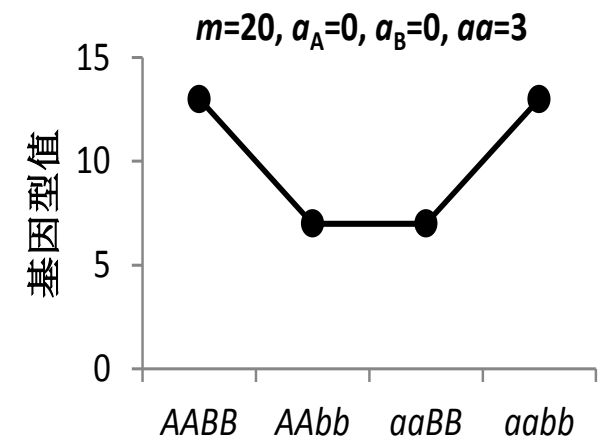
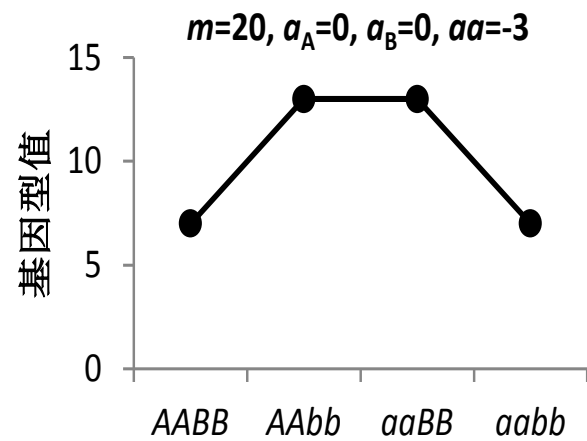
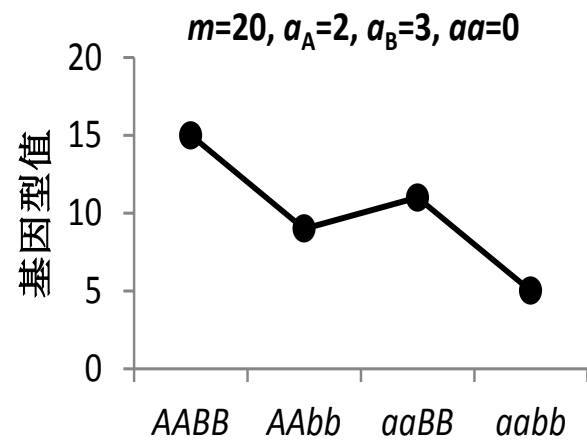
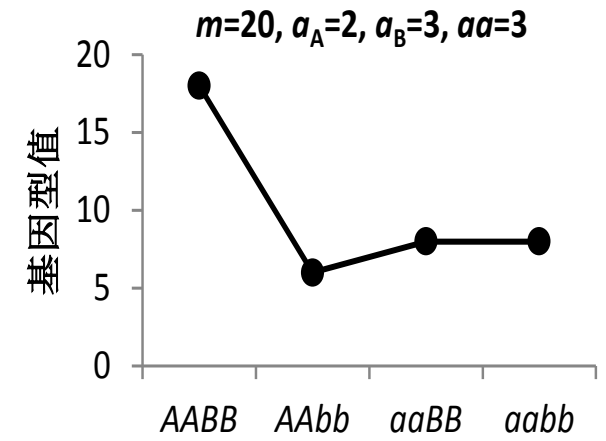
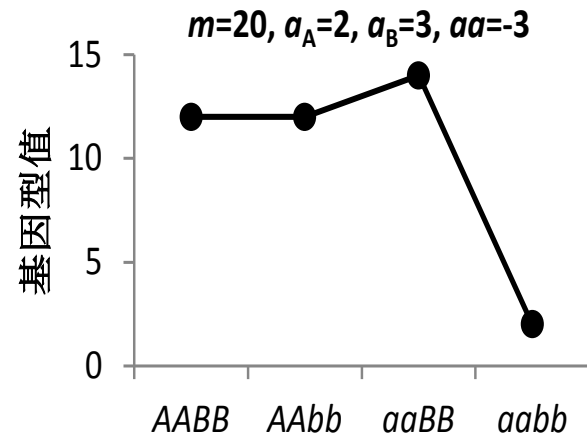
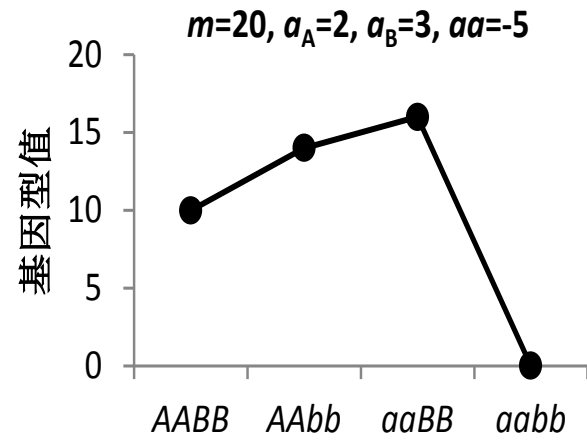
# 上位性互作的复杂性

- 一个基因要产生一定的功能和表型效应，离不开其它众多基因的作用。从分子水平来看，任何一个生化过程都不可能由一个座位上的基因单独完成。因此，基因间的互作在遗传学上应该是普遍存在的。
- 但是，从前面的内容我们看到，上位性互作的分析还是相当复杂的。如果控制性状的基因多于2个，还可能存在更高阶的基因间互作，这时互作的分析和检测会变得更加困难。
- 就目前来看，对于绝大多数数量性状来说，关于基因之间是如何相互作用这方面的了解还十分有限。上位性互作的研究，离不开新的遗传学方法和分子生物学技术，也离不开适宜的遗传材料，如前面的近等基因系。

# 上位性互作对育种的重要性

- 两个座位上的基因之间不存在互作时，联合基因型效应等于单座位基因型效应之和。因此，座位A上效应高的基因型与座位B上效应高的基因型结合在一起，联合基因型就会产生较高的效应。这时，最优基因型可以从单个座位上基因型是否最优进行判断。
- 但是，当存在上位性互作时，两个座位上的最优基因型结合在一起不一定是最优的。

# 不同遗传效应模型下，两对独立遗传座位上4种纯合基因型值的示意图



# 互作对基因型值的作用

- 图中，两个座位的加性效应在保持恒定，分别为2和3。因此，从单个座位来看， $AA$ 的表现高于 $aa$ 、 $BB$ 的表现高于 $bb$ 。
- 当 $aa=0$ 时，4种基因型值分别为 $AABB=15$ 、 $AAbb=9$ 、 $aaBB=11$ 、 $aabb=5$ 。不论座位B上是 $BB$ 还是 $bb$ ，基因型 $AA$ 与 $aa$ 之间的差异都是4；不论座位A上是 $AA$ 还是 $aa$ ，基因型 $BB$ 与 $bb$ 之间的差异都是6。
- 因此，基因型 $AA$ 与 $aa$ 之间的差异与座位B无关；基因型 $BB$ 与 $bb$ 之间的差异与座位A也无关； $AA$ 与 $BB$ 结合在一起基因型 $AABB$ 在4种基因型中有最高的表现。

# 互作对基因型值的作用

- 图上，互作效应 $aa$ 的三个取值分别为-5、-3、3。当 $aa \neq 0$ 时，基因型 $AA$ 与 $aa$ 之间的差异与座位B的基因型有关，同时基因型 $BB$ 与 $bb$ 之间的差异与座位A的基因型也有关。
- 当互作达到一定程度时，单个座位上表现较好的基因型 $AA$ 与 $BB$ 结合在一起的基因型 $AABB$ ，不一定就在4种基因型中有最高的表现。例如，图7.7上左、上中， $aa = -5$ 、 $-3$ ，最高表现的基因型不是 $AABB$ ，而是 $aaBB$ 。这时，只根据单个座位上的基因型高低进行选择，就无法选到表现最高的基因型。

# 互作对基因型值的作用

- 对于只存在互作而不存在加显性效应的座位来说（如图7.6的10:6分离比），从单个座位来看，似乎基因型之间并没有差异。但是，不同座位的基因型结合在一起时，能产生出不同的效应。
- 例如，图7.7下中、下右，座位A和B的加性效应都是0，但是它们的加加互作效应不为0。对于负向加加互作，基因型*AAbb*和*aaBB*有更高的表现；对于正向加加互作，基因型*AABB*和*aabb*有更高的表现。如果单独分析每个座位，则会发现两个基因型之间没有任何表型差异，因此就可能得出这两个座位都与表型无关的错误结论。这样的上位性互作，也只能通过两个座位的联合分析才能发现出来。