第 24 卷 第 4 期
1998 年 7 月

作 物 学 报
ACTA AGRONOMICA SINICA

Vol. 24, No. 4
July, 1998

利用回交或 $F_{2:3}$ 家系世代鉴定数量性状主基因-多基因混合遗传模型*

盖钧镒** 王建康***

（南京农业大学大豆研究所，江苏南京，210095）

提 要 主-多基因性状在分离世代中表现出多峰或偏态现象，具有混合分布的特征。本文利用混合分布理论提出从回交或 $F_{2,3}$ 家系个别分离世代鉴定主基因的存在并估计其遗传效应的统计分析方法。利用所提出的方法可以鉴别主基因的存在并估计其遗传效应和方差等遗传参数。

关键词 数量性状；主基因和多基因混合遗传；混合分布；EM 算法

大量试验数据和 QTL(quantitative trait loci)作图结果表明：控制数量性状的基因体系中既有遗传效应较大的主基因，又有遗传效应较小的微效多基因，称之为主基因-多基因混合遗传[3, 5, 6, 11, 13, 16]。Elkind 和 Cahaner[8, 9]提出单基因-多基因(single gene-polygene，SG-PG)遗传模型，其单基因是可鉴定的。莫惠栋[3]分析了一对主基因存在时，主基因-多基因混合遗传性状在各个世代的遗传组成以及遗传参数的估计问题，并把这类性状称为质量-数量性状，由于 F_2 代的分组趋势不明显，作者建议采用后裔测验的方法，然后通过聚类分析确定 F_2 个体的主基因基因型。Loisel[13]研究了在 F_2 世代中探测主基因存在的似然比统计量的渐近性质，由此可对主基因是否存在、主基因是否表现为加性或完全显性进行检验。姜长鉴[1]将 Loisel[13]的结果用于大麦矮秆突变体与正常秆品系间杂交产生的 F_2 代株高性状的遗传分析。王建康和盖钧镒[5, 6]提出利用混合分布理论在 F_2 群体中鉴定数量性状主基因的存在和估计遗传参数的方法。这些结果虽然可以检验主基因的加显性并估计主基因的加显性效应和主基因遗传方差，但 F_2 世代受环境影响较大，检测主基因的功效不高，对多基因的存在无法作进一步鉴定。本文提出利用混合分布理论在回交和 $F_{2,3}$ 家系世代中鉴定主基因的存在并估计其遗传效应的统计分析方法。

1 主基因-多基因混合遗传模型

设来自杂种分离群体个体的表现型可分解为：

$$x = m + g + c + e$$

其中 m 为群体平均值，g 为主基因效应值，c 为多基因效应值，e 为环境效应值。此处假定主

* 国家 863 项目部分内容和河南省科委攻关项目
致谢：南京农业大学农学系朱立宏教授惠允使用水稻南京 6 号与广丛回交世代株高分布的资料，谨致谢忱。
** 联系作者，*E-mail*：*sri@njau. edu. cn*
*** 现在河南省农业科学院科学实验中心工作，河南郑州，450002
收稿日期：1997-03-12，收到修改稿日期：1997-10-29

基因与多基因之间不存在互作，基因型效应与环境效应是相互独立的。主基因效应 g 为固定值，主基因方差用 σ^2_{mg}表示；多基因效应 $c\sim N(0, \sigma^2_{pg})$是随机变量，$\sigma^2_{pg}$为多基因效应值的方差；环境效应 $e\sim N(0, \sigma^2_e)$是随机变量，σ^2_e 为环境方差，σ^2_p 表示总表型方差，这样遗传率可分解为主基因遗传率 $h^2_{mg}=\sigma^2_{mg}/\sigma^2_p$ 及多基因遗传率 $h^2_{pg}=\sigma^2_{pg}/\sigma^2_p$。如果主基因有 k 个不同的基因型值，那么分离群体表现为 k 个正态分布的混合，通过对混合群体的分解可将主基因的效应分解出来。除主基因之外的变异是多基因变异和环境变异的混合，在只有个别分离群体的情况下，不能将二者分解开来，记 $\sigma^2=\sigma^2_{pg}+\sigma^2_e$。

2　混合分布分析方法

混合分布理论在实际中有广泛的应用[4, 15]，混合分布是如下定义的：假定 X 为一随机变量，其概率密度函数可以表示为：

$$p(x) = \pi_1 f_1(x) + \pi_2 f_2(x) + \cdots + \pi_k f_k(x)$$

其中 $\pi_j>0$，$j=1, \cdots, k$，

$$\pi_1 + \pi_2 + \cdots + \pi_k = 1, \quad f_j(\cdot) \geqslant 0, \quad \int f_j(x)dx = 1, \ j = 1, \cdots, k,$$

则称 X 是一有限混合分布，$p(\cdot)$是有限混合分布的密度函数，参数 $\pi_1, \pi_2, \cdots, \pi_k$ 称为混合权数，$f_1(\cdot), f_2(\cdot), \cdots, f_k(\cdot)$为混合分布中各个成份(以下称为成分分布)的密度函数，k 为混合分布中所包含的成分分布的个数。

设 θ_j 为分布 $f_j(\cdot)$的参数，那么混合分布的密度函数也可用参数形式表示成：

$$\begin{aligned} p(x; \varphi) &= \pi_1 f_1(x; \theta_1) + \pi_2 f_2(x; \theta_2) + \cdots + \pi_k f_k(x; \theta_k) \\ &= \sum_{j=1}^{k} \pi_j f_j(x; \theta_j) \end{aligned}$$

其中 $\varphi=(\pi_1, \pi_2, \cdots, \pi_k, \theta_1, \theta_2, \cdots, \theta_k)$为混合分布 X 的参数向量。混合分布的研究内容主要是探讨如何由随机变量 X 经分离分析去获得其中所包含的各个成分分布的特征，联系本研究目的，是从分离世代的分布中区别出各主基因型成分分布并估计其数字特征。

2.1　混合分布的似然函数

假定 $x=(x_1, x_2, \cdots, x_n)$是来自总体 X 的一组样本，那么样本的似然函数为：

$$L(\varphi) = \prod_{i=1}^{n} p(x_i; \varphi) = \prod_{i=1}^{n} \sum_{j=1}^{k} \pi_j f_j(x_i; \theta_j)$$

极大似然估计一般通过求解对数似然函数的极大值点获得，在以后的分析中 $L(\varphi)$表示对数似然函数，即：

$$L(\varphi) = \sum_{i=1}^{n} \log\ p(x_i; \varphi)$$

2.2　混合分布中成分分布个数的估计

混合分布中所包含成分分布个数 k 是一个最基本的参数，确定成分分布数目的方法可分为图形方法和统计检验方法[15]两大类，本文利用 Akaike[7] 提出的最大熵(信息)准则(Akaike's information criterion，AIC)，从不同假设中选择一个最优假设，从而确定参数 k。AIC 值定义为：

$$AIC = AIC(k) = -2\ L(\hat{\varphi}) + 2\ N(k)$$

其中 $\hat{\varphi}$是假设"Ho：成分分布个数为 k"时混合分布中参数的极大似然估计，$N(k)$是混合模型

中独立参数的个数。通过比较不同 k 值下的 AIC 值，选择使 AIC 值达到最小的 k 值作为成分分布数的估计。

2.3 混合模型中其它参数的估计

混合模型的参数估计可利用 EM(expectation and maximization)算法来完成，分两个步骤进行[4, 6, 10]。

E 步骤：计算完全数据似然函数 $L_C(\varphi)$在初始值 $\varphi^{(0)}$下的期望值 $Q(\varphi, \varphi^{(0)})$，

$$Q(\varphi, \varphi^{(0)}) = E\{L_C(\varphi)|X; \varphi^{(0)}\}$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{k}\tau_{ij}^{(0)} \cdot [\log \pi_i + f_j(x_i; \theta)]$$

M 步骤：极大化 $Q(\varphi, \varphi^{(0)})$，并用极大值点处的 φ 值代替 $\varphi^{(0)}$作为下一轮循环的初始值。$Q(\varphi, \varphi^{(0)})$的极大值点由下式确定：

$$\pi_j = \sum_{i=1}^{n}\tau_{ij}^{(0)}/n \qquad j = 1, \cdots, k$$
$$\sum_{i=1}^{n}\sum_{j=1}^{k}\tau_{ij}^{(0)}\partial \log f_j(x_i; \theta)/\partial_\theta = 0$$

EM 算法有以下优良性质：(1) M 步骤在正态混合分布的情况下期望函数的极大值点可以用数学式子明确表示出来，而一般情况下企图通过对似然函数求导数来获得极大似然估计的明显数学表示是很困难的；(2) EM 迭代过程中，似然函数是单调增加的，即：$L(\varphi^{k+1}) \geqslant L(\varphi^k)$，$k \geqslant 0$ 表示第 k 次迭代。这意味着不论对于怎样的初始值，EM 算法最终总能获得一个极大值点。EM 算法的不足之处是收敛速度较慢，并且收敛速度对初始值的选择有一定的依赖性。

对于主基因-多基因混合遗传性状来说，假定成分分布为正态分布，用 μ_j 和 σ_j^2 表示分布的均值和方差，EM 算法具体过程是：

假定 $\varphi^{(0)} = (\pi_1^{(0)}, \cdots, \pi_k^{(0)}, \mu_1^{(0)}, \cdots, \mu_k^{(0)}, \sigma_1^{2(0)}, \cdots, \sigma_k^{2(0)})$

是初始值，$f(x; \mu, \sigma^2)$表示正态分布的密度函数，则：

$$\tau_{ij}^{(0)} = Pro(x_i \in G_j|x_i; \varphi^{(0)})$$
$$= \pi_j^{(0)} f(x_i; \mu_j^{(0)}, \sigma_j^{2(0)}) / \sum_{t=1}^{k}\pi_t^{(0)} f(x_i; \mu_t^{(0)}, \sigma_t^{2(0)})$$

$$\pi_j^{(1)} = \sum_{i=1}^{n}\tau_{ij}^{(0)}/n$$

$$\mu_j^{(1)} = \sum_{i=1}^{n}\tau_{ij}^{(0)} x_i/(n\pi_j^{(1)})$$

$$\sigma_j^{2(1)} = \sum_{i=1}^{n}\tau_{ij}^{(0)}(x_i - \mu_j^{(1)})^2/(n\pi_j^{(1)})$$

将求得的参数值 $\varphi^{(1)}$代替 $\varphi^{(0)}$开始下一轮 EM 迭代。在回交世代中，成分分布还具有相同的方差，用 σ^2 表示，这时，

$$\sigma^{2(1)} = \sum_{i=1}^{n}\sum_{j=1}^{k}\tau_{ij}^{(0)}(x_i - \mu_j^{(1)})^2/n$$

其它参数的表示式同上。

3　利用回交群体鉴定主基因-多基因混合遗传模型

3.1　主基因存在的鉴定及遗传效应的估计

若回交世代均为单一分布，则一般不存在主基因，此处不再讨论。以下分析一个主基因位点(A-a)的情形。假定亲本基因型为 AA(P_1)和 aa(P_2)$B_1=F_1\times P_1$，$B_2=F_1\times P_2$。

3.1.1　*主基因表现为完全显性*　如果通过对回交群体的分解发现回交群体 B_1 是单一正态分布，B_2 是两个正态分布的混合并且两个分布所占比例符合 1∶1 的分离比，此时认为存在有一个主基因(用 A-a 表示)位点，并表现为完全显性，B_2 群体中两个成分分布的主基因基因型分别为 aa 和 Aa，分布的均值用 μ_1 和 μ_2 表示，μ_1 和 μ_2 与主基因的加性效应 d 和群体平均数 m 的关系为：

$$\mu_1 = m - d$$
$$\mu_2 = m + d,\ \mu_1 < \mu_2$$

根据极大似然估计的线性不变性，便可获得 m 和 d 的极大似然估计：

$$\hat{m} = 0.5\ \hat{\mu}_1 + 0.5\ \hat{\mu}_2$$
$$\hat{d} = 0.5\ \hat{\mu}_2 - 0.5\ \hat{\mu}_1$$

3.1.2　*主基因表现为负向完全显性*　如果通过对回交群体的分解发现回交群体 B_2 是单一正态分布，B_1 是两个正态分布的混合并且两个分布所占比例符合 1∶1 的分离比，此时认为只存在有一个主基因(用 A-a 表示)，并表现为负向完全显性，B_1 群体中两个分布的主基因基因型分别为 Aa 和 AA，分布的均值用 μ_1 和 μ_2 表示，μ_1 和 μ_2 与主基因的加性效应 d 和群体平均数的关系为：

$$\mu_1 = m - d$$
$$\mu_2 = m + d$$

根据极大似然估计的线性不变性，便可获得 m 和 d 的极大似然估计：

$$\hat{m} = 0.5\ \hat{\mu}_1 + 0.5\ \hat{\mu}_2$$
$$\hat{d} = 0.5\ \hat{\mu}_2 - 0.5\ \hat{\mu}_1$$

3.1.3　*主基因表现为部分显性或无显性*　如果通过对回交群体的分解发现回交群体 B_1 和 B_2 均为两个正态分布的混合并且两个分布所占比例符合 1∶1 的分离比，此时认为只存在有一个主基因(用 A-a 表示)位点，并表现为部分显性或无显性，分别用 μ_{11}、μ_{12}、μ_{21}和 μ_{22}表示两个回交世代中分布的均值，B_1 群体中两个分布的主基因基因型分别为 Aa 和 AA，B_2 群体中两个分布的主基因基因型分别为 aa 和 Aa，分布的均值与主基因的加性效应 d、显性效应 h、群体平均数 m_1 和 m_2 的关系为：

$$\mu_{11} = m_1 + h$$
$$\mu_{12} = m_1 + d$$
$$\mu_{21} = m_2 - d$$
$$\mu_{22} = m_2 + h$$

根据极大似然估计的线性不变性，便可获得 m_1、m_2、d 和 h 的极大似然估计：

$$\hat{d} = 0.5(\hat{\mu}_{12} + \hat{\mu}_{22} - \hat{\mu}_{11} - \hat{\mu}_{21})$$
$$\hat{h} = 0.5(\hat{\mu}_{11} + \hat{\mu}_{22} - \hat{\mu}_{12} - \hat{\mu}_{21})$$

$$\hat{m}_1 = 0.5(\hat{\mu}_{11} + \hat{\mu}_{12} + \hat{\mu}_{21} - \hat{\mu}_{22})$$
$$\hat{m}_2 = 0.5(\hat{\mu}_{12} + \hat{\mu}_{21} + \hat{\mu}_{22} - \hat{\mu}_{11})$$

3.2 实例分析

对于一组回交试验数据，可按"2"节中混合分布分析方法对各种可能情况进行极大似然估计，根据 *AIC* 值判断成分分布个数，从而推论是否属于主-多基因混合遗传模型并最后估计其有关参数。表 1 所列为水稻南京 6 号(P_1)与广丛(P_2)的回交群体 B_1 和 B_2 株高分布数据。经按上述方法进行统计分析，表 2 给出不同成分分布个数下对数似然函数的极大值和 *AIC* 值，从中可以得到两个回交群体均为两个正态分布的混合。参数极大似然估计列于表3，由

表 1 水稻南京 6 号(P_1)与广丛(P_2)的回交世代 B_1 和 B_2 株高分布

Table 1 Distributions of plant height of backcrosses between two rice cultivars

	80—	85—	90—	95—	100—	105—	110—	115—	120—	125—	130—	135—	140—	145—	150—	155—	160—	165—	170—	175
P_1															4	5	12	22	2	
P_2				1	11	13	18	5												
F_1													3	17	8	4				
B_1													20	14	24	13	12	3		
B_2	1	10	18	27	18	11	2					1	3	14	19	22	11	4	1	

表 2 南京 6 号与广丛所产生的回交世代群体在不同成分分布个数时的 *AIC* 值

Table 2 The *AIC* values under various component numbers of backcrosses between two rice cultivars

成分分布数(k) Number of components(k)		1	2	3	4
独立参数的个数($N(k)$) Number of independent parameters		2	4	6	8
对数似然函数的极大值 $L(\hat{\varphi})$ Maximum of log likelihood	B_1	−292.84	−290.11	−289.15	−286.21
	B_2	−776.54	−648.49	−648.48	−648.45
AIC 值 *AIC* value	B_1	589.69	588.23	590.30	588.42
	B_2	1557.08	1304.97	1308.96	1312.90

表 3 回交群体 B_1 和 B_2 中各参数的极大似然估计

Table 3 The maximum likelihood estimates of parameters in the backcross populations B_1 and B_2

群体 Population	成分分布 Component	权重(π) Weight(π)	均值(m) Mean(m)	方差(σ^2) Variance(σ^2)
B_1	1	0.66	146.98	23.02
	2	0.34	158.54	23.02
B_2	1	0.54	97.13	44.14
	2	0.46	154.17	44.14

表 3 的结果可以对成分分布的权重比是否为 1∶1 进行检验。结果表明 B_1 群体中所包含分布的比例不符合 1∶1 的分离比($\chi^2=8.81^{**}$)，B_2 群体中所包含分布的比例符合 1∶1 的分离比($\chi^2=1.04$，$P>0.25$)。B_1 群体的奇异分离可能由于成分分布比较靠近所致。在对混合分布作分解时，如果其中成分分布的位置参数相差太小，参数估计的偏差是很大的。若仍认为 B_1 群体中的成分分布也具有 1∶1 的分离比，那么可以得到：在影响株高的基因中存在一个遗传效应较大的主基因，且表现为部分显性。进一步可以估计出主基因的加性效应 $d=34.30$，显性效应 $h=22.74$，主基因的显性度 $d/h=0.66$。回交世代 B_1 中主基因的遗传方差为：$\sigma^2_{mg}=(d-h)^2/4=33.41$，总变异 $\sigma^2_p=53.11$，因此主基因遗传率 $h^2_{mg}=62.91\%$。回交世代

B_2 中主基因的遗传方差为：$\sigma^2_{mg}=(d+h)^2/4=813.39$，总变异 $\sigma^2_p=853.26$，因此主基因遗传率 $h^2_{mg}=95.33\%$。

4　利用 $F_{2,3}$ 家系世代鉴定主基因-多基因混合遗传模型

4.1　主基因存在的鉴定及遗传效应的估计

首先分析一个主基因位点(A-a)的情形下 $F_{2,3}$ 家系的分布，然后考虑利用分离分析结果的遗传参数估计问题。假定亲本基因型为 AA(P_1)和 aa(P_2)，主基因的加显性效应用 d 和 h 表示，多基因符合加-显性模型，主基因效应和多基因效应相互独立，$F_{2,3}$ 家系按照主基因基因型可分为三类，即 aa 家系、Aa 家系和 AA 家系，家系平均数的方差应包含两部分，一部分是理论的家系平均数方差，另一部分是抽样方差，此处认为家系间附加的非遗传变异 E_b 为 0[14](详见表 4)。分别用 $N(\mu_1, \sigma_1^2)$、$N(\mu_2, \sigma_2^2)$和 $N(\mu_3, \sigma_3^2)$表示这三类家系平均数的分布。

表 4　$F_{2,3}$ aa 家系、Aa 家系和 AA 家系平均数的均值和方差

Table 4　The means and variances of various $F_{2,3}$ major gene genotypes

	aa 家系 aa families	Aa 家系 Aa families	AA 家系 AA families
家系平均数的均值 Mean of family means	$m-d$	$m+\frac{1}{2}h$	$m+d$
家系平均数的理论方差 Theoretical variance of family means	$\frac{1}{2}D+\frac{1}{16}H$	$\frac{1}{2}D+\frac{1}{16}H$	$\frac{1}{2}D+\frac{1}{16}H$
家系内的方差 Variance within family	$\frac{1}{4}D+\frac{1}{8}H+E$	$\frac{1}{2}d^2+\frac{1}{4}h^2+\frac{1}{4}D+\frac{1}{8}H+E$	$\frac{1}{4}D+\frac{1}{8}H+E$
家系平均数方差的估计(每个家系有 n 个观测值) Estimated variance of family means	$\frac{1}{2}D+\frac{1}{16}H+$	$\frac{1}{2}D+\frac{1}{16}H+$	$\frac{1}{2}D+\frac{1}{16}H+$
(n is the number of individuals in a family)	$(\frac{1}{4}D+\frac{1}{8}H+E)/n$	$(\frac{1}{2}d^2+\frac{1}{4}h^2+\frac{1}{4}D+\frac{1}{8}H+E)/n$	$(\frac{1}{4}D+\frac{1}{8}H+E)/n$

当只存在一个主基因位点并且多基因变异和环境变异较小时，对 $F_{2,3}$ 家系平均数构成的混合分布的分离分析应获得 3 个分布，这 3 个分布具有 1∶2∶1 的分离比，它们的均值分别用 μ_1、μ_2 和 μ_3 表示，根据上面的分析有以下关系：

$$\mu_1 = m - d$$
$$\mu_2 = m + (1/2)h$$
$$\mu_3 = m + d$$

因此，

$$m = (1/2)(\mu_1 + \mu_3)$$
$$d = (1/2)(\mu_3 - \mu_1)$$
$$h = 2\mu_2 - \mu_1 - \mu_3$$

但有时也可能由于多基因变异和环境变异的修饰作用，不能将分布 $N(\mu_1, \sigma_1^2)$和 $N(\mu_2, \sigma_2^2)$或 $N(\mu_2, \sigma_2^2)$和 $N(\mu_3, \mu_3^2)$区分开，分离分析的结果是：$F_{2,3}$ 家系群体是由两个分布构成的混合群体，两个分布所占比例满足 3∶1 的分离比，它们的均值分别用 μ_1 和 μ_2 表示，方差用 σ_1^2 和 σ_2^2 表示。下面以 $N(\mu_2, \sigma_2^2)$和 $N(\mu_3, \sigma_3^2)$不能区分为例说明遗传参数的估计方法。μ_1 和 σ_1^2 为分

布 $N(\mu_1, \sigma_1^2)$的均值和方差，μ_2 和 σ_2^2 为混合分布 $(2/3)N(\mu_2, \sigma_2^2)+(1/3)N(\mu_3, \sigma_3^2)$的均值和方差，因此有：

$$m - d = \mu_1$$
$$m + d/3 + h/3 = \mu_2$$
$$(2d - h)^2/18 + (2d^2 + h^2)/(6n) = \sigma_2^2 - \sigma_1^2$$

从中可以获得 m、d 和 h 的估计。但是，方差估计值的误差一般较大，因此第三个方程不可靠，这时可以假定 $d=h$(因为 3∶1 分离比)，从而只用前两个方程对遗传参数作大概的估计。

4.2 实例分析

表 5 列出大豆抗豆秆黑潜蝇遗传分析的数据，杂交组合的亲本为邳县天鹅蛋(P_1，感蝇)和 1138-2(P_2，抗蝇)，F_1 代平均表型与 P_2 亲本接近，F_2 代表现为偏态分布，峰不明显，从 127 棵 F_2 单株获得 $F_{2:3}$家系世代，每个家系观测 5 株。将 $F_{2:3}$家系平均数群体作为混合分布进行分解，表 6 给出不同分布个数下的极大似然函数及 *AIC* 值，当分布数为 2 时 *AIC* 取得最小值，因此认为 $F_{2:3}$家系是两个分布的混合，分解的结果列在表 7 中，分布的比例与 3∶1 的分离比无显著差异($\chi^2=0.78$，$P=0.38$)。利用：

$$\mu_1 = m + (1/3)d + (1/3)h = 12.08$$
$$\mu_2 = m - d = 8.62$$
$$d = h$$

可以得到 $m=10.00$，$d=-2.08$。总的方差 $\sigma_p^2=2.98$，主基因的方差为：

$\sigma_{mg}^2=d^2/4+h^2/8+(d^2/2+h^2/4)/n=2.27$，

因此主基因的遗传率为：$h_{mg}^2=\sigma_{mg}^2/\sigma_p^2=76.17\%$。

表 5 大豆邳县天鹅蛋(P_1)×1138-2(P_2)$F_{2:3}$家系茎秆虫量分布

Table 5 Distributions of number of maggots of beanfly in stem in a soybean cross

	4	5	6	7	8	9	10	11	12	13	14	15	16
P_1						1	2	7	2	3	3	0	2
P_2	1	2	2	7	4	3	1						
F_1		1	3	3	3	6	1	3					
$F_{2:3}$				4	17	50	18	4	19	1			

表 6 邳县天鹅蛋和 1138-2 衍生的 $F_{2:3}$家系群体在不同成分分布个数时的 *AIC* 值

Table 6 Akaike's information values of the $F_{2:3}$ populations derived from Pixiantianedan×1138-2

成分分布数(k) Component number k	1	2	3	4
独立参数的个数($N(k)$) Number of independent parameters $N(k)$	2	5	8	11
对数似然函数的极大值($L(\hat{\varphi})$) Maximum of log likelihood	−249.55	−214.40	−212.75	−208.96
AIC 值 *AIC* value	503.10	438.80	441.49	439.82

表 7 $F_{2:3}$世代混合分布中各参数的极大似然估计(括号内的为标准误)

Table 7 Maximum likelihood estimates of component distributions of the $F_{2:3}$ generation (standard errors in parentheses)

成分分布 Component	权 重(π) Weight	均 值(m) Mean	方 差(σ^2) Variance
1	0.28(0.04)	12.08(0.12)	0.44(0.17)
2	0.72(0.04)	8.62(0.09)	0.60(0.09)

5 讨论

主-多基因性状在分离世代中表现出多峰现象，具有混合分布的特征，因此可以利用混合分布理论鉴定主基因的存在并估计遗传效应。但是不同分离世代鉴定主基因的功效是不同的，F_2 世代的功效最低，回交世代和 $F_{2,3}$ 家系世代的功效较高；只利用个别分离世代的信息只能对主基因是否存在进行鉴定，无法对多基因的存在进行鉴定，多世代的联合分析方法可以对多基因的存在进行鉴定，并且可以判断多基因是否满足加-显性模型，并进一步估计多基因的遗传效应。

参 考 文 献

1 姜长鉴、莫惠栋，1995，作物学报，21(6)，641～648

2 马育华编著，1982，植物育种的数量遗传学基础，江苏科学技术出版社，南京

3 莫惠栋，1993，作物学报，19(1)，1～6

4 王建康、盖钧镒，1995，生物数学学报，10(4)，87～92

5 王建康、盖钧镒(导师)，1996，数量性状主基因-多基因混合遗传模型的鉴别和遗传参数估计的研究(博士学位论文)，南京农业大学

6 王建康、盖钧镒，1997，遗传学报，24(5)：432～440

7 Akaike, H., 1977, In: P. R. Krishnaiah(ed.) Application of Statistics, pp. 27～41, North-Holland Publishing Company, Amsterdam

8 Elkind, Y. and A. Cahaner, 1986, Theor. Appl. Genet., 72, 377～383

9 Elkind, Y. and A. Cahaner, 1990, Heredity, 64, 205～213

10 Dempster, A. P., N. M. Laird and D. B. Robin, 1977, J. R. Statist. Soc. B., 39, 1～38

11 Elston, R. C., 1984, Genetics, 108, 733～744

12 Goffinet, B., P. Loisel and B. Laurent, 1990, Biometrics, 46, 583～594

13 Loisel, P., B. Goffinit, H. Monod and G. M. De Oca, 1994, Biometrics, 50, 512～516

14 Mather, K. and J. L. Jinks, 1982, Biometrical Genetics, Chapman and Hall

15 McLachlan, G. J., 1988, Mixture Models: Inference and Applications to Clustering, Marcel Dekker, Inc.

16 Tanksley, S. D., 1993, Annu. Rev. Genet., 27, 205～233

Identification of Major Gene and Polygene Mixed Inheritance Model from Backcrosses or $F_{2,3}$ Families

Gai Junyi　Wang Jiankang

(*Soybean Research Institute, Nanjing Agricultural University, Nanjing* 210095)

Abstract　For major-polygene traits, the distribution of a segregating population demonstrates multimodality which is the characteristic of a mixture of more than one distributions. A statistical method to identify major gene and polygene mixed inheritance from backcrosses and $F_{2,3}$ families was developed by using the theory of mixture distribution. With the method, the existence of major gene and its genetic effects could be determined.

Key words　Quantitative trait; Major gene and polygene mixed inheritance; Mixture distribution; EM algorithm