遗 传 学 报，24(5)：432～440，1997
Acta Genetica Sinica

利用杂种 F_2 世代鉴定数量性状主基因-多基因混合遗传模型并估计其遗传效应

王建康① 盖钧镒
(南京农业大学大豆所 南京 210095)

摘 要 大量实验数据和 QTL 作图结果表明：控制数量性状的基因中既有遗传效应较大的主基因，又有遗传效应较小的多基因，其分离世代表现出多峰性，即出现多个分布混合的特征。本文利用混合模型理论的 AIC 信息准则在 F_2 世代中鉴定影响数量性状的主基因的存在，当主基因存在时通过分离分析估计主基因的遗传效应以及主基因的遗传变异占总变异的分量；同时还提出利用 P_1、P_2、F_1 和 F_2 4个世代鉴定多基因存在的方法。以大豆开花期性状为例说明了该方法的应用，在所分析的骨绿豆 × 上海红芒早杂交组合的 F_2 世代数据中发现有主基因的存在，主基因表现出完全显性(晚开花)，并且有多基因存在。

关键词 数量性状，主基因-多基因混合遗传，混合模型，EM 算法，大豆开花期

生物的性状分为质量性状和数量性状两类，一般认为质量性状受少数主基因控制，在表型上形成间断性变异，并且可以从表型或后裔测验中判断个体的基因型；数量性状受大量微效基因的控制，多个微效基因的集合形成一个多基因系统，对这类性状的研究需借助统计遗传学的方法[2,11,22]。然而在实际工作中还发现许多数量性状，在植物中如大豆的生育期[5]、水稻的株高[3,17]、水稻对白叶枯病和稻瘟病的抗性、小麦和大麦对叶锈病的抗性[7,8,20]等，在动物中如羊的影响排卵率的 *Booroola* 基因、牛的隐性矮小症基因、家禽中的矮基因等[15,18,19]，它们在分离世代中既有可分组的趋势，但又存在组间界线模糊现象，这类性状在遗传上同时受少数主基因和大量多基因的控制，称之为主基因-多基因混合遗传体系；同时大量 QTL 定位结果也表明控制数量性状的基因在效应大小上有很大的差异[28]。传统数量遗传研究的对象为数量性状，在分离世代中，数量性状的表现型为一近似正态分布，而主基因-多基因混合遗传体系控制的性状的表型呈现出多峰性，表现出多个分布混合的特征。因此在分析数量性状遗传的传统方法中，关于遗传型和表现型分布的基本假定对主基因-多基因混合遗传性状来说已不成立，需要相应的鉴别主基因和多基因的存在以及确定主基因遗传效应等参数估计的统计方法。

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

①现在河南省农业科学院工作

本文于1996-01-22收到，1996-10-31修回

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[21]推导了F_2世代中检验主基因存在的似然比统计量的渐进性质，姜长鉴[17]将Loisel[21]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。混合模型(Mixture Model)或有限混合分布(Finite Mixture Distribution)在现代统计的整个发展过程中作为一个模型得到广泛的研究和应用[4,8,13,23,29]。凡是数据结构具有以下特征，均可应用混合模型的理论：数据来自一组不同分布的混合，组成这一混合分布的单个分布是未知的。主基因-多基因混合遗传性状在分离世代中表现出混合分布的特征，因此本文研究利用混合分布理论在F_2世代中鉴定主基因存在的方法，当主基因存在时给出有关遗传参数的估计方法；同时还提出利用P_1、P_2、F_1和F_2 4个世代鉴定多基因存在的方法。

1 混合分布的一般理论

1.1 基本概念

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$\pi_j > 0$, $j = 1, \cdots, k$, $\pi_1 + \pi_2 + \cdots + \pi_k = 1$,

$$f_j(\cdot) \geqslant 0, \int_\Omega f_j(x)\,dx = 1, j = 1, \cdots, k,$$

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π_1, π_2, $\cdots$, π_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分(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$f_j(\cdot)$的参数向量，那么混合分布的密度函数也可用参数形式表示成:

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

$\varphi = (\pi_1, \pi_2, \cdots, \pi_k, \theta_1, \theta_2, \cdots, \theta_k)$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，具体说来有3个方面的内容: (1)确定混合分布中所包含的成分分布的个数; (2)成分分布的数字特征; (3)正态混合分布下的方差同质性和正态性检验。

1.2 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来获得，在以后的分析中$L(\varphi)$表示对数似然函数，即: $L(\varphi) = \sum_{i=1}^{n} \log p(x_i;\ \varphi)$

1.3 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，不知道k就无法进一步确

定其他参数；此外在实际应用中，k也是一个重要的指标，例如，从鱼群的体长数据中，是否能确定这些个体是从同一年产的卵发育而来还是从不同年份的卵发育而来？给定一个岩石样本，从颗粒大小数据中确定其中只包含一种矿石还是由多种不同矿石组成？在主基因-多基因复合遗传性状分析中，根据k的大小可以判定主基因的存在，如果$k=1$，则说明无主基因存在；如果$k\geq 2$，则说明有主基因存在。

确定成分分布数目的方法可分为图形方法[24,28]和统计检验方法[15,22,23]两大类，本文利用 Akaike[6]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：$AIC = AIC(k) = -2L(\hat{\varphi}) + 2N(k)$，其中$\hat{\varphi}$是假设"$H_0$：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$k$值作为成分分布个数的估计。

1.4 混合模型中其它参数的估计

混合模型的参数估计一般利用 EM 算法来完成，EM 算法分两个步骤进行[22,28]：

E步骤：计算$L_C(\varphi)$在初始值$\varphi^{(0)}$下的期望值$Q(\varphi, \varphi^{(0)})$，

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

$L_C(\varphi)$为完全数据的似然函数，$\tau_{ij}^{(0)}$的意义见下文。

M步骤：极大化$Q(\varphi, \varphi^{(0)})$，并用极大值点处的$\varphi$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：

$$\pi_i = \sum_{j=1}^{n}\tau_{ij}^{(0)} / n \quad i = 1, \cdots, k$$

$$\sum_{j=1}^{n}\sum_{i=1}^{k}\tau_{ij}^{(0)}\,\partial\log f_i(x_j;\ \theta) / \partial\theta = 0$$

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，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不可能的。

(2) EM迭代过程中，似然函数是单调增加的，即：

$$L(\varphi^{k+1}) \geq L(\varphi^{k})$$

$k\geq 0$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较大的依赖性，关于收敛速度与初始值的关系问题将在其他文章中加以探讨。对于主基因-多基因复合遗传性状来说，混合分布中的每个成分分布有相同的方差σ^2，EM算法具体过程是：

假定$\varphi^{(0)} = (\pi_1^{(0)}, \cdots, \pi_k^{(0)}, \mu_1^{(0)}, \cdots, \mu_k^{(0)}, \sigma^{2(0)})$是初始值，$f(x;\ \mu,\ \sigma^2)$表示正态分布的密度函数，则：

$$\tau_{ij}^{(0)} = Pro(x_j \in G_i \mid x_j;\ \varphi^{(0)})$$
$$= \pi_i^{(0)} f(x_j;\ \mu_i^{(0)},\ \sigma^{2(0)}) \Big/ \sum_{t=1}^{k} \pi_t^{(0)} f(x_j;\ \mu_t^{(0)},\ \sigma^{2(0)})$$

$$\pi_i^{(1)} = \sum_{j=1}^{n} \tau_{ij}^{(0)} / n$$

$$\mu_i^{(1)} = \sum_{j=1}^{n} \tau_{ij}^{(0)} x_j / (n\pi_i^{(1)})$$

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

将求得的参数值$\varphi^{(1)}$代替$\varphi^{(0)}$开始下一轮 EM 迭代。

2 主基因-多基因混合遗传模型的遗传分析

2.1 主基因-多基因混合遗传模型的建立

主基因和多基因混合遗传模型的提出把生统遗传学和孟德尔遗传学很好地统一起来[8,9,11,20]。来自 F_2世代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，c 为多基因效应，e 为环境效应。在以下的分析中我们假定主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 对相同的主基因基因型是固定的，为一固定值; 多基因效应 $c \sim N(0,\ \sigma_{pg}^2)$是随机变量; σ_{pg}^2为多基因效应值的方差;环境效应 $e \sim N(0,\ \sigma_e^2)$是随机变量，σ_e^2为环境方差。如果主基因有 k 个不同的基因型值，那么 F_2群体表现为 k 个正态分布的混合，通过分离分析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 F_2群体的情况下，不能将二者分解开来，记$\sigma^2 = \sigma_{pg}^2 + \sigma_e^2$。

2.2 主基因遗传效应的估计

以下对只有一个主基因位点存在的情况下，分析主基因的遗传效应，更复杂的情况需借助多个世代的联合分析方法。

2.2.1 主基因表现为部分显性　如果通过分离分析发现 F_2混合群体中包含 3 个分布，并且 3 个分布所占比例具有 1∶2∶1 的分离比，此时认为只存在有一个主基因(用 *A-a*表示)，并表现为部分显性(或无显性)，3 个分布中个体的主基因基因型分别为 *aa*、*Aa* 和 *AA*，分布的均值μ_1、μ_2和μ_3与主基因的加性效应 *d*、显性效应 *h* 和群体平均数 *m* 之间的关系为:

$$\mu_1 = m - d$$
$$\mu_2 = m + h$$
$$\mu_3 = m + d$$

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

$$\hat{m} = 0.5\hat{\mu}_1 + 0.5\hat{\mu}_3$$
$$\hat{d} = -0.5\hat{\mu}_1 + 0.5\hat{\mu}_3$$
$$\hat{h} = -0.5\hat{\mu}_1 + \hat{\mu}_2 - 0.5\hat{\mu}_3$$

因此 $\hat{m}$、$\hat{d}$ 和 $\hat{h}$ 的方差分别为:

$$\mathrm{Var}(\hat{m}) = 0.25\mathrm{Var}(\hat{\mu}_1) + 0.25\mathrm{Var}(\hat{\mu}_3) + 0.5\mathrm{Cov}(\hat{\mu}_1, \hat{\mu}_3)$$
$$\mathrm{Var}(\hat{d}) = 0.25\mathrm{Var}(\hat{\mu}_1) + 0.25\mathrm{Var}(\hat{\mu}_3) - 0.5\mathrm{Cov}(\hat{\mu}_1, \hat{\mu}_3)$$
$$\mathrm{Var}(\hat{h}) = 0.25\mathrm{Var}(\hat{\mu}_1) + \mathrm{Var}(\hat{\mu}_2) + 0.25\mathrm{Var}(\hat{\mu}_3) - \mathrm{Cov}(\hat{\mu}_1, \hat{\mu}_2) - \mathrm{Cov}(\hat{\mu}_2, \hat{\mu}_3) + 0.5\mathrm{Cov}(\hat{\mu}_1, \hat{\mu}_3)$$

主基因的遗传方差$\sigma^2_{mg} = 0.5d^2 + 0.25h^2$，主基因遗传方差占总变异的比例是总的表型变异中归属于主基因的那一部分，称为主基因遗传率，记为h^2_{mg}[18,19]：

$$h^2_{mg} = \sigma^2_{mg} / (\sigma^2_{mg} + \sigma^2) = (2d^2 + h^2) / (2d^2 + h^2 + 4\sigma^2)$$

2.2.2 主基因表现为完全显性 通过分离分析发现F_2混合群体中包含2个分布，并且所占比例具有1∶3的分离比，此时认为存在有一个表现为完全显性(或隐性)的主基因(用*A*–*a*表示)。2个分布中个体的主基因基因型分别为*aa*和*AA* + *Aa*，分布的均值用μ_1和μ_2表示，因此，

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

从而获得*m*和*d*的极大似然估计：

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

$\hat{m}$和$\hat{d}$的方差分别为：

$$\mathrm{Var}(\hat{m}) = 0.25\mathrm{Var}(\hat{\mu}_1) + 0.25\mathrm{Var}(\hat{\mu}_2) + 0.5\mathrm{Cov}(\hat{\mu}_1, \hat{\mu}_2)$$
$$\mathrm{Var}(\hat{d}) = 0.25\mathrm{Var}(\hat{\mu}_1) + 0.25\mathrm{Var}(\hat{\mu}_2) - 0.5\mathrm{Cov}(\hat{\mu}_1, \hat{\mu}_2)$$

主基因的遗传方差$\sigma^2_{mg} = 0.75d^2$，主基因遗传率为：

$$h^2_{mg} = \sigma^2_{mg} / (\sigma^2_{mg} + \sigma^2) = 3d^2 / (3d^2 + 4\sigma^2)$$

2.3 多基因存在的鉴定

在以上的分析中，如果主基因存在的话，我们已经可以将主基因的变异从总的表型变异中分离出来，但是仍需确定主基因之外的变异是来自多基因或是环境或是两者共同

表1 P_1，P_2，F_1和F_2群体的一些统计参数

Table 1 Parameters of P_1, P_2, F_1 and F_2 populations

群体 Poupulation	样本量 Sample size	样本值 Sample value	均值 Mean	方差 Variance	群体分布特征 Population distribution	概率密度函数 Density function
P_1	n_1	x_{1i}	μ_1	σ^2_e	单一正态分布 Normal distribution	$f(x; \mu_1, \sigma^2_e)$
P_2	n_2	x_{2i}	μ_2	σ^2_e	单一正态分布 Normal distribution	$f(x; \mu_2, \sigma^2_e)$
F_1	n_3	x_{3i}	μ_3	σ^2_e	单一正态分布 Normal distribution	$f(x; \mu_3, \sigma^2_e)$
F_2	n_4	x_{4i}			正态混合分布 Normal mixture	$p(x; \varphi)$[1)]

1): $p(x; \varphi) = \sum_{j=1}^{k} \pi_j f(x; \mu_j, \sigma^2)$

作用的结果，即对多基因的存在进行鉴定。表 1 给出 P_1、P_2、F_1和 F_2世代群体的一些统计参数，因此样本似然函数为:

$$L(\varphi)=\prod_{i=1}^{n_1} f(x_{1i};\ \mu_1,\ \sigma_e^2)\prod_{i=1}^{n_2} f(x_{2i};\ \mu_2,\ \sigma_e^2)\prod_{i=1}^{n_3} f(x_{3i};\ \mu_3,\ \sigma_e^2)\prod_{i=1}^{n_4} p(x_i;\ \varphi)$$

构造零假设 $H_0: \sigma^2=\sigma_e^2$(不存在多基因)和备择假设 $H_a: \sigma^2>\sigma_e^2$(有多基因存在，多基因的遗传方差为$\sigma_{pg}^2=\sigma^2-\sigma_e^2$)。通过计算两种假设下似然函数的最大值 L_0和 L_a构造出似然比统计量$\lambda=2(\log L_a-\log L_0)\sim x^2(1)$, 进而对上述假设进行显著性测验。

3　大豆开花期性状的遗传分析

3.1　数据来源

为了研究大豆生育期性状的光温反应特性，南京农业大学大豆研究所把不同生育期品种进行杂交，在不同播季下种植杂交后代,观察生育期性状在不同播季下的表现[5]。表 5 中的前两列给出杂交组合骨绿豆 × 上海红芒早在夏播条件下 158 个 F_2个体自出苗到开花的天数(按从小到大的顺序排列)。

3.2　结果分析

从图 1 的频数分布可以看出 F_2群体不具有单一分布的特征而是表现为两个或多个成分分布的混合，表 2 给出混合群体在不同成分分布个数的条件下对数似然函数的极大值和 *AIC* 值，从中看出: 对骨绿豆 × 上海红芒早产生的 F_2群体来说，当成分分布数 $k=2$ 时，*AIC* 值达到最小，由此确定这个混合群体中包含两个成分分布。确定了混合群体中的成分分布数后，就可通过分离分析确定各个成分分布的数字特征，表 3 给出混合群体中各参数的极大似然估计，括号内的数字代表标准误。由表 3 可以看出骨绿豆 × 上海红芒早产生的 F_2混合群体中，两个基本分布所占的比例接近 1∶3($\chi^2=0.07$, $P=0.85$)，说明存在有一个主基因位点，该主基因位点表现为完全显性。由表 3 可以进一步获得主基因遗传效应和遗传方差的估计，其结果列在表 4 中。由分离分析的结果还能对样本进行分类，分类时采用 Bayes 的分类方法，从而可以确定 F_2群体中不同个体的主基因基因型，分类的结果见表 5。利用由 F_2衍生的 F_3株行数据可以对上述结论作进一步的验证，表 4 最

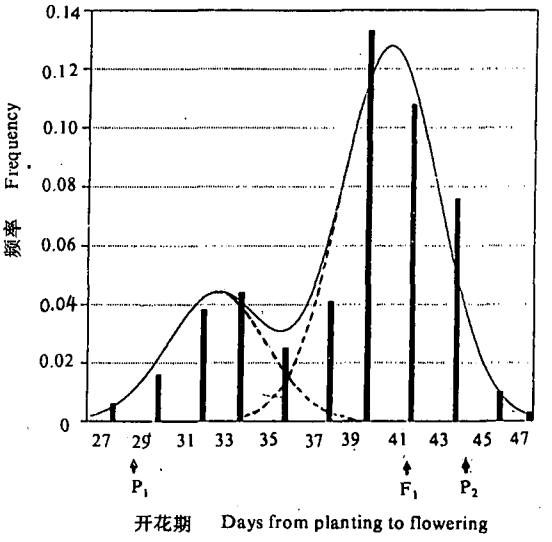

图 1　骨绿豆×上海红芒早 F_2 群体的分离分析图

条形图为样本的频数分布图，实线为混合分布的密度函数曲线，虚线为混合分布中所包含的两个正态分布的密度函数曲线，箭头所指方向是 P_1、F_1和 P_2群体的平均数

Fig. 1　The result of segregation analysis of the F_2 population derived from Guludou × Shanghaihongmangzao

The barline is the frequency distribution; the solid line is the mixture distribution; the dash lines are distribution curves of components in the mixture; and the arrows indicate the means of P_1, F_1 and P_2 respectively

表2 不同成分分布个数时的*AIC*值

Table 2 Akaike's information values under different component numbers

成分分布数(k) Component number(k)	1	2	3	4	5
独立参数的个数($N(k)$) Number of independent parameters($N(k)$)	2	4	6	8	10
对数似然函数极大值($L(\hat{\varphi})$) Maximum of log likelihood function ($L(\hat{\varphi})$)	−451.62	−433.61	−433.53	−432.06	−431.46
*AIC*值 *AIC* value	907.25	875.23	879.01	880.13	882.93

表3 F_2世代混合分布中各参数的极大似然估计(括号内的数字为标准误)

Table 3 Maximum likelihood estimates in F_2 population(Numbers in brackets are standard errors)

基本分布 Component No.	权 重(π) Weight(π)	均值(μ) Mean(μ)	方差(σ^2) Variance(σ^2)
1	0.26(0.04)	31.80(0.44)	5.35(0.69)
2	0.74(0.04)	39.87(0.24)	5.35(0.69)

表4 主基因遗传效应的极大似然估计和主基因遗传方差

Table 4 Maximum likelihood estimates of major gene effects and genetic variance of major gene

世 代 Progeny	群体平均数(m) Population mean(m)	加性效应(d) Additive effect(d)	主基因遗传方差(σ^2_{mg}) Genetic variance of major gene (σ^2_{mg})	主基因遗传率(h^2_{mg})(%) Heritability of major gene (h^2_{mg})
F_2	35.84(0.27)	4.04(0.23)	12.21	68.19
F_3家系 F_3 families	36.83(0.41)	3.45(0.57)	6.20	78.97

表5 骨绿豆×上海红芒早F_2世代不同个体的主基因基因型分类

Table 5 Classification of F_2 individuals of Guludou×Shanghaihongmangzao

个体序号 No.	开花期(天) Days to flowering(days)	后验概率[1] Posterior probability		主基因基因型 Genotype of major gene
		aa	*Aa*+*AA*	
1–9	26～30	1.00	0.00	*aa*
10–19	31	0.99	0.00	*aa*
20–23	32	0.99	0.01	*aa*
24–33	33	0.96	0.04	*aa*
34–38	34	0.85	0.16	*aa*
39–41	35	0.55	0.45	*aa*
42–48	36	0.21	0.79	*Aa*+*AA*
49–54	37	0.06	0.94	*Aa*+*AA*
55–74	38	0.02	0.98	*Aa*+*AA*
75–96	39	0.01	0.99	*Aa*+*AA*
97–158	40～47	0.00	1.00	*Aa*+*AA*

1) 样本来自不同成分分布的概率。 Probabilities of individual being a member of each component

后一列给出由骨绿豆 × 上海红芒早衍生 F_3世代数据(与 F_2世代同期种植)估计出的主基因的加性效应和遗传方差，主基因存在的鉴定结果与 F_2世代是一致的。利用 P_1、P_2、F_1和 F_2世代的数据资料，可以得到 $L_0 = -493.29$，$L_a = -486.67$，$\lambda = 13.23$，显著性概率 $P = 0.000275$，因此我们接收备择假设"H_a：多基因存在"。多基因的遗传方差 $\sigma^2_{pg} = \sigma^2 - \sigma^2_e = 3.55$，多基因遗传方差占总变异的比例是总的表型变异中归属于多基因的那一部分，称之为多基因遗传率，记为 $h^2_{pg} = 19.96\%$。

4 讨论

对于主基因-多基因混合遗传控制的数量性状，我们认为其 F_2分离世代为一正态混合分布，因此在采用混合分布理论之前首先应对成分分布的正态性进行检验，对于偏离正态的情况应考虑利用数据变换将其转换为正态分布。由分离分析获得的关于主基因数目和分离群体中个体的主基因基因型等结论可以利用实验数据作进一步的验证，例如可利用回交世代数据对主基因数目和显隐性进行验证，利用 F_3家系数据对 F_2个体的主基因基因型进行检验。同时考虑亲本和 F_1世代，还可对主基因之外的变异作进一步的分解，进而对多基因的存在进行鉴定，多基因的存在在一些作物抗性持久性的利用上有重要意义。

主基因-多基因混合遗传模型的提出实现了生统遗传学和孟德尔遗传学的统一。一个基因位点是主基因还是多基因是相对于它所产生的表型效应大小而言的，主基因可以产生较大的表型效应，以至在分离世代中出现多峰现象；但是一个基因位点是否表现为主基因可能还与环境有关，也就是说在某种环境条件下表现为主基因的位点在另一种环境下不一定仍表现为主基因[5]。在所分析的杂交组合中，主基因的变异占总表型变异的分量达 68.19%。如果主基因的效应过小，或是有多个遗传效应类似的主基因存在，都会使上述方法变得困难起来。利用 F_3家系平均数，可以检测到在 F_2世代检测不到的主基因，多个世代的联合分析可以更精确地估计各种遗传参数。

生育期性状是大豆的重要育种目标性状之一，亲本材料中主效基因的存在可为杂交育种的亲本选配和回交育种中供体亲本的选择提供指导。

参 考 文 献

1 姜长鉴，莫惠栋。作物学报，1995，21(6)：641～648
2 马育华编著。植物育种的数量遗传学基础，南京：江苏科学技术出版社，1982
3 莫惠栋。作物学报，1993，19(1)：1～6
4 王建康，盖钧镒。生物数学学报，1995，10(4)：87～92
5 杨永华，盖钧镒，马育华。中国农业科学，1994，27(3)：1～6
6 Akaike H. In: P R Krishnaiah(ed.) Application of Statistics, North-Holland Publishing Company, Amsterdam: 1977, 27～41
7 Barker H *et al.* Theor. Appl. Genet., 1994, 88: 754～758
8 Conner J. 1993, Genetica, 90: 41～45
9 Elkind Y *et al.* Theor. Appl. Genet., 1986, 72: 377～383
10 Elston R C. Genetics, 1984, 108: 733～744
11 Falconer D S. Introduction to Quantitative Genetics, and edn., Longman: London and New York: 1981
12 Fernando R L *et al.* Theor. Appl. Genet., 1994, 88: 573～580
13 Goffinet B *et al.* Biometrics, 1990, 46: 583～594
14 Gray G. Biometrics, 1994, 50: 457～470

15 Hoeschele I. Theor. Appl. Genet., 1988, 76: 311～319
16 Hsu Y S *et al.* Mathematical Geology, 1986, 18(2): 153～160
17 Jiang C *et al.* Genetics, 1994, 136: 383～394
18 Knott S A *et al.* Heredity, 1991, 68: 299～311
19 Knott S A *et al.* Heredity, 1991, 68: 313～320
20 Kreike C M *et al.* Theor. Appl. Genet., 1994, 88: 764～769
21 Loisel P *et al.* Biometrics, 1994, 50: 512～516
22 Mather S K, J L Jinks. Biometrical Genetics, Chapman and Hall, 1982
23 McLachlan G J. Appl. Statist., 1987, 36: 318～324
24 McLachlan G J. Mixture Models: Inference and Applications to Clustering, Marcel Dekker, Inc, 1988
25 Roeder K. Journal of the American Statistical Association, 1994, 89(426): 487～495
26 Shoukri M M, G L McLachlan. Biometrics, 1994, 50: 128～139
27 Suzuki D T *et al.* An Introduction to Genetic Analysis, W.H. Freeman and Company / New York: 1986
28 Tanksley S D. Annu. Rev. Genet., 1993, 27: 205～233
29 Titterington D M. Statistical Analysis of Finite Mixture Distributions, John Wiley & Sons, 1985
30 Tondini F *et al.* Euphytica, 1993, 69: 109～114

Identification of Major Gene and Polygene Mixed Inheritance Model and Estimation of Genetic Parameters of a Quantitative Trait from F_2 Progeny

WANG Jiankang GAI Junyi

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

Abstract

It has been proved by many field experiments and QTL mapping results that among genes affecting some quantitative traits there are some major genes with larger genetic effect and some polygenes with smaller genetic effect. For such traits, the distribution of segregating population demonstrates multimodality, and this is the characteristic of the mixture of more than one distributions. Mixtrue distribution models have been used extensively as models in a wide variety of practical situations where data can be viewed as arising from two or more populations mixed in certain proportions. Akaike's Information Criterion(AIC) has been used to identify the existence of major genes affecting quantitative traits. Under the existence of major genes, the genetic effects of these genes and their genetic variance were estimated through segregation analysis. The genotype of major gene of F_2 individuals of were determined by clustering using Bayesian criterion. With P_1, P_2, F_1 and F_2 populations, the likelihood ratio test was used to test the existence of polygenes. In the end, the inheritance of soybean flowering date is analyzed. One major gene was found in F_2 population derived from Guludou × Shanghaihongmangzao.

Key words Major gene and polygene mixed inheritance, Quantitative traits, Mixture model, EM algorithm, Soybean flowering date