

QTL 混合遗传模型扩展至 2 对主基因+ 多基因时的多世代联合分析*

盖钧镒** 章元明 王建康***

(南京农业大学大豆研究所, 农业部国家大豆改良中心, 江苏南京, 210095)

提 要 本文将 Gai and Wang (1998) 的 P_1 、 F_1 、 P_2 、 B_1 、 B_2 和 F_2 6 个世代联合分离分析方法从 A 、 B 、 C 、 D 4 类共 17 种遗传模型扩展至 E (2 对主基因+ 多基因) 共 5 类 24 种遗传模型。成分分布参数估计的方法由 EM 算法改进为迭代条件 EM 算法 ($IECM$)，其收敛性和所获结果优于 EM 算法。水稻株高例中 2 对主基因+ 多基因的遗传模型 ($E-2$) 优于原所获得的 1 对主基因+ 多基因的遗传模型 (D)。

关键词 数量性状; 主基因+ 多基因混合遗传; 多世代联合分析; $IECM$ 算法

A Joint Analysis of Multiple Generations for QTL Models Extended to Mixed Two Major Genes Plus Polygene

GA I Jun-Yi ZHANG Yuan-Ming WANG Jian-Kang

(Soybean Research Institute, Nanjing Agricultural University; National Center of Soybean Improvement, Ministry of Agriculture, Nanjing 210095)

Abstract The two major genes plus polygene Model (Model E) of QTL in a joint analysis of P_1 , F_1 , P_2 , B_1 , B_2 and F_2 was extended to based on Gai and Wang (1998). The major steps were to establish the genetic models and respective maximum likelihood functions, to estimate the distribution parameters through iterated expectation and conditional maximization ($IECM$) algorithm, to select the best genetic model through AIC value and tests for goodness of fit, to estimate the genetic parameters of the best fitted model through least squares method, and to classify the individuals into major gene genotypes in terms of Bayesian posterior probability. The example of the inheritance of rice plant height which was used by Gai and Wang (1998) was further analyzed to explain the procedures. It was shown through the example that the convergence of the parameter estimation in $IECM$ algorithm was better than that in EM algorithm, and the two major genes plus polygene model ($E-2$) was better than the previous one major gene plus polygene model (D).

Key words Quantitative trait; Major gene plus polygene mixed inheritance; Joint analysis of multiple generations; $IECM$ algorithm

对植物数量性状遗传体系的认识有一个发展过程^[1-2]。最早认为数量性状由微效多基因

* 国家 973 项目

** 联系作者, E-mail: sri@njau.edu.cn 或 nausri@public1.ptt.js.cn

*** 现在河南农业科学院科学实验中心工作

致谢: 南京农业大学农学系朱立宏教授惠允使用南京 6 号 × 广丛杂交组合 6 个基本世代群体株高资料, 谨致谢忱!

收稿日期: 1999-07-28, 接受日期: 1999-12-18

控制, 以后发现水稻矮秆等一些数量性状也受主基因控制; Elston and Steward (1973)^[3]提出了 1 对主基因+ 多基因混合遗传模型。后来大量研究发现控制数量性状的基因效应大小不一, 大者表现为主基因, 小者表现为微效多基因。因而盖钧镛等(1999)^[11]提出将主基因+ 多基因混合遗传看作植物数量性状遗传的普遍性模型, 将单纯多基因、单纯主基因看作其特例, 并在前人基础上^[3~5]发展了一套 *QTL* 遗传模型检测的分离分析方法, 将孟德尔用于研究主基因的遗传研究方法延伸到数量性状的通用模型。这种方法通过对分离世代分析可以检测 1 对主基因(A)、2 对主基因(B)、多基因(C)、1 对主基因+ 多基因(D)等 4 类多种遗传模型, 所用试验材料可以是单个分离世代及其亲本, 也可以是多个分离世代及其亲本^[6~9]。这种分离分析的主要理论基础是混合分布理论, 将分离世代的分布看作为多个主基因型在多基因和环境修饰下形成的多个正态分布的混合分布^[10], 其主要方法是通过极大似然法和 *EM* 算法对混合分布中有关成分分布参数作出估计, 然后通过 *AIC* 值判别及一组适合性测验从中选出最适遗传模型, 并由之估计相应的主基因和多基因的效应值、方差和有关遗传参数, 乃至对参试材料的主基因型作出后验概率判别。Gai and Wang (1998)^[9]提出的 P_1 、 F_1 、 P_2 、 B_1 、 B_2 、 F_2 6 世代联合分析方法考虑了上述 4 类 17 种遗传模型, 并在此基础上进一步拓展分离世代的类型, 包括个体世代的类型及可进行重复试验提高精确度的家系世代的类型, 以及拓展为更为复杂的遗传模型等方面改进提高这套方法。本文目的是将 Gai and Wang (1998)^[9]的 6 世代联合分析方法进一步拓展至 2 对主基因+ 多基因遗传模型的情况, 使 6 世代试验资料可鉴别 5 类共 24 种遗传模型, 并估计相应遗传参数。

1 理论推演

1.1 基本假定与遗传模型

本文对 P_1 、 F_1 、 P_2 、 B_1 、 B_2 及 F_2 *QTL* 体系及 A、B、C、D 4 类共 17 种遗传模型所作的基本假定、符号和参数与 Gai and Wang (1998)^[9]相同, 此处不再重复。现将 2 对主基因+ 多基因简称为 *E* 类模型, 其相应 6 世代的分布通式列于后, 观察值以个体或植株为单位, 此处假定 2 对主基因属独立遗传。至于主基因间连锁遗传有待进一步研究。

$$P_1: X_{1i} \sim N(\mu_1, \sigma^2) \quad F_1: X_{2i} \sim N(\mu_2, \sigma^2) \quad P_2: X_{3i} \sim N(\mu_3, \sigma^2)$$

$$B_1: X_{4i} \sim (1/4)N(\mu_{41}, \sigma_1^2) + (1/4)N(\mu_{42}, \sigma_1^2) + (1/4)N(\mu_{43}, \sigma_1^2) + (1/4)N(\mu_{44}, \sigma_1^2)$$

$$B_2: X_{5i} \sim (1/4)N(\mu_{51}, \sigma_2^2) + (1/4)N(\mu_{52}, \sigma_2^2) + (1/4)N(\mu_{53}, \sigma_2^2) + (1/4)N(\mu_{54}, \sigma_2^2)$$

$$F_2: X_{6i} \sim (1/16)N(\mu_{61}, \sigma_6^2) + (1/8)N(\mu_{62}, \sigma_6^2) + (1/16)N(\mu_{63}, \sigma_6^2) + (1/8)N(\mu_{64}, \sigma_6^2) + \\ (1/4)N(\mu_{65}, \sigma_6^2) + (1/8)N(\mu_{66}, \sigma_6^2) + (1/16)N(\mu_{67}, \sigma_6^2) + \\ (1/8)N(\mu_{68}, \sigma_6^2) + (1/16)N(\mu_{69}, \sigma_6^2)$$

以上 6 个群体的分布分别由单个、4 个和 9 个各不相同的成分分布组成。在两对主基因服从等加性、完全显性、等显性且多基因服从加性-显性模型下, B_1 、 B_2 和 F_2 群体的成分分布数分别为: $3(1 \ 2 \ 1)$ 、 $3(1 \ 2 \ 1)$ 和 $5(1 \ 4 \ 6 \ 4 \ 1)$; $1, 4(1 \ 1 \ 1 \ 1)$ 和 $4(9 \ 3 \ 3 \ 1)$; $1, 3(1 \ 2 \ 1)$ 和 $3(9 \ 6 \ 1)$ 。将上述结果列于表 1, A~D 类模型参见文献^[9]。表 1 中 *E* 为 2 对主基因+ 多基因时包括主基因与多基因的加性、显性和上位性的全部遗传效应的完全模型; $E-1 \sim E-6$ 的多基因均为加性-显性模型; $E-1$ 的主基因为完全模型; $E-2$ 的主基因为加性-显性模型; $E-3$ 的主基因为加性模型; $E-4$ 的主基因为等加性模型; $E-5$ 的主基因为加性-完全显性模型; $E-6$ 的主基因为等加性-等显性(完全显性)模型。综合起来, 本文考虑的遗传

模型共有 *A*、*B*、*C*、*D*、*E* 5 类 24 种。至于 3 对主基因模型和 3 对主基因+ 多基因模型等更复杂的情况还有待于进一步研究。

表 1 两对主基因+ 多基因模型下 6 个世代所包含的成分分布数及可估的遗传参数

Table 1 The number of component distributions and estimatable genetic parameters of *P*₁, *F*₁, *P*₂, *B*₁, *B*₂ and *F*₂ under two major genes plus polygene mixed inheritance models

模型代号 Code	成分分布数 No. component distribution	独立参数个数 No. independent parameters	一阶遗传参数 1st order genetic parameter	二阶分布参数 2nd parameter	约束条件数 No. restricted conditions
<i>E</i>	20	18	<i>m</i> , <i>d_a</i> , <i>d_b</i> , <i>h_a</i> , <i>h_b</i> , <i>i</i> , <i>j_{ab}</i> , <i>j_{ba}</i> , <i>l</i> , [<i>d</i>], [<i>h</i>], [<i>i</i>], [<i>j</i>], [<i>l</i>]	$\alpha_1^2, \alpha_2^2, \alpha_3^2, \sigma^2$	6
<i>E-1</i>	20	15	<i>m</i> , <i>d_a</i> , <i>d_b</i> , <i>h_a</i> , <i>h_b</i> , <i>i</i> , <i>j_{ab}</i> , <i>j_{ba}</i> , <i>l</i> , [<i>d</i>], [<i>h</i>]	$\alpha_1^2, \alpha_2^2, \alpha_3^2, \sigma^2$	9
<i>E-2</i>	20	11	<i>m</i> , <i>d_a</i> , <i>d_b</i> , <i>h_a</i> , <i>h_b</i> , [<i>d</i>], [<i>h</i>]	$\alpha_1^2, \alpha_2^2, \alpha_3^2, \sigma^2$	13
<i>E-3</i>	20	9	<i>m</i> , <i>d_a</i> , <i>d_b</i> , (<i>h_a</i> = <i>h_b</i> = 0), [<i>d</i>], [<i>h</i>]	$\alpha_1^2, \alpha_2^2, \alpha_3^2, \sigma^2$	15
<i>E-4</i>	14	8	<i>m</i> , <i>d</i> (= <i>d_a</i> = <i>d_b</i> , <i>h_a</i> = <i>h_b</i> = 0), [<i>d</i>], [<i>h</i>]	$\alpha_1^2, \alpha_2^2, \alpha_3^2, \sigma^2$	10
<i>E-5</i>	12	9	<i>m</i> , <i>d_a</i> (= <i>h_a</i>), <i>d_b</i> (= <i>h_b</i>), [<i>d</i>], [<i>h</i>]	$\alpha_1^2, \alpha_2^2, \alpha_3^2, \sigma^2$	7
<i>E-6</i>	10	8	<i>m</i> , <i>d</i> (= <i>d_a</i> = <i>d_b</i> = <i>h_a</i> = <i>h_b</i>), [<i>d</i>], [<i>h</i>]	$\alpha_1^2, \alpha_2^2, \alpha_3^2, \sigma^2$	6

1.2 6 世代联合似然函数和成分分布参数极大似然估计的 IECM 算法

利用 *P*₁、*F*₁、*P*₂、*B*₁、*B*₂ 和 *F*₂ 世代联合分析的样本似然函数为:

$$f(Y|\Theta) = \prod_{i=1}^{n_1} f(x_{1i}; \mu_1, \sigma^2) \prod_{i=1}^{n_2} f(x_{2i}; \mu_2, \sigma^2) \prod_{i=1}^{n_3} f(x_{3i}; \mu_3, \sigma^2) \prod_{i=1}^{n_4} \prod_{t=1}^{k_1} \Pi_{4it} f(x_{4i}; \mu_{4it}, \alpha_i^2) \prod_{i=1}^{n_5} \prod_{t=1}^{k_2} \Pi_{5it} f(x_{5i}; \mu_{5it}, \alpha_i^2) \prod_{i=1}^{n_6} \prod_{t=1}^{k_3} \Pi_{6it} f(x_{6i}; \mu_{6it}, \alpha_i^2) \quad (1)$$

其中, $f(x_{ji}; \mu, \sigma^2)$ 是正态分布 $N(\mu, \sigma^2)$ 的密度函数, k_1 、 k_2 和 k_3 分别是 *B*₁、*B*₂ 和 *F*₂ 群体的成分分布数。采用 IECM 算法获得(1)式中分布参数的极大似然估计值。IECM 算法采用迭代方法在条件 EM 算法(EM^[11, 14])基础上获得极大似然估计的方法。在多世代联合估计的 IECM 算法中, *E* 步骤的完全数据对数似然函数的期望函数为:

$$L_c(Y|\Theta) = \sum_{i=1}^{n_1} \lg f(x_{1i}; \mu_1, \sigma^2) + \sum_{i=1}^{n_2} \lg f(x_{2i}; \mu_2, \sigma^2) + \sum_{i=1}^{n_3} \lg f(x_{3i}; \mu_3, \sigma^2) + \sum_{i=1}^{n_4} \sum_{t=1}^{k_1} w_{4it} \lg f(x_{4i}; \mu_{4it}, \alpha_i^2) + \sum_{i=1}^{n_5} \sum_{t=1}^{k_2} w_{5it} \lg f(x_{5i}; \mu_{5it}, \alpha_i^2) + \sum_{i=1}^{n_6} \sum_{t=1}^{k_3} w_{6it} \lg f(x_{6i}; \mu_{6it}, \alpha_i^2)$$

其中, $w_{4it} \sim w_{4ik_1}$, $w_{5it} \sim w_{5ik_2}$, $w_{6it} \sim w_{6ik_3}$ 表示样本的后验概率。迭代 CM 步骤是分步骤地计算 $L_c(Y|\Theta)$ 的条件极大值和极大值点。分布平均数间无约束条件时可直接分步骤地对 $L_c(Y|\Theta)$ 求偏导数并令其等于 0 以得到新一轮参数估计值; 分布平均数间有约束条件时可利用 Lagrange 函数分步骤地确定 CM 步骤中 $L_c(Y|\Theta)$ 的条件极值。其具体过程为:

根据样本观测值选择一组分布参数初始值。

计算混合群体中样本观测值的后验概率 $w_{4it} \sim w_{4ik_1}$, $w_{5it} \sim w_{5ik_2}$, $w_{6it} \sim w_{6ik_3}$, 从而得到完全数据的似然函数值 $L_c(Y|\Theta)$ (*E* 步骤)。

分 2~ 3 步对 $L_c(Y|\Theta)$ 求条件极值, 用迭代方法得到分布平均数、多基因方差组分和环境方差的估计 (CM 步骤)。以模型 D、E 为例, 迭代 CM_1 步骤是在固定多基因方差组分和环境方差的条件下用迭代方法求分布平均数的条件极大似然估计; 迭代 CM_2 步骤是在固定环境方差和迭代 CM_1 步骤所获得的分布平均数条件下用迭代方法求多基因方差组分的条件极大似然估计; 迭代 CM_3 步骤是在固定迭代 CM_1 和 CM_2 步骤所获得的两组估计值条件下用迭代方法求环境方差的条件极大似然估计。模型 A、B、C 比较简单, 只分两步进行。

将得到的估计值作为初始值重复进行 和 步骤, 直到达到预定的精度为止。

为从 A ~ E 5 类共 24 种模型中找出最适合于所得试验数据的模型, 要按上述方法分别估计出这 5 类共 24 种模型的成分分布参数。

1.3 遗传模型的判定及其相应遗传参数的估计

从 5 类 24 种模型中判定最佳遗传模型的方法按 Gai and Wang (1998)^[9] 的方法分两步进行。第一步根据最大熵准则, 从最小 AIC 值找出最佳的 1 个或几个模型; 第二步对初选模型进行适合性检验, 以找出最适模型。 AIC 值的计算、判定方法以及入选模型的适合性检验包括 U_1^2 、 U_2^2 、 U_3^2 、Smirnov 的 nW^2 和 Kolmogorov 的 D_n 统计量检验方法均同 Gai and Wang (1998)^[9]。选定最佳遗传模型后, 使用该模型下的各成分分布参数估计值估算相应的遗传参数。以 E-6 模型为例说明分布平均数与一阶遗传参数的关系:

$$\begin{aligned} \mu_1 &= m + d + [d]; \quad \mu_2 = m + h + [h]; \quad \mu_3 = m - d - [d]; \quad \mu_4 = m + 2d + (1/2)[d] + (1/2)[h] \\ \mu_{51} &= m + 2d - (1/2)[d] + (1/2)[h]; \quad \mu_{52} = m - (1/2)[d] + (1/2)[h] \\ \mu_{53} &= m - 2d - (1/2)[d] + (1/2)[h]; \quad \mu_{61} = m + 2d + (1/2)[h] \\ \mu_{62} &= m + (1/2)[h]; \quad \mu_{63} = m - 2d + (1/2)[h] \end{aligned}$$

由此, 采用最小二乘法可估计出 m 、 d 、 $[d]$ 和 $[h]$ 。其它模型同理可得相应遗传参数估计值。

各世代表型方差 σ_p^2 由试验数据直接算得, 极大似然估计得到的 σ^2 为环境方差; α_a^2 、 α_b^2 和 α_c^2 分别为 B_1 、 B_2 和 F_2 各成分分布方差, 由多基因方差组分和环境方差构成; 群体的多基因方差 $\sigma_{pg}^2 = \sigma_p^2 - \alpha_a^2 - \alpha_b^2 - \alpha_c^2$, 主基因方差 α_{ng}^2 分别按下列公式求得:

$$\alpha_{ng(B_1)}^2 = \frac{1}{4}(d_a - h_a)^2 + \frac{1}{4}(d_b - h_b)^2 + \frac{1}{16}[(i - j_{ab})^2 + (i - j_{ba})^2 + (i - l)^2 + (j_{ab} - j_{ba})^2 + (j_{ab} - l)^2 + (j_{ba} - l)^2] + \frac{1}{4}(d_a - h_a) \quad (2a)$$

$$(i + j_{ab} - j_{ba} - l) + \frac{1}{4}(d_b - h_b)(i - j_{ab} + j_{ba} - l)$$

$$\alpha_{ng(B_2)}^2 = \frac{1}{4}(d_a + h_a)^2 + \frac{1}{4}(d_b + h_b)^2 + \frac{1}{16}[(i + j_{ab})^2 + (i + j_{ba})^2 + (i - l)^2 + (j_{ab} - j_{ba})^2 + (j_{ab} + l)^2 + (j_{ba} + l)^2] + \frac{1}{4}(d_a + h_a) \quad (2b)$$

$$(-i + j_{ab} - j_{ba} + l) + \frac{1}{4}(d_b + h_b)(-i - j_{ab} + j_{ba} + l)$$

$$\alpha_{ng(F_2)}^2 = \frac{1}{4}[d_a^2 + d_b^2 + i^2 + (d_a + j_{ab})^2 + (d_b + j_{ba})^2 + (h_a + \frac{1}{2}l)^2 + (h_b + \frac{1}{2}l)^2 + \frac{1}{4}l^2] \quad (2c)$$

主基因遗传率 $h_{mg}^2 = \alpha_{ng}^2/\sigma_p^2$ 和多基因遗传率 $h_{pg}^2 = \sigma_{pg}^2/\sigma_p^2$ 。D-1~ D-4 和 E-1~ E-6 模型还能把



B₁、B₂ 和 F₂ 的多基因方差组分 σ_{a0}^2 、 σ_{s0}^2 和 σ_{e0}^2 剖分为多基因加性效应方差 *D*、显性效应方差 *H* 以及 *F*:

$$D = 4\sigma_{e0}^2 - 2\sigma_{a0}^2 - 2\sigma_{s0}^2 \tag{3a}$$

$$H = 4(\sigma_{a0}^2 + \sigma_{s0}^2 - \sigma_{e0}^2) \tag{3b}$$

$$F = \sigma_{s0}^2 - \sigma_{a0}^2 \tag{3c}$$

由此, 可联合估计 B₁、B₂ 和 F₂ 群体的多基因遗传率:

$$h_{pg(B_1)}^2 = (D/4 + H/4 - F/2)/\sigma_P^2(B_1) \tag{4a}$$

$$h_{pg(B_2)}^2 = (D/4 + H/4 + F/2)/\sigma_P^2(B_2) \tag{4b}$$

$$h_{pg(F_2)}^2 = (D/2 + H/4)/\sigma_P^2(F_2) \tag{4c}$$

1.4 Bayes 后验概率与分离世代个体主基因型的归属

将分布参数估计值代入 *E* 步骤便可得到 Bayes 后验概率 $w_{4il} \sim w_{4ik_1}$, $w_{5il} \sim w_{5ik_2}$ 和 $w_{6il} \sim w_{6ik_3}$, B₁、B₂ 和 F₂ 群体单株的后验概率可表示为:

$$w_{mit} = \prod_{i=1}^{k_{m-3}} \text{Pr}(x_{mi}; \mu_{mi}, \sigma_m^2) / \prod_{r=1}^{k_{m-3}} \text{Pr}(x_{mi}; \mu_{mr}, \sigma_r^2) \quad (m = 4, 5, 6) \tag{5}$$

在 *A*、*D* 模型时, 主基因型数不多, 个体主基因型的判别较简单, 但在 *B*、*E* 模型时主基因型数增多, F₂ 代将有 9 种主基因型, 此时个体主基因型的后验概率判别弹性增大, 实际意义比 *A*、*D* 模型时小得多。

以上 *A* ~ *E* 5 类共 24 个遗传模型的分析计算 (*IECM* 算法) 均用 Turbo C++ 语言编写成程序名为 SN.EXE 的计算程序软件, 欢迎读者来函联系。

2 应用实例

Gai and Wang (1998)^[9] 文中南京 6 号 × 广丛杂交组合的 6 世代的数据在此作进一步分析。其次数分布数据

参见该文表 3。按本文 2.2 节所获得的 5 类 24 种模型的极大对数似然函数值和 *AIC* 值列于表 2。*E*-2 模型具有最小 *AIC* 值, 为最佳可能模型, 与之相近的还有 *D*-2、*D*-1、*D*、*E*-1 等。这与 Gai and Wang (1998)^[9] 的结果有所不同也有所相近, 说明可能属 2 对

表 2 用 *IECM* 算法估计各种遗传模型的极大对数似然函数值和 *AIC* 值
Table 2 The maximum log likelihood values and *AIC* values under various genetic models estimated through the iterated ECM (*IECM*) algorithm

模型 Model	极大对数似然值 Maximum log-likelihood	<i>AIC</i>	模型 Model	极大对数似然值 Maximum log-likelihood	<i>AIC</i>
A-1	- 3257.94	6523.87	<i>D</i>	- 3185.95	6395.90
A-2	- 3460.48	6926.97	<i>D</i> -1	- 3185.30	6388.60
A-3	- 3300.69	6607.37	<i>D</i> -2	- 3185.30	6386.60
A-4	- 3631.70	7269.40	<i>D</i> -3	- 3264.98	6545.97
B-1	- 3224.81	6469.63	<i>D</i> -4	- 3422.24	6860.48
B-2	- 3225.94	6463.88	<i>E</i>	- 3185.93	6407.87
B-3	- 3381.16	6770.32	<i>E</i> -1	- 3183.80	6397.60
B-4	- 3422.23	6864.46	<i>E</i> -2	- 3172.06	6366.12
B-5	- 3444.06	6902.12	<i>E</i> -3	- 3226.75	6471.50
B-6	- 3389.06	6784.12	<i>E</i> -4	- 3254.51	6525.03
<i>C</i>	- 3287.47	6582.93	<i>E</i> -5	- 3195.23	6408.46
<i>C</i> -1	- 3290.65	6587.30	<i>E</i> -6	- 3337.26	6690.51

主基因+ 多基因的混合遗传模型。但因与 1 对主基因+ 多基因模型的差异甚小, 可能第 2 对主基因的作用较小。经一组适合性检验的结果 (表 3), *E*-2 不仅最佳而且最适, 因而确认该组

合水稻株高为 2 对主基因+ 多基因的遗传。模型的扩展使分析不局限于 D ，更确切地反应了实际数据的性质。

表 3 E-2 模型的适合性检验

Table 3 Tests for goodness of fit of model E-2

模型	U_1^2	U_2^2	U_3^2	nV^2	D_n
P_1	0.48(0.49)	0.20(0.65)	0.82(0.37)	0.196(> 0.05)	0.143(> 0.05)
F_1	0.20(0.65)	0.61(0.44)	1.89(0.17)	0.199(> 0.05)	0.178(> 0.05)
P_2	0.20(0.89)	0.16(0.69)	1.06(0.30)	0.118(> 0.05)	0.116(> 0.05)
B_1	0.00(0.96)	0.03(0.88)	0.19(0.66)	0.088(> 0.05)	0.100(> 0.05)
B_2	0.10(0.76)	0.03(0.87)	0.29(0.59)	0.040(> 0.05)	0.068(> 0.05)
F_2	0.00(0.99)	0.00(1.00)	0.00(0.97)	0.049(> 0.05)	0.037(> 0.05)

E-2 模型时全部 1 阶、2 阶分布参数估计值列于表 4，由表 4 计算出 1 阶、2 阶遗传参数估计值列于表 5。将此结果与 Gai and Wang (1998)^[9] 的 D 模型结果相比，第 1 对主

基因的加性和显性效应、显性度(27.11、21.28、0.785)与 D 模型的 1 对主基因相近或略小些(29.15、24.12、0.83)。E-2 模型中增加了第 2 对主基因，其加性效应只有第 1 对主基因的 1/3，且显性不明显。多基因的效应虽然不大，但 E-2 模型和 D 模型一方面多基因效应的成分不同，且加性和显性效应有明显差别。E-2 模型主基因效应在 B_1 、 B_2 和 F_2 3 世代为 70.76 ~ 94.12(%)，多基因遗传率为 3.01 ~ 8.07(%)，比在 D 模型下较一致。总之，水稻株高性状主基因是起决定作用的，通过将模型扩展到 E ，增加了对第 2 对主基因的了解。原来只有 D 模型时，这第 2 对主基因的效应是混杂在多基因的效应中。

表 4 E-2 模型参数的极大似然估计值

Table 4 The maximum likelihood estimates in model E-2

参数 Parameter	估计值 Estimate	参数 Parameter	估计值 Estimate	参数 Parameter	估计值 Estimate	参数 Parameter	估计值 Estimate
μ_1	162.54	μ_{44}	146.76	μ_{62}	156.49	μ_{68}	102.26
μ_2	148.49	μ_{51}	154.56	μ_{63}	147.95	μ_{69}	93.73
μ_3	103.70	μ_{52}	146.03	μ_{64}	162.35	σ^2	19.23
μ_{41}	164.27	μ_{53}	106.17	μ_{65}	150.66	σ_1^2	22.69
μ_{42}	152.58	μ_{54}	97.63	μ_{66}	142.12	σ_2^2	44.08
μ_{43}	158.45	μ_{61}	168.18	μ_{67}	113.95	σ_6^2	48.36

3 讨论

通过 A-1 ~ D-4 模型共 17 个模型比较本文 IECM 算法与文献^[9]的 EM 算法，IECM 算法的极大对数似然函数值比相应 EM 算法的结果要大 300 ~ 600，IECM 算法的 AIC 值比相应的结果要小 600 ~ 1300。因此，IECM 算法的结果要比 EM 算法的结果更好，说明 IECM 算法

表 5 水稻南京 6 号(P_1) × 广丛(P_2) 株高的有关遗传参数估计值

Table 5 The estimates of genetic parameters of plant height of the rice cross between nanjing No. 6 and guangcong

1 阶参数 1st order parameter	估计值 Estimate	2 阶参数 2nd order parameter	估计值 Estimate		
			B_1	B_2	F_2
m	133.12	σ^2	53.11	853.26	563.86
d_a	27.11	σ_{mg}^2	42.65	603.78	532.55
d_b	10.11	σ_g^2	3.46	24.85	29.13
h_a	21.28	σ_e^2	19.23	19.23	19.23
h_b	- 1.58	h_{mg}^2 (%)	80.31	70.76	94.45
h_a/d_a	0.785	h_{pg}^2 (%)	8.07	3.01	5.31
h_b/d_b	- 0.156				
$[d]$	- 7.81				
$[h]$	- 4.34				

更易收敛。同时, 本文在估计环境方差时不像文献^[9]只用 P_1 、 F_1 、 P_2 3 个不分离世代, 将每一成分分布均做分解, 故应用了 6 个世代的有关信息, 这也是降低 AIC 值的一个原因。

分布平均数间的约束条件数是分布平均数减去一阶遗传参数, E 模型的分布平均数间有 6 个约束条件; D 模型有 2 个约束条件: $g_1 = \mu_{41} - \mu_{42} - \mu_{61} + \mu_{62} = 0$ 和 $g_2 = \mu_{51} - \mu_{52} - \mu_{62} + \mu_{63} = 0$, 这就补充了文献^[9]中 D 模型的分布平均数间无约束条件的不足。

比较 Gai and Wang (1998)^[9]的结果和本文结果, 在 A 、 B 、 C 和 D 类模型中只能发现遗传效应最大的主基因, 在拓展 E 类模型后则可发现遗传效应最大和次大的 2 个主基因。因此, 拓展 2 对主基因+ 多基因的混合遗传模型是必要的。比较利用个别世代的分离分析结果和利用多个世代的联合分析结果, 利用回交世代群体只能发现遗传效应最大的 1 对主基因和遗传效应很小的 1 对主基因, 这可能是由于利用个别世代的信息较少或环境方差的估计误差较大的缘故。因此, 有必要拓展利用环境误差较小的家系世代的重复试验的分析方法, 这将另文报道。此外, 在完成了 E 模型的基础上自然会考虑 3 对主基因+ 多基因模型。鉴于这种情况的推导更加复杂, 有待于今后的工作。但从孟德尔主基因的一般情况来看, 最常见还是 1、2 对主基因的实验, 更多的也已很少应用了。

参 考 文 献

- 1 盖钧镒, 管荣展, 王建康 世界科技研究与发展, 1999, 21(1): 34~ 40
- 2 马育华 植物育种的量遗传学基础 南京: 江苏科技出版社 1982
- 3 Elston R C, J Steward *Genetics*, 1973, 73: 695~ 711
- 4 莫惠栋 作物学报, 1993, 19 (1): 1~ 6
- 5 姜长鉴, 莫惠栋 作物学报, 1995, 21(6): 641~ 648
- 6 王建康, 盖钧镒 遗传学报, 1997, 24 (5): 432~ 440
- 7 盖钧镒, 王建康 作物学报, 1998, 24 (4): 402~ 409
- 8 王建康, 盖钧镒 作物学报, 1998, 24 (6): 651~ 659
- 9 Gai J, J Wang *Theor Appl Genet*, 1998, 97: 1162~ 1168
- 10 王建康, 盖钧镒 生物数学学报, 1995, 10 (3): 87~ 92
- 11 Dempster A P, N M Laird, D B Rubin *J. R. Statist B.*, 1977, 39: 1~ 38
- 12 王建康, 盖钧镒 生物数学学报, 1997, 12 (5): 540~ 548
- 13 Akaike H. In: P R Krishnaiah (ed) *A Application of Statistics*, Amsterdam: North-Holland Publishing Co. 1977. 27~ 41
- 14 Meng X L, D B Rubin *Biometrika*, 1993, 80 (2): 267~ 278