

主基因-多基因混合遗传数量性状的 单性状选择模型

罗凤娟,董晓萌,袁志发,郭满才

(西北农林科技大学 理学院,陕西 杨凌 712100)

【摘 要】 【目的】 分别估计分离世代数量性状主基因-多基因混合遗传中的多基因效应方差和环境方差,为育种工作者提供一种简单易行的单性状选择方法。【方法】 利用正态分布的极限误差,将混合正态分布数据划分到多个单一正态分布对应的区间,每一区间代表一种主基因型,采用有重复观测数据的单因素多元方差分析方法,分别估计各主基因型间的主基因效应协方差阵、多基因效应协方差阵和环境协方差阵,然后结合混合正态分布中各成分分布的比例,计算该数量性状的表型方差、遗传方差、环境方差和遗传力等遗传参数,并进行实例分析。【结果】 在相同选择强度下,遗传效率为 110.376 0%,较作为单一正态分布的普通数量性状选择效率提高了 10.376 0%。当选择强度 $i=0.5$ 时,遗传增益 $GS=28.41$ 。【结论】 应用建立的主基因-多基因混合遗传数量性状单性状选择模型计算,遗传效率有很大提高,表明该方法具有实用性。

【关键词】 数量性状;主基因-多基因混合遗传;AIC 准则;EM 算法;遗传增益;单性状选择模型

【中图分类号】 Q348;O29 【文献标识码】 A 【文章编号】 1671-9387(2008)09-0190-07

Single character selective model of major gene and polygene mixed inheritance model of quantitative trait

LUO Feng-juan, DONG Xiao-meng, YUAN Zhi-fa, GUO Man-cai

(College of Science, Northwest A&F University, Yangling, Shaanxi 712100, China)

Abstract: 【Objective】 The study was done to estimate the variance of polygene effect and environmental variance separately in the major gene and polygene mixed inheritance model of quantitative traits from segregating generation and provide a simple single character selective method for breeding operators. 【Method】 This paper used normal distribution's limit error to plot out the data of mixed normal distribution into the multiple sections which correspond single normal distribution, each section representing one major gene genotype, and the one-way analysis of variance of repeated data was applied to estimate the covariance matrix of major gene effect among major gene genotypes, covariance matrix of polygene effect and environmental covariance matrix, then the genetic parameters, namely the phenotypic variance, genetic variance and heritability of the quantitative trait, were calculated combining the proportion of the component distribution of the mixed normal distribution and an example was taken to analyse. 【Result】 Under the same selection intensity, the genetic efficiency was 110.376 0%, it was improved by 10.376 0% compared with the common quantitative trait which corresponds single normal distribution. The genetic gain GS was 28.41 when the selection intensity was 0.5. 【Conclusion】 Using the single character selective model of major gene and polygene mixed inheritance model of

* [收稿日期] 2007-09-23
[基金项目] 国家自然科学基金项目“空间协方差结构混合模型分析田间试验效果与技术”(30571072)
[作者简介] 罗凤娟(1982—),女,陕西宝鸡人,在读硕士,主要从事生物数学研究。E-mail: juanfengl@sina.com
[通讯作者] 郭满才(1963—),男,陕西宝鸡人,教授,博士,主要从事农业应用数学研究。

quantitative trait to calculate,the selection efficiency was higher,showing practicability of this method.

Key words:quantitative character;major gene and polygene mixed inheritance;AIC;EM;genetic gain;single character selective model

动植物的许多重要经济性状,在分离世代往往既具有可分组趋势,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微效基因控制,称之为主基因-多基因混合遗传体系,如植物中大豆的生育期、水稻的株高、小麦和大麦对叶锈病的抗性等,动物中影响羊排卵率的 Booroola 基因、家禽中的矮基因等。在分离世代中,由微效多基因控制的数量性状的表现型近似服从正态分布,而主基因-多基因混合遗传体系控制的性状表型呈现多峰性,表现出多个分布混合的特征。对于此类性状的研究,莫惠栋等^[1-2]、姜长鉴等^[3]、盖钧镒等^[4]和王建康等^[5-7]分别建立了无主基因存在和主基因存在时的主基因-多基因混合遗传数学模型,并对相应的参数进行了估计。

在分离世代数量性状主基因-多基因混合遗传模型分析中,前人的研究主要是基于无重复的完全随机样本,只能估计主基因效应方差 σ_{mg}^2 、多基因效应方差和环境方差之和 σ^2 ,若要进一步估计多基因效应方差 $\sigma_{pg}^2 = \sigma^2 - \sigma_e^2$,则必须加入不分离世代 P_1 、 P_2 和 F_1 代估计环境方差 σ_e^2 。此外,在此类性状的遗传分析中,前人的研究主要是对遗传参数、遗传效应的估计,以及对不同动植物的不同数量性状所属的遗传模型进行判别与分析,而未涉及选择问题,也并未提出分离世代主基因-多基因混合遗传的选择模型。针对以上两个问题,本研究采用等重复观测样本试验设计方案,在单个分离世代中能分别估计多基因效应方差和环境方差,并根据直接遗传进展建立了主基因-多基因混合遗传的单性状选择模型,以期 of 育种选择提供一种简便易行的选择方法。

1 主基因-多基因混合遗传模型简介

假设主基因-多基因混合遗传所控制的性状 $\mathbf{X}$ 有 $x_1, x_2, \dots, x_k$ 共 $k(k > 1)$ 个主基因型,表型向量记为 $\mathbf{P} = [x_1, x_2, \dots, x_k]^T$,其遗传模型为^[7]:

$$\mathbf{P} = m + g + e = m + t + c + e。$$

式中: m 为中亲值向量; g 为遗传效应向量,其协方差阵为 Σ_g ; e 为环境效应向量,且 $e \sim N_k(0, \Sigma_e)$,其中 Σ_e 为环境协方差阵; t 为主基因效应向量,其协方差阵为 Σ_t ; c 为多基因效应向量,且 $c \sim N_k(0, \Sigma_c)$,其中 Σ_c 为多基因效应协方差阵; t, c, e 之间相互独立,有 $\Sigma_p = \Sigma_g + \Sigma_e = \Sigma_t + \Sigma_c + \Sigma_e$,其中 Σ_p 为表型协方差阵。

为了分别估计主基因效应协方差阵、多基因效应协方差阵和环境协方差阵,本研究采用有重复的单因素多元方差分析,将主基因-多基因混合遗传的 k 个主基因型作为 k 个处理因素进行协方差阵分析。因此,应将实际观测数据进行预处理,并对样本个体的主基因型进行划分,将每种主基因型作为单因素方差分析的一个处理因素。

2 实际样本数据的预处理及样本个体主基因型的划分

2.1 原样本数据的预处理

在实际观测中,设观测了包括所有主基因型的样本个体数为 n 的全程样本,且每个样本重复观测 r 次,因此实际观测数据是 r 行 n 列的包括所有主基因型的全样本信息。将实际观测数据从小到大排序,经过整理,结果如表 1 所示。

表 1 实际观测的数据表

Table 1 Data of practical observation

重复观测次数 Repeated measure times	样本个数 Sample number				
	1	...	i	...	n
1	x_{11}	...	x_{i1}	...	x_{n1}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
j	x_{1j}	...	x_{ij}	...	x_{nj}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
r	x_{1r}	...	x_{ir}	...	x_{nr}
平均值 Average	$\bar{x}_1$	...	$\bar{x}_i$	...	$\bar{x}_n$

注: $x_{ij}(i=1,2,\dots,n;j=1,2,\dots,r)$ 表示第 i 个个体的第 j 次重复观测值; $\bar{x}_i(i=1,2,\dots,n)$ 表示第 i 个个体的平均观测值。
Note: $x_{ij}(i=1,2,\dots,n;j=1,2,\dots,r)$ is the j times repeated measure value of the i individual; $\bar{x}_i(i=1,2,\dots,n)$ is the average observation value of the individual.

该数量性状是主基因-多基因混合遗传,其表型观测值呈现多峰性,表现为多个正态分布的混合,故用混合分布理论来估计参数进行主基因型的划分。

2.2 混合分布参数的估计

混合分布中所包含成分分布的个数 k 利用 Akaike 提出的最大熵信息准则(AIC)^[4]来估计,混合分布中其他参数,即各成分分布的均值 μ_i 、权重 α_i 和方差 σ_i^2 ($i=1,2,\cdots,k$),采用 EM 算法^[8-10]估计。

因为样本的每组重复观测值间存在一定误差,同一个体的不同重复观测值可能划分到不同的区间,即属于不同的主基因型,这与实际情况不符,故参数的估计用其均值 $\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_n$, 即一组全样本信息进行计算。

2.3 观测样本个体主基因型的划分

假设测量值 $x \sim N(\mu, \sigma^2)$, 则由正态分布的极限误差^[11]可知, x 落在区间 $[\mu - 3\sigma, \mu + 3\sigma]$ 外的概率非常小, 因此由此极限误差划分各观测个体所属的区间, 即对应的主基因型类别, 信息损失较小。

设混合分布的成分分布个数为 k , 即有 k 种主基因型; 各成分分布的均值和方差分别为 μ_i 和 σ_i^2 ,

其中 $i=1,2,\cdots,k$, 则各成分分布的均值 μ_i ($i=1,2,\cdots,k$) 的 3σ 区间为 $[\mu_i - 3\sigma_i, \mu_i + 3\sigma_i]$ 。因此, 可将各观测个体的均值 $\bar{x}_i$ ($i=1,2,\cdots,n$) 按 $[\mu_i - 3\sigma_i, \mu_i + 3\sigma_i]$ ($i=1,2,\cdots,k$) 分别划分到相应的 k 个区间中, 进而将其对应的每组观测值 x_{ij} ($i=1,2,\cdots,n; j=1,2,\cdots,r$) 划分到对应的区间中, 即相应的主基因型 $x_1, x_2, \cdots, x_k$ 中。若其不属于某区间, 则将观测个体在此区间对应的主基因型中的观测数据记为 0, 这样就得到表 2 数据形式。

例如: 若第 1 个样本到第 p ($p < n$) 个样本的均值在 $[\mu_1 - 3\sigma_1, \mu_1 + 3\sigma_1]$ 区间, 则判定这些样本属于混合分布中的第 1 个成分分布, 其主基因型为 x_1 , 则其余样本的观测数在主基因型 x_1 中记为 0, 可认为这些样本在基因型为 x_1 时观测不到; 其他样本个体主基因型的划分与此类似。

对每个个体的重复观测值进行主基因型划分后, 实际观测数据可整理为表 2 的形式。这种观测设计可按重复中有重复的单因素完全随机设计进行离差阵分析。

表 2 等重复的单因素随机试验数据

Table 2 Data of the single element random experiment with equal repetition number

主基因型 Major gene genotype	样本个数 Sample number					和 Sum	平均 Average
	1	...	j	...	n		
x_1	x_{111}	...	x_{1j1}	...	x_{1n1}	$T_{1..}$	$\bar{x}_{1..}$
	x_{112}	...	x_{1j2}	...	x_{1n2}		
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		
	x_{11r}	...	x_{1jr}	...	x_{1nr}		
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	x_{i11}	...	x_{ij1}	...	x_{in1}	$T_{i..}$	$\bar{x}_{i..}$
	x_{i12}	...	x_{ij2}	...	x_{in2}		
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		
	x_{i1r}	...	x_{ijr}	...	x_{inr}		
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	x_{k11}	...	x_{kj1}	...	x_{kn1}	$T_{k..}$	$\bar{x}_{k..}$
	x_{k12}	...	x_{kj2}	...	x_{kn2}		
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		
	x_{k1r}	...	x_{kjr}	...	x_{knr}		
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
总平均 General average	$\bar{x}_{..1}$	...	$\bar{x}_{..j}$	...	$\bar{x}_{..n}$	$T_{...}$	$\bar{x}_{...}$

注: x_{ijt} ($i=1,2,\cdots,k; j=1,2,\cdots,n; t=1,2,\cdots,r$) 为第 j 个观测个体的第 t 次重复观测值的第 i 个主基因型值。
Note: x_{ijt} ($i=1,2,\cdots,k; j=1,2,\cdots,n; t=1,2,\cdots,r$) is the i major gene genotype value of the t times repeated measure value of the j individual.

3 主基因和多基因存在的检验

3.1 主基因存在的检验

对于主基因-多基因混合遗传的数量性状来说, 根据其表型混合分布中成分分布个数 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。但具体的主基因座位点之间的相互作用, 如加性效应、完全显性、部分显性等还需进行进一步的似然比假设检验, 才能作出相应推断, 本文不加以详述。

3.2 多基因存在的检验

如果主基因存在,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、环境、或是两者共同作用的结果,即应对多基因的存在进行鉴定。假设混合分布的成分分布个数为 k ,观测样本个数为 n ,则样本的似然函数为:

$$L(\varphi) = \prod_{i=1}^n p(x_i; \varphi) = \prod_{i=1}^n \sum_{j=1}^k \pi_j f_j(x_i; \mu_j, \sigma_j^2).$$

构造零假设 $H_0: \sigma_j^2 = \sigma_e^2$ (即不存在多基因),其中 $j=1, 2, \dots, k$; 备择假设 $H_a: \sigma_j^2 > \sigma_e^2$ (即有多基因存在),其中 $j=1, 2, \dots, k$, 且 σ_j^2 互不相等。通过计算两种假设下似然函数最大值 L_0 和 L_a , 构造出似然比统计量:

$$\lambda = 2[\log L_a - \log L_0] \sim \chi^2(k-1).$$

式中: $L_a = \prod_{i=1}^n \sum_{j=1}^k \pi_j f_j(x_i; \mu_j, \sigma_j^2)$, $L_0 = \prod_{i=1}^n \sum_{j=1}^k \pi_j f_j(x_i; \mu_j, \sigma_e^2)$, 进而对上述假设进行显著性检验。若 $\lambda > \chi_a^2(k-1)$, 则拒绝原假设, 说明在显著水平 α 下有多基因存在, 即除主基因之外的变异是多基因和环境共同作用的结果; 否则, 接受原假设, 在显著水平 α 下无多基因存在, 即除主基因之外的变异只是环境作用的结果。

4 主基因-多基因混合遗传的方差分析

4.1 方差分析模型的建立

设 x_{ijt} 是主基因型为 x_i 的第 j 个观测个体的第 t 次重复观测值, 主基因型 x_i 的均值为 μ_i , 第 j 个观测值的主效应为 β_{ij} , x_{ij} 的第 t 次重复观测的随机误差为 ϵ_{ijt} , 则方差分析的统计模型为^[12-13]:

$$\begin{aligned} x_{ijt} &= \mu_i + \beta_{ij} + \epsilon_{ijt}, \\ i &= 1, 2, \dots, k; \\ j &= 1, 2, \dots, n; \\ t &= 1, 2, \dots, r. \end{aligned}$$

主基因型 $x_1, x_2, \dots, x_k$ 的理论平均值分别为 $\mu_1, \mu_2, \dots, \mu_k$, 其均值 $\mu = \frac{1}{k} \sum_{i=1}^k \mu_i$ 。把离差 $\alpha_i = \mu_i - \mu$ ($i=1, 2, \dots, k$) 称为主基因型 x_i 的主效应, 则有:

$$\begin{aligned} x_{ijt} &= \mu + \alpha_i + \beta_{ij} + \epsilon_{ijt}, \\ i &= 1, 2, \dots, k; j = 1, 2, \dots, n; t = 1, 2, \dots, r. \end{aligned}$$

在这个模型中, α_i 间相互独立, 且均服从 $N_k(0, \Sigma_t)$, 其中 Σ_t 为主基因型间的主基因效应协方差阵; β_{ij} 相互独立, 且均服从 $N_k(0, \Sigma_c)$, 其中 Σ_c 为多基因效应协方差阵; ϵ_{ijt} 为随机误差, 且均服从 $N_k(0, \Sigma_e)$,

其中 Σ_e 为环境协方差阵。

以上参数的最小二乘估计分别为:

$$\begin{aligned} \hat{\mu} &= \bar{x}_{...}, \hat{\mu}_{ij} = \bar{x}_{ij}, \hat{\mu}_i = \bar{x}_{i..}, \\ \hat{\alpha}_i &= \bar{x}_{i..} - \bar{x}_{...}, \hat{\beta}_{ij} = \bar{x}_{ij} - \bar{x}_{i..}, \hat{\epsilon}_{ijt} = x_{ijt} - \bar{x}_{ij}. \end{aligned}$$

4.2 协方差阵的分解

对于每个观测值有重复的总离差阵 W , 可分解为主基因型间的离差阵即主基因效应离差阵 L_t 、主基因型内的离差阵即多基因效应离差阵 L_c 和随机误差的离差阵即环境离差阵 L_e 3 部分^[12-13]。

主基因型间的离差阵为:

$$L_t = n \times r \times \begin{bmatrix} l'_{11} & l'_{12} & \cdots & l'_{1k} \\ l'_{21} & l'_{22} & \cdots & l'_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ l'_{k1} & l'_{k2} & \cdots & l'_{kk} \end{bmatrix},$$

$$\text{其中} \begin{cases} l'_{ii} = (\bar{x}_{i..} - \bar{x}_{...})^2 \\ l'_{ij} = l'_{ji} = (\bar{x}_{i..} - \bar{x}_{...})(\bar{x}_{j..} - \bar{x}_{...}). \\ i, j = 1, 2, \dots, k, i \neq j \end{cases}$$

主基因型内的离差阵即多基因效应离差阵为:

$$L_c = r \times \begin{bmatrix} l^c_{11} & l^c_{12} & \cdots & l^c_{1k} \\ l^c_{21} & l^c_{22} & \cdots & l^c_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ l^c_{k1} & l^c_{k2} & \cdots & l^c_{kk} \end{bmatrix},$$

$$\text{其中} \begin{cases} l^c_{ii} = \sum_{j=1}^n (\bar{x}_{ij} - \bar{x}_{i..})^2 \\ l^c_{ip} = l^c_{pi} = \sum_{j=1}^n (\bar{x}_{ij} - \bar{x}_{i..})(\bar{x}_{pj} - \bar{x}_{p..}). \\ i, p = 1, 2, \dots, k, i \neq p \end{cases}$$

随机误差的离差阵即环境离差阵为:

$$L_e = \begin{bmatrix} l^e_{11} & l^e_{12} & \cdots & l^e_{1k} \\ l^e_{21} & l^e_{22} & \cdots & l^e_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ l^e_{k1} & l^e_{k2} & \cdots & l^e_{kk} \end{bmatrix},$$

$$\text{其中} \begin{cases} l^e_{ii} = \sum_{j=1}^n \sum_{t=1}^r (x_{ijt} - \bar{x}_{ij}.)^2 \\ l^e_{ip} = l^e_{pi} = \sum_{j=1}^n \sum_{t=1}^r (x_{ijt} - \bar{x}_{ij}.)(x_{pjt} - \bar{x}_{pj}.). \\ i, p = 1, 2, \dots, k, i \neq p \end{cases}$$

L_t, L_c 和 L_e 的自由度分别为:

$$f_{L_t} = k-1, f_{L_c} = k(n-1), f_{L_e} = kn(r-1);$$

L_t, L_c 和 L_e 的均方离差阵为:

$$SP_t = \frac{L_t}{k-1},$$

$$SP_c = \frac{L_c}{k(n-1)},$$

$$SP_e=\frac{L_e}{kn(r-1)};$$

$$ESP_c=\boldsymbol{\Sigma}_e+r\boldsymbol{\Sigma}_c,$$

$$ESP_e=\boldsymbol{\Sigma}_e。$$

各均方的期值望分别为：

$$ESP_t=\boldsymbol{\Sigma}_e+r\boldsymbol{\Sigma}_c+nr\boldsymbol{\Sigma}_t,$$

离差阵分析模式见表 3。

表 3 主基因型间离差阵的分析模式

Table 3 Covariance matrix analysis among the major gene genotypes

变异来源 Source	自由度 <i>df</i>	离差阵 <i>L</i>	均方离差阵 <i>SP</i>	期望均方离差阵 <i>ESP</i>
主基因型间 Among major gene genotypes	<i>k</i> −1	<i>L_t</i>	<i>SP_t</i>	<i>Σ_e</i> + <i>rΣ_c</i> + <i>nrΣ_t</i>
主基因型内 Within major gene genotypes	<i>k</i> (<i>n</i> −1)	<i>L_e</i>	<i>SP_e</i>	<i>Σ_e</i> + <i>rΣ_c</i>
误差 Error	<i>kn</i> (<i>r</i> −1)	<i>L_e</i>	<i>SP_e</i>	<i>Σ_e</i>
总变异 Total	<i>knr</i> −1	<i>W</i>		

当原假设 $H_{01}:\boldsymbol{\Sigma}_t=0,H_{02}:\boldsymbol{\Sigma}_c=0$ 成立时, L_e,L_c 与 L_t 相互独立,且 $L_e\sim W_k[kn(r-1),\boldsymbol{\Sigma}_e],f_e=kn(r-1);L_c\sim W_k[k(n-1),\boldsymbol{\Sigma}_e],f_c=k(n-1);L_t\sim W_k(k-1,\boldsymbol{\Sigma}_e),f_t=k-1$ 。

由 Wilks 统计量可知^[13]：

$$\Lambda_t=\frac{|L_e|}{|L_e+L_t|}\sim\Lambda(k,f_e,f_t),$$

$$\Lambda_c=\frac{|L_e|}{|L_e+L_c|}\sim\Lambda(k,f_e,f_c),$$

且根据 Bartlett 的研究结果,有

$$V_t=-[f_e+f_t-(k+f_t+1)/2]\ln\Lambda_t\sim\chi^2(kf_t),$$

$$V_c=-[f_e+f_c-(k+f_c+1)/2]\ln\Lambda_c\sim\chi^2(kf_c)。$$

由样本计算 V_t 和 V_c ,若 $V_t\leq\chi^2_{\alpha}(kf_t)$ 且 $V_c\leq\chi^2_{\alpha}(kf_c)$,则接受 H_{01} 和 H_{02} ,否则拒绝原假设。当 V_t 和 V_c 均显著时,可估计参数 $\hat{\boldsymbol{\Sigma}}_e,\hat{\boldsymbol{\Sigma}}_c$ 和 $\hat{\boldsymbol{\Sigma}}_t$,即：

$$\hat{\boldsymbol{\Sigma}}_e=SP_e,$$

$$\hat{\boldsymbol{\Sigma}}_c=\frac{1}{r}(SP_c-SP_e),$$

$$\hat{\boldsymbol{\Sigma}}_t=\frac{1}{nr}(SP_t-SP_e)。$$

5 主基因-多基因混合遗传体系的单性状选择模型

根据文献[8-10]的方法,计算出混合分布中成分分布个数为 k ,每个成分分布的权重分别为 $a_1,a_2,\cdots,a_k$,且 $\sum_{i=1}^ka_i=1$,则该数量性状 $\mathbf{X}$ 由 k 种主基因型控制,分别记为 $x_1,x_2,\cdots,x_k$,并有

$$\mathbf{X}=a_1x_1+a_2x_2+\cdots+a_kx_k=\mathbf{a}^T\mathbf{P}。$$

式中： $\mathbf{a}=[a_1,a_2,\cdots,a_k]^T,\mathbf{P}=[x_1,x_2,\cdots,x_k]^T$ 。

因此, $\mathbf{X}$ 的表型方差为：

$$\boldsymbol{\sigma}_{Xp}^2=[a_1,a_2,\cdots,a_k]\boldsymbol{\Sigma}_p$$

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix}=\mathbf{a}^T\boldsymbol{\Sigma}_p\mathbf{a},\text{其中 }\boldsymbol{\Sigma}_p$$

为各主基因型间的表型协方差阵。

主基因遗传方差为 $\boldsymbol{\sigma}_{Xt}^2=\mathbf{a}^T\boldsymbol{\Sigma}_t\mathbf{a}$;多基因遗传方差为 $\boldsymbol{\sigma}_{Xc}^2=\mathbf{a}^T\boldsymbol{\Sigma}_c\mathbf{a}$;环境方差为 $\boldsymbol{\sigma}_{Xe}^2=\mathbf{a}^T\boldsymbol{\Sigma}_e\mathbf{a}$;遗传方差为 $\boldsymbol{\sigma}_{Xg}^2=\mathbf{a}^T\boldsymbol{\Sigma}_ga=\mathbf{a}^T(\boldsymbol{\Sigma}_t+\boldsymbol{\Sigma}_c)\mathbf{a}$;数量性状的表型方差为 $\boldsymbol{\sigma}_{Xp}^2=\mathbf{a}^T\boldsymbol{\Sigma}_pa=\mathbf{a}^T(\boldsymbol{\Sigma}_g+\boldsymbol{\Sigma}_e)\mathbf{a}=\mathbf{a}^T(\boldsymbol{\Sigma}_t+\boldsymbol{\Sigma}_c+\boldsymbol{\Sigma}_e)\mathbf{a}。$

主基因遗传力为 $h_t^2=\frac{\boldsymbol{\sigma}_{Xt}^2}{\boldsymbol{\sigma}_{Xp}^2}$;多基因遗传力为 $h_c^2=\frac{\boldsymbol{\sigma}_{Xc}^2}{\boldsymbol{\sigma}_{Xp}^2}$;主基因与多基因遗传力即广义遗传力为： $h^2=\frac{\boldsymbol{\sigma}_{Xg}^2}{\boldsymbol{\sigma}_{Xp}^2}$ ；环境遗传力为： $h_e^2=\frac{\boldsymbol{\sigma}_{Xe}^2}{\boldsymbol{\sigma}_{Xp}^2}。$

遗传增益 GS 为： $GS=i\sigma_ph^2$,其中 i 为选择强度, σ_p 为表型方差, h^2 为遗传力^[14]。对主基因-多基因混合遗传控制的数量性状,其选择模型可用 GS 表示为：

$$GS=i\sigma_ph^2=i\sigma_{Xp}h_{(t+c)}^2。$$

在同一选择强度下,遗传效率

$$\eta=\frac{i\sigma_{Xp}h_{(t+c)}^2}{i\sigma_ph_c^2}。$$

式中： σ_{Xp} 是主基因-多基因混合遗传时的表型方差, σ_p 是将此数量性状作为普通数量性状,即表型值服从单一正态分布时的方差。

由以上分析,对于主基因-多基因混合遗传的数量性状的不同个体群,在一定选择强度下,将其遗传增益按大小顺序排序,选择遗传增益较大者即可。

6 实例分析

根据盖钧镒等^[4]对骨绿豆×上海红芒早 F_2 世代不同个体开花期的观测及计算结果,用 Matlab 中的 normrnd 随机产生两组数据,每组数据各有 200 个,均包含两个具有不同均值、相同方差的正态分布,两个正态分布的均值分别为 $\mu_1=3.18,\mu_2=39.5$,方差为 $\sigma_1^2=\sigma_2^2=5.35$,且混合分布中两个成分

分布比例为 1 : 3,对其产生的随机数据从小到大排列,频数分布如表 4 所示。

表 4 骨绿豆×上海红芒早 F₂ 世代开花期的频数分布

Table 4 Frequency distribution of the days from planting to flowering in F₂ generation derived from Guludou × Shanghai hongmangzao

开花期/d Day from planting to flowering	频数分布 Frequency distribution		
	第 1 次重复 First repetition	第 2 次重复 Second repetition	平均值 Average
26.5~27.5	4	4	4
27.5~28.5	1	1	1
28.5~29.5	4	3	4
29.5~30.5	5	7	5
30.5~31.5	5	6	6
31.5~32.5	12	10	11
32.5~33.5	8	9	8
33.5~34.5	4	5	6
34.5~35.5	8	6	6
35.5~36.5	6	7	7
36.5~37.5	18	18	17
37.5~38.5	21	22	22
38.5~39.5	26	27	26
39.5~40.5	27	26	27
40.5~41.5	23	19	22
41.5~42.5	11	10	11
42.5~43.5	10	12	10
43.5~44.5	4	5	4
44.5~45.5	3	3	3

从骨绿豆×上海红芒早 F₂ 世代开花期的频数分布可以看出,第 1 次重复、第 2 次重复和平均值均表现为两个成分分布的混合。根据平均值的频数分布,用 DPS 数据处理软件进行计算,在不同成分分布个数下,对数似然函数的极大值和 AIC 值如表 5。

从表 5 可以看出,当成分分布个数 $k=2$ 时,AIC 值达到最小,由此可确定该分离群体包含两个成分分布。成分分布个数确定后,可通过分离分析,根据两次重复观测的平均值确定各个成分分布的数字特征,计算结果见表 6。

表 5 不同成分分布个数时的 AIC 值

Table 5 Akaike's information values under different component numbers

成分分布个数(k) Component number (k)	独立参数的个数($N(k)$) Independent parameters ($N(k)$)	对数似然函数极大值($L(\hat{\varphi})$) Maximum of log likelihood function ($L(\hat{\varphi})$)	AIC 值 AIC value
1	2	-487.725 9	979.451 9
2	5	-482.364 4	974.728 7
3	8	-483.926 1	983.852 1
4	11	-481.106 2	984.212 3
5	14	-475.262 4	978.524 7

表 6 骨绿豆×上海红芒早 F₂ 群体混合分布中的参数估计

Table 6 Parameters' estimation of the mixed distribution in F₂ generation derived from Guludou × Shanghaihongmangzao

重复 Repetition	均值 Average		标准差 Standard variance	
	成分分布 1 First component distribution	成分分布 2 Second component distribution	成分分布 1 First component distribution	成分分布 2 Second component distribution
第 1 次重复 First repetition	31.032 7	39.220 9	2.154 1	3.067 4
第 2 次重复 Second repetition	30.951 8	39.326 6	2.052 9	3.076 1
平均值 Average	30.944 8	39.285 0	2.077 9	3.014 0

由于成分分布个数 $k=2$,且成分分布所占的比例接近 1 : 3,说明存在一个主基因位点,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、或是两者共同作用的结果,应对多基因的存在进行检验。

假设多基因存在时,似然函数最大值为 $L_a=-479.102\ 6$,无多基因存在时,似然函数最大值为 $L_0=-480.799\ 4$,似然比统计量为 $\lambda=2[\log L_a-\log L_0]=3.393\ 6>\chi^2_{0.1}(1)=2.706$,因此可认为,在置信度为 0.10 时,似然比统计量显著,即多基因存在。

通过以上分析,按本文介绍的方法,计算可得:主基因型间的离差阵为:

$$\mathbf{L}_t=10^4\times\begin{bmatrix}6.132 & -6.132 \\ -6.132 & 6.132\end{bmatrix};$$

多基因和环境的离差阵分别为:

$$\mathbf{L}_c=10^4\times\begin{bmatrix}9.867 & -3.201 \\ -3.201 & 3.749\end{bmatrix},$$

$$\mathbf{L}_e=\begin{bmatrix}1.841 & 1.566 \\ 1.566 & 3.557\end{bmatrix};$$

主基因型间的均方离差阵为:

$$\mathbf{SP}_t=10^4\times\begin{bmatrix}6.132 & -6.132 \\ -6.132 & 6.132\end{bmatrix};$$

多基因和环境的均方离差阵分别为:

$$\mathbf{SP}_c=\begin{bmatrix}247.921 & -80.419 \\ -80.419 & 94.198\end{bmatrix},$$

$$\mathbf{SP}_e=\begin{bmatrix}0.005 & 0.004 \\ 0.004 & 0.009\end{bmatrix}。$$

统计量 $V_t=4\ 691.2$, $V_c=10\ 108$, V_t , V_c 极显著,因而有

环境协方差阵为:

$$\hat{\mathbf{\Sigma}}_e=\begin{bmatrix}0.005 & 0.004 \\ 0.004 & 0.009\end{bmatrix},$$

多基因协方差阵为:

$$\hat{\mathbf{\Sigma}}_c=\begin{bmatrix}123.953 & -40.212 \\ -40.212 & 47.094\end{bmatrix},$$

主基因型间的协方差阵为:

$$\hat{\mathbf{\Sigma}}_t=\begin{bmatrix}152.684 & -153.103 \\ -153.103 & 153.068\end{bmatrix}。$$

各成分分布的权重向量为 $\mathbf{a}=\left[\frac{1}{4},\frac{3}{4}\right]^T$,即该数量性状只存在 1 个主基因位点,设为 A 和 a,且表现为完全显性,则 $aa:(Aa+AA)=1:3$,那么主基

因遗传方差为 $\sigma^2_{X_t}=\mathbf{a}^T\mathbf{\Sigma}_t\mathbf{a}=38.23$,多基因遗传方差为 $\sigma^2_{X_c}=\mathbf{a}^T\mathbf{\Sigma}_c\mathbf{a}=19.158$,环境方差为 $\sigma^2_{X_e}=\mathbf{a}^T\mathbf{\Sigma}_e\mathbf{a}=0.007$,遗传方差为 $\sigma^2_{X_g}=\mathbf{a}^T\mathbf{\Sigma}_g\mathbf{a}=\mathbf{a}^T(\mathbf{\Sigma}_t+\mathbf{\Sigma}_c)\mathbf{a}=57.389$,表型方差为 $\sigma^2_{X_p}=\mathbf{a}^T\mathbf{\Sigma}_p\mathbf{a}=\mathbf{a}^T(\mathbf{\Sigma}_t+\mathbf{\Sigma}_c+\mathbf{\Sigma}_e)\mathbf{a}=57.395$,广义遗传力为 $h^2=\frac{\sigma^2_{X_g}}{\sigma^2_{X_p}}=99\%$,其中主基因

遗传力为 $h^2_t=\frac{\sigma^2_{X_t}}{\sigma^2_{X_p}}=66\%$,多基因遗传力为 $h^2_c=\frac{\sigma^2_{X_c}}{\sigma^2_{X_p}}=33\%$ 。

在同一选择强度下,遗传效率 $\eta=110.376\ 0\%$,较作为单一正态分布的普通数量性状的选择效率提高了 10.376 0%。

当选择强度 $i=0.5$ 时,遗传增益 $GS=28.41$ 。

对不同个体群进行该数量性状的选择时,可直接根据遗传进展的大小,选择遗传进展较大的群体即可。

7 结 论

在主基因-多基因混合遗传模型的基础上,本研究采用等重复观测样本试验设计方案,在单个分离世代中估计了各主基因型对应性状的主基因效应协方差阵、多基因效应协方差阵和环境协方差阵,进而估计了该数量性状的主基因效应方差、多基因效应方差和环境方差,解决了以往单个分离世代不能分别计算多基因遗传方差和环境方差的问题。此外,通过计算个体群的直接遗传进展进行单性状的选择,选择效率有很大提高,为育种者提供了方便。

[参考文献]

[1] 莫惠栋. 质量-数量性状的遗传分析Ⅰ. 遗传组成和主基因基因型鉴别 [J]. 作物学报,1993,19(1):1-6.
Mo H D. Genetic analysis for qualitative-quantitative traits I. The genetic constitution of generation populations and the identification of major gene genotypes [J]. Acta Agronomica Sinica,1993,19(1):1-6. (in Chinese)

[2] 莫惠栋,徐辰武. 质量-数量性状的遗传分析Ⅲ. 受三倍体遗传控制的胚乳性状 [J]. 作物学报,1994,20(5):513-519.
Mo H D,Xu C W. Genetic analysis for qualitative-quantitative traits Ⅲ. Endosperm characters under triploid genetic control [J]. Acta Agronomica Sinica,1994,20(5):513-519. (in Chinese)

[3] 姜长鉴,莫惠栋. 质量-数量性状的遗传分析Ⅳ. 极大似然法的应用 [J]. 作物学报,1995,21(6):641-648.
Jiang C J,Mo H D. Genetic analysis for qualitative-quantitative traits Ⅳ. Application of the maximum likelihood method [J]. Acta Agronomica Sinica,1995,21(6):641-648. (in Chinese)