



## 双交组合种子性状的遗传模型及蒙特卡罗模拟分析

许自成<sup>1</sup> 丁永乐<sup>1</sup> 朱军<sup>2</sup>

(1 河南农业大学 农学院, 河南 郑州 450002; 2 浙江大学 农业与生物技术学院, 浙江 杭州 310029)

**摘要:** 将 Cockerham 广义遗传模型的建模原理应用于双交方式的交配设计, 针对二倍体种子和三倍体胚乳数量性状的遗传特点, 提出了 2 个适用于分析种子品质性状的加性-显性-细胞质-母体模型, 介绍了相应的统计分析方法, 并进行了蒙特卡罗模拟分析.

**关键词:** 双交设计; 种子性状; 加性-显性-细胞质-母体模型; 遗传效应分量

**中图分类号:** S333

**MR 分类号:** 62J10; 92E05

**文献标识码:** A

**文章编号:** 1001-9626(2005)03-0339-06

### 0 引言

在作物杂交育种和杂种优势利用中, 经常采用复交组合方式创造丰富的遗传变异, 育成综合多个亲本优良性状的新品种. Rawlings & Cockerham<sup>[1,2]</sup> 曾提出三交和双交设计的遗传模型, 用来分析不同组配方式杂交种的遗传效应. 但是, 上述复交方式的交配设计需要配制大量的组合, 容易导致缺失的非平衡数据, 其应用研究并不多见. 另外, 这些模型主要用于农艺性状的遗传研究, 与大多数农艺性状的遗传规律不同, 关于种子及幼苗性状的遗传往往存在母体效应. 这是因为种子的营养物质由母体植株提供, 种子品质性状可能同时受到种子核基因、细胞质基因和母体植株核基因的控制<sup>[3]</sup>. 朱军<sup>[3,4]</sup>、朱军等<sup>[5]</sup>、Zhu & Weir<sup>[6,7]</sup> 等从单交组合方式的双列杂交出发, 提出了包括上述 3 套遗传体系的模型, 已在一些作物中得到初步应用. 就复交方式而言, 许自成和朱军已较系统地提出了三交和双交组合的 ADM 模型<sup>[8]</sup>、ADAA 模型<sup>[9,10]</sup> 以及三交方式的种子模型<sup>[11]</sup>. 本文在以上研究的基础上, 针对双交组合方式的遗传特点, 提出双交设计种子数量性状的遗传模型, 为农作物品质性状的遗传改良提供理论依据.

### 1 双交组合二倍体种子性状的遗传模型构建

双子叶植物如棉花、大豆、油菜的种子没有  $3n$  胚乳, 其营养品质如蛋白质、油分等服从  $2n$  种子性状的遗传规律. 因此, 二倍体种子性状的遗传模型应包括胚及子叶、细胞质、母体植株 3 套体系遗传模型的控制<sup>[3,6]</sup>. 设  $p$  个纯合亲本按双交方式进行交配设计, 田间试验采用

收稿日期: 2003-04-20

基金项目: 河南省杰出青年科学基金资助项目 (0112000900); 河南省杰出人才创新基金资助项目 (0421001900)

作者简介: 许自成 (1964-), 男, 河南汝南人, 教授, 博士.

随机区组排列. 在考虑细胞质和母体效应的情况下, 若假定不存在父体效应和上位性效应, 则任意 4 个独立的亲本  $i, j, s, t$  组配的双交组合  $F_{1ijst}$  的第  $k$  种交配类型, 在第  $l$  个区组中的平均表现型值可用线性模型表示为:

$$y_{hijstkl} = \mu + E_h + G_{ijstk} + GE_{hijstk} + B_{l(h)} + \varepsilon_{hijstkl}$$

这里  $\mu$  是群体平均数, 固定效应;  $E_h$  是环境效应, 可设为固定效应;  $G_{ijstk}$  和  $GE_{hijstk}$  分别是第  $k$  种交配类型世代平均数的遗传效应和基因型  $\times$  环境互作效应;  $B_{l(h)}$  是区组随机效应,  $B_l \sim (0, \sigma_B^2)$ ;  $\varepsilon_{hijstkl}$  是残差效应,  $\varepsilon_{hijstkl} \sim (0, \sigma_e^2)$ ; 不同世代的遗传效应分量可进行如下分解:

当  $k=0, i=j=s$  时, 亲本  $P_i$  的遗传效应分量为

$$\begin{aligned} G_{iiii0} &= 2A_i + D_{ii} + C_i + 2Am_i + Dm_{ii} \\ GE_{hiiii0} &= 2AE_{hi} + DE_{hii} + CE_{hi} + 2AmE_{hi} + DmE_{hii} \end{aligned}$$

当  $k=1, i=j, s=t$  时, 单交一代  $F_{1iz} = (P_i \times P_s)$  的遗传效应分量为

$$\begin{aligned} G_{iiss1} &= A_i + A_s + D_{ii} + C_i + 2Am_i + Dm_{ii} \\ GE_{hiiss1} &= AE_{hi} + AE_{hs} + DE_{his} + CE_{hi} + 2AmE_{hi} + DmE_{hii} \end{aligned}$$

当  $k=2$  时, 双交一代  $F_{1ijst} = F_{1ij} \times F_{1st}$  的遗传效应分量为

$$\begin{aligned} G_{ijst2} &= \frac{1}{2}A_i + \frac{1}{2}A_j + \frac{1}{2}A_s + \frac{1}{2}A_t + \frac{1}{4}D_{is} + \frac{1}{4}D_{js} + \frac{1}{4}D_{it} + \frac{1}{4}D_{jt} + \\ &\quad C_i + Am_i + Am_j + Dm_{ij} \\ GE_{hijst2} &= \frac{1}{2}AE_{hi} + \frac{1}{2}AE_{hj} + \frac{1}{2}AE_{hs} + \frac{1}{2}AE_{ht} + \frac{1}{4}DE_{his} + \frac{1}{4}DE_{hjs} + \\ &\quad \frac{1}{4}DE_{hit} + \frac{1}{4}DE_{hjt} + CE_{hi} + AmE_{hi} + AmE_{hj} + DmE_{hij} \end{aligned}$$

式中  $A_i, A_j, A_s$  或  $A_t$  是直接加性效应,  $A_i, A_j, A_s$  或  $A_t \sim (0, \sigma_A^2)$ ;  $D_{ii}, D_{is}, D_{js}, D_{it}$  或  $D_{jt}$  是直接显性效应,  $D_{ii}, D_{is}, D_{js}, D_{it}$  或  $D_{jt} \sim (0, \sigma_D^2)$ ;  $C_i$  是细胞质效应,  $C_i \sim (0, \sigma_C^2)$ ;  $Am_i$  或  $Am_j$  是母体加性效应,  $Am_i$  或  $Am_j \sim (0, \sigma_{Am}^2)$ ;  $Dm_{ii}$  或  $Dm_{ij}$  是母体显性效应,  $Dm_{ii}$  或  $Dm_{ij} \sim (0, \sigma_{Dm}^2)$ ;  $AE_{hi}, AE_{hj}, AE_{hs}$  或  $AE_{ht}$  是加性  $\times$  环境互作效应,  $AE_{hi}, AE_{hj}, AE_{hs}$  或  $AE_{ht} \sim (0, \sigma_{AE}^2)$ ;  $DE_{hii}, DE_{his}, DE_{hjs}, DE_{hit}$  或  $DE_{hjt}$  是显性  $\times$  环境互作效应,  $DE_{hii}, DE_{his}, DE_{hjs}, DE_{hit}$  或  $DE_{hjt} \sim (0, \sigma_{DE}^2)$ ;  $CE_{hi}$  是细胞质  $\times$  环境互作效应,  $CE_{hi} \sim (0, \sigma_{CE}^2)$ ;  $AmE_{hi}$  或  $AmE_{hj}$  是母体加性  $\times$  环境互作效应,  $AmE_{hi}$  或  $AmE_{hj} \sim (0, \sigma_{AmE}^2)$ ;  $DmE_{hii}$  或  $DmE_{hij}$  是母体显性  $\times$  环境互作效应,  $DmE_{hii}$  或  $DmE_{hij} \sim (0, \sigma_{DmE}^2)$ . 根据以上亲本、单交一代、双交一代 3 个世代的试验资料, 即可分析 3 套遗传体系, 估计各类基因效应的遗传方差分量, 同时还可对各类基因效应值进行预测.

## 2 双交组合三倍体胚乳性状的遗传模型构建

单子叶禾谷类作物如玉米、水稻、小麦的种子具有  $3n$  胚乳, 其营养品质如蛋白质含量、赖氨酸含量、糊化温度(稻米)等, 应采用不同的遗传模型来描述其遗传特征. 因此, 三倍

体胚乳性状的遗传模型至少应包括胚乳、细胞质、母体植株3套遗传体系。这里, 对各世代的遗传效应分量相应调整如下:

当  $k=0, i=j=s$  时, 亲本  $P_i$  的遗传效应分量为

$$\begin{aligned} G_{iiii0} &= 3A_i + 3D_{ii} + C_i + 2Am_i + Dm_{ii} \\ GE_{hiiii0} &= 3AE_{hi} + 3DE_{hii} + CE_{hi} + 2AmE_{hi} + DmE_{hii} \end{aligned}$$

当  $k=1, i=j, s=t$  时, 单交一代  $F_{1is}=(P_i \times P_s)$  的遗传效应分量为

$$\begin{aligned} G_{iiss1} &= 2A_i + A_s + 2D_{ii} + D_{is} + C_i + 2Am_i + Dm_{ii} \\ GE_{hiiis1} &= 2AE_{hi} + AE_{hs} + 2DE_{hii} + DE_{hii} + CE_{hi} + 2AmE_{hi} + DmE_{hii} \end{aligned}$$

当  $k=2$  时, 双交一代  $F_{1ijst}=F_{1ij} \times F_{1st}$  的遗传效应分量为

$$\begin{aligned} G_{ijst2} &= A_i + A_j + \frac{1}{2}A_s + \frac{1}{2}A_t + \frac{1}{2}D_{ii} + \frac{1}{2}D_{jj} + \frac{1}{2}D_{is} + \frac{1}{2}D_{jt} + \\ &\quad \frac{1}{2}D_{it} + \frac{1}{2}D_{jt} + C_i + Am_i + Am_j + Dm_{ij} \\ GE_{hijst2} &= AE_{hi} + AE_{hj} + \frac{1}{2}AE_{hs} + \frac{1}{2}AE_{ht} + \frac{1}{2}DE_{hii} + \frac{1}{2}DE_{hjj} + \\ &\quad \frac{1}{2}DE_{his} + \frac{1}{2}DE_{hjs} + \frac{1}{2}DE_{hit} + \frac{1}{2}DE_{hjt} + CE_{hi} + \\ &\quad AmE_{hi} + AmE_{hj} + DmE_{hij} \end{aligned}$$

### 3 统计分析方法

无论是双子叶植物二倍体种子还是单子叶植物三倍体胚乳性状, 其表型观察值均可用混合线性模型表示为<sup>[11]</sup>,

$$\begin{aligned} y &= Xb + U_A e_A + U_D e_D + U_C e_C + U_{Am} e_{Am} + U_{Dm} e_{Dm} + U_{AE} e_{AE} + \\ &\quad U_{DE} e_{DE} + U_{CE} e_{CE} + U_{AmE} e_{AmE} + U_{DmE} e_{DmE} + U_{BE} e_{BE} + e_c = \\ &\quad Xb + \sum_{u=1}^{12} U_u e_u \sim \\ &\quad (Xb, V = \sum_{u=1}^{12} \sigma_u^2 U_u U_u^T) \end{aligned}$$

其中  $b$  是固定效应向量,  $X$  是固定效应向量的系数矩阵;  $e_u$  是第  $u$  项随机效应向量,  $e_u \sim (0, \sigma_u^2 I)$ ;  $U_u$  是随机效应向量  $e_u$  的系数矩阵 ( $u=1, 2, \dots, 12$ ). 采用设先验协方差为 0、先验方差为 1 的 MINQUE(0/1) 法<sup>[3,6,7]</sup> 可对每一性状的方差以及成对性状的遗传协方差分量进行无偏估计; 采用 AUP 法可对各遗传效应进行无偏预测。

### 4 蒙特卡罗模拟分析

双交组合种子模型的模拟试验设计采用随机区组设计, 3 次重复。试验设计 1 包括 7 个亲本、21 个单交组合 ( $F_1$ ) 和 105 个双交组合 ( $F_1$ ), 共计 133 个遗传材料, 在单个环境下进

行试验; 试验设计 2 包括 6 个亲本、15 个单交组合 ( $F_1$ ) 和 45 个双交组合 ( $F_1$ ), 共计 66 个遗传材料, 在两个环境下进行试验. 应用 MINQUE(0/1) 法分别按二倍体种子和三倍体胚乳两种遗传模型进行 500 次蒙特卡罗模拟, 结果见表 1.

表 1 双交组合种子模型方差分量的模拟分析 (MINQUE(0/1) 法)

Table 1 Simulation Analysis of Variance Components for Seed Model Based on the Double Cross Mating Design (MINQUE(0/1) method)

参数	真值 $\sigma^2$	二倍体种子			三倍体胚乳		
		偏差	均方误差	效益系数	偏差	均方误差	效益系数
单个环境 ( $p = 7$ )							
$\sigma_A^2$	60	1.89	1844.40	0.69	2.98	2515.44	0.80
$\sigma_D^2$	50	-2.39	314.78	0.34	-1.35	242.98	0.30
$\sigma_C^2$	30	2.61	1300.13	1.11	2.47	1335.85	1.13
$\sigma_{Am}^2$	20	0.74	275.29	0.80	0.69	361.27	0.92
$\sigma_{Dm}^2$	30	-1.14	158.01	0.40	-1.14	161.46	0.41
$\sigma_{A.Am}$	20	1.33	367.28	0.90	0.99	443.79	1.01
$\sigma_{D.Dm}$	20	-0.07	139.55	0.59	0.09	128.44	0.56
$\sigma_\epsilon^2$	20	1.78	18.85	0.20	0.79	16.31	0.19
两个环境 ( $p = 6$ )							
$\sigma_A^2$	60	-0.44	2607.29	0.84	3.27	4166.17	1.02
$\sigma_D^2$	50	-0.85	734.91	0.53	-0.66	583.17	0.48
$\sigma_C^2$	30	2.71	2410.11	1.50	2.67	3159.50	1.72
$\sigma_{Am}^2$	20	1.85	589.39	1.11	1.89	788.59	1.28
$\sigma_{Dm}^2$	30	-0.41	460.39	0.71	-0.43	481.83	0.72
$\sigma_{AE}^2$	20	0.40	300.87	0.85	1.85	549.53	1.07
$\sigma_{DE}^2$	30	-1.40	210.72	0.46	-0.42	139.26	0.39
$\sigma_{CE}^2$	15	1.15	576.00	1.49	1.53	678.24	1.58
$\sigma_{AmE}^2$	15	-0.32	152.39	0.81	-0.12	249.12	1.04
$\sigma_{DmE}^2$	20	-0.37	126.75	0.55	-0.33	132.56	0.57
$\sigma_{A.Am}$	20	1.52	683.20	1.21	1.88	877.44	1.35
$\sigma_{D.Dm}$	20	0.96	354.63	0.90	0.86	334.55	0.88
$\sigma_{AE.AmE}$	15	0.31	134.12	0.76	1.27	194.71	0.86
$\sigma_{DE.DmE}$	15	0.10	105.26	0.68	0.07	92.46	0.64
$\sigma_\epsilon^2$	20	1.73	14.78	0.18	1.54	14.19	0.17

就试验设计 1 而言, 当不考虑基因型  $\times$  环境互作效应时, 二倍体种子性状的细胞质方差、直接加性与母体加性协方差和三倍体胚乳性状的细胞质方差的估计值偏差小于方差参数的 10%, 其余各项遗传方差分量估计值的偏差均小于方差参数的 5%. 双交组合三倍体胚乳性状各项方差分量均方误差的大小在总体上呈现  $A > C > A.Am > Am > D > Dm > D.Dm > \epsilon$  的趋势, 与相应三交组合的  $MSE$  位次完全吻合; 双交组合二倍体种子性状各项方差分量  $MSE$  的位次高低略有出入, 总体上呈现  $A > C > A.Am > D > Am > Dm > D.Dm > \epsilon$  的趋势. 从效益系数的估值来看, 各项方差分量均具有较高的估算效益. 以上分析表明基于亲本、单交  $F_1$  和三交  $F_1$  三个世代, 应用 MINQUE(0/1) 法可以无偏估计各项遗传方差分量.

就试验设计 2 而言, 当两个环境的资料联合分析时, 全部方 (协方) 差分量及其与环境互作的方 (协方) 差分量的估计值与方 (协方) 差参数接近, 估计值的偏差均小于参数的 10%,

再次验证了 MINQUE(0/1) 法估计遗传方(协方)差分量的有效性. 各项方差分量  $MSE$  的位次高低相比较, 二倍体种子性状的趋势是:  $A > C > D > A.Am > Am > CE > Dm > D.Dm > AE > DE > AmE > AE.AmE > DmE > DE.DmE > \epsilon$ ; 三倍体胚乳性状的趋势是:  $A > C > A.Am > Am > CE > D > AE > Dm > D.Dm > AmE > AE.AmE > DE > DmE > DE.DmE > \epsilon$ .

表2 双交组合种子模型遗传效应的模拟分析

Table 2 Simulation Analysis of Genetic Effects for Seed Model Based on the Double Cross Mating Design

参数	二倍体种子			三倍体胚乳		
	均值偏差 ( $\times 10^{-4}$ )	方差 $\hat{\sigma}_u^2$	平均距离	均值偏差 ( $\times 10^{-4}$ )	方差 $\hat{\sigma}_u^2$	平均距离
单个环境 ( $p = 7$ )						
<i>A</i>	0.22	61.89	1.49	-0.47	62.98	1.92
<i>D</i>	0.01	47.61	0.81	-0.13	48.96	0.65
<i>C</i>	0.07	32.61	1.85	-0.31	32.47	1.90
<i>Am</i>	-0.04	20.74	1.06	-0.33	20.69	1.20
<i>Dm</i>	-0.02	28.86	0.74	-0.10	28.86	0.74
$\epsilon$	0.00	21.73	0.12	0.00	20.79	0.12
两个环境 ( $p = 6$ )						
<i>A</i>	0.14	59.56	2.06	2.10	63.27	2.49
<i>D</i>	0.05	49.15	1.16	0.69	49.34	1.05
<i>C</i>	-0.04	32.71	2.29	0.83	32.67	2.34
<i>Am</i>	-0.23	21.85	1.53	1.20	21.89	1.63
<i>Dm</i>	0.02	29.59	1.05	0.21	29.57	1.06
<i>AE</i>	0.14	20.40	1.17	0.84	21.85	1.31
<i>DE</i>	0.00	28.60	0.75	0.33	29.58	0.68
<i>CE</i>	0.04	16.15	1.25	0.31	16.53	1.27
<i>AmE</i>	0.00	14.68	0.95	0.59	14.88	1.04
<i>DmE</i>	0.03	19.63	0.66	0.13	19.67	0.67
$\epsilon$	0.00	21.73	0.13	0.01	21.53	0.14

表2列出了采用 AUP 法预测种子性状遗传效应的分析结果. 无论是单个环境还是两个环境下, 均以加性效应的平均距离最大, 细胞质效应的平均距离次之, 剩余项的最小. 一般地, 各遗传效应的平均距离均相应大于其与环境互作效应的平均距离, 即  $A > AE, D > DE, C > CE, Am > AmE, Dm > DmE$ ; 各遗传效应值的均值偏差接近零值, 效应预测值方差接近方差参数, 平均距离较小, 表明采用 AUP 法预测遗传效应值具有可靠性.

## 5 结语

根据  $2n$  种子和  $3n$  胚乳性状的遗传特点, 本节分别构建了双交设计关于种子、细胞质和母体 3 套不同体系的遗传模型, MINQUE(0/1) 法被用来求解各项方差和协方差分量, AUP 法被用来预测各项遗传效应值. 对于一般的双列杂交设计, 即使不包括双交世代, 只要有亲本、单交一代、单交二代 3 个基本的遗传世代的种子性状, 也可根据单交组合各遗传效应分量的预测值对双交世代进行间接地预测. 若不考虑细胞质和母体效应, 则本文推导的 2 个遗传模

型均回复为简单的 AD 模型。

当不存在直接加性与母体加性、直接显性与母体显性协方差时,可采用 MINQUE(1) 法有效地分析不包括协方差分量的种子模型。以 7 个亲本、21 个单交  $F_1$  和 105 个双交  $F_1$  为例,删除协方差分量的种子模型的模拟结果表明,方差分量估计值的偏差均小于方差真值的 10%; 均方误差的大小在总体上呈现  $A > C > Am > D > Dm > \varepsilon$  的趋势未变; 估算效益较高。同时对均值偏差、AUP 预测值方差和平均距离三项指标的模拟结果进行比较,也证实了 AUP 法预测遗传效应的有效性。

### 参 考 文 献

- [1] Rawlings J O, Cockerham C C. Triallel analysis[J]. *Crop Science*, 1962, 2(3):228-231.
- [2] Rawlings J O, Cockerham C C. Analysis of double cross hybrid populations[J]. *Biometrics*, 1962, 18(3): 229-244.
- [3] 朱军. 遗传模型分析方法 [M]. 北京: 中国农业出版社, 1997, 12-55, 104-111, 175-201.
- [4] 朱军. 包括基因型  $\times$  环境互作效应的种子遗传模型及其分析方法 [J]. 遗传学报, 1996, 23(1): 56-68.
- [5] 朱军, 许韶华. 胚乳性状的遗传模型及其分析方法 [J]. 作物学报, 1994, 20(3):264-270.
- [6] Zhu, J, Weir B S. Analysis of cytoplasmic and maternal effects. I. A genetic model for diploid plant seeds and animals[J]. *Theor Appl Genet*, 1994, 89(2-3): 153-159.
- [7] Zhu, J, Weir B S. Analysis of cytoplasmic and maternal effects. II. Genetic models for triploid endosperm[J]. *Theor Appl Genet*, 1994, 89(2-3): 160-166.
- [8] 许自成, 朱军. 三交和双交组合加性-显性-母体效应的遗传模型及蒙特卡罗模拟分析 [J]. 作物学报, 2000, 26(1): 40-46.
- [9] 许自成, 朱军. 双交组合农艺性状的 ADAA 模型及其分析方法 [J]. 遗传学报, 2000, 27(3): 247-256.
- [10] 许自成, 朱军. 三交组合农艺性状的遗传模型及杂种优势预测方法 [J]. 生物数学学报, 2001, 16(2): 448-455.
- [11] 许自成, 朱军. 三交设计种子数量性状的遗传模型 [J]. 生物数学学报, 1998, 13(6):1015-1021.

## Genetic Models and Monte Carlo Simulation Analysis for Quantitative Seed Traits Based on the Double-way Mating Design

XU Zi-cheng<sup>1</sup> DING Yong-le<sup>1</sup> ZHU Jun<sup>2</sup>

(1 College of Agronomy, Henan Agricultural University, Zhengzhou Henan 450002 China)

(2 College of Agriculture and Biotechnology, Zhejiang University, Hangzhou Zhejiang 310029 China)

**Abstract:** According to a double-way mating design and using principles of Cockerham's general genetic model, two models, one for diploid seeds and the other for triploid endosperms were proposed for the analysis of quantitative traits of seeds influenced by genes with cytoplasmic effects, maternal effects, as well as direct additive and dominance effects. Mixed model approaches were introduced for analyzing these genetic models. Some questions on the extension of above genetic models and the prediction of heterosis with genotype  $\times$  environment interaction effects were discussed briefly.

**Key words:** Double-way mating design ; Quantitative characters of seeds ; Additive, dominance model with cytoplasmic and maternal effects ; Components of genetic effects