

包括基因型×环境互作效应的种子 遗传模型及其分析方法

朱 军

(浙江农业大学农学系 杭州 310029)

摘要 本文提出了分析作物种子数量性状世代平均数的遗传模型。该模型包括了基因型效应(G)和基因型×环境互作效应(GE)。基因型效应(G)可以分解为种子直接遗传效应(G_0)、细胞质遗传效应(G)和母体植株遗传效应(G_m)。种子直接遗传效应(G_0)进一步被分解为直接加性(A)和直接显性(D)遗传分量。母体植株遗传效应(G_m)则可被分解为母体加性(A_m)、母体显性(D_m)遗传分量。基因型×环境互作效应(GE)也可作相应的分解,由种子直接遗传效应与环境效应的互作(G_0E)、细胞质遗传效应与环境效应的互作(CE)和母体植株遗传效应与环境效应的互作(G_mE)所组成。 G_0E 可进一步被分解为直接加性×环境互作效应(AE)和直接显性×环境互作效应(DE)遗传分量。 G_mE 则可被分解为母体加性×环境互作效应(A_mE)、母体显性×环境互作效应(D_mE)遗传分量。列出了亲本及其 F_1 、 F_2 和回交世代的遗传分量表达式。只需采用一组随机亲本及其正反交 F_1 、 F_2 的世代平均数,便可用该模型对种子数量性状进行有效的分析。用MINQUE(0/1)法可以无偏地估算该模型中的各项遗传方差和协方差分量,以及成对性状的各项遗传协方差分量。对该模型中的随机遗传效应可用MINQUE(0/1)法作调整无偏预测(Adjusted Unbiased Prediction, AUP)。用Jackknife法可计算各估计值(或预测值)及其标准误,再用 t 测验对各参数作统计检验。运用蒙特卡罗模拟的方法,验证了遗传模型的稳定性和统计分析方法的无偏性。

关键词 种子性状, 遗传模型, 基因型×环境互作, 遗传方差和协方差估算, 遗传效应值预测

农作物种子是农业生产的主要产品。作物种子中储存有人类和动物所需要的营养物质,如蛋白质、必需氨基酸、油分等。目前作物的品质改良已成为作物育种的一项重要内容。研究作物种子中营养成分的遗传规律,对于了解种子储存物质的遗传动态、提高作物品质育种的效率具有实际指导意义。

生长在母体植株上的种子是由雌雄配子受精发育而成的新世代个体。母体植株对种子不仅贡献了部分遗传物质,还提供了大部或全部的营养物质。由于种子的营养物质由母体植株所提供,有些种子性状的表现可能会同时受到种子核基因和母体植株核基因两套遗传体系的控制。另外细胞质基因也可能通过控制叶绿体(或线粒体)的合成而影响植株的光合(或呼吸)作用,从而间接控制种子性状的表现。这一复杂的遗传体系给种子性

国家自然科学基金资助项目 本文于1994年6月4日收到。

状的遗传分析带来很大的困难。能否提出具有生物学意义的种子遗传模型和简易、实用的统计分析方法, 是决定种子数量性状遗传研究能否取得成效的关键所在。

莫惠栋提出了三倍体胚乳数量性状的遗传模型和统计分析方法^[1-3]。Bogyo 等也曾提出过类似的三倍体模型^[8]。这些遗传模型都没有包括植株母体效应对胚乳性状的作用。如果胚乳的数量性状完全决定于胚乳细胞中的三倍体基因, 利用三倍体模型可以分析基因的直接加性效应(Direct Additive Effects)和直接显性效应(Direct Dominance Effects)。当存在母体或细胞质遗传效应时, 这些胚乳模型便可能会导致有偏的估算。

作物种子数量性状的遗传分析表明: 无论是双子叶作物的大豆^[11]、油菜^[16]、棉花^[9, 14], 还是单子叶作物的玉米^[21]、大麦^[20]、小麦^[10], 这些作物的某些种子性状都不同程度地受到母体效应的影响。因此, 有必要发展包括母体效应或细胞质效应的遗传模型, 深入研究控制种子数量性状的遗传规律。

Topham 曾提出可以分析母体效应的双列分析方法^[19]。Beavis 等则提出了包括细胞质效应的遗传模型^[7]。近年来, 同时包括种子、细胞质和母体植株 3 套遗传体系基因效应的一些遗传模型已被提出^[4, 6, 12, 22, 23]。这些种子模型都假设不存在基因型 × 环境的互作。由于生物体的基因是在环境条件作用下表达的, 环境条件的不同可能导致基因表达方式或程度的差异。已有研究表明一些种子品质性状存在基因型 × 环境的互作^[15, 18]。如果互作确实存在, 需要在不同的环境条件下实施遗传实验, 并采用包括互作效应的遗传模型, 才能无偏地分析基因效应以及基因与环境的互作效应。

本文的目的是: 在已有的种子模型^[4, 6, 22, 23]基础上, 发展包括基因型 × 环境互作效应的种子遗传模型; 提出相应的统计分析方法, 估算遗传方差分量及遗传协方差分量, 预测遗传效应值; 运用蒙特卡罗模拟的方法, 验证遗传模型的稳定性和统计分析方法的无偏性。

1 遗传模型和统计分析方法

如果控制种子数量性状的基因在不同环境条件下表现不一致, 需要在多种环境条件下分析性状的遗传规律, 估算基因效应与环境效应的相互作用。某一环境下种子性状的表现型值主要由环境效应(E)、基因型效应(G)和基因型 × 环境互作效应(GE)所决定。基因型效应(G)可以分解为种子核基因直接遗传效应(G_0)、细胞质遗传效应(C)和母体植株核基因遗传效应(G_m)。种子直接遗传效应(G_0)可以进一步被分解为直接加性(A)和直接显性(D)遗传分量。母体植株遗传效应(G_m)则可被分解为母体加性(A_m)、母体显性(D_m)遗传分量。基因型 × 环境互作效应(GE)也可作相应的分解, 由种子直接遗传效应与环境效应的互作(G_0E)、细胞质遗传效应与环境效应的互作(CE)和母体植株遗传效应与环境效应的互作(G_mE)所组成。 G_0E 可进一步被分解为直接加性 × 环境互作效应(AE)和直接显性 × 环境互作效应(DE)遗传分量。 G_mE 则可被分解为母体加性 × 环境互作效应(A_mE)、母体显性 × 环境互作效应(D_mE)遗传分量。

假定(1)不存在父体效应(Paternal Effect); (2)细胞质基因只通过母体稳定遗传; (3)不存在上位性效应, 种子数量性状表现型值可用以下线性模型表示。一组纯系材料作双列杂交遗传交配, 第 i 个母本与第 j 个父本的第 k 种交配类型在环境 h 内第 l 个区组中的世代平均数值(y_{hijkl})可以表示为

$$Y_{hijk} = \mu + E_h + G_{ijk} + GE_{hijk} + B_{l(h)} + e_{hijk} \quad (1)$$

其中 μ 是群体平均数, 固定效应; E_h 是环境效应, 固定效应; $B_{l(h)}$ 是环境内区组效应, 随机效应; e_{hijk} 是剩余机误, 随机效应; G_{ijk} 是遗传效应, GE_{hijk} 是遗传效应与环境效应的互作。

种子营养物质, 在双子叶作物主要储存在二倍体的子叶中, 而在单子叶作物主要储存在三倍体的胚乳中。因此 G_{ijk} 和 GE_{hijk} 的遗传组成决定于作物类型、不同亲本 (i, j) 及交配类型 (k)。对于双子叶作物种子数量性状, 各遗传世代基因效应 G_{ijk} 和基因与环境互作效应 GE_{hijk} 的分解如下式所示。

亲本 P_i 遗传效应分量 ($k=0$)

$$G_{ii0} + GE_{hii0} = 2A_i + D_{ii} + C_i + 2A_{mi} + D_{mii} \\ + 2AE_{hi} + DE_{hii} + CE_{hi} + 2A_mE_{hi} + D_mE_{hii}$$

杂种一代 F_{1ij} ($P_i \times P_j$) 遗传效应分量 ($k=1$)

$$G_{ij1} + GE_{hij1} = A_i + A_j + D_{ij} + C_i + 2A_{mi} + D_{mii} \\ + AE_{hi} + AE_{hj} + DE_{hij} + CE_{hi} + 2A_mE_{hi} + D_mE_{hii}$$

杂种二代 F_{2ij} 遗传效应分量 ($k=2$)

$$G_{ij2} + GE_{hij2} = A_i + A_j + 0.25D_{ii} + 0.25D_{jj} + 0.5D_{ij} + C_i \\ + A_{mi} + A_{mj} + D_{mij} + AE_{hi} + AE_{hj} + 0.25DE_{hii} \\ + 0.25DE_{hjj} + 0.5DE_{hij} + CE_{hi} + A_mE_{hi} + A_mE_{hj} + D_mE_{hij}$$

回交一代 $F_{1ij} \times P_j$ 遗传效应分量 ($k=3$)

$$G_{ij3} + GE_{hij3} = 0.5A_i + 1.5A_j + 0.5D_{jj} + 0.5D_{ij} + C_i + A_{mi} + A_{mj} \\ + D_{mij} + 0.5AE_{hi} + 1.5AE_{hj} + 0.5DE_{hjj} + 0.5DE_{hij} \\ + CE_{hi} + A_mE_{hi} + A_mE_{hj} + D_mE_{hij}$$

回交一代 $F_{1ij} \times P_i$ 遗传效应分量 ($k=4$)

$$G_{ij4} + GE_{hij4} = 1.5A_i + 0.5A_j + 0.5D_{ii} + 0.5D_{ij} + C_i + A_{mi} + A_{mj} \\ + D_{mij} + 1.5AE_{hi} + 0.5AE_{hj} + 0.5DE_{hii} + 0.5DE_{hij} \\ + CE_{hi} + A_mE_{hi} + A_mE_{hj} + D_mE_{hij}$$

回交一代 $P_j \times F_{1ij}$ 遗传效应分量 ($k=5$)

$$G_{ij5} + GE_{hij5} = 0.5A_i + 1.5A_j + 0.5D_{jj} + 0.5D_{ij} + C_j + 2A_{mj} + D_{mjj} \\ + 0.5AE_{hi} + 1.5AE_{hj} + 0.5DE_{hjj} + 0.5DE_{hij} \\ + CE_{hj} + 2A_mE_{hj} + D_mE_{hjj}$$

回交一代 $P_i \times F_{1ij}$ 遗传效应分量 ($k=6$)

$$G_{ij6} + GE_{hij6} = 1.5A_i + 0.5A_j + 0.5D_{ii} + 0.5D_{ij} + C_i + 2A_{mi} + D_{mii}$$

$$+ 1.5AE_{hi} + 0.5AE_{hj} + 0.5DE_{hii} + 0.5DE_{hij}$$

$$+ CE_{hi} + 2A_m E_{hi} + D_m E_{hii}$$

而对于单子叶作物胚乳数量性状, 各遗传世代基因效应 G_{ijk} 和基因与环境互作效应 GE_{hijk} 的分解如下式所示。

亲本 P_i 遗传效应分量 ($k=0$)

$$G_{ii0} + GE_{hij0} = 3A_i + 3D_{ii} + C_i + 2A_{mi} + D_{mii}$$

$$+ 3AE_{hi} + 3DE_{hii} + CE_{hi} + 2A_m E_{hi} + D_m E_{hii}$$

杂种一代 $F_{1ij}(P_i \times P_j)$ 遗传效应分量 ($k=1$)

$$G_{ij1} + GE_{hij1} = 2A_i + A_j + D_{ii} + 2D_{ij} + C_i + 2A_{mi} + D_{mii}$$

$$+ 2AE_{hi} + AE_{hj} + DE_{hii} + 2DE_{hij} + CE_{hi} + 2A_m E_{hi} + D_m E_{hii}$$

杂种二代 F_{2ij} 遗传效应分量 ($k=2$)

$$G_{ij2} + GE_{hij2} = 1.5A_i + 1.5A_j + D_{ii} + D_{jj} + D_{ij} + C_i + A_{mi} + A_{mj} + D_{mij}$$

$$+ 1.5AE_{hi} + 1.5AE_{hj} + DE_{hii} + DE_{hjj} + DE_{hij}$$

$$+ CE_{hi} + A_m E_{hi} + A_m E_{hj} + D_m E_{hij}$$

回交一代 $F_{1ij} \times P_j$ 遗传效应分量 ($k=3$)

$$G_{ij3} + GE_{hij3} = A_i + 2A_j + 0.5D_{ii} + 1.5D_{jj} + D_{ij} + C_i + A_{mi} + A_{mj} + D_{mij}$$

$$+ AE_{hi} + 2AE_{hj} + 0.5DE_{hii} + 1.5DE_{hjj} + DE_{hij}$$

$$+ CE_{hi} + A_m E_{hi} + A_m E_{hj} + D_m E_{hij}$$

回交一代 $F_{1ij} \times P_i$ 遗传效应分量 ($k=4$)

$$G_{ij4} + GE_{hij4} = 2A_i + A_j + 1.5D_{ii} + 0.5D_{jj} + D_{ij} + C_i + A_{mi} + A_{mj} + D_{mij}$$

$$+ 2AE_{hi} + AE_{hj} + 1.5DE_{hii} + 0.5DE_{hjj} + DE_{hij}$$

$$+ CE_{hi} + A_m E_{hi} + A_m E_{hj} + D_m E_{hij}$$

回交一代 $P_j \times F_{1ij}$ 遗传效应分量 ($k=5$)

$$G_{ij5} + GE_{hij5} = 0.5A_i + 2.5A_j + 2D_{jj} + D_{ij} + C_j + 2A_{mj} + D_{mjj}$$

$$+ 0.5AE_{hi} + 2.5AE_{hj} + 2DE_{hjj} + DE_{hij}$$

$$+ CE_{hj} + 2A_m E_{hj} + D_m E_{hjj}$$

回交一代 $P_i \times F_{1ij}$ 遗传效应分量 ($k=6$)

$$G_{ij6} + GE_{hij6} = 2.5A_i + 0.5A_j + 2D_{ii} + D_{ij} + C_i + 2A_{mi} + D_{mii}$$

$$+ 2.5AE_{hi} + 0.5AE_{hj} + 2DE_{hii} + DE_{hij}$$

$$+ CE_{hi} + 2A_m E_{hi} + D_m E_{hii}$$

本文虽然列出了 7 个世代 ($k=0, 1, 2, \dots, 6$) 平均数的遗传效应分量的公式, 在实际分析作物种子数量性状时, 并不需要包括全套遗传材料。最简便的遗传交配设计是采用

一组亲本、正反交 F_1 及其自交后代 F_2 。采用这一交配设计只需在田间种植亲本和 F_1 植株,由其自交产生亲本种子和 F_2 种子,亲本间杂交可获得 F_1 种子。

如果杂交亲本是从一个推断群体中抽取的随机样本,则以上各项遗传效应分量均为随机效应。此时遗传模型可以进一步改写成矩阵形式的混合线性模型 (Mixed Linear Model),

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\mathbf{b} + \mathbf{U}_A\mathbf{e}_A + \mathbf{U}_D\mathbf{e}_D + \mathbf{U}_C\mathbf{e}_C + \mathbf{U}_{Am}\mathbf{e}_{Am} + \mathbf{U}_{Dm}\mathbf{e}_{Dm} + \mathbf{U}_{AE}\mathbf{e}_{AE} \\ &\quad + \mathbf{U}_{DE}\mathbf{e}_{DE} + \mathbf{U}_{CE}\mathbf{e}_{CE} + \mathbf{U}_{AmE}\mathbf{e}_{AmE} + \mathbf{U}_{DmE}\mathbf{e}_{DmE} + \mathbf{U}_B\mathbf{e}_B + \mathbf{e}_e \\ &= \mathbf{X}\mathbf{b} + \sum_{u=1}^{12} \mathbf{U}_u\mathbf{e}_u \end{aligned} \quad (2)$$

其中 \mathbf{b} 是固定效应向量,第一个参数是 μ ,其余参数是环境效应; \mathbf{X} 是固定效应的系数矩阵,第一列全为 1; \mathbf{e}_u 是第 u 项独立随机变量向量 ($u=1, 2, \dots, 11$),具有平均数零、方差 $\sigma_u^2\mathbf{I}$; \mathbf{U}_u 是第 u 项随机效应的系数矩阵 ($u=1, 2, \dots, 11$), $\mathbf{U}_{12}=\mathbf{I}$ 是单位矩阵。

由于部分种子核基因来自母体植株,因而直接遗传效应与母体遗传效应及其对应的交互效应并不是相互独立的。直接加性效应与母体加性效应、直接显性效应与母体显性效应可能会存在协方差。直接加性 \times 环境交互效应与母体加性 \times 环境交互效应、直接显性 \times 环境交互效应与母体显性 \times 环境交互效应也可能会存在协方差。种子数量性状表现型向量 \mathbf{y} 的方差-协方差矩阵为

$$\begin{aligned} \text{Var}(\mathbf{y}) &= \sigma_A^2\mathbf{U}_A\mathbf{U}_A' + \sigma_D^2\mathbf{U}_D\mathbf{U}_D' + \sigma_C^2\mathbf{U}_C\mathbf{U}_C' + \sigma_{Am}^2\mathbf{U}_{Am}\mathbf{U}_{Am}' + \sigma_{Dm}^2\mathbf{U}_{Dm}\mathbf{U}_{Dm}' \\ &\quad + \sigma_{AE}^2\mathbf{U}_{AE}\mathbf{U}_{AE}' + \sigma_{DE}^2\mathbf{U}_{DE}\mathbf{U}_{DE}' + \sigma_{CE}^2\mathbf{U}_{CE}\mathbf{U}_{CE}' + \sigma_{AmE}^2\mathbf{U}_{AmE}\mathbf{U}_{AmE}' \\ &\quad + \sigma_{DmE}^2\mathbf{U}_{DmE}\mathbf{U}_{DmE}' + \sigma_B^2\mathbf{U}_B\mathbf{U}_B' \\ &\quad + \sigma_{A,Am}(\mathbf{U}_A\mathbf{U}_{Am}' + \mathbf{U}_{Am}\mathbf{U}_A') + \sigma_{D,Dm}(\mathbf{U}_D\mathbf{U}_{Dm}' + \mathbf{U}_{Dm}\mathbf{U}_D') \\ &\quad + \sigma_{AE,AmE}(\mathbf{U}_{AE}\mathbf{U}_{AmE}' + \mathbf{U}_{AmE}\mathbf{U}_{AE}') + \sigma_{DE,DmE}(\mathbf{U}_{DE}\mathbf{U}_{DmE}' + \mathbf{U}_{DmE}\mathbf{U}_{DE}') \\ &\quad + \sigma_e^2\mathbf{I} \end{aligned}$$

其中 σ_A^2 是种子核基因加性效应方差, σ_D^2 是种子核基因显性效应方差, σ_C^2 是细胞质基因效应方差, σ_{Am}^2 是母体核基因加性效应方差, σ_{Dm}^2 是母体核基因显性效应方差; σ_{AE}^2 是种子核基因加性 \times 环境交互效应方差, σ_{DE}^2 是种子核基因显性 \times 环境交互效应方差, σ_{CE}^2 是细胞质基因 \times 环境交互效应方差, σ_{AmE}^2 是母体核基因加性 \times 环境交互效应方差, σ_{DmE}^2 是母体核基因显性 \times 环境交互效应方差; σ_B^2 是区组效应方差; $\sigma_{A,Am}$ 是种子核基因加性效应与母体核基因加性效应的协方差, $\sigma_{D,Dm}$ 是种子核基因显性效应与母体核基因显性效应的协方差; $\sigma_{AE,AmE}$ 是种子核基因加性 \times 环境交互效应与母体核基因加性 \times 环境交互效应的协方差, $\sigma_{DE,DmE}$ 是种子核基因显性 \times 环境交互效应与母体核基因显性 \times 环境交互效应的协方差; σ_e^2 是机误效应方差分量。

采用 MINQUE (0/1) 法^[4, 22], 可以无偏地估算模型中的各项方差和协方差分量。MINQUE (0/1) 是一种最小范数二阶无偏估算法 (Minimum Norm Quadratic Unbiased Estimation, MINQUE 法)^[17], 它设所有的方差先验值为常数 1、协方差先验值为常数 0。

由 MINQUE (0/1) 法算得模型中的方差和协方差分量以后, 还需进一步计算作物种子的遗传方差和协方差分量。种子数量性状表现型方差 (V_p) 包括的遗传分量为,

$$\begin{aligned} V_p &= V_{G_0} + V_C + V_{G_m} + V_{G_0E} + V_{CE} + V_{G_mE} + 2C_{G_0, G_m} + 2C_{G_0E, G_mE} + V_e \\ &= (V_A + V_D) + V_C + (V_{A_m} + V_{D_m}) + (V_{A_E} + V_{D_E}) + V_{CE} + (V_{A_mE} + V_{D_mE}) \\ &\quad + 2(C_{A, A_m} + C_{D, D_m}) + 2(C_{A_E, A_mE} + C_{D_E, D_mE}) + V_e \end{aligned}$$

其中 V_{G_0} 是直接效应遗传方差, V_C 是胞质效应遗传方差, V_{G_m} 是母体植株核基因效应遗传方差, V_{G_0E} 是直接效应 × 环境互作遗传方差, V_{CE} 是胞质效应 × 环境互作遗传方差, V_{G_mE} 是母体植株核基因效应 × 环境互作遗传方差, C_{G_0, G_m} 是直接效应与母体植株核基因效应的遗传协方差, C_{G_0E, G_mE} 是直接效应 × 环境互作与母体植株核基因效应 × 环境互作的遗传协方差, V_e 是剩余机误方差。

这些遗传分量的计算公式因作物类型及种子世代而异。双子叶作物 F_2 种子的遗传方差和遗传协方差分量计算公式为: $V_A = 2\sigma_A^2$, $V_D = 0.375\sigma_D^2$, $V_C = \sigma_C^2$, $V_{A_m} = 2\sigma_{A_m}^2$, $V_{D_m} = \sigma_{D_m}^2$, $V_{A_E} = 2\sigma_{A_E}^2$, $V_{D_E} = 0.375\sigma_{D_E}^2$, $V_{CE} = \sigma_{CE}^2$, $V_{A_mE} = 2\sigma_{A_mE}^2$, $V_{D_mE} = \sigma_{D_mE}^2$, $C_{A, A_m} = 2\sigma_{A, A_m}$, $C_{D, D_m} = 0.5\sigma_{D, D_m}$, $C_{A_E, A_mE} = 2\sigma_{A_E, A_mE}$, $C_{D_E, D_mE} = 0.5\sigma_{D_E, D_mE}$, $V_e = \sigma_e^2$ 。

单子叶作物 F_2 胚乳的遗传方差和遗传协方差分量计算公式为: $V_A = 4.5\sigma_A^2$, $V_D = 3\sigma_D^2$, $V_C = \sigma_C^2$, $V_{A_m} = 2\sigma_{A_m}^2$, $V_{D_m} = \sigma_{D_m}^2$, $V_{A_E} = 4.5\sigma_{A_E}^2$, $V_{D_E} = 3\sigma_{D_E}^2$, $V_{CE} = \sigma_{CE}^2$, $V_{A_mE} = 2\sigma_{A_mE}^2$, $V_{D_mE} = \sigma_{D_mE}^2$, $C_{A, A_m} = 3\sigma_{A, A_m}$, $C_{D, D_m} = \sigma_{D, D_m}$, $C_{A_E, A_mE} = 3\sigma_{A_E, A_mE}$, $C_{D_E, D_mE} = \sigma_{D_E, D_mE}$, $V_e = \sigma_e^2$ 。

种子遗传实验一般需要分析若干项品质性状, 因此有必要评价种子不同性状之间的遗传相关性。采用 MINQUE (0/1) 法^[4], 也可以无偏地估算成对性状的各项协方差分量。所估算的遗传协方差分量分别定义为:

$\sigma_{A/A}$ 是种子核基因加性效应协方差, $\sigma_{D/D}$ 是种子核基因显性效应协方差, $\sigma_{C/C}$ 是细胞质基因效应协方差, σ_{A_m/A_m} 是母体核基因加性效应协方差, σ_{D_m/D_m} 是母体核基因显性效应协方差; σ_{A_E/A_E} 是种子核基因加性 × 环境互作效应协方差, σ_{D_E/D_E} 是种子核基因显性 × 环境互作效应协方差, $\sigma_{CE/CE}$ 是细胞质基因 × 环境互作效应协方差, σ_{A_mE/A_mE} 是母体核基因加性 × 环境互作效应协方差, σ_{D_mE/D_mE} 是母体核基因显性 × 环境互作效应协方差; $\sigma_{B/B}$ 是区组效应协方差; σ_{A/A_m} 是成对性状的种子核基因加性效应与母体核基因加性效应的协方差, σ_{D/D_m} 是成对性状的种子核基因显性效应与母体核基因显性效应的协方差; σ_{A_E/A_mE} 是成对性状的种子核基因加性 × 环境互作效应与母体核基因加性 × 环境互作效应的协方差, σ_{D_E/D_mE} 是成对性状的种子核基因显性 × 环境互作效应与母体核基因显性 × 环境互作效应的协方差; $\sigma_{e/e}$ 是机误效应协方差。

种子成对性状表现型协方差 (C_p) 的遗传分量包括,

$$\begin{aligned} C_p &= C_{G_0} + C_C + C_{G_m} + C_{G_0E} + C_{CE} + C_{G_mE} \\ &\quad + 2C_{G_0/G_m} + 2C_{G_0E/G_mE} + C_e \\ &= (C_A + C_D) + C_C + (C_{A_m} + C_{D_m}) \end{aligned}$$

$$+(C_{AE}+C_{DE})+C_{CE}+(C_{AmE}+C_{DmE})$$

$$+2(C_{A/Am}+C_{D/Dm})+2(C_{AE/AmE}+C_{DE/DmE})+C_e$$

其中 C_{G_0} 是直接效应遗传协方差, C_C 是胞质效应遗传协方差, C_{G_m} 是母体植株核基因效应遗传协方差, C_{G_0E} 是直接效应×环境互作遗传协方差, C_{CE} 是胞质效应×环境互作遗传协方差, C_{G_mE} 是母体植株核基因效应×环境互作遗传协方差, C_{G_0/G_m} 是直接效应与母体植株核基因效应的遗传协方差, C_{G_0E/G_mE} 是直接效应×环境互作与母体植株核基因效应×环境互作的遗传协方差, C_e 是剩余机误协方差。

协方差遗传分量的计算公式也因作物类型及种子世代而异。双子叶作物 F_2 种子成对性状的遗传协方差分量计算公式为: $C_A = 2\sigma_{A/A}$, $C_D = 0.375\sigma_{D/D}$, $C_C = \sigma_{C/C}$, $C_{Am} = 2\sigma_{Am/Am}$, $C_{Dm} = \sigma_{Dm/Dm}$, $C_{AE} = 2\sigma_{AE/AE}$, $C_{DE} = 0.375\sigma_{DE/DE}$, $C_{CE} = \sigma_{CE/CE}$, $C_{AmE} = 2\sigma_{AmE/AmE}$, $C_{DmE} = \sigma_{DmE/DmE}$, $C_{A/Am} = 2\sigma_{A/Am}$, $C_{D/Dm} = 0.5\sigma_{D/Dm}$, $C_{AE/AmE} = 2\sigma_{AE/AmE}$, $C_{DE/DmE} = 0.5\sigma_{DE/DmE}$, $C_e = \sigma_e/e$ 。

单子叶作物 F_2 胚乳成对性状的遗传方差和遗传协方差分量计算公式为: $C_A = 4.5\sigma_{A/A}$, $C_D = 3\sigma_{D/D}$, $C_C = \sigma_{C/C}$, $C_{Am} = 2\sigma_{Am/Am}$, $C_{Dm} = \sigma_{Dm/Dm}$, $C_{AE} = 4.5\sigma_{AE/AE}$, $C_{DE} = 3\sigma_{DE/DE}$, $C_{CE} = \sigma_{CE/CE}$, $C_{AmE} = 2\sigma_{AmE/AmE}$, $C_{DmE} = \sigma_{DmE/DmE}$, $C_{A/Am} = 3\sigma_{A/Am}$, $C_{D/Dm} = \sigma_{D/Dm}$, $C_{AE/AmE} = 3\sigma_{AE/AmE}$, $C_{DE/DmE} = \sigma_{DE/DmE}$, $C_e = \sigma_e/e$ 。

分别获得作物种子成对性状(性状 X 和 Y)的遗传方差分量和遗传协方差分量的无偏估计值以后,还可以进一步估算各项相关系数:

$$\begin{aligned} \text{直接加性相关 } r_A &= C_A / (V_{A(X)}V_{A(Y)})^{1/2}, \\ \text{直接显性相关 } r_D &= C_D / (V_{D(X)}V_{D(Y)})^{1/2}, \\ \text{细胞质相关 } r_C &= C_C / (V_{C(X)}V_{C(Y)})^{1/2}, \\ \text{母体加性相关 } r_{Am} &= C_{Am} / (V_{Am(X)}V_{Am(Y)})^{1/2}, \\ \text{母体显性相关 } r_{Dm} &= C_{Dm} / (V_{Dm(X)}V_{Dm(Y)})^{1/2}, \\ \text{直接加性效应} \times \text{环境互作相关 } r_{AE} &= C_{AE} / (V_{AE(X)}V_{AE(Y)})^{1/2}, \\ \text{直接显性效应} \times \text{环境互作相关 } r_{DE} &= C_{DE} / (V_{DE(X)}V_{DE(Y)})^{1/2}, \\ \text{细胞质效应} \times \text{环境互作相关 } r_{CE} &= C_{CE} / (V_{CE(X)}V_{CE(Y)})^{1/2}, \\ \text{母体加性效应} \times \text{环境互作相关 } r_{AmE} &= C_{AmE} / (V_{AmE(X)}V_{AmE(Y)})^{1/2}, \\ \text{母体显性效应} \times \text{环境互作相关 } r_{DmE} &= C_{DmE} / (V_{DmE(X)}V_{DmE(Y)})^{1/2}, \\ \text{剩余效应相关 } r_e &= C_e / (V_{e(X)}V_{e(Y)})^{1/2}. \end{aligned}$$

国际上通常采用 BLUP 法^[13]预测随机遗传效应值。BLUP 法需要用方差和协方差分量的全部参数值,而实际上这些参数是未知的。育种上常用方差和协方差分量的估计值代替 BLUP 预测公式中的方差和协方差参数,这时获得的所谓的“BLUP”预测值既失去了线性,也不能保证是无偏预测。预测种子模型中的遗传效应值,可以采用一种更有效的预测方法,即调整无偏预测(Adjusted Unbiased Prediction, AUP)方法,简称 AUP 法^[5,6]。

遗传分析的目的是在估算遗传方差、协方差分量,或预测遗传效应值的基础上,通过检验适当的统计假设,推断群体的遗传表现及性状的遗传变异规律。为了对遗传参数作

统计检验,需要计算参数估计值或预测值的标准误。采用 Jackknife 重复抽样技术^[4],可以计算各项遗传参数的估计值(或预测值)及其标准误。然后可用 t 测验对遗传参数作统计检验。

2 蒙特卡罗模拟结果

用蒙特卡罗模拟比较了二倍体种子模型和三倍体胚乳模型的分析效益。为了使不同交配设计分析效果具有可比性,本实验人为地设置了二种交配设计,并使它们的实验群体大小相等(有 66 种基因型)。交配设计 I 包括了 6 个亲本,30 个正反交 F_{1ij} 及其 30 个正反交 F_{2ij} ;交配设计 II 包括亲本 6 个,正交 $F_{1ij}(i < j)$ 、正交 $F_{2ij}(i < j)$ 、回交 $F_{1ij} \times P_j$ 和回交 $F_{1ij} \times P_i$ 各 15 个。这二种交配设计虽然群体大小相等,但杂交的工作量不同。在交配设计 I 中只有 F_1 种子需要杂交获得,而交配设计 II 中 F_1 和两种回交世代的种子都需由杂交产生。设定每种交配设计均在二年实施,每年有 3 个区组。以年份内区组作为 Jackknife 重复抽样单位,对每项设计模拟 400 次,计算了遗传参数(Φ)平均估计值($\hat{\Phi}$)、遗传参数的估计偏差($\text{Bias} = \hat{\Phi} - \Phi$)、效益系数($\text{C.E.} = [\text{Bias}^2 + \text{Var}(\hat{\Phi})]^{1/2} / (|\Phi| + |\text{Bias}|)$)和功效值(在 5% 水平下测出参数显著性的概率值)。并用这些统计量检验不同方法估算遗传方差、遗传协方差或预测遗传效应值的无偏性和有效性。随机遗传效应的预测值向量 $\hat{\mathbf{e}}$ 与实际值向量 $\tilde{\mathbf{e}}$ 之间的距离($\text{Dis.} = \|\hat{\mathbf{e}} - \tilde{\mathbf{e}}\|$)是比较预测效果的主要指标。用该指标比较了 BLUP 和 AUP 两种预测方法的效益。由于区组效应对遗传分析的价值不大,故没有列出区组效应的模拟结果。

估算方差和协方差分量的蒙特卡罗模拟结果表明(表 1),对于不同遗传模型的二种交配设计,采用 MINQUE(0/1)法都可获得方差和协方差分量的无偏估计(偏差 < 参数的 10%)。模型中各项分量的估算并不都是等效的,机误项的效益最好(C. E. = 0.12),显性效应的估算效益(C. E. \approx 0.60)超过加性效应的(C. E. \approx 1.00),基因与环境互作的估算效应一般优于基因效应的,而各项协方差的估算效益较差。虽然模拟实验的遗传群体大小相等,但二种交配设计的估算效益并不相等。与交配设计 II 相比,由交配设计 I 所估算的细胞值效应和母体遗传效应的方差分量以及这些效应与环境的交互方差分量,均表现较小的效益系数和较大的功效值,因而有较好的效益。在各种可能的交配设计中,交配设计 I 所需要的杂交制种工作量最少,但其对于种子数量性状遗传参数的估计效益却最好。

本文所提出的遗传模型要求杂交亲本是一个推断群体的随机样本。通过对亲本材料的统计分析,可以推断这个遗传群体中种子数量性状的规律。但是在实际应用时,育种工作者往往还希望对亲本或杂交组合的育种价值也能有所了解。因此有必要预测各项遗传效应值。

分别按二倍体种子模型和三倍体胚乳模型产生伪正态分布(平均数为零,方差值如表 2 所示)的随机数,运用蒙特卡罗模拟比较了两种无偏预测方法(BLUP 和 AUP)。分析结果表明,两种预测方法都能得到无偏的遗传效应预测值,但 AUP 的均值比 BLUP 的均值更接近遗传效应的均值参数(零)。用预测值计算的方差($\hat{\sigma}^2$)是否接近方差参数(σ^2),是判断预测有效性的一项指标。对于所有的遗传效应,用 BLUP 值计算的方差都偏低。这表明,遗传效应的 BLUP 绝对值比实际遗传效应的绝对值小。而遗传效应 AUP 值的

表1 二种组配方式估算方差和协方差分量的偏差、效益系数和功效值的比较

Table 1 Bias, C. E. and power value of estimated variance and covariance components by two types of mating design

参数 Parameter	真值 True value	交配设计 I Design I			交配设计 II Design II		
		偏差 Bias	效益系数 C.E.	功效值 Power	偏差 Bias	效益系数 C.E.	功效值 Power
		二倍体种子 Diploid seed					
σ_A^2	50	1.58	1.04	0.86	1.28	1.04	0.86
σ_D^2	30	-0.06	0.57	0.99	0.58	0.61	0.99
σ_C^2	25	0.23	1.00	0.88	0.81	2.05	0.65
σ_{Am}^2	50	-3.23	0.87	0.89	-3.23	0.93	0.85
σ_{Dm}^2	30	2.23	0.54	0.99	1.97	0.64	0.97
σ_{AE}^2	40	-1.04	0.67	0.99	-0.57	0.71	0.99
σ_{DE}^2	25	-0.11	0.38	1.00	-0.33	0.41	1.00
σ_{CE}^2	20	-0.22	0.67	0.99	-1.69	1.44	0.68
σ_{AmE}^2	40	0.95	0.69	1.00	1.59	0.79	0.98
σ_{DmE}^2	25	-0.56	0.39	1.00	-0.64	0.43	1.00
$\sigma_{A.Am}$	15	-0.61	2.26	0.65	-1.47	2.28	0.62
$\sigma_{D.Dm}$	15	0.81	0.80	0.92	1.08	0.86	0.90
$\sigma_{AE.AmE}$	10	0.83	1.86	0.75	0.85	1.86	0.70
$\sigma_{DE.DmE}$	10	-0.03	0.73	0.94	-0.35	0.78	0.89
σ_e^2	10	0.06	0.12	1.00	0.03	0.12	1.00
三倍体胚乳 Triploid endosperm							
σ_A^2	50	2.43	1.22	0.80	1.92	1.22	0.80
σ_D^2	30	0.09	0.55	0.99	0.35	0.57	0.99
σ_C^2	25	0.23	1.00	0.88	1.27	2.03	0.65
σ_{Am}^2	50	-3.09	0.88	0.88	-4.21	0.94	0.85
σ_{Dm}^2	30	2.23	0.54	0.99	2.06	0.67	0.96
σ_{AE}^2	40	-1.22	0.80	0.95	-0.51	0.84	0.93
σ_{DE}^2	25	-0.21	0.34	1.00	-0.24	0.35	1.00
σ_{CE}^2	20	-0.21	0.67	0.99	-2.48	1.48	0.66
σ_{AmE}^2	40	0.74	0.71	1.00	1.28	0.84	0.96
σ_{DmE}^2	25	-0.57	0.39	1.00	-0.71	0.42	1.00
$\sigma_{A.Am}$	15	-1.03	2.47	0.61	-2.10	2.46	0.63
$\sigma_{D.Dm}$	15	0.78	0.79	0.92	1.03	0.86	0.91
$\sigma_{AE.AmE}$	10	0.90	2.13	0.69	0.92	2.17	0.68
$\sigma_{DE.DmE}$	10	-0.08	0.70	0.95	-0.27	0.79	0.91
σ_e^2	10	0.05	0.12	1.00	0.02	0.12	1.00

表 2 采用 BLUP 法和 AUP 法对二种遗传模型预测的遗传效应的均值、方差及距离

Table 2 Mean, variance and distance for predicted genetic effects in two genetic models

by BLUP and AUP methods

遗传效应 Genetic effect	方差真值 σ^2	BLUP			AUP		
		均值($\times 10^{-4}$) Mean	方差 $\hat{\sigma}^2$	距离 Dis.	均值($\times 10^{-4}$) Mean	方差 $\hat{\sigma}^2$	距离 Dis.
二倍体种子 Diploid seed							
A	50	1.08	33.45	10.99	-0.13	53.19	12.08
D	30	0.15	15.79	17.25	-0.03	30.00	18.75
C	25	0.23	17.49	7.69	-0.05	25.59	8.27
A _m	50	1.15	31.88	10.95	-0.24	48.29	12.01
D _m	30	0.15	16.51	17.23	-0.03	32.20	18.89
AE	40	0.47	21.05	15.32	-0.10	39.05	16.46
DE	25	0.05	11.06	24.36	-0.02	25.01	27.03
CE	20	0.09	10.89	10.64	-0.03	19.94	11.53
A _m E	40	0.46	21.46	15.24	-0.14	40.95	16.52
D _m E	25	0.05	11.20	24.09	-0.02	24.64	26.61
三倍体胚乳 Triploid endosperm							
A	50	13.74	27.83	12.18	0.00	55.93	13.77
D	30	2.25	17.04	16.53	-0.03	30.12	17.88
C	25	1.86	17.50	7.69	-0.02	25.99	8.27
A _m	50	10.62	31.68	10.99	0.07	48.65	12.13
D _m	30	1.62	16.64	17.16	-0.01	32.22	18.89
AE	40	5.36	17.41	16.68	-0.00	39.19	18.44
DE	25	0.92	12.40	23.12	-0.01	24.82	25.28
CE	20	0.73	10.89	10.64	-0.01	19.94	11.53
A _m E	40	4.01	20.93	15.41	0.03	40.62	16.80
D _m E	25	0.59	11.24	24.05	-0.00	24.58	26.60

方差是方差参数的无偏估计。所以,遗传效应的 AUP 值平均值和方差都是无偏的, BLUP 值只能得到无偏的平均数。

遗传效应的预测值与实际值之间的距离是度量预测有效性的最重要指标,距离越小,表明预测值越接近实际值。在所有的线性无偏预测中, BLUP 值的距离应该是最小的。在本实验中各项遗传效应 AUP 值的距离都非常接近(略大于)BLUP 值的距离。在实际应用时,由于方差参数是未知的,因而 BLUP 值是无法算得的。在估算遗传方差分量的同时,不需太多的计算便可获得 AUP 值。AUP 值虽然不是遗传效应的线性预测,却是无偏和有效的预测。

3 讨论

对于简单的加性—显性遗传模型,通常定义狭义遗传率为加性方差占表现型方差的比率($h^2 = V_A / V_P$)。由于种子数量性状同时受到三套遗传体系(种子核基因、细胞质基因和母体植株核基因)的控制,这些基因效应与环境效应还存在不同程度的互作。因此,为了适应遗传体系的复杂性和选择育种的实际需要,狭义遗传率的概念也应作相应的修正,狭义遗传率宜定义为:有累加性遗传效应的方差占表现型方差的比率。在种子遗传模型中,有累加性遗传效应的方差除了包括加性效应方差以外,还应包括胞质效应方差、母体加性效应方差,以及这些效应与环境互作的方差。当存在基因型×环境互作效应时,遗传率应包括普通遗传率和互作遗传率。普通遗传率是有累加性的基因效应方差分量占表现型方差的比率,适用于各种环境。互作遗传率是有累加性的基因效应与环境效应互作的方差分量占表现型方差的比率,只适用于特定的环境。如果不存在基因型×环境互作效应,互作方差分量均为零,此时互作遗传率也为零。

在种子数量遗传分析中,普通遗传率包括普通直接遗传率(h_G^2)、普通胞质遗传率(h_C^2)和普通母体遗传率(h_M^2)3项。各项普通遗传率的计算公式为,

$$h_G^2 = (V_A + C_{A.Am}) / V_P$$

$$h_C^2 = V_C / V_P$$

$$h_M^2 = (V_{Am} + C_{A.Am}) / V_P$$

总的普通遗传率为

$$\begin{aligned} h_G^2 &= h_G^2 + h_C^2 + h_M^2 \\ &= (V_A + V_C + V_{Am} + 2C_{A.Am}) / V_P \end{aligned}$$

互作遗传率包括互作直接遗传率(h_{OE}^2)、互作胞质遗传率(h_{CE}^2)和互作母体遗传率(h_{ME}^2)3项。各项互作遗传率的计算公式为,

$$h_{OE}^2 = (V_{AE} + C_{AE.AmE}) / V_P$$

$$h_{CE}^2 = V_{CE} / V_P$$

$$h_{ME}^2 = (V_{AmE} + C_{AE.AmE}) / V_P$$

总的互作遗传率则为

$$\begin{aligned} h_{GE}^2 &= h_{OE}^2 + h_{CE}^2 + h_{ME}^2 \\ &= (V_{AE} + V_{CE} + V_{AmE} + 2C_{AE.AmE}) / V_P \end{aligned}$$

本文提出的种子遗传模型中的各项参数都不能被传统的方差分析方法(ANOVA法)所估算。运用混合线性模型的其他分析方法,如限制性最大似然法(Restricted Maximum Likelihood,简称REML法),也可以分析本文提出的种子遗传模型中的各项参数。由于REML法需要进行迭代运算,并且要求线性模型符合正态分布,而MINQUE法运算简便,也无需正态分布的限定,因此我们认为MINQUE法比REML法更为简便和优越。

MINQUE法是方差分量的无偏估算方法,其估计值可能出现负数。在分析遗传资

料时,可以设所有的负数估计值为零,并在此基础上进一步估算遗传率。如果方差参数接近零,对于非零估计值的限定可能导致方差分量的有偏估算。

虽然种子模型(1)和(2)中包括了区组效应,如果种子遗传实验只有重复而未设置区组,可以删除模型中的区组效应。

Jackknife 重复抽样技术是计算参数估计值(或预测值)及其标准误的有效方法。重复抽样的方法决定于抽样单位的选择。对于多环境下完全随机区组设计的遗传资料,可选用环境内的区组作为抽样单位。但如果实验未设区组或重复,则可对基因型进行抽样。

参 考 文 献

- 1 莫惠栋, 1988. In: Weir B. S. *et al.* (eds) Proc of the 2nd international Conference on Quantitative Genetics. Sinauer, Sunderland / MA, 478—487
- 2 莫惠栋, 1989. 遗传学报, 16(2): 111—117
- 3 莫惠栋, 1989. 遗传学报, 16(3): 333—341
- 4 朱 军, 1992. 生物数学学报, 7(1): 1—11
- 5 朱 军, 1993. 生物数学学报, 8(1): 32—44
- 6 朱 军、许馥华, 1994. 作物学报, 20(3): 264—270
- 7 Beavis W D *et al.*, 1987. Theor. Appl. Genet., 74: 571—578
- 8 Bogyo T P *et al.*, 1988, Heredity, 60: 61—67
- 9 Dani R G and R J Kohel, 1989. Theor. Appl. Genet., 77: 569—575
- 10 Dhaliwal H S, 1977. Theor. Appl. Genet., 50: 235—239
- 11 Erickson E A *et al.*, 1988. Crop Sci., 28: 644—646
- 12 Foolad M R and R A Jones, 1992. Theor. Appl. Genet., 83: 360—366
- 13 Henderson C R, 1963. In: Hanson W D, and H F Robinson (eds.) Statistical Genetics and Plant Breeding. Natl. Acad. Sci.—N R C, 982: 141—163
- 14 Kohel R J, 1980. Crop Sci., 20: 784—787
- 15 Peltone J, 1992. Acta Agricultura Scandinavica, Section B. Soil and Plant Science, 42: 111—117
- 16 Pleines S. and W Friedt, 1989. Theor. Appl. Genet., 78: 793—797
- 17 Rao C R, 1971. J. Multivar. Anal., 257—275
- 18 Schnebly S R and W R Fehr, 1993. Crop Science, 33: 716—719
- 19 Topham P B, 1966. Heredity, 21: 665—674
- 20 Ullrich S E and R F Eslick, 1978. Barley Genet News, 8: 108—109
- 21 Widstrom N W *et al.*, 1975. Crop Sci., 15: 467—470
- 22 Zhu J and B S Weir, 1994. Theor. Appl. Genet., 89: 153—159
- 23 Zhu J and B S Weir, 1994. Theor. Appl. Genet., 89: 160—166

Analytic Methods for Seed Models with Genotype \times Environment Interactions

Zhu Jun

(Agronomy Department Zhejiang Agricultural University, Hangzhou 310029)

Abstract

Genetic models with genotype effect (G) and genotype \times environment interaction effect (GE) are proposed for analyzing generation means of seed quantitative traits in crops. The total genetic effect (G) is partitioned into seed direct genetic effect (G_0), cytoplasm genetic effect (C), and maternal plant genetic effect (G_m). Seed direct genetic effect (G_0) can be further partitioned into direct additive (A) and direct dominance (D) genetic components. Maternal genetic effect (G_m) can also be partitioned into maternal additive (A_m) and maternal dominance (D_m) genetic components. The total genotype \times environment interaction effect (GE) can also be partitioned into direct genetic by environment interaction effect (G_0E), cytoplasm genetic by environment interaction effect (CE), and maternal genetic by environment interaction effect (G_mE). G_0E can be partitioned into direct additive by environment interaction (AE) and direct dominance by environment interaction (DE) genetic components. G_mE can also be partitioned into maternal additive by environment interaction (A_mE) and maternal dominance by environment interaction (D_mE) genetic components. Partitions of genetic components are listed for parent, F_1 , F_2 and backcrosses. A set of parents, their reciprocal F_1 and F_2 seeds is applicable for efficient analysis of seed quantitative traits. MINQUE(0/1) method can be used for estimating variance and covariance components. Unbiased estimation for covariance components between two traits can also be obtained by the MINQUE(0/1) method. Random genetic effects in seed models are predictable by the Adjusted Unbiased Prediction (AUP) approach with MINQUE(0/1) method. The jackknife procedure is suggested for estimation of sampling variances of estimated variance and covariance components and of predicted genetic effects, which can be further used in a t -test for parameter. Unbiasedness and efficiency for estimating variance components and predicting genetic effects are tested by Monte Carlo simulations.

Key words Seed traits, Genetic models, Genotypes by environment interaction, Estimation of genetic variances and covariances, Prediction of genetic effects

This project was supported by National Natural Science Foundation of China. Received on June 4, 1994.