Theor Appl Genet (2008) 116:769–776
DOI 10.1007/s00122-008-0709-3

ORIGINAL PAPER

Analysis of genetic effects of nuclear–cytoplasmic interaction on quantitative traits: Genetic models for seed traits of plants

Lide Han · Haiming Xu · Jun Zhu · Xiangyang Lou

Received: 18 May 2007 / Accepted: 8 January 2008 / Published online: 19 February 2008

Abstract Two Genetic models (an embryo model and an endosperm model) were proposed for analyzing genetic effects of nuclear genes, cytoplasmic genes, maternal genes, and nuclear–cytoplasmic interaction (NCI) as well as their genotype by environment interaction for quantitative traits of plant seed. In these models, the NCI effects were partitioned into direct additive and dominance NCI components. Mixed linear model approaches were employed for statistical analysis. For both balanced and unbalanced diallel cross designs, Monte Carlo simulations were conducted to evaluate unbiasedness and precision of estimated variance components of these models. The results showed that the proposed methods work well. Random genetic effects were predicted with an adjusted unbiased prediction method. Seed traits (protein content and oil content) of Upland cotton (*Gossypium hirsutum* L.) were analyzed as worked examples to demonstrate the use of the models.

Communicated by J.-L. Jannink.

L. Han · H. Xu · J. Zhu (✉)
Institute of Bioinformatics, Zhejiang University,
Hangzhou, Zhejiang 310029, People's Republic of China
e-mail: jzhu@zju.edu.cn

L. Han
e-mail: hldahau@126.com

H. Xu
e-mail: hmxu@zju.edu.cn

L. Han
College of Agronomy, Anhui Agricultural University,
Hefei 230036, People's Republic of China

X. Lou
Department of Psychiatry and Neurobehavioral Sciences,
University of Virginia, Charlottesville, VA 22911, USA
e-mail: xlsn@virginia.edu

Introduction

Plant seed is a reproductive organ and one of important food resources for human being. It mainly consists of maternal tissue (testa and pericarp), diploid offspring tissue (embryo) or triploid offspring tissue (endosperm). Its compositions, such as protein, oil, and carbohydrate, are the metabolic products. Photosynthesis and respiration, which are regulated by enzymes encoded by the genomes in cytoplasm and nuclear compartments, are partially responsible for its compositions expression. Nuclear genes regulate cytoplasmic gene expression and affect cytoplasmic genome organization (Ramana et al. 2002; Strand 2004). During plant seed development, these compositions are affected by maternal plant or tissue, which supplies it with nutrition. A recent report emphasized that nuclear–cytoplasmic interaction (NCI) played essential roles during the development of onion (*Allium cepa* L.) and wheat (*Triticum aestivum* L.) seed (Galloway and Fenster 1999; Gokce and Havey 2006; Mizumoto et al. 2004; Murai et al. 2002). To investigate the genetic determination of these traits, it is necessary to extend genetic models to include nuclear-cytoplasmic interaction (NCI) effects. Beavis et al. (1987) has proposed a method for estimating variance components of nuclear genes, cytoplasmic genes, and NCI by reciprocal mating designs. However, the main disadvantage of their approach was that they were unable to obtain unique estimation of cytoplasmic effects that were confounded with the maternal genetic effects. In view of this, Mosjidis et al. (1989) employed weighted least square (WLS) method to estimate cytoplasmic effects and NCI effects; the problem of using this method is that the accompanying experimental design is hardly used in practice. In addition, there have been several other genetic models proposed for the analysis of seed quantitative traits

(Bogyo et al. 1988; Foolad and Jones 1992; Lou and Zhu 2002; Mo 1987; Pooni et al. 1992; Zhu and Weir 1994a, b), but NCI was not considered.

Furthermore, to breed crop varieties with wide geographical adaptability, genotype by environment (*GE*) interaction has been an important issue to breeders and quantitative geneticists, and the *GE* interaction on seed traits were found in some plants (Peterson 1992; Tao et al. 2004). Therefore, in addition to the aforementioned genetic effects, *GE* interaction should also be taken into account in the genetic models.

In the present research, we proposed two genetic models (an embryo model and an endosperm model) for studying the inheritance of quantitative traits of plant seed that are controlled by genetic effects from nuclear genes, cytoplasmic genes, maternal genes, and their *GE* interaction effects. Mixed linear model approaches are applied for the statistical analysis. The performance of the approach was evaluated by Monte Carlo simulations, and two seed quantitative traits of Upland cotton (*Gossypium hirsutum* L.) were analyzed as examples for demonstration of the models.

Models and methodologies

Genetic models

Consider a set of genetic materials in a modified diallel design that may include F_2, backcross and/or other progeny generations derived from the traditional diallel matings. If the genetic experiments are conducted in multiple environments, the phenotypic value (y_{hijkl}) of mating-type k of the combination from maternal parent i and paternal parent j in block l within environment h can be expressed by a linear model as follows:

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

where μ is the fixed population mean, E_h is the environmental effect, G_{ijk} is the total genetic effect, GE_{hijk} is the $G_{ijk} \times E_h$ interaction effect, $B_{l(h)} \sim (0,\sigma_B^2)$ is the random effect of block l within environment h, and $e_{hijkl} \sim (0,\sigma_e^2)$ is the residual effect. If the experiment is designed in replications but not in a randomized complete block, the block effect $B_{l(h)}$ should be dropped.

For simplification, we assume the following: (1) absence of paternal effects; (2) absence of epistatic effects between different nuclear genes; (3) constant inheritance of cytoplasmic genes through maternal lines; (4) independent genetic variation between maternal tissue and cytoplasm; and (5) diverse cytoplasms existed within the population.

For a quantitative trait of the plant seed jointly controlled by the nuclear genes, cytoplasmic genes, and maternal genes, the total genetic effect (G) in model (1) can be further partitioned into the effects of nuclear genes (*Go*), cytoplasmic genes (C), maternal genes (*Gm*), and NCIs. If the additive-dominance model is employed, the *Go* can be further partitioned into the direct additive (A) and dominance (D) effects, the NCIs into direct additive nuclear–cytoplasmic interaction effects (AC) and direct dominance nuclear–cytoplasmic interaction effects (DC), and the *Gm* into maternal additive (*Am*) and maternal dominance (*Dm*) effects.

The partition of genetic effect of G_{ijk} depends on specific generations and genetic entries. For diploid embryo traits, the G_{ijk} effect of parent P_i ($k = 0$) can be partitioned into

$$G_{ij0} = 2A_i + D_{ii} + C_i + 2AC_{ii} + DC_{iii} + 2Am_i + Dm_{ii}$$

For F_{1ij} ($k = 1$) from maternal parent $i \times$ paternal parent j, it is

$$G_{ij1} = A_i + A_j + D_{ij} + C_i + AC_{ii} + AC_{ij} + DC_{iij} + 2Am_i + Dm_{ii}$$

For F_{2ij} ($k = 2$) selfed from F_{1ij}, it is

$$G_{ij2} = A_i + A_j + 0.25D_{ii} + 0.25D_{jj} + 0.5D_{ij} + C_i + AC_{ii} + AC_{ji} + 0.25DC_{iii} + 0.25DC_{jji} + 0.5DC_{iji} + Am_i + Am_j + Dm_{ij}$$

For BC_{1ij} ($k = 3$) ($F_{1ij} \times P_i$), it is

$$G_{ij3} = 1.5A_i + 0.5A_j + 0.5D_{ii} + 0.5D_{ij} + C_i + 1.5AC_{ii} + 0.5AC_{ji} + 0.5DC_{iii} + 0.5DC_{iji} + Am_i + Am_j + Dm_{ij}$$

For BC_{2ij} ($k = 4$) ($F_{1ij} \times P_j$), it is

$$G_{ij4} = 0.5A_i + 1.5A_j + 0.5D_{jj} + 0.5D_{ij} + C_i + 0.5AC_{ii} + 1.5AC_{ji} + 0.5DC_{iji} + 0.5DC_{iji} + Am_i + Am_j + Dm_{ij}$$

For RBC_{1ij} ($k = 5$) ($P_i \times F_{1ij}$), it is

$$G_{ij5} = 1.5A_i + 0.5A_j + 0.5D_{ii} + 0.5D_{ij} + C_i + 1.5AC_{ii} + 0.5AC_{ji} + 0.5DC_{iii} + 0.5DC_{iji} + 2Am_i + Dm_{ii}$$

For RBC_{2ij} ($k = 6$) ($P_j \times F_{1ij}$), it is

$$G_{ij6} = 0.5A_i + 1.5A_j + 0.5D_{jj} + 0.5D_{ij} + C_j + 0.5AC_{ii} + 1.5AC_{jj} + 0.5DC_{jjj} + 0.5DC_{ijj} + 2Am_j + Dm_{jj}$$

For triploid endosperm traits, the preceding formulas should be changed as follows.

$$G_{ij0} = 3A_i + 3D_{ii} + C_i + 3AC_{ii} + 3DC_{iii} + 2Am_i + Dm_{ii}$$

$$G_{ij1} = 2A_i + A_j + 2D_{ij} + D_{ii} + C_i + 2AC_{ii} + AC_{ji} + 2DC_{iji} + DC_{iii} + 2Am_i + Dm_{ii}$$

$$G_{ij2} = 1.5A_i + 1.5A_j + D_{ii} + D_{jj} + D_{ij} + C_i + 1.5AC_{ii} + 1.5AC_{ji} + DC_{iii} + DC_{jji} + DC_{iji} + Am_i + Am_j + Dm_{ij}$$

$$G_{ij3} = 2A_i + A_j + 1.5D_{ii} + 0.5D_{jj} + D_{ij} + C_i + 2AC_{ii} + AC_{ji} + 1.5DC_{iii} + 0.5DC_{jji} + DC_{iji} + Am_i + Am_j + Dm_{ij}$$

$$G_{ij4} = A_i + 2A_j + 0.5D_{ii} + 1.5D_{jj} + D_{ij} + C_i + AC_{ii} + 2AC_{ji} + 0.5DC_{iii} + 1.5DC_{jji} + DC_{iji} + Am_i + Am_j + Dm_{ij}$$

$$G_{ij5} = 2.5A_i + 0.5A_j + 2D_{ii} + D_{ij} + C_i + 2.5AC_{ii} + 0.5AC_{ji} + 2DC_{iii} + DC_{iji} + 2Am_i + Dm_{ii}$$

$$G_{ij6} = 0.5A_i + 2.5A_j + 2D_{jj} + D_{ij} + C_j + 0.5AC_{ii} + 2.5AC_{jj} + 2DC_{jjj} + DC_{ijj} + 2Am_j + Dm_{jj}$$

Assuming that the inbred parents are randomly sampled from a population, each of the above genetic and *GE* interaction effects in the models is treated as random effect. A_i (or A_j) $\sim (0,\sigma_A^2)$ is the direct additive effect from line i (or line j); D_{ii}, D_{ij} or D_{jj} $\sim (0,\sigma_D^2)$ is the direct dominance effect from line $i \times$ line j ($i \leq j$); C_i (or C_j) $\sim (0,\sigma_C^2)$ is the cytoplasmic effect from line i (or line j); AC_{ii}, AC_{ij}, AC_{ji}, or AC_{jj} $\sim (0,\sigma_{AC}^2)$ is the AC between A_i(orA_j) and C_i(or C_j); DC_{iii}(DC_{iij}, DC_{iji}, DC_{ijj}, DC_{jji} or DC_{jjj}) $\sim (0, \sigma_{DC}^2)$ is the DC between the D_{ii}(D_{ij} or D_{jj}) and C_i(or C_j); Am_i(or Am_j) $\sim (0, \sigma_{Am}^2)$ is the maternal additive effect from line i (or j); Dm_{ii} (Dm_{ij} or Dm_{jj}) $\sim (0,\sigma_{Dm}^2)$ is the maternal dominance effect from line $i \times$ line j ($i \leq j$). Similarly, the total interaction effect GE_{hijk} can be partitioned into the corresponding interaction effects between the environment and the aforementioned genetic effects.

Therefore, the phenotypic variance (V_P) can be expressed as

$$\begin{aligned} \boldsymbol{V}_P &= \boldsymbol{V}_G + \boldsymbol{V}_{GE} + \boldsymbol{V}_\varepsilon \\ &= (\boldsymbol{V}_A + \boldsymbol{V}_D + \boldsymbol{V}_C + \boldsymbol{V}_{AC} + \boldsymbol{V}_{DC} + \boldsymbol{V}_{Am} + \boldsymbol{V}_{Dm}) \\ &\quad + (\boldsymbol{V}_{AE} + \boldsymbol{V}_{DE} + \boldsymbol{V}_{CE} + \boldsymbol{V}_{ACE} + \boldsymbol{V}_{DCE} + \boldsymbol{V}_{AmE} \\ &\quad + \boldsymbol{V}_{DmE}) + \boldsymbol{V}_\varepsilon \end{aligned} \quad (2)$$

where V_G is the total genetic variance with components of A variance V_A, D variance V_D, C variance V_C, AC variance V_{AC}, DC variance V_{DC}, Am variance V_{Am}, Dm variance V_{Dm}. V_{GE} is the total GE variance with components of direct additive interaction (AE) variance V_{AE}, direct dominance interaction (DE) variance V_{DE}, cytoplasm interaction (CE) variance V_{CE}, maternal additive interaction (AmE) variance V_{AmE}, and maternal dominance interaction (DmE) variance V_{DmE}, the interaction of AC and environment (ACE) variance V_{ACE}, the interaction of DC and environment (DCE) variance V_{DCE}, and V_ε is the residual variance.

Variance component estimation and effect prediction

Variance components of the mixed linear model can be estimated by the methods of maximum likelihood estimation (ML) (Hartley and Rao 1967), restriction maximum likelihood estimation (REML) (Patterson and Thomson 1971), minimum norm quadratic unbiased estimation (MINQUE) (Rao 1971), and expectation-maximization (EM) algorithm (Anderson and Hinde 1988). Among these methods, MINQUE approach has advantages of unbiasedness, no assumption of normality, and less computation over the other methods. Hence, the MINQUE(1) was used in the study (Zhu and Weir 1996). An adjusted unbiased prediction (AUP) method was adopted to predict the random genetic effects.

Sampling variances for estimated variance and predicted genetic effects can be estimated by the Jackknife technique. Assume that the samples are divided into L subsets. If $\hat{\theta}$ is an estimate of a genetic parameter θ from all L subsets of samples, and $\hat{\theta}_l$ is the estimate with the lth subset omitted, then the lth pseudovalue is

$$J_l(\hat{\theta}) = L\hat{\theta} - (L-1)\hat{\theta}_{(l)} \quad (3)$$

The jackknife estimator $J(\hat{\theta})$ of parameter θ is the mean of the pseudovalues. If L is not very large, $(J(\hat{\theta}) - \theta)/SE(J(\hat{\theta}))$ is approximately distributed as a t distribution with $(L-1)$ degrees of freedom, parameter θ is zero under the null hypotheses here (Zhu and Weir 1994a). In this study, genetic entries are served as resampling units. All analyses were conducted using home-made programs.

Simulation study

In this simulation study, we considered two mating designs of modified diallel crosses, a balanced mating design, which consisted of all parents, F_1s, reciprocal F_1s, F_2s and reciprocal F_2s from seven inbred lines, and an unbalanced mating design, in which there were 13 inbred lines that were divided into two mutually exclusive groups—group 1 was composed of seven lines and group 2 contained six lines. The unbalanced design was composed of all parents in group 1, all F_1s (not reciprocal F_1s) between the lines in group 1 and those in group 2 and the corresponding F_2s. Both of the two designs had a total of 91 genetic entries and the genetic experiments were conducted in three different environments.

Results

Monte Carlo simulations

For simplicity, a randomized complete block design with three replications in each environment was used in this simulation study. On the basis of the genetic models either for an embryo trait (Model I) or for an endosperm trait (Model II) under the two designs, we generated all phenotypic values with each vector of effects following its normal distribution with different parameter scenarios. For each case, 200 simulations were carried out to obtain sample means of the estimates and mean square errors (MSE) for each variance component.

The bias and MSE for these variance components were presented in Table 1. The absolute values of most biases were less than 5% of the corresponding parameter values, which was the conventional criterion of unbiasedness in parameter estimation. Few values, with biases between 5 and 10% of the parameter values, were regarded to be well estimated. From the comparison between simulations with different sample sizes, all the counterparts of MSE showed a tendency to decrease in the larger sample case. As suggested, a large enough size should be required to obtain a reliable estimation of variance. In the unbalanced design, each variance component showed larger MSEs than that in the balanced design for the same model. For different allocations of samples into mating types (such as some missing sample in the balanced design, different mating parents in the unbalanced design), there showed similar bias and MSE (data not shown). Under different modes of inheritance, the other effects were well detected with similar bias and MSE as the NCI ($\sigma^2_{AC} = \sigma^2_{DC} = \sigma^2_{ACE} = \sigma^2_{DCE} = 0$) or dominance variations ($\sigma^2_D = \sigma^2_{Dm} = \sigma^2_{DC} = \sigma^2_{DE} = \sigma^2_{DmE} = \sigma^2_{DCE} = 0$) were present (data not shown). Additionally, a tendency was revealed by simulations that the counterparts of MSE were increased with the decrease in heritabilities (Table 2); all effects still showed unbiased estimation. It indicted that the proposed method was reliable in estimating variance components.

Worked example

Two seed traits of cotton, i.e. oil content and protein content, were analyzed as examples. The experiments were conducted at Zhejiang University Experimental Farmer (Hangzhou, China) with two randomized complete blocks in 1994, 1995, and 1996. Crosses and reciprocal crosses of five Upland cotton varieties (denoted by P_1, P_2, P_3, P_4, and P_5) were arranged in a 5 × 5 diallel design with some missing entries. The seeds of all parents, F_1s, RF_1s, F_2s and RF_2s, were produced by hand-pollination and collected at maturity. As the seed is mainly made of embryo, the Model I was employed.

Table 1 Bias and MSE from simulations by MINQUE(1) method with the Jackknife procedure for modified diallel crosses under the heritability $h^2_G = 32\%$

Parameter	True value	Model I				Model II			
		Balanced design		Unbalanced design		Balanced design		Unbalanced design	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
σ^2_A	20	0.54	88.53	1.02[a]	196.85	0.49	81.44	1.72[a]	188.58
σ^2_D	20	0.12	89.67	0.14	94.90	0.20	88.57	0.50	102.27
σ^2_C	20	−0.19	93.68	1.10[a]	195.85	−0.19	92.07	0.72	102.35
σ^2_{AC}	20	−0.43	93.05	−1.03[a]	202.22	−0.06	93.19	0.43	92.16
σ^2_{DC}	30	−0.37	203.58	−0.76	250.43	−0.08	166.60	1.61[a]	324.95
σ^2_{Am}	30	−0.13	209.73	1.48	216.02	0.58	201.77	2.24[a]	316.23
σ^2_{Dm}	20	−0.03	90.98	−0.38	97.09	0.03	97.30	0.26	99.58
σ^2_{AE}	15	0.86	47.17	1.03[a]	59.73	−0.15	48.65	0.31	53.93
σ^2_{DE}	10	−0.36	25.33	−0.49	26.77	0.20	23.30	−0.56	25.53
σ^2_{CE}	20	−0.31	80.05	−0.61	94.14	0.19	89.67	1.03[a]	194.42
σ^2_{ACE}	15	0.42	47.79	0.49	54.40	−0.03	49.76	0.12	51.33
σ^2_{DCE}	15	−0.24	49.47	−0.13	57.18	0.03	54.95	−1.22[a]	60.92
σ^2_{AmE}	20	0.52	73.84	0.51	86.09	0.78	84.85	1.70[a]	195.54
σ^2_{DmE}	10	−0.16	24.04	−0.18	25.83	−0.09	25.50	0.35	26.98
σ^2_e	235	−1.07	186.10	1.51	202.32	−1.45	209.03	−1.69	221.21

[a] Bias between 5 and 10% of the true value

Table 2 Bias and MSE from simulations by MINQUE(1) method with the Jackknife procedure for modified diallel crosses under the heritability $h_G^2 = 25\%$

Parameter	True value	Model I				Model II			
		Balanced design		Unbalanced design		Balanced design		Unbalanced design	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
σ_A^2	20	0.58	90.49	1.29[a]	199.32	0.54	82.15	1.75[a]	195.16
σ_D^2	20	0.19	92.90	0.38	101.01	0.52	90.02	−1.11[a]	205.40
σ_C^2	20	−0.41	94.26	1.22[a]	197.31	0.56	99.00	0.91	103.84
σ_{AC}^2	20	−0.56	94.18	−1.14[a]	206.70	0.15	97.23	0.88	98.06
σ_{DC}^2	30	−0.46	218.96	1.23	266.54	−1.29	180.43	1.74[a]	349.73
σ_{Am}^2	30	0.44	214.36	1.62[a]	320.82	1.78[a]	308.49	2.40[a]	325.49
σ_{Dm}^2	20	0.16	96.73	0.59	102.61	0.49	101.11	0.57	102.87
σ_{AE}^2	15	0.95[a]	48.84	1.19[a]	62.77	0.32	48.05	0.54	56.68
σ_{DE}^2	10	−0.66[a]	25.93	−0.87[a]	29.35	−0.43	23.47	−0.81	26.67
σ_{CE}^2	20	0.44	85.76	0.69	97.31	0.60	93.78	1.19[a]	195.51
σ_{ACE}^2	15	0.28	53.05	0.81	57.97	0.27	50.42	−0.33	52.97
σ_{DCE}^2	15	0.16	53.75	0.39	61.76	0.32	56.64	−1.25[a]	62.68
σ_{AmE}^2	20	0.56	81.89	0.68	90.87	0.79	84.11	1.58[a]	197.56
σ_{DmE}^2	10	0.57[a]	24.43	0.64[a]	26.49	−0.15	27.57	−0.63[a]	28.26
σ_e^2	375	−1.65	473.67	2.41	515.20	−2.70	563.29	−2.32	532.28

[a] Bias between 5 and 10% of the true value

The estimated variance components were presented in Table 3. The contributions of V_G and V_{GE} to the total genetic and interaction variance were 77.09 and 22.91% for protein content, and 76.20 and 23.80% for oil content, respectively; so, these traits were controlled mainly by genetic effects and slightly by *GE* interaction effects. For genetic effects, there were significant variances of NCI, accounting for about 16.56 and 38.23% of the total genetic variances for the protein content and oil content, respectively, which indicated that protein content was influenced by *DC* and the oil content controlled by *AC*.

Table 3 Estimation of variance components for genetic effects and *GE* interaction effects of seed traits in Upland cotton (Estimate ± SE)

Parameter	Protein content	Oil content
V_A	0.754** ± 0.144	0.060 ± 0.808
V_D	0.390** ± 0.092	0.170** ± 0.034
V_C	3.462** ± 0.887	0.203 ± 0.399
V_{AC}	0.000 ± 0.326	0.268* ± 0.140
V_{DC}	1.036** ± 0.419	0.000 ± 0.197
V_{Am}	0.380** ± 0.068	0.000 ± 0.039
V_{Dm}	0.235** ± 0.060	0.000 ± 0.038
V_{AE}	0.215** ± 0.052	0.028 ± 0.029
V_{DE}	0.348 ± 0.217	0.000 ± 0.178
V_{CE}	1.083** ± 0.386	0.140 ± 0.190
V_{ACE}	0.000 ± 0.157	0.000 ± 0.102
V_{DCE}	0.000 ± 0.094	0.000 ± 0.115
V_{AmE}	0.081** ± 0.016	0.051 ± 0.029
V_{DmE}	0.133 ± 0.099	0.000 ± 0.168
V_ε	0.211** ± 0.050	1.491** ± 0.463

* and ** indicate significance at the 0.05 and 0.01 levels, respectively

The predicted *DC* effects and their standard errors for protein content (Table 4) showed that the higher predicted values were DC_{111}, DC_{222}, DC_{333}, DC_{444} and DC_{555}, which showed that the genetic homogeneity lines gave rise to a superior *DC* than the heterogeneity hybrid ones. For oil content, the predicted *AC* effects and their standard errors were also shown in Table 5. Since the *AC* interaction variance (V_{AC}) was the most important, the genetic merits of hybridized combinations could be evaluated mainly based on $\widehat{AC}_{ij}$. The highest predicted value $\widehat{AC}_{53}$ indicated that P_5 (male parent) and P_3 (female parent) might be suitable parents for improving oil content, although the two parents owned the lower homogenous *AC* effects.

Discussion

The influence of NCI on seed traits had been studied in many plants, such as cotton, wheat, oilseed rape (*Brassica napus* L.), maize (*Zea mays* L.), rice (*Oryza satiua* L.), etc. (Gazyantz and Zhai 1992; Murai et al. 2002; Pathania et al. 2003; Rao and Fleming 1978; Tao et al. 2004). In *Arabidopsis thaliana* (L.) Heynh. (Columbia) seeds,

Table 4 Predicated *DC* effect for protein content of seed in Upland cotton (Estimate ± SE)

Parameter	Protein content
DC_{111}	1.484** ± 0.029
DC_{131}	0.263** ± 0.020
DC_{141}	−1.789** ± 0.054
DC_{151}	−0.629** ± 0.007
DC_{331}	0.039** ± 0.010
DC_{441}	−0.667** ± 0.008
DC_{551}	−0.715** ± 0.009
DC_{222}	5.726** ± 0.121
DC_{232}	1.981** ± 0.058
DC_{242}	0.164** ± 0.018
DC_{252}	2.957** ± 0.032
DC_{332}	0.789** ± 0.010
DC_{442}	1.108** ± 0.018
DC_{552}	0.399** ± 0.007
DC_{113}	0.202** ± 0.004
DC_{133}	1.960** ± 0.037
DC_{223}	0.762** ± 0.018
DC_{233}	2.843** ± 0.060
DC_{333}	1.957** ± 0.026
DC_{353}	1.884** ± 0.067
DC_{553}	0.766** ± 0.003
DC_{114}	0.156** ± 0.040
DC_{144}	3.124** ± 0.042
DC_{224}	0.688** ± 0.005
DC_{234}	1.175** ± 0.057
DC_{444}	2.855** ± 0.023
DC_{454}	2.872** ± 0.055
DC_{554}	0.580** ± 0.037
DC_{555}	5.772** ± 0.019

** indicate significance at the 0.01 level

biosynthesis of seed oil is controlled by three nuclear genes and one cytoplasm gene, nuclear genes coordinate cytoplasmic gene expression (Ke et al. 2000; Leon et al. 1998), and the development of seed oil is also influenced by maternal effects (Tsuchiya et al. 2004). As *GE* interaction is a common characteristic for quantitative traits, *GE* interaction is also likely to exist for seed oil. These provide empirical proofs for justifying the proposed genetic models. In our example, genetic analysis was made for two seed quality traits in Upland cotton (oil content and protein content). The result showed that NCI effects were involved in their inheritance; *AC* and *DC* effects played important roles in genetic variation of oil content and protein content, respectively.

The seed traits are controlled by different genetic mechanisms in different plants. The genetic models for seed traits of plants include Models I and II. In the present study, we assume the choice of models depends on the compositional traits and the types of seeds (embryo or endosperm trait), e.g., the seed traits of cotton were employed with the Model I, because the large volume of the cotton seed consists of embryo. On the other hand, Model II may be appropriate for seed traits of rice, since the major part of the rice seed is endosperm. If the information about seed composition or genetic mechanism was not clear, we suggest that a model selection procedure is needed to test which model fits the data best. The minimum of Akaike's information (AIC) is a simple and very useful criterion for selecting the best model among alternative models. The AIC can be obtained by maximum likelihood estimation such as EM algorithm and REML (Wada and Kashiwagi 1990).

In breeding practice, the exploitation of NCI has been a component in improving seed yield and its components, and so it is important to correctly estimate NCI variation. The new model can partition the total genetic effect and interaction effect into effects of nuclear genes, cytoplasmic genes, maternal genes, NCI, and their *GE* interaction. The results of the Monte Carlo simulation showed that this method can well estimate each variance component. Here we consider the NCI effects that have not been included by the previous seed models (Lou and Zhu 2002; Zhu and Weir 1994a, b). As for the model proposed by Beavis et al. (1987), if maternal effect was small enough to be ignored for reciprocal mating design in breeding experiments, reciprocal effects could be attributed to cytoplasmic effects

Table 5 Predicted *AC* effect for oil content of seed in Upland cotton (Estimate ± SE)

	AC_{k1}	AC_{k2}	AC_{k3}	AC_{k4}	AC_{k5}
$k = 1$	0.128** ± 0.014		0.086** ± 0.018	0.035* ± 0.016	
$k = 2$		0.195** ± 0.029	0.155** ± 0.081	0.087* ± 0.029	
$k = 3$	0.008 ± 0.014	0.103** ± 0.013	0.067** ± 0.012		
$k = 4$	0.191** ± 0.047	0.212** ± 0.065		0.189** ± 0.064	
$k = 5$	0.129** ± 0.042	0.010 ± 0.017	0.280** ± 0.041	0.079** ± 0.017	0.408** ± 0.067

* and ** indicate significance at the 0.05 and 0.01 levels, respectively

k represents the subscript for additive effect source (according to the male parent code)

(Golmirzaie and Ortiz 2003); moreover, the unique cytoplasmic variance could not be unbiasedly estimated by quadratic analysis (Beavis et al. 1987; Cockerham and Weir 1977). If there were no maternal effects, the model proposed by Mosjidis et al. (1989) would provide unbiased estimation on the effects of nuclear, cytoplasm and NCI for seed quantitative traits. In addition, genetic merits of breeding material are sometimes more interesting to breeder than the variance components in plant breeding; therefore, the genetic effects predicted by the AUP method were of importance.

To conduct genetic research, plant breeders usually produce hybrid seeds by artificial emasculation and pollination, which is laborious for many cereal crops. Therefore, it is difficult to conduct experiments for the reciprocal crosses mating design with all possible hybrid combinations. Furthermore, interspecific and higher order crossing is possible in only one direction because of intercrossing barriers of some plants in nature (Eijlander et al. 2000). CMS (cytoplasmic male-sterile lines) is not only a good example of NCI in genetics, but it is easy to make hybridization at a large scale (Françoise et al. 2003; Schnable and Wise 1998; Touzet et al. 2004). The present models can estimate the variance components and predict genetic effects of the quantitative traits of the plant seed by using CMS and CMS fertility restorer (R) lines. Among the six generations of parents, F_1s, F_2s, BC_1s, BC_2s, RBC_1s, and RBC_2s, any three generations, such as parents (CMS and R lines), F_1s, F_2s, can be chosen to construct the unbalanced mating crosses design. Monte Carlo simulations showed that variance components could be well and efficiently estimated for this type of design.

Acknowledgments This work was supported in part by the 973 project of National Natural Science Foundation of China. We are grateful to Dr. Jannink Jean-Luc and two anonymous reviewers for useful comments and suggestions on the earlier version of the manuscript.

References

Anderson DA, Hinde JP (1988) Random effects in generalized linear models and the EM algorithm. Commun Statis Theory Meth 17:3847–3856

Beavis WD, Pollak E, Frey KJ (1987) A theoretical model for quantitatively inherited traits influenced by nuclear–cytoplasmic interaction. Theor Appl Genet 74:571–578

Bogyo TP, Lance RCM, Chevalier P, Nilan RA (1988) Genetic models for quantitatively inherited endosperm characters. Heredity 60:61–67

Cockerham CC, Weir BS (1977) Quadratic analysis of reciprocal crosses. Biometrics 33:187–203

Eijlander R, Laak WT, Hermsen JGT, Ramanna MS (2000) Occurrence of self-compatibility, self-incompatibility and unilateral incompatibility after crossing diploid *S. tuberosum* (SI) with *S. verrucosum* (SC): I. Expression and inheritance of self-compatibility. Euphytica 115:127–139

Foolad MR, Jones RA (1992) Models to estimate maternally controlled genetic variation in quantitative seed characters. Theor Appl Genet 83:360–366

Françoise B, Pascal T, Rosine DP (2003) The nucleo-mitochondrial conflict in cytoplasmic male sterilities revisited. Genetica 117:3–16

Galloway LF, Fenster CB (1999) The effect of nuclear and cytoplasmic genes on fitness and local adaptation in an annual legume, *Chamaecrista fasciculata*. Evolution 53:1734–1743

Gazyantz CM, Zhai XJ (1992) The physiological bases for interspecific hybrids of intensive type cotton. Acta Agriculturae Boreali-Sinica 7:55–59

Gokce AF, Havey MJ (2006) Selection at the Ms locus in open pollinated onion (*Allium cepa L.*) populations possessing S-cytoplasm or mixtures of N- and S-cytoplasms. Genet Resour Crop Evol 53:1495–1499

Golmirzaie AM, Ortiz R (2003) Reciprocal effects in true potato seed breeding in short-day length environments. Plant Breed 122:372–374

Hartley HO, Rao JNR (1967) Maximum-likelihood estimation for the mixed analysis of variance model. Biometrics 54:93–108

Ke JS, Wen TN, Nikolau BJ, Wurtele ES (2000) Coordinate regulation of the nuclear and plastidic genes coding for the subunits of the Heteromeric Acetyl-Coenzyme A Carboxylase. Plant Physiol 122:1057–1071

Leon P, Arroyo A, Mackenzie S (1998) Nuclear control of plastid and mitochondrial development in higher plants. Annu Rev Plant Physiol Plant Mol Biol 49:453–480

Lou XY, Zhu J (2002) Analysis of genetic effects of major genes and polygenes on quantitative traits. II. Genetic models for seed traits of crops. Theor Appl Genet 105:964–971

Mizumoto K, Murai K, Nakamura C, Takumi S (2004) Preferential expression of a HLP homolog encoding a mitochondrial L14 ribosomal protein in stamens of common wheat. Gene 343:281–289

Mo HD (1987) Genetic expression for endosperm traits. In: Weir B, Eisen E, Goodman M, Namkoong G (eds) Proceedings of the 2nd International Conference on Quantitative Genetics. Sinauer Associates, Sunderland, MA, pp 478–487

Mosjidis JA, Waines JG, Yermanos DM, Rosielle AA (1989) Methods for the study of cytoplasmic effects on quantitative traits. Theor Appl Genet 77:195–199

Murai K, Takumi S, Koga H, Ogihara Y (2002) Pistillody, homoeotic transformation of stamens into pistil-like structures, caused by nuclear–cytoplasmic interaction in wheat. Plant J 29:169–181

Pathania A, Bhat SR, Kumar VD, Ashutosh, Kirti PB, Prakash S, Chopra VL (2003) Cytoplasmic male sterility in alloplasmic *Brassica juncea* carrying *Diplotaxis catholica* cytoplasm: molecular characterization and genetics of fertility restoration. Theor Appl Genet 107:455–461

Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. Biomatrika 58:545–554

Peterson CJ (1992) Genotype and environment effects on quality characteristics of hard red winter wheat. Crop Sci 32:98–103

Pooni HS, Kumar I, Khush GS (1992) A comprehensive model for disomically inherited metrical traits expressed in triploid tissues. Heredity 69:166–174

Ramana S, Sharma-Natu P, Ghildiyal MC (2002) Cytoplasmic effects on photosynthesis and ribulose-1,5-bisphosphate carboxylase activity in *Brassica* species. Euphytica 123:361–365

Rao CR (1971) Estimation of variance and covariance components MINQUE theory. J Multivar Anal 1:257–275

Rao AP, Fleming AA (1978) Cytoplasmic-genetypic effect in the GT 112 maize inbred with four cytoplasms. Crop Sci 18:935–937

Schnable PS, Wise RP (1998) The molecular basis of cytoplasmic male sterility and fertility restoration. Trends Plant Sci 3:175–180

Strand A (2004) Plastid-to-nucleus signaling. Curr Opin Plant Biol 7:621–625

Tao D, Hu F, Yang J, Yang G, Yang Y, Xu P, Li J, Ye C, Dai L (2004) Cytoplasm and cytoplasm-nucleus interactions affect agronomic traits in *japonica* rice. Euphytica 135:129–134

Touzet P, Hueber N, Brkholz A, Barnes S, Cuguen J (2004) Genetic analysis of male fertility restoration in wild cytoplasmic male sterility G of beet. Theor Appl Genet 109:240–247

Tsuchiya Y, Nambara E, Naito S, McCourt P (2004) The FUS3 transcription factor functions through the epidermal regulator TTG1 during embryogenesis in Arabidopsis. Plant J 37:73–81

Wada Y, Kashiwagi N (1990) Selecting statistical models with information statistics. J Dairy Sci 73:3573–3582

Zhu J, Weir BS (1994a) Analysis of cytoplasmic and maternal effects: I. a genetic model for diploid plant seeds and animals. Theor Appl Genet 89:153–159

Zhu J, Weir BS (1994b) Analysis of cytoplasmic and maternal effects: II. genetic models for triplod endosperms. Theor Appl Genet 89:160–166

Zhu J, Weir BS (1996) Diallel analysis for sex-linked and maternal effects. Theor Appl Genet 92:1–9