3

New Approaches for Analyzing Quantitative Traits and Their Applications in Cotton

Jun Zhu

1. INTRODUCTION

Most agricultural and economically important traits of cotton are quantitative traits, which are controlled by polygenes with different genetic effects and are affected by the environment. Many quantitative genetic analyses have been conducted in cotton since the 1950s (Meredith, 1984). Generation mean analysis (Mather and Jinks, 1982) and diallel cross analysis (Yates, 1947; Hayman, 1954; Griffing, 1956; Gardner and Eberhart, 1966) have been the most widely used methods for quantitative genetic analysis in cotton. The generation mean analysis calculates the mean and variance for populations of parental and segregating generations. Since this method needs to measure individuals of many generations, it cannot handle complicated models with genotype × environment (*GE*) interactions. Diallel cross analysis utilizes analysis of variance (ANOVA), which has a few deficiencies in analyzing advanced genetic models containing unbalanced data, non-integer values of coefficients, and correlated random factors.

Since the 1970s, some new statistical methods have been developed for analyzing mixed linear models. Mixed linear model approaches can overcome the shortcomings of ANOVA methods while handling unbalanced data and complicated models. Development of mixed linear model approaches and its application in quantitative genetics have created enormous challenges for quantitative geneticists in dealing with complicated genetic problems.

In this chapter, some new genetic models and their corresponding analyses for quantitative genetics will be reviewed. Methods recently

Department of Agronomy, Zhejiang University, Hangzhou 310029, China.

developed for mixed linear models with their applications will be evaluated to show the ways of solving complicated problems in quantitative genetics for crops such as cotton. Since the mixed model approaches involve enormous matrix computation, an appropriate software has been developed for analyzing experimental data (Zhu, 1997), which is available at 'http://statgen.ncsu.edu/zhu/index.html' and can also be downloaded.

2. ANALYSIS METHODS FOR AGRONOMIC TRAITS

Quantitative traits are controlled by polygenes that have small effects and are easily affected by the environment. Genotype × environment interactions have been detected for many quantitative traits of crops. Understanding the genotypic effects, as well as *GE* interaction effects on agronomic traits, is of great importance for plant breeding and sustainable agricultural production.

When genetic experiments are conducted in multiple environments, the phenotypic performance of a genetic entry in one environment can be expressed by the following genetic model:

$$y = \mu + E + G + GE + B + \varepsilon \tag{2.1}$$

where, μ = population mean, E = environment effect, G = genotypic effect, GE = genotype × environment interaction effect, B = block effects (if present), and ε = residual effect.

Cockerham (1980) proposed a general genetic model for partitioning the total genotypic effect (G). Zhu (1994) extended Cockerham's general genetic model by including a GE interaction. If there are only additive (A) and dominance (D) effects, the effects can be partitioned as: $G = A + D$ and $GE = AE + DE$. Therefore, the total phenotype variance (V_P) for each trait, as well as the total phenotype covariance (C_P) between the two traits, can be partitioned as:

$$V_P = V_A + V_D + V_{AE} + V_{DE} + V_\varepsilon \tag{2.2a}$$

$$C_P = C_A + C_D + C_{AE} + C_{DE} + C_\varepsilon \tag{2.2b}$$

Instead of using F_1 progenies in a diallel mating or factorial mating design (Hallauer and Miranda, 1981), F_2's and the parental lines can be used for analyzing A, D, AE and DE effects (Zhu, 1993a). A genetic model for parent ($i = j$) and F_2 ($i \neq j$) in the k-th block within the h-th environment is given by:

$$\begin{aligned} y_{hijk} = \mu + E_h + A_i + A_j + \frac{1}{4} D_{ii} + \frac{1}{4} D_{jj} + \frac{1}{2} D_{ij} \\ + AE_{hi} + AE_{hj} + \frac{1}{4} DE_{hii} + \frac{1}{4} DE_{hjj} + \frac{1}{2} DE_{hij} + B_{hk} + \varepsilon_{hijk} \end{aligned} \tag{2.3}$$

where, this model cannot be analyzed by ANOVA approaches since the coefficients for *D* and *DE* are non-integer for F_2's. Variance and covariance components can be estimated in an unbiased manner by the Minimum Norm Quadratic Unbiased Estimation MINQUE (1) method (Zhu and Weir, 1996b). For genetic models with *GE* interaction effects, the total heritability (h^2) can be partitioned into genetic heritability ($h^2_G = V_A/V_P$) and interaction heritability ($h^2_{GE} = V_{AE}/V_P$) (Zhu, 1997). General heritability, which is applicable to multiple environments, is the ratio of variances of accumulated heritable genotypic effects to phenotypic variance. Interaction heritability, which is only applicable to specific environments, is the ratio of variances of accumulated heritable *GE* interaction effects to phenotypic variance.

Since phenotypic variance (V_P) and covariance (C_P) can be partitioned into components for *G* and *GE* effects, phenotypic correlation between two traits (1 and 2) is also contributed by correspondent components of correlation:

$$r_p = \sqrt{\frac{V_{A(1)}}{V_{P(1)}}}\sqrt{\frac{V_{A(2)}}{V_{P(2)}}}r_A + \sqrt{\frac{V_{D(1)}}{V_{P(1)}}}\sqrt{\frac{V_{D(2)}}{V_{P(2)}}}r_D + \sqrt{\frac{V_{AE(1)}}{V_{P(1)}}}\sqrt{\frac{V_{AE(2)}}{V_{P(2)}}}r_{AE}$$

$$+ \sqrt{\frac{V_{DE(1)}}{V_{P(1)}}}\sqrt{\frac{V_{DE(2)}}{V_{P(2)}}}r_{DE} + \sqrt{\frac{V_{\varepsilon(1)}}{V_{P(1)}}}\sqrt{\frac{V_{\varepsilon(2)}}{V_{P(2)}}}r_E \qquad (2.4)$$

where, $r_A = C_A/\sqrt{V_{A(1)}\,V_{A(2)}}$ is additive correlation, $r_A = C_A/\sqrt{V_{A(1)}\,V_{A(2)}}$ is dominance correlation, $r_A = C_A/\sqrt{V_{A(1)}\,V_{A(2)}}$ is *AE* interaction correlation, $r_{DE} = C_{DE}/\sqrt{V_{DE(1)}V_{DE(2)}}$ is *DE* interaction correlation, and $r_A = C_A/\sqrt{V_{A(1)}\,V_{A(2)}}$ is and residual correlation.

Prediction of genetic merit can be obtained by the linear unbiased prediction (LUP) method (Zhu and Weir, 1994; Zhu and Weir, 1996a) or the adjusted unbiased prediction (AUP) method (Zhu, 1993a; Zhu and Weir, 1996b). Predicted genotypic effects and *GE* interaction effects can be further used in analyzing heterosis of different generations (Zhu, 1997). Heterosis in specific environments consists of two components. General heterosis is due to genotypic effects and can be expected in all environments, while interaction heterosis is a component of *GE* interaction unique to specific environments. The two components of heterosis expressed as a proportion of the midparent or better parent can be calculated as:

General heterosis of F_n over midparent: $H_M(F_n) = \left(\frac{1}{2}\right)^{n-1} \Delta_D$

Interaction heterosis of F_n over midparent: $H_{ME}(F_n) = \left(\frac{1}{2}\right)^{n-1} \Delta_{DE}$

General heterosis of F_n over better parent (P_i):

$$H_B(F_n) = \left(\frac{1}{2}\right)^{n-1} \Delta_D - \frac{1}{2}\overline{\omega}_G$$

Interaction heterosis of F_n over better parent (P_i):

$$H_{BE}(F_n) = \left(\frac{1}{2}\right)^{n-1} \Delta_{DE} - \frac{1}{2}\overline{\omega}_{GE}$$

where, $\Delta_D = D_{ij} - \frac{1}{2}(D_{ii} + D_{jj})$ is dominance heterosis, $\Delta_{DE} = D_{hij} - \frac{1}{2}(DE_{hii} + D_{hjj})$ is *DE* interaction heterosis, $\overline{\omega}_G = |G(P_i) - G(P_j)|$ is parental genotypic difference, and $\overline{\omega}_{GE} = |GE(P_i) - GE(P_j)|$ is parental interaction difference.

Heterosis, based on population mean (i.e. $H_{PM} = \frac{1}{\mu} H_M$, $H_{PME} = \frac{1}{\mu} H_{ME}$, $H_{PB} = \frac{1}{\mu} H_B$ or $H_{PBE} = \frac{1}{\mu} H_{BE}$), can be used to compare the proportion of heterosis for different traits. In an analysis for Parents and F_1's from a 6 × 6 diallel cross of cotton, average heterosis of F_2's ($H_{PM}(F_2)$) was 8.12 per cent for bolls per plant and 5.43 per cent for boll size (Zhu, 1993a). One F_2 progeny exhibited 11.6 per cent heterosis over the parent with more bolls, and the F_3 of this cross was expected to have 5 per cent more bolls than the better parent. Wu *et al.* (1995) analyzed ten parents and their 20 F_1's and F_2's of cotton in two years. They found that hoterosis averaged 10 per cent for bolls per plant and 3 per cent for boll size. Tang *et al.* (1996) analyzed 64 F_2 hybrids resulting from crosses of four commercial cultivars and 16 pest-resistant germplasm lines for five fiber and four yield traits in four environments at Mississippi State, MS, USA. They found that dominance variance accounted for the major proportion of the phenotypic variances for lint yield, lint percentage and boll size. A low proportion of additive variance for fiber traits and significant additive × environment variance components were observed.

For some traits of crops, epistatic effects could be important components of total genotypic effects. Among three types of epistasis [additive

by additive (*AA*), dominance by dominance (*DD*), and additive by dominance (*AD*)], *AA* is most important for its response to selection. Zhu (1994) suggested using three generations (Parents, F_1's, and F_2's) from a diallel mating to estimate *A*, *D*, and *AA* effects and their environment interactions, *AE*, *DE*, and *AAE* effects. The partitioning of *G* and *GE* effects for these three generations in the *h*-th environment is defined as:

$$G(P_i) + GE(P_i) = 2A_i + D_{ii} + 4AA_{ii} + 2AE_{hi} + DE_{hii} + 4AAE_{hii}$$

$$G(F_{1ij}) + GE(F_{1ij}) = A_i + A_j + D_{ij} + AA_{ii} + AA_{jj} + 2AA_{ij} + AE_{hi} + AE_{hj} + DE_{hij} + AAE_{hii} + AAE_{hjj} + 2AAE_{hij} \quad (2.5)$$

$$G(F_{2ij}) + GE(F_{2ij}) = A_i + A_j + \frac{1}{4} D_{ii} + \frac{1}{4} D_{jj} + \frac{1}{2} D_{ij} + AA_{ii} + AA_{jj} + 2AA_{ij} + AE_{hi} + AE_{hj} + \frac{1}{4} DE_{hii} + \frac{1}{4} DE_{hjj} + \frac{1}{2} DE_{hij} + AAE_{hii} + AAE_{hjj} + 2AAE_{hij}.$$

The epistatic model can be analyzed by the MINQUE(1) method (Zhu and Weir, 1996b) for estimating variances and covariances, and by the LUP (Zhu and Weir, 1994) or AUP method (Zhu, 1993a; Zhu and Weir, 1996b) for predicting genetic effects. The partitioning of total phenotype variance (V_P) and covariance (C_P) are as follows:

$$V_P = V_A + V_D + V_{AA} + V_{AE} + V_{DE} + V_{AAE} + V_\varepsilon \quad (2.6a)$$

$$C_P = C_A + C_D + C_{AA} + C_{AE} + C_{DE} + C_{AAE} + C_\varepsilon \quad (2.6b)$$

When the *AA* epistasis effects are included in the model, predicted heterosis will contain extra components ($2\Delta_{AA}$ and $2\Delta_{AAE}$) (Xu and Zhu, 1999), such that:

$$H_M(F_n) + H_{ME}(F_n) =$$

$$\left[\left(\frac{1}{2}\right)^{n-1} \Delta_D + 2\Delta_{AA}\right] + \left[\left(\frac{1}{2}\right)^{n-1} \Delta_{DE} + 2\Delta_{AAE}\right]$$

$$H_B(F_n) + H_{BE}(F_n) =$$

$$\left[\left(\frac{1}{2}\right)^{n-1} \Delta_D - \frac{1}{2}\bar{\omega}_G + 2\Delta_{AA}\right] + \left[\left(\frac{1}{2}\right)^{n-1} \Delta_{DE} - \frac{1}{2}\bar{\omega}_{GE} + 2\Delta_{AAE}\right]$$

where, $\Delta_{AA} = AA_{ij} - \frac{1}{2}(AA_{ii} + AA_{jj})$ is epistatic heterosis, and $\Delta_{AAE} = AAE_{hij} - \frac{1}{2}(AAE_{hii} + AAE_{hjj})$ is epistatic × environment heterosis.

It is suggested that heterosis due to additive epistasis (*AA* and *AAE*) could be passed to later generations, but that due to dominance (*D* and *DE*) should be reduced by half for each succeeding generation. Using the *AD* + *AA* model with *GE* interaction effects, Xu and Zhu (1999) predicted the general and interaction heterosis of cotton over two years. General heterosis was significant for bolls per plant [$H_{PM}(F_1) \hat{=} 11.8\%$ and $H_{PM}(F_2) \hat{=} 6.5\%$] but not for boll size. The reverse was true for interaction heterosis. For boll size, $H_{PME}(F_1) \hat{=} 6.3\%$ in 1992, and $H_{PME}(F_1) \hat{=} 7.2\%$ and $H_{PME}(F_2) \hat{=} 2.8\%$ in 1993.

For analyzing agronomy traits, programs of 'GENAD.EXE' for AD model and 'GENADE.EXE' for AD + AA model can be used for analyzing data file 'filename.txt' in order to construct design matrix of experimental data. Estimated variance components and predicted genetic effects can be obtained with the program 'GENVAR1R.EXE' for jackknifing over blocks or 'GENVAR1C.EXE' for jackknifing over entry means. Covariance components can be estimated with the program 'GENCOV1R.EXE' or 'GENCOV1C.EXE'. Heterosis can be predicted by the program 'GENHET1R.EXE' or 'GENHET1C.EXE'.

3. ANALYSIS METHODS FOR SEED TRAITS

An important breeding objective is improvement of crop quality. Genetic models with applicable statistical methods for developing seeds are of importance for efficient analysis of seed quantitative traits. Seed traits could be simultaneously controlled by seed direct genetic effects (G_O), cytoplasm genetic effects (G_C), and maternal nuclear genetic effects (G_M), as well as their respective *GE* interaction effects (G_OE, G_CE, and G_ME). Therefore model (2.1) needs to be expanded to include more genomic systems for seed traits (Zhu, 1996) as follows:

$$\begin{aligned} y &= \mu + E + G_O + G_C + G_M + G_OE + G_CE + G_ME + B + \varepsilon \\ &= \mu + E + A + D + C + Am + Dm + AE + DE + CE \\ &\quad + AmE + DmE + B + \varepsilon \end{aligned} \tag{3.1}$$

where, A = direct additive effect, D = direct dominance effect, C = cytoplasmic effect, Am = maternal additive effect, Dm = maternal dominance effect, AE = direct additive × environment interaction effect, DE = direct dominance × environment interaction effect, CE = cytoplasmic × environment interaction effect, AmE = maternal additive × environment interaction effect, and DmE = maternal dominance × environment interaction

effect. The components of the total phenotype variance (V_P) consist of the following variances and covariances:

$$\begin{aligned} V_P &= V_{G_O} + V_{G_C} + V_{G_M} + V_{G_OE} + V_{G_CE} \\ &\quad + V_{G_ME} + 2C_{G_O.G_M} + 2C_{G_OE.G_ME} + V_\varepsilon \\ &= V_A + V_D + V_C + V_{Am} + V_{Dm} + V_{AE} + V_{DE} + V_{CE} + V_{AmE} \\ &\quad + V_{DmE} + 2C_{A.Am} + 2C_{D.Dm} + 2C_{AE.AmE} + 2C_{DE.DmE} + V_\varepsilon \end{aligned} \tag{3.2}$$

where, $C_{G_O.G_M} = C_{A.Am} + C_{D.Dm}$ is the covariance between the G_O effect (A and D effects) and the G_M effect (Am and Dm effects) of the same trait, $C_{G_OE.G_ME} = C_{AE.AmE} = C_{DE.DmE}$ is the covariance between the G_OE effect (AE and DE effects) and the G_ME effect (AmE and DmE effects) of the same trait.

The total phenotype covariance (C_P) between two seed traits can also be partitioned the same way as for phenotype variance:

$$\begin{aligned} C_P &= C_{G_O} + C_{G_C} + C_{G_M} + C_{G_OE} + C_{G_CE} + C_{G_ME} \\ &\quad + 2C_{G_O/G_M} + 2C_{G_OE/G_ME} + C_\varepsilon \\ &= C_A + C_D + C_C + C_{Am} + C_{Dm} + C_{AE} + C_{DE} + C_{CE} + C_{AmE} \\ &\quad + C_{DmE} + 2C_{A/Am} + 2C_{D/Dm} + 2C_{AE/AmE} + 2C_{DE/DmE} + C_\varepsilon \end{aligned} \tag{3.3}$$

where, $C_{G_O/G_M} = C_{A/Am} + C_{D/Dm}$ is the covariance between the G_O effect (A and D effects) of one trait and the G_M effect (Am and Dm effects) of another trait, $C_{G_OE/G_ME} = C_{AE/AmE} + C_{DE/DmE}$ is the covariance between the G_OE effect (AE and DE effects) of one trait and the G_ME effect (AmE and DmE effects) of another trait.

Based on this extension, experiments of a diallel cross with three generations (Parents, F_1's, and F_2's) conducted in multiple environments can be appropriately analyzed (Zhu and Weir, 1994; Zhu, 1996). For cotton seeds the partition of the total $G + GE$ effect for three generations is

$$\begin{aligned} G(P_i) + GE(P_i) &= 2A_i + D_{ii} + C_i + 2Am_i + Dm_{ii} + 2AE_{hi} \\ &\quad + DE_{hii} + CE_{hi} + 2AmE_{hi} + DmE_{hii} \\ G(F_{1ij}) + GE(F_{1ij}) &= A_i + A_j + D_{ij} + C_i + 2Am_i + 2Dm_{ii} + AE_{hi} \\ &\quad + AE_{hj} + DE_{hij} + CE_{hi} + 2AmE_{hi} + DmE_{hii} \end{aligned} \tag{3.4}$$

$$G(F_{2ij}) + GE(F_{2ij}) = A_i + A_j + \frac{1}{4} D_{ii} + \frac{1}{4} D_{jj} + \frac{1}{2} D_{ij} + C_i + Am_i + Am_j$$
$$+ Dm_{ij} + AE_{hi} + AE_{hj} + \frac{1}{4} DE_{hii} + \frac{1}{4} DE_{hjj} + \frac{1}{2} DE_{hij}$$
$$+ CE_{hi} + AmE_{hi} + AmE_{hj} + DmE_{hij}$$

Other generations, such as BC_1's and BC_2's and their reciprocals (RBC_1's and RBC_2's, respectively), can also be used for analyzing seed traits (Zhu and Weir, 1994; Zhu, 1996). Variances and covariances of the models for seeds can be estimated by the MINQUE (0/1) method (Zhu and Weir, 1994). Therefore, other parameters (e.g. heritability and selection response) derived from variance and covariance components can be obtained. In seed genetic models, the total heritability (h^2) can be partitioned into general heritability (h^2_G) and interaction heritability (h^2_{GE}) as:

$$h^2 = h^2_G + h^2_{GE}$$
$$= h^2_O + h^2_C + h^2_M + h^2_{OE} + h^2_{CE} + h^2_{ME} \quad (3.5)$$

where, $h^2_O = (V_A + C_{A.Am})/V_P$ is direct general heritability, $h^2_C = V_C/V_P$ is cytoplasmic general heritability, and $h^2_M = (V_{Am} + C_{A.Am})/V_P$ is maternal general heritability; and $h^2_{OE} = (V_{AE} + C_{AE.AmE})/V_P$ is direct interaction heritability, $h^2_{CE} = V_{CE}/V_P$ is cytoplasmic interaction heritability, and $h^2_{ME} = (V_{AmE} + C_{AE.AmE})/V_P$ is maternal interaction heritability (Zhu, 1997).

Heritability is often used in predicting selection response. Since heritability consists of several components for the seed model, the total selection response ($R = ih^2\sqrt{V_P}$) can also be partitioned into several components (Zhu, 1997) as:

$$R = R_G + R_{GE} \quad (3.6)$$
$$= (R_O + R_C + R_M) + (R_{OE} + R_{CE} + R_{ME})$$

where, $R_G = ih^2_G\sqrt{V_P}$ is general response, which consists of direct general response ($R_O = ih^2_O\sqrt{V_P}$), cytoplasmic general response ($R_C = ih^2_C\sqrt{V_P}$), and maternal general response ($R_M = ih^2_M\sqrt{V_P}$); $R_{GE} = ih^2_{GE}\sqrt{V_P}$ is interaction response, which consists of direct interaction response ($R_{OE} = ih^2_{OE}\sqrt{V_P}$), cytoplasmic interaction response ($R_{CE} = ih^2_{CE}\sqrt{V_P}$), and maternal interaction response ($R_{ME} = ih^2_{ME}\sqrt{V_P}$).

Plant breeders usually want to improve seed quality traits while retaining the genetic merit of agronomic traits. Therefore, understanding the genetic relationship between seed quality traits and agronomic traits is of importance. While seed models include direct ($A_{(S)}$ and $D_{(S)}$), cytoplasm ($C_{(S)}$) and maternal effects ($Am_{(S)}$ and $Dm_{(S)}$), models for the plant only have cytoplasmic ($C_{(P)}$) and plant nuclear effects ($Am_{(P)}$ and $Dm_{(P)}$). These two kinds of models have unequal design matrices. Zhu (1993b, 1997) developed a new method for estimating covariance components between seed traits and maternal plant traits. Applying this method, Wang *et al.* (1996) analyzed data from a two-year experiment for five upland cotton parents and their F_1's and F_2's. Covariances between four kernel quality traits (oil content, oil index, protein content, and protein index) and four yield traits (lint yield, boll number per plant, boll size, and lint percentage) were estimated. No significant covariances were found between lint yield per plant and kernel quality. Covariances were negative for boll number per plant but positive for boll size with oil content, oil index, and protein index for cytoplasmic effects, and maternal additive and dominance effects, as well as their environmental interaction effects. There were positive cytoplasmic and maternal relationships between lint percentage and protein index.

Genetic effects in the seed model can be predicted by the LUP (Zhu and Weir, 1994) or AUP method (Zhu, 1993a; Zhu and Weir, 1996b). Heterosis of seed traits also consists of components due to direct, cytoplasmic and maternal effects as well as their *GE* interaction effects (Zhu, 1997). For quantitative traits of seed, two types of heterosis can be evaluated: Heterosis over midparent (H_M) and heterosis over female parent (H_F):

$$H_M(F_n) + H_{ME}(F_n) = H_{MO} + H_{MC} + H_{MM} + H_{MOE} + H_{MCE} + H_{MME}$$

$$= \left(\frac{1}{2}\right)^{n-1} \Delta_O + \frac{1}{2}\bar{\omega}_C + \left(\frac{1}{2}\right)^{n-2} \Delta_M$$

$$+ \left(\frac{1}{2}\right)^{n-1} \Delta_{OE} + \frac{1}{2}\bar{\omega}_{CE} + \left(\frac{1}{2}\right)^{n-2} \Delta_{ME}$$

$$H_M(F_n) + H_{FE}(F_n) = H_{FO} + H_{FM} + H_{FOE} + H_{FME}$$

$$= \left[\left(\frac{1}{2}\right)^{n-1} \Delta_O - \frac{1}{2}\bar{\omega}_O\right] + \left[\left(\frac{1}{2}\right)^{n-2} \Delta_M - \frac{1}{2}\bar{\omega}_M\right]$$

$$+ \left[\left(\frac{1}{2}\right)^{n-1} \Delta_{OE} - \frac{1}{2}\bar{\omega}_{OE}\right] + \left[\left(\frac{1}{2}\right)^{n-2} \Delta_{ME} - \frac{1}{2}\bar{\omega}_{ME}\right]$$

where, $\Delta_O = D_{ij} - \frac{1}{2}(D_{ii} + D_{jj})$, $\Delta_M = Dm_{ij} - \frac{1}{2}(Dm_{ii} + Dm_{jj})$, $\bar{\omega}_O = 2(A_i - A_j) + (D_{ii} - D_{jj})$, $\bar{\omega}_C = C_i - C_j$, $\bar{\omega}_M = 2(Am_i - Am_j) + (Dm_{ii} - Dm_{jj})$.

Zhu *et al.* (1997) analyzed three years of data (1991-1993) for kernel nutrient quality traits of upland cotton. The experiment consisted of nine generations (P_i and P_j, $F_1 = P_i \times P_j$, $RF_1 = P_j \times P_i$, F_2, $BC_i = F_1 \times P_i$, $BC_j = F_1 \times P_j$, $RBC_i = P_i \times F_1$, $RBC_j = P_j \times F_1$) derived from five parents and their six F_1 crosses. The results for kernel oil content and protein content are presented in Table 1. The total *GE* interaction variance ($V_{GE} = V_{AE} + V_{DE} + V_{CE} + V_{AmE} + V_{DmE}$) was almost as important as the total genotypic variance ($V_G = V_A + V_D + V_C + V_{Am} + V_{Dm}$) for oil content ($V_G \hat{=} 5.065$, $V_{GE} \hat{=} 5.112$) and protein content ($V_G \hat{=} 6.414$, $V_{GE} \hat{=} 4.390$). Seed effects ($V_O + V_{OE} \hat{=} 3.138$ for oil content and 3.766 for protein content) appeared to be less important than cytoplasmic and maternal effects ($V_C + V_M + V_{CE} + V_{ME} \hat{\approx} 7.040$ for oil content and protein content). Variances of dominance effects ($V_D + V_{Dm} + V_{DE} + V_{DmE} \hat{=} 2.403$ for oil content and 1.255 for protein content) were much smaller than those of heritable effects ($V_A + V_C + V_{Am} + V_{AE} + V_{CE} + V_{AmE} \hat{=} 7.774$ for oil content and 9.549 for protein content). The covariances ($C_{A.Am}$ and $C_{D.Dm}$) suggested that the genotypic effects of seed and plant genomes were positively correlated for protein content but not for oil content.

Table 1. Estimates of genetic parameters for kernel oil content and protein content in upland cotton

Trait	V_A	V_D	V_C	V_{Am}	V_{Dm}	V_{AE}	V_{CE}	V_{DmE}
Oil %	1.485**	0.426**	0.000	3.154**	0.000	1.227**	1.908**	1.977**
Protein %	1.586**	0.478**	1.692**	1.881**	0.777**	1.702**	2.688**	0.000
	$C_{A.Am}$	$C_{D.Dm}$	V_ε	h^2_O	h^2_C	h^2_M	h^2_{OE}	h^2_{CE}
Oil %	-1.083	0.000	0.956**	0.045	0.000	0.231**	0.137**	0.213**
Protein %	0.673**	0.170*	0.591**	0.173**	0.129**	0.195**	0.130**	0.206**
	r_P	r_A	r_D	r_{Am}	r_{AE}	r_{CE}	r_ε	
Oil % & Protein %	-0.130**	0.278**	-0.093*	-0.327**	0.309**	-0.266**	-0.415**	

*, ** Significantly different from zero at the 0.05 and 0.01 levels of probability, respectively.

The total general heritability ($h^2_G = h^2_O + h^2_C + h^2_M$) was higher than the total interaction heritability ($h^2_{GE} = h^2_{OE} + h^2_{CE} + h^2_{ME}$) for protein content ($h^2_G \hat{=} 0.497 > h^2_{GE} \hat{=} 0.336$). But the reverse was true for oil content ($h^2_G \hat{=} 0.231 < h^2_{GE} \hat{=} 0.350$). Therefore, it would be more difficult to improve oil content than protein content. For improvement of seed quantitative traits, selection should be based on measuring individual seeds (single-seed selection) or samples of seeds from maternal plants (mater-

nal-plant selection). For oil content and protein content, maternal plant selection could be more efficient than single seed selection, because the major component of h_G^2 was h_M^2 for oil content and $(h_M^2 + h_C^2)$ for protein content. As h_{CE}^2 was the main components of h_{GE}^2, selection based on cytoplasmic interaction effects could result in genetic gain in specific environments. Between the two traits a very weak negative r_p was observed, which was primarily due to negative r_D, r_{Am}, r_{CE} and r_ε. The positive additive (r_A) and AE interaction (r_{AE}) correlations suggested that oil and protein content might be simultaneously improved by single-seed selection for these two traits.

For analyzing diploid seed traits, programs of 'GENDIPLD.EXE' can be used for analyzing data file 'filename.txt' to construct design matrix of experimental data. Estimated variance components and predicted genetic effects can be obtained with the program 'GENVAR0R.EXE' for jackknifing over blocks or 'GENVAR0C.EXE' for jackknifing over entry means. Covariance components are estimable by the program 'GENCOV0R.EXE' or 'GENCOV0C.EXE'. Heterosis can be predicted by the program 'GENHET0R.EXE' or 'GENHET0C.EXE'.

4. ANALYSIS METHODS FOR DEVELOPMENTAL TRAITS

Understanding the gene expression for quantitative traits at specific times is an important objective in developmental quantitative genetics. Cotton, like many other crops, has a long period of blooming and boll-setting. The final yield of cotton is determined by the number of bolls setting at different stages of growth. For an additive and dominance model with GE interaction effects, the phenotypic value of quantitative trait measured at time t can be partitioned as:

$$\begin{aligned} y_{(t)} &= \mu_{(t)} + E_{(t)} + G_{(t)} + GE_{(t)} + B_{(t)} + \varepsilon_{(t)} \\ &= \mu_{(t)} + E_{(t)} + A_{(t)} + D_{(t)} + AE_{(t)} + DE_{(t)} + B_{(t)} + \varepsilon_{(t)} \end{aligned} \tag{4.1}$$

The partitioning of phenotypic variance at time t in model (3.1) can be expressed as:

$$\begin{aligned} V_{P(t)} &= V_{G(t)} + V_{GE(t)} V_{B(t)} + V_{\varepsilon(t)} \\ &= V_{A(t)} + V_{D(t)} + V_{AE(t)} + V_{DE(t)} + V_{B(t)} V_{\varepsilon(t)} \end{aligned} \tag{4.2}$$

Analyzing random effects in model (4.1) and variance components in equation (4.2) will reveal the genetic properties of accumulated genotypic effects and GE interaction effects from initial time to time t (i.e., $0 \rightarrow t$). In order to study the net gene effects in the period from time $t-1$ to time t (i.e., $t-1 \rightarrow t$), Zhu (1995) developed a mixed model approach for analyzing conditional genotypic effects and conditional variance components. The conditional genetic model for the phenotypic mean

measured at time t conditional upon the phenotypic mean measured at time $t-1$ is:

$$y_{(t|t-1)} = \mu_{(t|t-1)} + E_{(t|t-1)} + G_{(t|t-1)} + GE_{(t|t-1)} + B_{(t|t-1)} + \varepsilon_{(t|t-1)} \quad (4.3)$$
$$= \mu_{(t|t-1)} + E_{(t|t-1)} + A_{(t|t-1)} + D_{(t|t-1)} + AE_{(t|t-1)} + DE_{(t|t-1)} + B_{(t|t-1)} + \varepsilon_{(t|t-1)}$$

with conditional phenotypic variance partitioned as:

$$V_{P(t|t-1)} = V_{G(t|t-1)} + V_{GE(t|t-1)} + V_{B(t|t-1)} + V_{\varepsilon(t|t-1)} \quad (4.4)$$
$$= V_{A(t|t-1)} + V_{D(t|t-1)} + V_{AE(t|t-1)} + V_{DE(t|t-1)} + V_{B(t|t-1)} + V_{\varepsilon(t|t-1)}$$

The genotypic effects and GE interaction effects at time t conditional upon the causal effects at time $(t-1)$ will imply the new effects of genes which are independent of the causal effects. The changes in conditional genotypic and GE interaction variance can be used to measure the epigenetic effects of the causal components on the dynamic variability of developmental behavior.

Zhu (1995) applied these conditional approaches to analyze fruiting behavior of upland cotton in 1981 and 1985. From 1 July to 3 September in each year, bolls per plant were counted every 5 days on 10 plants from each plot. Cell means for the two-year data were analyzed by the additive-dominance model including GE interactions. Genotypic variance ($V_G = V_A + V_D$) was more important than GE interaction variance ($V_{GE} = V_{AE} + V_{DE}$). The major contribution of genetic variation for number of bolls per plant was due to the dominance variance component in the early blooming period, but was due to the additive variance component in the later blooming period. Additive effects (A_2 for full-season variety DPL15, A_3 for early-season variety GL5) and additive variances are summarized in Table 2.

Table 2. Estimated additive variance components and predicted additive effects for bolls per plant of upland cotton, by date

Month/Day	$V_{A(t)}$	$V_{A(t\|t-1)}$	$A_{2(t)}$+	$A_{3(t)}$	$A_{2(t\|t-1)}$	$A_{3(t\|t-1)}$
7/05	0.11**	0.11**	−0.27**	0.30**	−0.27**	0.30**
7/10	0.19*	0.00	−0.35*	0.32*	0.00	0.00
7/15	0.00	0.00	0.00	0.00	0.00	0.00
7/20	0.00	0.01	0.00	0.00	0.04	0.03
7/25	0.00	0.00	0.00	0.00	0.00	0.00
7/30	0.00	0.11*	0.00	0.00	0.12*	−0.29*
8/04	0.00	0.60**	0.00	0.00	0.44**	−0.79**
8/09	2.25*	0.53**	0.38*	−1.31**	0.66**	−0.56**
8/14	4.47**	0.10*	0.81*	−2.06**	0.27*	−0.14
8/19	6.27**	0.07*	1.18**	−2.47**	0.25	−0.03
8/24	7.26**	0.00	1.40**	−2.67**	0.00	0.00
8/29	7.78**	0.01*	1.54**	−2.78**	0.06	−0.02
9/03	8.18**	0.02**	1.64**	−2.87**	0.07	−0.05

*, ** Significantly different from zero at the 0.05 and 0.01 levels of probability, respectively.
+ A_2 = DPL15, A_3 = GL5.

At the early stages (7/05 to 7/10), there was significant unconditional additive variance ($V_{A(t)}$). Unconditional additive effects ($A_{(t)}$) were positive for GL5 but negative for DPL15. Conditional additive effects ($A_{(7/10|7/05)}$) and conditional variance ($V_{A(7/10|7/05)}$) were not significant at 7/10. This implied that there was no additional gene expression for additive effects from time 7/05 to 7/10. Detection of significant unconditional additive variation at 7/10 was due to the accumulated additive effects of genes expressed at the initial time. Even though unconditional additive variance was not detectable at times 7/30 and 8/04, conditional additive effects and variance were already detected. This implies that expression of quantitative genes might start several days before the accumulated genetic effects for developmental behavior can be detected by unconditional genetic analysis. Unconditional additive variance was redetectable and increased steadily after 8/04. The unconditional additive effects were negative for GL5 but positive for DPL15. The increase in unconditional additive variance was mostly contributed by gene expression for additive effects during time 8/04 to 8/09. After 8/09 there were very few new additive effects.

Developmental genetic behavior of quantitative traits for nutrients in cotton seed can also be analyzed by conditional methodology. The phenotype measured at time t for seed can be expressed as a function of different component effects similar to model (3.1):

$$\begin{aligned} y_{(t)} = {} & \mu_{(t)} + E_{(t)} + A_{(t)} + D_{(t)} + C_{(t)} + Am_{(t)} + Dm_{(t)} \\ & + AE_{(t)} + DE_{(t)} + CE_{(t)} + AmE_{(t)} + DmE_{(t)} + \varepsilon_{(t)} \end{aligned} \tag{4.5}$$

Phenotypic variance can be partitioned as:

$$\begin{aligned} V_{P(t)} = {} & V_{A(t)} + V_{D(t)} + V_{C(t)} + V_{Am(t)} + V_{Dm(t)} + V_{AE(t)} \\ & + V_{DE(t)} + V_{CE(t)} + V_{AmE(t)} + V_{DmE(t)} + 2C_{A.Am(t)} \\ & + 2C_{D.Dm(t)} + 2C_{AE.AmE(t)} + 2C_{DE.DmE(t)} + V_{\varepsilon(t)} \end{aligned} \tag{4.6}$$

The conditional effects can be analyzed by the mixed model approaches based on the following conditional seed model:

$$\begin{aligned} y_{(t|t-1)} = {} & \mu_{(t|t-1)} + E_{(t|t-1)} + A_{(t|t-1)} + D_{(t|t-1)} + C_{(t|t-1)} + Am_{(t|t-1)} \\ & + Dm_{(t|t-1)} + AE_{(t|t-1)} + DE_{(t|t-1)} + CE_{(t|t-1)} + AmE_{(t|t-1)} \\ & + DmE_{(t|t-1)} + \varepsilon_{(t|t-1)} \end{aligned} \tag{4.7}$$

Conditional phenotypic variance can be partitioned as:

$$\begin{aligned} V_{P(t|t-1)} = {} & V_{A(t|t-1)} + V_{D(t|t-1)} + V_{C(t|t-1)} + V_{Am(t|t-1)} + V_{Dm(t|t-1)} \\ & + V_{AE(t|t-1)} + V_{DE(t|t-1)} + V_{CE(t|t-1)} + V_{AmE(t|t-1)} \end{aligned}$$

$$+ V_{DmE(t|t-1)} + 2C_{A.Am(t|t-1)} + 2C_{D.Dm(t|t-1)} + 2C_{AE.AmE(t|t-1)} + 2C_{DE.DmE(t|t-1)} + V_{\varepsilon(t|t-1)} \quad (4.8)$$

Recently we analyzed a 2-year data set of eight parents and their F_1, reciprocal F_1, and F_2 progenies. This experiment was conducted with two replications in 1994 and one replication in 1995 at Hangzhou, China. Lysine index (cg of lysine per 100 seeds) and oil index (g oil per 100 seeds) were measured on seed kernels sampled from four stages (20, 30, 40, and 50 days after blooming). Estimates of unconditional variances at the initial stage and conditional variances at successive stages are presented in Table 3.

Table 3. Estimates of variance components for lysine index and oil index of upland cotton

	Lysine index ($\times 10^{-1}$)				Oil index ($\times 10^{-3}$)			
Var	20D	30D|20D	40D|30D	50D|40D	20D	30D|20D	40D|30D	50D|40D
V_A	1.42 **	0.00	30.89 **	0.00	4.66 **	0.00	48.72 **	0.00
V_D	0.00	1.85 **	0.00	10.83 **	0.46 **	9.66 **	0.00	0.00
V_C	0.79 **	0.00	17.39 **	0.00	0.00	0.00	0.00	0.00
V_{Am}	0.72 **	9.46 **	14.64 **	0.00	2.60 **	48.65 **	131.89 **	167.45 **
V_{Dm}	0.50 **	6.67 **	0.00	31.25 **	2.75 **	29.15 **	52.20 **	92.72 **
V_{AE}	0.96 **	23.97 **	36.39 **	0.00	0.00	35.38 **	0.00	0.00
V_{DE}	0.50 **	0.00	12.53 **	0.00	2.86 **	6.13 **	22.98 **	18.45 **
V_{CE}	0.87 **	15.55 **	0.00	56.28 **	6.68 **	49.70 **	0.00	0.00
V_{AmE}	0.00	5.28 **	0.00	0.00	0.00	0.00	0.00	25.59 **
V_{DmE}	1.20 **	0.00	48.21**	0.00	0.00	55.46 **	32.37 **	0.00
V_ε	0.54 **	44.42 **	16.20 **	152.65 **	8.33 **	23.49 **	61.62**	85.71 **

** Significantly different from zero at the 0.01 level of probability.

Behavior of genetic effects varied across developmental stages for different traits. For two traits studied, genotypic variance (V_G) was similar to GE interaction variance (V_{GE}) at the initial stage (20D), but conditional $V_{G(30D|20D)}$ was smaller than conditional $V_{GE(30D|20D)}$ from 20D to 30D. At the third ($30D \rightarrow 40D$) and fourth ($40D \rightarrow 50D$) intervals, the gene expression tended to be more stable ($V_{G(40D|30D)} > V_{GE(40D|30D)}$, $V_{G(50D|40D)} > V_{GE(50D|40D)}$) for oil index, but less stable for lysine index. Gene expression of cytoplasmic and maternal genomes (G_C, G_CE, G_M, and G_ME) had much larger effects than those of the seed genome (G_O, G_OE) for oil index, but the difference was not as large for lysine index. At later stages (30D $\rightarrow$ 40D $\rightarrow$ 50D), oil index had larger general heritable effects (G_A, G_C, and G_{Am}) than interaction heritable effects (G_{AE}, G_{CE}, and G_{AmE}).

For analyzing developmental quantitative traits, design matrix of experimental data should be constructed by using programs described in section one for agronomy traits or in section two for seed traits. The

original data files 'filename.txt' should be converted into conditional data files 'filename.doc' by running the program 'GENCOND1.EXE' for agronomy traits or 'GENCOND0.EXE' for seed traits. The conditional files can then be analyzed by the procedures described in the preceding sections for estimating variances and covariances, predicting genetic effects and heterosis.

5. ANALYSIS METHODS FOR QTL MAPPING

Recent developments in molecular marker techniques and quantitative trait loci (QTL) mapping methodology have provided possibilities for identifying individual gene effects for various important agronomic traits. To date, QTLs have been evaluated mostly by the interval mapping (IM) method (Lander and Botstein, 1989) using MAPMAKER/QTL software (Lincoln *et al.* 1993), or by the composite interval mapping (CIM) method (Zeng, 1994) using QTL Cartographer software (Basten *et al.* 1996). These two methods are based on maximum likelihood approaches for regression models with different underlying genetic assumptions. Genetic variation is assumed to be due to one searched QTL ($G = G_Q$) in the IM method, but is assumed to be due to one searched QTL with markers linked to other QTLs ($G = G_Q + G_M$) in the CIM method. Both IM and CIM methods cannot directly handle QTL mapping data derived from multi-environment experiments, because random environmental effects and QTL × environment (QE) interaction effects cannot be included in the regression models.

Because quantitative traits are usually controlled by multiple QTLs with genetic main effects as well as QE interaction effects, the following mixed-model-based CIM (MCIM) method was proposed (Zhu, 1998; Zhu and Weir, 1998) for analyzing the following genetic model:

$$y = \mu + G_Q + E + G_QE + G_M + G_ME + \varepsilon \qquad (5.1)$$

where, G_Q = QTL main effect (fixed), E = environment effect, G_QE = QTL × environment interaction effect, G_M = genetic main effect of markers linked to other QTLs, and G_ME = marker × environment interaction effect.

When applying model (5.1) for searching a putative QTL within two flanking markers (M_{i-} and M_{i+}), the phenotypic value of a quantitative trait measured on the j-th individual in the h-th environment can be expressed as:

$$y_{hj} = \mu + ax_{A_j} + dx_{D_j} + u_{E_{hj}}\, e_{E_h} + u_{AE_{hj}}\, e_{AE_h} + u_{DE_{hj}}\, e_{DE_h} \qquad (5.2)$$

$$+ \sum_{k \neq i-,i+} u_{M_{kj}}\, e_{M_k} + \sum_{k \neq i-,i+} u_{ME_{hkj}}\, e_{ME_{hk}} + \varepsilon_{hj}$$

where, a and d are the additive and dominance main effects for the searched QTL; x_{A_j} and x_{D_j} are coefficients for genotypic effects; e_{E_h} is the h-th environment effect with its coefficient $u_{E_{hj}}$; e_{AE_h} is the additive × environment interaction effect with its coefficient $u_{AE_{kj}}$; e_{DE_h} is the dominance × environment interaction effect with its coefficient $u_{DE_{hj}}$; e_{M_k} is the random main effect for the k-th marker genotype with its coefficient $u_{M_{kj}}$; $e_{ME_{hk}}$ is the marker × environment interaction effect with its coefficient $u_{ME_{hkj}}$.

If double haploids (DH) or recombinant inbred lines (RIL) populations are used for mapping QTLs in multiple environments genotypic effects for A and AA, as well as GE interaction for AE and AAE can be simultaneously analyzed by the MCIM method base on the following mixed linear model (Zhu, 1998):

$$\begin{aligned} y_{hk} &= \mu + a_i x_{A_{ik}} + a_j x_{A_{jk}} + aa_{ij} x_{AA_{ijk}} \\ &+ u_{E_{hk}}\ e_{E_h} + u_{A_iE_{hk}}\ e_{A_iE_h} + u_{A_jE_{hk}}\ e_{A_jE_h} + u_{AA_{ij}E_{hk}}\ e_{AA_{ij}E_h} \\ &+ \sum_f u_{M_{fk}}\ e_{M_f} + \sum_l u_{MM_{lk}}\ e_{MM_l} + \sum_p u_{ME_{hpk}}\ e_{ME_{hp}} \\ &+ \sum_q u_{MME_{hqk}}\ e_{MME_{hq}} + \varepsilon_{hk} \end{aligned} \tag{5.3}$$

where, a_i and a_j are the additive effects of loci Q_i and Q_j, respectively; aa_{ij} is the AA epistasis effect of loci Q_i and Q_j; $x_{A_{ik}}$, $x_{A_{jk}}$ and $x_{AA_{ijk}}$ are coefficients of these QTL effects; e_{E_h} is the random effect of environment h with coefficient $u_{E_{hk}}$; $e_{A_iE_h}$ (or $e_{A_jE_h}$) is the AE interaction effect with coefficient $u_{A_iE_{hk}}$ (or $u_{A_jE_{hk}}$) for Q_i (or Q_j); $e_{AA_{ij}E_h}$ is the AAE interaction effect with coefficient $u_{AA_{ij}E_{hk}}$; e_{M_f} is the f-th marker effect with coefficient $u_{M_{fk}}$; e_{MM_l} is the l-th effect of marker × marker interaction with coefficient $u_{MM_{lk}}$; $e_{ME_{hp}}$ is the ME interaction effect with coefficient $u_{ME_{hpk}}$; and $e_{MME_{hq}}$ and is the MME interaction effect with coefficient $u_{MME_{hqk}}$.

Models (5.2) and (5.3) can be expressed as a matrix form of the general mixed linear model as:

$$y = Xb + \sum_u U_u e_u \sim N\left(Xb, V = \sum_u \sigma_u^2 U_u R_u U_u^T\right) \tag{5.4}$$

where, y is a vector of phenotypic values; b is a vector of the fixed effects with coefficient matrix x; e_u is a vector of random effects with coefficient matrix U_u; and R_u is a constant matrix describing the relationship of e_u.

The likelihood ratio statistic (LR) can be calculated by:

$$LR = 2l_1\ (\hat{b}, \hat{V}, \hat{r}_{M_i\text{-}Q}) - 2l_0\ (\hat{b}, \hat{V}, r_{M_i\text{-}Q} = 0.5) \tag{5.5}$$

for testing the null hypothesis H_0: $r_{M_i\text{-}Q} = 0.5$ vs. the alternative hypothesis H_1: $r_{M_i\text{-}Q} < 0.5$. The LR has an approximate χ^2 distribution with $df = 4$ for model (5.2) and $df = 6$ for model (5.3).

The QTL effects (additive, dominance, and epistasis) in b can be estimated by

$$\hat{b} = (X^T\ \hat{V}^{-1}\ X)^{-1}\ X^T\ \hat{V}^{-1}\ y \tag{5.6}$$

with sampling variance matrix:

$$\mathrm{Var}\,(\hat{b}) = (X^T\ \hat{V}^{-1}\ X)^{-1} \tag{5.7}$$

The QTL × environment interaction effects (*ae*, *de*, and *aae*) e_u can be predicted by the LUP (Zhu and Weir, 1994) or AUP method (Zhu, 1993a; Zhu and Weir, 1996b).

Shappley *et al.* (1998) applied the MCIM method for mapping QTLs of 19 agronomic and fiber traits based on 31 linkage groups in upland cotton. There were 100 QTLs located on 60 positions in 24 linkage groups. Several QTLs influenced more than one trait.

A software 'QTLMapper Version 1.0' has been developed by Wang *et al.* (1999) for mapping QTLs with additive effects, additive × additive effects as well as their QTL × E interaction effects. This software and a PDF file of the user manual are freely available to users from the authors.

6. ANALYSIS METHODS FOR REGIONAL TRIALS

Among the various genetic models described in the previous sections, further partitioning of the phenotype value can be imposed on the total genotypic effect (G) as well as the genotype × environment interaction effect (GE). The purpose of quantitative genetic analysis is primarily to provide strategic selection information to breeders based on the components of G and GE. In crop cultivar development, breeding stocks or germplasm accessions are usually tested in a wide range of environments to identify superior genotypes in specific environments.

Analysis of experimental data from regional trials is based on the following linear model, which regard the genotypic effects (G) as fixed, and further partitions the random E ($E = Y + L + YL$) and GE interaction effects ($GE = GY + GL + GYL$):

$$\begin{aligned} y &= \mu + G + E + GE + B + \varepsilon \\ &= \mu + G + Y + L + YL + GY + GL + GYL + B + \varepsilon \end{aligned} \tag{6.1}$$

where, Y = year effect, L = location effect, YL = year × location interaction effect, GY = genotype × year effect, GL = genotype × location effect, GYL = genotype × year × location interaction effect, and B = block effect.

The genotypes included in regional trials are of specific interest and are defined as fixed in model (6.1). The other effects in model (6.1) are usually treated as random effects. Balanced data of the regional trials can be easily analyzed by ANOVA methods. But the experiment data of regional trial are quite often unbalanced, due to missing some genotype records in specific locations and years. Zhu *et al.* (1993a, 1993b) developed mixed model approaches for analyzing unbalanced data of single trait and multiple traits from regional trials. The phenotypic data of trait $f(f = 1,2,\ldots..,\ t)$ can be expressed as a matrix form of mixed linear model:

$$\begin{aligned} y_{(f)} &= Xb_{(f)} + U_Y e_{Y(f)} + U_L e_{L(f)} + U_{YL} e_{YL(f)} + U_{GY} e_{GY(f)} \\ &\quad + U_{GL} e_{GL(f)} + U_{GYL} e_{GYL(f)} + U_B e_{B(f)} + e_{e(f)} \\ &= Xb_{(f)} + \sum_{u=1}^{8} U_u e_{u(f)} \end{aligned} \tag{6.2}$$

with variance matrix:

$$\begin{aligned} \mathrm{Var}(y_{(f)}) &= \sigma^2_{Y(f)}\, U_Y U_Y^T + \sigma^2_{L(f)}\, U_L U_L^T + \sigma^2_{YL(f)}\, U_{YL} U_{YL}^T \\ &\quad + \sigma^2_{GY(f)}\, U_{GY} U_{GY}^T + \sigma^2_{GL(f)} U_{GL} U_{GL}^T \\ &\quad + \sigma^2_{GYL(f)}\, U_{GYL} U_{GYL}^T + \sigma^2_{B(f)} U_B U_B^T + \sigma^2_{\varepsilon(f)} I \\ &= \sum_{u=1}^{8} \sigma^2_{u(f)} U_u U_u^T = V_{(f)} \end{aligned}$$

The covariance between traits $y_{(f)}$ and $y_{(f')}$ is $C_{(ff')} = \sum_{u=1}^{8} \sigma_{u(ff')}\, U_u U_u^T$. Both variance and covariance matrices can be estimated by the MINQUE(1) method (Zhu and Weir, 1996b). Comparison of genotypes for trait f can be conducted by testing linear contrast among genotype effects $\left(\sum_{h=1}^{g} c_h\, G_{h(f)}\right)$. The linear contrast can be estimated by

$$C_{(f)} = c^T \hat{b} = c^T\, (X^T \hat{V}^{-1}_{(f)}\, X)^- X^T \hat{V}^{-1}_{(f)} y_{(f)} \tag{6.3}$$

with sampling variance $\hat{\sigma}^2(C_{(f)}) = c^T (X^T \hat{V}_{(f)}^{-1} X)^- c$. If $|C_{(f)}/\hat{\sigma}(C_{(f)})|$ $> z_{(\alpha/2)}$, reject the null hypothesis $H_0 \sum_{h=1}^{g} c_h G_{h(f)} = 0$ and accept the alternative hypothesis $H_1 \sum_{h=1}^{g} c_h G_{h(f)} \neq 0$ at a significant level = α.

To compare the weighted genotypic merits of t traits $\left(\sum_{f=1}^{t} w_f G_{h(f)} \right)$, the weighted linear contrast can be estimated by

$$C_W \sum_{h=1}^{g} c_h \sum_{f=1}^{t} w_f \hat{G}_{h(f)} = \sum_{f=1}^{t} w_f C_{(f)}$$

with sampling variance

$$\sigma^2(C_W) = \sum_{f=1}^{t} w_f^2 \sigma^2(C_{(f)}) + 2 \sum_{f=1}^{t-1} \sum_{f'=f+1}^{t} w_f w_{f'} \sigma(C_{(f)}, C_{(f')})$$

where, $\hat{\sigma}(C_{(f)}, C_{(f')}) = c^T (X^T \hat{C}_{(ff')}^{-1} X)^- c$ is the covariance between $C_{(f)}$ and $C_{(f')}$. If $|C_W/\hat{\sigma}(C_W)| > z_{(\alpha/2)}$, reject the null hypothesis $H_0 \sum_{h=1}^{g} c_h \sum_{f=1}^{t} w_f G_{h(f)} = 0$ and accept the alternative hypothesis $H_1 \sum_{h=1}^{g} c_h \sum_{f=1}^{t} w_f G_{h(f)} \neq 0$ at a significant level = α.

For analyzing data from regional trials, the program 'GENTEST.EXE' can be used to construct design matrix of experiments. First, run the program 'GENTESTM.EXE' for analyzing each trait. Then the program 'GENTESTW.EXE' can be used for combining analysis of all traits studied.

7. REFERENCES

Basten, C.J., Weir, B.S. and Zeng, Z.B. (1996). QTL Cartographer. North Carolina State University, Raleigh, NC.

Cockerham, C.C. (1980). Random and fixed effects in plant genetics. *Theor. Appl. Genet.* **56**, 119-131.

Gardner, C.O. and Eberhart, S.A. (1966). Analysis and interpretation of the variety cross diallel and related populations. *Biometrics* **22**, 439-452.

Griffing, B. (1956). Concept of general and specific combining ability in relation to diallel crossing systems. *Aust. J. Biol. Sci.* **9**, 463-493.

Hallauer, A.R. and Miranda, J.B. (1981). Quantitative Genetics in Maize Breeding. Iowa State University Press, Ames, Iowa. p 1-468

Hayman, B.I. (1954). The theory and analysis of diallel crosses. *Genetics* **39**, 789-809.

Lander, E.S. and Botstein, D. (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185-199.

Lincoln, S.E., Daly, M.J. and Lander, E.S. (1993). Mapping Genes Controlling Quantitative Traits Using MAPMAKER/QTL1.1. Whitehead Institute Technical Report, 2nd ed.

Mather, K. and Jinks, J.L. (1982). Biometrical Genetics, 3rd ed. Chapman and Hall, London, UK.

Meredith, W.R. (1984). Quantitative genetics. *In*: Cotton. R.J. Kohel and C.F. Lewis, (Eds). ASA, Inc., CSSA, Inc., and SSSA, Inc., Publishers, Madison, Wisconsin, USA, 131-150.

Shappley, Z.W., Jenkins, J.N., Zhu, J. and McCarty, J.C. (Jr). (1998). Quantitative trait loci associated with agronomic and fiber traits of upland cotton. *Journal of Cotton Science* **2**(4), 153-163.

Tang, B., Jenkins, J.N., Watson, C.E., McCarty, J.C. (Jr) and Creech, R.G. (1996). Evaluation of genetic variances, heritabilities, and correlations for yield and fiber traits among cotton F_2 hybrid populations. *Euphytica* **91**, 315-322.

Wang, G.J., Zhu, J., Zang, R.C., Xu, F.H. and Ji, D.F. (1996). Analysis of covariance components between seed and agronomy traits in upland cotton. *Acta Gossypii Sinica* **8**(6), 301-304.

Wang, D.L., Zhu, J., Li, Z.K. and Paterson, A.H. (1999). QTLMapper Version 1.0. Zhejiang University, Hangzhou, China.

Wu, J.X., Zhu, J., Ji, D.F. and Xu, F.H. (1995). Analysis of genetic effects × environment interactions for yield traits in upland cotton. (Chinese). *Hereditas* **17**(5), 1-4.

Xu, Z.C. and Zhu, J. (1999). A new approach for predicting heterosis based on an additive, dominance and additive × additive model with environment interaction. *Heredity* (*In Press*)

Yates, F. (1947). Analysis of data from all possible reciprocal crosses between a set of parental lines. *Heredity* **1**, 287-301.

Zeng, Z.-B. (1994). Precision mapping of quantitative trait loci. *Genetics* **136**, 1457-1468.

Zhu, J. (1993a). Methods of predicting genotype value and heterosis for offspring of hybrids. (Chinese). *Journal of Biomathematics* **8**(1), 32-44.

Zhu, J. (1993b). Mixed model approaches for estimating covariances between two traits with unequal design matrices. *Journal of Biomathematics* (Chinese) **8**(3), 24-30.

Zhu, J. (1994). General genetic models and new analysis methods for quantitative traits. (Chinese). *Journal of Zhejiang Agricultural University* **20**(6), 551-559.

Zhu, J. (1995). Analysis of conditional genetic effects and variance components in developmental genetics. *Genetics* **141**, 1633-1639.

Zhu, J. (1996). Analysis methods for seed models with genotype × environment interactions. (Chinese). *Acta Genetica Sinica* **23**(1), 56-68.

Zhu, J. (1997). Analysis Methods for Genetic Models. Agricultural Publishing House of China, Beijing.

Zhu, J. (1998). Mixed model approaches of mapping genes for complex quantitative traits. *In:* Proceedings of Genetics and Crop Breeding of China, L.Z. Wang and J.R. Dai (Eds). Chinese Agricultural Science and Technology Publication House, Beijing, 19-20.

Zhu, J. and Weir, B.S. (1994). Analysis of cytoplasmic and maternal effects: I. A genetic model for diploid plant seeds and animals. *Theor. Appl. Genet.* **89**, 153-159.

Zhu, J. and Weir, B.S. (1996a). Mixed model approaches for diallel analysis based on a bio-model. *Genetical Research* **68**, 233-240.

Zhu, J. and Weir, B.S. (1996b). Diallel analysis for sex-linked and maternal effects. *Theor. Appl. Genet.* **92**, 1-9.

Zhu. J., and B.S. Weir. 1998. Mixed model approaches for genetic analysis of quantitative traits. In Advanced Topics in Biomathematics: Proceedings of International Conference on Mathematical Biology. pp: 321-330. Edited by L.S. Chen, S .G. Ruan, and J. Zhu, World Scientific Publishing Co., Singapore.

Zhu, J., Lai, M.G. and Xu, F.H. (1993a). Analysis methods for unbalanced data from regional trial of crop variety: analysis for multiple traits. (Chinese). *Journal of Zhejiang Agricultural University* **19**(3), 241-247.

Zhu, J., Xu, F.H. and Lai, M.G. (1993b). Analysis methods for unbalanced data from regional trials of crop variety: analysis for single trait. (Chinese). *Journal of Zhejiang Agricultural University* **19**, 7-13.

Zhu, J., Wang, G.J. and Zang, R.C. (1997). Genetic analysis on gene effects and GE interaction effects for kernel nutrient quality traits of upland cotton. *Journal of Biomathematics*, **12**(2), 111-120.