



Genetic association of yield with its component traits in a recombinant inbred line population of cotton*

Jixiang Wu¹, Johnie N. Jenkins^{2,*}, Jack C. McCarty Jr.² & Jun Zhu³

¹Department of Plant and Soil Sciences, Mississippi State University, Mississippi State, MS 39762, U.S.A.;

²Crop Science Research Laboratory, USDA-ARS, Mississippi State, MS 39762, U.S.A.; ³College of Agriculture & Biotechnology, Zhejiang University, Hangzhou 310029, China;

(*author for correspondence: e-mail: jjenkins@msa-msstate.ars.usda.gov)

Received 5 December 2003; accepted 13 August 2004

Key words: contribution effect, contribution ratio, lint yield, mixed linear model approach, multivariable conditional analysis, upland cotton

Summary

Lint yield of upland cotton (*Gossypium hirsutum* L.) is determined by its component traits, boll number, boll weight, and lint percentage. Selecting high yielding lines is based on the ability to manipulate component traits. In this study, 188 recombinant inbred lines and two parental lines were grown in 1999 and 2000 at Mississippi State University. Lint yield and its three component traits were measured and analyzed by an extended conditional mixed linear model approach. Boll number unit-area⁻¹ made the largest contribution to genotypic and genotype × environment ($G \times E$) variations for lint yield. Both boll number and lint percentage, and boll number and boll weight jointly accounted for more than 70% of the genotypic and $G \times E$ variations in lint yield. Ninety-nine percent of the genetic and phenotypic variation in lint yield could be explained by the three component traits, indicating that lint yield was mainly dependent on its three component traits. Small phenotypic variation in lint yield could be accounted for by effects of genotype, $G \times E$ interactions of boll number or boll number combined with other component trait(s) (Table 5). For boll number unit-area⁻¹ a wider distribution of genotypic contribution effects was detected than for lint percentage and boll weight in this study. Boll number and boll weight interacted to affect lint yield, indicating that balanced selection for boll weight and boll number is needed in high-yielding line development. Comparative results with other approaches were also discussed in this study.

Introduction

A complex trait like cotton (*Gossypium hirsutum* L.) lint yield depends on the joint contribution of its several component traits. The existence of correlations between a complex trait and its components is an indication of gene association or pleiotropism (Kebede et al., 2001; Dilday et al., 1990). Correlation analysis, multiple linear regression analysis, and path coefficient analysis are three common methods used to evaluate the relationships between a complex trait and its component traits (Bora et al., 1998; Ball et al., 2001; Cramer & Wehner, 2000; Samonte et al., 1998). A simple

phenotypic relationship between a complex trait and each of its component traits can be detected by correlation analysis. Multiple linear regression analysis can reveal single or joint contributions from component traits to the complex trait. Simple correlation coefficients can be partitioned into direct and indirect effects to the target trait by the path analysis (Wright, 1920). In most cases, both the complex trait and its component traits are random variables. Jobson (1991) proposed a conditional multiple linear regression method based on the normal conditional distribution theory. Several approaches have been proposed for analyzing a complex trait with multiplicative component traits (Sparnaaij

*The U.S. Government's right to retain a non-exclusive, royalty-free license in and to any copyright is acknowledged.

& Bos, 1993; Melchinger et al., 1994; Piepho, 1995). Both the complex trait and its component traits could be influenced by several factors such as environment, genotype, and genotype by environment interaction effects. Thus, it is necessary to partition the phenotypic values of the component traits into contributions due to different effects; however, the above methods are unable to analyze data for these complicated cases.

The total variation for a complex trait consists of variations in the component trait(s) as well as other factors (Jobson, 1991). Conditional analysis can remove the variation due to component trait(s) and obtain the remaining variation (conditional variances) (Graybill, 1976; Krzanowski, 1988; Jobson, 1991). Variation due to component traits can be obtained accordingly (Jobson, 1991). The expectation and the variance-covariance matrix for a conditional random vector can be derived (Graybill, 1976; Krzanowski, 1988); however, neither the conditional variance components nor the conditional effects under a mixed linear model are directly obtainable. Zhu (1995) proposed a method that could be used to analyze the conditional effects and the conditional variance components for single developmental traits. This conditional model approach also helped analyze the contribution of each component trait to a complex trait; however, this conditional model approach is only for single-variable analysis.

Cotton lint yield is determined by boll number unit-area⁻¹, boll weight, and lint percentage. Worley et al. (1974) reported that boll number unit-land-area⁻¹ was the largest contributor to lint yield. Maintaining a high lint percentage was necessary to ensure high lint yield (Culp & Harrell, 1974). Covariance component analyses have shown that lint yield was significantly dependent on each of these three yield components due to different genetic effects (Wu et al., 1995; Tang et al., 1996; McCarty et al., 1998). Conditional analysis by a mixed linear model showed that number of bolls per plant contributed ~45% of the variation in lint yield due to additive genetic effects and additive genetic by environment interaction effects, while 2% of the variation in lint yield was due to dominance effects and dominance by environment interaction effects (Zhu, 1995). If component traits independently contribute to a complex trait, then Zhu's (1995) approach can be used repeatedly to obtain multivariable conditional variance components. Previous research, however, has shown that yield components have significant genetic correlations (Wu et al., 1995; Tang et al., 1996; McCarty et al., 1998), which complicates the multivariable conditional analysis. Therefore, a more general conditional model

approach that allows for multiple conditional analyses for more than one component trait is needed.

In this study, Zhu's (1995) conditional model is extended for multiple conditional variance and conditional effect analyses. Two-year data for lint yield and three yield components from 188 upland cotton recombinant inbred (RI) lines and two parental lines were analyzed based on the extended model. Conditional variance components, contribution ratios, and contribution effects for different component traits were determined. These data should provide a better understanding of lint yield components and their relationship with lint yield and provide important information for the improvement of lint yield and other complex traits in breeding programs.

Materials and methods

Materials

One hundred eighty-eight RI (F₈) lines were developed by the single-hill procedure (Fehr, 1987) from the cross HS46 (P₁) × MARCABUCAG8US-1-8 (P₂). These RI lines with their two parental lines were planted at the Plant Science Research Center, Mississippi State, MS in 1999 and 2000. The experimental design was a randomized complete block with four replicates. Plot size was two rows 12 m in length with a between-row spacing of 0.97 m and an intra-row spacing of approximately 10 cm. The soil type was a Leeper silty clay loam (Fine, smectics, nonacid, thermic Vertic Epiaquepts). Planting date was 12 May for each of the two years. Standard cultural practices were followed throughout the growing season. A 50-boll sample was collected from each plot before machine harvest to determine boll weight (BW, g) and lint percentage (LP, %). Boll number per ha (BN) was calculated by dividing seed cotton yield per ha by average boll weight (Tang et al., 1996). Lint yield per ha (LY, kg) was determined by multiplying seed cotton yield by lint percentage. Thus the relationship between LY and three component traits is LY = BN × BW × LP.

Genetic model and statistical methods

The random effects linear model was used for data analysis for each trait,

$$\begin{aligned} \mathbf{y} &= \mathbf{1}\mu + \mathbf{U}_E\mathbf{e}_E + \mathbf{U}_G\mathbf{e}_G + \mathbf{U}_{GE}\mathbf{e}_{GE} + \mathbf{U}_B\mathbf{e}_B + \mathbf{e} \\ &= \mathbf{X}\mathbf{b} + \sum_{u=1}^5 \mathbf{U}_u\mathbf{e}_u \end{aligned} \quad (1)$$

where, μ is the fixed population mean; $\mathbf{1}$ is a vector with all 1520 elements equal to 1; \mathbf{e}_E is the random environmental effect vector with dimension of 2, $\mathbf{e}_E \sim N(\mathbf{0}, \sigma_E^2 \mathbf{I}_E)$; \mathbf{U}_E is the known design matrix for \mathbf{e}_E with dimension of 1520×2 ; \mathbf{e}_G is the random genotype effect vector with dimension of 190, $\mathbf{e}_G \sim N(\mathbf{0}, \sigma_G^2 \mathbf{I}_G)$; \mathbf{U}_G is the known design matrix for \mathbf{e}_G with dimension of 1520×190 ; \mathbf{e}_{GE} is the random $G \times E$ interaction effect vector with dimension of 380, $\mathbf{e}_{GE} \sim N(\mathbf{0}, \sigma_{GE}^2 \mathbf{I}_{GE})$; \mathbf{U}_{GE} is the known design matrix for \mathbf{e}_{GE} with dimension of 1520×380 ; \mathbf{e}_B is the random block effect vector with dimension of 8, $\mathbf{e}_B \sim N(\mathbf{0}, \sigma_B^2 \mathbf{I}_B)$; \mathbf{U}_B is the known design matrix for \mathbf{e}_B with dimension of 1520×8 ; \mathbf{e} is the random error with dimension of 1520, $\mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$; and \mathbf{I} is an identity matrix with different sizes.

Suppose that lint yield and the s component vectors have a joint normal distribution and that each of them follows the same mixed linear model (Equation 1), then the conditional vector of target trait $\mathbf{y}_{(T)}$ on the s component traits of $\mathbf{y}_{(1)}, \mathbf{y}_{(2)}, \dots, \mathbf{y}_{(s)}$ ($s = 1, 2, 3$) is defined as $\mathbf{y}_{(T|1,2,\dots,s)}$. This vector can be expressed in terms of matrices and vectors as follows:

$$\mathbf{y}_{(T|1,2,\dots,s)} = \mathbf{X}\mathbf{b}_{(T|1,2,\dots,s)} + \sum_{u=1}^5 \mathbf{U}_u \mathbf{e}_{u(T|1,2,\dots,s)} \quad (2)$$

where $\mathbf{y}_{(T|1,2,\dots,s)}$ is normally distributed with expectation

$$\begin{aligned} \mathbf{X}\mathbf{b}_{(T|1,2,\dots,s)} &= \mathbf{X}\mathbf{b}_{(T)} + (\mathbf{C}_{(T,1)} \mathbf{C}_{(T,2)} \dots \mathbf{C}_{(T,s)}) \\ &\times \sum_{(1,2,\dots,s)}^{-1} \begin{pmatrix} \mathbf{y}_{(1)} - \mathbf{X}\mathbf{b}_{(1)} \\ \dots \\ \mathbf{y}_{(s)} - \mathbf{X}\mathbf{b}_{(s)} \end{pmatrix} \end{aligned} \quad (3)$$

and variance-covariance matrix

$$\begin{aligned} \mathbf{V}_{(T|1,2,\dots,s)} &= \mathbf{V}_{(T)} - (\mathbf{C}_{(T,1)} \mathbf{C}_{(T,2)} \dots \mathbf{C}_{(T,s)}) \sum_{(1,2,\dots,s)}^{-1} \\ &\times (\mathbf{C}_{(1,T)} \mathbf{C}_{(2,T)} \dots \mathbf{C}_{(s,T)})^T = \sum_{u=1}^5 \sigma_{u(T|1,2,\dots,s)}^2 \mathbf{U}_u \mathbf{U}_u^T \end{aligned} \quad (4)$$

where $\mathbf{C}_{(i,j)} = \text{cov}(\mathbf{y}_{(i)}, \mathbf{y}_{(j)}) = \sum_{u=1}^5 \sigma_{u(i,j)} \mathbf{U}_u \mathbf{U}_u^T$ is the covariance matrix between traits $\mathbf{y}_{(i)}$ and $\mathbf{y}_{(j)}$ with dimension size of 1520×1520 ;

$$\sum_{(1,2,\dots,s)} = \text{var} \begin{pmatrix} \mathbf{y}_{(1)} \\ \dots \\ \mathbf{y}_{(s)} \end{pmatrix} = \begin{pmatrix} \mathbf{V}_{(1)} & \dots & \mathbf{C}_{(1,s)} \\ \dots & \dots & \dots \\ \mathbf{C}_{(s,1)} & \dots & \mathbf{V}_{(s)} \end{pmatrix},$$

which is a $sn \times sn$ symmetric matrix and has an inverse matrix (in this study, n is 1520 and s could be 1, 2, and 3); and $\mathbf{V}_{(i)} = \sum_{u=1}^5 \sigma_{u(i)}^2 \mathbf{U}_u \mathbf{U}_u^T$ is variance matrix of trait $\mathbf{y}_{(i)}$ (either target or component trait) with dimension size of 1520×1520 and $\sigma_{u(i)}^2$ is the variance component for trait $\mathbf{y}_{(i)}$.

The conditional variance in (4) and conditional random effects in (2) cannot be directly obtained due to five items of random effects in (2); however, the new vector in (5) is independent of all s ($s = 1, 2, 3$) component variables with the same multiple conditional variance and covariance matrix as in (4)

$$\begin{aligned} \mathbf{y}_{(T|1,2,\dots,s)}^* &= \mathbf{y}_{(T)} - (\mathbf{C}_{(T,1)} \mathbf{C}_{(T,2)} \dots \mathbf{C}_{(T,s)}) \\ &\times \sum_{(1,2,\dots,s)}^{-1} \begin{pmatrix} \mathbf{y}_{(1)} - \mathbf{X}\mathbf{b}_{(1)} \\ \dots \\ \mathbf{y}_{(s)} - \mathbf{X}\mathbf{b}_{(s)} \end{pmatrix} \end{aligned} \quad (5)$$

The vector in (5) can be expressed in terms of matrices and vectors as in Equation (1):

$$\begin{aligned} \mathbf{y}_{(T|1,2,\dots,s)}^* &= \mathbf{X}\mathbf{b}_{(T|1,2,\dots,s)}^* + \sum_{u=1}^5 \mathbf{U}_u \mathbf{e}_{u(T|1,2,\dots,s)}^* \\ &\sim N \left(\mathbf{X}\mathbf{b}_{(T|1,2,\dots,s)}^*, \sum_{u=1}^5 \sigma_{u(T|1,2,\dots,s)}^{*2} \mathbf{U}_u \mathbf{U}_u^T \right) \end{aligned} \quad (6)$$

Thus, our statistical idea is to construct the new vector in (5) to obtain $\sigma_{u(T|1,2,\dots,s)}^{*2}$ and $\mathbf{e}_{u(T|1,2,\dots,s)}^*$ in (6), which can be regarded as equivalent to the conditional variance component $\sigma_{u(T|1,2,\dots,s)}^2$ and the conditional effect vector $\mathbf{e}_{u(T|1,2,\dots,s)}$. However, $\mathbf{X}\mathbf{b}_{(T|1,2,\dots,s)}^*$ is not the conditional expectation of $\mathbf{y}_{(T|1,2,\dots,s)}$. In practice, it is computationally difficult to directly obtain the new vector in (5) because of the large inverse matrix of $\Sigma_{(1,\dots,s)}$ when s is greater than 1, a recursive approach is recommended (Wu, 2003). The new vector in (5) obtained by the recursive method can be used for detection of conditional variance components and conditional random effects. In this study, the new vector in (5) was obtained by the recursive approach (Wu, 2003; Wu et al., in press).

Conditional and unconditional variance components were estimated by minimum norm quadratic unbiased estimation (MINQUE) in which all prior values were set 1.0 (Zhu, 1989). Conditional and unconditional effects were predicted by the adjusted unbiased prediction (AUP) approach (Zhu, 1993). The phenotypic variance (V_P) was defined as $V_P = \sigma_G^2 + \sigma_{GE}^2 + \sigma_e^2$ where, σ_G^2 for genotypic effects, σ_{GE}^2 for $G \times E$ interaction effects, and σ_e^2 for random errors. Based on

unconditional and conditional variance and random effects, the following parameters were derived: (1) the quantity $1.0 - V_{P(LY|component(s))}/V_{P(LY)}$ is defined as the phenotypic contribution ratio $CR_{P(component(s) \rightarrow LY)}$ from single or multiple component traits to lint yield; (2) the quantity $1.0 - \sigma_{u(LY|component(s))}^2/\sigma_{uLY}^2$ is defined as the contribution ratio $CR_{u(component(s) \rightarrow LY)}$ from single or multiple component traits for the u -th random effect to lint yield (Zhu, 1995); (3) the ratio $(\sigma_{u(LY)}^2 - \sigma_{u(LY|component(s))}^2)/V_{P(LY)}$ is defined as the proportional contribution ratio $PCR_{u(component(s) \rightarrow LY)}$ to the phenotypic variance for lint yield from single or multiple component traits for the u -th random effect; and (4) the vector $\mathbf{e}_{u(LY)} - \mathbf{e}_{u(LY|component(s))}$ is defined as the u -th contribution effect vector, $\mathbf{e}_{u(component(s) \rightarrow LY)}$, from single or joint yield components to lint yield, where the arrow \rightarrow means “contributed to lint yield due to a single or multiple component trait(s)”. A re-sampling (jackknifing) method was applied to calculate the standard error (SE) for each parameter by removal of each block within year (Miller, 1974). An approximate t -test was used to evaluate the significance of each parameter (Miller, 1974). In this study the degrees of freedom were 7 because there were four replications in each of two years. All data analyses were conducted using a self-written program in C++.

Results

Means for lint yield and component traits within years and across years are summarized in Table 1. P₁ had more bolls ha⁻¹, higher lint percentage, and greater lint yield but a smaller boll weight than P₂. Mean lint yield and means of the three yield components over all recombinant inbred lines were between the two parental lines; however, they were more similar to P₂ than to P₁.

All correlation coefficients between the constructed random variables of lint yield (vector in (5)) and the component trait(s) were small ($-0.01 \sim 0.01$). This suggested that the new constructed random variables were independent of all component traits. Therefore, they were appropriate for multivariable conditional analyses.

Variance components and contribution ratios

Lint yield and the three component traits were controlled by both genotypic and $G \times E$ interaction effects (Table 2). Genotype effects were significantly

Table 1. Means for lint yield and yield components

Genotype	Year	BN(10 ⁵ /ha)	LP(%)	BW(g)	LY(kg/ha)
P ₁	1999	10.16	38.48	5.00	1951
	2000	8.21	38.23	4.93	1549
	Mean	9.18	38.36	4.97	1750
P ₂	1999	7.06	35.16	5.23	1293
	2000	6.17	35.61	5.40	1188
	Mean	6.62	35.39	5.32	1240
RI	1999	7.78	35.80	5.14	1427
	2000	6.68	35.50	5.25	1245
	Mean	7.23	35.65	5.20	1336

P₁: HS46; P₂: MARCABUCAG8US-1-88; RI: recombinant inbred; BN: boll number; LP: lint percentage; BW: boll weight; LY: lint yield.

Table 2. Proportions of estimated variance components to phenotypic variance for lint yield and yield component traits

Ratio	BN	LP	BW	LY
V_G/V_P	0.32**	0.69**	0.46**	0.32**
V_{GE}/V_P	0.18**	0.05*	0.07*	0.18**
V_e/V_P	0.50**	0.26**	0.47**	0.51**

BN: boll number; LP: lint percentage; BW: boll weight; LY: lint yield.

*Significant at 0.05 probability.

**Significant at 0.01 probability.

more important than $G \times E$ interaction effects for all traits, indicating these RI lines were relatively stable across the two years with respect to these traits. Experimental errors accounted for approximately 50% of the phenotypic variance for all traits except lint percentage (26%).

Conditional genotypic (or $G \times E$) variance for lint yield given component trait(s) measures the amount of genotypic (or $G \times E$) variance in lint yield not explained by the genotypic (or $G \times E$) effects of the component trait(s). Compared with the unconditional variances for lint yield, the conditional variances (both variance components and phenotypic variance) for lint yield given boll number were much smaller than those given lint percentage or boll weight (Table 3). For example, only 42% (5168/12327) of genotypic variance in lint yield was not explained by genotypic effects of boll number, while 89% (10931/12327) or 96% (11883/12327) of the genotypic variance in lint yield was not explained by genotypic effects of lint percentage or boll weight. The data suggested that boll number plays a more important

Table 3. Unconditional and conditional variance components for lint yield

	LY	LY-BN	LY-LP	LY-BW
σ_G^2	12327**	5168**	10931**	11883**
σ_{GE}^2	6884**	1103**	6866**	6184**
σ_e^2	19915**	4619**	19036**	18257**
V_P	39126**	10890**	36833**	36324**
	LY-BN&LP	LY-BN&BW	LY-LP&BW	LY-BN&LP&BW
σ_G^2	3327**	1890**	10412**	10*
σ_{GE}^2	590**	230**	6177**	85**
σ_e^2	3623**	865**	17573**	172**
V_P	7540**	2991**	34161**	276**

LY: lint yield; BN: boll number; LP: lint percentage; BW: boll weight σ_G^2 , σ_{GE}^2 , and σ_e^2 for LY are the unconditional genotypic, genotype \times environment interaction, and residual variance components for lint yield, respectively; V_P for LY is the unconditional phenotypic variance for lint yield.

σ_e^2 for LY—component trait(s) = genotypic variance in lint yield that cannot be explained by the genotypic effects of component trait(s); σ_{GE}^2 for LY—component trait(s) = genotype \times environment interaction variance in lint yield that cannot be explained by the genotype \times environment interaction effects of component trait(s); σ_e^2 for LY—component trait(s) = residual variance in lint yield that cannot be explained by the residual of component trait(s); V_P for LY—component trait(s) = phenotypic variance in lint yield that cannot be explained by phenotypic values of component trait(s).

*Significant at 0.05 probability level.

**Significant at 0.01 probability level.

role in lint yield than the other two component traits.

Contribution ratio (CR) for the u -th effects measures the proportion of variance in lint yield for the u -th effects explained by the respective effects in component trait(s). Boll number contributed 58% of the genotypic variance and 84% of $G \times E$ variance in lint yield (Table 4). Less than 20% of genotypic and $G \times E$

variance in lint yield was explained by lint percentage and boll weight, separately and jointly (Table 4). More than 70% of genotypic variance in lint yield was explained jointly by boll number and lint percentage (73%) or boll number and boll weight (85%). More than 90% of $G \times E$ variance in lint yield was explained jointly by boll number and lint percentage (91%) or boll number and boll weight (97%). Ninety nine percent of genotypic and $G \times E$ variances in lint yield was accounted for by the three component traits, suggesting that lint yield in this RI population was mainly dependent on these three component traits. A large proportion of residual variance in lint yield was related to the residual variance in boll number alone or boll number combined with other component traits (Table 4). For example, 77% of the residual variance in lint yield was contributed by boll number. The residual variance in lint yield could be explained by boll number and lint percentage, or boll number and boll weight with 82 and 96% of the variance, respectively. A large proportion of phenotypic variance in lint yield was due to boll number (72%) or boll number combined with other component trait(s) (81, 92, or 99% for lint percentage, boll weight, and lint percentage with boll weight, respectively).

Based on the conditional and unconditional variances for lint yield, the phenotypic contribution ratio (CR_P) can be partitioned into different proportional contribution ratios (PCR). The parameter PCR_u measures the percentage of the phenotypic variance explained by the u -th effects in component trait(s). Genotypic and $G \times E$ effects of boll number contributed 18 and 15% of the phenotypic variance in lint yield; however, 39% of the phenotypic variance in lint yield was from the residual effects for boll number (Table 5). Genotypic effects for boll number and lint percentage, and boll number and boll weight made similar

Table 4. Contribution ratios (CR) to lint yield from yield components

CR	BN	LP	BW	BN & LP	BN & BW	LP & BW	BN & LP & BW
G	0.58**	0.11*	0.04	0.73**	0.85**	0.16**	1.00**
GE	0.84**	0.00	0.10*	0.91**	0.97**	0.10	0.99**
e	0.77**	0.04**	0.08**	0.82**	0.96**	0.12**	0.99**
P	0.72**	0.06**	0.07**	0.81**	0.92**	0.13**	0.99**

BN: boll number; LP: lint percentage; BW: boll weight; CR_u for component trait(s) = proportion of variance in lint yield for u -th effects accounted by the u -th effects of component trait(s); example for calculation: CR_G for BN = $1.0 \times 5168 / 12327 = 0.58$.

*Significant at 0.05 probability level.

**Significant at 0.01 probability level.

Table 5. Proportional contribution ratios (PCR) to lint yield from yield components

PCR	BN	LP	BW	BN & LP	BN & BW	LP & BW	BN & LP & BW
G	0.18**	0.04**	0.01	0.23**	0.27**	0.05**	0.31**
GE	0.15**	0.00	0.02*	0.16**	0.17**	0.02	0.17**
e	0.39**	0.02**	0.04**	0.42**	0.49**	0.06**	0.50**

BN: boll number; LP: lint percentage; BW: boll weight.

PCR_{*u*} for component trait(s) = proportion of phenotypic variance in lint yield accounted by the *u*-th effects of component trait(s); example for calculation: PCR_G for BN = (12327–5168)/39125.96 = 0.18.

*Significant at 0.05 probability level.

**Significant at 0.01 probability level.

contribution to the phenotypic variation in lint yield (23 and 27%, respectively). This was also observed for the $G \times E$ effects.

Contribution effects

Genotype and $G \times E$ interaction contribution effects to lint yield were predicted. Correlation coefficients between unconditional effects for lint yield and contribution effects to lint yield are summarized in Table 6. The squared correlation coefficients were similar to the contribution ratios for the yield component traits (Table 4), indicating that the contribution effect prediction was in good agreement with the contribution ratio estimation. This suggests the model and the method used in this study were appropriate.

Genotype contribution effects to lint yield due to boll number, lint percentage, and boll weight ranged between –263 and 337 (kg/ha), –93 and 112 (kg/ha), and –71 and 69 (kg/ha), respectively. The ranges of joint contribution effects for boll number and other component trait(s) were similar to the range for boll number. The genotypic effect range due to boll number was wider than those due to the other two component traits (data not provided). The range for joint genotypic contribution effects was wider than those due to a single trait. The results were in agreement with the conditional variance estimates (Table 3).

Discussion

Cotton lint yield is determined by boll number, boll weight, and lint percentage. Both the unconditional linear regression method (Myers, 1990) and conditional linear regression method (Jobson, 1991) can be used to dissect the relationship between lint yield and its component trait(s). Numerically, the coefficients of multiple determination (R^2) obtained by these two linear regression methods should be equivalent although the mean square errors (MSE) are different; however, the R^2 values obtained by these two methods only represent the phenotypic relationships between lint yield and its component trait(s). In this study, both lint yield and its component traits were significantly affected by genotype, $G \times E$ interaction, and experimental error (Table 2); however, the above methods were not able to analyze the data under the mixed linear model used in this study. A mixed model based conditional approach was used in this study to determine the conditional variance components, contribution ratio, proportional contribution ratio, and the contribution effects. The phenotypic contribution ratios obtained by this mixed-model-based conditional approach (bottom line of Table 4) are equivalent to the R^2 values obtained from both conditional and unconditional linear regression analyses (first two lines of Table 7). The slight differences might be due to the fact that environmental and block variance components were not included in the

Table 6. Squared coefficients of correlation between contribution effects for component traits and unconditional effects of lint yield

Effect	BN	LP	BW	BN & LP	BN & BW	LP & BW	BN & LP & BW
Genotype	0.60	0.11	0.04	0.74	0.86	0.11	1.00
$G \times E$	0.83	0.06	0.14	0.91	0.97	0.15	0.99

Note: all correlation coefficients are significant at 0.01 level of probability; BN: boll number per unit area; LP: lint percentage; BW: boll weight.

Table 7. Coefficients of determination (R^2) for lint yield on yield component(s) obtained by conditional and unconditional linear regression models

	BN	LP	BW	BN & LP	BN & BW	LP & BW	LP & BN & BW
Unconditional	0.65	0.10	0.06	0.77	0.88	0.15	0.99
Conditional ^a	0.65	0.10	0.06	0.77	0.88	0.15	0.99
Piepho's model	0.77	0.13	0.16	0.89	0.92	0.28	1.00

BN: boll number per unit area; LP: lint percentage; BW: boll weight.

^aUsing Jobson (1991) model.

phenotypic variance. Our results were also compared with those obtained by Piepho's approach (1995) (bottom line of Table 7). The contribution ratios obtained by his approach were in a good agreement with our phenotypic contribution ratios for boll number or boll number combined with other component traits (Tables 4 and 7), indicating that Piepho's approach is workable for detecting the relationship between a complex trait and its multiplicative component traits using the logarithmic transformation. The slight difference between Piepho's model and regression models or our model may be due to the fact that additive models rather than a multiplicative model were applied to regression models and our model.

Contribution ratio (CR) for the u -th effects measures the proportion of variance in lint yield for the u -th effects explained by the respective effects in component trait(s). Boll number unit-area⁻¹ made the largest contribution to genotypic and genotype \times environment ($G \times E$) variance for lint yield. Both boll number and lint percentage, and boll number and boll weight jointly accounted for 73% or more of the contributions to genotypic and $G \times E$ variations in lint yield. The results agreed with the report by Worley et al. (1974). Based on the conditional and unconditional variance components for lint yield, the phenotypic contribution ratio (CR_p), Table 4, could be partitioned into proportional contribution ratios (PCR). The majority of phenotypic variation in lint yield could be explained by phenotypic variation of boll number, or boll number with other component trait(s); however, more than 50% of the variation in lint yield was due to residual effects in boll number (0.39/0.72), or boll number with other component trait(s) (0.42/0.81, 0.49/0.92, or 0.50/0.99, for lint percentage, boll weight, and lint percentage with boll weight, respectively) (Tables 4 and 5). The data indicated that it could be less efficient to predict the phenotypic value of lint yield for a genotype than to the predict genotypic value based on genotypic values of boll number, or boll number with other component traits because

there were large residual contribution ratios and large residual variation in lint yield ($0.51 = 19915/39126$) (Table 3).

Another important contribution of this study is that the conditional model allows the prediction of the genotype and $G \times E$ interaction contribution effects for yield component trait(s). In this study the genotype contribution effects for boll number were greater than those of boll weight and lint percentage (Table 4). Correlation coefficients between unconditional effects for lint yield and contribution effects for component trait(s) were similar to the contribution ratios for the yield component traits (Tables 4 and 6). This indicates that prediction of contribution effects was in a good agreement with estimation of contribution ratios.

The mixed-model-based conditional approach used in this study offers several advantages over previous methods: (1) the mixed linear model is extendable to the other genetic models such as the additive-dominance model (Cockerham, 1980), or the plant seed model (Zhu & Weir, 1994), which may provide information relevant for selection study; (2) this conditional approach allows for the partition of the phenotypic contribution ratio into different contribution ratios and prediction of contribution effects under a specific mixed linear model; (3) this approach may provide an important way to evaluate the relationship between a complex trait and its hierarchical component traits (Worley et al., 1976; Coyle & Smith, 1997).

If two component traits having contributions to lint yield are due to a specific genetic effects (i.e. genotypic effects) and are independent, then the joint contribution ratio due to this genetic component will be approximately equal to the summation of the two single contribution ratios. Joint contribution ratios from boll number and boll weight were greater than the sum of single contribution ratios from boll weight and boll number for genotypic effects (Table 4). However, lint percentage appeared to be independent of boll number or boll weight for its effect on lint yield because their summation nearly equaled the joint

contribution ratio. On the other hand, deviation of the joint contribution effect for two component traits from the sum of the two single contribution effects can be equivalently considered as the interaction contribution effects, namely, $e_{u(\text{Component1, Component2} \rightarrow LY)} - e_{u(\text{Component1} \rightarrow LY)} - e_{u(\text{Component2} \rightarrow LY)}$. Genotypic interaction contribution effects between boll number unit-area⁻¹ and boll weight could be seen (data not shown), indicating that balanced selection for boll weight and boll number is important for developing high yielding lines, although boll weight itself only made a small contribution to lint yield.

In practice, it is almost impossible to evaluate all plants for boll number for each plot with such a large number of RI lines. In this study, the total boll number ha⁻¹ was calculated following the method used by Tang et al. (1996). Boll number could also be obtained by multiplying the number of standing plants unit-area⁻¹ with mean boll number per plant. The results may differ from this study if different genetic materials are used; however, this study provides a new way to evaluate the genetic relationship for complex traits in plant breeding.

Acknowledgements

The authors want to express their special appreciations to Drs. Daryl Bowman, Gerald Myers, Jinfa Zhang, Clarence Watson, and Fred Bourland for careful peer review and helpful suggestions and comments on the earlier version of this manuscript. The authors also like to thank one editor and two anonymous reviewers for their useful comments and suggestions for improving the quality of this paper.

References

- Ball, R.A., R.W. McNew, E.D. Vories, T.C. Keisling & L.C. Purcel, 2001. Path analyses of population density effects on short-season soybean yield. *Agron J* 93: 187–195.
- Bora, G.C., S.N. Gupta, Y.S. Tomer & S. Singh, 1998. Genetic variability, correlation and path analysis in faba bean (*Vicia faba*). *Indian J Agric Sci* 68: 212–214.
- Cockerham, C.C., 1980. Random and fixed effects in plant genetics. *Theor Appl Genet* 56: 119–131.
- Coyle, G.G. & C.W. Smith, 1997. Combining ability for within-boll yield components in cotton, *Gossypium hirsutum* L. *Crop Sci* 37: 1118–1122.
- Culp, T.W. & D.C. Harrell, 1975. Influence of lint percentage, boll size, and seed size on lint yield of upland cotton with high fiber strength. *Crop Sci* 15: 741–746.
- Cramer, C.S. & T.C. Wehner, 2000. Path analysis of the correlation between fruit number and plant traits of cucumber populations. *Hort Sci* 35: 708–711.
- Dilday, R.H., M.A. Mgonja, S.A. Amosilpa, F.C. Collins & B.R. Wells, 1990. Plant height vs. mesocotyl and coleoptile elongation in rice: Linkage or pleiotropism. *Crop Sci* 30: 815–818.
- Fehr, W., 1987. Principles of Cultivar Development: Volume 1, Theory and Technique, Macmillan Publishing Co., New York.
- Graybill, F.A., 1976. Theory and Application of Linear Model, Duxbury Press, Boston, MA.
- Jobson, J.D., 1991. Applied Multivariate Data Analysis Volume II: Categorical and Multivariate Methods. Springer-Verlag, New York.
- Kebede, H., P.K. Subudhi, D.T. Rosenow & H.T. Nguyen, 2001. Quantitative trait loci influencing drought tolerance in grain sorghum (*Sorghum bicolor* L. Moench). *Theor Appl Genet* 103: 266–276.
- Krzyszowski, W.J., 1988. Principles of Multivariate Analysis. Oxford Science Publications, New York.
- McCarty, J.C. Jr., J.N. Jenkins & J. Zhu, 1998. Introgression of day-neutral genes in primitive cotton accessions: I. Genetic variances and correlations. *Crop Sci* 38: 1425–1428.
- Melchinger A.E., M. Singh, W. Link, H.F. Utz & E. von Kittlitz, 1994. Heterosis and gene effects of multiplicative characters: Theoretical relationships and experimental results from *Vicia faba* L. *Theor Appl Genet* 88: 343–348.
- Miller, R.G., 1974. The jackknife: A review. *Biometrika* 61: 1–15.
- Myers, R.H., 1990. Classical and Modern Regression with Applications, PWS-KENT Publ. Co., Boston, MA.
- Piepho, H.P., 1995. A simple procedure for yield component analysis. *Euphytica* 84: 43–48.
- Samonte, S.O.P., L.T. Wilson & A.M. McClung, 1998. Path analyses of yield and yield-related traits of fifteen diverse rice genotypes. *Crop Sci* 38: 1130–1136.
- Sparnaaij, L.D. & I. Bos, 1993. Component analysis of complex characters in plant breeding I. Proposed method for quantifying the relative contribution of individual components to variation of the complex character. *Euphytica* 70: 225–235.
- Tang, B., J.N. Jenkins, C.E. Watson, J.C. McCarty & R.G. Creech, 1996. Evaluation of genetic variances, heritabilities, and correlations for yield and fiber traits among cotton F₂ hybrid populations. *Euphytica* 91: 315–322.
- Worley, S., T.W. Culp & D.C. Harrell, 1974. The relative contributions of yield components to lint yield of upland cotton, *Gossypium hirsutum* L. *Euphytica* 23: 399–403.
- Worley, S., H.H. Ramey, D.C. Harrell & T.W. Culp, 1976. Ontogenetic model of cotton yield. *Crop Sci* 16: 30–34.
- Wright, S., 1920. The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. *Proc Nat Acad Sci USA* 6: 320–332.
- Wu, J., 2003. Genetic variation, conditional analysis, and QTL mapping for agronomic and fiber traits in upland cotton. Ph.D. Dissertation, Mississippi State University, MS.
- Wu, J., D. Wu, J.N. Jenkins & J.C. McCarty Jr., in press. A recursive approach to detect multivariate conditional variance components and random effects. *Comput Stat Data Anal*.
- Wu, J., J. Zhu, F. Xu & D. Ji, 1995. Analysis of genetic effect × environment interactions for yield traits in upland cotton (Chinese). *Heredity* 17(5): 1–4.

- Zhu, J., 1989. Estimation of genetic variance components in the general mixed model. Ph.D. Dissertation, North Carolina State University, Raleigh, NC.
- Zhu, J., 1993. Methods of predicting genotype value and heterosis for offspring of hybrids. *J Biomath (Chinese)* 8(1): 32–44.

- Zhu, J., 1995. Analysis of conditional effects and variance components in developmental genetics. *Genetics* 141: 1633–1699.
- Zhu, J. & B.S. Weir, 1994. Analysis of cytoplasmic and maternal effects: I. A genetic model for diploid plant seeds and animals. *Theor Appl Genet* 89: 153–159.