

# Multivariable Conditional Analysis for Complex Trait and Its Components

WEN Yong-Xian<sup>1,2</sup>, ZHU Jun<sup>1,①</sup>

(1. Department of Agronomy, Zhejiang University, Hangzhou 310029, China;

2. College of Computer and Information Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China)

**Abstract**: Methods of multivariable conditional analysis were proposed for analyzing contribution of component traits to complex target trait based on mixed linear model approaches. The contribution ratio and the contributed genetic effect were defined. The contribution ratio could measure the proportion of genetic variation of the given component trait on the target trait. The contributed genetic effect could quantify genetic effect of the given component traits contributed to the target trait. A worked example from the cotton data is given to illustrate the application of the new methods for analyzing the contribution of three yield components to lint yield.

**Key words**: conditional genetic variance components; conditional genetic effects; contribution ratio, contributed genetic effect; complex traits

## 综合性状及其分量的多元条件分析

温永仙<sup>1,2</sup>, 朱 军<sup>1,①</sup>

(1. 浙江大学农学系, 杭州 310029;

2. 福建农林大学计算机与信息科学学院, 福州 350002)

**摘 要**: 提出基于混合线性模型的多元条件分析方法,用于分析复杂综合性状的分量对其目标性状的贡献。定义了贡献率和贡献遗传效应两个概念,贡献率测定给定分量性状的遗传变异对目标性状的贡献比率,贡献遗传效应衡量给定分量性状的遗传效应对目标性状的贡献值。运用所提出的新方法分析了棉花籽棉产量的3个分量对产量的贡献。

**关键词**: 条件方差分量; 条件遗传效应; 贡献率; 贡献遗传效应; 产量性状

中图分类号: Q348 文献标识码: A 文章编号: 0379-4172(2005)03-0289-08

In agronomy, a complex quantitative trait like yield is usually consisted of several component traits. The components, which are correlated each other, have their contribution in complicated ways to the target complex trait. In order to efficiently

improve the complex trait, it is of importance to understand the contribution due to the component traits.

Up to now, analysis methods of correlation, regression, and path analysis have been used for

收稿日期 2004-02-09; 修回日期 2004-07-23

基金项目: 国家自然科学基金(编号: 39970473) [Supported by the National Natural Science Foundation of China(No. 39970473)]

作者简介: 温永仙(1966-), 女, 副教授, 浙江大学进修教师

① 通讯作者。朱 军(1949-), 男, 教授, 博导, 研究方向: 数量遗传学。E-mail: jzhu@zju.edu.cn; Tel: 0571-86971731

studying relationship between a target complex trait and its components<sup>[1-4]</sup>. Correlation and covariance analysis are usually concerned with the relationship for the complex trait with its components, rather than with the contribution of component traits to the complex trait. The multiple linear regression method is superior to correlation analysis for studying the predictability of the complex trait by its component traits. When using regression approaches for analyzing determination of component traits, there will be two limitations: Firstly, the contribution is measured by partial regression coefficient which depends on other variables in the regression model; Secondly, the magnitude of each regression coefficients also depends on the unit of independent variables. The relative importance can not be compared by the regression coefficients of component traits. Path analysis was developed by Wright<sup>[5,6]</sup> in order to reveal the correlation relationships among several components and their effects on a complex trait. By path analysis, simple correlation coefficient between a complex trait and its component can be divided into direct and indirect effects on the complex trait. However, path coefficients are standardized partial regression coefficients in multiple regression model. The values of partial regression coefficients are partially determined by the number of independent variables included in the model.

In the 1970's, statisticians developed some methods for analyzing mixed linear models, which can be used for dealing with complicated genetic problems in quantitative genetics. Zhu<sup>[7]</sup> proposed a method for analyzing conditional genetic effects and variance components by mixed linear model approaches, which can be applied in making inference on contribution of single component to a complex target trait. Since there is correlation among genetic effects of component traits due to the pleiotropy and linkage of genes, the total contribution of multiple component traits is not just a sum of their single contributions.

In the present paper, statistical methods were

proposed for analyzing multivariable conditional genetic effects and variance components based on mixed linear models. Methods of measuring the contribution of component traits to the complex target trait were developed. An example of yield traits in cotton was used to illustrate the use of these new methodology and their applications in plant breeding.

## 1 Model and Methodology

Most genetic models can be expressed by a general form of mixed linear model

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \sum_{u=1}^{m-1} \mathbf{U}_u \mathbf{e}_u + \mathbf{e}_m \quad (1)$$

$$\sim MVN(\mathbf{X}\mathbf{b}, \mathbf{V} = \sum_{u=1}^{m-1} \sigma_u^2 \mathbf{U}_u \mathbf{R}_u \mathbf{U}'_u + \sigma_m^2 \mathbf{I})$$

where  $\mathbf{y}$  is the vector ( $n \times 1$ ) of phenotypic value with mean  $\mathbf{X}\mathbf{b}$  and variance  $\mathbf{V}$ ;  $\mathbf{b}$  is the vector ( $p \times 1$ ) of fixed effects;  $\mathbf{X}$  is the known incidence matrix ( $n \times p$ ) relating to the fixed effects;  $\mathbf{e}_u$  ( $u = 1, 2, \dots, m-1$ ) is the vector ( $q_u \times 1$ ) of the  $u$ -th random factor,  $\mathbf{e}_u \sim (0, \sigma_u^2 \mathbf{R}_u)$ ,  $\mathbf{R}_u$  is the constant matrix describing the relationship of  $\mathbf{e}_u$ ;  $\mathbf{U}_u$  is the known coefficient matrix relating to the random vector  $\mathbf{e}_u$ ,  $\mathbf{U}'_u$  is the transpose matrix of  $\mathbf{U}_u$ ;  $\mathbf{e}_m$  is the vector ( $n \times 1$ ) of the residual random effects with  $\mathbf{e}_m \sim (0, \sigma_m^2 \mathbf{I})$ .

For two traits, covariance between observation vectors  $\mathbf{y}_1$  and  $\mathbf{y}_2$  are as follows,

$$\text{Cov}(\mathbf{y}_1, \mathbf{y}'_2) = \sum_{u=1}^{m-1} \sigma_{u_1/u_2} \mathbf{U}_u \mathbf{R}_u \mathbf{U}'_u + \sigma_{\sigma_{m1/m2}} \mathbf{I}$$

$$= \mathbf{C}_{(1,2)} \quad (2)$$

where  $\mathbf{C}_{1,2}$  is symmetric ( $\mathbf{C}_{(1,2)} = \mathbf{C}_{(2,1)}$ ).

The variance components in model (1) can be estimated by mixed linear model approaches such as maximum likelihood (ML)<sup>[8]</sup>, restricted maximum likelihood (REML)<sup>[9,10]</sup> or minimum norm quadratic unbiased estimation (MINQUE)<sup>[11,12]</sup>. The method suggested by Zhu and Weir<sup>[13-15]</sup> can be used for estimating covariance components between different traits.

The estimates of fixed effects  $\mathbf{b}$  can be usual-

ly obtained by the general least squares estimation method  $\hat{\mathbf{b}}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ . Random genetic effects  $\mathbf{e}_u$  can be predictable without bias by the best linear unbiased prediction (BLUP) method<sup>[16]</sup>, the linear unbiased prediction (LUP) method<sup>[13,17]</sup>, and the adjusted unbiased prediction (AUP) method<sup>[14,15,18]</sup>.

If a trait  $\mathbf{y}_{(C)}$  is one component of a target complex trait  $\mathbf{y}_{(T)}$ , conditional random variables of  $\mathbf{y}_{(T)} | \mathbf{y}_{(C)}$  for the phenotypic mean of the target trait  $\mathbf{y}_{(T)}$  conditional upon the phenotypic mean of component trait  $\mathbf{y}_{(C)}$  have conditional distribution

$$\mathbf{y}_{(T)} | \mathbf{y}_{(C)} = \mathbf{X}\mathbf{b}_{(T|C)} + \sum_{u=1}^m \mathbf{U}_u \mathbf{e}_{u(T|C)} \sim MVN(\mathbf{X}\mathbf{b}_{(T|C)}, \mathbf{V} = \sum_{u=1}^{m-1} \sigma_{u(T|C)}^2 \mathbf{U}_u \mathbf{U}_u' + \sigma_{m(T|C)}^2 \mathbf{I}). \quad (3)$$

Since conditional  $\mathbf{y}_{(T)} | \mathbf{y}_{(C)}$  is independent of  $\mathbf{y}_{(C)}$ , conditional random effects  $\mathbf{e}_{u(T|C)}$  and conditional variance components  $\sigma_{u(T|C)}^2$  contain extra genetic variation without the influence of the given component trait  $\mathbf{y}_{(C)}$  on the target trait  $\mathbf{y}_{(T)}$ . Zhu<sup>[7]</sup> proposed a mixed model approach for analyzing conditional variance components and random effects.

The proportion of  $\sigma_{u(T|C)}^2 / \sigma_{u(T)}^2$  might uncover the contribution of gene effects without the influence of the given component trait  $\mathbf{y}_{(C)}$  on the target complex trait  $\mathbf{y}_{(T)}$ , and  $CR_{u(C \rightarrow T)} = 1.0 - \sigma_{u(T|C)}^2 / \sigma_{u(T)}^2$  defined as contribution ratio could be used for measuring the contribution proportion of the  $u$ -th genetic variation of the given component trait  $\mathbf{y}_{(C)}$  on the target trait  $\mathbf{y}_{(T)}$ .  $\mathbf{e}_{u(C \rightarrow T)} = \mathbf{e}_{u(T)} - \mathbf{e}_{u(T|C)}$  defined as contributed effects could be used for measuring the  $u$ -th vector of genetic effects of the given component trait  $\mathbf{y}_{(C)}$  on the target trait  $\mathbf{y}_{(T)}$ .

For the case of two components of a complex trait, conditional random vector  $\mathbf{y}_{(T)} | \mathbf{y}_{(C_1)}, \mathbf{y}_{(C_2)}$  is independent of both phenotypic values  $\mathbf{y}_{(C_1)}$  and  $\mathbf{y}_{(C_2)}$ . Therefore, conditional random variables can be used for detecting conditional genetic effects as well as conditional variance components.

Conditional phenotypic vector  $\mathbf{y}_{(T|C_1, C_2)}$  is defined as

$$\mathbf{y}_{(T|C_1, C_2)} = \mathbf{X}\mathbf{b}_{(T|C_1, C_2)} + \sum_{u=1}^m \mathbf{U}_u \mathbf{e}_{u(T|C_1, C_2)} \sim MVN(\mathbf{X}\mathbf{b}_{(T|C_1, C_2)}, \mathbf{V}_{(T|C_1, C_2)} = \sum_{u=1}^{m-1} \sigma_{u(T|C_1, C_2)}^2 \mathbf{U}_u \mathbf{R}_u \mathbf{U}_u' + \sigma_{m(T|C_1, C_2)}^2 \mathbf{I}), \quad (4)$$

where  $\mathbf{b}_{(T|C_1, C_2)}$  is the vector ( $p \times 1$ ) of conditional fixed effects,  $\mathbf{e}_{u(T|C_1, C_2)}$  ( $u=1, 2, \dots, m$ ) is the vector ( $q_u \times 1$ ) of the  $u$ -th conditional random effects,  $\mathbf{e}_{u(T|C_1, C_2)} \sim (\mathbf{0}, \sigma_{u(T|C_1, C_2)}^2)$ . Since  $Cov(\mathbf{e}_{u(T|C_1, C_2)}, \mathbf{e}'_{u'(C_1)}) = Cov(\mathbf{e}_{u(T|C_1, C_2)}, \mathbf{e}'_{u'(C_2)}) = 0$ , conditional genetic effect  $\mathbf{e}_{u(T|C_1, C_2)}$  and conditional variance  $\sigma_{u(T|C_1, C_2)}^2$  can be used for illuminating the effects of gene expression without the influence of the given component traits  $\mathbf{y}_{(C_1)}$  and  $\mathbf{y}_{(C_2)}$  on the target traits  $\mathbf{y}_{(T)}$ .

Estimation of conditional genetic variance components or prediction of conditional genetic effects can be derived indirectly for analyzing a new random vector  $\mathbf{y}_{(***)}$  defined as

$$\mathbf{y}_{(***)} = \mathbf{y}_{(T)} - [\mathbf{C}_{(T, C_1)} \quad \mathbf{C}_{(T, C_2)}] \begin{bmatrix} \mathbf{V}_{(C_1)} & \mathbf{C}_{(C_1, C_2)} \\ \mathbf{C}_{(C_2, C_1)} & \mathbf{V}_{(C_2)} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y}_{(C_1)} - \mathbf{X}\mathbf{B}_{(C_1)} \\ \mathbf{y}_{(C_2)} - \mathbf{X}\mathbf{B}_{(C_2)} \end{bmatrix}. \quad (5)$$

with variance-covariance matrix

$$Var(\mathbf{y}_{(***)}) = \mathbf{V}_{(T)} - [\mathbf{C}_{(T, C_1)} \quad \mathbf{C}_{(T, C_2)}] \begin{bmatrix} \mathbf{V}_{(C_1)} & \mathbf{C}_{(C_1, C_2)} \\ \mathbf{C}_{(C_2, C_1)} & \mathbf{V}_{(C_2)} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{C}_{(C_1, T)} \\ \mathbf{C}_{(C_2, T)} \end{bmatrix}, \quad (6)$$

which is equivalent to the conditional variance-covariance matrix of two-variable conditional random vector  $\mathbf{y}_{(T)} | \mathbf{y}_{(C_1)}, \mathbf{y}_{(C_2)}$ . Since  $Cov(\mathbf{y}_{(***)}, \mathbf{y}'_{(C_1)}) = Cov(\mathbf{y}_{(***)}, \mathbf{y}'_{(C_2)}) = \mathbf{0}$ , random vector  $\mathbf{y}_{(***)}$  is independent to both phenotypic values  $\mathbf{y}_{(C_1)}$  and  $\mathbf{y}_{(C_2)}$ . Unknown parameters (variance matrix, covariance matrix and fixed effect vector) in Equation (6) can be replaced by their unbiased estimates in practice.

Random vector  $\mathbf{y}_{(***)}$  is then fitted to the mixed linear model,

$$\mathbf{y}_{(***)} = \mathbf{X}\mathbf{b}_{(***)} + \sum_{u=1}^m \mathbf{U}_u \mathbf{e}_{u(***)}$$

$$\begin{aligned} & \sim MVN(\mathbf{Xb}_{(***)}, \mathbf{V}_{(***)}) \\ & = \sum_{u=1}^{m-1} \sigma_{u(***)}^2 \mathbf{U}_u \mathbf{R}_u \mathbf{u}_u^T + \sigma_{m(***)}^2 \mathbf{I} \end{aligned} \quad (7)$$

which turns out to have variance-covariance matrix equivalent to the two-variable conditional variance-covariance matrix  $\mathbf{V}_{(T|C_1, C_2)}$ .

Estimated conditional variance  $\hat{\sigma}_{u(***)}^2$  can be served as unbiased estimate of the conditional variance components  $\sigma_{u(T|C_1, C_2)}^2$ . Since  $\sigma_{u(***)}^2$  is equivalent to  $\sigma_{u(T|C_1, C_2)}^2$ , genetic effects  $\mathbf{e}_{u(***)}$  also have an equivalency to the conditional genetic effects  $\mathbf{e}_{u(T|C_1, C_2)}$ .

The two-variable proportion of  $\sigma_{u(T|C_1, C_2)}^2 / \sigma_{u(T)}^2$ , the contribution ratio  $CR_{u(C_1, C_2 \rightarrow T)} = 1.0 - \sigma_{u(T|C_1, C_2)}^2 / \sigma_{u(T)}^2$  and the contributed effects  $\mathbf{e}_{u(C_1, C_2 \rightarrow T)} = \mathbf{e}(T) - \mathbf{e}_{(T|C_1, C_2)}$  are defined the same ways as for the one-variable conditional analysis.

## 2 A worked example

Yield data<sup>[19]</sup> from an experiment on upland cotton (*Gossypium hirsutum* L.) conducted in two years were used as an example to illustrate the application of the new analysis method for studying the contribution to lint yield (LY) by three yield components: boll number per plant (BN), boll size (BW) and lint percentage (LP). Ten parents (five as females and another five as males) and their 20  $F_1$ 's and  $F_2$ 's were conducted according to NC design II with three randomized complete blocks carried out in the experimental station of Zhejiang Agricultural University.

Variance and covariance components were estimated by MINQUE (1) method<sup>[15]</sup>. Genetic correlation coefficients were calculated from estimated variances and covariances. Jackknifing over blocks within year was used for obtaining estimates and standard errors of parameters<sup>[15, 20]</sup>. A *t*-test with *df* = 5 was employed for testing significance of genetic parameters studied.

The estimated proportion of genetic variance to phenotype variance was presented in Table 1

for lint yield and its components. It was suggested that lint yield controlled by genetic main effects (additive and dominance) as well as *GE* interaction. The dominance environment variance of boll number was larger as compared to other genetic variance components. It was indicated that the utilization of heterosis should be considered in special environments. For boll size, the proportion of dominance variance to phenotype variance was larger than other genetic components, so that heterosis could be expected. Lint percentage was mostly controlled by additive effects, indicating that selection in early generation could result in apparent genetic gain. The residual effects were the major cause for variation of lint yield and its two components (boll number and boll size).

**Table 1 Estimated proportion of genetic variance to phenotype variance for lint yield and its components boll number, boll size and lint percentage**

Parameter	Lint yield	Boll number	Boll size	Lint%
$V_A/V_P$	0.193 **	0.152 **	0.000	0.561 **
$V_D/V_P$	0.233 **	0.083 *	0.351 **	0.026
$V_{AE}/V_P$	0.110 *	0.046	0.073 *	0.034
$V_{DE}/V_P$	0.137 *	0.286 *	0.079 *	0.095 *
$V_e/V_P$	0.326 **	0.432 **	0.497 **	0.284 **

\* and \*\* significant at 0.05 and 0.01 levels, respectively. Variances represent:  $V_A$  = additive variance,  $V_D$  = dominance variance,  $V_{AE}$  = additive × environment interaction variance,  $V_{DE}$  = dominance × environment interaction variance,  $V_e$  = residual variance,  $V_P$  = phenotypic variance.

The genetic correlation coefficients were estimated among lint yield and its components (Table 2). Lint yield had highly significant positive phenotypic correlation to boll number, but their additive and dominance correlation was much larger. Although, there was large phenotypic correlation between lint yield and lint percentage, correlation coefficients were positive for additive effects and *DE* interaction effects but negative for dominance effects between these two traits. Even though there was small phenotypic correlation between lint yield and boll size, there still existed strong positive correlation of dominance effects for these two traits.

There was highly positive dominance correlation between boll number and boll size. Phenotypic correlation was not large between boll number and lint percentage, which could be due to the reason of negative correlation of dominance

effects and *AE* interaction effects but positive correlation of additive effects and *DE* interaction effects. There existed significantly negative correlation between boll size and lint percentage for dominance effects and *AE* interaction effects.

**Table 2 Estimation of genetic correlation coefficients between lint yield (LY) and its components : boll number (BN), boll size (BW) and lint percentage (LP).**

Parameter	LY & BN	LY & BW	LY & LP	BN & BW	BN & LP	BW & LP
$r_A$	0.993 **	—	1.000 **	—	0.778 **	—
$r_D$	0.721 *	0.717 **	-0.677 *	0.768 **	-1.000 **	-0.327 *
$r_{AE}$	0.104	-0.544	0.272	-0.017	-1.000 **	-0.928 **
$r_{DE}$	0.363	0.653	0.702 *	0.333	0.684 **	0.514
$r_P$	0.495 **	0.205 *	0.524 **	0.052	0.292 **	0.003

\* and \*\* significant at 0.05 and 0.01 levels, respectively. Correlation coefficients represent:  $r_A$  = additive variance,  $r_D$  = dominance variance,  $r_{AE}$  = additive × environment interaction variance,  $r_{DE}$  = dominance × environment interaction variance,  $r_e$  = residual variance,  $r_P$  = phenotypic variance.

Contribution ratios of single yield component trait to lint yield were presented in Table 3. The phenotypic contribution ratios were smaller for all the three yield component traits. However, there was very large contribution of additive effects due to boll number and lint percentage. The contribution ratio of dominance effects was larger than that of *DE* interaction effect for boll number. The contribution ratios of dominance effects and *DE* interaction effects were around 45% for boll size. For lint percentage, the contribution ratio of *AE* interaction effects was positive, and that of *DE* interaction effects was larger and than other two cases (boll number and boll size). The contribution ratios of

two component traits to lint yield were also estimated (Table 3). All contribution ratios of two component traits to lint yield were highly positive significant except for  $CR_{AE(BN, BW \rightarrow LY)}$ . The phenotypic contribution ratios of two component traits to lint yield were all around 45%. However, the additive contribution ratios were the largest (>93%) for all of pairwise two component traits, while  $CR_{D(BN, LP \rightarrow LY)}$ ,  $CR_{AE(BW, LP \rightarrow LY)}$ , and  $CR_{DE(BN, BW \rightarrow LY)}$  were relatively small as compared with other contribution ratios.  $CR_{D(BN, BW \rightarrow LY)}$ ,  $CR_{AE(BN, LP \rightarrow LY)}$ ,  $CR_{DE(BN, LP \rightarrow LY)}$  and  $CR_{DE(BW, LP \rightarrow LY)}$  were larger than 50%.

**Table 3 Estimated contribution ratios of one/two components to lint yield for boll number (BN), boll size (BW) and lint percentage (LP).**

Parameter	BN	BW	LP	Parameter	BN & BW	BN & LP	BW & LP
$CR_{A(C \rightarrow T)}$	0.909 **	-0.090	1.000 **	$CR_{A(C_1, C_2 \rightarrow T)}$	0.931 **	1.000 **	1.000 **
$CR_{D(C \rightarrow T)}$	0.389 **	0.493 **	-0.176	$CR_{D(C_1, C_2 \rightarrow T)}$	0.571 **	0.277 **	0.389 **
$CR_{AE(C \rightarrow T)}$	-0.007	-0.054	0.151 **	$CR_{AE(C_1, C_2 \rightarrow T)}$	-0.032	0.534 **	0.113 **
$CR_{DE(C \rightarrow T)}$	0.110 **	0.409 **	0.645 **	$CR_{DE(C_1, C_2 \rightarrow T)}$	0.383 **	0.541 **	0.756 **
$CR_{R(C \rightarrow T)}$	0.342 **	0.140 **	0.284 **	$CR_{R(C_1, C_2 \rightarrow T)}$	0.429 **	0.457 **	0.426 **

\* and \*\* significant at 0.05 and 0.01 levels, respectively.  $CR_{C \rightarrow T}$  represents contribution ratios of one component to lint yield,  $CR_{C_1, C_2 \rightarrow T}$  represents contribution ratio of two components to lint yield.

Predicted additive effects of lint yield and contributed additive effects of one/two component to lint yield were presented in Table 4. Since con-

ditional additive effects were zero for component trait(s)  $A_{(LY|LP)}$ ,  $A_{(LY|BN, LP)}$  and  $A_{(LY|BW, LP)}$ , the additive effects of lint yield could be wholly contribu-

ted due to these component trait( s ) (  $A_{(LY)} = A_{(LP \rightarrow LY)} = A_{(BN, LP \rightarrow LY)} = A_{(BW, LP \rightarrow LY)}$  ) so that these contributed additive effects were not presented in Table 4. There were negative additive effects of lint yield for parents 1, 3 and 10, but positive for parents 5, 7 and 8. These parents had additive effects (  $A_{(LY)}$  ) very closed to the contributed additive effects (  $A_{(BN \rightarrow LY)}$  and  $A_{(BN, BW \rightarrow LY)}$  ) except for parent 7. These results were matched to the facts that  $CR_{A_{(BN \rightarrow LY)}}$  and  $CR_{A_{(BN, BW \rightarrow LY)}}$  were very large. The contributed additive effects due to boll size tended to have small magnitude, revealing less importance of  $A_{(BW)}$  contributed to  $A_{(LY)}$ . It was suggested, by the fact of four parents (  $i = 1, 5, 7, 8$  ) having reverse values of  $A_{(BW \rightarrow LY)}$  as compared with  $A_{(LY)}$ , that boll size could have suppressing effects on additive behavior of lint yield.

**Table 4 Predicted additive effects of lint yield as well as the contributed additive effects of boll number ( BN ), boll size ( BW ) and lint percentage ( LP ) to lint yield**

Parent	$A_i$	$A_{(C \rightarrow T)}$		$A_{(C_1, C_2 \rightarrow T)}$
	LY	BN	BW	BN & BW
1	-4.040 **	-3.343 **	0.887 **	-3.968 **
2	-1.840	0.672 **	0.558 **	0.543 **
3	-2.170 **	-2.706 **	-0.644 **	-2.283 **
4	1.888 *	1.453 **	0.414 **	1.110
5	1.910 *	1.173 **	-0.906 **	1.304 **
6	1.194	1.214 **	1.055 **	0.582 **
7	6.160 **	3.791 **	-1.518 **	4.317 **
8	3.904 *	3.126 **	-0.017 **	3.611 **
9	-1.070	-0.631 **	0.209 **	-0.597 *
10	-5.940 **	-4.749 **	-0.042 **	-4.618 **

\* and \*\* significant at 0.05 and 0.01 levels, respectively.

$A_{(C \rightarrow T)}$  represents contributed additive effect of one component to lint yield,  $A_{(C_1, C_2 \rightarrow T)}$  represents contributed additive effect of two components to lint yield.

Predicted dominance effects of lint yield and the contributed dominance effects of yield component traits to lint yield were presented in Table 5. The dominance effects of cross 3 × 9 and 4 × 8 were positive for lint yield, which could be due to the reason of very large contributions of boll number. For cross 5 × 10, there was larger positive effect of ( BW → LY ). The dominance effects of lint

yield were no significant for cross 3 × 10, 4 × 6 and 4 × 10, but the contributed dominance effects of three component traits were negative. The contributed dominance effects of two component traits to lint yield were also predicted ( Table 5 ). Noticeably, the contributed dominance effects (  $D_{(BN, BW \rightarrow LY)}$ ,  $D_{(BN, LP \rightarrow LY)}$ , and  $D_{(BW, LP \rightarrow LY)}$  ) were quite large for cross 2 × 8. For cross 3 × 7 and 4 × 8, the dominance effects of lint yield were resulted from the contributed effects of ( BW, LP → LY ) and of ( BN, LP → LY ), respectively. The dominance effect of lint yield was so large for cross 3 × 9. The reason was that all contributed dominance effects of two component traits were positive and there were very large contributed effects for ( BN, BW → LY ) and ( BN, LP → LY ). For cross 3 × 10, 4 × 6 and 4 × 10, there were negative contributed dominance effects of both one component trait and two component traits, indicating genetic suppression of these component traits on the dominance behavior of these crosses. The result of cross 5 × 10 having  $D_{(C_1, C_2 \rightarrow T)} > D_{(T)}$  for ( BN, BW → LY ) and ( BW, LP → LY ) suggested that two component traits ( BN & BW, BW & LP ) could jointly augment the dominance behavior for cross 5 × 10.

### 3 Discussion

Path analysis has been extensively applied in quantitative genetics. Since path coefficient is the standardized partial regression coefficient in multiple regression model, its value is varied according to other variables in the model. For example, when boll number and boll size are considered as cause variables, the estimated path coefficients of boll number and boll size are 0.653 \*\* ( ± 0.045 ) and 0.166 \*\* ( ± 0.016 ), respectively. When three component traits are selected as cause variables, the estimated path coefficient will be 0.650 \*\* ( ± 0.042 ) for boll number, -0.024 ( ± 0.032 ) for boll size, and 0.153 \*\* ( ± 0.023 ) for lint percentage, respectively. However, the contribution ratio is only concerned with the conditional variance

**Table 5 Predicted dominance effects of lint yield as well as the contributed dominance effects of boll number ( BN ),boll size ( BW ) and lint percentage ( LP ) to lint yield**

Cross	$D_{ij}$	$D_{j(C \rightarrow T)}$			$D_{j(C_1, C_2 \rightarrow T)}$		
	LY	BN	BW	LP	BN & BW	BN & LP	BW & LP
1 × 6	1.356	0.923	-1.792**	0.681	-1.702**	0.928	-2.497**
1 × 7	6.283*	3.021*	-0.250**	-0.668**	1.482**	3.150	-0.476**
1 × 9	3.891*	-2.938**	1.552*	-0.904**	-0.840**	-3.785**	0.594**
1 × 10	6.050	-1.451**	5.153**	-1.740**	2.572*	-2.432**	4.172**
2 × 6	6.413**	1.840**	3.908**	-1.029**	3.715**	1.057**	3.152**
2 × 7	3.461	3.262	0.089**	2.145	1.756**	2.919*	-0.157**
2 × 8	0.006	4.149**	3.683**	-1.113**	5.591**	4.087**	3.888**
2 × 9	0.844	1.526**	1.704**	-1.446**	2.467**	0.937*	1.175**
2 × 10	0.699	-1.381**	-0.075	-1.614**	-0.556*	-2.388**	-1.130**
3 × 7	4.329*	-0.924**	2.925**	0.174**	2.119	-0.556**	4.010**
3 × 9	9.875*	8.335**	4.385	-1.385**	7.767**	8.498**	3.941*
3 × 10	3.713	-2.120**	-2.960**	-1.545**	-3.316**	-2.524**	-4.054**
4 × 6	2.629	-2.649**	-0.839**	-0.150*	-2.121**	-2.667**	-0.515**
4 × 8	5.372*	4.061**	-1.442**	0.399**	1.182**	4.205**	-1.676**
4 × 9	-3.580	2.870**	-0.794	1.792**	0.550**	3.372**	-0.524**
4 × 10	0.100	-3.287**	-2.136**	-1.283*	-3.367**	-3.651**	-2.725**
5 × 7	1.979	1.273	-0.032**	0.705	1.261	1.831**	1.034
5 × 8	7.747*	2.423**	3.732*	-0.440**	4.603**	2.265**	4.052*
5 × 9	6.163	0.763**	1.052**	-0.447**	1.376**	0.333**	0.673**
5 × 10	3.713*	1.802*	7.720**	-0.915**	6.567**	1.508**	7.728**

\* and \*\* significant at 0.05 and 0.01 levels, respectively.  $D_{j(C \rightarrow T)}$  represents contributed dominance effect of one component to lint yield,  $D_{j(C_1, C_2 \rightarrow T)}$  represents contributed dominance effect of two components to lint yield.

components and unconditional variance components. It will not be affected by the number of the component traits involved.

The contribution ratio proposed by Zhu<sup>[7]</sup> is a new concept to measure the influence of the component trait( s ) on a complex target trait. It can uncover the proportion of genetic variation of the given component trait( s ) to a complex trait. In some cases, there could sometimes have negative values for the contribution ratio, due to the conditional variance of complex trait being larger than the unconditional variance (  $\sigma_{U(T|C)}^2 > \sigma_{U(T)}^2$  ). It was suggested by negative contribution that the genetic effects on the complex trait could be larger when excluding the influence of the given component( s ). This implied that the expression of quantitative genes for the complex trait might be constrained by the expression of quantitative genes controlling the given component( s ). Then, the negative contribution ratio can be used for measuring the proportion of interference due to genetic

effects of the given component( s ) on the complex trait.

Multivariable conditional analysis for the contributions due to one or two component traits to the complex target trait provides a new approach for studying the complicated relationship for traits like yield and their components. If there is no correlation between two component traits, the genetic contribution ratio of two component traits will be equal to the sum of the genetic contribution ratio of each component trait. When the correlation coefficient between two components was positive ( or negative ), the genetic contribution ratio of these two component traits was smaller ( or larger ) than the sum of the genetic contribution ratio of each component trait. Since there are complicated relationship between genetic effects of component traits, the genetic contribution ratio of two component traits to a complex trait is usually not equal to the sum of the genetic contribution ratio of each component trait, and also the contributed

genetic effects of two component traits on a complex trait is not simple accumulated by that of single component trait on complex trait.

In analyzing genetic models of conditional genetic effects and variance components, it is possible to extend two component traits to three or more component traits. For example, if the contribution of three component traits are needed to be analyzed, the three-variable conditional phenotypic value can be obtained by

$$\mathbf{y}_{(***)} = \mathbf{y}_{(T)} - [\mathbf{C}_{(T,C_1)} \quad \mathbf{C}_{(T,C_2)} \quad \mathbf{C}_{(T,C_3)}] \begin{bmatrix} \mathbf{V}_{(C_1)} & \mathbf{C}_{(C_1,C_2)} & \mathbf{C}_{(C_1,C_3)} \\ \mathbf{C}_{(C_2,C_1)} & \mathbf{C}_{(C_2)} & \mathbf{C}_{(C_2,C_3)} \\ \mathbf{C}_{(C_3,C_1)} & \mathbf{C}_{(C_3,C_2)} & \mathbf{V}_{(C_3)} \end{bmatrix} \begin{bmatrix} \mathbf{y}_{(C_1)} - \mathbf{Xb}_{(C_1)} \\ \mathbf{y}_{(C_2)} - \mathbf{Xb}_{(C_2)} \\ \mathbf{y}_{(C_3)} - \mathbf{Xb}_{(C_3)} \end{bmatrix}$$

which can result in unbiased estimates of conditional variances  $\sigma_{\mu(T|C_1,C_2,C_3)}^2$  and predictors of conditional effects  $\mathbf{e}_{(T|C_1,C_2,C_3)}$ .

## References :

- [ 1 ] Bora G C ,Gupta S N ,Tomer Y S ,Singh S. Genetic variability ,correlation and path analysis in faba bean ( *Vicia faba* ). *Indian J Agric Sci* ,1998 68( 4 ) : 212 ~ 214.
- [ 2 ] Giridharan M P ,Jindal P C. Correlation regression and path-coefficient analysis of biochemical parameters associated with duration of crop growth in grape ( *Vitis vinifera* ). *Indian J Agric Sci* ,1999 69( 4 ) : 261 ~ 264.
- [ 3 ] Ogg Jr A G ,Seefeldt S S. Characterizing traits that enhance the competitiveness of winter wheat ( *Triticum aestivum* ) against jointed goatgrass ( *Aegilops cylindrica* ). *Weed Sci* ,1999 47( 1 ) : 74 ~ 80.
- [ 4 ] Razi H ,Assad M T. Comparison of selection criteria in normal and limited irrigation in sunflower. *Euphytica* ,1999 ,105 ( 2 ) : 83 ~ 90.
- [ 5 ] Wright S. The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. *Proc Natl Acad Sci USA* ,1920 6 : 320 ~ 332.
- [ 6 ] Wright S. Correlation and causation. *J Agric Res* ,1921 20 : 557 ~ 585.
- [ 7 ] Zhu J. Analysis of conditional genetic effects and variance components in developmental genetics. *Genetics* ,1995 , 141( 4 ) : 1633 ~ 1639.
- [ 8 ] Hartley H D ,Rao J N K. Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika* ,1967 54 ( 1 ~ 2 ) : 93 ~ 108.
- [ 9 ] Patterson H D ,Thompson R. Recovery of inter-block information when block sizes are unequal. *Biometrika* ,1971 58 ( 3 ) : 545 ~ 554.
- [ 10 ] Corbeil R R ,Searle S R. Restricted maximum likelihood ( REML ) estimation of variance components in the mixed model. *Technometrics* ,1976 18( 1 ) : 31 ~ 38.
- [ 11 ] Rao C R. Estimation of heteroscedastic variances in linear models. *J Am Stat Assoc* ,1970 65( 329 ) : 161 ~ 172.
- [ 12 ] Rao C R. Estimation of variance and covariance components MINQUE theory. *J Multivar Anal* ,1971 ,1 : 257 ~ 275.
- [ 13 ] Zhu J ,Weir B S. Analysis of cytoplasmic and maternal effects : I. A genetic model for diploid plant seeds and animals. *Theor Appl Genet* ,1994 89( 2 ~ 3 ) : 153 ~ 159.
- [ 14 ] Zhu J ,Weir B S. Diallel analysis for sex-linked and maternal effects. *Theor Appl Genet* ,1996 92( 1 ) : 1 ~ 9.
- [ 15 ] Zhu J ,Weir B S. Mixed model approaches for diallel analysis based on a bio-model. *Genet Res* ,1996 68( 3 ) : 233 ~ 240.
- [ 16 ] Henderson C R. Selection index and expected genetic advance. In : Hanson W D ,Robinson H F ( eds ). *Statistical Genetics and Plant Breeding*. National Academy of Science ,National Research Council ,Washington D C ,Publication 982 ,1963 ,141 ~ 163.
- [ 17 ] ZHU Jun. Mixed model approaches for estimating genetic variances and covariances. *J Biomath ( Chinese )* ,1992 , 7( 1 ) : 1 ~ 11.  
朱 军. Mixed model approaches for estimating genetic variances and covariances. *生物数学学报* ,1992 ,7( 1 ) : 1 ~ 11.
- [ 18 ] ZHU Jun. Methods of predicting genotype value and heterosis for offspring of hybrids. *J Biomath ( Chinese )* ,1993 8 ( 1 ) : 32 ~ 44.  
朱 军. 作物杂种后代基因型值和杂种优势的预测方法. *生物数学学报* ,1993 8( 1 ) : 32 ~ 44.
- [ 19 ] WU Ji-Xiang ,ZHU Jun ,JI Dao-Fan ,XU Fu-Hua. Analysis of genetic effects  $\times$  environment interactions for yield traits in upland cotton. *Hereditas ( Beijing )* ,1995 ,17( 5 ) : 1 ~ 4.  
吴吉祥 ,朱 军 ,季道藩 ,许馥华. 陆地棉产量性状的遗传及其与环境互作的分析. *遗传* ,1995 ,17( 5 ) : 1 ~ 4.
- [ 20 ] Miller R G. The jackknife : a review. *Biometrika* ,1974 61 ( 1 ) : 1 ~ 15.