A new approach to dissecting complex traits by combining quantitative trait transcript (QTT) mapping and diallel cross analysis

YANG DaiGang^{1†}, YE ChengYin^{2†}, MA XiongFeng¹, ZHU ZhiHong², ZHOU XiaoJian¹, WANG HaiFeng¹, MENG QingQin¹, PEI XiaoYu¹, YU ShuXun^{1*} & ZHU Jun^{2*}

¹ Cotton Research Institute, Chinese Academy of Agricultural Sciences, State Key Laboratory of Cotton Biology, Anyang 455000, China;

² Department of Agronomy, Zhejiang University, Hangzhou 310058, China

Received December 15, 2011; accepted April 6, 2012; published online May 22, 2012

A promising way to uncover the genetic architectures underlying complex traits may lie in the ability to recognize the genetic variants and expression transcripts that are responsible for the traits' inheritance. However, statistical methods capable of investigating the association between the inheritance of a quantitative trait and expression transcripts are still limited. In this study, we described a two-step approach that we developed to evaluate the contribution of expression transcripts to the inheritance of a complex trait. First, a mixed linear model approach was applied to detect significant trait-associated differentially expressed transcripts. Then, conditional analysis were used to predict the contribution of the differentially expressed genes to a target trait. Diallel cross data of cotton was used to test the application of the approach. We proposed that the detected differentially expressed transcripts with a strong impact on the target trait could be used as intermediates for screening lines to improve the traits in plant and animal breeding programs. It can benefit the discovery of the genetic mechanisms underlying complex traits.

new approach, complex traits, quantitative trait transcript (QTT), mapping, diallel cross analysis

Citation: Yang D G, Ye C Y, Ma X F, et al. A new approach to dissecting complex traits by combining quantitative trait transcript (QTT) mapping and diallel cross analysis. *Chin Sci Bull*, 2012, 57: 2695–2700, doi: 10.1007/s11434-012-5196-x

The partition and estimation of the genetic variance components underlying quantitative traits has long been recognized as an important problem in the field of quantitative genetics. The diallel cross design is one of the most popularly used genetic mating designs. It is widely used by plant and animal breeders, as well as by geneticists, to investigate the genetic properties of quantitative traits. The ANOVA and general linear model (GLM) methods have been widely used to estimate the genetic variance components in balanced diallel cross designs, but their limitations in analyzing unbalanced data have also been recognized [1]. Mixed linear model approaches to estimate genetic variance components and predict genetic effects under various types of ge-

netic designs have been proposed [2]. However, these traditional methods have failed to uncover the underlying genetic mechanisms and architectures, and were unable to identify the genes that significantly contribute to these genetic effects.

Recently rapid advances in high throughput technologies and statistical methods have helped accelerating the detection of differentially expressed genes in various species. Microarray techniques have been widely used to investigate the dynamic patterns of transcripts in both the temporal and spatial scales [3,4]. In various studies, the differentially expressed genes underlying distinct phenotypic categories have been detected; however, their mode of genetic inheritance is yet to be explored [5,6]. To genetically analyze gene expression traits, mapping studies of expression quantitative trait loci (eQTL) have been undertaken to identify

[†]These authors contributed equally to this work.

*Corresponding authors (email: jzhu@zju.edu.cn; yu@cricaas.com.cn)

the *cis/trans*-regulatory genomic variations that control variations in gene expression traits and to model the corresponding regulation networks [7–10]. Although by combining the analysis of expression and genomic variations a better understanding of the underlying genetic inheritance potentially could be obtained, the relationship between expression patterns and the genetic variations that contribute to phenotypic quantitative traits are still understudied and are not understood.

Thus, to investigate the inheritance of quantitative traits at the gene expression level, several studies have analyzed the F_1 hybrids of diallel cross data and performed correlation analyses to evaluate the association between gene expression patterns and quantitative traits; heterosis and marker heterozygosity have also been investigated [11–14]. However, these studies were not designed to estimate directly the contribution of gene expression patterns to the inheritance of the quantitative traits.

Conditional analysis methods have been developed to directly obtain the conditional variation of a complex trait by excluding the contribution of a component trait. These methods estimate the extra effects and variance components associated with the complex phenotypic trait that are independent of any existing component trait. Conditional analysis methods have also been applied to study the dynamic behavior of developmental traits on time-series datasets in both plants and animals [15,16]. Furthermore, these methods have been used to estimate the contribution of one particular component trait as well as the combination of multiple component traits to the variations of a phenotypic complex trait [17,18].

In this study, we extended the conditional analysis model to the gene expression level. This new approach has made the model capable of estimating the conditional genetic variance components of a target phenotypic trait that can be attributed to certain differentially expressed genes. The model can also predict the contribution of the genetic effects of the differentially expressed genes to the target trait. A general approach was employed in this study. First, significant trait-associated differentially expressed transcripts were detected, and then the conditional analysis was applied to evaluate the contributions of these transcripts to the inheritance of the target complex trait. Diallel cross data of cotton were used as a test dataset to demonstrate the application of the new approach.

1 Materials and methods

1.1 Upland cotton lines and field experiments

Using eight upland cotton inbreeding lines, we conducted an incomplete diallel cross with 8 parents and 10 F_1 crosses. The experimental design was a randomized complete block with three replicates. The experiments were conducted in clay soil in seven different environments. Each experiment

was followed by medium fertility and standard field management was employed. Agronomic traits, including lint yield, boll number per plant, boll weight, and lint percentage, were measured.

1.2 RNA extraction and the Affymetrix Gene Chip experiment

Cotton flower buds were collected separately from each of the seven environments after 10 d in the budding stage. Total RNA was extracted and three replicates from each of the parents and F_1 crosses were pooled and applied to 18 cotton chips. For the Affymetrix Gene Chip (Affymetrix, Santa Clara, CA, USA) analysis, 8 μg of the total RNA from each cotton root sample was used to make biotin-labeled cRNA targets. The cDNA and cRNA synthesis, cRNA fragmentation, hybridization, washing and staining, and scanning were all performed as described in the standard sample preparation Affymetrix gene chip protocol (Eukaryotic Target Preparation, Affymetrix). The poly-A RNA control kit and the One-Cycle cDNA Synthesis Kit were used as described at www.affymetrix.com/support/technical/manuals.affx. The signal intensity of each probe set on the Gene Chip was read using Affymetrix GCOS software, and the TGT (target mean value) was scaled to 500 for each chip. The Student's *t*-tests and the log2-transformed signal ratios of each probe set were carried out using the Partek Genomics Suite (version 6.3). The *q*-value of each probe set was calculated using the SAM software (Significance Analysis of Microarrays).

1.3 Statistical analysis

(i) Detection of differentially expressed genes. To detect the trait-associated expression transcripts, we used a mixed-model based approach that considered the transcript and block effects as random effects and searched for the significant differentially expressed transcripts in one dimension. The phenotypic value of an agronomic trait ($\mathbf{y}_{(p)}$) can be described as

$$\mathbf{y}_{(p)} = \mathbf{1}\mu + \mathbf{U}_Q \mathbf{e}_Q + \mathbf{U}_B \mathbf{e}_B + \mathbf{e}_\varepsilon \\ \sim N(\mathbf{1}\mu, \sigma_Q^2 \mathbf{U}_Q \mathbf{U}_Q^T + \sigma_B^2 \mathbf{U}_B \mathbf{U}_B^T + \sigma_\varepsilon^2 \mathbf{I}), \quad (1)$$

where $\mathbf{e}_Q$ is the vector of QTT effects, $\mathbf{e}_B$ is the vector of block effects, and $\mathbf{e}_\varepsilon$ is the vector of residual errors.

(ii) Conditional analysis of yield traits and differentially expressed gene transcripts. In the general unconditional genetic analysis of quantitative traits of diallel crosses in multiple environments, $\mathbf{y}_{(p)}$ can be represented by a mixed linear model as

$$\mathbf{y}_{(p)} = \mathbf{X}_E \mathbf{b}_E + \mathbf{U}_A \mathbf{e}_A + \mathbf{U}_D \mathbf{e}_D + \mathbf{U}_{AE} \mathbf{e}_{AE} + \mathbf{U}_{DE} \mathbf{e}_{DE} + \mathbf{U}_B \mathbf{e}_B + \mathbf{e}_\varepsilon \\ \sim N(\mathbf{1}\mu, \sigma_A^2 \mathbf{U}_A \mathbf{U}_A^T + \sigma_D^2 \mathbf{U}_D \mathbf{U}_D^T + \sigma_{AE}^2 \mathbf{U}_{AE} \mathbf{U}_{AE}^T \\ + \sigma_{DE}^2 \mathbf{U}_{DE} \mathbf{U}_{DE}^T + \sigma_B^2 \mathbf{U}_B \mathbf{U}_B^T + \sigma_\varepsilon^2 \mathbf{I}), \quad (2)$$

where $\mathbf{b}_E$ is the vector of environment effects, $\mathbf{e}_A$, $\mathbf{e}_D$, $\mathbf{e}_{AE}$ and $\mathbf{e}_{DE}$ are the vectors of additive, dominance, additive $\times$ environment effects and dominance $\times$ environment effects, respectively; $\mathbf{e}_B$ is the vector of block effects, and $\mathbf{e}_e$ is the vector of residual errors.

The total phenotypic variance (V_P) consists of variance components of additive (A) and dominance (D) effects, and their interactions with the environment (AE and DE) (i.e., $V_P = V_A + V_D + V_{AE} + V_{DE} + V_e$).

We defined the conditional vector of a trait (P) on the i -th differentially expressed gene (Q_i) as $\mathbf{y}_{(P|Q_i)}$. The conditional model for this vector can be derived as

$$\begin{aligned} \mathbf{y}_{(P|Q_i)} &= \mathbf{X}_{E(P|Q_i)} \mathbf{b}_{E(P|Q_i)} + \mathbf{U}_{A(P|Q_i)} \mathbf{e}_{A(P|Q_i)} + \mathbf{U}_{D(P|Q_i)} \mathbf{e}_{D(P|Q_i)} \\ &+ \mathbf{U}_{AE(P|Q_i)} \mathbf{e}_{AE(P|Q_i)} + \mathbf{U}_{DE(P|Q_i)} \mathbf{e}_{DE(P|Q_i)} + \mathbf{U}_{B(P|Q_i)} \mathbf{e}_{B(P|Q_i)} + \mathbf{e}_{e(P|Q_i)} \\ &\sim N\left(\mathbf{1}\mu_{(P|Q_i)}, \sigma_{A(P|Q_i)}^2 \mathbf{U}_{A(P|Q_i)} \mathbf{U}_{A(P|Q_i)}^T + \sigma_{D(P|Q_i)}^2 \mathbf{U}_{D(P|Q_i)} \mathbf{U}_{D(P|Q_i)}^T \right. \\ &+ \sigma_{AE(P|Q_i)}^2 \mathbf{U}_{AE(P|Q_i)} \mathbf{U}_{AE(P|Q_i)}^T + \sigma_{DE(P|Q_i)}^2 \mathbf{U}_{DE(P|Q_i)} \mathbf{U}_{DE(P|Q_i)}^T \\ &\left. + \sigma_{B(P|Q_i)}^2 \mathbf{U}_{B(P|Q_i)} \mathbf{U}_{B(P|Q_i)}^T + \sigma_{e(P|Q_i)}^2 \mathbf{I}\right), \end{aligned} \quad (3)$$

The variances of all these conditional genetic effects (additive $\mathbf{e}_{A(P|Q_i)}$, dominance $\mathbf{e}_{D(P|Q_i)}$, dominance $\times$ environment $\mathbf{e}_{AE(P|Q_i)}$, dominance $\times$ environment $\mathbf{e}_{DE(P|Q_i)}$, block $\mathbf{e}_{B(P|Q_i)}$, and residual $\mathbf{e}_{e(P|Q_i)}$) can be estimated by the minimum norm quadratic unbiased estimation (MINQUE) method by setting all prior values to 1.0 [2].

The unconditional and conditional random genetic effects were predicted using the adjusted unbiased prediction (AUP) method [19]. The contributed additive ($A_{(Q-P)}$) and dominance ($D_{(Q-P)}$) effects can be estimated as $A_{(Q-P)} = A_{(P)} - A_{(P|Q)}$ and $D_{(Q-P)} = D_{(P)} - D_{(P|Q)}$ [17]. The jackknife resampling method was applied to calculate the standard error (SE) for each parameter [20] and an approximate t -test was used to evaluate the significance of each parameter.

Gene mapping was conducted by using a software QTXNetwork (<http://ibi.zju.edu.cn/software/QTXNetwork/>) for genome-wide association study (GWAS), which can be used for mapping quantitative trait transcript (QTT) as well as quantitative trait SNP (SNP). Conditional genetic effects were predicted by a software QGASStation (<http://ibi.zju.edu.cn/software/qga/>) for quantitative trait analysis.

2 Results

2.1 Detection of differentially expressed genes

An association analysis was conducted using equation (1) to compare one-environmental data for the phenotypic trait of lint yield in cotton flower buds with the expression transcripts identified in the Affymetrix Gene Chip experiment.

Eleven significantly associated quantitative trait transcripts (QTTs) were detected, with a total heritability of 89.6%. The mean lint yield in the population tested was 108.3 kg, and the phenotypic variance reached 583.8. From the statistically associated QTTs, we selected six QTTs that had relatively large narrow heritabilities ($>5.0\%$) to investigate their role in the genetic architecture of lint yield. The six QTTs and their GenBank ID with putative gene descriptions are presented in Table 1. Four of the QTTs had negative genetic associations with lint yield, suggesting that lint yield could be increased when these QTTs had low expression levels. The other two QTTs had positive effects on lint yield, indicating that high expression levels of these QTTs could result in increased lint yield.

2.2 Contributed genetic effects of the QTTs on lint yield and its component traits

The predicted additive effects and the contributed effects of the six selected QTTs on lint yield and its component traits in four of the parents are presented in Table 2. The largest positive additive effect ($\hat{A}_5=8.52$) was observed in parent 5. The main contributions to this effect were from the decreasing expression of Q1951 ($\hat{A}_{5(Q1951 \rightarrow \text{LintYld})}=4.81$) and the increasing expression of Q1521 ($\hat{A}_{5(Q1521 \rightarrow \text{LintYld})}=3.29$). In parent 5, these two QTTs also made large contributions to increased boll numbers ($\hat{A}_{5(Q1951 \rightarrow \text{Bolls})}=0.51$ and $\hat{A}_{5(Q1521 \rightarrow \text{Bolls})}=1.12$). For parent 1, the additive effect on lint yield ($\hat{A}_1=2.62$) was largely contributed by the decreasing expressions of three QTTs (Q1999, Q1951, and Q1271); Q1999 and Q1271 also contributed significantly to an increase of lint percentage ($\hat{A}_{1(Q1999 \rightarrow \text{Lint}\%)=1.07}$) and ($\hat{A}_{1(Q1271 \rightarrow \text{Lint}\%)=1.01}$). In parent 8, all six QTTs contributed significantly ($\hat{A}_{8(Q \rightarrow \text{LintYld})} \approx -1.93$ – -6.10) to the negative additive effect on lint yield ($\hat{A}_8=-6.86$), while only three of the QTTs ($\hat{A}_{8(Q \rightarrow \text{Bolls})} \approx -2.21$ – -3.04) were the main contributors to the negative additive effect on boll number ($\hat{A}_8=-3.22$). These results indicated

Table 1 Expressed genes in cotton flower buds that were significantly associated with lint yield

QTT ID	GenBank ID	Gene description	Effect	h_Q^2 (%)
Q1999	DT049282	Amine oxidase, putative	-13.59	31.62
Q1951	CO120827	Nodulin MtN3-like protein	-10.64	19.38
Q1271	CO127124	UDP-glycosyltransferase 83A1-like	-7.44	9.49
Q1708	CD486429	Short chain alcohol dehydrogenase	6.39	6.99
Q1521	DW505577.1	Skp1, putative	6.35	6.9
Q1778	DT466217	WRKY transcription factor 48	-5.43	5.05

Table 2 Contributed additive effects of individual QTTs on lint yield and its component traits^{a)}

Parent	Trait	Additive effect	Contributed additive effect ($A_{(Q-P)}$)					
			Q1999	Q1951	Q1271	Q1780	Q1521	Q1778
A ₁	LintYld	2.62 [*]	2.12 [‡]	1.09 [‡]	1.01 [‡]	0.12 [‡]	−1.15 [‡]	0.08 [‡]
	Lint%	2.26 [‡]	1.07 [‡]	−0.25 [‡]	1.01 [‡]	−0.02 [‡]	−0.78 [‡]	−0.51 [‡]
	BollWt	0.17 [‡]	−0.09 [‡]	0.05 [‡]	−0.13 [‡]	0.06 [‡]	0.26 [‡]	
A ₃	LintYld	−1.61 [*]	0.22 [‡]	0.21 [‡]	0.07 [‡]	−0.49 [‡]	0.84 [‡]	0.10 [‡]
	Lint%	−0.35 [‡]	−0.25 [‡]	0.03 [‡]	1.07 [‡]	0.72 [‡]	0.29 [‡]	0.00 [‡]
	Bolls	1.18 [‡]	0.12 [‡]	0.13 [‡]			0.37 [‡]	−0.01 [‡]
	BollWt	−0.4 [‡]	−0.03 [‡]	−0.02 [‡]	−0.03 [‡]	0.06 [‡]	−0.08 [‡]	
A ₅	LintYld	8.52 [†]	1.78 [‡]	4.81 [‡]	0.46 [‡]	0.45 [‡]	3.29 [‡]	
	Lint%	−0.3 [‡]		0.92 [‡]			0.99 [‡]	1.28 [‡]
	Bolls	0.82 [†]	0.57 [‡]	0.51 [‡]	0.45 [‡]		1.12 [‡]	−0.22 [‡]
	BollWt	0.08 ^{**}	−0.06 [‡]	−0.13 [‡]	0.01 [‡]	0.00 [‡]	−0.22 [‡]	
A ₈	LintYld	−6.86 [†]	−4.88 [‡]	−5.04 [‡]	−2.19 [‡]	−1.93 [‡]	−6.10 [‡]	−2.92 [‡]
	Bolls	−3.22 [‡]	−2.21 [‡]		−2.79 [‡]		−3.04 [‡]	
	BollWt	0.47 [‡]		0.12 [‡]	0.40 [‡]	0.45 [‡]	0.45 [‡]	

a) [‡] $P < 0.001$, [†] $P < 0.005$, ^{**} $P < 0.01$, ^{*} $P < 0.05$.

that, in different parents, individual QTTs can have dissimilar additive effects on lint yield and its component traits.

For the dominant effects on lint yield and its component traits, the predicted positive effects ($\hat{D}_{i \times j}$) and the contributed effects ($\hat{D}_{i \times j(Q \rightarrow P)}$) in the significant F_1 crosses are presented in Table 3. The results showed that all the predicted overall dominant effects on lint yield ($\hat{D}_{i \times j}$) had relatively large values that ranged from 6.57 to 19.07. For different F_1 crosses, the six selected QTTs contributed differently to the dominant effect on lint yield. For the three F_1 crosses, $F_{1(1 \times 2)}$, $F_{1(5 \times 7)}$ and $F_{1(5 \times 8)}$, with the largest dominance effects, the main contributors to the positive effects in $F_{1(1 \times 2)}$ and for $F_{1(5 \times 8)}$ were Q1999 ($\hat{D}_{1 \times 2(Q1999 \rightarrow \text{LintYld})} = 10.19$), Q1521 ($\hat{D}_{1 \times 2(Q1521 \rightarrow \text{LintYld})} = 13.63$, $\hat{D}_{5 \times 8(Q1521 \rightarrow \text{LintYld})} = 16.80$), and Q1951 ($\hat{D}_{5 \times 8(Q1951 \rightarrow \text{LintYld})} = 19.69$), and the main contributors to the negative effect in $F_{1(5 \times 7)}$ were Q1951 ($\hat{D}_{5 \times 7(Q1951 \rightarrow \text{LintYld})} = -7.41$) and Q1521 ($\hat{D}_{5 \times 7(Q1521 \rightarrow \text{LintYld})} = -9.25$). Similar to the analysis of the additive effects, the dominance mechanisms that influence lint yield and its component traits were inferred by evaluating the contributed dominant effects of these QTTs in certain crosses. For instance, in $F_{1(5 \times 8)}$, the two QTTs that contributed most to lint yield (Q1951 and Q1521) also contributed the largest dominant effects to lint percentage ($\hat{D}_{5 \times 8(Q1951 \rightarrow \text{Lint}\%) } = 0.90$, $\hat{D}_{5 \times 8(Q1521 \rightarrow \text{Lint}\%) } = 0.38$) and boll weight ($\hat{D}_{5 \times 8(Q1951 \rightarrow \text{BollWt}) } = 0.43$, $\hat{D}_{5 \times 8(Q1521 \rightarrow \text{BollWt}) } = 0.48$). This result implied that these two component traits share a genetic pathway that is similar to the pathway for

lint yield. The same analysis can be applied to $F_{1(1 \times 2)}$, where Q1999 and Q1521 made large contributions to the dominant effects on both lint yield and lint percentage. Thus, similar to the results obtained for the contributed additive effects, the six selected QTTs all contributed to the dominance effects. While they contributed positively to some traits in some F_1 crosses, they could contribute negatively to other traits in other F_1 crosses. These results implied that significantly associated quantitative trait transcripts can only be identified appropriately for specific traits with particular genotypes.

3 Discussion

In this study, we described and tested a new approach to evaluate the contribution of gene expression to the inheritance of complex traits. The association between gene expression transcripts and complex traits was investigated. After obtaining the significant trait-associated differentially expressed transcripts, conditional diallel analysis was applied to the significant transcripts. From the results of the analysis, the additive and dominance variations of the phenotypic trait were decomposed to obtain the genetic contributions of specific differentially expressed transcripts. Using this information, their contributed additive and dominant effects on the phenotypic quantitative trait were predicted. The proposed approach is different from existing methods that only analyze the correlation between heterosis and gene expression patterns. Our approach directly estimates the genetic impact of each of the differentially expressed transcripts on the inheritance of phenotypic traits, thereby exploring the

Table 3 Contributed dominant effects of individual QTTs on lint yield and its component traits^{a)}

Cross	Trait	Dominance effect	Contributed dominance effect ($D_{(Q-p)}$)					
			Q1999	Q1951	Q1271	Q1780	Q1521	Q1778
D _{1×2}	LintYld	19.07 [†]	10.19 [‡]	0.38 [‡]	2.00 [‡]	0.50 [‡]	13.63 [‡]	1.08 [‡]
	Lint%	0.68 [†]	0.73 [‡]	−0.01 [‡]	0.28 [‡]	0.12 [‡]	0.42 [‡]	
	BollWt	0.19 [*]	0.07 [‡]	0.05 [‡]	−0.04 [‡]	0.03 [‡]	0.10 [‡]	0.02 [‡]
D _{1×3}	LintYld	14.27 [*]		0.38 [‡]	2.00 [‡]	0.50 [‡]	13.63 [‡]	1.08 [‡]
	Lint%	0.90 [†]		−0.01 [‡]	0.28 [‡]	0.12 [‡]	0.42 [‡]	
D _{2×5}	LintYld	13.39 ^{**}	3.22 [‡]	6.32 [‡]	0.95 [‡]	−0.94 [‡]	2.69 [‡]	6.89 [‡]
	BollWt	0.19 [†]	0.07 [‡]	0.09 [‡]			0.01 [‡]	−0.05 [‡]
D _{3×5}	LintYld	6.84 [*]	1.48 [‡]	1.44 [‡]	3.42 [‡]	1.03 [‡]	−2.52 [‡]	5.83 [‡]
	Lint%	0.62 [†]	0.20 [‡]	0.11 [‡]		0.16 [*]	0.02 [‡]	1.14 [‡]
D _{5×6}	LintYld	6.57 [*]	12.51 [‡]	7.66 [‡]	−1.04 [‡]	−0.51 [‡]	−5.48 [‡]	4.93 ^{**}
	Lint%	0.84 [‡]	0.93 [‡]	0.30 [‡]	−0.19 [‡]	0.01 [‡]	−0.01 [‡]	0.84 [‡]
D _{5×7}	LintYld	16.72 [†]	−0.25 [‡]	−7.41 [‡]	1.49 [‡]	−2.26 [‡]	−9.25 [‡]	0.39 [‡]
	Lint%	0.36 [*]	−0.01 [‡]	−0.24 [‡]	−0.12 [‡]	−0.22 [‡]	−0.18 [‡]	0.08 [‡]
D _{5×8}	LintYld	18.77 [‡]	4.99 [‡]	19.69 [‡]	−5.95 [‡]	5.12 [‡]	16.80 [‡]	−0.55 [‡]
	Lint%	0.41 [*]	0.10 [‡]	0.90 [‡]	−0.57 [‡]	0.25 [*]	0.38 [*]	−0.19 [‡]
	BollWt	0.53 [‡]	0.21 [‡]	0.43 [‡]	0.07 [‡]	0.24 [‡]	0.48 [‡]	0.02 [‡]

a) [‡] $P < 0.001$, [†] $P < 0.005$, ^{**} $P < 0.01$, ^{*} $P < 0.05$.

underlying genetic architecture of phenotypic traits.

We used the cotton diallel cross data to analyze a real data application using our approach. After selecting the six most significantly associated transcripts from the association model of the cotton lint yield trait, we evaluated the contributed additive and dominant effects of each of the transcript to lint yield and its component traits in eight parent lines and 10 F₁ crosses. The results showed that the six selected transcripts had high genetic effects on lint yield and its component traits. However, the effects varied among the parent lines and their F₁ crosses. We propose that these predicted contributed additive and dominant effects could be used as indicators to screen specific parent lines and their F₁ crosses in breeding selection to improve the quality of traits in offspring. Furthermore, by comparing the contributions of the transcripts among traits in a specific parent line or its F₁ crosses, we could evaluate the molecular relationship of a particular trait and its component traits and come to a preliminary understanding of the molecular mechanisms of gene expression that might be shared among different traits.

If both diallel cross data and gene expression data from multiple environments were available, then our conditional approach could be extended to, for example, a genetic model with G×E effects. However, generating microarray data for multiple environments can be quite costly. In our study, we used the microarray data for one environment and diallel cross data for seven environments. Therefore, we actually evaluated the contribution of gene expression from one environment to the quantitative traits collected from seven environments. The contributions of the transcripts to G×E effects were not estimated. Therefore, when our approach was used to analyze the real cotton diallel cross data, the

results obtained were actually a measure of the genetic contribution of transcripts whose expressions were not sensitive to the changes of environments. We proposed that these data could be valuable for breeding programs when extended to various environments.

Transcripts that significantly contribute to the genetic properties of quantitative traits could be further investigated for their functional roles in processes that may be involved in the formation of the phenotypic trait and its correlated traits. Furthermore, we suggest that the transcripts with strong impacts on the genetic variation of a complex trait that were detected in this study could be used as intermediates in plant and animal breeding programs, to screen parents with either high general or high specific combination abilities. By applying our proposed method to the analysis of genetic mating and microarray datasets, a better understanding of the contribution of gene expression to the genetic mechanisms of complex traits (e.g., lint yield) at the molecular level could be developed. Further, an understanding of the genetic relationship between a certain trait and its correlated traits could help build the genetic architecture of these traits and guide the improvement of agronomic and yield traits in breeding programs.

This work was supported by the National Basic Research Program of China (2011CB109306), the National High Technology Research and Development Program of China (2009ZX08009-004B, 2011AA10A102), the CNTC (110200701023), the YNTC (08A05) and the earmarked fund for Modern Agro-industry Technology Research System (CARS-18-05).

- 1 Wu J, Jenkins J, McCarthy J, et al. Variance component estimation using the additive, dominance, and additive x additive model when

- genotypes vary across environments. *Crop Sci*, 2006, 46: 174–179
- 2 Zhu J. General genetic models and new analysis methods for quantitative traits (in Chinese). *J Zhejiang Agri Univ*, 1994, 20: 551–559
 - 3 Allison D B, Cui X, Page G P, et al. Microarray data analysis: From disarray to consolidation and consensus. *Nat Rev Genet*, 2006, 7: 55–65
 - 4 Hoheisel J D. Microarray technology: Beyond transcript profiling and genotype analysis. *Nat Rev Genet*, 2006, 7: 200–210
 - 5 Slonim D K. From patterns to pathways: Gene expression data analysis comes of age. *Nat Genet*, 2002, 32(Suppl): 502–508
 - 6 Wang Y, Wu C, Ji Z, et al. Non-parametric change-point method for differential gene expression detection. *PLoS ONE*, 2011, 6: e20060
 - 7 Cookson W, Liang L, Abecasis G, et al. Mapping complex disease traits with global gene expression. *Nat Rev Genet*, 2009, 10: 184–194
 - 8 Druka A, Potokina E, Luo Z, et al. Expression quantitative trait loci analysis in plants. *Plant Biotech J*, 2010, 8: 10–27
 - 9 Li H, Deng H. Systems genetics, bioinformatics and eQTL mapping. *Genetica*, 2010, 138: 915–924
 - 10 Nica A C, Parts L, Glass D, et al. The architecture of gene regulatory variation across multiple human tissues: The MuTHER study. *PLoS Genet*, 2011, 7: e1002003
 - 11 Chen X, Li M, Shi J, et al. Gene expression profiles associated with intersubgenomic heterosis in *Brassica napus*. *Theor Appl Genet*, 2008, 117: 1031–1040
 - 12 Hoecker N, Keller B, Muthreich N, et al. Comparison of maize (*Zea mays* L.) F₁-hybrid and parental inbred line primary root transcriptomes suggests organ-specific patterns of nonadditive gene expression and conserved expression trends. *Genetics*, 2008, 179: 1275–1283
 - 13 Stupar R M, Gardiner J M, Oldre A G, et al. Gene expression analyses in maize inbreds and hybrids with varying levels of heterosis. *BMC Plant Biol*, 2008, 8: 33
 - 14 Zhan H, Chen X, Xu S. A stochastic expectation and maximization algorithm for detecting quantitative trait-associated genes. *Bioinformatics*, 2011, 27: 63–69
 - 15 Zhu J. Analysis of conditional genetic effects and variance components in developmental genetics. *Genetics*, 1995, 141: 1633–1639
 - 16 Atchley W R, Zhu J. Developmental quantitative genetics, conditional epigenetic variability and growth in mice. *Genetics*, 1997, 147: 765–776
 - 17 Wen Y X, Zhu J. Multivariable conditional analysis for complex trait and its components. *Acta Genet Sin*, 2005, 32: 289–296
 - 18 Wu J, Jenkins J, McCarty J, et al. Genetic association of yield with its component traits in a recombinant inbred line population of cotton. *Euphytica*, 2004, 140: 171–179
 - 19 Zhu J. Methods of predicting genotype value and heterosis for offspring of hybrids (in Chinese). *J Biomath*, 1993, 8: 32–44
 - 20 Miller R G. The jackknife: A review. *Biometrika*, 1974, 61: 1–15

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.