

Framework for dissection of complex cytonuclear epistasis by a two-dimensional genome scan

TANG ZaiXiang^{1,2}, HU ZhiQiu³, YANG ZeFeng¹, YU Bo² & XU ChenWu^{1*}

¹ Jiangsu Provincial Key Laboratory of Crop Genetics and Physiology, and Key Laboratory of Plant Functional Genomics of Ministry of Education, Yangzhou University, Yangzhou 225009, China;

² School of Public Health, Medical College of Soochow University, Suzhou 215123, China;

³ Departments of Botany and Plant Sciences, University of California, Riverside, California 92521, USA

Received November 5, 2011; accepted March 8, 2012; published online April 5, 2012

Epistasis between cytoplasmic and nuclear genes is the primary genetic component of complex quantitative traits. Genetic dissection of cytonuclear epistasis is fundamentally important to understand the genetic architecture of complex traits. In this study, a two-dimensional genome scan strategy was employed to evaluate the contribution of cytoplasm, quantitative trait loci (QTL), QTL×QTL interactions and QTL×QTL×cytoplasm interactions to the phenotypic variation. The *p*-value and parameter value for each genetic effect were calculated by multiple regression analysis. A stepwise approach was suggested to build confidence in candidate QTL on the basis of *q*-value estimation, false discovery rate calculation and Bonferroni adjustment. A fine-scale grid scan strategy was proposed for further analysis of peaks of interest. Plant height in maize was used as an example to illustrate the efficiency of the two-dimensional genome scan strategy.

cytonuclear interaction, quantitative trait loci, reciprocal cross, maize

Citation: Tang Z X, Hu Z Q, Yang Z F, et al. Framework for dissection of complex cytonuclear epistasis by a two-dimensional genome scan. *Chin Sci Bull*, 2012, 57: 2675–2680, doi: 10.1007/s11434-012-5116-0

The nucleus and cytoplasm are separate entities in a cell, but they cooperate to maintain cellular functions and to determine the organismal phenotypes. The cytoplasm contains a variety of organelles, of which the mitochondria and chloroplasts contain their own DNA, and provides an environment for nuclear gene expression and cellular metabolic reactions. Cytonuclear interactions are an important contributor to phenotypic variation. A significant contribution of nuclear-cytoplasm interaction has been detected in model organisms such as maize, rice, mice, yeast, and *Drosophila* [1–5]. The close relationship between the nucleus and cytoplasm has attracted much attention from macro- and micro-evolutionary perspectives. Research on cytonuclear co-evolution and coadaptation suggests cytonuclear epistasis contributes substantially to phenotypic variation [6]. However, in most gene-mapping studies that feature a single

cytoplasmic background, incorporation of this important component is difficult. Nichols et al. [7] pioneered research on subtle QTL×cytoplasm interactions. These authors used double haploids produced by androgenesis using eggs from different females in rainbow trout. Both the maternal cytoplasmic environment and QTL×cytoplasm effects contributed to variability in development rate, but QTL×cytoplasm interactions were minor and only detected at small-effect QTL. Subsequently, Tang et al. [8] proposed a mapping model based on a reciprocal-cross design to detect the cytonuclear epistatic QTL and evaluate the contribution of the cytoplasm and QTL×cytoplasm interactions to phenotypic variation.

1 Statistical model and methods

The proposed model is an extension of composite interval

*Corresponding author (email: qtls@yzu.edu.cn)

mapping model [9]. Obviously, this additive-dominant genetic model cannot be used to detect and evaluate the interactions among different loci in the genome. However, epistasis among nuclear genes is a primary genetic component of quantitative traits in several model organisms [10–12]. Therefore, interactions among loci should not be ignored in complex trait studies. To address this problem, we developed a novel strategy to perform a simultaneous two-dimensional (2D) genome-wide search under a flexible model of epistasis.

Taking a F₂ population as an example, based on a reciprocal-cross design and the basic concept of a two-dimensional scan approach, the phenotypic value of the *j*th (*j* = 1 to *n*) individual in the F₂ mapping population can be expressed by the following genetic model:

$$y_j = m + x_{c_j}c + x_{1j}a_1 + x_{2j}d_1 + x_{3j}a_2 + x_{4j}d_2 + x_{5j}i_{a_1a_2} + x_{6j}i_{a_1d_2} + x_{7j}i_{a_2d_1} + x_{8j}i_{d_1d_2} + x_{9j}i_{a_1c} + x_{10j}i_{d_1c} + x_{11j}i_{a_2c} + x_{12j}i_{d_2c} + x_{13j}i_{a_1a_2c} + x_{14j}i_{a_1d_2c} + x_{15j}i_{a_2d_1c} + x_{16j}i_{d_1d_2c} + e_j.$$

This is a two-QTL model, where *m* is the overall mean of the population; *c* is the cytoplasmic effect; *a*₁ and *d*₁ are the additive and dominance effects, respectively, of one QTL; *a*₂ and *d*₂ denote the effects of a second QTL; *i*_{*a*₁*a*₂}, *i*_{*a*₁*d*₂}, *i*_{*a*₂*d*₁} and *i*_{*d*₁*d*₂} are the additive-by-additive, additive-by-dominance, dominance-by-additive and dominance-by-dominance interaction effects of two QTL; the cytonuclear interaction effects are denoted by *i*_{*a*₁*c*}, *i*_{*d*₁*c*}, *i*_{*a*₂*c*}, *i*_{*d*₂*c*},

*i*_{*a*₁*a*₂*c*}, *i*_{*a*₁*d*₂*c*}, *i*_{*a*₂*d*₁*c*} and *i*_{*d*₁*d*₂*c*}, which are the interactions among cytoplasmic effect and *a*₁, *d*₁, *a*₂, *d*₂, *i*_{*a*₁*a*₂}, *i*_{*a*₁*d*₂}, *i*_{*a*₂*d*₁} and *i*_{*d*₁*d*₂}; *x*_{*cj*} is an indicator variable, for which *x*_{*cj*} = 1 and *x*_{*cj*} = –1 denote the cytoplasmic types of the two parents; *x*_{*1j*} to *x*_{*16j*} are the indicator variables that describe different QTL genotypes and their interactions and are defined in Table 1; and *e*_{*j*} is the residual error with a *N*(0, *σ*_{*e*}²) distribution. All 16 effects that involved nuclear QTL and cytonuclear interaction are partial regression coefficients of *y*_{*j*} on the indicator variables *x*_{*1j*} to *x*_{*16j*}.

The genotype indicator variable for a QTL was not observable. However, we assumed that each QTL was linked to a marker, and thus the genotype of the QTL was replaceable by that of the corresponding marker. The proposed method was essentially a two-marker analysis implemented via a regression approach [13]. Hereafter, we use the terms marker and QTL interchangeably.

The proposed statistical model incorporated important interactions within the nuclear genome and interactions among nuclear loci and the cytoplasmic background. A general multiple regression analysis was implemented for each QTL pair to obtain the *F*-test statistic of each two-QTL model, the *t*-test statistic of each genetic component in the model and their estimates. To perform the 2D genome scan, (*k*²–*k*)/2 tests would be implemented in the case of *k* markers in the whole genome. Consequently, the risk of false discoveries would be usually very high because of this high-throughput experiment. Therefore, it was necessary to establish an appropriate significance threshold in 2D genome

Table 1 QTL genotypes and their genetic components in the F₂ population with two cytoplasmic backgrounds

Cyto-genotypes ^{a)}	<i>m</i>	<i>c</i>	<i>a</i> ₁	<i>d</i> ₁	<i>a</i> ₂	<i>d</i> ₂	<i>a</i> ₁ <i>a</i> ₂	<i>a</i> ₁ <i>d</i> ₂	<i>a</i> ₂ <i>d</i> ₁	<i>d</i> ₁ <i>d</i> ₂	<i>i</i> _{<i>a</i>₁<i>c</i>}	<i>i</i> _{<i>d</i>₁<i>c</i>}	<i>i</i> _{<i>a</i>₂<i>c</i>}	<i>i</i> _{<i>d</i>₂<i>c</i>}	<i>i</i> _{<i>a</i>₁<i>a</i>₂}	<i>i</i> _{<i>a</i>₁<i>d</i>₂}	<i>i</i> _{<i>a</i>₂<i>d</i>₁}	<i>i</i> _{<i>d</i>₁<i>d</i>₂}
<i>C</i> ₁ (<i>Q</i> ₁ <i>Q</i> ₁ <i>Q</i> ₂ <i>Q</i> ₂)	1	1	1	0	1	0	1	0	0	0	1	0	1	0	1	0	0	0
<i>C</i> ₁ (<i>Q</i> ₁ <i>Q</i> ₁ <i>Q</i> ₂ <i>q</i> ₂)	1	1	1	0	0	1	0	1	0	0	1	0	0	1	0	1	0	0
<i>C</i> ₁ (<i>Q</i> ₁ <i>Q</i> ₁ <i>q</i> ₂ <i>q</i> ₂)	1	1	1	0	–1	0	–1	0	0	0	1	0	–1	0	–1	0	0	0
<i>C</i> ₁ (<i>Q</i> ₁ <i>q</i> ₁ <i>Q</i> ₂ <i>Q</i> ₂)	1	1	0	1	1	0	0	0	1	0	0	1	1	0	0	0	1	0
<i>C</i> ₁ (<i>Q</i> ₁ <i>q</i> ₁ <i>Q</i> ₂ <i>q</i> ₂)	1	1	0	1	0	1	0	0	0	1	0	1	0	1	0	0	0	1
<i>C</i> ₁ (<i>Q</i> ₁ <i>q</i> ₁ <i>q</i> ₂ <i>q</i> ₂)	1	1	0	1	–1	0	0	0	–1	0	0	1	–1	0	0	0	–1	0
<i>C</i> ₁ (<i>q</i> ₁ <i>q</i> ₁ <i>Q</i> ₂ <i>Q</i> ₂)	1	1	–1	0	1	0	–1	0	0	0	–1	0	1	0	–1	0	0	0
<i>C</i> ₁ (<i>q</i> ₁ <i>q</i> ₁ <i>Q</i> ₂ <i>q</i> ₂)	1	1	–1	0	0	1	0	–1	0	0	–1	0	0	1	0	–1	0	0
<i>C</i> ₁ (<i>q</i> ₁ <i>q</i> ₁ <i>q</i> ₂ <i>q</i> ₂)	1	1	–1	0	–1	0	1	0	0	0	–1	0	–1	0	1	0	0	0
<i>C</i> ₂ (<i>Q</i> ₁ <i>Q</i> ₁ <i>Q</i> ₂ <i>Q</i> ₂)	1	–1	1	0	1	0	1	0	0	0	–1	0	–1	0	–1	0	0	0
<i>C</i> ₂ (<i>Q</i> ₁ <i>Q</i> ₁ <i>Q</i> ₂ <i>q</i> ₂)	1	–1	1	0	0	1	0	1	0	0	–1	0	0	–1	0	–1	0	0
<i>C</i> ₂ (<i>Q</i> ₁ <i>Q</i> ₁ <i>q</i> ₂ <i>q</i> ₂)	1	–1	1	0	–1	0	–1	0	0	0	–1	0	1	0	1	0	0	0
<i>C</i> ₂ (<i>Q</i> ₁ <i>q</i> ₁ <i>Q</i> ₂ <i>Q</i> ₂)	1	–1	0	1	1	0	0	0	1	0	0	–1	–1	0	0	0	–1	0
<i>C</i> ₂ (<i>Q</i> ₁ <i>q</i> ₁ <i>Q</i> ₂ <i>q</i> ₂)	1	–1	0	1	0	1	0	0	0	1	0	–1	0	–1	0	0	0	–1
<i>C</i> ₂ (<i>Q</i> ₁ <i>q</i> ₁ <i>q</i> ₂ <i>q</i> ₂)	1	–1	0	1	–1	0	0	0	–1	0	0	–1	1	0	0	0	1	0
<i>C</i> ₂ (<i>q</i> ₁ <i>q</i> ₁ <i>Q</i> ₂ <i>Q</i> ₂)	1	–1	–1	0	1	0	–1	0	0	0	1	0	–1	0	1	0	0	0
<i>C</i> ₂ (<i>q</i> ₁ <i>q</i> ₁ <i>Q</i> ₂ <i>q</i> ₂)	1	–1	–1	0	0	1	0	–1	0	0	1	0	0	–1	0	1	0	0
<i>C</i> ₂ (<i>q</i> ₁ <i>q</i> ₁ <i>q</i> ₂ <i>q</i> ₂)	1	–1	–1	0	–1	0	1	0	0	0	1	0	1	0	–1	0	0	0

a) *C*₁ and *C*₂ denote the cytoplasm types of two parents, respectively.

scans to determine the overall significance of the results.

Several adjustment procedures have been suggested to establish the significance and control the number of false positives in multiple tests [14–17]. Similar to choosing a significance threshold in a single test, the choice of a multiple testing correction method to use depends on the research objectives and costs associated with false positives. In accordance with Carlborg and Haley's suggestion [18], a stepwise strategy was suggested to build confidence in these QTL in the present study, which integrated multiple testing correction methods and possible external information. Firstly, a 2D genome scan was implemented to estimate the contribution of each marker combination to the interesting trait. The complete results included the F -test statistic of each two-QTL model, the t -test statistic of each genetic component in the model and their estimates. Secondly, from these results, a set of potentially interesting QTL was selected on the basis of q -value estimation, false discovery rate (FDR) calculation and Bonferroni adjustment for each genetic component. In the present study, the q -value was a Bayesian posterior p -value, which was used to control the positive false discovery rate (pFDR). Storey [15] proposed definition of the q -value as an analog of the p -value that incorporates FDR-based multiple testing correction.

The aim of the procedure was to separate the set of potential QTL into high-, moderate- and low-confidence QTL, which were simultaneously detected by three, two and one correction methods, respectively. In addition, if strong external information was available, it was preferred to help with classification of these potential QTL. This information can be obtained from previous research, other analysis models and even other independent data sources. This stepwise strategy allowed the detection of QTL with a risk of false positive and false negative detection, but was supported by strong external information. Subsequently, the interesting main-effect QTL, QTL×QTL interactions, and QTL×cytoplasm interactions can be selected for further research.

2 Simulation study

The applicability of the proposed method was further demonstrated by a simulation experiment. Two QTL, denoted QTL1 and QTL2, were designed to locate at the 63 cM (7th marker) and 140 cM (15th marker) positions of a mimic chromosome, which was 195 cM in length and covered by 20 evenly spaced markers (Figure 1). The effects of the two QTL were $a_1 = 1.0$, $d_1 = 1.0$, $a_2 = 0.0$, $d_2 = 1.0$, $i_{a1c} = 1.5$, $i_{d1c} = 0.0$, $i_{a2c} = 1.5$, $i_{d2c} = 1.0$, $i_{a1a2} = 1.0$, $i_{d1d2} = 1.0$, $i_{a2d1} = 0.0$, $i_{a1d2} = 1.5$, $i_{a1a2c} = 1.5$, $i_{d1d2c} = 1.5$, $i_{a2d1c} = 1.0$, and $i_{a1d2c} = 0.0$. The population mean and cytoplasmic effect were set as $m = 5$ and $c = 0.5$, respectively. We generated data with 100 direct F_2 individuals and 100 reciprocal F_2 individuals. The QTL heredities, calculated with the formula

$(a^2/2 + d^2/4) / \sigma_p^2$, were 9.7% and 3.2% for QTL1 and QTL2, respectively. These data were analyzed using the proposed method with 100 replicates. The threshold in the simulation study was obtained by Bonferroni adjustment, and the whole-genome significance level was 0.05. The statistical power was calculated in accordance with the threshold and estimated position. Only the QTL that were located in close proximity to the seventh and fifteenth markers and reached the threshold were considered to be effective QTL. The statistical powers of QTL detection and estimates of the QTL effects are summarized in Table 2. The proposed method provided a reasonably precise and accurate estimation of QTL effects, especially the interaction among the two QTL and the cytoplasmic background. The estimated population mean was 4.98 ± 0.17 and the estimated cytoplasmic effect was 0.48 ± 0.16 , which were very similar to the true values. The results from a 2D chromosome scan of simulated data are presented in Figure 1. The $-\log(p)$ scores were obtained with the proposed multiple regression model. The hot region suggested the existence of QTL. The results presented in Figure 1 suggested that two QTL were located at the seventh and fifteenth markers. Interaction effects between the two QTL were also apparent. Further detailed analysis would help to establish a novel interaction model, for example, for additive or dominant effects that interacted with the cytoplasmic background.

3 Application of the model with empirical data

To assess the efficiency of the model, a data set for plant height in maize was compiled. Using two parental lines, JB and Y53, we constructed a bulked F_2 mapping population with 120 individuals from the direct cross (JB × Y53) and 120 individuals from the reciprocal cross (Y53 × JB). A genetic linkage map was constructed that contained 154 simple sequence repeat markers, which covered 1735.0 cM of the maize genome with an average marker spacing of 11.3 cM. It was noted that in field experiments a plethora of differences in growth, development, morphology, and production were exhibited between the direct and reciprocal F_1 hybrid populations, which suggested a difference in the cytoplasmic background and possible cytonuclear interaction. Plant height in the direct and reciprocal populations differed significantly ($p < 0.01$). A cytoplasmic effect was significant at -4.82 cm, which indicated that the cytoplasm from JB caused an average 4.82 cm reduction in plant height. The contribution of a cytoplasmic effect to variation in plant height was almost 9.0%, which was calculated with the formula $h_c^2 = \sigma_c^2 / v_p = c^2 / v_p$, where σ_c^2 and v_p denote variance caused by the cytoplasm and phenotypic variance, respectively.

For convenience of analysis, the linkage groups were

Table 2 Statistical power of QTL detection and estimates of QTL effects obtained from simulated data

Component	Genetic effect	True value	Estimates value ^{a)}	Standard deviation	Statistical power (%) ^{b)}
QTL1	a_1	1.0	0.98	0.15	88**
	d_1	1.0	1.06	0.19	
QTL2	a_2	0.0	–	–	81**
	d_2	1.0	1.14	0.20	
QTL1×Cytoplasm	i_{a1c}	1.5	1.51	0.15	100**
	i_{d1c}	0.0	–	–	
QTL2×Cytoplasm	i_{a2c}	1.5	1.52	0.15	72**
	i_{d2c}	1.0	1.12	0.14	
QTL1×QTL2	i_{a1a2}	1.0	1.03	0.15	100
	i_{d1d2}	1.0	1.57	0.27	81
	i_{a2d1}	0.0	–	–	–
QTL1×QTL2×Cytoplasm	i_{a1a2c}	1.5	1.08	0.18	81
	i_{d1d2c}	1.5	1.52	0.15	100
	i_{d1d2c}	1.5	1.57	0.28	89
	i_{a2d1c}	1.0	1.04	0.17	95
	i_{a1d2c}	0.0	–	–	–

a) ‘–’ indicates that the QTL effects were not detected by the proposed method. b) **, The statistical power calculated as the number of detected effective QTL divided by the 100 replicates in the simulated study. The effective QTL is defined as the QTL found in the assumed marker position with a significant additive and dominant effect simultaneously.

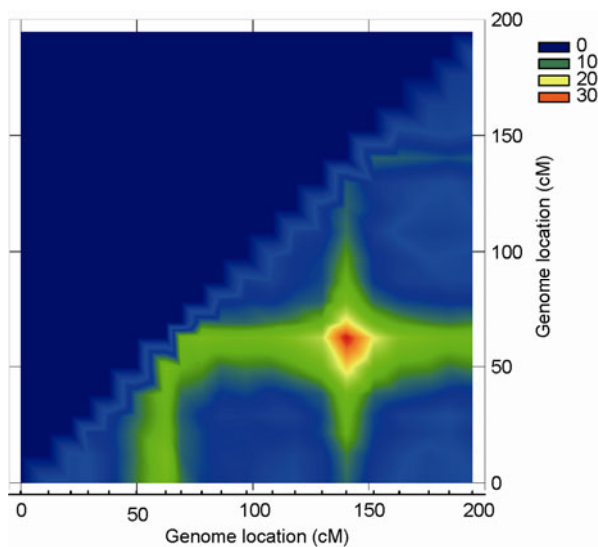


Figure 1 The $-\log(p)$ values estimated by two-dimensional genome-wide scan of simulated data shown in the lower right triangle. Calculations were performed by multiple regression of each pair of markers along the simulated chromosome. The p -values of the two-QTL model were obtained for each marker pair. The results suggest that two QTL were located at the seventh and fifteenth markers and the interaction effects between the seventh and fifteenth markers are also apparent.

linked as a single chromosome. We performed the 2D scan by multiple linear regression analysis and computing the q -value for the F -statistic of the two-QTL model for each marker pair, to form a 2D irregular grid of marker coordinates across the genome on the basis of $-\log(p)$ scores (Figure 2(a), lower right triangle). The highest peak over the 2D surface was obtained for two loci on chromosomes 3 and 4

(p -value 2.72×10^{-20}). It was interesting that chromosomes 3, 5 and 6 were indicated to be vitally important because the hot spot was always associated with these chromosomes. A single-QTL genome scan also showed evidence of the importance of chromosomes 3, 5 and 6 (Figure 2(b)). These peaks over the 2D surface suggested that at least one genetic component in the two-QTL model substantially contributes to plant height. Some peaks in the 2D surface provided evidence for intranuclear and cytonuclear interactions that significantly contributed to plant height. Some peaks suggested that QTL pairs interacted, but these loci did not have significant effects on plant height in the single-QTL genome scan (Figure 2(b), red line). In this analysis, the single-QTL genome scan was performed with the cytonuclear model, written in an interval mapping form, $y_j = m + x_{jc} + x_{1j}a + x_{2j}d + x_{3j}i_{ac} + x_{4j}i_{dc} + e_j$ ($j = 1$ to n), which was a shortened version of the model that we proposed previously [8].

The complete results included all of these genetic components screened by q -value estimation, FDR calculation and Bonferroni adjustment to control the overall significance threshold of $\alpha = 0.05$. The QTL of potential biological interest were selected and classified as high-, moderate- and low-confidence QTL on the basis of the p -value only. In this study, Bonferroni adjustment retained its conservative property and detected the fewest but most highly significant QTL, whereas the q -value estimation approach was less critical and thus identified the highest number of QTL. The FDR estimation approach seemed moderate compared with the other two methods. The most highly significant QTL and various interactions identified by the conservative Bonferroni adjustment method are summarized in Tables S1 and S2.

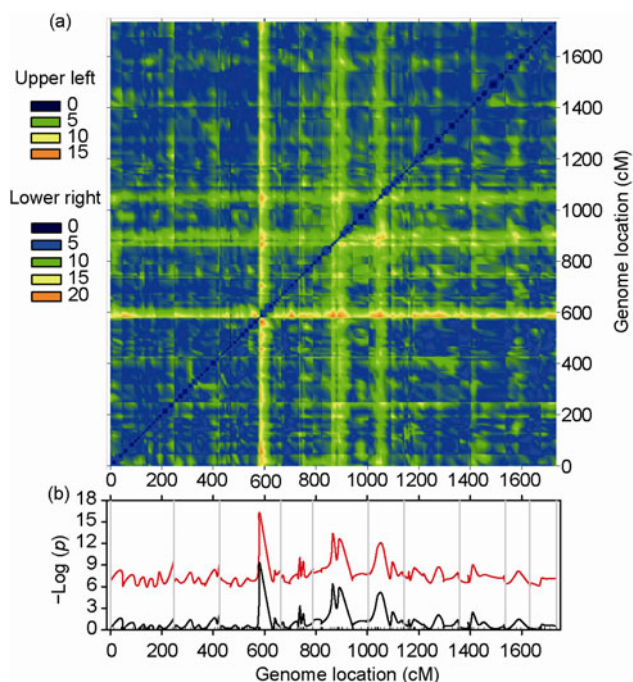


Figure 2 Genome scan for plant height in maize. (a) Two-dimensional genome-wide scan estimating $-\log(p)$ for a two-QTL genetic model. Calculations were performed by multiple regression of each pair of markers in the genome. The p -values of the two-QTL model were obtained for each marker pair. The $-\log(p)$ scores are presented in the lower right and upper left triangles for a full model and reduced model with no cytoplasmic effect included, respectively. (b) Single-QTL scan for the model suggested by Tang et al. [8]. The $-\log(p)$ profiles were obtained from single-QTL analysis of the genome for a full model (red line) and reduced model with no cytoplasmic effect included (black line).

4 Discussion

In this article, we proposed a framework for dissection of complex cytonuclear epistasis based on a reciprocal mating design. The proposed model incorporates different genetic effects, such as intranuclear interactions and QTL×QTL×cytoplasm interactions. However, this model does not provide a conclusive solution and many questions remain to be answered. An important issue for future research to resolve involves its applicability for marker analysis or QTL mapping. In the present study, we implemented the model under the assumption that each marker was linked to a QTL. Thus, it is a method for marker analysis rather than true QTL mapping. However, if the marker density is relatively high, for example with an average interval of less than 5 cM, we are confident that the QTL involved can be detected with this approach. Furthermore, the LOD score, instead of the p -value used in the present study, can be obtained by the method suggested by Haley and Knott [13]. In addition, maximum likelihood estimation and a Bayesian method can be incorporated in the model for QTL mapping.

Cytoplasmic effect acts as an important and significant covariate and should be taken into account for complex trait mapping. Consideration of cytoplasmic effect reduces re-

sidual variation and thus enhances the ability to detect QTL. In our previous model, this covariate was treated as an additive effect and was easily incorporated into an interval mapping model [8]. In the current study, cytoplasmic effect was treated as an interactive covariate and allowed to interact with both QTL, as well as with the QTL×QTL interaction. However, cytoplasmic effect, as a fixed effect, can be treated in two ways in a multiple regression analysis. One method is illustrated by the above model. Cytoplasmic effect (c) was estimated once in a regression analysis for each QTL pair, and was estimated $(k^2 - k)/2$ times, assuming the presence of k markers in the whole genome. The other method uses the equation $y' = y - c$ to calculate a new dependent variable, and then regression is implemented on the reduced model without inclusion of the c effect. The two methods have a very weak influence on parameter estimation, especially for parameters with a significant contribution (Figure 2(a)). The $-\log(p)$ profile of the 2D genome scan on the reduced model is plotted on the upper left triangle of Figure 2(a). Note that when c is reduced from the full model, the F -test statistic for the model decreases and the p -value increases correspondingly about 10^5 times. The $-\log(p)$ profile is plotted in Figure 2(b) (black line) for the single-QTL genome scan under the reduced model without inclusion of the c effect. The full and reduced models produced almost identical results, regardless of whether a 2D genome scan or single-QTL genome scan was used.

In practice, a data set often contains missing data. Two situations occurred in the present study, namely missing individual marker genotypes and a missing genotype combination for a QTL pair. Several existing methods have been suggested for instances where a marker genotype is missing, such as a hidden Markov model. However, when a genotype combination is missing, a non-full rank model produces biased results because of complex multicollinearity. To deal with this problem, variables that are a linear combination of other variables can be deleted from the regression model. For example, if only genotype $Q_1Q_1Q_2Q_2$ was missing in a QTL pair, $i_{a_2d_c}$ would confound with other effects and should be removed from the model for construction of a full rank model. Although this strategy may not be the best solution, it enabled us to obtain estimates for the other effects.

Further analysis of interesting peaks and gaps in the linkage group is required. For these interesting genomic regions, a fine-scale grid scan can be performed by insertion of a dummy marker at a fixed interval with a hidden Markov model. In our linkage map, a 40 cM gap was located on the end of chromosome 3, which interacted with chromosome 5 detected by the 2D genome scan (Figure 2). For this interesting peak, a fine-scale grid scan was implemented by inserting dummy markers at an average spacing of 5 cM under the full model (Figure 3). The highest peak over the 2D surface was increased, and the corresponding p -value was 2.73×10^{-21} , whereas the p -value was 1.01×10^{-19} in the

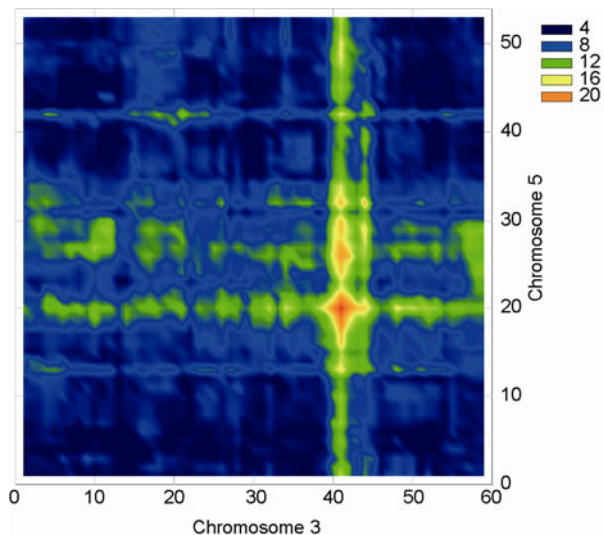


Figure 3 Fine grid two-dimensional genome-scan analysis at an average spacing of 5 cM on chromosome 3 versus chromosome 5 under the full model. The highest peak over the 2D surface was increased, with a p -value of 2.73×10^{-21} . The p -value was 1.01×10^{-19} in the same chromosome region in Figure 2(a) (lower right triangle).

same region in Figure 2(a) (lower right triangle), which indicated that a fine-scale grid scan can increase the confidence with which QTL are detected.

This work was supported by the National Basic Research Program of China (2011CB100106), the National Natural Science Foundation of China (30971846 and 31171187) and the Vital Project of Natural Science of Universities in Jiangsu Province (09KJA210002) to C. Xu, and from the National Natural Science Foundation of China (31100882) to Z. Tang.

- 1 Rand D M, Fry A, Sheldahl L. Nuclear-mitochondrial epistasis and drosophila aging: Introgression of *Drosophila simulans* mtDNA modifies longevity in *D. melanogaster* nuclear backgrounds. *Genet-*

- ics, 2006, 172: 329–341
- 2 Roubertoux P L, Sluyter F, Carlier M, et al. Mitochondrial DNA modifies cognition in interaction with the nuclear genome and age in mice. *Nat Genet*, 2003, 35: 65–69
- 3 Tao D, Hu F, Yang J, et al. Cytoplasm and cytoplasm-nucleus interactions affect agronomic traits in japonica rice. *Euphytica*, 2004, 135: 129–134
- 4 Zeyl C, Anderson B, Weninck E. Nuclear-mitochondrial epistasis for fitness in *Saccharomyces cerevisiae*. *Evolution Int J Org Evolution*, 2005, 59: 910–914
- 5 Allen J O. Effect of teosinte cytoplasmic genomes on maize phenotype. *Genetics*, 2005, 169: 863–880
- 6 Rand D M, Haney R A, Fry A J. Cytonuclear coevolution: the genomics of cooperation. *Trends Ecol Evol*, 2004, 19: 645–653
- 7 Nichols K M, Broman K W, Sundin K, et al. Quantitative Trait Loci by maternal cytoplasmic environment interaction for development rate in *Oncorhynchus mykiss*. *Genetics*, 2006, 175: 335–347
- 8 Tang Z, Wang X, Hu Z, et al. Genetic dissection of cytonuclear epistasis in line crosses. *Genetics*, 2007, 177: 669–672
- 9 Zeng Z B. Precision mapping of quantitative trait loci. *Genetics*, 1994, 136: 1457–1468
- 10 Eta-Ndu J T, Openshaw S J. Epistasis for grain yield in two F₂ populations of maize. *Crop Sci*, 1999, 39: 346–352
- 11 Li Z K, Luo L J, Mei H W, et al. Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. *Genetics*, 2001, 158: 1737–1753
- 12 Yu S B, Li J X, Xu C G, et al. Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proc Natl Acad Sci USA*, 1997, 94: 9226–9231
- 13 Haley C S, Knott S A. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, 1992, 69: 315–324
- 14 Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J Roy Stat Soc B*, 1995, 57: 289–300
- 15 Storey J D. The positive false discovery rate: A Bayesian interpretation and the q -value. *Annal Statist*, 2003, 31: 2013–2035
- 16 van den Oord E J. Controlling false discoveries in genetic studies. *Am J Med Genet B Neuropsychiatr Genet*, 2008, 147B: 637–644
- 17 van den Oord E J, Sullivan P F. False discoveries and models for gene discovery. *Trends Genet*, 2003, 19: 537–542
- 18 Carlborg O, Haley C S. Epistasis: Too often neglected in complex trait studies? *Nat Rev Genet*, 2004, 5: 618–625

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.

Supporting Information

- Table S1** Most-significant QTL main effects and their interactions with cytoplasm for plant height detected in the reciprocal F₂ population of maize
Table S2 Most-significant pairs of loci and the effects of different interaction types for plant height detected in the reciprocal F₂ population of maize

The supporting information is available online at csb.scichina.com and www.springerlink.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.