

Mapping Quantitative Trait Loci Interactions From the Maternal and Offspring Genomes

Yuehua Cui,* George Casella* and Rongling Wu*,†,1

*Department of Statistics, University of Florida, Gainesville, Florida 32611 and †College of Life Sciences, Zhejiang Forestry University, Lin'an, Zhejiang 311300, People's Republic of China

Manuscript received November 8, 2003
Accepted for publication March 15, 2004

ABSTRACT

The expression of most developmental or behavioral traits involves complex interactions between quantitative trait loci (QTL) from the maternal and offspring genomes. The maternal-offspring interactions play a pivotal role in shaping the direction and rate of evolution in terms of their substantial contribution to quantitative genetic (co)variation. To study the genetics and evolution of maternal-offspring interactions, a unifying statistical framework that embraces both the direct and indirect genetic effects of maternal and offspring QTL on any complex trait is developed. This model is derived for a simple backcross design within the maximum-likelihood context, implemented with the EM algorithm. Results from extensive simulations suggest that this model can provide reasonable estimation of additive and dominant effects of the QTL at different generations and their interaction effects derived from the maternal and offspring genomes. Although our model is framed to characterize the actions and interactions of maternal and offspring QTL affecting offspring traits, the idea can be readily extended to decipher the genetic machinery of maternal traits, such as maternal care. Our model provides a powerful means for studying the evolutionary significance of indirect genetic effects in any sexually reproductive organisms.

FOR many characteristics, particularly those expressed early in life, an individual's phenotype may be influenced by two kinds of genetic effects: *direct genetic effects* of the genes carried by that individual on its own traits and *indirect genetic effects* of genes carried by others (WOLF *et al.* 1998; LI *et al.* 1999; WOLF 2000, 2003; AGRAWAL *et al.* 2001; HAGER and JOHNSTONE 2003). Direct genetic effects occur when genes possessed by an individual directly influence that individual's phenotype, whereas indirect genetic effects occur when genes expressed in one individual have phenotypic effects in another individual. Notable indirect genetic effects were observed in many taxa, such as mice (FALCONER 1965; HAGER and JOHNSTONE 2003), fish (REZNICK 1991), plants (ROACH and WULFF 1987), and insects (MOUSSEAU and DINGLE 1991; AGRAWAL *et al.* 2001). Their simultaneous occurrence with direct genetic effects can create new sources of genetic (co)variation (MOUSSEAU and FOX 1998; WOLF 2003) and therefore can change the rate and direction of evolution (CHEVERUD 2003).

A number of quantitative genetic models have been proposed to quantify the contribution of direct genetic effects to the phenotypic variance (LYNCH and WALSH 1998). These models have now been refined to characterize and map individual genetic loci [or quantitative trait loci (QTL)] affecting complex phenotypes, using

molecular linkage maps (LANDER and BOTSTEIN 1989; WU *et al.* 2002; reviewed in JANSEN 2000). While these refined models are instrumental in the genetic mapping of QTL triggering direct effects on a variety of traits, none of them can be effectively used to map QTL of indirect genetic loci because of their failure to take into account gene actions and interaction from different genomes. Many of the current results from the mapping of direct-effect QTL may be insufficient or misleading in the characterization of genetic architecture due to the omission of the potentially important sources of genetic variance contributed by indirect genetic effects (see WOLF *et al.* 2002).

The presence of indirect genetic effects means that the phenotypic effect of a gene in an individual relies on the genes possessed by its social partners. Perhaps the most important indirect genetic effects occur between parents and their offspring (WOLF 2003). An enormous body of work on parent-offspring conflict suggests that life history traits and other development-related traits are affected by genes expressed in mothers and by maternally and paternally inherited genes expressed in offspring (AGRAWAL *et al.* 2001; HAGER and JOHNSTONE 2003). Using simple mapping approaches, some researchers have begun to map QTL for indirect effects from maternal interactions (PERIPATO and CHEVERUD 2002; PERIPATO *et al.* 2002; WOLF *et al.* 2002). For example, WOLF *et al.* (2002) identified several direct and maternal-effect QTL for offspring preweaning growth in a cross of LG/J and SM/J inbred mouse strains. They

¹Corresponding author: Department of Statistics, 533 McCarty Hall C, University of Florida, Gainesville, FL 32611. E-mail: rwu@stat.ufl.edu

further found that maternal-effect loci contribute much more strongly to the genetic variance in offspring growth than do direct-effect loci. However, results from simple mapping approaches are premature because they were not designed to understand the genetics underlying parent-offspring interactions.

Here we present a novel statistical model for characterizing the genetic effects of the maternal genome on offspring traits. In doing this, we incorporate into our mapping model epistatic interactions between one QTL from the maternal genome and other QTL from the offspring genome. In the QTL mapping study, the direct- and maternal-effect QTL mapped to distinct portions of the genome (WOLF *et al.* 2002). Epistasis, expressed as the effect of one gene contingent upon the expression of other genes, is thought to play a central role in trait evolution and speciation (DOEBLEY *et al.* 1995; LARK *et al.* 1995; WHITLOCK *et al.* 1995; PHILLIPS 1998; WADE 1998). Despite their considerable evolutionary significance (WOLF 2003), theoretical models of epistatic genetic effects resulting from maternal and offspring genomes are virtually lacking in the literature. To illustrate the idea of our mapping approach, we base our analysis on a simple backcross genetic design for an autogamous plant system. It is possible for our model to be extended to more complex designs, such as the F_2 , full-sib family, natural populations, and allogamous systems.

EXPERIMENTAL DESIGN

Suppose a backcross population is of size n , initiated with two contrasting inbred lines in autogamous plants. We use t to denote the generation of the backcross and $t + 1$ to denote the generation of offspring derived from the autogamous pollination of the backcross plants. Thus, leaves, stems, or roots of the backcross represent generation t , whereas the embryos and endosperms of its seeds represent generation $t + 1$. We use this backcross population to construct a genetic linkage map with molecular markers. Each of the members in the backcross is measured for a quantitative trait. For the traits with the same generation as the backcross plant (mother), such as plant height, traditional mapping approaches (LANDER and BOTSTEIN 1989) can be used directly to map the underlying QTL affecting plant height on the map. But for offspring traits (resulting from the embryo) with different generations from the mother, the direct use of traditional approaches may lead to biased estimation of genetic architecture because these traits, particularly when they are expressed in the mother plant, are confounded by genes from both the maternal genome and the maternally and paternally contributed offspring genome. We propose to use marker genotypes derived from the backcross population to map epistatically interacting maternal and offspring QTL for the embryo trait.

Consider two flanking markers, $\mathcal{M}_\eta(t)$ and $\mathcal{M}_{\eta+1}(t)$, derived from sporophytic plants of generation t in the

TABLE 1

Conditional probabilities of maternal QTL genotypes given maternal marker genotypes in a backcross design

Marker genotype	$Qq(t)$	$qq(t)$
$M_\eta m_\eta M_{\eta+1} m_{\eta+1}$	$\frac{(1 - r_1)(1 - r_2)}{1 - r}$	$\frac{r_1 r_2}{1 - r}$
$M_\eta m_\eta m_{\eta+1} m_{\eta+1}$	$\frac{(1 - r_1) r_2}{r}$	$\frac{r_1 (1 - r_2)}{r}$
$m_\eta m_\eta M_{\eta+1} m_{\eta+1}$	$\frac{r_1 (1 - r_2)}{r}$	$\frac{(1 - r_1) r_2}{r}$
$m_\eta m_\eta m_{\eta+1} m_{\eta+1}$	$\frac{r_1 r_2}{1 - r}$	$\frac{(1 - r_1)(1 - r_2)}{1 - r}$

r_1 and r_2 are the recombination fractions between the QTL and marker \mathcal{M}_η and $\mathcal{M}_{\eta+1}$, respectively, and r is the recombination fraction between the two markers.

backcross population. The recombination fraction between the two markers is denoted by r . A putative maternal QTL, denoted by $\mathcal{Q}(t)$ and located between these two markers (measured by the recombination fraction r_1 with \mathcal{M}_η or r_2 with $\mathcal{M}_{\eta+1}$), may affect an embryo trait of interest. Table 1 gives the conditional probabilities (π_{ij}) of backcross individual i ($i = 1, \dots, n$) carrying a maternal QTL genotype j [$j = 1$ for $Qq(t)$ and 0 for $qq(t)$], given the four marker genotypes of the backcross population, $M_\eta m_\eta M_{\eta+1} m_{\eta+1}(t)$, $M_\eta m_\eta m_{\eta+1} m_{\eta+1}(t)$, $m_\eta m_\eta M_{\eta+1} m_{\eta+1}(t)$, and $m_\eta m_\eta m_{\eta+1} m_{\eta+1}(t)$. This embryo trait is also affected by a QTL on the offspring genome of generation $(t + 1)$, denoted by $\mathcal{Q}(t + 1)$. The offspring QTL may be located on either the same marker interval or a different marker interval, $\mathcal{M}_{\eta'}(t)$ and $\mathcal{M}_{\eta'+1}(t)$. Unlike the maternal QTL, the offspring QTL will segregate into three genotypes, $QQ(t + 1)$, $Qq(t + 1)$, and $qq(t + 1)$, through the autogamous pollination of the backcross plants. We derive the conditional probabilities ($\pi_{ij'}$, $j' = 2, 1, 0$) of these three offspring QTL genotypes given four different marker genotypes of the backcross, $M_{\eta'} m_{\eta'} M_{\eta'+1} m_{\eta'+1}(t)$, $M_{\eta'} m_{\eta'} m_{\eta'+1} m_{\eta'+1}(t)$, $m_{\eta'} m_{\eta'} M_{\eta'+1} m_{\eta'+1}(t)$, and $m_{\eta'} m_{\eta'} m_{\eta'+1} m_{\eta'+1}(t)$ (Table 2). The conditional probabilities of joint maternal-offspring QTL genotypes given the marker genotypes of two different marker intervals can be expressed as the product of the corresponding conditional probabilities, *i.e.*,

$$\pi_{i(jj')} = \pi_{ij} \pi_{ij'}$$

If the maternal and offspring QTL are located within the same marker interval, we need to derive the conditional probabilities of six joint maternal-offspring QTL genotypes, $Qq(t)QQ(t + 1)$, $Qq(t)Qq(t + 1)$, $Qq(t)qq(t + 1)$, $qq(t)QQ(t + 1)$, $qq(t)Qq(t + 1)$, and $qq(t)qq(t + 1)$, given four marker genotypes of the backcross (Table 3).

Traditional quantitative genetic theory (CHEVERUD 2000) is used to model the genetic values of six joint maternal-offspring QTL genotypes. Because a backcross is considered in which there are two genotypes at a

TABLE 2

Conditional probabilities of offspring QTL genotypes given maternal marker genotypes in a backcross design

Marker genotype	$QQ(t+1)$	$Qq(t+1)$	$qq(t+1)$
$M_{\eta'}m_{\eta'}M_{\eta'+1}m_{\eta'+1}$	$\frac{(1-r'_1)(1-r'_2)}{4(1-r')}$	$\frac{(1-r'_1)(1-r'_2)}{2(1-r')}$	$\frac{(1-r'_1)(1-r'_2)+4r'_1r'_2}{4(1-r')}$
$M_{\eta'}m_{\eta'}m_{\eta'+1}m_{\eta'+1}$	$\frac{(1-r'_1)r'_2}{4r'}$	$\frac{(1-r'_1)r'_2}{2r'}$	$\frac{(1-r'_1)r'_2+4r'_1(1-r'_2)}{4r'}$
$m_{\eta'}m_{\eta'}M_{\eta'+1}m_{\eta'+1}$	$\frac{r'_1(1-r'_2)}{4r'}$	$\frac{r'_1(1-r'_2)}{2r'}$	$\frac{r'_1(1-r'_2)+4(1-r'_1)r'_2}{4r'}$
$m_{\eta'}m_{\eta'}m_{\eta'+1}m_{\eta'+1}$	$\frac{r'_1r'_2}{4(1-r')}$	$\frac{r'_1r'_2}{2(1-r')}$	$\frac{r'_1r'_2+4(1-r'_1)(1-r'_2)}{4(1-r')}$

r'_1 , r'_2 , and r' are defined similarly in Table 1.

locus, we can model only the additive effect of the maternal QTL. But a heterozygous backcross is selfed to generate three possible QTL genotypes in the embryo, which makes it possible to model the additive and dominant effects of the offspring QTL. Therefore, the maternal-offspring QTL interactions that we can model will include the additive \times additive and additive \times dominant effects. Let $a(t)$ be the additive effect of the maternal QTL, $\mathcal{Q}(t)$, and $a(t+1)$ and $d(t+1)$ be the additive and dominant effects of the offspring QTL, $\mathcal{Q}(t+1)$. These two QTL interact to form the additive \times additive (i_{aa}) and the additive \times dominant epistatic effects (i_{ad}), respectively, between the two different genomes. The genotypic means of the six joint QTL are expressed as

$$\begin{aligned}
\mu_1 &= \mu + a(t) + a(t+1) + i_{aa}, & \text{for } Qq(t)QQ(t+1), \\
\mu_2 &= \mu + a(t) + d(t+1) + i_{aa}, & \text{for } Qq(t)Qq(t+1), \\
\mu_3 &= \mu + a(t) - a(t+1) - i_{ad}, & \text{for } Qq(t)qq(t+1), \\
\mu_4 &= \mu - a(t) + a(t+1) - i_{aa}, & \text{for } qq(t)QQ(t+1), \\
\mu_5 &= \mu - a(t) + d(t+1) - i_{ad}, & \text{for } qq(t)Qq(t+1), \\
\mu_6 &= \mu - a(t) - a(t+1) + i_{aa}, & \text{for } qq(t)qq(t+1).
\end{aligned}$$

(1)

The additive, dominant, and epistatic effects of the two QTL can be estimated from the estimated genotypic means by solving the above regular equations, as can be seen in WU *et al.* (2002).

STATISTICAL METHOD

The mixture model: Statistical methods for mapping QTL based on a mixture model have been previously developed (LANDER and BOTSTEIN 1989). In the mixture model, each observation y is assumed to have arisen from one of k (k possibly unknown but finite) components, each component being modeled by a density from the parametric family f ,

$$p(y|\pi, \phi, \eta) = \pi_1 f_1(y; \phi_1, \eta) + \dots + \pi_k f_k(y; \phi_k, \eta), \quad (2)$$

where $\pi = (\pi_1, \dots, \pi_k)$ are the mixture proportions that are constrained to be nonnegative and sum to unity; $\phi = (\phi_1, \dots, \phi_k)$ are the component-specific parameters, with ϕ_i being specific to component i ; and η is a parameter (*i.e.*, residual variance) that is common to all components.

TABLE 3

Conditional probabilities of joint maternal-offspring QTL genotypes given maternal marker genotypes of the same interval in a backcross design

Marker genotype	$Qq(t)QQ(t+1)$	$Qq(t)Qq(t+1)$	$Qq(t)qq(t+1)$	$qq(t)QQ(t+1)$	$qq(t)Qq(t+1)$	$qq(t)qq(t+1)$
$M_{\eta}m_{\eta}M_{\eta+1}m_{\eta+1}$	$\frac{p_1}{4(1-r)}$	$\frac{p_1}{2(1-r)}$	$\frac{(p_1+4p_2)}{4(1-r)}$	$\frac{p_3}{4(1-r)}$	$\frac{p_3}{2(1-r)}$	$\frac{(p_3+4p_4)}{4(1-r)}$
$M_{\eta}m_{\eta}m_{\eta+1}m_{\eta+1}$	$\frac{p_5}{4r}$	$\frac{p_5}{2r}$	$\frac{(p_5+4p_6)}{4r}$	$\frac{p_7}{4r}$	$\frac{p_7}{2r}$	$\frac{(p_7+4p_8)}{4r}$
$m_{\eta}m_{\eta}M_{\eta+1}m_{\eta+1}$	$\frac{p_8}{4r}$	$\frac{p_8}{2r}$	$\frac{(4p_7+p_8)}{4r}$	$\frac{p_6}{4r}$	$\frac{p_6}{2r}$	$\frac{(4p_5+p_6)}{4r}$
$m_{\eta}m_{\eta}m_{\eta+1}m_{\eta+1}$	$\frac{p_4}{4(1-r)}$	$\frac{p_4}{2(1-r)}$	$\frac{(4p_3+p_4)}{4(1-r)}$	$\frac{p_2}{4(1-r)}$	$\frac{p_2}{2(1-r)}$	$\frac{(4p_1+p_2)}{4(1-r)}$

r_1 , r_2 , and r_3 are the recombination fractions between marker M_{η} and the maternal QTL, between the maternal and offspring QTL, and between the offspring QTL and $M_{\eta+1}$, respectively, and r is the recombination fraction between the two markers. $p_1 = \frac{1}{2}(1-r_1)(1-r_2)(1-r_3)$, $p_2 = \frac{1}{2}(1-r_1)r_2r_3$, $p_3 = \frac{1}{2}r_1r_2(1-r_3)$, $p_4 = \frac{1}{2}r_1(1-r_2)r_3$, $p_5 = \frac{1}{2}(1-r_1)(1-r_2)r_3$, $p_6 = \frac{1}{2}(1-r_1)r_2$, $p_7 = \frac{1}{2}r_1r_2r_3$, and $p_8 = \frac{1}{2}r_1(1-r_2)(1-r_3)$.

Modeling maternal-offspring interactions contains two major tasks: (1) Derive the mixture proportions (π) expressed as the conditional probabilities of joint maternal-offspring QTL genotypes given marker genotypes (Tables 1–3) and the density functions expressed in terms of six QTL genotypic means and residual variance; (2) provide estimates of unknown parameters $\Theta = (\pi, \phi, \eta)$ included in the mixture model using statistical and computational algorithms, such as the maximum-likelihood method (LANDER and BOTSTEIN 1989; WU *et al.* 2002) and the Bayesian approach (SATAGOPAN *et al.* 1996; ROBERT and CASELLA 1999). Here, we implement the expectation-maximization (EM) algorithm to obtain the maximum-likelihood estimates (MLEs) of the unknown parameters.

EM algorithm: On the basis of the mixture model (2), we formulate the likelihood of the marker data and embryo trait phenotypes controlled by the putative QTL as

$$\ell(\Theta) = \prod_{i=1}^n \left[\sum_{j=1}^6 \pi_{ij} f_j(y_i) \right], \quad (3)$$

where the unknown vector Θ contains QTL effects (Equation 1), QTL positions (Tables 1–3), and residual variance (σ^2). These parameters are solved using the EM algorithm (DEMPSTER *et al.* 1977) described below.

In the E step, the conditional probabilities (priors) of the QTL genotypes given the marker genotypes and the normal distribution function are used to calculate

$$\Pi_{ij} = \frac{\pi_{ij} f_j(y_i)}{\sum_{j=1}^6 \pi_{ij} f_j(y_i)}, \quad (4)$$

which could be thought of as a posterior probability that the j th seed of the i th backcross plant has the k th joint maternal-offspring QTL genotype.

In the M step, the calculated posterior probabilities were used to solve the unknown parameters

$$\hat{\mu}_j = \frac{\sum_{i=1}^n y_i \Pi_{ij}}{\sum_{i=1}^n \Pi_{ij}}, \quad (5)$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n \sum_{j=1}^6 (y_i - \hat{\mu}_j)^2 \Pi_{ij}}{n}. \quad (6)$$

Iterations are repeated between Equations 4 and 6 until convergence. The values at convergence are the MLEs. With the MLEs of μ_j 's, the MLEs of the overall mean, the additive, dominant, and epistatic effects of the QTL, as indicated in Equation 1, can be obtained by solving a system of regular equations. The estimation of the QTL positions can be obtained using a grid approach. This approach views the QTL positions as a known parameter in the likelihood function (3) by scanning the QTL over all marker intervals. The positions corresponding to the maximum of the log-likelihood ratio across a linkage group are the MLEs of the QTL positions.

Hypothesis tests: Our model provides hypothesis tests

for the presence of QTL affecting the expression of an offspring trait and their additive, dominant, and maternal-offspring epistatic interactions. The presence of QTL can be tested by using the following hypotheses:

$$H_0: a(t) = a(t+1) = d(t+1) = i = j = 0$$

$$H_1: \text{at least one of the parameters is not equal to zero.} \quad (7)$$

The test statistic for testing the above hypotheses is calculated as the log-likelihood ratio (LR) of the full model (H_1) over the reduced model (H_0),

$$\text{LR} = -2 \log[L(\hat{\Omega}) - L(\hat{\Omega})],$$

where $\hat{\Omega}$ and $\hat{\Omega}$ denote the MLEs of the unknown parameters under H_0 and H_1 , respectively. The LR is asymptotically χ^2 distributed with 5 d.f. However, for a practical data set with a limited sample size, the critical threshold value for declaring the presence of QTL can be empirically calculated on the basis of permutation tests (CHURCHILL and DOERGE 1994).

We can also test whether the maternal QTL exerts a significant additive (main) effect on the offspring trait by simply formulating

$$H_0: a(t) = 0$$

$$H_1: a(t) \neq 0, \quad (8)$$

whose log-likelihood-ratio test statistic is asymptotically χ^2 distributed with 1 d.f. The hypothesis about the effect of offspring's own QTL on the offspring trait is tested using

$$H_0: a(t+1) = d(t+1) = 0$$

$$H_1: \text{at least one of the parameters above is not equal to zero.} \quad (9)$$

Increasing evidence has been obtained for the role of maternal-offspring interactions in changing the rate and direction of evolution (WOLF 2003). The significance of maternal-offspring interactions (*i.e.*, i_{aa} and i_{ad} in our model) can be tested by formulating the hypotheses

$$H_0: i_{aa} = i_{ad} = 0$$

$$H_1: \text{at least one of the parameters above is not equal to zero.} \quad (10)$$

The likelihood value under H_1 of Equation 10 is the same as that under H_1 of Equation 7. The likelihood value under H_0 of Equation 10 is calculated by plugging the MLEs of μ , $a(t)$, $a(t+1)$, $d(t+1)$, and σ^2 under this null hypothesis into the likelihood function (3). The characterization of critical thresholds for hypotheses 8–10 can be empirically obtained through simulation studies.

RESULTS

We performed simulation studies to investigate the theoretical properties of our model. A linkage group

of 80 cM, composed of five equidistant markers, is simulated for a backcross population. For a given marker, there are two (maternal) genotypes, one heterozygous and the other homozygous, in the backcross. Through selfed pollination, the heterozygote produces three genotypes and the homozygote produces one homozygote in the progeny population, which leads to four joint maternal-offspring genotypes. We use a two-stage hierarchical bifurcation strategy to simulate genotypes at the five markers for a given sample size of n . The upper-hierarchical bifurcation of this strategy concerns the polynomial distribution of 2^5 marker genotypes in the backcross population, whereas the lower-hierarchical bifurcation concerns the polynomial distribution of 4^5 joint maternal-offspring marker genotypes in the autogamous progeny of the backcross plants. These two hierarchies are connected through Mendelian inheritance laws. This simulation strategy is illustrated by Figure 1 in which three genes (A , B , C) are cosegregating in a backcross and cotransmitted into the offspring generation of the backcross.

Assume that there are two QTL for an offspring trait, one from the maternal genome and the other from the offspring genome, which together account for a proportion (R^2) of the total observed variance in this trait. We consider the two QTL to be located in two different marker intervals (\mathcal{L}_1) or the same marker interval (\mathcal{L}_2). In this simulation, we consider 12 different schemes each representing a combination of three different sample sizes ($n = 200, 400, \text{ and } 800$), two different QTL-effect sizes ($R^2 = 0.1 \text{ and } 0.4$), and two different QTL locations (\mathcal{L}_1 and \mathcal{L}_2). For \mathcal{L}_1 , the maternal QTL is located at 16 cM from the left marker of the first interval and the offspring QTL is located at 8 cM from the left marker of the third interval (Figure 2). For \mathcal{L}_2 , the maternal and offspring markers are located at 8 and 16 cM from the left marker of the first interval, respectively (Figure 3). Knowing the locations of these assumed QTL, joint QTL-marker genotypes can be simulated by viewing the QTL as “markers” using the two-stage hierarchical bifurcation strategy described above (Figure 1). The phenotypic values of an assumed offspring trait are simulated as the sum of the QTL genotypic means (determined by both the maternal and offspring QTL) and normally distributed random errors.

Tables 4 and 5 give the hypothesized values of the QTL-effect parameters as shown in Equation 1, as well as their MLEs under different simulation schemes. First, both the maternal and offspring QTL can be well identified. As an example, we draw the landscapes of the log-likelihood-ratio test statistics, calculated using Equation 3, as a function of the location of these two QTL only for two extreme situations, one having small R^2 and n (Figures 2A and 3A) and the other having large R^2 and n (Figures 2B and 3B). The peaks of the landscapes from each of 100 simulation replicates are all beyond the critical thresholds at $\alpha = 0.001$ obtained from per-

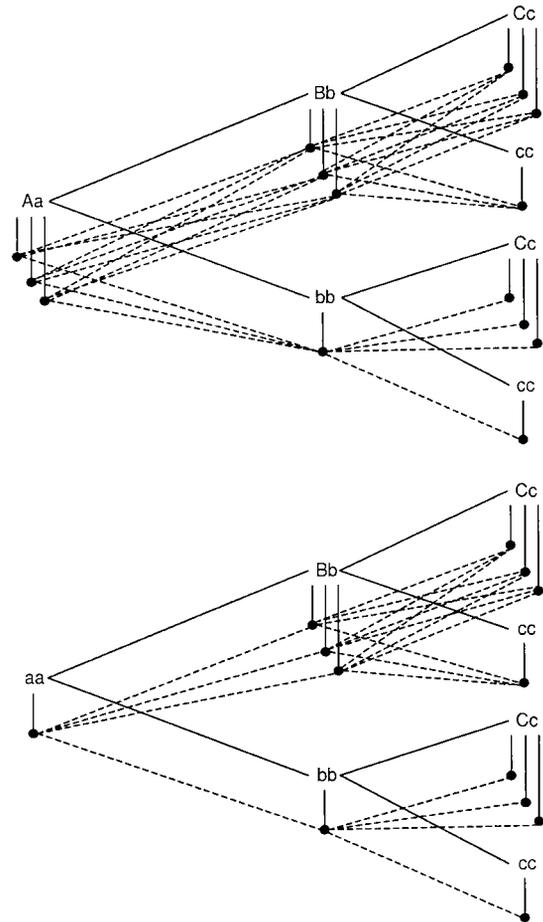


FIGURE 1.—The geometry of a two-stage hierarchical bifurcation simulation strategy for three genes, A , B , and C . The top hierarchy presents cosegregation among the three genes in the backcross (generation t), whereas the bottom hierarchy presents cosegregation among the three genes in the autogamous offspring (generation $t + 1$) of the backcross through their cotransmission from generation t to $t + 1$.

mutation tests, suggesting that our model has a power of 100% to detect the QTL hidden in our simulated dataset. The locations of both the maternal and offspring QTL can be reasonably estimated (Figures 2 and 3). Second, in general, all parameters can be quite reliably estimated using our model constructed with marker information purely from the maternal population. As expected, the additive genetic effects of the two QTL can be more reliably estimated than the dominant effects. Of different epistatic components, the additive \times additive genetic effect can be estimated better than the additive \times dominant effect. The overall mean and residual variance can be estimated precisely.

Increased R^2 and sample sizes can always improve the precision of the parameter estimation (Tables 4 and 5). But relative to the effect of sample sizes used, the increase of R^2 from 0.1 to 0.4 can lead to more significant improvement for the estimation precision, especially for the dominant and epistatic effects, than the increase of n from 200 to 400 to 800. For example, the squared

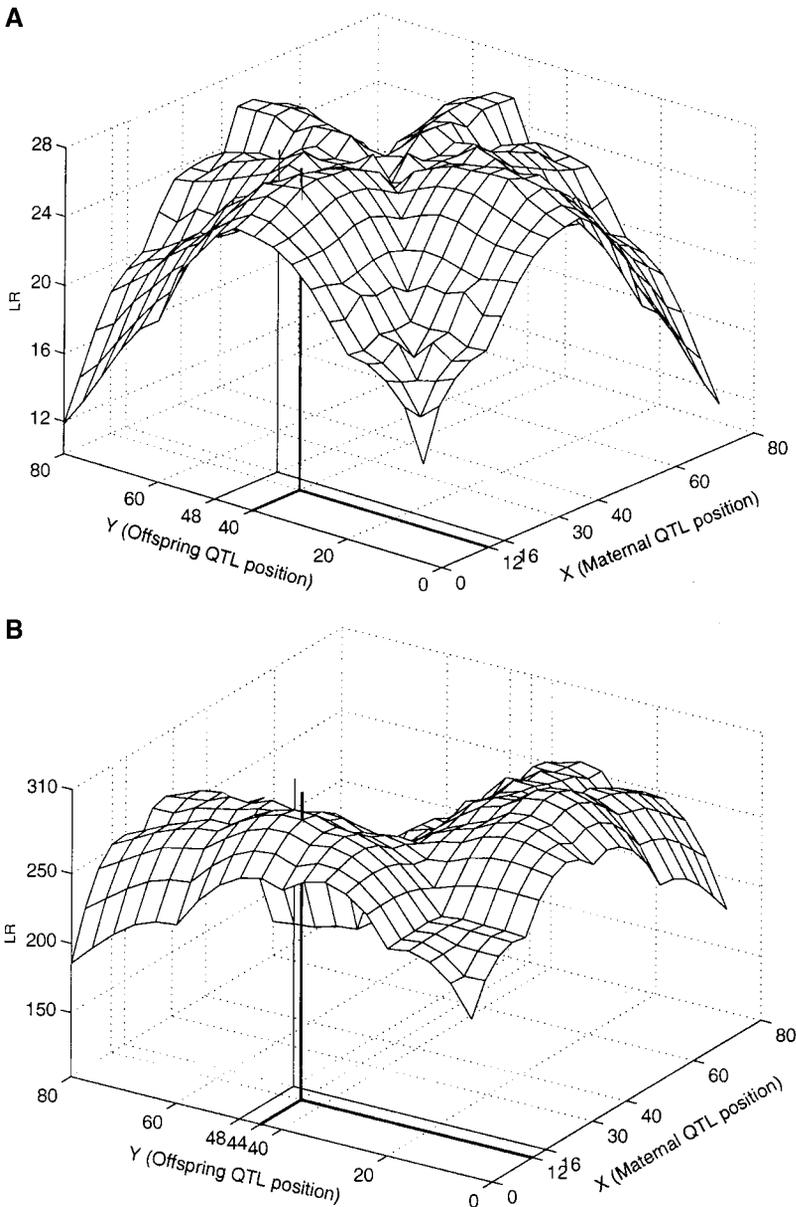


FIGURE 2.—The landscapes of log-likelihood-ratio (LR) test statistics, averaged over five randomly chosen simulation replicates, specifying the epistatic interaction between one maternal QTL (x -axis) and one offspring QTL (y -axis) located on two different marker intervals. The maximum of the estimated LR has an increased consistency with the true QTL as R^2 and n are increased from 0.1 and 200 (A) to 0.4 and 800 (B). The estimated location is indicated by the thick lines, whereas the true location is indicated by the thin lines. The critical thresholds for the existence of QTL are empirically estimated as 22 and 145 for sample sizes 200 and 800, respectively.

root of the mean square error of the MLE of the additive \times dominant genetic effect (i_{ad}) between two QTL located at different marker intervals would decrease by 217% when R^2 is increased from 0.1 to 0.4 but only by 66% when n is increased from 200 to 800. The corresponding numbers are 29% and 77% for the two QTL located at the same interval. This suggests that in practice well-managed experiments, through which residual errors are reduced and therefore R^2 is increased, are more important than simply increasing sample sizes.

When both R^2 and n are small, the estimates of the QTL parameters display similar precision for the two cases in which QTL are located at different (\mathcal{L}_1 ; Table 4) or the same marker intervals (\mathcal{L}_2 ; Table 5). The relative advantage of \mathcal{L}_1 over \mathcal{L}_2 for the precision of parameter estimation becomes more remarkable with the increased R^2 value and sample size. It seems that the maternal QTL location can be estimated more precisely

than the offspring QTL location when they are located at either different (Table 4) or the same marker intervals (Table 5). When R^2 is small, the estimation precision of the QTL locations is not very responsive to sample size. But the QTL locations can be more precisely estimated with increased sample size at a large R^2 value.

It is interesting to compare the results for genetic mapping of offspring traits from previous models and our model. LANDER and BOTSTEIN's (1989) model associates the marker genotypes with the phenotypes measured at the same generation, whereas WU *et al.*'s (2002) model associates the marker genotypes at the maternal generation with the phenotypes at the offspring generation. Unlike our model, these two models do not consider possible interactions between the QTL from the maternal and offspring genomes. The simulated data including the maternal-offspring interactions on each of the above 12 schemes were analyzed by the three

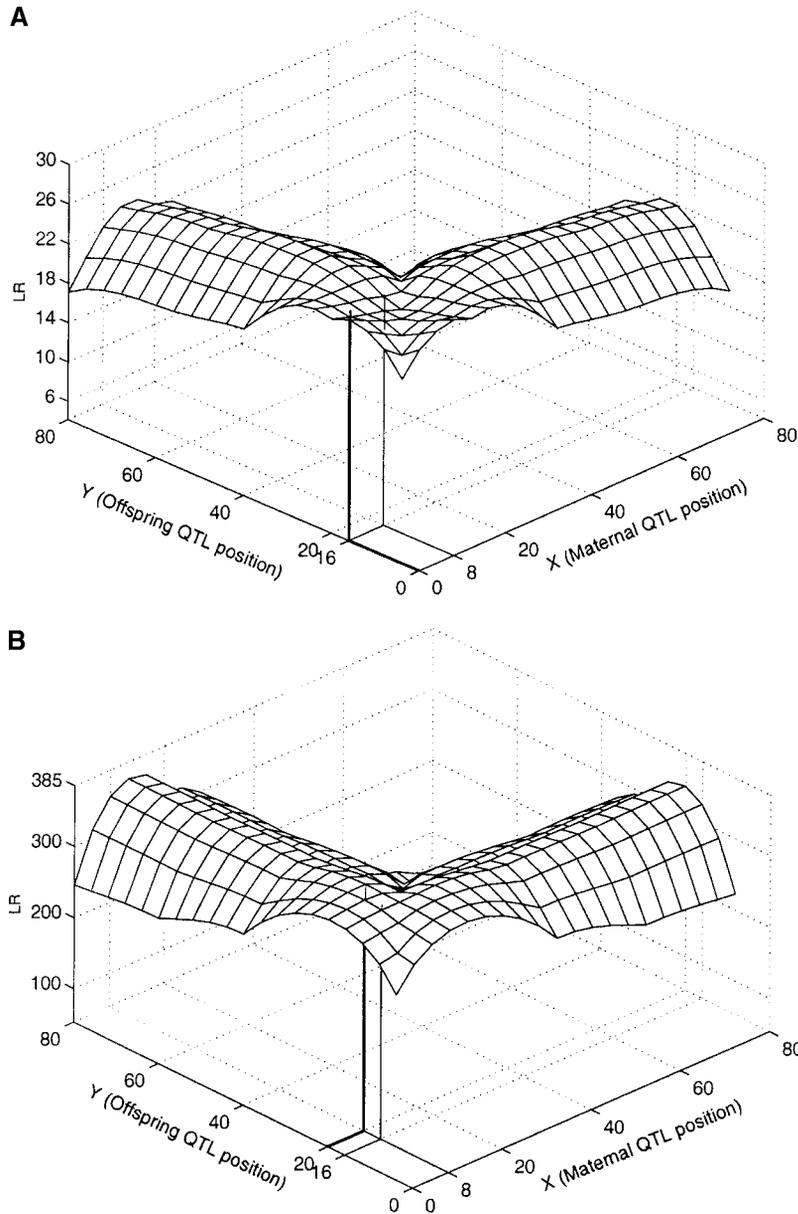


FIGURE 3.—The landscapes of log-likelihood-ratio (LR) test statistics, averaged over five randomly chosen simulation replicates, specifying the epistatic interaction between one maternal QTL (x -axis) and one offspring QTL (y -axis) located on the same marker interval. The maximum of the estimated LR has an increased consistency with the true QTL as R^2 and n are increased from 0.1 and 200 (A) to 0.4 and 800 (B). The estimated location is indicated by the thick lines, whereas the true location is indicated by the thin lines. The critical thresholds for the existence of QTL are empirically estimated as 14 and 130 for sample sizes 200 and 800, respectively.

models, aimed at mapping the QTL for the embryo. As expected, the two previous models cannot provide reasonable estimation of the embryo QTL parameters (results not shown). Also, their power to detect significant QTL was much reduced compared to our model.

We performed the hypothesis test for the significance of the additive \times additive (i_{aa}) and additive \times dominant epistatic effects (i_{ad}) between different genomes. Using the hypothesized values in Tables 4 and 5, i_{aa} and i_{ad} are detected to be significant for 80 of 100 simulation replicates at $R^2 = 0.1$ and $n = 200$. This proportion is increased markedly when the QTL explain a larger proportion of the observed variance and/or when there is a larger sample size (results not shown).

Different gene action modes are suggested to affect the estimates of QTL parameters. An additional simulation study was conducted to investigate the impact of weak main (additive and dominant) effects *vs.* strong

maternal-offspring interaction effects on the precision of parameter estimation. We consider an intermediate sample size (400), two different R^2 values, and two different QTL distribution patterns (Table 6). The simulation results are summarized below. First, both maternal-offspring additive \times additive (i_{aa}) and additive \times dominant effects (i_{ad}) can be better estimated when these effects are large. Second, large maternal-offspring interaction effects have no marked effects on the estimation of the additive or dominant effects. Third, large maternal-offspring effects would reduce the estimation precision of the two QTL located at the same interval.

DISCUSSION

We have proposed a new theoretical model for estimating the epistatic effects between maternal and offspring genomes on a complex trait expressed in the

TABLE 4

The MLEs of the QTL position and effect parameters between a maternal QTL and an offspring QTL each at a different interval based on 100 simulation replicates

R^2	n	Positions (16, 48)	$\mu = 4$	$a(t) = 0.5$	$a(t + 1) = 0.5$	$d(t + 1) = 0.3$	$i_{aa} = 0.2$	$i_{ad} = 0.2$	σ_ϵ^2
0.1	200	(16.56, 46.64) (6.3051, 7.7526)	3.9834 (0.5513)	0.2967 (0.6374)	0.4390 (0.6745)	0.3140 (0.9601)	-0.1250 (0.7814)	0.5035 (1.1930)	3.9340 (0.7324)
	400	(16.52, 48.56) (5.8394, 6.3815)	3.9769 (0.4731)	0.4633 (0.5360)	0.4406 (0.5118)	0.3044 (0.8729)	0.0935 (0.5504)	0.2069 (0.9507)	4.1451 (0.5106)
	800	(16.92, 49.04) (5.3477, 5.6272)	3.9484 (0.3334)	0.4740 (0.3920)	0.4285 (0.3649)	0.3423 (0.5807)	0.1449 (0.4097)	0.2479 (0.7202)	4.2651 (0.3395)
0.4	200	(16.12, 47.04) (5.0050, 6.4315)	3.8889 (0.2522)	0.5162 (0.2152)	0.3753 (0.2910)	0.4466 (0.4419)	0.2085 (0.2240)	0.2006 (0.3763)	0.6698 (0.1162)
	400	(16.84, 48.12) (3.5722, 3.8350)	3.9194 (0.1811)	0.4799 (0.1719)	0.4012 (0.2064)	0.4512 (0.3636)	0.1748 (0.2007)	0.2388 (0.2975)	0.7143 (0.0801)
	800	(16.24, 47.88) (2.2740, 2.5741)	3.9386 (0.1512)	0.4893 (0.1253)	0.4356 (0.1624)	0.4071 (0.2651)	0.1828 (0.1286)	0.2144 (0.2515)	0.7218 (0.0575)

The square roots of the mean squared errors of the MLEs are given in parentheses. The locations of the two QTL are described by the map distances (in centimorgans) from the first marker of the linkage group (80 cM long). The hypothesized σ_ϵ^2 value is 4.4550 for $R^2 = 0.1$ and 0.7425 for $R^2 = 0.4$.

offspring. A considerable body of literature in animals supports the view that the maternal and offspring genomes interact to determine the developmental processes of offspring (LI *et al.* 1999; AGRAWAL *et al.* 2001; HAGER and JOHNSTONE 2003; WOLF 2003). Maternal-offspring interactions have long been thought to be

more ignorable in plants than in animals, but increasing evidence suggests that they are more common and of greater evolutionary importance in plants than originally appreciated (VIELLE-CALZADA *et al.* 2000; WEIJERS *et al.* 2001). Our model will have great implications for estimating maternal-offspring interactions in any sexu-

TABLE 5

The MLEs of the QTL position and effect parameters between a maternal QTL and an offspring QTL at the same marker interval based on 100 simulation replicates

R^2	n	Positions (6, 20)	$\mu = 4$	$a(t) = 0.5$	$a(t + 1) = 0.5$	$d(t + 1) = 0.3$	$i_{aa} = 0.2$	$i_{ad} = 0.2$	σ_ϵ^2
0.1	200	(4.76, 20.68) (6.7551, 6.6743)	4.4348 (0.7811)	0.3281 (0.5977)	0.8042 (0.7402)	-0.4346 (1.3426)	-0.0799 (0.7301)	0.3225 (1.0737)	3.9297 (0.7634)
	400	(5.60, 19.84) (6.6502, 6.4708)	4.3127 (0.6108)	0.3932 (0.5619)	0.6732 (0.7336)	-0.4312 (1.2568)	-0.0476 (0.6069)	0.4346 (1.0666)	4.2286 (0.4763)
	800	(6.12, 20.20) (6.1861, 6.1219)	4.3145 (0.5093)	0.3713 (0.4963)	0.6400 (0.5465)	-0.2878 (1.0472)	-0.1276 (0.5486)	0.3455 (0.8304)	4.3284 (0.3325)
0.4	200	(5.28, 18.28) (5.9836, 4.9674)	4.2278 (0.3662)	0.4313 (0.3247)	0.6401 (0.3559)	-0.0609 (0.6551)	0.0396 (0.3419)	0.2098 (0.6072)	0.6845 (0.1244)
	400	(6.64, 18.64) (3.5020, 4.9165)	4.2431 (0.3462)	0.3761 (0.2909)	0.6800 (0.3188)	-0.0190 (0.4770)	0.0112 (0.2909)	0.2912 (0.4671)	0.7163 (0.0767)
	800	(6.48, 17.28) (2.9502, 3.5934)	4.1396 (0.2384)	0.3994 (0.2481)	0.6282 (0.2490)	0.1541 (0.3895)	0.0868 (0.2327)	0.2551 (0.3332)	0.7318 (0.0510)

The squared roots of the mean square errors of the MLEs are given in parentheses. The locations of the two QTL are described by the map distances (in centimorgans) from the first marker of the linkage group (80 cM long). The hypothesized σ_ϵ^2 value is 4.4550 for $R^2 = 0.1$ and 0.7425 for $R^2 = 0.4$.

TABLE 6

The MLEs of the QTL position and effect parameters between a maternal QTL and an offspring QTL with weak main effect and strong interaction effects based on 100 simulation replicates

R^2	n	Positions (16, 48)	$\mu = 4$	$a(t) = 0.3$	$a(t + 1) = 0.3$	$d(t + 1) = 0.2$	$i_{aa} = 0.5$	$i_{ad} = 0.3$	σ_{ϵ}^2
0.1	400	(14.72, 47.40) (6.0154, 6.5441)	3.9745 (0.4028)	0.2627 (0.3361)	0.2603 (0.3970)	0.2596 (0.7584)	0.4362 (0.3782)	0.3659 (0.6447)	2.4511 (0.2837)
0.4	400	(15.92, 47.76) (3.5505, 4.0201)	3.9824 (0.1545)	0.2252 (0.1665)	0.2808 (0.1611)	0.2307 (0.2848)	0.4200 (0.1702)	0.4235 (0.2839)	0.4173 (0.0470)
R^2	n	Positions (8, 16)	$\mu = 4$	$a(t) = 0.3$	$a(t + 1) = 0.3$	$d(t + 1) = 0.2$	$i_{aa} = 0.5$	$i_{ad} = 0.3$	σ_{ϵ}^2
0.1	400	(7.32, 11.32) (8.3264, 9.5273)	4.4601 (0.6416)	0.1718 (0.4636)	0.5070 (0.5498)	-0.5781 (1.1378)	0.1085 (0.6158)	0.4049 (0.7679)	2.4517 (0.2986)
0.4	400	(5.16, 9.16) (5.3408, 8.2003)	4.2395 (0.3089)	0.0757 (0.3053)	0.5212 (0.3047)	-0.0455 (0.3936)	0.2536 (0.3187)	0.5164 (0.3845)	0.4235 (0.0475)

The square roots of the mean squared errors of the MLEs are given in parentheses. The locations of the two QTL are described by the map distances (in centimorgans) from the first marker of the linkage group (80 cM long). The hypothesized σ_{ϵ}^2 value is 2.6125 for $R = 0.1$ and 0.4354 for $R = 0.4$.

ally reproductive organisms and make it possible to decipher the genetic basis of this important phenomenon (AGRAWAL *et al.* 2001; HAGER and JOHNSTONE 2003; WOLF 2003) at the individual QTL level.

Unlike traditional mapping approaches that were all developed to estimate direct genetic effects, our model incorporates both direct and indirect genetic effects into a QTL mapping framework and displays two significant advantages: (1) The results from our model should be closer to biological reality, given the fact that indirect genetic effects account for a great proportion (50% or more) of the genetic variance (WOLF *et al.* 2002; WOLF 2003), and (2) the model is statistically more powerful because more genetic information (indirect genetic effects) is used. As shown by extensive simulation studies, our model is quite robust in that the QTL parameters can be reliably estimated for a modest sample size (200) and low R^2 value (0.1). Increased sample sizes and R^2 values can improve the power to detect QTL and the estimation precision of QTL parameters. It is interesting to note that increasing R^2 may be more important for precise parameter estimation than increasing sample sizes to the same extent. The increased R^2 values may result from more precise phenotyping of a quantitative trait or weaker interaction between the underlying QTL and the environment.

Our strategy for mapping epistatic QTL from the maternal and offspring genomes is based on the marker genotypes derived purely from the maternal genome. Although it is no problem for the maternal markers to infer the maternal QTL, as demonstrated in traditional mapping approaches (LANDER and BOTSTEIN 1989), such a one-stage sampling strategy is limited to predict the offspring QTL. This is because the marker information and offspring QTL are at different generations and

undergo Mendelian transmission. In a study of mapping the endosperm of cereals, WU *et al.* (2002) proposed a two-stage hierarchical genotyping strategy in which both the maternal and offspring genomes are genotyped at the same set of molecular markers. It is likely that the incorporation of Wu *et al.*'s two-stage hierarchical genotyping strategy can provide more unique information to capture gene transmission.

Our model proposed here is based on a simple backcross design for an autogamous plant system. It is not difficult to extend the model to other reproductive systems, such as allogamous and mixed-pollinated and other mapping populations. More interesting, our model can also be modified to characterize the genes affecting maternal care (PERIPATO and CHEVERUD 2002). The amount of care provided by parents is determined through a complex interaction of offspring signals and responses by parents to those signals (AGRAWAL *et al.* 2001). Variation in maternal care results from two distinct genetic sources: variation among offspring in their ability to elicit care and variation among parents in their response to offspring signals. By estimating the genetic loci for maternal care derived from these two sources, we can test for one of the most fundamental genetic assumptions in the family conflict theory: *i.e.*, offspring signaling is negatively genetically correlated with maternal provisioning (AGRAWAL *et al.* 2001; WOLF 2003).

Understanding the maternal and paternal genetic regulation of offspring development and/or maternal care helps to answer many fundamental evolutionary questions in both animals and plants. Our model can be applied immediately to evolutionary genetic studies. But it needs to consider the patterns of gene segregation and transmission in natural plant populations. Parameters characterizing population structure and organiza-

tion, such as allele frequencies, linkage disequilibrium, and haplotype frequencies, should be incorporated into our mapping model. LOU *et al.* (2003) have recently derived a closed-form solution for precise and efficient estimates of haplotype frequencies. By incorporating Lou *et al.*'s model, we will be closer to unraveling the genetic basis of embryogenesis and organ development in natural animal and plant populations.

This work is supported by an Outstanding Young Investigator Award of the National Natural Science Foundation of China (30128017), a University of Florida Research Opportunity Fund (02050259), and a University of South Florida Biodefense grant (7222061-12) to R.W. The publication of this manuscript was approved as Journal Series no. R-10064 by the Florida Agricultural Experiment Station.

LITERATURE CITED

- AGRAWAL, A. F., E. D. BRODIE, III and J. BROWN, 2001 Parent-offspring coadaptation and the dual genetic control of maternal care. *Science* **292**: 1710–1712.
- CHEVERUD, J. M., 2000 Detecting epistasis among quantitative trait loci, pp. 58–81 in *Epistasis and the Evolutionary Process*, edited by J. B. WOLF, E. D. BRODIE, III and M. J. WADE. Oxford University Press, New York.
- CHEVERUD, J. M., 2003 Evolution in a genetically heritable social environment. *Proc. Natl. Acad. Sci. USA* **100**: 4357–4359.
- CHURCHILL, G. A., and R. W. DOERGE, 1994 Empirical threshold values for quantitative trait mapping. *Genetics* **138**: 963–971.
- DEMPSTER, A. P., N. M. LAIRD and D. B. RUBIN, 1977 Maximum likelihood from incomplete data via EM algorithm. *J. R. Stat. Soc. Ser. B* **39**: 1–38.
- DOEBLEY, J., A. STEG and C. GUSTUS, 1995 *teosinte branched1* and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics* **141**: 333–346.
- FALCONER, D. S., 1965 Maternal effects and selection response, pp. 763–774 in *Genetics Today: Proceedings of the 11th International Congress of Genetics*, edited by S. J. GEERTS. Pergamon Press, Oxford.
- HAGER, R., and R. A. JOHNSTONE, 2003 The genetic basis of family conflict resolution in mice. *Nature* **421**: 533–535.
- JANSEN, R. C., 2000 Quantitative trait loci in inbred lines, pp. 567–597 in *Handbook of Statistical Genetics*, edited by D. J. BALDING, M. BISHOP and C. CANNINGS. John Wiley & Sons, New York.
- LANDER, E. S., and D. BOTSTEIN, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185–199.
- LARK, K. G., K. CHASE, F. ADLER, L. M. MANSUR and J. H. ORF, 1995 Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. *Proc. Natl. Acad. Sci. USA* **92**: 4656–4660.
- LI, L.-L., E. B. KEVERNE, S. A. APARICIO, F. ISHINO, S. C. BARTON *et al.*, 1999 Regulation of maternal behavior and offspring growth by paternally expressed *Peg3*. *Science* **284**: 330–333.
- LOU, X.-Y., G. CASELLA, R. C. LITTELL, M. K. C. YANG and R. L. WU, 2003 A haplotype-based algorithm for multilocus linkage disequilibrium mapping of quantitative trait loci with epistasis in natural populations. *Genetics* **163**: 1533–1548.
- LYNCH, M., and B. WALSH, 1998 *Genetics and Analysis of Quantitative Traits*. Sinauer, Sunderland, MA.
- MOUSSEAU, T. A., and H. DINGLE, 1991 Maternal effects in insect life histories. *Annu. Rev. Entomol.* **36**: 511–534.
- MOUSSEAU, T. A., and C. FOX, 1998 *Maternal Effects as Adaptations*. Oxford University Press, New York.
- PERIPATO, A. C., and J. M. CHEVERUD, 2002 Genetic influences on maternal care. *Am. Nat.* **160**: S173–S185.
- PERIPATO, A. C., R. A. DE BRITTO, T. T. VAUGHN, L. S. PLETSCHER, S. R. MATIOLIA *et al.*, 2002 Quantitative trait loci for maternal performance for offspring survival in mice. *Genetics* **162**: 1341–1353.
- PHILLIPS, P. C., 1998 The language of gene interaction. *Genetics* **149**: 1167–1171.
- REZNICK, D. N., 1991 Maternal effects in fish life histories, pp. 780–793 in *The Unity of Evolutionary Biology: Proceedings of the Fourth International Congress of Systematic and Evolutionary Biology*, Vol. II, edited by E. C. DUDLEY. Dioscorides Press, Portland, OR.
- ROACH, D. A., and R. D. WULF, 1987 Maternal effects in plants. *Annu. Rev. Ecol. Syst.* **18**: 209–235.
- ROBERT, C. P., and G. CASELLA, 1999 *Monte Carlo Statistical Methods*. Springer, New York.
- SATAGOPAN, J. M., Y. S. YANDELL, M. A. NEWTON and T. C. OSBORN, 1996 A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. *Genetics* **144**: 805–816.
- VIELLE-CALZADA, J.-P., R. BASKAR and U. GROSSNIKLAUS, 2000 Delayed activation of the paternal genome during seed development. *Nature* **404**: 91–94.
- WADE, M. J., 1998 The evolutionary genetics of maternal effects, pp. 5–21 in *Maternal Effects as Adaptations*, edited by T. MOUSSEAU and C. FOX. Oxford University Press, Oxford.
- WEIJERS, D., N. GELDNER, R. OFFRINGA and G. JORGENS, 2001 Early paternal gene activity in *Arabidopsis*. *Nature* **414**: 709–710.
- WHITLOCK, M. C., P. C. PHILLIPS, F. B. G. MOORE and S. J. TONSOR, 1995 Multiple fitness peaks and epistasis. *Annu. Rev. Ecol. Syst.* **26**: 601–629.
- WOLF, J. B., 2000 Gene interactions from maternal effects. *Evolution* **54**: 1882–1898.
- WOLF, J. B., 2003 Genetic architecture and evolutionary constraint when the environment contains genes. *Proc. Natl. Acad. Sci. USA* **100**: 4655–4660.
- WOLF, J. B., E. D. BRODIE, III, J. M. CHEVERUD, A. J. MOORE and M. J. WADE, 1998 Evolutionary consequences of indirect genetic effects. *Trends Ecol. Evol.* **13**: 64–69.
- WOLF, J. B., T. T. VAUGHN, L. S. PLETSCHER and J. M. CHEVERUD, 2002 Contribution of maternal effect QTL to genetic architecture of early growth in mice. *Heredity* **89**: 300–310.
- WU, R. L., C.-X. MA, M. GALLO-MEAGHER, R. C. LITTELL and G. CASELLA, 2002 Statistical methods for dissecting triploid endosperm traits using molecular markers: an autogamous model. *Genetics* **162**: 875–892.

Communicating editor: J. B. WALSH