Genet. Res., Camb. (2005), **86**, *pp.* 65–75. *With 2 figures.*
doi:10.1017/S0016672305007615 Printed in the United Kingdom

Statistical model for characterizing epistatic control of triploid endosperm triggered by maternal and offspring QTLs

YUEHUA CUI AND RONGLING WU*
Department of Statistics, University of Florida, Gainesville, FL 32611, USA

(*Received 14 March 2005 and in revised form 13 May 2005*)

Summary

To study the effects of maternal and endosperm quantitative trait locus (QTL) interaction on endosperm development, we derive a two-stage hierarchical statistical model within the maximum-likelihood context, implemented with an expectation-maximization algorithm. A model incorporating both maternal and offspring marker information can improve the accuracy and precision of genetic mapping. Extensive simulations under different sampling strategies, heritability levels and gene action modes were performed to investigate the statistical properties of the model. The QTL location and parameters are better estimated when two QTLs are located at different intervals than when they are located at the same interval. Also, the additive effect of the offspring QTLs is better estimated than the additive effect of the maternal QTLs. The implications of our model for agricultural and evolutionary genetic research are discussed.

1. Introduction

As a pivotal process in the life cycle of an angiosperm, seed development is initiated by a process of double fertilization in which the embryo and the endosperm develop (Chaudhury *et al*., 2001). In diploid higher plants, the embryo (a diploid zygote) results from the fertilization of the haploid egg by one of the sperm cells. By contrast, the triploid endosperm develops when the maternal homodiploid central cell is fertilized by another sperm cell. The embryo and endosperm tissues, each with a different ploidy, develop in a coordinated manner inside the maternal ovule tissues surrounded by the diploid sporophytic layers of inner and outer integuments. Thus, the interplay of different genomes from the parental and offspring generations defines seed development (Chaudhury & Berger, 2001). The understanding of gene action and interaction in tissues of different ploidies using molecular linkage maps has become possible with the recent advent of powerful statistical models.

Lander and Botstein (1989) developed a maximum-likelihood-based approach for mapping the genes responsible for a quantitative trait (a quantitative trait locus (QTL)) based on two flanking markers in a segregating population. This approach assumes that the marker genotypes are collected from the tissue of the same generation on which a putative QTL is expressed and therefore cannot be directly used to map the embryo or endosperm QTLs with marker information from the maternal (sporophytic) plants. Wu *et al*. (2002*a*) modified Lander and Botstein's approach to characterize the genetic effects of endosperm QTLs on endosperm traits by considering the generation difference between the endosperm and sporophytic plant. More recently, the joint control of the maternal and offspring genomes on embryo traits, as documented in empirical studies (Chaudhury *et al*., 2001), has been incorporated in the QTL-mapping framework (Cui *et al*., 2004), aimed at the precise characterization of the genetic mechanisms underlying seed development.

In this article, we develop a new statistical model for mapping QTLs that affect triploid endosperm traits in higher plants. Unlike the previous model by Cui *et al*. (2004), the model presented here adopts a two-stage hierarchical sampling strategy for genotyping markers from both the maternal and offspring generations. We are especially interested in characterizing epistatic effects between different QTLs

* Corresponding author. Department of Statistics, 533 McCarty Hall C, University of Florida, Gainesville, FL 32611, USA. Tel: +1 (352) 392 3806. Fax: +1 (352) 392 8555. e-mail: rwu@stat.ufl.edu

from the maternal and offspring genomes, because QTL interactions are thought to play a central role in development and evolution (Doebley *et al*., 1995; Lark *et al*., 1995; Whitlock *et al*., 1995; Phillips, 1998; Cheverud, 2000). We derive a mixture-based likelihood function to model the gene action and interaction effects on endosperm traits between the QTLs from different genomes, and implement the expectation maximization (EM) algorithm proposed by Dempster *et al*. (1977) to provide a solution to the likelihood equations. The model incorporating both the maternal and offspring marker information can improve the accuracy and precision of genetic mapping. Extensive simulations were performed to investigate the statistical behaviour of the model. Our model provides a powerful tool with which to study the genetic epistatic effects of the maternal and offspring genomes on endosperm traits in agricultural crops, and can also be used to study the genetic significance of double fertilization in the evolution of higher plants.

2. Modelling maternal and offspring QTLs

(i) *Experimental design*

Consider a backcross population of size n derived from two contrasting inbred lines for an autogamous species. For backcross plant i ($i=1, \ldots, n$), n_i seeds are collected to measure an endosperm trait from each seed and to screen the marker genotypes of this plant and its self-pollinated offspring. Given the complexity of genotyping the triploid endosperm, we will screen marker genotypes from the seedlings that develop from the diploid embryos. In total, the number of seeds is $m=\sum_{i=1}^{n} n_i$. We denote the (maternal) backcross plant as generation t and the embryo and endosperm as generation $t+1$. To identify the QTL affecting the triploid endosperm, we can either use a simple one-stage design in which marker genotypes are derived only from the diploid maternal plants or use a two-stage hierarchical design in which marker genotypes are derived from both the maternal plants and their embryos. Clearly, the two-stage hierarchical design is more precise because it considers the within-family variation of a backcross plant (Wu *et al*., 2002*a*). As a result of this, our analysis and modelling of QTL interactions from the maternal and offspring genomes is based on the two-stage hierarchical model.

Consider two flanking markers $\mathbf{M_1}(t)$ and $\mathbf{M_2}(t)$ genotyped from the backcross plants. The recombination fraction between these two markers is denoted by r. Suppose there is a maternal QTL, denoted $Q(t)$ and located between these two markers (measured by the recombination fraction r_1 with $\mathbf{M_1}(t)$ and r_2 with $\mathbf{M_2}(t)$), which affects the endosperm trait. The maternal QTL has two genotypes expressed as $Qq(t)$ and $qq(t)$. The conditional probability $\pi_{ij_1}(t)$, of a backcross plant i carrying a maternal QTL genotype j_1, conditional upon the four maternal marker genotypes $M_1m_1M_2m_2(t)$, $M_1m_1m_2m_2(t)$, $m_1m_1M_2m_2(t)$ and $m_1m_1m_2m_2(t)$, can be derived and is shown in Table 1.

Table 1. *The frequencies of material marker genotypes and joint frequencies of maternal marker–QTL genotypes in a backcross design, used for the calculation of the conditional probabilities of maternal QTL genotypes given marker genotypes*

Marker		QTL*	
Genotype	Frequency	$Qq(t)$	$qq(t)$
$M_1m_1M_2m_2(t)$	$1-r$	$(1-r_1)\,(1-r_2)$	r_1r_2
$M_1m_1m_2m_2(t)$	r	$(1-r_1)r_2$	$r_1(1-r_2)$
$m_1m_1M_2m_2(t)$	r	$r_1(1-r_2)$	$(1-r_1)r_2$
$m_1m_1m_2m_2(t)$	$1-r$	r_1r_2	$(1-r_1)\,(1-r_2)$

* r_1 and r_2 are the recombination fractions between marker $\mathbf{M_1}$ and the QTL, and between the QTL and $\mathbf{M_2}$, respectively, and r is the recombination fraction between the two markers.

Suppose the endosperm trait is also affected by an offspring QTL, denoted by $Q(t+1)$, on the endosperm genome of the backcross population. During double fertilization, this offspring QTL generates four different triploid genotypes $QQQ(t+1)$, $QQq(t+1)$, $Qqq(t+1)$ and $qqq(t+1)$ of equal frequency if the maternal plant is a heterozygote $Qq(t)$, or one triploid genotype $qqq(t+1)$ if the maternal plant is a homozygote $qq(t)$. The offspring QTL can be predicted by a pair of flanking markers derived from the maternal and offspring genomes. Joint maternal (in generation t) and embryonic (in generation $t+1$) marker genotypes at two flanking markers, $\mathbf{M_1}$ and $\mathbf{M_2}$, form the basic framework of a two-stage hierarchical model as described by Wu *et al*. (2002*a*). We derive the conditional probability $\pi_{k_ij_2}(t+1)$, of an endosperm k ($k=1, \ldots, n_i$) from backcross plant i carrying endosperm QTL genotype j_2, conditional upon joint maternal and embryo marker genotypes (Table 2).

Both the maternal ($Q(t)$) and endosperm ($Q(t+1)$) QTLs can be located either on the same marker interval or on different marker intervals. These two QTLs form eight joint genotypes across the two different generations ($Qq(t)QQQ(t+1)$, $Qq(t)QQq(t+1)$, $Qq(t)Qqq(t+1)$, $Qq(t)qqq(t+1)$, $qq(t)QQQ(t+1)$, $qq(t)QQq(t+1)$, $qq(t)Qqq(t+1)$ and $qq(t)qqq(t+1)$) numbered 1–8, respectively. We assume that these two QTLs located on the genomes with different generations epistatically affect the endosperm trait. If they are located on different marker intervals, $\mathbf{M_1}$–$\mathbf{M_2}$ for $Q(t)$ and $\mathbf{M_1'}$–$\mathbf{M_2'}$ for $Q(t+1)$, the conditional probability matrix, $\pi_{k_ij_1j_2}$, of the joint QTL genotype for endosperm k derived from

https://doi.org/10.1017/S0016672305007615 Published online by Cambridge University Press

Table 2. *The frequencies of two-stage hierarchical marker genotypes from the backcross plants and their embryos and joint frequencies of marker–endosperm QTL genotypes in a backcross design, used for the calculation of the conditional probabilities of endosperm QTL genotypes, conditional upon two-stage hierarchical marker genotypes from the backcross plants and their embryos*

Two-stage hierarchical marker genotype			Joint marker–endosperm QTL genotype frequency*			
Maternal	Embryo	Frequency	$QQQ(t+1)$	$QQq(t+1)$	$Qqq(t+1)$	$qqq(t+1)$
$M_1'm_1'M_2'm_2'(t)$	$M_1'M_1'M_2'M_2'(t+1)$	θ^3	$\theta_1^3\theta_2^3$	$r_1r_2\theta_1^2\theta_2^2$	$r_1r_2\theta_1^2\theta_2^2$	$r_1r_2(r_1r_2\theta_1\theta_2+\theta^2)$
	$M_1'M_1'M_2'(t+1)$	$2r\theta^2$	$2r_2\theta_1^3\theta_2^2$	$r_1\theta_1^2\theta_2\varphi_2$	$r_1\theta_1^2\theta_2\varphi_2$	$2r_1r_2(r_1\theta_1\theta_2^2+r\theta)$
	$M_1'M_1'm_2'm_2'(t+1)$	$r^2\theta$	$r_2^2\theta_1^3\theta_2$	$r_1r_2\theta_1^2\theta_2^2$	$r_1r_2\theta_1^2\theta_2^2$	$r_1(r_1\theta_1\theta_2^3+r_2r^2)$
	$M_1'm_1'M_2'M_2'(t+1)$	$2r\theta^2$	$2r_1\theta_1^2\theta_2^3$	$r_2\theta_1\theta_2^2\varphi_1$	$r_2\theta_1\theta_2^2\varphi_1$	$2r_1r_2(r_2\theta_1^2\theta_2+r\theta)$
	$M_1'm_1'M_2'm_2'(t+1)$	2ϕ	$4r_1r_2\theta_1^2\theta_2^2$	$\theta_1\theta_2\varphi_1\varphi_2$	$\theta_1\theta_2\varphi_1\varphi_2$	$2r_1r_2(2\theta_1^2\theta_2^2+\varphi)$
	$m_1'm_1'm_2'm_2'(t+1)$	$2r\theta^2$	$2r_1r_2^2\theta_1^2\theta_2$	$r_2\theta_1\theta_2^2\varphi_1$	$r_2\theta_1\theta_2^2\varphi_1$	$2r_1(\theta_1^2\theta_2^3+r_2r\theta)$
	$m_1'm_1'M_2'M_2'(t+1)$	$r^2\theta$	$r_1^2\theta_1\theta_2^3$	$r_1\theta_1^2r_2\theta_2^2$	$r_1\theta_1^2r_2\theta_2^2$	$r_2(r_2\theta_1^3\theta_2+r_1r^2)$
	$m_1'm_1'M_2'm_2'(t+1)$	$2r\theta^2$	$2r_1^2r_2\theta_1\theta_2^2$	$r_1\theta_1^2\theta_2\varphi_2$	$r_1\theta_1^2\theta_2\varphi_2$	$2r_2(\theta_1^3\theta_2^2+r_1r\theta)$
	$m_1'm_1'm_2'm_2'(t+1)$	θ^3	$r_1^2r_2^2\theta_1\theta_2$	$r_1r_2\theta_1^2\theta_2^2$	$r_1r_2\theta_1^2\theta_2^2$	$\theta_1^3\theta_2^3+r_1r_2\theta^2$
$M_1'm_1'm_2'm_2'(t)$	$M_1'M_1'm_2'm_2'(t+1)$	r	$r_2\theta_1^3$	$r_1r_2\theta_1^2$	$r_1r_2\theta_1^2$	$r_1(r_1r_2\theta_1+\theta_2)$
	$M_1'm_1'm_2'm_2'(t+1)$	$2r$	$2r_1r_2\theta_1^2$	$r_2\theta_1\varphi_1$	$r_2\theta_1\varphi_1$	$2r_1(r_2\theta_1^2+\theta_2)$
	$m_1'm_1'm_2'm_2'(t+1)$	r	$r_1^2r_2\theta_1$	$r_1r_2\theta_1^2$	$r_1r_2\theta_1^2$	$r_2\theta_1^3+r_1\theta_2$
$m_1'm_1'M_2'm_2'(t)$	$m_1'm_1'M_2'M_2'(t+1)$	r	$r_1\theta_2^3$	$r_1r_2\theta_2^2$	$r_1r_2\theta_2^2$	$r_2(r_1r_2\theta_2+\theta_1)$
	$m_1'm_1'M_2'm_2'(t+1)$	$2r$	$2r_1r_2\theta_2^2$	$r_1\theta_2\varphi_2$	$r_1\theta_2\varphi_2$	$2r_2(r_1\theta_2^2+\theta_1)$
	$m_1'm_1'm_2'm_2'(t+1)$	r	$r_1r_2^2\theta_2$	$r_1r_2\theta_2^2$	$r_1r_2\theta_2^2$	$r_1\theta_2^3+r_2\theta_1$
$m_1'm_1'm_2'm_2'(t)$	$m_1'm_1'm_2'm_2'(t+1)$	4θ	r_1r_2	r_1r_2	r_1r_2	$4\theta_1-4r_2+5r_1r_2$

* r_1, r_2 and r are the recombination fractions between marker $\mathbf{M}_1'$ and the endosperm QTL, between the endosperm QTL and marker $\mathbf{M}_2'$, and between the two flanking markers, respectively. $\varphi_1=1-2r_1+2r_1^2$, $\varphi_2=1-2r_2+2r_2^2$, $\varphi=1-2r+2r^2$, $\theta_1=1-r_1$, $\theta_2=1-r_2$ and $\theta=1-r$.

backcross plant i, conditional upon the two different marker intervals can be expressed as

$$\pi_{k_ij_1j_2}=\pi_{ij_1}\otimes\pi_{k_ij_2},$$

where $\otimes$ is the matrix Kronecker product operation. If $Q(t)$ and $Q(t+1)$ are linked and located within the same marker interval, the joint conditional probabilities of the two QTLs, conditional upon the joint maternal–embryonic marker genotypes of the flanking markers, are derived differently and are shown in Table 3.

(ii) *Quantitative genetic models*

The phenotypic value of an endosperm trait caused by the two putative QTLs $Q(t)$ and $Q(t+1)$ can be modelled using traditional quantitative genetic theory (Lynch & Walsh, 1998). Let $\mu_1-\mu_2$ denote the genotypic means of the eight joint maternal-endosperm QTL genotypes, respectively. Let $a(t)$ be the additive effect of the maternal QTL on the endosperm trait, $a(t+1)$ be the additive effect of the endosperm QTL, $d_1(t+1)$ be the first dominant effect of the endosperm QTL (i.e. the dominant effect of $QQ(t+1)$ to $q(t+1)$ when $Q(t+1)$ is dominant or that of $q(t+1)$ to $QQ(t+1)$ when $q(t+1)$ is dominant), $d_2(t+1)$ be second dominant effect of the endosperm QTL (i.e. the dominant effect of $Q(t+1)$ to $qq(t+1)$ when $Q(t+1)$ is dominant or $qq(t+1)$ to $Q(t+1)$ when $q(t+1)$ is dominant), ξ be the $a(t)\times a(t+1)$ interaction of these two QTLs, ζ_1 be the $a(t)\times d_1(t+1)$ interaction, and ζ_2 be the $a(t)\times d_2(t+1)$ interaction. The eight genotypic means can be expressed in terms of these genetic effects as

$$\begin{cases}
\mu_1=\mu+\frac{1}{2}a(t)+\frac{3}{2}a(t+1)+\frac{3}{4}\xi, & \text{for } Qq(t)QQQ(t+1)\\
\mu_2=\mu+\frac{1}{2}a(t)+\frac{1}{2}a(t+1)+d_1(t+1)+\frac{1}{4}\xi+\frac{1}{2}\zeta_1, & \text{for } Qq(t)QQq(t+1)\\
\mu_3=\mu+\frac{1}{2}a(t)-\frac{1}{2}a(t+1)+d_2(t+1)-\frac{1}{4}\xi+\frac{1}{2}\zeta_2, & \text{for } Qq(t)Qqq(t+1)\\
\mu_4=\mu+\frac{1}{2}a(t)-\frac{3}{2}a(t+1)-\frac{3}{2}\xi, & \text{for } Qq(t)qqq(t+1)\\
\mu_5=\mu-\frac{1}{2}a(t)+\frac{3}{2}a(t+1)-\frac{3}{4}\xi, & \text{for } qq(t)QQQ(t+1)\\
\mu_6=\mu-\frac{1}{2}a(t)+\frac{1}{2}a(t+1)+d_1(t+1)-\frac{1}{4}\xi-\frac{1}{2}\zeta_1, & \text{for } qq(t)QQq(t+1)\\
\mu_7=\mu-\frac{1}{2}a(t)-\frac{1}{2}a(t+1)+d_2(t+1)+\frac{1}{4}\xi-\frac{1}{2}\zeta_2, & \text{for } qq(t)Qqq(t+1)\\
\mu_8=\mu-\frac{1}{2}a(t)-\frac{3}{2}a(t+1)+\frac{3}{4}\xi, & \text{for } qq(t)qqq(t+1)
\end{cases} \quad (1)$$

https://doi.org/10.1017/S0016672305007615 Published online by Cambridge University Press

Table 3. *The frequencies of joint maternal–embryo marker genotypes of the same markers interval in a backcross design and joint, frequencies of maternal–endosperm QTL genotypes in a backcross design, used for the calculation of conditional probabilities of joint maternal–endosperm QTL genotypes given joint maternal–embryo marker genotypes of the same markers interval*

Joint maternal–embryo marker genotype			Joint maternal–endosperm QTL genotype frequency*			
Maternal	Embryo	Frequency	$Qq(t)QQQ(t+1)$	$Qq(t)QQq(t+1)$	$Qq(t)Qqq(t+1)$	$Qq(t)qqq(t+1)$
$M_1'M_1'M_2'm_2'(t)$	$M_1'M_1'M_2'M_2'(t+1)$	θ^3	$2\theta_{12}^2\theta_3^2P_1$	$2r_{12}\theta_{12}r_3\theta_3P_1$	$2r_{12}\theta_{12}r_3\theta_3P_1$	$2r_{12}^2r_3^2P_1+2\theta^2P_2$
	$M_1'M_1'M_2'm_2'(t+1)$	$r\theta^2$	$2\theta_{12}^2r_3\theta_3P_1$	$r_{12}\theta_{12}\varphi_3P_1$	$r_{12}\theta_{12}\varphi_3P_1$	$2r_{12}^2r_3\theta_3P_1+2r\theta P_2$
	$M_1'M_1'm_2'm_2'(t+1)$	$r\theta^2$	$2\theta_{12}^2r_3^2P_1$	$2r_{12}\theta_{12}r_3\theta_3P_1$	$2r_{12}\theta_{12}r_3\theta_3P_1$	$2r_{12}^2\theta_3^2P_1+2r^2P_2$
	$M_1'm_1'M_2'M_2'(t+1)$	$r\theta^2$	$2r_{12}\theta_{12}\theta_3^2P_1$	$r_3\theta_3\varphi_{12}P_1$	$r_3\theta_3\varphi_{12}P_1$	$2r_{12}\theta_{12}r_3^2P_1+2r\theta P_2$
	$M_1'm_1'M_2'm_2'(t+1)$	$\theta\varphi$	$4r_{12}\theta_{12}r_3\theta_3P_1$	$\varphi_{12}\varphi_3P_1$	$\varphi_{12}\varphi_3P_1$	$4r_{12}\theta_{12}r_3\theta_3P_1+2(r^2+\theta^2)P_2$
	$M_1'm_1'm_2'm_2'(t+1)$	$r\theta^2$	$2r_{12}\theta_{12}r_3^2P_1$	$r_3\theta_3^2\varphi_{12}P_1$	$r_3\theta_3^2\varphi_{12}P_1$	$2r_{12}\theta_{12}\theta_3^2P_1+2r\theta P_2$
	$m_1'm_1'M_2'M_2'(t+1)$	$r^2\theta$	$2r_{12}^2\theta_3^2P_1$	$2r_{12}\theta_{12}r_3\theta_3P_1$	$2r_{12}\theta_{12}r_3\theta_3P_1$	$2\theta_{12}^2r_3^2P_1+2r^2P_2$
	$m_1'm_1'M_2'm_2'(t+1)$	$r\theta^2$	$2r_{12}^2r_3\theta_3^2P_1$	$r_{12}\theta_{12}\varphi_3P_1$	$r_{12}\theta_{12}\varphi_3P_1$	$2\theta_{12}^2r_3\theta_3P_1+2r\theta P_2$
	$m_1'm_1'm_2'm_2'(t+1)$	θ^3	$2r_{12}^2r_3^2P_1$	$2r_{12}\theta_{12}r_3\theta_3P_1$	$2r_{12}\theta_{12}r_3\theta_3P_1$	$2\theta_{12}^2r_3^2P_1+2\theta^2P_2$
$M_1'm_1'm_2'm_2'(t)$	$M_1'M_1'm_1'm_2'(t+1)$	r	$2\theta_{12}^2P_5$	$2r_{12}\theta_{12}P_5$	$2r_{12}\theta_{12}P_5$	$2r_{12}^2P_5+2P_6$
	$M_1'm_1'm_2'm_2'(t+1)$	r	$2r_{12}\theta_{12}P_5$	$\varphi_{12}P_5$	$\varphi_{12}P_5$	$2r_{12}\theta_{12}P_5+2P_6$
	$m_1'm_1'm_2'm_2'(t+1)$	r	$2r_{12}^2P_5$	$2r_{12}\theta_{12}P_5$	$2r_{12}\theta_{12}P_5$	$2\theta_{12}^2P_5+2P_6$
$m_1'm_1'M_2'm_2'(t)$	$m_1'm_1'M_2'M_2'(t+1)$	r	$2\theta_3^2P_8$	$2r_3\theta_3P_8$	$2r_3\theta_3P_8$	$2r_3^2P_8+2P_7$
	$m_1'm_1'M_2'm_2'(t+1)$	r	$2r_3\theta_3P_8$	φ_3P_8	φ_3P_8	$2r_3\theta_3P_8+2P_7$
	$m_1'm_1'm_2'm_2'(t+1)$	r	$2r_3^2P_8$	$2r_3\theta_3P_8$	$2r_3\theta_3P_8$	$2\theta_3^2P_8+2P_7$
$m_1'm_1'm_2'm_2'(t)$	$m_1'm_1'm_2'm_2'(t+1)$	2θ	P_4	P_4	P_4	P_4+2P_3
Maternal	Embryo	Frequency	$qq(t)QQQ(t+1)$	$qq(t)QQq(t+1)$	$qq(t)Qqq(t+1)$	$qq(t)qqq(t+1)$
$M_1'm_1'M_2'm_2'(t)$	$M_1'M_1'M_2'M_2'(t+1)$	θ^3	$2\theta_{12}^2\theta_3^2P_3$	$2r_{12}\theta_{12}r_3\theta_3P_3$	$2r_{12}\theta_{12}r_3\theta_3P_3$	$2r_{12}^2r_3^2P_3+2\theta^2P_4$
	$M_1'M_1'M_2'm_2'(t+1)$	$r\theta^2$	$2\theta_{12}^2r_3\theta_3P_3$	$r_{12}\theta_{12}\varphi_3P_3$	$r_{12}\theta_{12}\varphi_3P_3$	$2r_{12}^2r_3\theta_3P_3+2r\theta P_4$
	$M_1'M_1'm_2'm_2'(t+1)$	$r^2\theta$	$2\theta_{12}^2r_3^2P_3$	$2r_{12}\theta_{12}r_3\theta_3P_3$	$2r_{12}\theta_{12}r_3\theta_3P_3$	$2r_{12}^2\theta_3^2P_3+2r^2P_4$
	$M_1'm_1'M_2'M_2'(t+1)$	$r\theta^2$	$2r_{12}\theta_{12}\theta_3^2P_3$	$r_3\theta_3\varphi_{12}P_3$	$r_3\theta_3\varphi_{12}P_3$	$2r_{12}\theta_{12}r_3^2P_3+2r\theta P_4$
	$M_1'm_1'M_2'm_2'(t+1)$	$\theta\varphi$	$4r_{12}\theta_{12}r_3\theta_3P_3$	$\varphi_{12}\varphi_3P_3$	$\varphi_{12}\varphi_3P_3$	$4r_{12}\theta_{12}r_3\theta_3P_3+2(r^2+\theta^2)P_4$
	$M_1'm_1'm_2'm_2'(t+1)$	$r\theta^2$	$2r_{12}\theta_{12}r_3^2P_3$	$r_3\theta_3^2\varphi_{12}P_3$	$r_3\theta_3^2\varphi_{12}P_3$	$2r_{12}\theta_{12}\theta_3^2P_3+2r\theta P_4$
	$m_1'm_1'M_2'M_2'(t+1)$	$r^2\theta$	$2r_{12}^2\theta_3^2P_3$	$2r_{12}\theta_{12}r_3\theta_3P_3$	$2r_{12}\theta_{12}r_3\theta_3P_3$	$2\theta_{12}^2r_3^2P_3+2r^2P_4$
	$m_1'm_1'M_2'm_2'(t+1)$	$r\theta^2$	$2r_{12}^2r_3\theta_3^2P_3$	$r_{12}\theta_{12}\varphi_3P_3$	$r_{12}\theta_{12}\varphi_3P_3$	$2\theta_{12}^2r_3\theta_3P_3+2r\theta P_4$
	$m_1'm_1'm_2'm_2'(t+1)$	θ^3	$2r_{12}^2r_3^2P_3$	$2r_{12}\theta_{12}r_3\theta_3P_3$	$2r_{12}\theta_{12}r_3\theta_3P_3$	$2\theta_{12}^2r_3^2P_3+2\theta^2P_4$
$M_1'm_1'm_2'm_2'(t)$	$M_1'M_1'm_2'm_2'(t+1)$	r	$2\theta_{12}^2P_7$	$2r_{12}\theta_{12}P_7$	$2r_{12}\theta_{12}P_7$	$2r_{12}^2P_7+2P_8$
	$M_1'm_1'm_2'm_2'(t+1)$	r	$2r_{12}\theta_{12}P_7$	$\varphi_{12}P_7$	$\varphi_{12}P_7$	$2r_{12}\theta_{12}P_7+2P_8$
	$m_1'm_1'm_2'm_2'(t+1)$	r	$2r_{12}^2P_7$	$2r_{12}\theta_{12}P_7$	$2r_{12}\theta_{12}P_7$	$2\theta_{12}^2P_7+2P_8$
$m_1'm_1'M_2'm_2'(t)$	$\mathrm{m_1'm_1'}M_2'M_2'(t+1)$	r	$2\theta_3^2P_6$	$2r_3\theta_3P_6$	$2r_3\theta_3P_6$	$2r_3^2P_6+2P_5$
	$\mathrm{m'_1m_1'}M_2'm_2'(t+1)$	r	$2r_3\theta_3P_6$	φ_3P_6	φ_3P_6	$2r_3\theta_3P_6+2P_6$
	$m_1'm_1'm_2'm_2'(t+1)$	r	$2r_3^2P_6$	$2r_3\theta_3P_6$	$2r_3\theta_3P_6$	$2\theta_3^2P_6+2P_5$
$m_1'm_1'm_2'm_2'(t)$	$m_1'm_1'm_2'm_2'(t+1)$	2θ	P_2	P_2	P_2	P_2+2P_1

* r_1, r_2, r_{12}, r_3 and r are the recombination fractions between marker $\mathbf{M}_1'$ and the maternal QTL, between the maternal and offspring QTL, between the marker $\mathbf{M}_1'$ and the offspring QTL and between the offspring QTL and $\mathbf{M}_2'$ and between the two markers, respectively. $P_1=\frac{1}{2}(1-r_1)(1-r_2)(1-r_3)$, $P_2=\frac{1}{2}(1-r_1)r_2r_3$, $P_3=\frac{1}{2}r_1r_2(1-r_3)$, $P_4=\frac{1}{2}r_1(1-r_2)r_3$, $P_5=\frac{1}{2}(1-r_1)(1-r_2)r_3$, $P_6=\frac{1}{2}(1-r_1)r_2(1-r_3)$, $P_7=\frac{1}{2}r_1r_2r_3$, $P_8=\frac{1}{2}r_1(1-r_2)(1-r_3)$. $\varphi_3=1-2r_3+2r_3^2$, $\varphi_{12}=1-2r_{12}+2r_{12}^2$, $\theta_3=1-r_3$, $\theta_{12}=1-r_{12}$.

where μ is the overall population mean. This can be written in matrix form as

$$\mathbf{m} = \mathbf{De},$$

where $m = (\mu_1, \ldots, \mu_8)^T$ is the vector of the QTL genotypic means, $\mathbf{e} = [a(t), a(t+1), d_1(t+1), d_2(t+1), \xi, \zeta_1, \zeta_2]^T$ is the vector of the genetic effect and **D** is an (8×8) design matrix relating **m** and **e**. If the genotypic means are known, we can estimate the genetic effect using

$$\mathbf{e} = \mathbf{D}^{-1}\mathbf{m}. \tag{2}$$

3. Statistical method

(i) *Mixture model*

The statistical foundation of QTL mapping is based on the mixture model, in which each observation is assumed to have arisen from one of a known or unknown number of components, each component being modelled by a density from the parametric family *f*. With eight possible maternal-endosperm QTL genotypes, this mixture model for the observation (y_{k_i}) of endosperm k derived from backcross plant i is expressed as

$$\begin{aligned} y_{k_i} \approx f(y_{k_i}|\pi, \phi, \sigma^2) = \pi_{k_i 1} f_1(y_{k_i};\, \phi_1, \sigma^2) + \cdots \\ + \pi_{k_i 8} f_8(y_{k_i};\, \phi_8, \sigma^2) \end{aligned} \tag{3}$$

where $\pi = (\pi_{k_i 1}, \ldots, \pi_{k_i 8})$ are the mixture proportions, which are constrained to be non-negative and to sum to unity, $\phi = (\phi_1, \ldots, \phi_8)$ are the component-specific parameters, with ϕ_l being specific to component l ($l = 1, \ldots, 8$) (in our case, ϕ_j can be specified by the genotypic mean vector **m**, which contains the parameters **e** to be estimated) and σ^2 is the residual variance, which is assumed to be the same among all the components. The likelihood function of the marker data and the endosperm trait values controlled by the putative QTL, $Q(t)$ and $Q(t+1)$, based on the mixture model in Eqn 3 can be expressed as

$$L(\mathbf{\Omega}) = \prod_{i=1}^{n} \prod_{k=1}^{n_i} \left[\sum_{l=1}^{8} \pi_{k_i} f_l(y_{k_i}) \right] \tag{4}$$

where $\mathbf{\Omega} = (\mathbf{m}^T, \theta, \sigma^2)$ or $(\mathbf{e}^T, \theta, \sigma^2)$ is the vector for the unknown QTL effect parameters, QTL positions θ (measured by the recombination fraction between the QTL and its flanking markers) and residual variance σ^2, and $f_l(y_{k_i})$ is the normal density corresponding to the lth QTL genotype with mean μ_l and variance σ^2.

(ii) *The EM algorithm*

To obtain the maximum-likelihood estimates (MLEs) of **Ω**, we implement the EM algorithm. The log-likelihood function of Eqn 4 is given by

$$\log L(\mathbf{\Omega}) = \sum_{i=1}^{n} \sum_{k=1}^{n_i} \log \left[\sum_{l=1}^{8} \pi_{k_i} f_l(y_{k_i}) \right] \tag{5}$$

with the derivative for an unknown parameter $\mathbf{\Omega}_\varphi$,

$$\begin{aligned} \frac{\partial}{\partial \mathbf{\Omega}_\varphi} \log L(\mathbf{\Omega}) &= \sum_{i=1}^{n} \sum_{k=1}^{n_i} \sum_{l=1}^{8} \frac{\pi_{k_i l} \frac{\partial}{\partial \mathbf{\Omega}\varphi} f_l(y_{k_i})}{\sum_{l=1}^{8} \pi_{k_i} f_l(y_{k_i})} \\ &= \sum_{i=1}^{n} \sum_{k=1}^{n_i} \sum_{l=1}^{8} \frac{\pi_{k_i} f_l(y_{k_i})}{\sum_{l=1}^{8} \pi_{k_i} f_l(y_{k_i})} \\ &\quad \times \frac{\partial}{\partial \mathbf{\Omega}\varphi} \log f_l(y_{k_i}) \\ &= \sum_{i=1}^{n} \sum_{k=1}^{n_i} \sum_{l=1}^{8} \Pi_{k_i l} \frac{\partial}{\partial \mathbf{\Omega}\varphi} \log f_l(y_{k_i}) \end{aligned} \tag{6}$$

where we define

$$\Pi_{k_i l} = \frac{\pi_{k_i l} f_l(y_{k_i})}{\sum_{l=1}^{8} \pi_{k_i l} f_l(y_{k_i})} \tag{7}$$

which is thought of as a posterior probability of joint maternal-endosperm QTL genotype l for the k_ith endosperm derived from the ith backcross plants, given joint maternal–offspring marker genotypes. The conditional probabilities of the QTL genotypes given the marker genotypes described in Tables 1–3 are viewed as the prior probabilities. Given the initial values for the unknown parameters **Ω** and marker and phenotypic observations, we can update $\Pi_{k_i l}$ (E step). The estimated posterior probabilities are used to obtain the new MLEs of Ω (M step) based on the log-likelihood equations

$$\hat{\mu}_l = \frac{\sum_{i=1}^{n} \sum_{k=1}^{n_i} y_{k_i} \Pi_{k_i l}}{\sum_{i=1}^{n} \sum_{k=1}^{n_i} \Pi_{k_i l}} \tag{8}$$

$$\hat{\sigma}^2 = \frac{1}{m} \sum_{i=1}^{n} \sum_{k=1}^{n_i} \sum_{l=1}^{8} (y_{k_i} - \mu_l)^2 \Pi_{k_i l} \tag{9}$$

which are derived by letting the derivative (Eqn 6) equal zero. This iterative process is repeated until the specified convergence criterion is satisfied. The values at convergence are regarded as the MLEs.

In the procedure described above for the EM algorithm, we treated the positions of QTLs as known parameters even though their MLEs can also be obtained through iterative steps. We can use a grid approach to estimate the QTL positions. By hypothesizing a pair of maternal and endosperm QTLs every 4 cM at marker intervals, we can draw the landscape of log-likelihood test statistics throughout the entire genome. The positions corresponding to the peak of the landscape across a linkage group are the MLEs of the QTL positions.

The MLEs of the genotypic mean vector **m** can be used to obtain the estimates of the genetic effects

https://doi.org/10.1017/S0016672305007615 Published online by Cambridge University Press

contained in vector **e**. As long as the MLEs of **m** are uniquely estimated for specific QTL genotypes, the MLEs of **e** can be uniquely determined. However, because QTL genotypes $Qq(t)QQq(t+1)$ and $Qq(t)Qqq(t+1)$ have the same conditional probabilities (as do QTL genotypes $qq(t)QQq(t+1)$ and $qq(t)Qqq(t+1)$), the MLEs of the genotypic means μ_2 and μ_3, and the means μ_6 and μ_7 are not identifiable. For this reason, we cannot uniquely estimate dominant effects $d_1(t+1)$ and $d_2(t+1)$ or the epistatic effects ζ_1 and ζ_2. However, we can estimate the sums of each of these two pairs of effects, in which case the unknown vector is $\mathbf{m}^+ = [\mu, a(t), a(t+1), d_1(t+1)+d_2(t+1), \xi, \zeta_1+\zeta_2]$.

Based on Eqn 1, it is interesting that the estimates of genetic effect parameters $\mathbf{e}^* = [\mu, a(t), a(t+1), \xi]$ are only dependent on genotypic means $\mathbf{m}^* = (\mu_1, \mu_4, \mu_5, \mu_8)$. Let

$$\mathbf{d} = \begin{bmatrix} 1 & \frac{1}{2} & \frac{3}{2} & \frac{3}{4} \\ 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} \\ 1 & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{4} \\ 1 & \frac{1}{2} & -\frac{3}{2} & -\frac{3}{4} \end{bmatrix}. \tag{10}$$

The MLEs of $\mathbf{e}^*$ can be obtained by solving the equation $\mathbf{e}^* = \mathbf{d}^{-1}\,\mathbf{m}^*$. The MLEs for the summed dominant effect and the summed additive × dominant effect can be obtained by solving

$$\begin{cases} d_1(t+1)+d_2(t+1) = \frac{1}{2}(\mu_2+\mu_3+\mu_6+\mu_7-4\mu) \\ \zeta_1+\zeta_2 = \mu_2+\mu_3+(\mu_6+\mu_7)-2a(t) \end{cases} \tag{11}$$

(iii) *Hypothesis tests*

Several statistical hypothesis tests can be performed to analyse the genetic control of endosperm traits. The presence of QTLs affecting an endosperm trait can be tested for by formulating the hypotheses

$$\begin{cases} H_0\text{: } a(t)=a(t+1)=d_1(t)=d_2(t+1)=\xi=\zeta_1=\zeta_2=0 \\ H_1\text{: at least one of the equalities above does not hold} \end{cases} \tag{12}$$

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

$$LR = -2[\log L(\tilde{\boldsymbol{\Omega}}) - \log L(\hat{\boldsymbol{\Omega}})], \tag{13}$$

where $\tilde{\boldsymbol{\Omega}}$ and $\hat{\boldsymbol{\Omega}}$ denote the MLEs of the unknown parameters under H_0 and H_1, respectively. The LR is asymptotically χ^2 distributed with seven degrees of freedom. However, the critical threshold value for declaring the existence of the testing QTL is generally calculated on the basis of permutation tests (Churchill & Doerge, 1994).

To test the additive effects of the maternal QTL and the endosperm QTL on the endosperm traits, we formulate the hypotheses

$$\begin{cases} H_0\text{: } a(t)=0 \\ H_1\text{: } a(t)\neq 0 \end{cases}, \tag{14}$$

and

$$\begin{cases} H_0\text{: } a(t+1)=0 \\ H_1\text{: } a(t+1)\neq 0 \end{cases}, \tag{15}$$

where the likelihood-ratio test statistic for both the tests is asymptotically χ^2 distributed with one degree of freedom. To obtain the MLEs of the parameters under the null hypotheses above, we need to pose the constraints $\mu_1+\mu_4=\mu_5+\mu_8$ for the hypotheses in Eqn 14 and $\mu_1+\mu_5=\mu_4+\mu_8$ for the hypotheses in Eqn 15 in the M step.

The significance of maternal-endosperm QTL additive × additive interaction can be tested by comparing the following hypotheses

$$\begin{cases} H_0\text{: } \xi=0 \\ H_1\text{: } \xi\neq 0 \end{cases}. \tag{16}$$

The estimates of the parameter under the null hypothesis are based on the constraint $\mu_1+\mu_8=\mu_4+\mu_5$.

Similarly, the significance of the summed dominant effect and the summed additive × dominant effect can be tested, with the constraints derived from Eqn 11. The critical values for declaring the significance of various genetic effects can be determined on the basis of the simulated data in which no particular genetic effect is assumed to exist.

4. Results

We performed a series of simulation studies to examine the statistical properties of the model. Five equidistant markers are ordered as $\mathbf{M}_1$–$\mathbf{M}_5$ on a linkage group with length 80 cM. These five markers were simulated for a backcross population based on the recombination fractions between all pairs of two adjacent markers. The Haldane map function was used to convert the map distance into the recombination fraction. In the simulation experiments, different levels of heritability contributed jointly by maternal and endosperm QTL ($H^2 = 0{\cdot}1$ vs $0{\cdot}4$) and different total sample sizes ($m = 200$ vs 400) are examined. Different sampling strategies are designed on the basis of the allocations of a given sample size between the backcross plants and their progeny (seeds): (1) 10×20 or 10×40 (20 or 40 seeds are sampled from each of ten backcross plants); (2) 200×1 or 400×1 (one seed is sampled from each of 200 or 400 backcross plants). Such sampling strategies not only allow us to examine the possible effects of the parameters estimation but also provide useful guidance for practical molecular studies (Wu *et al.*, 2002*b*).

https://doi.org/10.1017/S0016672305007615 Published online by Cambridge University Press

Table 4. *The MLEs of the QTL position and effect parameters exerted by a maternal QTL and an endosperm QTL each on a different interval derived from 100 simulation replicates. The squared roots of the mean square errors of the MLEs are given in the second parenthesized row of the MLEs**

H^2	n	Positions (8, 48)	μ (10)	$a(t)$ (0·7)	$a(t+1)$ (0·5)	ξ (0·3)	$d_1(t+1)+d_2(t+1)$ (0·5)	$\zeta_1+\zeta_2$ (0·3)	σ^2
0·1	10 × 20	(11·10, 49·52)	10·11	0·52	0·56	0·11	−0·13	1·21	5·41
		(10·29, 5·24)	(0·63)	(1·34)	(0·45)	(0·89)	(2·12)	(4·08)	(0·82)
	10 × 40	(9·46, 49·44)	10·08	0·48	0·53	0·17	0·15	1·08	5·51
		(9·88, 4·92)	(0·45)	(1·00)	(0·30)	(0·60)	(1·59)	(3·42)	(0·60)
	200 × 1	(9·74, 49·56)	9·95	0·81	0·44	0·38	0·61	0·07	5·32
		(9·69, 5·19)	(0·47)	(0·93)	(0·31)	(0·67)	(1·49)	(3·15)	(0·88)
	400 × 1	(9·13, 49·36)	9·95	0·81	0·43	0·33	0·55	0·08	5·49
		(8·93, 4·69)	(0·29)	(0·62)	(0·20)	(0·44)	(1·02)	(1·99)	(0·61)
0·4	10 × 20	(8·41, 47·84)	10·00	0·68	0·49	0·28	0·45	0·35	0·91
		(8·48, 3·55)	(0·23)	(0·51)	(0·15)	(0·33)	(0·74)	(1·56)	(0·12)
	10 × 40	(8·28, 48·20)	9·9801	0·7047	0·47887	0·2925	0·4889	0·4326	0·9408
		(7·42, 2·87)	(0·22)	(0·46)	(0·15)	(0·31)	(0·72)	(1·56)	(0·10)
	200 × 1	(9·84, 47·92)	9·96	0·84	0·46	0·38	0·58	0·14	0·92
		(7·72, 3·31)	(0·17)	(0·34)	(0·12)	(0·22)	(0·64)	(1·23)	(0·13)
	400 × 1	(8·30, 48·08)	10·01	0·70	0·50	0·30	0·45	0·34	0·92
		(6·19, 2·67)	(0·13)	(0·23)	(0·08)	(0·16)	(0·46)	(0·82)	(0·13)

* The locations of the two QTL are described by the map distances (in cM) from the first marker of the linkage group (80 cM long). The hypothesized σ^2 value is 5·8725 for $H^2=0·1$ and 0·9788 for $H^2=0·4$.

Table 5. *The MLEs of the QTL position and effect parameters between a maternal QTL and an endosperm QTL both on the same interval derived from 100 simulation replicates. The squared roots of the mean square errors of the MLEs are given in the second parenthesized row of the MLEs**

H^2	n	Positions (8, 16)	μ (10)	$a(t)$ (0·7)	$a(t+1)$ (0·5)	ξ (0·3)	$d_1(t+1)+d_2(t+1)$ (0·5)	$\zeta_1+\zeta_2$ (0·3)	σ^2
0·1	10 × 20	(5·73, 17·21)	10·11	0·24	0·52	0·00	−0·12	1·58	5·29
		(6·90, 5·73)	(0·80)	(1·78)	(0·56)	(1·22)	(2·18)	(4·86)	(0·91)
	10 × 40	(4·96, 17·24)	10·06	0·58	0·55	0·15	0·52	0·09	5·65
		(6·52, 4·05)	(0·75)	(1·58)	(0·51)	(1·09)	(2·15)	(4·39)	(0·59)
	200 × 1	(3·81, 17·99)	10·11	0·46	0·56	0·05	0·51	0·29	5·38
		(7·21, 4·83)	(0·61)	(1·34)	(0·46)	(0·88)	(2·17)	(4·27)	(0·86)
	400 × 1	(3·69, 16·99)	9·98	0·68	0·48	0·26	0·68	0·15	5·67
		(6·73, 3·89)	(0·59)	(1·19)	(0·39)	(0·80)	(1·68)	(3·26)	(0·48)
0·4	10 × 20	(4·64, 16·53)	10·10	0·52	0·56	0·14	0·53	0·27	0·95
		(6·16, 2·91)	(0·37)	(0·70)	(0·23)	(0·48)	(1·14)	(2·15)	(0·13)
	10 × 40	(4·36, 16·57)	10·04	0·58	0·53	0·23	0·54	0·39	0·94
		(5·79, 2·59)	(0·31)	(0·62)	(0·21)	(0·41)	(0·91)	(1·66)	(0·08)
	200 × 1	(2·45, 16·84)	10·08	0·48	0·55	0·17	0·63	0·06	0·92
		(6·87, 2·96)	(0·29)	(0·68)	(0·20)	(0·43)	(0·97)	(2·01)	(0·12)
	400 × 1	(4·04, 16·61)	10·04	0·64	0·52	0·26	0·49	0·24	0·96
		(6·06, 2·43)	(0·24)	(0·48)	(0·16)	(0·33)	(0·71)	(1·46)	(0·08)

* The locations of the two QTL are described by the map distances (in cM) from the first marker of the linkage group (80 cM long). The hypothesized σ^2 value is 5·8725 for $H^2=0·1$ and 0·9788 for $H^2=0·4$.

Suppose there are two different QTLs that affect a quantitative endosperm trait of interest, one from the maternal genome and the other from the endosperm genome. The two QTLs are hypothesized either on the same marker interval (L_1) or on a different marker interval (L_2). For L_1, the maternal and endosperm QTLs are located 8 cM and 16 cM from the marker $\mathbf{M}_1$, respectively. For L_2, the maternal QTL is located 8 cM from marker $\mathbf{M}_1$, whereas the endosperm QTL is located 8 cM from marker $\mathbf{M}_3$. Two sets of hypothesized parameter values are hypothesized, including large additive effects vs small interaction effects (Tables 4, 5) and small additive effects vs large interaction effects (Table 6). The endosperm trait

https://doi.org/10.1017/S0016672305007615 Published online by Cambridge University Press

Table 6. *The MLEs of the QTL position and effect parameters between a maternal QTL and an endosperm QTL on different and the same interval derived from 100 simulation replicates. The squared roots of the mean square errors of the MLEs are given in the second parenthesized row of the MLEs**

H^2	n		μ (10)	$a(t)$ (0·3)	$a(t+1)$ (0·3)	ξ (0·5)	$d_1(t+1)+d_2(t+1)$ (0·4)	$\zeta_1+\zeta_2$ (0·5)	σ^2
		Positions (8, 48)							
0·4 (D)‡	10 × 20	(8·12, 47·48)	9·98	0·25	0·29	0·48	0·53	0·28	0·44
		(7·68, 3·53)	(0·18)	(0·40)	(0·1297)	(0·26)	(0·60)	(1·26)	(0·06)
	10 × 40	(7·92, 48·20)	10·01	0·27	0·31	0·47	0·39	0·51	0·45
		(6·33, 2·81)	(0·13)	(0·28)	(0·09)	(0·19)	(0·40)	(0·79)	(0·04)
	200 × 1	(7·40, 48·08)	10·01	0·26	0·30	0·45	0·36	0·51	0·42
		(6·52, 3·55)	(0·12)	(0·25)	(0·08)	(0·17)	(0·39)	(0·92)	(0·06)
	400 × 1	(7·60, 48·32)	10·01	0·29	0·30	0·49	0·39	0·49	0·44
		(4·48, 2·60)	(0·09)	(0·18)	(0·06)	(0·12)	(0·29)	(0·66)	(0·04)
		Positions (8, 16)							
0·4 (S)‡	10 × 20	(4·56, 15·30)	10·12	0·06	0·38	0·34	0·38	0·58	0·42
		(6·19, 3·22)	(0·24)	(0·48)	(0·16)	(0·35)	(0·60)	(1·15)	(0·07)
	10 × 40	(6·48, 15·95)	10·10	0·13	0·36	0·37	0·37	0·50	0·44
		(4·32, 2·35)	(0·22)	(0·44)	(0·15)	(0·30)	(0·58)	(1·14)	(0·04)
	200 × 1	(4·64, 15·62)	10·08	0·10	0·36	0·37	0·45	0·45	0·43
		(5·92, 2·95)	(0·21)	(0·44)	(0·14)	(0·31)	(0·66)	(1·29)	(0·06)
	400 × 1	(5·56, 15·79)	10·08	0·12	0·35	0·38	0·36	0·63	0·44
		(4·96, 2·07)	(0·19)	(0·39)	(0·13)	(0·26)	(0·55)	(1·13)	(0·05)

* The locations of the two QTLs are described by the map distances (in cM) from the first marker of the linkage group (80 cM long). The hypothesized σ^2 value is 0·4547 for $H^2=0·4$.
‡ D refers to two QTLs in different location and S refers to two QTLs in the same location.

values for each seed were simulated as the summation of the joint maternal-endosperm QTL genotypic means and random errors that follow a normal distribution with mean zero and variance σ^2 calculated on the basis of given genetic effect values and different heritability levels.

In general, our model can provide reasonable estimates of the QTL positions and effects of various kind, with estimate precision depending on heritability, sample size, sampling strategy, gene action mode and QTL location. Our model has excellent power to detect epistatically interacting maternal and endosperm QTLs. In all cases of different sample sizes and heritabilities, the maximum values of the *LR* landscapes from 100 simulation replicates are beyond the critical thresholds at the $\alpha=0·001$ level determined from 1000 permutation tests for the simulated data. Fig. 1 and 2 are examples of the shapes of the *LR* landscapes for two contrasting sample sizes, heritabilities and sampling strategies when the QTLs are located at different marker intervals or at the same interval, respectively. The thicker line indicates the given QTL positions, whereas the thin line indicates the estimated QTL positions. The small differences between these two lines suggest that our model can accurately estimate the genomic positions of the QTLs.

The precision of parameter estimation is evaluated in terms of the square root of the mean squared errors of the MLEs. The QTL positions and effects can be better estimated when the endosperm trait has higher rather than lower heritability or when the sample size is larger rather than smaller (Tables 4–6). However, the increase of H^2 from 0·1 to 0·4 leads to more significant improvement of the estimation precision than the increase of m from 200 to 400. The square roots of the MSEs of the genetic parameters reduce by more than half when H^2 is increased from 0·1 to 0·4, whereas such reduction is much smaller when m is increased from 200 to 400. This suggests that, in practice, it is more important to manage experiments to reduce the residual errors (increase H^2) than to increase the sample size.

In addition to these two predictable effects by H^2 and m, we also found the following. First, different sampling strategies have effects on parameter estimation. For a given sample size, the sampling strategy of taking more backcross plants and fewer seeds for each backcross tends to provide more precise estimates of all parameters than the sampling strategy of taking fewer backcross plants but more seeds for each backcross. This is an important finding for designing an optimal molecular experiment. Second, the position and additive effect can be better estimated for

https://doi.org/10.1017/S0016672305007615 Published online by Cambridge University Press

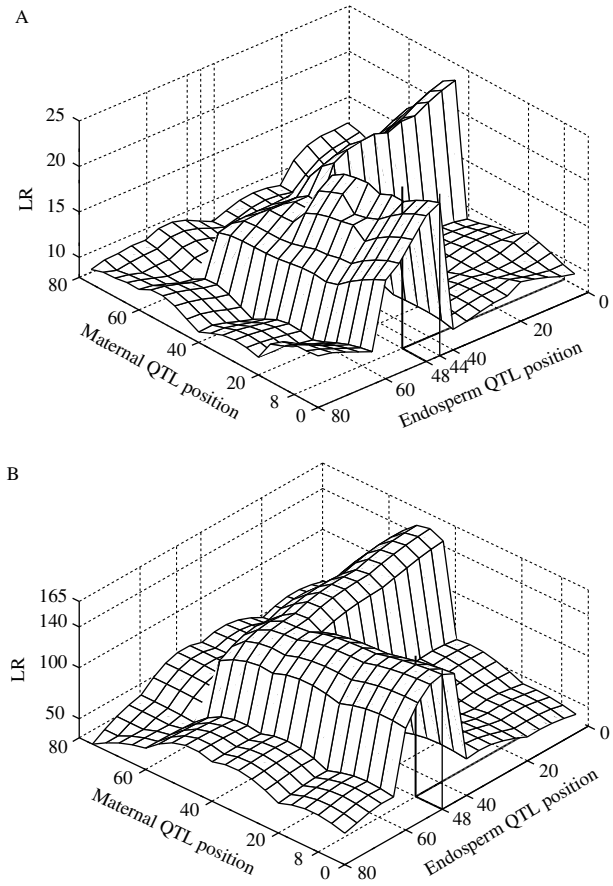

Fig. 1. The landscapes of the log-likelihood ratio (*LR*) test statistics calculated for the hypothesis test of the existence of QTLs against the maternal andendosperm QTL locations on the assumed linkage group when the two QTLs are located at different marker intervals. (A) The landscape for a heritability of 0·1 and a sampling strategy of ten backcross plants × 20 seeds for each backcross. (B) The landscape for a heritability of 0·4 and a sampling strategy of 400 backcross plants × one seed for each backcross.

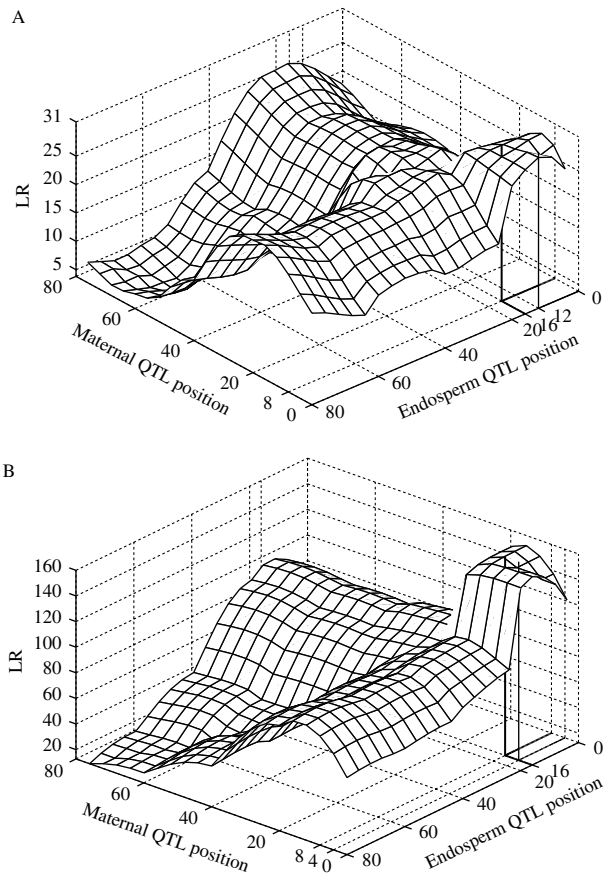

Fig. 2. The landscapes of the log-likelihood ratio (*LR*) test statistics calculated for the hypotesis test of the existence of QTLs against the maternal and endosperm QTL locations on the assumed linkage group when the two QTLs are located at the same marker interval. (A) The landscape for a heritability of 0·1 and a sampling strategy of ten backcross plants × 20 seeds for each backcross. (B) The landscape for a heritability of 0·4 and a sampling strategy of 400 backcross plants × one seed for each backcross.

the endosperm QTLs than the maternal QTLs, even if the maternal QTL has a larger effect than the endosperm QTL (Tables 4, 5). Also, the additive × additive effect between the maternal and endosperm QTLs can be well estimated, with better precision than that for the additive effect of the maternal QTLs. As expected, the estimation precision of the additive and additive × additive effects is better than that of the dominant effect, with the latter better than the estimation precision of the additive × dominant epistatic effects.

Third, the precision of parameter estimation is better when the maternal and endosperm QTLs are located at different marker intervals (Table 4) than when they are located at the same interval (Table 5). Thus, to avoid the analysis of two different QTLs located at the same interval, a high-density map is needed. Fourth, our model can well estimate the parameters for different gene action modes (large additive vs small interaction effect (Tables 4, 5) and small additive vs large interaction effect (Table 6)). However, in the latter case, it is possible to estimate precisely the interaction effects. Also, in this case, there is no marked difference in estimate precision between the two cases in which the two QTLs are located at different intervals or at the same interval.

5. Discussion

Endosperm is the result of double fertilization in flowering plants and plays a vital role in the early stage of embryo development (Chaudhury & Berger, 2001; Chaudhury *et al*., 2001). The expression of most quantitative traits in the endosperm results from direct (offspring) and indirect (maternal) genetic effects (Roach & Wulff, 1987), and involves complex interactions between QTLs from the maternal and offspring genomes (Chaudhury *et al*., 2001). Such

https://doi.org/10.1017/S0016672305007615 Published online by Cambridge University Press

maternal–offspring interactions have been well documented in animals (Falconer, 1965; Mousseau & Dingle, 1991; Reznick, 1991; Li *et al*., 1999) and are likely to have an effect on animal development and evolution (Mousseau & Fox, 1998; Wade, 1998; Agrawal *et al*., 2001; Cheverud, 2003; Hager & Johnstone, 2003; Wolf, 2000, 2003). Cheverud and colleagues have attempted a mapping approach for detecting specific QTLs that are associated with maternal care (Peripato & Cheverud, 2002; Peripato *et al*., 2002; Wolf *et al*., 2002). In this article, we develop a new statistical model for studying the genetic architecture of endosperm development contributed by the interactions between the maternal and endosperm QTLs.

Our model was founded on the developmental mechanisms of higher plants, which are characterized by a complex life cycle that consists of alternating haploid and diploid generations. These mechanisms culminate in five distinct phases (Chaudhury & Berger, 2001): (1) the diploid sporophytic mother; (2) the haploid female gametophyte; (3) the haploid male gametophyte; (4) the developing diploid embryo; and (5) the developing triploid endosperm. The development of the embryo sac and the seed are under the direct maternal control of both the sporophytic and the female gametophytic origin. The paternal gametophytic and postfertilization sporophytic controls are other levels in the complex genetic interactions that govern seed development. Cui *et al*. (2004) proposed an analytical model for characterizing QTL interactions from the sporophytic maternal and embryo genomes. The model proposed here is designed to detect interactions between QTLs derived from the maternal and endosperm genomes.

Our model has implemented two fundamental biological phenomena – maternal effects and epistasis – into QTL mapping models. Recent data suggest that these two phenomena might have been of greater importance in shaping the evolutionary process of organisms than was originally appreciated (Mousseau & Fox, 1998; Cheverud, 2003; Wolf *et al*., 2002). We expect that this model will have great implications for the study of evolutionary genetic problems related to seed development in higher plants. From a statistical perspective, by contrast, this model should be able to provide biologically more realistic results than many existing models because it integrates information about gene segregation and transmission from the maternal to offspring generations at both the marker and the QTLs.

We have conducted extensive computer simulations to investigate the statistical properties of this model. It is robust, in that it can provide reasonable estimation of QTL position and effect parameters at modest sample sizes and heritability levels. The result about the effect of different sampling strategies suggests that, for a given sample size, the inclusion of more maternal plants is more important for increased parameter-estimate precision than the inclusion of more seeds for each maternal plant. The simulation studies have also provided information about the impact on the precision of parameter estimates of different gene action modes, different origins of the QTLs and different QTL locations. For example, although both the maternal and endosperm QTLs are important for seed development, estimates of genetic effects of the endosperm QTLs is much more precise than that of the maternal QTLs.

The control of female and male gametophytes on seed development has been identified in many studies. Recent studies in particular show that the paternal and maternal genomes play unequal roles during early embryo and endosperm development (Vielle-Calzada *et al*., 2000; Weijers *et al*., 2001). For example, none of the paternally inherited alleles of 20 loci identified in *Arabidopsis* by Vielle-Galzada *et al*. (2000); is expressed during early seed development. These genes whose expression depends on the origin of parents are called imprinting genes (Li *et al*., 1999). Our model presented here provides an important step toward incorporating the control of imprinting genes within a QTL mapping framework.

We thank W. G. Hill and two anonymous reviewers for their constructive comments on this manuscript. 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-10583 by the Florida Agricultural Experiment Station.

References

Agrawal, A. F., Brodie, E. D. III & Brown, J. (2001). Parent–offspring coadaptation and the dual genetic control of maternal care. *Science* **292**, 1710–1712.

Chaudhury, A. M. & Berger, F. (2001). Maternal control of seed development. *Seminars in Cell and Developmental Biology* **12**, 381–386.

Chaudhury, A. M., Koltunow, A., Payne, T., Luo, M., Tucker, M. R., Dennis, E. S. & Peacock, W. J. (2001). Control of early seed development. *Annual Reviews in Cell and Developmental Biology* **17**, 677–699.

Cheverud, J. M. (2000). Detecting epistasis among quantitative trait loci. In *Epistasis and the Evolutionary Process* (ed. Wolf, J. B., Brodie, E. D. III & Wade, M. J.), pp. 58–81. New York: Oxford University Press.

Cheverud, J. M. (2003). Evolution in a genetically heritable social environment. *Proceedings of the National Academy of Sciences of the USA* **100**, 4357–4359.

Churchill, G. A. & Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963–971.

Cui, Y. H., Casella, G. & Wu, R. L. (2004). Mapping quantitative trait locus interactions from the maternal and offspring genomes. *Genetics* **167**, 1017–1026.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via EM algorithm. *Journal of the Royal Statistical Society Series B* **39**, 1–38.

Doebley, J., Stec, A. & Gustus, C. (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. In *Genetics Today: Proceedings of the 11th International Congress of Genetics* (ed. Geerts, S. J.), pp. 763–774. Oxford, UK: Pergamon Press.

Hager, R. & Johnstone, R. A. (2003). The genetic basis of family conflict resolution in mice. *Nature* **421**, 533–535.

Lander, E. S. & D. Botstein, (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199.

Lark, K. G., Chase, K., Adler, F., Mansur, L. M. & Orf, J. H. (1995). Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. *Proceedings of the National Academy of Sciences of the USA* **92**, 4656–4660.

Li, L.-L., Keverne, E. B., Aparicio, S. A., Ishino, F., Barton, S. C. & Surani, M. A. (1999). Regulation of maternal behavior and offspring growth by paternally expressed *Peg3*. *Science* **284**, 330–333.

Lynch, M. & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer.

Mousseau, T. A. & Dingle, H. (1991). Maternal effects in insect life histories. *Annual Reviews in Entomology* **36**, 511–534.

Mousseau, T. A. & Fox, C. (1998). *Maternal Effects as Adaptations*. New York: Oxford University Press.

Peripato, A. C. & Cheverud, J. M. (2002). Genetic influences on maternal care. *American Naturalist* **160**, S173–S185.

Peripato, A. C., de Brito, R. A., Vaughn, T. T., Pletscher, L. S., Matioli, S. R. & Cheverud, J. M. (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. In *The Unity of Evolutionary Biology: Proceedings of the Fourth International Congress of Systematic and Evolutionary Biology*, Vol. II (ed. Dudley, E. C.), pp. 780–793. Portland, OR: Dioscordes Press.

Roach, D. A. & R. D. Wulff (1987). Maternal effects in plants. *Annual Reviews of Ecological Systems* **18**, 209–235.

Vielle-Calzada, J.-P., Baskar, R. & Grossniklaus, U. (2000). Delayed activation of the paternal genome during seed development. *Nature* **404**, 91–94.

Wade, M. J. (1998). The evolutionary genetics of maternal effects. In *Maternal Effects as Adaptations* (ed. Mousseau, T. and Fox, C.), pp. 5–21. Oxford, UK: Oxford University Press.

Weijers, D., Geldner, N., Offringa, R. & Jorgens, G. (2001). Early paternal gene activity in *Arabidopsis*. *Nature* **414**, 709–710.

Whitlock, M. C., Phillips, P. C., Moore, F. B. G. & Tonsor, S. J. (1995). Multiple fitness peaks and epistasis. *Annual Reviews of Ecological Systems* **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. *Proceedings of the National Academy of Sciences of the USA* **100**, 4655–4660.

Wolf, J. B., Vaughn, T. T., Pletscher, L. S. & Cheverud, J. M. (2002). Contribution of maternal effect QTL to genetic architecture of early growth in mice. *Heredity* **89**, 300–310.

Wu, R. L., Ma, C.-X., Gallo-Meagher, M., Littell, R. C. & Casella, G. (2002*a*). Statistical methods for dissecting triploid endosperm traits using molecular markers: an autogamous model. *Genetics* **162**, 875–892.

Wu, R. L., Lou, X.-Y., Ma, C.-X., Wang, X. L., Larkins, B. A. & Casella, G. (2002*b*). An improved genetic model generates high-resolution mapping of QTL for protein quality in maize endosperm. *Proceedings of the National Academy of Sciences of the USA* **99**, 11281–11286.