Genet. Res., Camb. (2006), **87**, *pp.* 61–71. © 2006 Cambridge University Press
doi:10.1017/S0016672306007956 Printed in the United Kingdom

Modelling epistatic effects of embryo and endosperm QTL on seed quality traits

YUEHUA CUI[1,3], JIANGUO WU[2], CHUNHAI SHI[2]*, RAMON C. LITTELL[1] AND RONGLING WU[1]*

[1] *Department of Statistics, University of Florida, Gainesville, FL 32611, USA*
[2] *Department of Agronomy, Zhejiang University, Hangzhou, Zhejiang 310029, People's Republic of China*
[3] *Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA*

(*Received 14 June 2005 and in revised form 12 December 2005*)

Summary

Coordinated expression of embryo and endosperm tissues is required for proper seed development. The coordination among these two tissues is controlled by the interaction between multiple genes expressed in the embryo and endosperm genomes. In this article, we present a statistical model for testing whether quantitative trait loci (QTL) active in different genomes, diploid embryo and triploid endosperm, epistatically affect a trait expressed on the endosperm tissue. The maximum likelihood approach, implemented with the EM algorithm, was derived to provide the maximum likelihood estimates of the locations of embryo- and endosperm-specific QTL and their main effects and epistatic effects. This model was used in a real example for rice in which two QTL, one from the embryo genome and the other from the endosperm genome, exert a significant interaction effect on gel consistency on the endosperm. Our model has successfully detected *Waxy*, a candidate gene in the embryo genome known to regulate one of the major steps of amylose biosynthesis in the endosperm. This model will have great implications for agricultural and evolutionary genetic research.

1. Introduction

The developing seed contains endosperm and embryo tissues nourished by maternal tissues. The genetic and physiological balance between the embryo and endosperm tissues is crucial for proper seed development (Chaudhury *et al*., 2001; Walbot & Evans, 2003; Shi *et al*., 1999, 2000). A classic view about the function of the endosperm emphasizes its role in nourishing the embryo (Brink & Cooper, 1947). But the function of the endosperm is beyond simple nutrient delivery to the embryo. The endosperm is a source of signals involved in embryogenesis and interacts with the embryo in a coordinated way to regulate seed development (Olsen, 1998; van Hengel *et al*., 1998; Opsahl-Ferstad *et al*., 1997).

Substantial evidence shows that the coordinated expression of the maternal, embryo and endosperm tissues is under genetic control (Walbot & Evans, 2003). Much research has been performed to study the interaction effect between maternal and offspring zygotic genes on seed development (Evans & Kermicle, 2001; Dilkes *et al*., 2002). Despite its paramount importance in seed biology and crop breeding, however, the knowledge of the genetic control of the co-regulation between the embryo and endosperm is limited. Genetic mapping based on molecular linkage maps has proven powerful for identifying individual loci, known as quantitative trait loci (QTL), that affect complex phenotypes. Lander & Botstein (1989) proposed an interval mapping approach for mapping QTL on a particular chromosomal interval bracketed by two flanking markers. This approach was later improved by including markers from other intervals as covariates to control the overall genetic background (Jansen & Stam, 1994; Zeng, 1994). The improved method, called composite interval mapping by Zeng (1994), displays increased power in QTL detection because of reduced residual variance. Kao *et al*.

* Corresponding authors. Tel: +1 (352) 3923806. Fax: +1 (352) 3928555. e-mail: rwu@stat.ufl.edu; Tel: -86(571)-6971691, e-mail: chhshi@zju.edu.cn

(1999) proposed using multiple marker intervals simultaneously to map multiple QTL of epistatic interactions throughout a linkage map. The advantage of these mapping approaches is that they allow for a genome-wide search for the existence of QTL and estimates of their chromosomal locations, genetic actions and interactions.

Taking into account the unique trisomic inheritance nature of the endosperm, several statistical methods have been proposed to map endosperm-specific QTL that are expressed on the endosperm itself (Wu *et al*., 2002*a*, *b*; Xu *et al*., 2003; Kao, 2004). These methods have also considered the generation difference between the maternal sporophyte and the embryo/endosperm (offspring) tissue. The markers used may be genotyped solely from the maternal tissue or from both the maternal and embryo tissues. More recently, we have constructed a general framework model for characterizing the interaction between different QTL from the maternal and embryo genomes (Cui *et al*., 2004; Cui & Wu, 2005*a*, *b*). Simulation studies were used to examine the statistical properties of this interactive model.

In this article, we develop a new statistical model for unravelling the interactive network between the QTL expressed in the embryo and endosperm genomes and estimating the effects of these tissue-specific QTL on quantitative traits expressed on the endosperm. Our model is derived within the maximum likelihood context and implemented with the EM algorithm (Dempster *et al*., 1977). Computer simulation studies are performed to investigate the accuracy and precision of parameter estimates from this new model and test its power to detect QTL interactions under different sample sizes and heritability levels. The successful detection of significant interactive QTL effects from the embryo and endosperm genomes on endosperm regulation in rice has validated the usefulness of our model.

2. Methodology

(i) *Genetic design*

Consider an F_1 heterozygote, *Aa*, derived from two homologous inbred lines, high-valued P_1 (*AA*) and low-valued P_2 (*aa*), in plants. This F_1 as the female parent is crossed with each of the two parents to generate seeds known as a backcross progeny, i.e. $F_1 \times P_1$ and $F_1 \times P_2$. These seeds each contain two different organizations: the diploid embryo and the triploid endosperm resulting from the combination of the polar nucleus of two central cells and a sperm nucleus. In this article, we postulate two QTL active in the embryo or endosperm genomes to jointly regulate endosperm-specific traits. Because it is more difficult to genotype a triploid than a diploid, a linkage map is constructed with the diploid embryo. The two backcrosses are assumed to share an identical linkage map. As the most important source of staple grains, the endosperm is measured for a phenotypic trait of interest. Our aim is to develop a statistical model for mapping the QTL from both the embryo and endosperm that epistatically affect the endosperm trait using molecular markers from the embryo.

Suppose there are two interacting QTL, labelled **A** with two alleles *A* and *a*, and **B** with two alleles *B* and *b*, for an endosperm trait. These two QTL can have three different patterns of expression:

(1) both are located on the embryo genome (embryo–embryo),
(2) both are located on the endosperm genome (endosperm–endosperm),
(3) one is located on the embryo genome and the other on the endosperm genome (embryo–endosperm).

Regardless of their genome locations, the two QTL always generate four different genotypes in each of the backcrosses $F_1 \times P_1$ and $F_1 \times P_2$. The value of a backcross QTL genotype (j) under QTL expression pattern k can be generally expressed as

$$\mu_{kj} = \mu_k + \xi_{k1} a_{k1} + \xi_{k2} a_2 + \xi_{k1} \xi_{k2} I \tag{1}$$

where μ_k is the overall mean, a_{k1} and a_{k2} are the additive effects of QTL **A** and **B**, respectively, I_k is the additive × additive interaction between these two QTL (Lynch & Walsh, 1998) and ξ_{k1} and ξ_{k2} are the indicator variables that define the QTL genotypes for the two backcrosses under QTL location pattern k. Assuming that the capital letters *A* and *B* are the favourable alleles at two different QTL, respectively, ξ_{k1} and ξ_{k2} are defined as in Table 1 for different QTL location patterns in the two backcrosses.

(ii) *The likelihood and parameter estimation*

To fully use the genetic information from different backcross populations, we construct a joint model for integrating likelihood functions of two different backcrosses under the same pattern of QTL expression. Suppose there are n and n' members in backcross $F_1 \times P_1$ and $F_1 \times P_2$, respectively, whose endosperm-specific trait (y or y') is affected by the two QTL. For location pattern k, the likelihood function of unknown QTL effects given the phenotypic values of the endosperm for the two backcross populations is formulated, on the basis of a finite mixture model, as

$$L_k(\Omega_k) = \prod_{i=1}^{n} \left[\sum_{j=1}^{4} \varpi_{kj|i} f_{kj}(y_i) \right] \prod_{i=1}^{n'} \left[\sum_{j'=1}^{4} \varpi_{kj'|i} f_{kj'}(y_i') \right] \tag{2}$$

where $\Omega_k = (\theta_k, \beta_k, \sigma_k^2, \sigma_k'^2)$ contains unknown QTL position (θ_k) and QTL-effect parameters (β_k) and the residual variances (σ_k^2, $\sigma_k'^2$) being estimated, $\varpi_{kj|i}$ is the conditional probability of the jth joint QTL genotype

https://doi.org/10.1017/S0016672306007956 Published online by Cambridge University Press

Table 1. *Genotypic compositions of four QTL genotypic values for each backcross under three different patterns of QTL expression*

		QTL		Coefficients			
Expression pattern	Backcross	Genotype	Value	μ_k	a_{k1}	a_{k2}	I_k
Embryo–embryo ($k=1$)	$F_1 \times P_1$	$(AA)(BB)$	μ_{11}	1	$^1/_2$	$^1/_2$	$^1/_4$
		$(AA)(Bb)$	μ_{12}	1	$^1/_2$	$-^1/_2$	$-^1/_4$
		$(Aa)(BB)$	μ_{13}	1	$-^1/_2$	$^1/_2$	$-^1/_4$
		$(Aa)(Bb)$	μ_{14}	1	$-^1/_2$	$-^1/_2$	$^1/_4$
	$F_1 \times P_2$	$(Aa)(Bb)$	μ'_{11}	1	$-^1/_2$	$-^1/_2$	$^1/_4$
		$(Aa)(bb)$	μ'_{12}	1	$-^1/_2$	$-^3/_2$	$^3/_4$
		$(aa)(Bb)$	μ'_{13}	1	$-^3/_2$	$-^1/_2$	$^3/_4$
		$(aa)(bb)$	μ'_{14}	1	$-^3/_2$	$-^3/_2$	$^9/_4$
Endosperm–endosperm ($k=2$)	$F_1 \times P_1$	$(AAA)(BBB)$	μ_{21}	1	$^3/_2$	$^3/_2$	$^9/_4$
		$(AAA)(Bbb)$	μ_{22}	1	$^3/_2$	$-^1/_2$	$-^3/_4$
		$(Aaa)(BBB)$	μ_{23}	1	$-^1/_2$	$^3/_2$	$-^3/_4$
		$(Aaa)(Bbb)$	μ_{24}	1	$-^1/_2$	$-^1/_2$	$^1/_4$
	$F_1 \times P_2$	$(AAa)(BBb)$	μ'_{21}	1	$^1/_2$	$^1/_2$	$^1/_4$
		$(AAa)(bbb)$	μ'_{22}	1	$^1/_2$	$-^3/_2$	$-^3/_4$
		$(aaa)(BBb)$	μ'_{23}	1	$-^3/_2$	$^1/_2$	$-^3/_4$
		$(aaa)(bbb)$	μ'_{24}	1	$-^3/_2$	$-^3/_2$	$^9/_4$
Embryo–endosperm ($k=3$)	$F_1 \times P_1$	$(AA)(BBB)$	μ_{31}	1	$^1/_2$	$^3/_2$	$^3/_4$
		$(AA)(Bbb)$	μ_{32}	1	$^1/_2$	$-^1/_2$	$-^1/_4$
		$(Aa)(BBB)$	μ_{33}	1	$-^1/_2$	$^3/_2$	$-^3/_4$
		$(Aa)(Bbb)$	μ_{34}	1	$-^1/_2$	$-^1/_2$	$^1/_4$
	$F_1 \times P_2$	$(Aa)(BBb)$	μ'_{31}	1	$-^1/_2$	$^1/_2$	$-^1/_4$
		$(Aa)(bbb)$	μ'_{32}	1	$-^1/_2$	$-^3/_2$	$^3/_4$
		$(aa)(BBb)$	μ'_{33}	1	$-^3/_2$	$^1/_2$	$-^3/_4$
		$(aa)(bbb)$	μ'_{34}	1	$-^3/_2$	$^3/_2$	$^9/_4$

for QTL **A** and **B** given a marker interval for individual i in backcross $F_1 \times P_1$ and $f_{kj}(y_i)$ is the normal density corresponding to the jth genotype with mean μ_{kj} and variance σ_k^2 under QTL expression pattern k. Similar definitions of $\varpi_{kj'|i}$, $f_{kj'}(y_i')$, $\mu_{kj'}$ and $\sigma_k'^2$ can also be given.

Consider two flanking markers, $\mathbf{M_1}$ and $\mathbf{M_2}$, derived from the embryos of a backcross seed, whose recombination fraction is denoted by r. A putative embryo QTL (**A**) that exerts an effect on the endosperm trait is located between these two markers, as measured by the recombination fraction r_1 with $\mathbf{M_1}$ and r_2 with $\mathbf{M_2}$. The conditional probabilities of an embryo QTL genotype, conditional upon the four embryo marker genotypes in the backcross can be derived, as shown in Cui *et al.* (2004). This conditional probability matrix is denoted by Φ_1. The endosperm trait is also affected by the endosperm QTL (**B**). Accordingly, the conditional probabilities of a putative endosperm QTL genotype, conditional on the four embryo marker genotypes is derived (Cui *et al.*, 2004) and denoted by Φ_2. Both the embryo and endosperm QTL epistatically affect an endosperm-specific trait of interest and they could be located either on the same marker interval or on different marker intervals. If they are located on different intervals, the conditional probability matrix (Φ) of the joint embryo–endosperm QTL genotypes, conditional upon two different marker intervals, can be expressed as $\Phi = \Phi_1 \otimes \Phi_2$, where $\otimes$ refers to the matrix direct product operation. If two linked QTL are located within the same marker interval, the joint conditional probabilities (Φ) of the two QTL conditional upon the embryo marker genotypes of the flanking markers (bracketing two putative QTL) should be re-derived (Table 2).

We derive the standard EM algorithm to obtain the maximum likelihood estimates (MLEs) of Ω_k in the mixture model (2). As usual, we use a grid approach to estimate the QTL locations by fixing the putative QTL at particular locations between two flanking markers. The genetic effects (β_k) of QTL that comprise the genotypic values μ_{kj} (Table 1) and the residual variances are estimated using the algorithm given in the Appendix. In this particular study, we assume $\sigma_k^2 = \sigma_k'^2$.

By assuming different QTL expression patterns, the corresponding plug-in likelihoods are calculated. A most likely pattern is determined which corresponds to a maximum likelihood.

3. Materials

The F_1 heterozygote between two rice inbred lines, ZS97 and MH63, was self-crossed for nine

https://doi.org/10.1017/S0016672306007956 Published online by Cambridge University Press

Table 2. *Conditional probabilities of joint embryo–endosperm QTL genotypes given embryo marker genotypes of the same interval in a backcross design* $F_1 \times P_1$

	QTL genotype			
Marker genotype	$\{Aa\}\ \{BBb\}$	$\{Aa\}\ \{bbb\}$	$\{aa\}\ \{BBb\}$	$\{aa\}\ \{bbb\}$
$M_1m_1M_2m_2$	$\frac{(1-r_{1A})(1-r_{AB})(1-r_{B2})}{1-r}$	$\frac{(1-r_{1A})r_{AB}r_{B2}}{1-r}$	$\frac{r_{1A}r_{AB}(1-r_{B2})}{1-r}$	$\frac{r_{1A}(1-r_{AB})r_{B2}}{1-r}$
$M_1m_1m_2m_2$	$\frac{(1-r_{1A})(1-r_{AB})r_{B2}}{r}$	$\frac{(1-r_{1A})r_{AB}(1-r_{B2})}{r}$	$\frac{r_{1A}r_{AB}r_{B2}}{r}$	$\frac{r_{1A}(1-r_{AB})r_{B2}}{r}$
$m_1m_1M_2m_2$	$\frac{r_{1A}(1-r_{AB})(1-r_{B2})}{r}$	$\frac{r_{1A}r_{AB}r_{B2}}{r}$	$\frac{(1-r_{1A})r_{AB}(1-r_{B2})}{r}$	$\frac{(1-r_{1A})(1-r_{AB})r_{B2}}{r}$
$m_1m_1m_2m_2$	$\frac{r_{1A}(1-r_{AB})r_{B2}}{1-r}$	$\frac{r_{1A}r_{AB}(1-r_{B2})}{1-r}$	$\frac{(1-r_{1A})r_{AB}r_{B2}}{1-r}$	$\frac{(1-r_{1A})(1-r_{AB})(1-r_{B2})}{1-r}$

The markers and QTL are assumed to have order $\mathbf{M_1ABM_2}$. r_{1A}, r_{AB} and r_{B2} are the recombination fractions between marker $\mathbf{M_1}$ and the embryo QTL, between the embryo and endosperm QTL, and between the endosperm QTL and $\mathbf{M_2}$, respectively, and r is the recombination fraction between the two markers.

generations to produce 241 recombinant inbred lines (RILs) for high-resolution genetic mapping of genes influencing endosperm traits. Those RILs that are homozygous for the alternative alleles have been genotyped for 221 polymorphic markers distributed throughout the genome to construct a molecular linkage map composed of 12 chromosomes (Fig. 1). These RILs as the female parent were backcrossed toward the two original inbred lines as the male parent, which generate two backcross populations, RIL × ZS97 and RIL × MH63, each containing 241 plants. All the RIL and backcross plants have been evaluated for gel consistency in their endosperm tissues to determine any major QTL segregating in this material. Gel consistency, measured by gel length (mm) from the bottom of the tube to the front of the gel migration, is one of the most important traits for the cooking and eating quality of rice. Gel consistency was determined according to Cagampang *et al*. (1973).

4. Results

A molecular linkage map was constructed with 221 polymorphic markers distributed throughout the genome with 12 chromosomes (Fig. 1) for two backcrosses RIL × ZS97 and RIL × MH63 in rice. The newly developed two-QTL epistatic model was used to analyse marker (embryo) and phenotypic (endosperm) data collected in these two backcross populations. By scanning across the entire genome, maximum log-likelihood ratio (LR) values for gel consistency expressed on the endosperm were found to be 197, 228 and 630 through the joint analysis of the two backcrosses when both QTL are expressed on the embryo genome or on the endosperm genome or when one QTL is expressed on the embryo genome and the other on the endosperm genome, respectively. As a result, we suggest that QTL location pattern 3 is optimal for this dataset, implying that two putative QTL are located on different genomes. Pattern 3 is statistically significant because its LR value is markedly larger than the critical value empirically obtained from permutation tests.

The two detected significant QTL are at a similar location in the genome for the embryo and endosperm, i.e. 10–12 cM from the first marker on the top of chromosome 6 (Fig. 2). The QTL location detected is near a candidate gene, *Waxy*, that is associated with a critical step in amylose biosynthesis (Okagaki & Wessler, 1988). It has been known that *Waxy* is expressed in the embryo genome, which suggests that the first QTL detected by our model may be derived from the embryo genome.

We estimated the additive and additive × additive epistatic effects of the detected QTL on gel consistency in both backcross populations (Table 3). Further hypotheses were performed for the significance tests of the additive and epistatic genetic effects. The MLEs of the genetic effect parameters a_{31}, a_{32} and I_3 are −46·32, 32·55 and −4·93, with the respective LR values suggesting that they all are highly significant. Different directions of a_{31} and a_{32} suggest that the embryo and endosperm QTL affect the endosperm trait differently. Whereas the endosperm QTL exerts a positive effect on the endosperm trait, the embryo QTL displays a negative effect. The two QTL from different tissues together explain about two-thirds of the observed variance for gel consistency.

https://doi.org/10.1017/S0016672306007956 Published online by Cambridge University Press

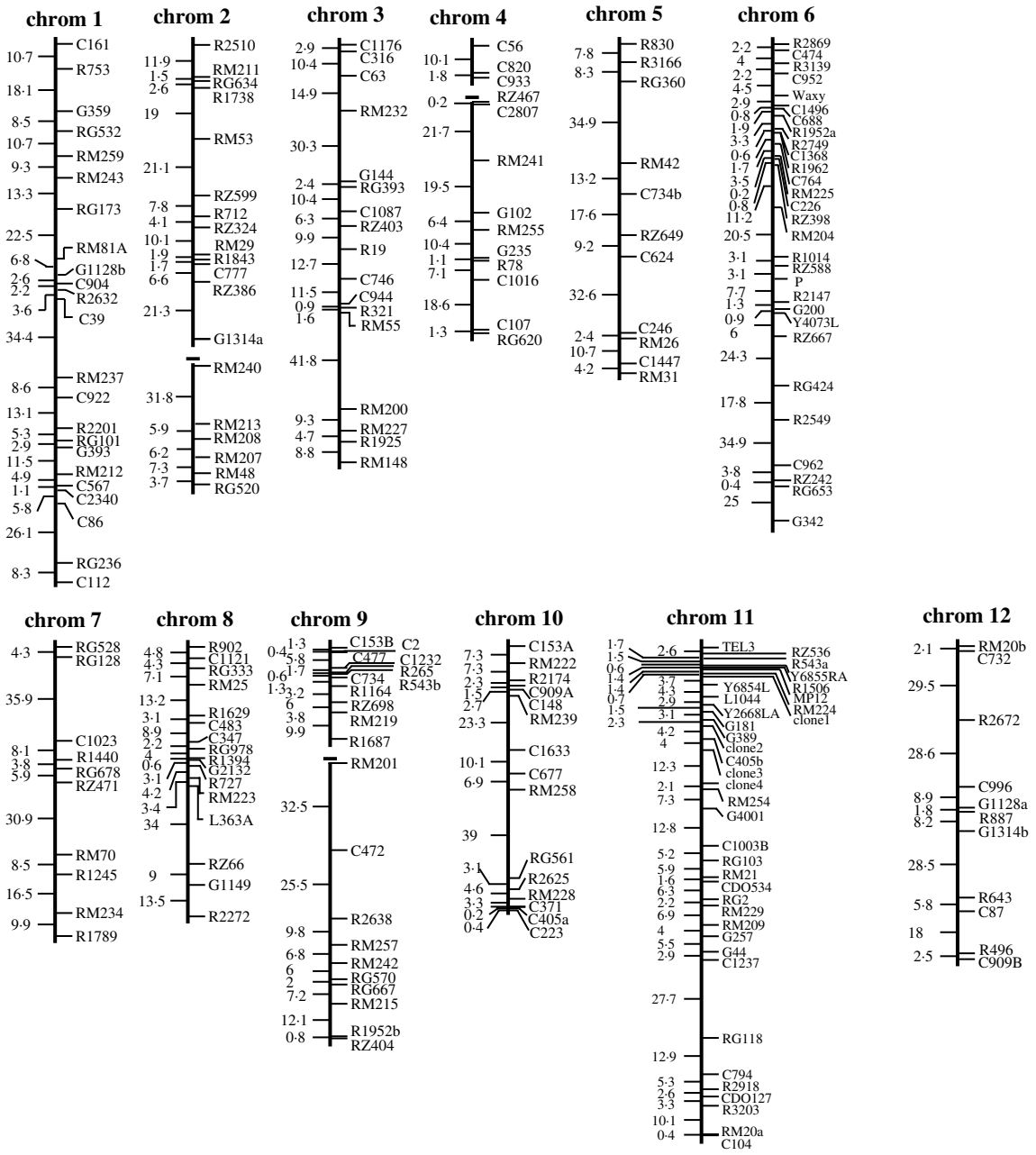

Fig. 1. A molecular linkage map of the rice genome using 241 RILs derived from two inbred lines, ZS97 and MH63. Some gaps on chromosomes 2, 4 and 9 are indicated by dotted lines.

5. Monte Carlo simulation

We proposed a series of simulation studies to examine the statistical properties of the model. Five equidistant markers are simulated from the embryo population and are ordered as $\mathbf{M_1}$–$\mathbf{M_5}$ on a linkage group with the length of 80 cM. The Kosambi map function was used to convert the map distance into the recombination fraction. Different heritability levels ($H^2 = 0{\cdot}1$ and $0{\cdot}4$) and different sample size ($n = 200$ and 400) were considered in the simulation study to examine the model performance under different situations.

Suppose there are two different putative QTL that affect a quantitative endosperm trait of interest, one expressed in the embryo genome and the other expressed in the endosperm genome. The two QTL could either be linked together and located on the same marker interval ($\mathcal{L}_1$) or located on different marker intervals ($\mathcal{L}_2$). For the $\mathcal{L}_1$ case, the embryo and endosperm QTL locations are hypothesized at 8 cM and 16 cM from the marker $\mathbf{M_1}$ respectively. For $\mathcal{L}_2$, the embryo QTL location is hypothesized at 12 cM from marker $\mathbf{M_1}$ and the endosperm QTL is hypothesized at 8 cM from marker $\mathbf{M_3}$. Two sets of

https://doi.org/10.1017/S0016672306007956 Published online by Cambridge University Press

Table 3. *MLEs of the additive genetic effect of the embryo* (a_{31}) *and endosperm QTL* (a_{32}) *and their additive × additive epistatic interaction effect* (I_3) *on gel consistency in the endosperm for two backcross progenies derived from two inbred lines in rice*

				Additive effects			Epistatic effect				
QTL	Marker interval	Location	Mean	MLE	LR	*P* value	MLE	LR	*P* value	σ^2	R^2
Embryo	C952–*Waxy*	1·8	23·3	−46·3	517	$1{\cdot}8\times10^{-113}$	−4·9	138	$1{\cdot}4\times10^{-30}$	38·1	0·97
Endosperm	C952–*Waxy*	3·8		32·5	56	$6{\cdot}8\times10^{-122}$					

Note: The locations of detected QTL are described in centimorgans from the first marker of the interval.

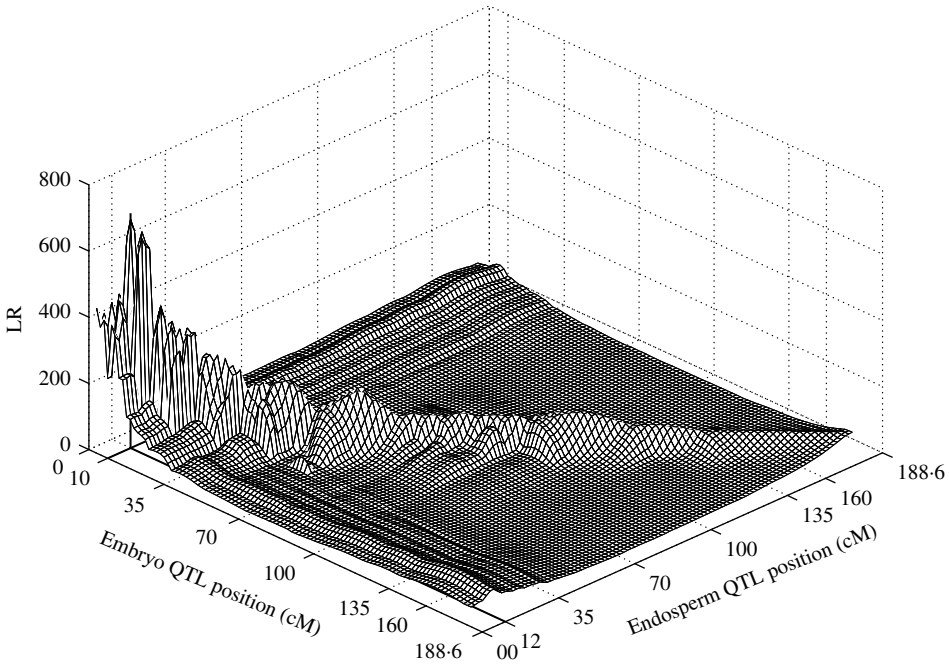

Fig. 2. The landscape of LR values for the existence of interactive QTL derived from the embryo and endosperm genomes throughout chromosome 6. The locations of the embryo and endosperm QTL are indicated.

parameter values are hypothesized, which include large additive effects versus small interaction effects (Tables 4 and 5) and small additive effects versus large interaction effects (Table 6). The endosperm trait values for each seed were simulated from a normal distribution with different joint genetic means and a residual variance.

In general, our model can provide reasonable estimates of the QTL positions and effects of various kinds, with estimation precision depending on heritability, sample size, sampling strategy, gene action mode and QTL location. Our model has excellent power to detect epistatically interacting embryo and endosperm QTL effects. In all cases of different sample sizes and heritabilities, the maximum values of the LR landscapes from 200 simulation replicates are beyond the critical thresholds at the $\alpha=0{\cdot}001$ level determined from 1000 permutation tests for the simulated data. Figs 3 and 4 are examples of the shapes of the LR landscapes for two contrasting sample sizes, heritabilities and sampling strategies when the QTL are located at different marker intervals or at the same interval, respectively. Small differences between these two lines suggest that our model can accurately estimate the genomic positions of the QTL.

The precision of parameter estimation is evaluated in terms of the square roots of the mean squared errors (SRMSEs) of the MLEs. The QTL positions and effects can be better estimated when the endosperm trait has higher rather than a lower heritability or when sample size is larger rather than smaller (Tables 4 and 5). But the increase in H^2 from 0·1 to 0·4 leads to more significant improvement for the estimation precision than the increase in m from 200 to 400. For example, the SRMSEs of the MLEs of the

https://doi.org/10.1017/S0016672306007956 Published online by Cambridge University Press

Table 4. *The MLEs of the QTL position and effect parameters between an embryo QTL and an endosperm QTL each at a different interval derived from 200 simulation replicates. The square roots of the mean square errors of the MLEs are given in parentheses*

H^2	N	Positions 12, 48	$\mu=5$	$a_1=0{\cdot}5$	$a_2=0{\cdot}5$	$I=0{\cdot}3$	σ^2
0·1	200	13·12, 47·62	4·9952	0·5346	0·4723	0·3779	1·1487
		(9·7812, 8·3971)	(0·1390)	(0·2913)	(0·2959)	(0·6809)	(0·1268)
	400	12·98, 48·86	4·9894	0·5239	0·4560	0·3400	1·1600
		(8·3392, 7·5606)	(0·0849)	(0·2104)	(0·2251)	(0·4144)	(0·0818)
0·4	200	12·12, 48·08	4·9977	0·4990	0·4918	0·3046	0·1936
		(4·7953, 4·3375)	(0·0410)	(0·0973)	(0·0987)	(0·1903)	(0·0221)
	400	12·04, 47·82	5·0010	0·4963	0·4961	0·3057	0·1945
		(3·5642, 3·2945)	(0·0326)	(0·0701)	(0·0730)	(0·1413)	(0·0168)

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 1·1756 for $H^2=0{\cdot}1$ and 0·1959 for $H^2=0{\cdot}4$.

Table 5. *The MLEs of the QTL position and effect parameters between an embryo QTL and an endosperm QTL both at the same interval derived from 200 simulation replicates. The square roots of the mean square errors of the MLEs are given in parentheses*

H^2	n	Positions 8, 16	$\mu=5$	$a_1=0{\cdot}5$	$a_2=0{\cdot}5$	$I=0{\cdot}3$	σ^2
0·1	200	10·80, 41·22	5·0453	0·8387	0·1409	0·1226	1·0965
		(8·5702, 34·2914)	(0·2489)	(0·7179)	(0·7724)	(1·1846)	(0·1526)
	400	9·48, 38·14	5·0139	0·7946	0·1978	0·2601	1·1388
		(6·7212, 32·5040)	(0·1882)	(0·6707)	(0·6976)	(0·8769)	(0·0903)
0·4	200	7·30, 31·30	5·0387	0·6455	0·3567	0·1528	0·1886
		(5·2818, 26·3084)	(0·1058)	(0·3579)	(0·3751)	(0·4685)	(0·0237)
	400	7·16, 27·54	5·0272	0·5959	0·4059	0·1995	0·1938
		(4·5893, 22·4216)	(0·0843)	(0·2924)	(0·2984)	(0·3628)	(0·0157)

See Table 4 for the explanations.

genetic parameters reduce by more than one-fold when H^2 is increased from 0·1 to 0·4 for a fixed sample size, whereas the reduction is much smaller when n is increased from 200 to 400 for a fixed heritability. This suggests that in practice it is more important to manage experiments so as to reduce residual errors (increase H^2) than simply to increase the sample size.

As expected, the estimation precision of the additive effects is better than that of the interaction effect. There is no great difference in the estimation of the embryo and endosperm additive effects (Tables 4 and 5). The precision of parameter estimation is better estimated when the embryo and endosperm QTL are located at different marker intervals (Tables 4) than when they are located at the same interval (Table 5). Thus, to avoid the analysis of two different QTL located at the same interval, a high-density map is needed. Our model can estimate well the parameters for different gene action modes, large additive versus small interaction effect (Tables 4 and 5) and small additive versus large interaction effect (Table 6). In all the cases, the estimates of the genetic effect parameters can be very biased when different QTL are located at the same marker interval.

6. Discussion

Coordinated interaction between the embryo and the endosperm is fundamental to seed development (Chaudhury *et al*., 2001; Walbot & Evans, 2003). The size of the endosperm that contains essential protein, starch and other nutrients is an important criterion for crop breeding programmes. Historically, quantitative genetic analyses of seed development are conducted separately for maternal, embryo and endosperm tissues and have failed to consider joint influences of these different but physiologically co-regulated tissues on the expression of a trait. Cui *et al*. (2004) made a first attempt to derive a unifying model for estimating the effects of different QTL derived from the maternal and offspring (embryo) genomes. The model described in this article can unravel the genetic control mechanisms of endosperm-specific

https://doi.org/10.1017/S0016672306007956 Published online by Cambridge University Press

Table 6. *The MLEs of the QTL position and effect parameters between an embryo QTL and an endosperm QTL at different intervals and at same interval derived from 200 simulation replicates under the heritability of 0·4. The square roots of the mean square errors of the MLEs are given in parentheses*

n	QTL positions	$\mu=5$	$a_1=0{\cdot}4$	$a_2=0{\cdot}4$	$I=0{\cdot}9$	$\sigma^2=0{\cdot}1959$
			Different intervals			
200	11·66, 48·08	5·0041	0·4027	0·4019	0·8970	0·1935
	(4·2343, 4·5897)	(0·0423)	(0·0943)	(0·0883)	(0·1754)	(0·0214)
400	11·88, 48·04	4·9990	0·3982	0·4054	0·8971	0·1935
	(2·9468, 2·7491)	(0·0293)	(0·0620)	(0·0647)	(0·1324)	(0·0155)
			The same interval			
200	8·14, 32·02	5·0778	0·5088	0·2896	0·6060	0·1895
	(4·9358, 27·9482)	(0·1509)	(0·3042)	(0·3193)	(0·6329)	(0·0212)
400	7·52, 23·00	5·0544	0·4593	0·3445	0·6909	0·1939
	(3·9696, 17·6964)	(0·1173)	(0·2721)	(0·2712)	(0·4903)	(0·0156)

The positions of the two QTL are described by the map distances, (12, 48) for the QTL at different intervals and (8, 16) for the QTL at the same interval, from the first marker of the linkage group (80 cM long).

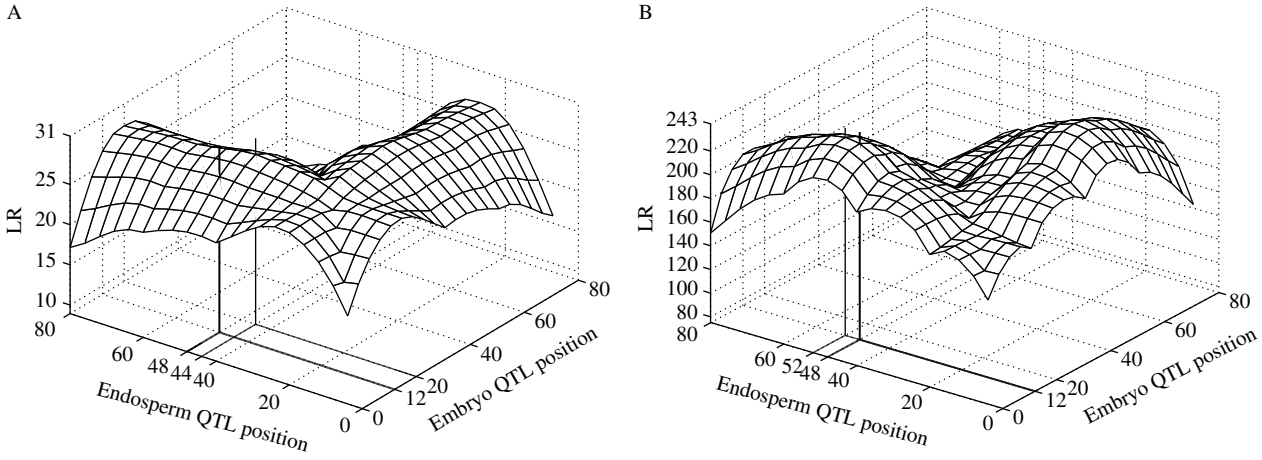

Fig. 3. The landscapes of the log-likelihood ratio (LR) test statistics calculated for the hypothesis test of the existence of QTL against the embryo and endosperm QTL locations at different marker intervals for a heritability of 0·1 and a sample size of 200 (*A*) and for a heritability of 0·4 and a sample size of 400 (*B*). The true and estimated QTL positions are indicated by the thick and thin lines, respectively.

traits conferred by the embryo and endosperm QTL, and can thus be thought of as complementary to Cui *et al.*'s (2004) and Cui & Wu's (2005*a*, *b*) models.

Our model integrates parameters that account for genetic epistasis between the embryo and endosperm genomes into a general QTL mapping model for endosperm traits. A large body of evidence has suggested that genome–genome interactions can provide supplementary genetic variation in adaptation to a wide spectrum of environments and, therefore, may have played a more important role in shaping the evolutionary process of organisms than originally appreciated (Mousseau & Fox, 1998). Given such a feature, this model can be expected to have great implications for the study of evolutionary genetic problems related to seed development in higher plants (Walbot & Evans, 2003). From a statistical perspective, this model should be able to provide biologically more realistic results than many existing models because it integrates the information about gene segregation and transmission from the maternal to offspring generations at both the embryo and endosperm QTL.

We have conducted extensive computer simulations to investigate the statistical properties of this model. It is robust in that it can provide a reasonable estimate of QTL position and effect parameters at modest sample sizes and heritability levels. The simulation studies have also provided information about the impact of different gene action modes, different origins of QTL, and different QTL locations on the precision of parameter estimation. Our model presented here provides an important step towards incorporating the control of seed development within a QTL mapping framework.

https://doi.org/10.1017/S0016672306007956 Published online by Cambridge University Press

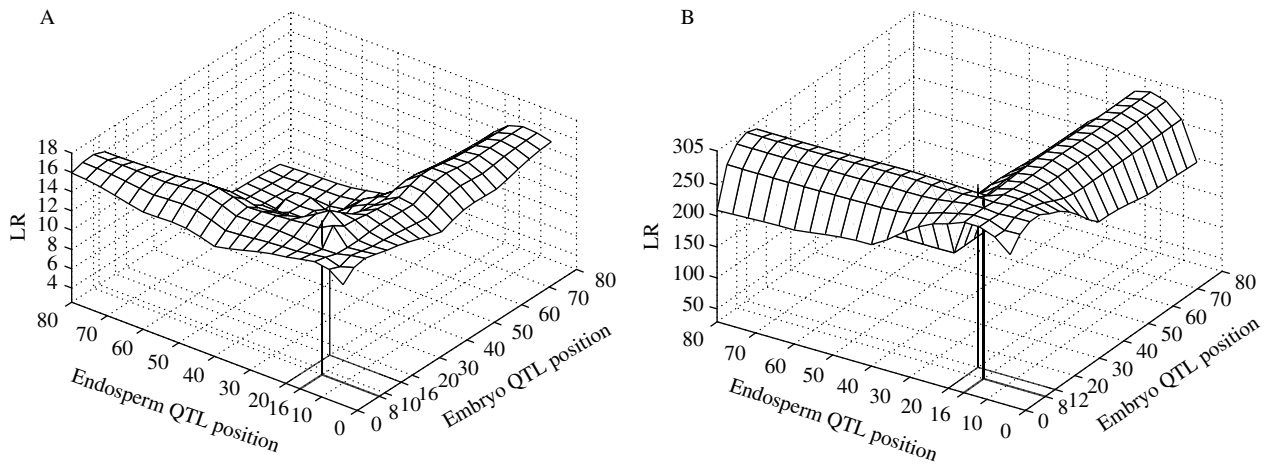

Fig. 4. The landscapes of the log-likelihood ratio (LR) test statistics calculated for the hypothesis test of the existence of QTL against the embryo and endosperm QTL locations at the same marker interval for a heritability of 0·1 and a sample size of 200 (*A*) and for a heritability of 0·4 and a sample size of 400 (*B*). The true and estimated QTL positions are indicated by the thick and thin lines, respectively.

Perhaps the most important aspect of this model is its successful discovery of significant QTL that trigger considerable effects on differentiation in an endosperm trait. Our model identified an embryo QTL that interacts with an endosperm QTL at a similar chromosomal location to regulate gel consistency in the endosperm for two RIL-derived backcross populations in rice. This embryo QTL was detected at almost the exact position of a candidate gene, known as the *Waxy* gene, on the short arm of chromosome 6 (Terada *et al*., 2002). The *Waxy* gene that is responsible for the synthesis of amylose in the endosperm and pollen is genetically well characterized in many grasses including maize and rice. This consistency with the *Waxy* gene convincingly demonstrates the power of our model to probe significant QTL hidden in real datasets. The further functional analysis of this detected embryo QTL will accelerate its usefulness for improving the quality and quantity of rice grains. It should be pointed out that the embryo QTL for gel consistency near the *Waxy* gene was also detected by single-marker analysis and interval mapping approaches that do not incorporate QTL interactions from the embryo and endosperm genomes (Tan *et al*., 1999). But our model possesses the unique power to unravel how these two different genomes are coordinated to regulate the expression of an endosperm trait. Thus, beyond the traditional methods, our model will help to shed light on the genetic and physiological mechanisms underlying seed development.

Our model combines two backcross mapping populations initiated from two inbred lines, in which the same QTL can be assumed to be segregating. Zou *et al*. (2001) showed increased power for QTL detection when different related crosses are combined. In order to clearly present our idea for mapping genome–genome interactions based on a combined analysis, we have derived the model within the context of interval mapping (Lander & Botstein, 1989). More sophisticated models that incorporate the ideas of composite interval mapping and multiple interval mapping should be developed (see Kao *et al*., 1999). Such more comprehensive approaches, in conjunction with increasingly accumulating genetic and genomic data, will make our model more useful and powerful in practice.

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. We thank Dr Qifa Zhang of Huazhong Agricultural University for providing molecular marker data. The publication of this manuscript was approved as Journal Series No. R-10585 by the Florida Agricultural Experiment Station.

Appendix. EM algorithm

To obtain the MLEs of genetic effects and residual variance, $\Theta_k=(\beta_k, \sigma_k^2)$, at a given QTL location, we implement the EM algorithm within the mixture-based likelihood function. The log-likelihood function described by equation (2), where the conditional probabilities of QTL genotypes given marker genotypes are a function of QTL positions expressed as θ_k, are given, under expression pattern k, by

$$\log L_k(\theta_k, \boldsymbol{\Theta}_k)=\sum_{i=1}^{n}\log\left[\sum_{j=1}^{4}\varpi_{kj|i}f_{kj}(y_i)\right]+\sum_{i=1}^{n'}\log\left[\sum_{j'=1}^{4}\varpi_{kj'|i}f_{kj'}(y_i')\right] \quad \text{(A1)}$$

with the derivative for the φth unknown parameter Ω_φ,

https://doi.org/10.1017/S0016672306007956 Published online by Cambridge University Press

$$\frac{\partial}{\partial \Omega_\varphi}\log L_k(\theta_k,\boldsymbol{\Theta}_k)$$
$$=\sum_{i=1}^{n}\sum_{j=1}^{4}\left[\frac{f_{kj}(y_i)\frac{\partial \varpi_{kj|i}}{\partial \theta_k}}{\sum_{\tilde{j}=1}^{4}\varpi_{k\tilde{j}|i}f_{k\tilde{j}}(y_i)}+\frac{\varpi_{kj|i}\frac{\partial}{\partial \boldsymbol{\Theta}_k}f_{kj}(y_i)}{\sum_{k\tilde{j}=1}^{4}\varpi_{k\tilde{j}|i}f_{k\tilde{j}}(y_i)}\right]+\sum_{i=1}^{n'}\sum_{j'=1}^{4}\left[\frac{f_{kj'}(y_i')\frac{\partial}{\partial \theta_k}\varpi_{kj'|i}}{\sum_{k\tilde{j}'=1}^{4}\varpi_{k\tilde{j}'|i}f_{k\tilde{j}'}(y_i')}+\frac{k\varpi_{j'|i}\frac{\partial}{\partial \boldsymbol{\Theta}_k}f_{kj'}(y_i')}{\sum_{k\tilde{j}'=1}^{4}\varpi_{k\tilde{j}'|i}f_{k\tilde{j}'}(y_i')}\right]$$
$$=\sum_{i=1}^{n}\sum_{j=1}^{4}\frac{k\varpi_{j|i}f_{kj}(y_i)}{\sum_{\tilde{j}=1}^{4}\varpi_{k\tilde{j}|i}f_{k\tilde{j}}(y_i)}\left[\frac{1}{\varpi_{kj|i}}\frac{\partial \varpi_{kj|i}}{\partial \theta_k}+\frac{\partial}{\partial \boldsymbol{\Theta}_k}\log f_{kj}(y_i)\right]+\sum_{i=1}^{n'}\sum_{j'=1}^{4}\frac{\varpi_{kj'|i}f_{kj'}(y_i')}{\sum_{\tilde{j}'=1}^{4}\varpi_{k\tilde{j}'|i}f_{k\tilde{j}'}(y_i')}\left[\frac{1}{\varpi_{kj'|i}}\frac{\partial \varpi_{kj'|i}}{\partial \theta_k}+\frac{\partial}{\partial \boldsymbol{\Theta}_k}\log f_{kj'}(y_i')\right]$$
$$=\sum_{i=1}^{n}\sum_{j=1}^{4}\Pi_{j|i}\left[\frac{1}{\varpi_{kj|i}}\frac{\partial \varpi_{kj|i}}{\partial \theta_k}+\frac{\partial}{\partial \boldsymbol{\Theta}_k}\log f_{kj}(y_i)\right]+\sum_{i=1}^{n'}\sum_{j'=1}^{4}\Pi_{kj'|i}\left[\frac{1}{\varpi_{kj'|i}}\frac{\partial \varpi_{kj'|i}}{\partial \theta_k}+\frac{\partial}{\partial \boldsymbol{\Theta}_k}\log f_{kj'}(y_i')\right]$$

where we define

$$\Pi_{kj|i}=\frac{\varpi_{kj|i}f_{kj}(y_i)}{\sum_{\tilde{j}=1}^{4}\varpi_{k\tilde{j}|i}f_{k\tilde{j}}(y_i)},\tag{A2}$$

$$\Pi_{kj'|i}=\frac{\varpi_{j'|i}f_{kj'}(y_i')}{\sum_{k\tilde{j}'=1}^{4}\varpi_{k\tilde{j}'|i}f_{k\tilde{j}'}(y_i')},\tag{A3}$$

which can be regarded as the posterior probability of joint QTL genotype j (or j') given a marker genotype for a backcross seed i under the expresison pattern k. Like the conditional (prior) probability $\varpi_{kj|i}$ (or $\varpi_{kj'|i}$), the posterior probability $\Pi_{kj|i}$ (or $\Pi_{kj'|i}$) is backcross-specific and, thus, forms an $n\times 4$ matrix Π and an $n'\times 4$ matrix $\boldsymbol{\Pi}'$.

Let $\beta_k=(\mu_k,\ a_{k1},\ a_{k2},\ I_k)'$ and $D_{k,\ell}$ be the design matrix (Table 1) for cross-combination $\mathrm{F}_1\times P_\ell$ ($\ell=1$, 2) under the expression pattern k. The EM algorithm is formulated as follows:

In the E step, calculate the posterior probability matrix for a given QTL genotype and endosperm trait value using given values for $\boldsymbol{\Theta}_k$ and QTL locations. The initial values are obtained through the least squares method.

In the M step, the calculated posterior probabilities are used to estimate unknown parameters by solving for the zeros of $\frac{\partial}{\partial \Omega_\varphi}\log L_k(\theta_k,\boldsymbol{\Theta}_k)$, which lead to

$$\hat{\beta}_{kv}=\frac{\mathbf{Y}_1'\Pi_{k,1}D_{k,1}^{v}-\mathbf{1}'\Pi_{k,1}\mathrm{Diag}\{D_{k,1}^{(-v)}\beta^{(-v)}\}D_{k,1}^{v}+\mathbf{Y}_2'\Pi_{k,2}D_{k,2}^{v}-\mathbf{1}'\Pi_{k,2}\mathrm{Diag}\{D_{k,2}^{(-j)}\beta^{(-j)}\}D_{k,2}^{j}}{\mathbf{1}'\Pi_{k,1}(D_{k,1}^{v})^2+\mathbf{1}'\Pi_{k,2}(D_{k,2}^{v})^2}\tag{A4}$$

$$\hat{\sigma}^2=\frac{\sum_{i=1}^{n}\sum_{j=1}^{4}\Pi_{kj|i}(y_i-D_{k,1}^{v}\hat{\boldsymbol{\beta}})^2+\sum_{i=1}^{n'}\sum_{j'=1}^{4}\Pi_{kj'|i}(y'_i-D_{k,2}^{v}\hat{\boldsymbol{\beta}})^2}{n+n'}$$
$$=\frac{1}{n+n'}\sum_{\ell=1}^{2}\{\mathbf{Y}'_\ell\mathbf{Y}_\ell-2\mathbf{Y}'_\ell\Pi_\ell D_{k,\ell}\hat{\boldsymbol{\beta}}+\mathbf{1}'\Pi_\ell(D_{k,\ell}\hat{\boldsymbol{\beta}})^2\}\tag{A5}$$

where $\Pi_{k,\ell}=\{\Pi_{kj,\ell|i}\}_{n_\ell\times 4}$ is the posterior probability matrix; $D_{k,l}^{(-v)}(l=1,\ 2)$ is the design matrix without the vth column; $\boldsymbol{\beta}^{(-v)}$ is the genetic parameter vector without the vth entry; $D_{k,1}^{v}$ is a vector which is the vth column of the design matrix $\mathbf{D}_k$; and Diag means diagonalizing the vector as a diagonal matrix.

The calculations are iterated between the E (equations A2 and A3) and M step (equations A4 and A5) until the estimates converge. The converged values are the MLEs of the unknown parameters.

Under the null hypothesis, with the parameters $a_1=a_2=I=0$, the estimates of μ and σ^2 are given by

$$\hat{\mu}=\frac{\mathbf{1}'\mathbf{Y}_1+\mathbf{1}'\mathbf{Y}_2}{n}$$

and

$$\hat{\sigma}^2=\frac{\sum_{\ell=1}^{2}(\mathbf{Y}_\ell-\hat{\mu})'(\mathbf{Y}_\ell-\hat{\mu})}{n}$$

https://doi.org/10.1017/S0016672306007956 Published online by Cambridge University Press

References

Brink, R. A. & Cooper, D. C. (1947). The endosperm in seed development. *Botanical Review* **13**, 423–541.

Cagampang, G. B., Perez, C. M. & Juliano, B. O. (1973). A gel consistency test for eating quality of rice. *Journal of Science and Food Agriculture* **24**, 1589–1594.

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 Review of Cell and Developmental Biology* **17**, 677–699.

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

Cui, Y. H. & Wu, R. L. (2005*a*). Statistical model for characterizing epistatic control of triploid endosperm triggered by maternal and offspring QTLs. *Genetical Research* **86**, 65–75.

Cui, Y. H. & Wu, R. L. (2005*b*). Mapping genome–genome epistasis: a high-dimensional model. *Bioinformatics* **21**, 2447–2455.

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 Royal Statistical Society*, *Series B* **39**, 1–38.

Dilkes, B. P., Dante, R. A., Coelho, C. & Larkins, B. A. (2002). Genetic analyses of endoreduplication in *Zea mays* endosperm: evidence of sporophytic and zygotic maternal control. *Genetics* **160**, 1163–1177.

Evans, M. M. S. & Kermicle, J. L. (2001). Interaction between maternal effect and zygotic effect mutations during maize seed development. *Genetics* **159**, 303–315.

Jansen, R. C. & Stam, P. (1994). High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* **136**, 1447–1455.

Kao, C. H. (2004). Multiple-interval mapping for quantitative trait loci controlling endosperm traits. *Genetics* **167**, 1987–2002.

Kao, C. H., Zeng, Z. B. & Teasdale, R. D. (1999). Multiple interval mapping for quantitative trait loci. *Genetics* **152**, 1203–1216.

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

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

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

Okagaki, R. J. & Wessler, S. R. (1988). Comparison of non-mutant and mutant waxy genes in rice and maize. *Genetics* **120**, 1137–1143.

Olsen, O. A. (1998). Endosperm developments. *Plant Cell* **10**, 485–488.

Opsahl-Ferstad, H. G., Le Deunff, E., Dumas, C. & Rogowsky, P. M. (1997). ZmEsr, a novel endosperm-specific gene expressed in a restricted region around the maize embryo. *Plant Journal* **12**, 235–246.

Shi, C. H., Zhu, J., Yang, X. E., Yu, Y. G. & Wu, J. G. (1999). Genetic analysis for protein content in indica rice. *Euphytica* **107**, 135–140.

Shi, C. H., Zhu, J. & Wu, J. G. (2000). Genetic and genotype × environment interaction effects from embryo, endosperm, cytoplasm and maternal plant for rice shape traits of indica rice. *Field Crops Research* **68**, 191–198.

Tan, Y. F., Li, J. X., Yu, S. B., Xing, Y. Z., Xu, C. G. & Zhang, Q. (1999). The three important traits for cooking and eating quality of rice grains are controlled by a single locus in an elite rice hybrid, Shanyou 63. *Theoretical Applied Genetics* **99**, 642–648.

Terada, R., Urawa, H., Inagaki, Y., Tsugane, K. & Iida, S. (2002). Efficient gene targeting by homologous recombination in rice. *Nature Biotechnology* **20**, 1030–1034.

van Hengel, A. J., Guzzo, F., van Kammen, A. & de Vries, S. C. (1998). Expression pattern of the carrot EP3 endochitinase genes in suspension cultures and in developing seeds. *Plant Physiology* **117**, 43–53.

Walbot, W. & Evans, N. M. S. (2003). Unique features of the plant life cycle and their consequences. *Nature Review Genetics* **4**, 369–379.

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.

Xu, C., He, X. & Xu, S. (2003). Mapping quantitative trait loci underlying triploid endosperm traits. *Heredity* **90**, 228–235.

Zeng, Z. B. (1994). Precision mapping of quantitative trait loci. *Genetics* **136**, 1457–1468.

Zou, F., Yandell, B. S. & Fine, J. P. (2001). Statistical issues in the analysis of quantitative traits in combined crosses. *Genetics* **158**, 1339–1346.