Theor Appl Genet (1999) 99:1255–1264

D.L. Wang · J. Zhu · Z.K. Li · A.H. Paterson

Mapping QTLs with epistatic effects and QTL×environment interactions by mixed linear model approaches

Received: 23 October 1998 / Accepted: 11 May 1999

Abstract A new methodology based on mixed linear models was developed for mapping QTLs with digenic epistasis and QTL×environment (QE) interactions. Reliable estimates of QTL main effects (additive and epistasis effects) can be obtained by the maximum-likelihood estimation method, while QE interaction effects (additive×environment interaction and epistasis×environment interaction) can be predicted by the-best-linear-unbiased-prediction (BLUP) method. Likelihood ratio and *t* statistics were combined for testing hypotheses about QTL effects and QE interactions. Monte Carlo simulations were conducted for evaluating the unbiasedness, accuracy, and power for parameter estimation in QTL mapping. The results indicated that the mixed-model approaches could provide unbiased estimates for both positions and effects of QTLs, as well as unbiased predicted values for QE interactions. Additionally, the mixed-model approaches also showed high accuracy and power in mapping QTLs with epistatic effects and QE interactions. Based on the models and the methodology, a computer software program (QTLMapper version 1.0) was developed, which is suitable for interval mapping of QTLs with additive, additive×additive epistasis, and their environment interactions.

Communicated by H.C. Becker

D.L. Wang[1] · J. Zhu (✉)
Department of Agronomy, Zhejiang Agricultural University, Hangzhou, Zhejiang 310029, China
e-mail: jzhu@zjau.edu.cn

Z.K.L. Li[2] · A.H. Paterson
Department of Soil and Crop Sciences, Texas A&M University, College Station, Texas 77843-2474, USA

Present addresses:
[1] Department of Statistics and Mathematics, Anhui Agricultural University, Hefei, Anhui 230036, China
[2] Plant Breeding, Genetics, and Biochemistry Division, International Rice Research Institute, P.O. Box 933, 1099 Manila, The Philippines

Key words QTL mapping method · Epistasis · QTL×environment interaction · Mixed linear model · Monte Carlo simulations

Introduction

Recent technical advances in DNA marker technology has triggered the "molecular dissection" of complex traits. As a result, many marker-based statistical methods have been developed for mapping quantitative trait loci (QTLs) (Weller 1986; Lander and Botstein 1989; Haley and Knott 1992; Moreno-Gonzalez 1992; Jansen 1992, 1993; Zeng 1993, 1994; Jansen and Stam 1994). QTLs affecting a wide range of quantitative traits have been mapped in plants and animals (Stuber et al. 1987; Paterson et al. 1988, 1991; Bradshaw and Stettler 1995; Neale and Sederoff 1996; Tanksley and Hewitt 1996). However, two important issues, epistasis and QTL×environment interaction (QE), in QTL mapping methodology remain largely unsolved.

The importance of epistasis has been suggested from numerous classic quantitative genetic studies (Spickett and Thoday 1966; Falconer 1981; Mather and Jinks 1982; Pooni et al. 1987; Allard 1988). Epistasis as an important genetic basis of complex phenotypes has also been revealed in several recent QTL mapping studies (Doebley et al. 1995; Lark et al. 1995; Wu et al. 1995; Fu and Ritland 1996; Li et al. 1997a, b; Routman and Cheverud 1997; Yu et al. 1997). However, current marker-based analyses for dissecting QTL effects usually assume absence of epistasis among QTLs (Weller 1986; Lander and Botstein 1989; Jansen 1993; Zeng 1993, 1994). This assumption was made largely for simplification of the statistical models, but violation of this assumption may result in biased estimates of the positions and effects of QTLs and a lower precision and power for QTL detection.

Statistical approaches of two-way ANOVA and multiple regression have been applied to analyzing epistasis for quantitative traits based on genetic markers (Wu et al. 1995; Xiao et al. 1995; Li et al. 1997a, b; Yu et al.

1997; Holland 1998). These methods, however, are only capable of detecting interaction between markers, by which epistasis between QTLs could be indicated but not reliably estimated. Therefore, some statistical models and methods based on marker linkage maps have been suggested for analyzing QTLs with epistatic effects (Haley and Knott 1992; Jansen 1992; Moreno-Gonzalez 1992; Jansen and Stam 1994; Cockerham and Zeng 1996; Fijneman et al. 1996; Chase et al. 1997). However, the accurate quantification and characterization of epistasis in the presence of genotype×environment (GE) interaction still remains unsolved.

GE interactions have been a very important issue for breeders and quantitative geneticists. With DNA markers and appropriate experimental designs, GE interactions can be further dissected into components of QE interactions, which are of great importance for marker-assisted selection in crop improvement. QE interactions have been considered in mapping QTLs without epistasis (Jansen 1992; Jansen et al. 1995; Tinker and Mather 1995; Jiang and Zeng 1995; Cockerham and Zeng 1996; Utz and Melchinger 1996; Sari-Gorla et al. 1997; Yan et al. 1998), but these methods do not allow reliable and efficient dissection of GE interactions into their contributing QE components in the presence of epistasis.

Mixed-linear-model approaches have been extensively used in animal breeding for estimating polygenic breeding values, and recently also for QTL-related analysis in animals (Fernando and Grossman 1989; Goddard 1992; Meuwissen and Goddard 1997). However, these applications of mixed-linear- model approaches have not been intended to solve the above problems. In the present paper, we propose a new methodology for systematically mapping QTLs involved in digenic epistasis as well as QE interactions based on the mixed-model approaches (Zhu 1998; Zhu and Weir 1998). The statistical models and analysis methods are described. Simulation results are provided for demonstrating the reliability, power and precision of the methods. Computer software has also been developed based on these approaches. Issues related to the mixed model approaches are discussed.

Genetic models and analysis methodology

For simplicity, we use a doubled-haploid (DH) population from a cross between two inbred lines to demonstrate this method. Extensions can be obtained for other mapping populations derived from line crosses (e.g., recombinant inbred lines, backcross, F_2, etc.).

Genetic models

According to the definitions of genetic effects (additive, dominance and digenic epistasis) given by Mather and Jinks (1982), a mixed linear model for the simultaneous search of two interacting QTLs (Q_i between flanking markers M_{i-} and M_{i+}, and Q_j between flanking markers M_{j-} and M_{j+}), under the assumption of no QE interaction, can be expressed as follows:

$$y_k = \mu + a_i x_{A_{ik}} + a_j x_{A_{jk}} + aa_{ij} x_{AA_{ijk}} + \sum_f u_{M_{fk}} e_{M_f} + \sum_l u_{MM_{lk}} e_{MM_l} + \varepsilon_k, \quad (1)$$

Table 1 Model coefficients ($x_{A_{ik}}$ or $x_{A_{jk}}$) of QTL additive effects for a DH population

Marker genotype[a]	Expected frequency		$x_{A_{xk}}$
	Q_xQ_x	q_xq_x	
$M_{x-}M_{x+}$	1	0	1
$M_{x-}m_{x+}$	$1-p_x$	p_x	$1-2p_x$
$m_{x-}M_{x+}$	p_x	$1-p_x$	$-(1-2p_x)$
$m_{x-}m_{x+}$	0	1	-1

[a] M_{x-} and M_{x+} ($x=i$ or j) are two flanking marker loci defining the interval. For simplicity, only one of the two identical gamete genotypes is used. The markers and the putative QTL (Q_x) are lined as $M_{x-}\rightarrow Q_x\rightarrow M_{x+}$. $p_x=r_{M_{x-}Q_x}/r_{M_{x-}M_{x+}}\cdot r_{M_{x-}M_{x+}}$ is the recombination fraction between M_{x-} and M_{x+}; $r_{M_{x-}Q_x}$ is the recombination fraction between marker M_{x-} and Q_x. It is assumed that Q_xQ_x has a positive effect on the trait, while q_xq_x has a negative effect

where y_k is the phenotypic value of a quantitative trait measured on the k-th individual (k=1, 2,…, n); μ is the population mean; a_i and a_j are the additive effects (fixed) of the two putative QTLs (Q_i and Q_j), respectively; aa_{ij} is the additive×additive epistatic effect (fixed) between Q_i and Q_j; $x_{A_{ik}}$, $x_{A_{jk}}$ and $x_{AA_{ijk}}$ are coefficients of QTL effects derived according to the observed genotypes of the markers (M_{i-}, M_{i+} and M_{j-}, M_{j+}) and the test positions ($r_{M_{i-}Q_i}$ and $r_{M_{i-}Q_j}$) (Tables 1, 2); $e_{M_f}\sim N(0, \sigma^2_M)$ is the random effect of marker f with indicator coefficient $u_{M_{fk}}$ (1 for M_fM_f and -1 for m_fm_f); $e_{MM_l}\sim N(0, \sigma^2_{MM})$ is the random effect of the l-th marker interaction (between marker K_l and marker L_l) with indicator coefficient $u_{MM_{lk}}$ (1 for $M_KM_KM_LM_L$ or $m_Km_Km_Lm_L$, and -1 for $M_KM_Km_Lm_L$ or $m_Km_KM_LM_L$); and $\varepsilon_k\sim N(0, \sigma^2_\varepsilon)$ the random residual effect.

The inclusion of e_{M_f} and e_{MM_l} in the model is intended to absorb the additive and epistatic effects of background QTLs (additional segregating QTLs other than the loci examined) for controlling the noise caused by these background QTLs.

Model (1) can be written as a matrix form of the mixed linear model

$$\begin{aligned}\mathbf{y} &= \mathbf{Xb} + \mathbf{U}_M\mathbf{e}_M + \mathbf{U}_{MM}\mathbf{e}_{MM} + \mathbf{e}_\varepsilon \\ &= \mathbf{Xb} + \sum_{u=1}^{3}\mathbf{U}_u\mathbf{e}_u \\ &\sim N\left(\mathbf{Xb}, \mathbf{V} = \sum_{u=1}^{3}\sigma^2_u\mathbf{U}_u\mathbf{R}_u\mathbf{U}'_u\right),\end{aligned} \quad (2)$$

where $\mathbf{y}$ is an $n\times1$ vector of phenotypic values; $\mathbf{b}=(\mu, a_i, a_j, aa_{ij})'$ is a 4×1 vector of fixed effects; $\mathbf{e}_1=\mathbf{e}_M\sim N(\mathbf{0},\sigma^2_M\mathbf{R}_M)$ is a random vector of main marker effects; $\mathbf{e}_2=\mathbf{e}_{MM}\sim N(\mathbf{0},\sigma^2_{MM}\mathbf{R}_{MM})$ is a random vector of interaction marker effects; $\mathbf{e}_3=\mathbf{e}_\varepsilon\sim N(\mathbf{0},\sigma^2_\varepsilon\mathbf{I})$ is an $n\times1$ vector of random residuals. $\mathbf{X}$, $\mathbf{U}_1=\mathbf{U}_M$ and $\mathbf{U}_2=\mathbf{U}_{MM}$ are known incidence matrices, respectively. $\mathbf{U}_3=\mathbf{I}$ is an identity matrix. $\mathbf{R}_1=\mathbf{R}_M$ and $\mathbf{R}_{MM}$ are known symmetric matrices of incidence coefficients that can be obtained from the linkage relationships between the main-effect markers and between the pairs of interacting markers, respectively:

$$\mathbf{R}_M=[\rho_{ff'}], \quad (3)$$

where $\rho_{ff'}=1-2r_{ff'}$ is the correlation coefficient between e_{M_f} and $e_{M_{f'}}$. $r_{ff'}$ is the recombination fraction between markers f and f';

$$\mathbf{R}_{MM}=[\rho_{ll'}], \quad (4)$$

where

$$\rho_{ll'} = \rho_{IJ\cdot KL} = \frac{\rho_{AB}\rho_{CD} - \rho_{IJ}\rho_{KL}}{\sqrt{(1-\rho^2_{IJ})(1-\rho^2_{KL})}}, \qquad I<J, K<L \subset (A,B,C,D)$$

is the correlation coefficient between e_{MM_l} (the interaction between markers I and J) and $e_{MM_{l'}}$ (the interaction between markers K and L) under the Haldane mapping function (Haldane 1919). Markers A, B, C and D are ordered as $A…B…C…D$ on the genome, and at least two of them are different. $\mathbf{R}_3=\mathbf{I}$ is an identity matrix.

Table 2 Model coefficients ($x_{AA_{ijk}}$) of epistatic effects for a DH population

Marker genotypes[a]	Expected frequency				$x_{AA_{ijk}}$
	$Q_iQ_iQ_jQ_j$	$q_iq_iq_jq_j$	$Q_iQ_iq_jq_j$	$q_iq_iQ_jQ_j$	
$M_{i-}M_{i+}M_{j-}M_{j+}$	1	0	0	0	$2r_{ij}$
$m_{i-}m_{i+}m_{j-}m_{j+}$	0	1	0	0	$2r_{ij}$
$M_{i-}m_{i+}m_{j-}m_{j+}$	0	p_i	$1-p_i$	0	$2(p_i+r_{ij}-1)$
$m_{i-}M_{i+}M_{j-}M_{j+}$	p_i	0	0	$1-p_i$	$2(p_i+r_{ij}-1)$
$M_{i-}M_{i+}m_{j-}m_{j+}$	0	0	1	0	$-2(1-r_{ij})$
$m_{i-}m_{i+}M_{j-}M_{j+}$	0	0	0	1	$-2(1-r_{ij})$
$M_{i-}M_{i+}M_{j-}m_{j+}$	$1-p_j$	0	p_j	0	$2(r_{ij}-p_j)$
$m_{i-}m_{i+}m_{j-}M_{j+}$	0	$1-p_j$	0	p_j	$2(r_{ij}-p_j)$
$M_{i-}m_{i+}M_{j-}M_{j+}$	$1-p_i$	0	0	p_i	$2(r_{ij}-p_i)$
$m_{i-}M_{i+}m_{j-}m_{j+}$	0	$1-p_i$	p_i	0	$2(r_{ij}-p_i)$
$M_{i-}m_{i+}m_{j-}M_{j+}$	$(1-p_i)p_j$	$p_i(1-p_j)$	$(1-p_i)(1-p_j)$	p_ip_j	$2(\eta_{ij}+r_{ij}-1)$
$m_{i-}M_{i+}M_{j-}m_{j+}$	$p_i(1-p_j)$	$(1-p_i)p_j$	p_ip_j	$(1-p_i)(1-p_j)$	$2(\eta_{ij}+r_{ij}-1)$
$M_{i-}M_{i+}m_{j-}M_{j+}$	p_j	0	$1-p_j$	0	$2(p_j+r_{ij}-1)$
$m_{i-}m_{i+}M_{j-}m_{j+}$	0	p_j	0	$1-p_j$	$2(p_j+r_{ij}-1)$
$M_{i-}m_{i+}M_{j-}m_{j+}$	$(1-p_i)(1-p_j)$	p_ip_j	$(1-p_i)p_j$	$p_i(1-p_j)$	$2(r_{ij}-\eta_{ij})$
$m_{i-}M_{i+}m_{j-}M_{j+}$	p_ip_j	$(1-p_i)(1-p_j)$	$p_i(1-p_j)$	$(1-p_i)p_j$	$2(r_{if}-\eta_{ij})$

[a] Only one of the two identical gamete genotypes is used. The arrangement of markers and QTLs is $M_{i-}\rightarrow Q_i\rightarrow M_{i+}\ldots M_{j-}\rightarrow Q_j\rightarrow M_{j+}$ is the recombination fraction between the two putative QTLs i and j; $\eta_{ij}=p_i+p_j-2p_ip_j$; p_i and p_j are calculated in the same way as in Table 1

Model (1) can be further extended to include QTL×environment interactions for experiments conducted in multiple environments. If a DH population is used for mapping QTLs with additive and additive×additive epistatic effects, as well as their environment interaction effects, the phenotypic value of the k-th DH line in environment h can be expressed as the following mixed linear model (h=1, 2,…, s; k=1, 2,…,n_h):

$$\begin{aligned} y_{hk} &= \mu + a_i x_{A_{ik}} + a_j x_{A_{jk}} + aa_{ij} x_{AA_{ijk}} \\ &\quad + u_{E_{hk}} e_{E_h} + u_{A_iE_{hk}} e_{A_iE_h} + u_{A_jE_{hk}} e_{A_jE_h} + u_{AA_{ij}E_{hk}} e_{AA_{ij}E_h} \\ &\quad + \sum_{f(h)} u_{M_{fk(h)}} e_{M_{f(h)}} + \sum_{l(h)} u_{MM_{lk(h)}} e_{MM_{l(h)}} + \varepsilon_{hk}, \end{aligned} \tag{5}$$

where μ, a_i, a_j, aa_{ij}, $x_{A_{ik}}$, $x_{A_{jk}}$ and $x_{AA_{ijk}}$ have the same meanings as in model (1); e_{E_h} is the random effect of environment h with coefficient $u_{E_{hk}}$; $e_{A_iE_h}$ (or $e_{A_jE_h}$) is the additive×environment interaction effect with coefficient $u_{A_iE_{hk}}$ (or $u_{A_jE_{hk}}$) for Q_i (or Q_j); $e_{AA_{ij}E_h}$ is the epistasis×environment interaction effect with coefficient $u_{AA_{ij}E_{hk}}$; $e_{M_{f(h)}}$ is the effect of marker f nested within the h-th environment with coefficient $u_{M_{fk(h)}}$; $e_{MM_{l(h)}}$ is the effect of marker×marker interaction nested within the h-th environment with coefficient $u_{MM_{lk(h)}}$; and ε_{hk} is the residual effect.

Model (5) can be expressed as the following matrix form:

$$\begin{aligned} \mathbf{y} &= \mathbf{Xb} + \mathbf{U}_E\mathbf{e}_E + \mathbf{U}_{A_iE}\mathbf{e}_{A_iE} + \mathbf{U}_{A_jE}\mathbf{e}_{A_jE} + \mathbf{U}_{AA_{ij}E}\mathbf{e}_{AA_{ij}E} \\ &\quad + \mathbf{U}_M\mathbf{e}_M + \mathbf{U}_{MM}\mathbf{e}_{MM} + \mathbf{e}_\varepsilon \\ &= \mathbf{Xb} + \sum_{u=l}^{7}\mathbf{U}_u\mathbf{e}_u \\ &\sim N\left(\mathbf{Xb}, \mathbf{V} = \sum_{u=1}^{7}\sigma_u^2\mathbf{U}_u\mathbf{R}_u\mathbf{U}_u'\right), \end{aligned} \tag{6}$$

where $\mathbf{y}$ is the vector of phenotypic values; $\mathbf{b}$ is the fixed parameter vector for population mean (μ) and QTL effects (a_i, a_j and aa_{ij}); $\mathbf{X}$ is the known incidence matrix of the fixed parameters; $\mathbf{e}_1=\mathbf{e}_E\sim N(\mathbf{0}, \sigma_E^2\mathbf{I})$ is the vector of random environment effects; $\mathbf{e}_2=\mathbf{e}_{A_iE}\sim N(\mathbf{0}, \sigma_{A_iE}^2\mathbf{I})$ is the vector of random $A_i\times E$ interaction effects; $\mathbf{e}_3=\mathbf{e}_{A_jE}\sim N(\mathbf{0}, \sigma_{A_jE}^2\mathbf{I})$ is the vector of random $A_j\times E$ interaction effects; $\mathbf{e}_4=\mathbf{e}_{AA_{ij}E}\sim N(\mathbf{0}, \sigma^2_{AA_{ij}E}\mathbf{I})$ is the vector of random $AA_{ij}\times E$ interaction effects; $\mathbf{e}_5=\mathbf{e}_M\sim N(\mathbf{0}, \sigma_M^2\mathbf{R}_M)$ is the vector of main marker effects; $\mathbf{e}_6=\mathbf{e}_{MM}\sim N(\mathbf{0}, \sigma_{MM}^2\mathbf{R}_{MM})$ is the vector of interaction marker effects; $\mathbf{e}_7=\mathbf{e}_\varepsilon\sim N(\mathbf{0}, \sigma_\varepsilon^2\mathbf{I})$ is the vector of residual effects; $\mathbf{U}_1-\mathbf{U}6$ are known incidence matrices of the random effects and $\mathbf{U}_7=\mathbf{I}$. $\mathbf{R}_1-\mathbf{R}_4$, and $\mathbf{R}_7$ are identity matrices; $\mathbf{R}_5=\mathbf{R}_M$ and $\mathbf{R}_6=\mathbf{R}_{MM}$ are matrices with symmetric diagonal blocks $\mathbf{R}_{M_h}$ and $\mathbf{R}_{MM_h}$, respectively.

$$\mathbf{R}_M = \begin{bmatrix} \mathbf{R}_{M_1} & & \\ & \ddots & \\ & & \mathbf{R}_{M_s} \end{bmatrix}, \qquad \mathbf{R}_{M_h} = \left[\rho_{f_hf_{h'}}\right], \qquad (h=1,2,\cdots,s) \tag{7}$$

$$\mathbf{R}_{MM} = \begin{bmatrix} \mathbf{R}_{MM_1} & & \\ & \ddots & \\ & & \mathbf{R}_{MM_s} \end{bmatrix}, \qquad \mathbf{R}_{MM_h} = \left[\rho_{l_hl_{h'}}\right], \qquad (h=1,2,\cdots,s) \tag{8}$$

Parameter estimation and hypothesis test

The likelihood function (L) for the parameters of fixed effects $\mathbf{b}$ and variance components $[\sigma_u^2]$ in model (1) or (5) is:

$$L(\mathbf{b},\mathbf{V}) = (2\pi)^{-\frac{n}{2}}|\mathbf{V}|^{-\frac{1}{2}}\exp\left[-\tfrac{1}{2}(\mathbf{y}-\mathbf{Xb})'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{Xb})\right], \tag{9}$$

with the log of the likelihood function (l)

$$l(\mathbf{b},\mathbf{V}) = -\tfrac{n}{2}\ln(2\pi) - \tfrac{1}{2}\ln|\mathbf{V}| - \tfrac{1}{2}(\mathbf{y}-\mathbf{Xb})'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{Xb}) \tag{10}$$

where $\mathbf{V}^{-1}$ is the inverse of $\mathbf{V}$.

When variance components of the model are known, the maximum-likelihood estimates of QTL effects in $\mathbf{b}$ can be obtained by

$$\hat{\mathbf{b}}=(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}.$$

The variance-covariance matrix of $\hat{\mathbf{b}}$ is obtained as

$$\mathrm{Var}(\hat{\mathbf{b}})=(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}. \tag{12}$$

QE interaction effects (additive×environment interaction $\mathbf{e}_{A_iE}$ and $\mathbf{e}_{A_jE}$, epistasis×environment interaction $\mathbf{e}_{AA_{ij}E}$) can be obtained by the best-linear-unbiased-prediction (BLUP) method,

$$\hat{\mathbf{e}}_u=\sigma_u^2\mathbf{U}_u'\mathbf{Q}\mathbf{y}, \tag{13}$$

where $\mathbf{Q}=\mathbf{V}^{-1}-\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$. The variance-covariance matrix of $\hat{\mathbf{e}}_u$ is obtained as

$$\mathrm{Var}(\hat{\mathbf{e}}_u)=\sigma_u^4\mathbf{U}_u'\mathbf{Q}\mathbf{U}_u \tag{14}$$

For searching QTLs (Q_i and Q_j) within intervals defined by (M_{i-}, M_{i+}) and (M_{j-}, M_{j+}) at two prior test positions ($r_{M_{i-}Q_i}$ and $r_{M_{j-}Q_j}$), the null hypothesis for genetic parameters (QTL main effects and QE interaction effects) can be tested by the likelihood-ratio statistic (LR),

$$LR=2l_1(\hat{\mathbf{b}}_1, \mathbf{V}_1)-2l_0(\hat{\mathbf{b}}_0, \mathbf{V}_0). \tag{15}$$

where the variance components in $\mathbf{V}$ can usually be replaced by their unbiased estimates,

$$\hat{\mathbf{V}} = \sum_u \hat{\sigma}_u^2\mathbf{U}_u\mathbf{R}_u\mathbf{U}_u', \tag{16}$$

For model (1) without QE interaction, we can test QTL effects by setting H_0: $a_i=a_j=aa_{ij}=0$. In this case, LR has an approximate χ^2 distribution with df=3. This hypothesis is equivalent to testing no QTLs in the two intervals at the test positions. The alternative hy-

pothesis, H_1, is that not all of the QTL main effects are equal to zero. For model (5) with QE interaction, we can test both QTL main effects and QE interaction effects by setting H_0: $a_i=a_j=aa_{ij}=0$ and $\sigma^2_{A_iE}=\sigma^2_{A_jE}=\sigma^2_{A_{ij}E}=0$. This hypothesis is equivalent to testing no QTLs in the two intervals across all the environments involved. The alternative hypothesis, H_1, is that not all of the QTL main effects and QE interaction effects are equal to zero. Therefore, the *LR* has an approximate χ^2 distribution with *df*=6. Additional hypothesis tests for other combinations of the genetic parameters can also be conducted.

The rejection of H_0 indicates that at least one of the QTL effects is not equal to zero. Then the estimates of QTL main effects can be obtained by equation (11). An hypothesis test for the deviation of QTL main effects from zero can be conducted by a *t*-test:

$$t_{cal}=\hat{b}_i/\mathrm{SE}(\hat{b}_i) \text{ with degrees of freedom } df=n-\text{rank }(\mathbf{X}), \quad (17)$$

where $\hat{b}_i$ is the estimate of a specific QTL main effect (a_i, a_j, or aa_{ij}) to be tested; $\mathrm{SE}(\hat{b}_i)=\sqrt{[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}]_{ii}}$ is the standard error for $\hat{b}_i$, rank (**X**)=4 for models (1) and (5).

An hypothesis test for deviation of QE interaction effects from zero can be conducted by a *z*-test based on the standard normal distribution:

$$z_{cal}=[\hat{\mathbf{e}}_u]_h/\mathrm{SE}([\hat{\mathbf{e}}_u]_h), \quad (18)$$

where $[\hat{\mathbf{e}}_u]_h$ is the predicted QE interaction effect ($e_{A_iE_h}$, $e_{A_jE_h}$, or $e_{AA^{ij}E_h}$) to be tested; $\sigma^2_u\sqrt{[\mathbf{U}'_u\mathbf{Q}\mathbf{U}_u]_{hh}}$ is the standard error for $[\hat{\mathbf{e}}_u]_h$.

Estimates of putative QTL positions can be obtained based on the above significance tests at peak points of the statistics (*LR* and/or *t*) along chromosomes. For a significant aa_{ij} and/or $\sigma^2_{AA_{ij}E}$, the two prior test positions ($r_{M_i\text{-}Q_i}$ and $r_{M_j\text{-}Q_j}$) are taken as the estimates of the positions of Q_i and Q_j. When a_i (or a_j) and/or $\sigma^2_{A_iE}$ (or $\sigma^2_{A_jE}$) are significant, $r_{M_i\text{-}Q_i}$ (or $r_{M_j\text{-}Q_j}$)) is taken as the estimate of the position of Q_i (or Q_j).

Results

Simulation results for mapping QTLs in one environment

Monte Carlo simulations for a DH population under one environment were conducted to study the properties of the mixed-model approaches described above. In all simulations, we employed three genomes sharing the same marker linkage map with four chromosomes and a total of 64 evenly distributed (10 cM between two adjacent markers) markers. Each of the genomes has four QTLs with a different linkage relationship among the QTLs. In genome A, all QTLs are independently inherited at 7.8, 123.7, 78.7, and 99.4 cM on chromosomes I–IV (from the left ends of the chromosomes), respectively. In genome B, two QTLs are independent on chromosome I (7.8 cM) and chromosome III (78.7 cM), while the other two are loosely linked on chromosome II at 55.3 and 123.7 cM, respectively. In genome C, all four QTLs are linked on chromosome III at 10.2, 48.7, 80.8, and 120.4 cM, respectively. In all genomes, the QTLs, according to their locations, were numbered as 1, 2, 3, and 4 from chromosomes I to IV and/or from left to right on the chromosomes (Tables 3, 4). The additive effects of the QTLs range from –1.21 to 4.07 units, and the digenic additive×additive epistatic effects from –1.71 to 1.31 units (Table 4).

For generating each data set based on the above genomic information, the expected genetic variance V_G was calculated first:

$$V_G=\sum_f\sum_{f'}\rho_{ff'}a_fa_{f'}+\sum_{i<}\sum_j\sum_{i'<}\sum_{j'}\omega_{ij\cdot i'j'}aa_{ij}aa_{i'j'},$$
$$f,f'=1,\cdots 4;i,i'=1,2,3;j,j'=2,3,4),$$

Table 3 Simulation results for studying the properties of the mixed-model approaches in determining QTL positions

Genome	QTL *i*	QTL *j*	Site *i* (cM)			Site *j* (cM)		
			Para [a]	Est. (SE) [b]		**Para**	*Est. (SE)*	
A	1	2	**7.8**	8.3	(0.35)	**123.7**	123.6	(0.33)
	1	3	**7.8**	8.2	(0.45)	**78.7**	78.5	(0.26)
	1	4	**7.8**	5.4	(1.40)	**99.4**	105.1	(2.43)
	2	3	**123.7**	123.2	(0.58)	**78.7**	78.4	(1.13)
	2	4	**123.7**	123.6	(0.20)	**99.4**	100.8	(0.17)
	3	4	**78.7**	77.6	(0.22)	**99.4**	101.7	(1.03)
B	1	2	**7.8**	7.7	(0.38)	**55.3**	55.1	(0.54)
	1	3	**7.8**	7.7	(0.31)	**123.7**	123.4	(0.13)
	1	4	**7.8**	4.8	(1.70)	**78.7**	80.1	(2.11)
	2	3	**55.3**	60.9	(0.97)	**123.7**	121.7	(0.95)
	2	4	**55.3**	54.7	(0.36)	**78.7**	78.6	(0.30)
	3	4	**123.7**	122.2	(0.36)	**78.7**	79.6	(1.37)
C	1	2	**10.2**	9.3	(0.37)	**48.7**	50.4	(0.48)
	1	3	**10.2**	8.7	(0.52)	**80.8**	79.1	(0.35)
	1	4	**10.2**	13.9	(3.03)	**120.4**	126.2	(2.85)
	2	3	**48.7**	44.9	(3.51)	**80.8**	77.4	(1.97)
	2	4	**48.7**	50.2	(0.36)	**120.4**	121.6	(0.30)
	3	4	**80.8**	79.6	(0.57)	**120.4**	119.1	(1.04)

[a] Parameter of QTL positions on the genome

[b] Estimated means and standard errors for positions of QTLs involved in epistasis detected in 300 simulations. The estimates of QTL positions were obtained within the pair-wise marker intervals each having at least one marker flanking the QTLs. Sample size=200; heritability=0.50

Table 4 Simulation results for studying the properties of the mixed model approach in the dissection of QTL effects

Genome	QTL *i*	QTL *j*	a_i[a]				a_j[b]				aa_{ij}[c]			
			Para[d]	Est.	(*SE*)[e]	Power[f]	**Para**	Est.	(*SE*)	Power	**Para**	Est.	(*SE*)	Power
A	1	2	**0.60**	0.59	(0.021)	0.09	**−1.21**	−1.22	(0.022)	0.64	**1.31**	1.34	(0.023)	0.73
	1	3	**0.60**	0.59	(0.021)	0.09	**4.07**	4.03	(0.023)	1.00	**1.08**	1.10	(0.023)	0.48
	1	4	**0.60**	0.59	(0.021)	0.11	**0.00**	0.02	(0.021)	0.00	**−0.38**	−0.39	(0.023)	0.03
	2	3	**−1.21**	−1.22	(0.022)	0.64	**4.07**	4.03	(0.023)	1.00	**0.00**	0.03	(0.022)	0.00
	2	4	**−1.21**	−1.22	(0.022)	0.64	**0.00**	0.01	(0.021)	0.00	**−1.71**	−1.68	(0.021)	0.94
	3	4	**4.07**	4.03	(0.023)	1.00	**0.00**	0.01	(0.021)	0.00	**−0.41**	−0.40	(0.022)	0.03
B	1	2	**0.60**	0.58	(0.022)	0.10	**−1.21**	−1.15	(0.024)	0.49	**1.31**	1.34	(0.024)	0.68
	1	3	**0.60**	0.57	(0.022)	0.08	**4.07**	3.97	(0.022)	1.00	**1.08**	1.25	(0.024)	0.62
	1	4	**0.60**	0.57	(0.022)	0.08	**0.00**	−0.01	(0.022)	0.00	**−0.38**	−0.44	(0.023)	0.05
	2	3	**−1.21**	−1.17	(0.024)	0.54	**4.07**	4.07	(0.023)	1.00	**0.00**	0.01	(0.024)	0.01
	2	4	**−1.21**	−1.15	(0.024)	0.52	**0.00**	−0.01	(0.021)	0.00	**−1.71**	−1.74	(0.022)	0.92
	3	4	**4.07**	3.95	(0.022)	1.00	**0.00**	−0.01	(0.022)	0.00	**−0.41**	−0.59	(0.024)	0.10
C	1	2	**0.60**	0.60	(0.025)	0.08	**−1.21**	−1.23	(0.032)	0.37	**1.31**	1.48	(0.029)	0.53
	1	3	**0.60**	0.53	(0.025)	0.08	**4.07**	3.97	(0.026)	1.00	**1.08**	1.17	(0.027)	0.37
	1	4	**0.60**	0.58	(0.025)	0.08	**0.00**	0.39	(0.024)	0.03	**−0.38**	−0.42	(0.028)	0.04
	2	3	**−1.21**	−1.11	(0.029)	0.37	**4.07**	4.10	(0.026)	1.00	**0.00**	−0.05	(0.030)	0.01
	2	4	**−1.21**	−1.17	(0.031)	0.35	**0.00**	0.40	(0.024)	0.03	**−1.71**	−1.91	(0.026)	0.93
	3	4	**4.07**	3.94	(0.028)	1.00	**0.00**	0.00	(0.025)	0.01	**−0.41**	−0.77	(0.031)	0.14

[a] Additive effect of QTL *i*
[b] additive effect of QTL *j*
[c] additive×additive epistatic effect
[d] parameter of QTL effects
[e] estimated means and standard errors of QTL effects across all 300 simulations
[f] power of detection (by *t* test) in 300 simulations. Test positions were set exactly at the positions of QTLs. Sample size=200; heritability=0.50

where $\rho_{ff'}=1-2r_{ff'}$, $\omega_{ij\cdot i'j'}=\rho_{ab}\rho_{cd}-\rho_{ij}\rho_{i'j'}$, [*a…b…c…d* are ordered as *a…b…c…d* on the genome; *i, i′, j, j′*⊂(*a, b, c, d*)] under the Haldane mapping function (Haldane 1919). r_{ij} is the recombination fraction between Q_i and Q_j. Then, the random error variance is calculated based on V_G and a given heritability h^2:

$$V_\varepsilon=V_G(1-h^2)/h^2.$$

The genotypic value G_k of the *k*-th individual was obtained by summing all genetic effects (Mather and Jinks 1982). The phenotypic value of the *k*-th individual was calculated as $G_k+\varepsilon_k$, where ε_k was obtained by generating a pseudo-random normal deviate with zero mean and variance V_ε.

We compared the power and efficiencies of two-way ANOVA and stepwise regression in the selection of the markers. It was shown that stepwise regression was much more powerful than the ANOVA method, not only in the selection of individual markers as indicated by Hackett (1994), but also in the selection of interaction markers (data not shown). Moreover, stepwise regression is also a relatively quick method and is suitable for the analyses of large genomes and large data sets. Therefore, stepwise regression was used for selecting both main-effect and interaction markers in the model. The selection of interaction markers by stepwise regression was conducted after main-effect markers had been selected. The probability for entering and dropping markers (or marker pairs) was 0.005. Restricted maximum-likelihood estimation (REML) (Patterson and Thompson 1971, 1974) was used to obtain unbiased estimates of variance components for replacing parameters in equation (16). The methods of two-dimensional searches for digenic epistatic QTLs are described in the following relevant sections, separately.

Simulations for determining QTL positions

To test the unbiasedness and accuracy of QTL positions determined by the mixed model-approaches, we conducted a genome-wide search for QTLs with digenic epistatic effects. The test positions were set in pairwise marker intervals that were separated by at least one other marker interval; the distance between adjacent test positions (or walking speed) in each marker interval is 2.0 cM. The results from 300 simulations are summarized in Table 3.

It was shown that the mixed-model approaches provided basically unbiased and accurate estimates for the positions of QTLs with relatively large additive and/or epistatic effects (Table 3). Considerable biases and standard errors in estimated QTL positions were associated with QTLs of small additive and epistatic effects (e.g. QTL #1 and QTL #4), which were almost undetectable (Table 4) under the threshold probability (α=0.005, equivalent to LOD=2.79 here). Significant epistasis could improve the unbiasedness and accuracy of the estimated positions of both QTLs involved. For instance, the SE of the estimated position of QTL #1 was 2.4 cM when it was involved in a non-significant epistasis with QTL #4 in genome A. However, this deviation became

Table 5 Simulation results for studying the properties of the mixed model approaches in the prediction of QE interaction effects. E_1, E_2, and E_3 are three environments used in the simulations. $\hat{e}_{A_iE_h}(\hat{e}_{A_jE_h})$ and $\hat{e}_{AA_{ij}E_h}$ are averages of additive×environment interactions and epistasis×environment interaction effects predicted by the BLUP method across 100 simulations. The bold numbers are parameters for the corresponding predicted values; the sets of parameters were used in 100 simulations

Genome	QTL i	QTL j	E_1						E_2						E_3					
			$e_{A_iE_1}$	$\hat{e}_{A_iE_1}$	$e_{A_jE_1}$	$\hat{e}_{A_jE_1}$	$e_{AA_{ij}E_1}$	$\hat{e}_{AA_{ij}E_1}$	$e_{A_iE_2}$	$\hat{e}_{A_iE_2}$	$e_{A_jE_2}$	$\hat{e}_{A_jE_2}$	$e_{AA_{ij}E_2}$	$\hat{e}_{AA_{ij}E_2}$	$e_{A_iE_2}$	$\hat{e}_{A_iE_3}$	$e_{A_jE_3}$	$\hat{e}_{A_jE_3}$	$e_{AA_{ij}E_3}$	$\hat{e}_{AA_{ij}E_3}$
A	1	2	**–2.73**	–2.75	**3.42**	3.42	**–3.03**	–3.17	**0.92**	0.85	**3.03**	3.19	**2.55**	2.76	**1.81**	1.90	**–6.45**	–6.60	**0.48**	0.42
	1	3	**–2.73**	–2.75	**1.48**	1.24	**–2.94**	–2.87	**0.92**	0.86	**–1.24**	–1.01	**0.82**	0.90	**1.81**	1.89	**–0.24**	–0.23	**2.12**	1.97
	1	4	**–2.73**	–2.62	**2.00**	1.88	**0.82**	0.56	**0.92**	0.82	**–0.51**	–0.40	**–1.31**	–1.09	**1.81**	1.79	**–1.50**	–1.48	**0.50**	0.53
	2	3	**3.42**	3.42	**1.48**	1.37	**–0.12**	–0.09	**3.03**	3.18	**–1.24**	–1.02	**2.52**	2.40	**–6.45**	–6.60	**–0.24**	–0.35	**–2.40**	–2.31
	2	4	**3.42**	3.37	**2.00**	1.87	**0.41**	0.33	**3.03**	3.19	**–0.51**	–0.39	**–2.04**	–1.98	**–6.45**	–6.56	**–1.50**	–1.48	**1.64**	1.65
	3	4	**1.48**	1.24	**2.00**	1.82	**2.29**	2.06	**–1.24**	–1.03	**–0.51**	–0.38	**–5.86**	–5.63	**–0.24**	–0.21	**–1.50**	–1.44	**3.57**	3.57
B	1	2	**–2.73**	–2.71	**3.42**	3.12	**–3.03**	–3.01	**0.92**	0.82	**3.03**	3.09	**2.55**	2.68	**1.81**	1.89	**–6.45**	–6.21	**0.48**	0.33
	1	3	**–2.73**	–2.81	**1.48**	1.51	**–2.94**	–3.24	**0.92**	0.79	**–1.24**	–1.02	**0.82**	1.02	**1.81**	2.02	**–0.24**	–0.48	**2.12**	2.22
	1	4	**–2.73**	–2.74	**2.00**	1.91	**0.82**	0.77	**0.92**	0.77	**–0.51**	–0.45	**–1.31**	–1.24	**1.81**	1.97	**–1.50**	–1.46	**0.50**	0.47
	2	3	**3.42**	3.19	**1.48**	1.51	**–0.12**	–0.23	**3.03**	3.11	**–1.24**	–0.99	**2.52**	2.45	**–6.45**	–6.30	**–0.24**	–0.52	**–2.40**	–2.22
	2	4	**3.42**	3.18	**2.00**	1.90	**0.41**	0.13	**3.03**	3.05	**–0.51**	–0.46	**–2.04**	–1.82	**–6.45**	–6.24	**–1.50**	–1.44	**1.64**	1.69
	3	4	**1.48**	1.42	**2.00**	1.93	**2.29**	2.43	**–1.24**	–0.98	**–0.51**	–0.47	**–5.86**	–5.98	**–0.24**	–0.44	**–1.50**	–1.46	**3.57**	3.55
C	1	2	**–2.73**	–2.69	**3.42**	3.49	**–3.03**	–2.96	**0.92**	0.85	**3.03**	3.09	**2.55**	2.08	**1.81**	1.84	**–6.45**	–6.59	**0.48**	0.88
	1	3	**–2.73**	–2.70	**1.48**	1.71	**–2.94**	–3.13	**0.92**	0.87	**–1.24**	–1.18	**0.82**	1.21	**1.81**	1.83	**–0.24**	–0.53	**2.12**	1.92
	1	4	**–2.73**	–2.64	**2.00**	1.98	**0.82**	0.67	**0.92**	0.85	**–0.51**	–0.68	**–1.31**	–1.35	**1.81**	1.80	**–1.50**	–1.30	**0.50**	0.68
	2	3	**3.42**	3.35	**1.48**	1.69	**–0.12**	–0.23	**3.03**	3.10	**–1.24**	–1.37	**2.52**	2.21	**–6.45**	–6.45	**–0.24**	–0.32	**–2.40**	–1.98
	2	4	**3.42**	3.39	**2.00**	1.99	**0.41**	0.46	**3.03**	3.10	**–0.51**	–0.65	**–2.04**	–1.83	**–6.45**	–6.49	**–1.50**	–1.34	**1.64**	1.37
	3	4	**1.48**	1.37	**2.00**	1.91	**2.29**	2.52	**–1.24**	–0.91	**–0.51**	–0.55	**–5.86**	–6.34	**–0.24**	–0.47	**–1.50**	–1.36	**3.57**	3.82

less than 0.5 cM when QTL #1 was involved in significant epistasis with QTL #2 or QTL #3.

Simulations for testing QTL effects

Two key properties, the power of QTL detection and the unbiasedness and accuracy of estimating QTL effects (both additive and epistatic effects), of the mixed-model approaches were also investigated by using computer simulations. The power of detecting QTL effects was obtained based on *t* tests after the null hypothesis (H_0: $a_i=a_j=aa_{ij}=0$) was rejected by *LR* tests at $\alpha=0.005$. The results from 300 simulations were summarized in Table 4.

Under the conditions of no linkage (genome A) or loose linkage (genome B), the mixed-model approaches produced unbiased and accurate (SE≤0.03) estimates of the additive and epistatic effects for the preset QTLs. When there are close linkages between QTLs (genome C), the estimated QTL effects were only slightly biased with increased SE.

Table 4 also indicated that the mixed model approaches were powerful in detecting QTLs with relatively large additive and/or epistatic effects (coefficient of determination R^2>3%). For instance, QTL #2 with an additive effect of $R^2\approx$3~4% were detected 35~64% of the times. QTL #3 having the largest additive effect of $R^2\approx$33% was always detected in the three genomes. The epistatic effect between QTL #2 and QTL #4 was ~6% in R^2, and was detected 92% of the time. Expectedly, small additive or epistatic effects (R^2<1%) were largely undetectable (e.g. additive effects of QTLs #1 and #4; epistasis effects between QTL #2 and QTL #3).

Linkage, particularly close linkage between QTLs (genome C), had a significant impact on the detection power in QTL mapping using the mixed-model approaches. The nature of this impact depends on the directions of QTL main effects (including epistatic effects) of linked QTLs. When effects of the linked QTLs are in coupling (in the same direction), the QTLs tended to be detected with increased power. However, when in repulsion (having effects of opposite direction), the QTLs tended to be detected with decreased power.

Simulations results for mapping QTLs in multiple environments

Monte Carlo simulations were also conducted for studying the properties of the mixed-model approaches in the estimation of QTL main effects (additive effects and epistatic effects) and in the prediction of QE interaction effects. The information for the three genomes (A, B, C) was still used in the simulations for three environments (s=3). The phenotypic values of individuals within environment h included an additional environment effect (E_h) and QE interactions ($e_{A_iE_h}$, $e_{A_jE_h}$ and $e_{AA_{ij}E_h}$). One-hundred simulations were performed for each of the genomes in a sample of 200 DH individuals and a total of 600 phenotypic observations at the preset positions of the QTLs. The simulation results indicated that the mixed-model approaches could still provide unbiased estimates of QTL effects (additive effects and epistatic effects), as well as unbiased predicted values for additive×environment interactions ($e_{A_iE_h}$ and $e_{A_jE_h}$) and epistasis×environment interactions ($e_{AA_{ij}E_h}$) (in Table 5). In the presence of QE interaction, however, estimated QTL effects based on models without QE interaction tended to result in biased estimates of QTL parameters.

Discussion

Epistasis, an important genetic component underlying quantitative trait variation, has not been well characterized due largely to lack of an appropriate methodology (Li 1997). One major difficulty in developing a powerful statistical approach for mapping QTLs with epistatic effects is the treatment of many parameters for multiple QTLs involved in the statistical model. In the present study, we have developed the mixed-model approaches and the corresponding computer software for simultaneous interval mapping of QTLs with both additive and epistatic effects as well as QTL×environment interactions. Several important properties of the methods will be discussed below.

Background genetic variation (BGV) control

To-date, two-dimensional searches for digenic epistasis using two-way ANOVA or regression models have been the most common way for detecting epistatic QTLs (Haley and Knott 1992; Li et al. 1997a, b). However, this simple method of detecting epistasis without BGV control may suffer a high probability of false positives (Li 1997; Li et al. 1997a, b).

Marker cofactors have been successfully used for controlling the influence of BGV in mapping QTLs excluding epistasis (Jansen 1992, 1993; Zeng 1993, 1994; Jansen and Stam 1994). This is because the inclusion of main-effect markers in the models can absorb a considerable portion of the BGV through their linkages to the undetected main-effect QTLs segregating in the population. In the presence of epistasis, however, control of only main-effect markers is insufficient, because the epistatic effects of QTLs will also show influences, particularly for complex phenotypes (Li et al. 1997a, b). Thus, inclusion of interaction markers closely linked to epistatic QTLs in the statistical models is expected to improve the power and accuracy of QTL mapping, which has been proven to be the case by our simulations. Our results indicated that without BGV control, accuracy, precision and power in mapping QTLs were significantly reduced, and that using both main-effect markers and interaction markers [Model (1)] was much more superior than using main-effect markers alone in BGV control. BGV control by this way will not cause over-parameterization since

the mixed-model approaches could shrink the parameters of marker cofactors to be estimated-only two additional variance components (σ^2_M and σ^2_{MM}) need to be estimated for marker cofactors in model (1). In the simulation and software developed, we only used markers selected by statistical analysis (e.g. stepwise regression) as cofactors for BGV control, which is slightly different from other methods. This is because only the markers closely linked to QTLs are important sources of BGV. Irrelevant markers contributed little to BGV but could reduce the precision in estimating residual errors. Selection of important markers, therefore, is the key for effective control of BGV. It should be pointed out that stepwise regression analysis may select different sets of markers when different entering and dropping probabilities are used. In real-data analysis, less stringent probabilities tend to produce a larger set of selected markers, and vice versa. Appropriate probabilities should be decided based on the effective population size, trait heritabilities, and error control in the experiments.

As a convention, most current methods for QTL analysis treat marker effects as fixed in the statistical models. However, some others also take marker effects as random (e.g. Fernando and Grossman 1989; Goddard 1992). In the mixed-model approaches for QTL mapping, marker effects are assumed to be random. There are two major reasons why marker effects can be considered random. First, there are almost unlimited markers in the whole genomic region and the markers observed are only a subset of all possible markers. Second, markers are usually used as a tool, but not the targets, in QTL detection. When compared with the methods of interval mapping (Lander and Botstein 1989) and composite interval mapping (Zeng 1994) for mapping QTLs without epistasis using simulation (data not shown), all three methods produced virtually identical results under no BGV control. For detecting multiple genetic main effects of QTLs, the mixed-model approach yielded almost the same results, with similar power and precision, as composite interval mapping with the same sets of marker cofactors.

Power of detection, relative magnitudes of genetic effects, and significance threshold

The simulation results indicated that the power of the mixed model approach in detecting QTLs and epistasis is highly correlated with the relative magnitudes (R^2) of the QTL main/epistatic effects under a given threshold. In other words, QTLs with large additive and/or epistatic effects (R^2 >6% in the simulations) can almost always be detected, and their positions and effects are also accurately estimated. In contrast, QTLs with small additive/epistatic effects (R^2<2% in the simulations) are largely undetectable. This is consistent with the empirical results of many QTL mapping efforts that tend to find QTLs with an R^2 of 5% as having the minimum variance detectable. The relative magnitudes (R^2) of genetic effects depend on the magnitudes of the effects themselves, as well as variation caused by the experimental errors, which is specific for individual traits and experiments. This implies that the power and precision in QTL mapping should be greatly improved by reducing the experimental errors, which can be achieved in a number of ways such as increasing population size and replicating measurements of traits, etc.

It is noted that there will be several different estimates for the position/additive effect of a QTL when it is involved in more than one epistasis. An average of these estimates weighted by detection powers should represent the appropriate QTL position/additive effect. But in real-data analysis, the arithmetic mean will have to be used since the detection powers are not available.

The *LR* threshold of *P*=0.005 (equivalent to LOD=2.79 for *df*=3) used in our simulations is considered as an typical one for most QTL mapping studies, but it gave us a consistent high power in detecting QTLs of moderate additive/epistatic effects (R^2~5%). However, increasing the threshold to 0.001 in our simulations resulted in a significantly reduced power of detecting QTLs with moderate additive/epistatic effects. In analyzing the data of five mapping populations with around 200 individuals (two BC, one RI and two test populations) using the mixed model-method, a threshold between 0.005 and 0.001 (depending on traits) was found to produce consistent results (Li et al., unpublished data).

Detection of QTL × environment interactions

The mixed linear model (1) developed in the present study provides a basic framework, which can be easily extended to cover more complex experimental designs of QTL mapping by the inclusion of additional factors (such as environments, QE interactions, and/or randomized complete blocks). Model (5) represents a direct extension of model (1) to include QE interactions ($e_{A_iE_h}$, and $e_{A_iE_h}$ and $e_{AA_{ij}E_h}$). Monte Carlo simulations showed that model (5) maintained the most important property of model (1), the unbiasedness of estimated QTL main effects (additive and epistatic effects) and predicted QE interaction effects. However, the *z*-tests for predicted QE interaction effects using the corresponding variance components ($\sigma^2_{A_iE}$, $\sigma^2_{A_jE}$ and $\sigma^2_{AA_{ij}E}$) indicated a very low power in detecting significant QE interaction effects. Alternatively, use of re-sampling techniques (e.g. Jackknife) to obtain the sampling variances for *t*-tests of QE parameters, or to estimate the variance components of QE interactions by REML, may also be helpful in quantifying specific additive and/or epistatic effects for QTLs in different environments, which remains to be examined.

Environmental effects (e.g. soil types, day length, and general temperature regimes, etc.) could also be taken as fixed effects in practice. When environments are fixed in the mixed linear model, QE interactions would also be fixed. In this case, however, QTL effects (additive and

epistatic effects), environmental effects and QE interactions in vector **b**, might be confounded and unestimable due to the singularity of the **X** matrix. But the estimable linear function (**c′b**) could be obtained for a general linear test of the QTL effects and QE interactions. This needs further investigation.

Development of computer software

Computing time has not been a factor for mapping QTLs without epistatic effects since a one-dimensional genome-wide search for several traits is within the current capacity of most computers. However, for mapping QTLs with epistatic effects and QE interactions, the computing time for a genome-wide search could take weeks. Very often, reliable mapping results may take several runs of computation in real-data analysis. Therefore, efficient statistical algorithms and mapping strategies are important factors in developing the software. Based on these considerations and computer simulation, we have developed a user-friendly 32-bit PC program (QTLMapper Version 1.0) based on the mixed-model approaches described in this paper. This software treats missing and dominant marker data in the algorithm suggested by Jiang and Zeng (1997). It can analyze data sets from DH and RI (recombinant inbred lines) populations, and has preliminarily been proven to be efficient by simulations and data analyses for a few real QTL mapping studies. This software is freely available to users from any of the co-authors.

Acknowledgements We are grateful to Drs. P. Wu, M. D. Burow, R. G. Ming, Y. Si, X. Draye, C. X. Jiang and two anonymous referees for their helpful comments on the manuscript, and to Dr. L. J. Luo and Mr. H. W. Mei for providing rice data for testing the related computer program. This research was funded jointly by a grant from the Natural Science Foundation of China to J. Zhu, and by a grant from the Rockefeller Foundation to Z. K. Li/A. H. Paterson. D. L. Wang was also supported by a Dissertation Fellowship from the Rockefeller Foundation.

References

Allard RW (1988) Future directions in plant population genetics, evolution and breeding. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) Plant population genetics and germplasm resources. Sinauer Associates Inc, Sunderland, Massachusetts, pp 1–19

Bradshaw HD, Stettler RF (1995) Molecular genetics of growth and development in *Populus*. IV. Mapping QTLs with large effects on growth, form, and phenology traits in a forest tree. Genetics 139:963–973

Chase K, Adler FR, Lark KG (1997) EPISTAT: a computer program for identifying and testing interactions between pairs of quantitative trait loci. Theor Appl Genet 94:724–730

Cockerham CC, Zeng ZB (1996) Design III with marker loci. Genetics 143:1437–1456

Doebley J, Stec A, Gustus C (1995) *teosinte branched l* and the origin of maize: evidence for epistasis and the evolution of dominance. Genetics 141:333–346

Falconer DS (1981) Introduction to quantitative genetics, 2nd edn. Longman Press, New York

Fernando RL, Grossman M (1989) Marker-assisted selection using best linear unbiased prediction. Genet Selec Evol 21:467–477

Fijneman RJA, Devries SS, Jansen RC, Demant P (1996) Complex interactions of new quantitative trait loci, *sluc1*, *sluc2*, *sluc3* and *sluc4*, that influence the susceptibility to lung-cancer in the mouse. Nature Genet 14:465–467

Fu YB, Ritland K (1996) Marker-based inferences about epistasis for genes influencing inbreeding depression. Genetics 144: 339–348

Goddard ME (1992) A mixed model for analyses of data on multiple genetic markers. Theor Appl Genet 83:878–886

Hackett CA (1994) Selection of markers linked to quantitative trait loci by regression techniques. In: Van Ooijen JW, Jansen J (eds) Biometrics in plant breeding: applications of molecular markers. Proc 9th EUCARPIA Meeting. Section Biometrics in Plant Breeding. Wageningen, The Netherlands, pp 99–106

Haldane JBS (1919) The combination of linkage values and the calculation of distance between the loci of linked factors. J Genet 8:299–309

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315–324

Holland JB (1998) EPISTACY: A SAS program for detecting two-locus epistatic interactions using genetic marker information. J Hered 89:374–375

Jansen RC (1992) A general mixture model for mapping quantitative trait loci by using molecular markers. Theor Appl Genet 85:252–260

Jansen RC (1993) Interval mapping of multiple quantitative trait loci. Genetics 135:205–211

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

Jansen RC, Van Ooijen JW, Stam P, Lister C, Dean C (1995) Genotype by environment interaction in genetic mapping of multiple quantitative trait loci. Theor Appl Genet 91:33–37

Jiang CJ, Zeng ZB (1995) Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics 140:1111–1127

Jiang CJ, Zeng ZB (1997) Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. Genetica 101:47–58

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Lark KG, Chase K, Adler FR, Mansur LM, Orf JJ (1995) Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. Proc Natl Acad Sci USA 92:4656–4660

Li ZK (1997) Molecular analysis of epistasis. In: Paterson AH (ed) Molecular dissection of complex traits. CRC Press, Boca Raton, Florida, pp 119–130

Li ZK, Pinson SRM, Park WD, Paterson AH, Stansel JW (1997a) Epistasis for three grain yield components in rice (*Oryza sativa* L.). Genetics 145:453–465

Li ZK, Pinson SRM, Paterson AH, Park WD, Stansel JW (1997b) Genetics of hybrid sterility and hybrid breakdown in an intersubspecific rice (*Oryza sativa* L.) population. Genetics 145: 1139–1148

Mather K, Jinks JL (1982) Biometrical genetics, 3rd edn. Chapman and Hall, London

Meuwissen THE, Goddard ME (1997) Estimation of effects of quantitative trait loci in large complex pedigrees. Genetics 146:409–416

Moreno-Gonzalez J (1992) Genetic models to estimate additive and non-additive effects of marker-associated QTLs using multiple regression techniques. Theor Appl Genet 85:435–444

Neale DB, Sederoff RR (1996) Genome mapping in gymnosperms: a case study in loblolly pine (*Pinus taeda* L.). In: Paterson AH, Landes RG (eds) Genome mapping in plants. Academic Press, Georgetown, Texas, pp 308–319

Paterson AH, Damon S, Hewitt JD, Zamir D, Rabinowitch HD, Lincoln SE et al. (1988) Resolution of quantitative traits into Mendelian factors by using a complete map of restriction fragment length polymorphisms. Nature 335:721–726

Paterson AH, Deverna JW, Lanini B, Tanksley SD (1991) Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. Genetics 127:181–197

Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. Biometrika 58:545–554

Patterson HD, Thompson R (1974) Maximum likelihood estimation of components of variance. Proc 8th Int Biometric Conf. Biometric Society, Washington, D.C

Pooni HS, Coombs DJ, Jinks PS (1987) Detection of epistatsis and linkage of interacting genes in the presence of reciprocal differences. Heredity 58:257–266

Routman EJ, Cheverud JM (1997) Gene effects on a quantitative trait: two-locus epistatic effect measured at microsatellite markers and at estimated QTLs. Evolution 51:1654–1662

Sari-Gorla M, Calinski T, Kaczmarek Z, Krajewski P (1997) Detection of QTL×environment interaction in maize by a least squares interval mapping method. Heredity 78:146–157

Spickett SG, Thoday JM (1966) Regular response to selection. 3. Interaction between located polygenes. Genet Res **7**:96–121

Stuber CW, Edwards MD, Wendel JF (1987) Molecular marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics 116:113–248

Tanksley SD, Hewitt JD (1996) Use of molecular markers in breeding for soluble solids in tomato – a re-examination. Theor Appl Genet 75:811–823

Tinker NA, Mather DE (1995) Methods for QTL analysis with progeny replicated in multiple environments. JQTL (http://probe.gig.usda.gov:8000/otherdocs/jqtl/) 1 (1)

Utz HF, Melchinger AE (1996) PLABQTL: a program for composite interval mapping of QTLs. JQTL (http://probe.gig.usda.gov:8000/otherdocs/jqtl/) 2 (1)

Weller JI (1986) Maximum-likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. Biometrics 42:627–640

Wu P, Zhang G, Huang N, Ladha JK (1995) Non-additive interaction conditioning spikelet sterility in an F_2 population of an *indica*/*japonica* cross in rice. Theor Appl Genet 91:825–829

Xiao J, Li J, Yuan L, Tanksley SD (1995) Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. Genetics 140:745–754

Yan JQ, Zhu J, He CX, Benmoussa M, Wu P (1998). Molecular dissection of developmental behavior of plant height in rice (*Oryza sativa* L.). Genetics 150:1257–1265

Yu SB, Li JX, Xu CG, Tan YF, Gao YJ, Li XH, et al. (1997) Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. Proc Natl Acad Sci USA 94:9226–9231

Zeng ZB (1993) Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci. Proc Natl Acad Sci USA 90:10972–10976

Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics 136:1457–1468

Zhu J (1998) Mixed-model approaches of mapping genes for complex quantitative traits. In: Wang LZ, Dai JR (eds) Proc Genet and Crop Breed of China. Chinese Agricultural Science and Technology Publication House, Beijing, pp 19–20

Zhu J, Weir BS (1998) Mixed-model approaches for genetic analysis of quantitative traits. In: Chen LS, Ruan SG, and Zhu J (eds) Advanced topics in biomathematics. Proc Inter Conf on Mathematical Biol. World Scientific Publishing Co., Singapore, pp 321–330