Theor Appl Genet (2007) 115:325–333
DOI 10.1007/s00122-007-0564-7

ORIGINAL PAPER

Mapping QTLs with digenic epistasis under multiple environments and predicting heterosis based on QTL effects

Yong-Ming Gao · Jun Zhu

Received: 22 August 2005 / Accepted: 23 April 2007 / Published online: 30 May 2007

Abstract Mixed linear model approach was proposed for mapping QTLs with the digenic epistasis and QTL by environment (QE) interaction as well as additive and dominant effects. Monte Carlo simulations indicated that the proposed method could provide unbiased estimations for both positions and genetic main effects of QTLs, as well as unbiased predictions for QE interaction effects. A method was suggested for predicting heterosis based on individual QTL effects. The immortalized F_2 (IF_2) population constructed by random mating among RI or DH lines is appropriate for mapping QTLs with epistasis and their QE interaction. Based on the models and methodology proposed, we developed a QTL mapping software, QTLMapper 2.0 on the basis of QTLmapper 1.0, which is suitable for analyzing populations of DH, RIL, F_2 and IF_2. Data of thousand grain weight of IF_2 population with 240 lines derived from elite hybrid rice Shanyou 63 were analyzed as a worked example.

Communicated by H. C. Becker.

Y.-M. Gao · J. Zhu (✉)
Department of Agronomy, Zhejiang University,
Hangzhou 310029, People's Republic of China
e-mail: jzhu@zju.edu.cn

Present Address:
Y.-M. Gao
Institute of Crop Science /National Key Facility for Crop Gene,
Resources and Genetic Improvement,
Chinese Academy of Agricultural Science,
Beijing 100081, People's Republic of China
e-mail: gaoym@caas.net.cn

Keywords QTL mapping · Heterosis prediction · Epistasis · QTL by environment interaction · Immortalized F_2

Introduction

The dissection of complex traits has been greatly facilitated with the advent of molecular markers since the 1980s. Lander and Botstein (1989) proposed an interval mapping (IM) method providing good estimations for position and effects of single QTL. To deal with multiple QTL problems, Jansen (1993, 1994) and Zeng (1993, 1994) independently proposed methods to detect QTL by conditioning the test on other selected markers, which yield better power than IM for mapping multiple QTLs. As an alternative approach to dissect quantitative traits, mixed linear model has been applied to map QTLs in humans and animals (Fernando and Grossman 1989; Xu and Atchley 1995; Meuwissen and Goddaod 1997).

Recent investigations suggested that epistasis might be an important genetic basis underlying complex traits (Li et al. 1997; Yu et al. 1997; Ohno et al. 2000). Many researchers have been focusing on statistical models and methods for analyzing QTLs with epistatic effects (Cockerham and Zeng 1996; Fijneman et al. 1996; Chase et al 1997; Lukens and Doebley 1999). Kao et al. (1999) and Zeng et al. (1999) have put forth a strategy of multiple interval mapping (MIM) to identify QTLs with epistasis. Jannink and Jansen (2001) suggest mapping QTLs with epistasis between QTLs and backgrounds using one-dimensional genome search.

QTL × environment (QE) interaction is another important issue in QTL mapping. QE interactions were analyzed in QTL mapping without epistasis (Jansen et al.

1995; Jiang and Zeng 1995; Cockerham and Zeng 1996; Yan et al. 1998; Vieira et al. 2000, Piepho 2000). Lukens and Doebley (1999) estimated the extent and variation of phenotypic plasticity of different genotypes between the high- and low-density environments using ANOVA. Wang et al. (1999) proposed mixed linear model approach for effectively estimating QTL effects of additive, additive × additive epistasis, as well as predicting their QE interaction effects. However, it remains unresolved to estimate dominance-related epistasis between QTLs and predict their interaction with environments.

Among mapping populations, F_2 is most informative in genetic analysis. However, since F_2 individuals are heterozygous, each genotype is different from the others in genetic construction, which makes it difficult to assess the reliability of the data and estimate the effects of QE interactions due to the inability to carry out replicated tests in different environments. As an alternative way for partly resolving this problem, the evaluation of F_3 families, usually referred to as $F_{2:3}$ populations, has been employed in QTL analysis (Edwards et al. 1987; Yu et al. 1997). Gains in precision, however, are partly sacrificed due to the reduction of genetic heterogeneity of replicates by additional selfing in F_3 families (Paterson 1997). Another disadvantage of using F_2 or $F_{2:3}$ population is that the data of marker genotypes cannot be repeatedly used. Li et al. (2000), using a vegetative replication of F_2 population, analyzed QTLs for yield of rice, without considering epistasis and QE interactions. However, the tolerance to ratooning displays difference with genotypes of individuals, and the estimates of QTL effects may have suffered from confounded effects of reduction in productivity resulting from ratooning. The immortalized F_2 (IF_2) can be derived from random mating among lines of RI or DH (Hua et al. 2002), so that a replicated population could be obtained with similar genetic architecture as in F_2 population.

In the present paper, we proposed a methodology for mapping QTLs with digenic epistasis (additive × additive, additive × dominance, and dominance × dominance) as well as additive and dominant effects and their QE interaction based on mixed linear model approaches for analyzing IF_2 population, it can also be used to analyze F_2 population in one environment without QE interactions. Simulations were conducted to testify the efficiency and unbiasedness of the algorithm. We also proposed a strategy to predict heterosis based on QTL analysis for elucidating the genetic basis of heterosis with the rationale provided by Xu and Zhu (1999). The data of thousand grain weight collected from IF_2 population of rice were analyzed as a worked example.

Genetic models and analysis methods for QTL mapping

For IF_2 population, a full mixed linear model for mapping QTLs with digenic epistasis and QE interactions could be written as

$$\begin{aligned} y_{hk} = {} & \mu + a_i x_{A_{ik}} + d_i x_{D_{ik}} + a_j x_{A_{jk}} + d_j x_{D_{jk}} \\ & + aa_{ij} x_{AA_{ijk}} + ad_{ij} x_{AD_{ijk}} + ad_{ji} x_{AD_{jik}} \\ & + dd_{ij} x_{DD_{ijk}} + e_h + ae_{ih} u_{A_i E_{hk}} + de_{ih} u_{D_i E_{hk}} \\ & + ae_{jh} u_{A_j E_{hk}} + de_{jh} u_{D_j E_{hk}} + aae_{ijh} u_{AA_{ij} E_{hk}} + ade_{ijh} u_{AD_{ij} E_{hk}} \\ & + ade_{jih} u_{AD_{ji} E_{hk}} + dde_{ijh} u_{DD_{ij} E_{hk}} \\ & + \sum_{f(h)} m_{f(h)} u_{M_{fk(h)}} + \sum_{l(h)} mm_{l(h)} u_{MM_{lk(h)}} + \varepsilon_{hk} \end{aligned} \quad (1)$$

where y_{hk} is the phenotypic observation of the k-th IF_2 genotype in environment h; μ is the population mean; a_i and a_j are the additive effects (fixed effects) of two putative QTLs (Q_i and Q_j), respectively; d_i and d_j are the dominance effects (fixed effects) in heterozygote of Q_i and Q_j; aa_{ij}, ad_{ij}, ad_{ji}, and dd_{ij} are the fixed effects of digenic epistasis of $A_i \times A_j$, $A_i \times D_j$, $A_j \times D_i$, and $D_i \times D_j$ between Q_i and Q_j; As coefficients of QTL effects, $x_{A_{ik}}, x_{A_{jk}}, x_{D_{ik}}, x_{D_{jk}}$ could be derived according to the observed genotypes of the markers (marker M_{i-} and M_{i+} flanking Q_i; marker M_{j-} and M_{j+} flanking Q_j) and the recombinant value ($r_{M_{i-}Q_i}$ and $r_{M_{j-}Q_j}$) between the testing points and left markers (M_{i-} and M_{j-}) (Table 1), $x_{AA_{ijk}} = x_{A_{ik}} x_{A_{jk}} - 0.5 + r_{ij}, x_{AD_{ijk}} = x_{A_{ik}} (x_{D_{jk}} + 0.5)$, $x_{AD_{jik}} = (x_{D_{ik}} + 0.5) x_{A_{jk}}$, $x_{DD_{ijk}} = (x_{D_{ik}} + 0.5)(x_{D_{jk}} + 0.5) - 0.5 + r_{ij} - r_{ij}^2$, r_{ij} is the recombinant value between Q_i and Q_j; e_h is the random effects of environment h with coefficient $u_{E_{hk}}$; ae_{ih} (or ae_{jh}) is the additive × environment interaction effect with coefficient $u_{A_i E_{hk}}$ (or $u_{A_j E_{hk}}$) for Q_i (or Q_j); de_{ih} (or de_{jh}) is the dominance × environment interaction effect with coefficient $u_{D_i E_{hk}}$ (or $u_{D_j E_{hk}}$) for Q_i (or Q_j); aae_{ijh}, ade_{ijh}, ade_{jih}, dde_{ijh} are the interaction effects between four kinds of epistasis and environments with coefficient $u_{AA_{ij} E_{hk}}, u_{AD_{ij} E_{hk}}, u_{AD_{ji} E_{hk}}, u_{DD_{ij} E_{hk}}$; $m_{f(h)}$ is the effect of marker f nested within the h-th environment with coefficient $u_{M_{fk(h)}}$ which takes value 1 for $M_f M_f$ marker genotype, 0 for $M_f m_f$, –1 for $m_f m_f$; $mm_{l(h)}$ is the effect of marker × marker interaction nested within the h-th environment with coefficient $u_{MM_{lk(h)}}$ which takes value 1, 0, or –1 depending on the genotypes of interaction markers; $m_{f(h)}$ and $mm_{l(h)}$ are used to control the genetic background of those QTLs outside the searching intervals; ε_{hk} is the residual effect.

Model (1) can be analyzed by mixed linear model approach (Wang et al. 1999), defining fixed factors for QTL main effects, and random factors for effects of environments, QTL by environment (QE) interactions, marker cofactors, and residuals.

Table 1 Coefficients ($x_{A_{xk}}$ or $x_{D_{xk}}$) of QTL additive and dominance effects for a F_2 or IF_2 population

Marker genotype	Expected Frequency			$x_{A_{xk}}$	$x_{D_{xk}}$
	$Q_x\ Q_x$	$Q_x\ q_x$	$q_x\ q_x$		
$M_{x-}\ M_{x-}\ M_{x+}\ M_{x+}$	1	0	0	1	–0.5
$M_{x-}\ M_{x-}\ M_{x+}\ m_{x+}$	$1–p_x$	p_x	0	$1–p_x$	p_x –0.5
$M_{x-}\ M_{x-}\ m_{x+}\ m_{x+}$	$(1–p_x)^2$	$2p_x\ (1–p_x)$	p_x^2	$1–2p_x$	$2p_x\ (1–p_x)$–0.5
$M_{x-}\ m_{x-}\ M_{x+}\ M_{x+}$	p_x	$1–p_x$	0	p_x	$0.5–p_x$
$M_{x-}\ m_{x-}\ M_{x+}\ m_{x+}$	$\delta_x\ p_x\ (1–p_x)$	$1–2\delta_x\ p_x\ (1–p_x)$	$\delta_x\ p_x\ (1–p_x)$	0	$0.5–2\delta_x\ p_x\ (1–p_x)$
$M_{x-}\ m_{x-}\ m_{x+}\ m_{x+}$	0	$1–p_x$	p_x	$–p_x$	$0.5–p_x$
$m_{x-}\ m_{x-}\ M_{x+}\ M_{x+}$	p_x^2	$2p_x\ (1–p_x)$	$(1–p_x)^2$	$–(1–2p_x)$	$2p_x\ (1–p_x)$–0.5
$m_{x-}\ m_{x-}\ M_{x+}\ m_{x+}$	0	p_x	$1–p_x$	$–(1–p_x)$	p_x –0.5
$m_{x-}\ m_{x-}\ m_{x+}\ m_{x+}$	0	0	1	–1	–0.5

The putative QTLx ($x = i$ or $x = j$) and its flanking markers which define the interval are lined in chromosome as $M_{x-} \rightarrow Q_x \rightarrow M_{x+}$. $p_x = r_{M_{x-}Q_x} / r_{M_{x-}M_{x+}}$, $\delta_x = r^2_{M_{x-}M_{x+}} \Big/ [(1 - r_{M_{x-}M_{x+}})^2 + r^2_{M_{x-}M_{x+}}]$. $r_{Mx–Mx+}$ is the recombination fraction between M_{x-} and M_{x+};$r_{Mx\text{-}Qx}$ is the recombination fraction between the left marker M_{x-} and Q_x. Assumed that $Q_x\ Q_x$ has a positive effect on the trait, while $q_x\ q_x$ has a negative effect. Double recombination is ignored

The QTL fixed effects can be estimated by generalized least squares method, and the QE interaction effects predicted by the best linear unbiased prediction (BLUP) method (Henderson 1963) or linear unbiased prediction (LUP) method (Zhu and Weir 1996). To test the significance of single fixed effect of QTLs, a t-statistic, with degrees of freedom $df = n–rank(X)$, is used, while a single random effect of QE interaction can be tested by the z-statistic based on standard normal distribution (Wang et al. 1999).

The method proposed by Lander and Kruglyak (1995) was applied to determine the critical value for claiming QTL detection. The point-wide significance level of $\alpha\ (X)$ was obtained from the equation $\alpha_0\ (X) = (C + 2\rho\ GX)\alpha\ (X)$ by iterative calculation. In the above equation, X is the LR value with significance level $\alpha\ (X)$ (Type I error),C is the number of chromosome, G is the total length of genome with unit M,ρ reflects the total crossing over rate between the genotypes being compared.

Heterosis prediction based on QTL effects

The total heterosis under specific environment is composed of general heterosis due to QTL main effects (dominance and epistasis) and interaction heterosis due to QE interaction effects (dominance by environment and epistasis by environment). The estimation of QTL main effects and the prediction of QE interaction effects could be used to construct model for predicting general heterosis and interaction heterosis of F_1 and the following generations, which would be helpful to uncover the genetic basis of heterosis.

The overall general heterosis (H_M) and QE interaction heterosis (H_{ME_h}) over mid-parent in F_1 can be predicted with following model:

$$H_M(\mathrm{F}_1) = \sum_{i=1}^{p} d_i + \sum_{k=1}^{p-1}\sum_{l=k+1}^{p} (dd_{kl} - aa_{kl})$$

$$H_{ME_h}(\mathrm{F}_1) = \sum_{i=1}^{p} de_{ih} + \sum_{k=1}^{p-1}\sum_{l=k+1}^{p} (dde_{klh} - aae_{klh})$$

in F_2:

$$H_M(\mathrm{F}_2) = \frac{1}{2}\sum_{i=1}^{p} d_i + \sum_{k=1}^{p-1}\sum_{l=k+1}^{p}\left(\frac{1}{2}\rho_{kl} - 1\right)aa_{kl} + \frac{1}{4}\sum_{k=1}^{p-1}\sum_{l=k+1}^{p}(1 + \rho_{kl}^2)dd_{kl}$$

$$H_{ME_h}(\mathrm{F}_2) = \frac{1}{2}\sum_{i=1}^{p} de_{ih} + \sum_{k=1}^{p-1}\sum_{l=k+1}^{p}\left(\frac{1}{2}\rho_{kl} - 1\right)aae_{klh} + \frac{1}{4}\sum_{k=1}^{p-1}\sum_{l=k+1}^{p}(1 + \rho_{kl}^2)dde_{klh}$$

and in F_n:

$$H_M(F_n) = \frac{1}{2^{n-1}}\sum_{i=1}^{p} d_i + \sum_{k=1}^{p-1}\sum_{l=k+1}^{p}\left(\sum_{m=1}^{n-1}\frac{1}{2^m}\rho_{kl}^m - 1\right)aa_{kl} + \sum_{k=1}^{p-1}\sum_{l=k+1}^{p}\frac{1}{4^{n-1}}(1 + \rho_{kl}^2)^{n-1}dd_{kl}$$

$$H_{ME_h}(\mathrm{F}_n) = \frac{1}{2^{n-1}}\sum_{i=1}^{p} de_{ih} + \sum_{k=1}^{p-1}\sum_{l=k+1}^{p}\left(\sum_{m=1}^{n-1}\frac{1}{2^m}\rho_{kl}^m - 1\right)aae_{klh} + \sum_{k=1}^{p-1}\sum_{l=k+1}^{p}\frac{1}{4^{n-1}}(1 + \rho_{kl}^2)^{n-1}dde_{klh}$$

where n is the selfing generation, p is the number of QTLs, $\rho_{kl} = 1–2r_{kl}$ is the correlation coefficient between QTL k and QTL l,r_{kl} is the recombinant value between QTL k and QTL l.

Simulation studies

QTL setting for simulations

A total of 100 Monte Carlo simulations for an IF_2 population with 200 individuals were conducted to testify the unbiasedness of estimations for QTL positions and genetic main effects, and of predictions for QE interaction effects. A genome of four chromosomes was constructed, with 40 evenly distributed markers at 10 cM space. Among a total of four QTLs, one was placed on chromosome I, one on chromosome II, and the other two linked on chromosome IV. QTL setting included effects of additive, dominance and epistasis of additive × additive, additive × dominance, dominance × dominance (Table 3), the corresponding QE effects (Table 4) were set for multiple environments. The heritability was assumed to be 0.5.

Analysis of QTL position

The positions of pair-wise QTLs involved in epistasis without QE interactions were estimated by two-dimensional search along the genome at every 2 cM position. The point-wide significance level of $\alpha_{(X)} = 0.000149$ was gained by setting the genome-wide significance level $\alpha_{0(X)}$ as 0.05 (Type I error) according to the method proposed by Lander and Kruglyak (1995). The results (Table 2) included the mean and standard error of estimated QTL positions and the power for identifying the interaction QTLs. It was shown that the mixed linear model approaches could result in unbiased and accurate estimation for the QTL positions. The deviations of all estimates for QTL positions were less than 2 cM except QTL 4. The results revealed that the power of identifying QTLs ranged from 0.37 to 0.98, suggesting the effectiveness of the new method in identifying pairs of epistatic QTLs. The impact of genetic main effects on the power of QTL mapping could not be found due to the complicacy of these effects in IF_2 population. High stability of the mixed linear model approaches in estimation of QTL position is evident from the low standard error values.

Estimation of genetic main effects

Monte Carlo simulations were also conducted to investigate the accuracy of estimated genetic main effects of QTLs. Table 3 provided the mean and standard error of the estimated QTL main effects. The results indicated that unbiased estimation could be obtained for QTL main effects at the preset QTL positions, except for some small effects with quite large deviations, such as the epistatic effects of additive × dominance between QTL 1 and QTL 2, and between QTL 3 and QTL 2, dominance × dominance between QTL 2 and QTL 4. Among the QTL main effects, the standard errors of estimated dominance effects were larger than that for additive effects. Estimation for epistatic effects with small standard error values was observed for additive × additive, followed by additive × dominance, and the highest for the dominance × dominance.

Prediction of QE interaction effects

The genome setting information above was also used in the simulations for the prediction of QE interaction effects. In the present study, LUP was used to predict QE interaction effects under three environments. The simulation results

Table 2 QTL positions estimated by mixed linear model approaches

QTL i	QTL j	Site i (cM)			Site j (cM)			Power [d]
		Par.[a]	Est.[b]	SE[c]	Par.	Est.	SE	
1	2	48.1	47.1	0.21	51.5	52.1	0.33	0.70
1	3	48.1	46.9	0.17	2.0	2.9	0.28	0.98
1	4	48.1	47.0	0.23	70.0	73.4	0.39	0.63
2	3	51.5	51.6	0.26	2.0	2.6	0.30	0.76
2	4	51.5	51.8	0.29	70.0	72.6	0.42	0.54
3	4	2.0	2.4	0.36	70.0	73.1	0.53	0.37

[a] Parameter of QTL positions apart from left end of the chromosome involved

[b] Estimates of QTL positions in 100 simulations

[c] standard errors for estimates of QTL positions in 100 simulations calculated only by QTLs detected

[d] Probability for identifying the QTLs by two-dimentional search. The estimates of QTL positions were obtained within the pair–wise marker intervals each having at least one marker flanking the QTLs

Table 3 Parameters and estimates of genetic main effects of QTLs in 100 simulations

QTL*i*	Additive and dominance effects						QTL*i*	QTL*j*	Epistatic effects											
	a_i[a]			d_i					aa_{ij}			ad_{ij}			ad_{ji}			dd_{ij}		
	Par.[b]	Est.[c]	SE[d]	Par.	Est.	SE			Par.	Est.	SE	Par.	Est.	SE	Par.	Est.	SE	Par.	Est.	SE
1	1.58	1.51	0.04	0.49	0.52	0.06	1	2	0.67	0.74	0.04	0.36	0.48	0.06	0.00	–0.05	0.05	0.21	0.18	0.08
2	–0.97	–0.97	0.04	0.63	0.62	0.05	1	3	–0.42	–0.40	0.04	0.00	–0.12	0.04	–0.38	–0.41	0.06	0.00	0.01	0.08
3	0.82	0.78	0.04	0.68	0.75	0.06	1	4	0.51	0.53	0.04	–0.68	–0.70	0.05	0.00	–0.12	0.06	0.00	0.00	0.07
4	0.60	0.56	0.04	–0.47	–0.50	0.05	2	3	1.06	0.98	0.04	0.00	–0.04	0.06	–0.22	–0.06	0.05	0.00	0.01	0.07
							2	4	–0.88	–0.83	0.04	0.00	0.11	0.05	0.51	0.45	0.05	–0.37	–0.50	0.07
							3	4	0.00	–0.09	0.04	0.41	0.39	0.05	–0.59	–0.58	0.06	0.30	0.30	0.08

[a] a_i= additive effect of QTL*i*;d_i= dominance effect of QTL*i*; $aa_{ij} = A_i \times A_j$, $ad_{ij} = A_i \times D_j$, $ad_{ji} = A_j \times D_i$, and $dd_{ij} = D_i \times D_j$, respectively

[b] Parameters of QTL genetic main effects

[c] Estimates of QTL genetic main effects

[d] Standard errors for estimates of QTL genetic main effects

indicated that mixed linear model approaches could effectively predict QE interaction effects (Table 4). All the standard errors of QE interaction effects were quite small, which supported the effectiveness of the proposed methods.

A worked example of mapping QTLs and predicting heterosis for thousand grain weight in rice

The population of recombinant inbred line (RIL) of rice derived from Zhenshan 97B × Minghui 63 was provided by Qifa Zhang of Huazhong Agricultural University. There were 241 lines in the RIL population with a genetic linkage map consisted of 221 markers which covered 1,796.58 cM of the rice genome. We constructed an IF_2 population of rice with 240 F_1 hybrids by randomly mating among the RIL lines based on the rationale proposed by Hua et al. (2002). The IF_2 population was evaluated by a randomized complete block design with two replications in 1999 and 2000 at Zhejiang University. All hybrids and parental RILs were planted in the seedling nursery on 19 May and transplanted into the plots with three rows for the hybrids and five rows for RILs on 15 June in 1999, and 16 June in 2000. Each row within a plot had 12 plants with a space of 17 cm × 26 cm. Five typical plants from the middle of each plot were evaluated for thousand grain weight (Kw).

According to method proposed by Lander and Kruglvak (1995), when significance level of genome scan was set at 0.05, the point-wise significance level of QTL mapping can be gained by iterative computation. For IF_2 population of rice constructed in the present study, it was 0.000018 with the corresponding threshold of likelihood ratio (LR) setting as 50.6. Data were analyzed using software QTLMapper 2.0.

QTL main effects and interactions with environments

A total of 9 pairs of epistatic loci involved 15 QTLs were identified for thousand grain weight in rice. The significance of QTL main effects was tested by a *t*-statistic, and the significance of QE interaction effects was tested by Jackknife sampling techniques (Zhu 1997). There were ten QTLs showed significant or extremely significant additive effects, and five QTLs showed significant or extremely significant dominance effects (Table 5). Both of additive and dominance effects of Kw3-1, Kw 5-1, Kw 6-3 and Kw7-1 were significant. Few interactions of additive and dominance with environments were observed, among which interactions between additive and environments for Kw3-2 and Kw6-3 were significant. No significant interaction between dominance and environments was observed.

Digenic epistasis and their interactions with environments

Digenic epistatic effects of nine pairs of loci were estimated, corresponding QE interactions were predicted (Table 6). The significant effects were detected for four *aa* effects, six *ad* effects, and three *dd* effects. The *QE* interaction effects for epistasis showed less significance.

Prediction on heterosis over mid-parent

The thousand grain weight showed high positive general heterosis over mid-parent (2.80 g) in F_1 (Table 7), which was close to the real performance of heterosis over mid-parent of Shan-you 63 (2.45 g) (Hua et al. 2003). It was indicated that the QTLs identified for thousand grain weight showed high contribution to heterosis. The interaction

Table 4 Prediction of *QE* interaction under three environments in 100 simulations

Env	QTL*i*	ae_{ih}[a]			de_{ih}			QTL*i*	QTL*j*	aae_{ijh}			ade_{ijh}			ade_{jih}			dde_{ijh}		
		Par[b]	Pre.[c]	SE[d]	Par.	Pre.	SE			Par.	Pre.	SE	Par.	Pre.	SE	Par.	Pre.	SE	Par.	Pre.	SE
E_1	1	–0.72	–0.69	0.01	0.61	0.66	0.02	1	2	0.62	0.60	0.01	0.00	–0.03	0.02	0.00	0.00	0.02	0.27	0.24	0.02
	2	0.63	0.65	0.01	–0.83	–0.75	0.02	1	3	–0.67	–0.63	0.01	0.00	–0.01	0.02	–0.24	–0.22	0.01	0.00	0.08	0.02
	3	–0.34	–0.41	0.01	1.25	1.03	0.02	1	4	0.71	0.67	0.02	0.00	–0.02	0.02	0.00	–0.02	0.02	0.00	0.02	0.02
	4	0.26	0.19	0.01	0.17	0.32	0.02	2	3	0.00	0.01	0.01	0.43	0.39	0.01	0.00	–0.12	0.02	–0.49	–0.36	0.02
								2	4	0.00	0.01	0.01	–0.33	–0.25	0.02	–0.41	–0.39	0.02	0.37	0.28	0.02
								3	4	–0.91	–0.89	0.01	0.00	–0.03	0.02	0.26	0.26	0.02	0.00	0.04	0.02
E_2	1	1.08	1.04	0.01	–0.35	–0.45	0.01	1	2	0.59	0.55	0.01	0.00	0.01	0.02	0.00	0.01	0.02	–0.35	–0.30	0.02
	2	0.21	0.16	0.01	0.49	0.55	0.01	1	3	–0.34	–0.32	0.01	0.00	0.05	0.01	–0.28	–0.21	0.01	0.00	–0.04	0.02
	3	0.55	0.53	0.01	–0.96	–0.66	0.01	1	4	–0.25	–0.25	0.01	0.00	0.02	0.02	0.00	–0.09	0.02	0.00	0.02	0.02
	4	–0.61	–0.45	0.01	0.54	0.44	0.01	2	3	0.00	0.01	0.01	–0.12	–0.09	0.01	0.00	0.13	0.01	0.75	0.63	0.02
								2	4	0.00	0.00	0.01	0.00	–0.01	0.02	0.41	0.38	0.01	–0.16	–0.06	0.02
								3	4	0.74	0.72	0.01	0.27	0.28	0.02	0.00	–0.03	0.02	0.00	0.00	0.02
E_3	1	–0.36	–0.35	0.01	–0.26	–0.21	0.01	1	2	–1.21	–1.15	0.01	0.00	0.01	0.02	0.00	–0.01	0.02	0.08	0.06	0.02
	2	–0.84	–0.81	0.01	0.32	0.21	0.01	1	3	1.01	0.95	0.01	0.00	–0.03	0.01	0.44	0.43	0.01	0.00	–0.05	0.02
	3	–0.21	–0.12	0.01	–0.29	–0.37	0.01	1	4	–0.46	–0.42	0.01	0.00	0.00	0.01	0.00	0.11	0.01	0.00	–0.04	0.02
	4	0.35	0.26	0.01	–0.71	–0.76	0.01	2	3	0.00	–0.02	0.01	–0.31	–0.30	0.01	0.00	–0.01	0.02	–0.26	–0.27	0.02
								2	4	0.00	0.00	0.01	0.33	0.27	0.01	0.00	0.01	0.01	–0.21	–0.22	0.02
								3	4	0.17	0.17	0.01	–0.27	–0.26	0.01	–0.26	–0.22	0.01	0.00	–0.04	0.02

[a] $e_{A_iE_h}$ and $e_{D_iE_h}$ are the interaction effects of additive and dominance for QTL*i* with environment *h* respectively; $e_{AA_{ij}E_h}$, $e_{AD_{ij}E_h}$, $e_{AD_{ji}E_h}$ and $e_{DD_{ij}E_h}$ are the interaction effects of $A_i \times A_j$, $A_i \times D_j$, $A_j \times D_i$, and $D_i \times D_j$ with environment *h*, respectively

[b] Parameters of *QE* interaction effects

[c] Predicted *QE* interaction effects

[d] Standard errors for estimates of *QE* interaction effects

Table 5 Analysis of QTL effects of additive, dominance and their interactions with environments for thousand grain weight of rice

QTL*i*[a]	Flanking Markers	a_i[b]	d_i	ae_{i1}
Kw1-1	G359-RG532	–0.79***		
Kw1-2	C39-RM237	–0.44***		
Kw2-2	C777-RZ386	0.34*		
Kw3-1	C1087-RZ403	0.72**	–0. 40*	
Kw3-2	C746-C944		–0.65*	–0.41*
Kw5-1	RG360-RM42	–1.04***	–0.35*	
Kw6-2	Y4073L-RZ667	0.52***		
Kw6-3	RG653-G342	0.62***	–0.58**	–0.38**
Kw7-1	RG128-C1023	0.66**	0.82*	
Kw7-2	C1023-R1440	0.64***		
Kw10-2	RM228-C371	–0.33*		

[a] Kw is the abbreviation of thousand grain weight, the first number is the chromosome the QTL located, the number after '–' is the series number of QTLs detected in the corresponding chromosome

[b] a_i are additive and dominance effects of QTL*i*, respectively;ae_{i1} ($ae_{i2} = -ae_{i1}$) is the additive by environment interaction of QTL*i* in 1999

*, ** and *** mark significance level at 0.05, 0.01 and 0.005 respectively

heterosis was negative in 1999 and positive in 2000. The general heterosis decreased in F_2, while the interaction heterosis increased in 1999 and decreased in 2000.

Discussion

It has been proved that epistasis between different loci exists universally during the long period of plant evolution (Clegg et al. 1972; Armbruster et al. 1997; Ungerer et al. 2003). Recently, QTL mapping suggested that epistasis was main genetic basis of complex traits (Li et al. 1997; Yu et al. 1997; Ohno et al. 2000). However, the study on dominance-related epistasis, which might play a key role in heterosis, leaved much to be desired. The strategy proposed in the present paper can unbiasedly estimate three types of digenic epistasis using F_2 or IF_2 population. Though the genetic effects of additive and dominance can also be estimated with no bias when ignoring epistasis, the power of QTL detection would be decreased (data not showed).

Specific expression of genes in various environments is widely accepted by biologists. Some researches compared mapping results in different environments (Paterson et al. 1991; Stuber et al. 1992; Lu et al. 1997). However, the differences in mapping results among different environments could not precisely indicate the existence of QE interaction and vice versa (Jansen et al. 1995). The mixed-model approaches for QTL mapping can provide unbiased prediction on QE interaction as well as digenic epistasis when the experiment was conducted under multiple environments. Monte Carlo simulation suggested that, when ignoring QE interaction, the estimation on QTL main effects might be biased because the QTL main effects and QE interaction were mixed in that case (data not showed). The power of QTL identification will be increased by using data collected from the experiments under multiple environments.

Table 7 Prediction on general heterosis and interaction heterosis of thousand grain weight in rice

Generation	H_M	H_{ME_1}	H_{ME_2}
F_1	2.80	–2.54	2.58
F_2	0.64	–1.03	1.04

H_M is the prediction on general heterosis over mid-parent; H_{ME_1} and H_{ME_2} are the prediction on interaction heterosis over mid-parent in 1999 and 2000, respectively

Table 6 Analysis of QTL epistatic effects and their interactions with environments for thousand grain weight of rice

QTL*i*	QTL*j*	aa_{ij}[a]	ad_{ij}	ad_{ji}	dd_{ij}	aae_{ij1}	ade_{ij1}	ade_{ji1}	dde_{ij1}
Kw1-2	Kw2-2	0.63*	0.49*	–0.49*					
Kw1-3	Kw3-1		0.73***	0.86***					
Kw1-3	Kw5-1				0.85*		0.63*		
Kw2-1	Kw7-1	0.98*		–0.97**					
Kw3-1	Kw11-1	–0.93***							
Kw3-2	Kw10-1		0.78*	0.79**	1.23*				
Kw5-1	Kw7-1								–1.35**
Kw6-2	Kw6-3			–0.78***	0.84***	0.57***			
Kw7-2	Kw10-2	–0.60***	–0.63***					–0.61**	–0.62*

[a] aa_{ij}, ad_{ij}, (ad_{ji}), and dd_{ij} are the epistatic effects of additive × additive, additive × dominance, and dominance × dominance between QTL*i*and QTL*j*, respectively;aae_{ij1},ade_{ij1},$ade_{ji1}$and$dde_{ij1}$are the QTL by environment interactions in 1999, and the opposite values of *QE* effects are expected in 2000

*, ** and *** denote significance level at 0.05, 0.01 and 0.005

The model for predicting heterosis in F_1, F_2 and following generations can illuminate the genetic basis of heterosis and inbreeding depression. It was suggested by the present study that *d*, *dd* and *aa* played an important role in heterosis. According to the prediction model for heterosis proposed in this study, effects of *aa* and *aae* increased with generation, while effects of *d*, and *dd* and their interaction with environments decreased with generation. In case when the increments of the former exceed the reduction of the later, the heterosis might not decrease with generation.

For further improvement on parents of hybrid, the heterosis performance of single locus and/or pair-wise epistatic loci may be more attractable. Based on the information of heterosis-related QTLs, it can be revealed which locus and/or locus pair could be improved in further breeding program by manipulating the related QTLs (Yang and Zhu 2005). There might exist overdominant loci for some specific cross, however partial dominance with negative or positive effects could be observed more frequently. That explained why the analysis of correlation between heterozygosity and heterosis could not obtain consistent conclusions (Lee et al. 1989; Smith et al. 1990; Stuber et al. 1992; Zhang et al. 1994; Xiao et al. 1995; Saghai et al. 1997).

A QTL mapping software, QTLMapper 2.0 written by C/C++ language was developed on the basis of QTLmapper 1.0 (Wang et al. 1999). The software could analyze populations of DH, RIL, F_2 and immortalized F_2, and could provide unbiased estimation for genetic main effects and/or unbiased prediction for the corresponding QE interaction effects. The heterosis over mid-parents of single QTL, pair of epstatic QTLs and overall QTLs could also be predicted based on QTL effects. The software is available at http://www.cab.zju.edu.cn/english/ics/faculty/zhujun.htm.

Acknowledgments We greatly thank two anonymous reviewers for their useful comments and suggestions on the earlier version of the manuscript. This work was supported in part by ''The key project of National Natural Science Foundation of China'' and the National Hi-Tech Research and Development Program (863) of China.

References

Armbruster P, Bradshaw WE, Holzapeel CM (1997) Evolution of the genetic architecture underlying fitness in the pitcher-plant mosquito, *Wyeomyia smithii*. Evolution 51:451–458

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

Clegg MT, Allard RW, Kahler AL (1972) Is the gene the unit of selection? Evidence from two experimental plant populations. Proc Natl Acad Sci USA 69:2474–2478

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

Edwards MD, Stuber CW, Wendel JF (1987) Molecular marker facilitated investigations of quantitative trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics 116:113–125

Fernando RL, Grossman M (1989) Marker—assisted selection using best linear unbiased prediction. Genet Sel 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. Nat Genet 14:465–467

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

Henderson CR (1963) Selection index and expected genetic advance. In: Hanson WD, Robinson HF (eds) Statistical genetics and plant breeding. Natl Acad Sci N R C 982:141–163

Hua JP, Xing YZ, Xu CG, Sun SL, Yu SB, Zhang QF (2002) Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantageous for performance. Genetics 162:1885–1895

Hua JP, Xing YZ, Wu WR, Xu CG, Sun XL, Yu SB, Zhang Q (2003) Single-locus heterotic effects and dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. Proc Natl Acad Sci USA 100:2574–2579

Jannink J-L, Jansen R, (2001) Mapping epistatic quantitative trait loci with one-dimensional genome searches. Genetics 157:445–454

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

Jansen RC (1994) Controlling the type I and type II errors in mapping quantitative trait loci. Genetics 138:871–881

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 Z–B (1995) Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics 140:1111–1127

Kao C-H, Zeng Z-B, Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. Genetics 152:1203–1216

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

Lander E and Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11:241–247

Lee M, Godshalk K, Lamkey KR, Woodman WW, (1989) Association of restriction fragment length polymorphism among maize inbreeds with agronomic performance of their crosses. Crop Sci 29:1067–1071

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

Li JX, Yu SB, Xu CG, Tan YF, Gao YJ, Li XH, Zhang Q (2000) Analyzing quantitative trait loci for yield using a vegetatively replicated F_2 population from a cross between the parents of an elite rice hybrid. Theor Appl Genet 101:248–254

Lu CF, Shen LS, Tan ZB, Xu YB, He P, Chen Y, Zhu LH (1997) Comparative mapping of QTLs for agronomic traits of rice across environments by using a doubled-haploid population. Theor Appl Genet 94:145–150

Lukens LN, Doebley J (1999) Epistatic and environmental interactions for quantitative trait loci involved in maize evolution. Genet Res 74:291–302

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

Ohno Y, Tanase H, Nabika T, Otsuda K, Sasaki T, Suzawa T, Korii T, Yamori Y, Saruta T (2000) Selective genotyping with epistasis

can be utilized for a major quantitative trait locus mapping in hypertension in rats. Genetics 155:785–792

Paterson AH (1997) Comparative mapping of plant phenotypes. Plant Breed Rev 14:13–37

Paterson AH, Damon S, Hewitt JD, Zamir D, Rabinowitch HD, Lincoln SE, Lander ES, Tanksley SD (1991) Mendelian factors underlying quantitative traits in tomato: Comparison across species, generations, and environments. Genetics 127:181–197

Piepho H-P (2000) A mixed-model approach to mapping quantitative trait loci in barley on the basis of multiple environment data. Genetics 156:2043–2050

Saghai Maroof MA, Yang GP, Zhang Q, Garvois KA (1997) Correlation between molecular marker distance and rice. Crop Sci 37:145–150

Smith OS, Smith JSC, Bowen SL, Tenborg TA, Wall SJ (1990) Similarities among a group of elite maize inbreds as measured by pedigree, F1 grain yield, grain yield heterosis and RFLPs. Theor Appl Genet 80:833–840

Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander ES (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132:823–839

Ungerer MC, Linder CR, Rieseberg LH (2003) Effects of genetic background on response to selection in experimental populations of *Arabidopsis thaliana*. Genetics 163:277–286

Vieira C, Pasyukova EG, Zeng Z-B, Hackett JB, Lyman RF, Machay TFC (2000) Genotype–environment interaction for quantitative trait loci affecting life span in *Drosophila melanogaster*. Genetics 154:213–227

Wang DL, Zhu J, Li ZK, Paterson AH (1999) Mapping QTLs with epistatic effects and QTL × environment interactions by mixed linear model approaches. Theor Appl Genet 99:1255–1264

Xiao J, Li J, Yuan LP, 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

Xu S, Atchley WR (1995) A random model approach to interval mapping of quantitative trait loci. Genetics 141:1189–1197

Xu ZC and Zhu J (1999) An approach for predicting heterosis based on an additive, dominance and additive × additive model with environment interaction. Heredity 82:510–517

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

Yang J and Zhu J (2005) Methods for predicting superior genotypes under multiple environments based on QTL effects. Theor Appl Genet 110:1268–1274

Yu S-B, Li J-X, Xu C-G, Tan Y-F, Gao Y-J, Li X-H, Zhang Q (1997) Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. Proc Natl Acad Sci USA 94:9226–9231

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

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

Zeng Z-B, Kao C–H, Basten CJ (1999) Estimating the genetic architecture of quantitative traits. Genet Res 74:279–289

Zhang Q, Gao YJ, Yang SH, Tagab RA, Saghai Maroof MA, Li ZB (1994) A diallel analysis of heterosis in elite hybrid rice based in RFLPs and microsatellites. Theor Appl Genet 89:185–192

Zhu J (1997) Analytical method of genetic model. Agriculture Press of China, Beijing

Zhu J, Weir BS (1996) Diallel analysis for sex-linked and maternal effects. Theor Appl Genet 92:1–9