浙江大学学报(农业与生命科学版)　25(5)：453～461，1999
Journal of Zhejiang University (Agric. & Life Sci.)

Article ID: 1008-9209(1999)05-0453-09

Simulation Studies on Identification of Interaction Markers Representing QTL Epistasis by Stepwise Regression

Wang Dao-long[1], ZHU Jun[2], LI Zhi-kang[3] (1. *Dept. of Statistics and Mathematics, Anhui Agricultural Univ., Hefei* 230036, *China*; 2. *Dept. of Agronomy, Zhejiang Univ., Hangzhou* 310029, *China*; 3. *Plant Breeding, Genetics, and Biochemistry Dept., International Rice Research Inst.*, 1099 *Manila, Philippines*)

Abstract: The usefulness of stepwise regression in identifying interaction markers that represent QTL epistasis was examined in detail by Monte Carlo simulations. It was indicated that stepwise regression method was powerful in identifying correct interaction markers, and better than two-way ANOVA that is currently used for the same purpose. Close linkage tended to reduce the resolution of identifying interaction markers and distort detection power. Large sample size and/or higher heritability could generally increase the chance of correct interaction markers being identified. The chance of identifying a specific marker interaction relied heavily on the relative contribution of the QTL epistasis that the marker interaction represented. Significance level was also an important factor affecting the power of identifying correct marker interactions. Several rounds of analyses under different significance levels were suggested with additional consideration of heritability and sample size used.

Key words: Monte Carlo simulation; interaction markers; QTL epistasis; stepwise regression

---

王道龙[1]，朱 军[2]，黎志康[3](1. 安徽农业大学 数理统计系，安徽 合肥 230036；2. 浙江大学 农学系，浙江 杭州 310029；3. 国际水稻所 作物遗传育种与生化系，菲律宾 马尼拉 1099)
用逐步回归鉴别上位性 QTL 互作标记的模拟研究(英文). 浙江大学学报(农业与生命科学版)，1999，25(5)：453～461

摘　要：利用蒙特卡罗模拟对逐步回归分析在鉴别代表 QTL 上位性的互作遗传标记上的应用进行了详细研究. 结果表明，逐步回归分析能够有效、准确地鉴别出反映 QTL 上位性的互作遗传标记，并优于当前常用的双向方差分析. QTL 之间的紧密连锁会导致互作标记鉴别的分辨率降低，并使鉴别的效率出现扭曲. 大样本和/或高遗传率可以从总体上提高互作标记准确鉴别的效率，而特定的互作标记鉴别的效率在很大程度上取决于它所代表的 QTL 上位性的相对贡献率. 显著水平也是影响互作标记鉴别效率的一个重要因子. 本文建议在几种不同的显著水平下进行若干次逐步回归分析，并在确定适当的显著水平时考虑性状的遗传率及样本容量.

关 键 词：蒙特卡罗模拟；互作标记；QTL 上位性；逐步回归分析
中图分类号：O212.1；S330　　**文献标识码**：A

Received date: 1999-04-26
Foundation item: China Natural Science Foundation (39893350)
Biography: Wang Dao-long, male, born in 1963, Ph. D. in Genetics and Plant Breeding, associate professor.

1 Introduction

Since the advent of DNA markers in the 1980s, complete molecular linkage maps have been constructed in many organisms[1]. One of the most important applications of DNA markers and molecular linkage maps is to dissect genetic variation of quantitative traits into individual Mendelian factors or quantitative trait loci (QTLs). The dissection of unobserved QTLs (or QTL mapping) with DNA markers is based on marker-trait associations resulting from linkage disequilibrium. Statistical analysis is necessary for identifying such associations and further mapping QTLs.

There have been some powerful complex statistical methods[2~5] proposed for mapping QTLs. These methods are important for understanding the exact position and genetic effects of QTLs. In practice, however, simple statistical methods are still useful for rough understanding of distributions and effects of QTLs on the genome by detecting marker-trait associations, instead of accurate mapping of QTLs. For instance, only the knowledge of markers that are important for the traits studied is usually enough for marker-assisted selection in genetic improvement of plants and animals. Besides, identification of important markers may also be helpful for accurately mapping QTLs as cofactors in statistical models[5].

Of many statistical methods, single QTL models (one-way ANOVA, simple regression, or t tests) are most commonly used for identifying associations between markers and QTLs[6~8]. These methods work well only when QTLs affecting the same traits in a mapping population are independent from one another. In other words, there is no linkage and/or epistasis between QTLs. Violation of this assumption could result in two serious related problems. The first one is the failure to detect epistasis, which may be an important genetic basis for many complex phenotypes[9~13]. The second is the reduced power, accuracy and resolution in marker detection, which results from background genetic variation (BGV), or noises arising from the influences on the same traits of multiple segregating QTLs[5, 14].

In many previous QTL mapping studies, two-way ANOVA has been regularly used to detect epistasis between QTLs[15~19]. In these cases, BGV control and efficient identification of pairwise interaction markers related to digenic epistasis remain as challenging problems. Hackett *et al*. compared several statistical methods in identification of markers by computer simulations, and found that multiple regression had the highest power and resolution in identifying correct markers related to single QTLs due to its effective control of BGV[20]. Nevertheless, the rationality and power of multiple regression in identifying interaction markers has not yet been examined theoretically.

In the present paper, we studied the properties of stepwise regression in the identification of interaction markers under different conditions (sample size, trait heritabilities, significance levels and linkage relationships, etc.) by Monte Carlo simulations, with the objective of learning the usefulness of stepwise regression in identification of epistasis-related interaction markers. Some related issues are also discussed.

2 Method Description

Assume there is a complete linkage map with a total of Φ markers covering the whole genome of a diploid species. Of the Φ markers, there are φ($\varphi \ll \Phi$) markers most closely linked with segregating QTLs affecting a quantitative trait in a mapping population. Then, a multiple regression model for the quantitative trait can be

written as

$$y_k = b_0 + \sum_h b_{M_h} \chi_{M_{hk}} + \sum_{i<j}\sum b_{MM_{ij}} \chi_{MM_{ijk}} + \varepsilon_k \quad (1)$$

where y_k is the phenotypic value of individual k ($k=1, \cdots, n$); b_0 is the intercept of the regression model, b_{M_h} is the partial regression coefficient for the hth main-effect marker ($h = 1, \cdots, \varphi$); $b_{MM_{ij}}$ is the partial regression coefficient for marker interaction between marker M_i and marker M_j ($i = 1, \cdots, \varphi-1$; $j=i+1, \cdots, \varphi$); indicator variable $\chi_{M_{hk}}$ depends on the observed genotype and the population type (Table 1); $\chi_{MM_{ijk}} = \chi_{M_{ik}} \chi_{M_{jk}}$; $\varepsilon_k \sim N(0, \sigma_\varepsilon^2)$; is the random residual composed of random error and a small portion of genetic effects that are not absorbed by the main-effect markers and interaction markers.

Table 1 Indicator variable ($\chi_{M_{hk}}$) for main-effect markers in multiple-regression model

Marker Genotype	Population type[1)]		
	F_2[2)]	BC[3)]	DH/RI[4)]
M_hM_h	1	0.5	1
M_hm_h	0	−0.5	—
m_hm_h	−1	—	−1

Note: 1) Populations are derived from a cross between two inbred lines;

2) F_2 is derived by selfing or sib-mating F_1;

3) BC is derived by backcrossing F_1 with one of the parents;

4) DH is doubled haploid lines from F_1; RI is recombinant inbred lines obtained by selfing and random selecting for several consecutive generations starting from F_2.

Under a given significance threshold, small main effects or epistatic effects of QTLs are not expected to be detected, and thus their corresponding main-effect and interaction markers will be unidentifiable and missed from model (1). Therefore, in the process of stepwise regression analysis, the regression equation would become

$$\hat{y}_k = \hat{b}_0 + \sum_h \hat{b}_{M_h} \chi_{M_{hk}} + \sum_i \hat{b}_{MM_i} \chi_{MM_{ik}},$$
$$(h=1, \cdots, \phi;\ i=1, \cdots, \delta') \quad (2)$$

where ϕ is the number of identified main-effect markers with ($\phi \le \varphi$); δ' is the number of identified marker interactions with ($\delta' \le \varphi(\varphi-1)/2$).

The formulas for estimation and significance tests of the partial regression coefficients could be found in many books on statistics. But for easy comparisons, partial determination coefficient for each of the selected main-effect markers or interaction marker pairs can be calculated in a slightly different way:

$$R^2_{\chi|\Delta_{-\chi}} = R^2_\Delta - R^2_{\Delta_{-\chi}} \quad (3)$$

where $R^2_{\chi|\Delta_{-\chi}}$ is the relative contribution for the χth main-effect marker or marker interaction conditional on $\Delta_{-\chi}$, which represents all selected main-effect and interaction markers except the χth main-effect marker or marker interaction; R^2_Δ is the general determination coefficient for all the selected main-effect markers and marker interactions (Δ); $R^2_{\Delta_{-\chi}}$ is the general determination coefficient for $\Delta_{-\chi}$.

3 Simulation Studies

Monte Carlo simulations were conducted in order to understand the accuracy, power and resolution of multiple regression in identifying interaction markers that represent epistatic QTLs, and impacts of factors such as trait heritability, sample size, significance level, etc. on the identification.

3.1 Methods for simulations

A doubled haploid (DH) population was used in the simulation studies. 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 had four QTLs with a different linkage relationship among the QTLs (Fig. 1). The digenic additive (additive epistatic effects ranged from -1.71 to 1.31 units (Table 2). For easy description, epistasis of Q_1 vs. Q_2, Q_1 vs. Q_3, Q_1 vs. Q_4, Q_2 vs. Q_3, Q_2

written as

$$y_k = b_0 + \sum_h b_{M_h} \chi_{M_{hk}} + \sum_{i<j}\sum b_{MM_{ij}} \chi_{MM_{ijk}} + \varepsilon_k \quad (1)$$

where y_k is the phenotypic value of individual k ($k=1, \cdots, n$); b_0 is the intercept of the regression model, b_{M_h} is the partial regression coefficient for the hth main-effect marker ($h = 1, \cdots, \varphi$); $b_{MM_{ij}}$ is the partial regression coefficient for marker interaction between marker M_i and marker M_j ($i = 1, \cdots, \varphi-1$; $j = i+1, \cdots, \varphi$); indicator variable $\chi_{M_{hk}}$ depends on the observed genotype and the population type (Table 1); $\chi_{MM_{ijk}} = \chi_{M_{ik}} \chi_{M_{jk}}$; $\varepsilon_k \sim N(0, \sigma_\varepsilon^2)$; is the random residual composed of random error and a small portion of genetic effects that are not absorbed by the main-effect markers and interaction markers.

Table 1 Indicator variable ($\chi_{M_{hk}}$) for main-effect markers in multiple-regression model

Marker Genotype	Population type[1)]		
	F_2[2)]	BC[3)]	DH/RI[4)]
M_hM_h	1	0.5	1
M_hm_h	0	−0.5	—
m_hm_h	−1	—	−1

Note: 1) Populations are derived from a cross between two inbred lines;
2) F_2 is derived by selfing or sib-mating F_1;
3) BC is derived by backcrossing F_1 with one of the parents;
4) DH is doubled haploid lines from F_1; RI is recombinant inbred lines obtained by selfing and random selecting for several consecutive generations starting from F_2.

Under a given significance threshold, small main effects or epistatic effects of QTLs are not expected to be detected, and thus their corresponding main-effect and interaction markers will be unidentifiable and missed from model (1). Therefore, in the process of stepwise regression analysis, the regression equation would become

$$\hat{y}_k = \hat{b}_0 + \sum_h \hat{b}_{M_h} \chi_{M_{hk}} + \sum_i \hat{b}_{MM_i} \chi_{MM_{ik}}, \quad (h=1, \cdots, \phi;\ i=1, \cdots, \delta') \quad (2)$$

where ϕ is the number of identified main-effect markers with ($\phi \leq \varphi$); δ' is the number of identified marker interactions with ($\delta' \leq \varphi(\varphi-1)/2$).

The formulas for estimation and significance tests of the partial regression coefficients could be found in many books on statistics. But for easy comparisons, partial determination coefficient for each of the selected main-effect markers or interaction marker pairs can be calculated in a slightly different way:

$$R^2_{\chi|\Delta_{-\chi}} = R^2_\Delta - R^2_{\Delta_{-\chi}} \quad (3)$$

where $R^2_{\chi|\Delta_{-\chi}}$ is the relative contribution for the χth main-effect marker or marker interaction conditional on $\Delta_{-\chi}$, which represents all selected main-effect and interaction markers except the χth main-effect marker or marker interaction; R^2_Δ is the general determination coefficient for all the selected main-effect markers and marker interactions (Δ); $R^2_{\Delta_{-\chi}}$ is the general determination coefficient for $\Delta_{-\chi}$.

3 Simulation Studies

Monte Carlo simulations were conducted in order to understand the accuracy, power and resolution of multiple regression in identifying interaction markers that represent epistatic QTLs, and impacts of factors such as trait heritability, sample size, significance level, etc. on the identification.

3.1 Methods for simulations

A doubled haploid (DH) population was used in the simulation studies. 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 had four QTLs with a different linkage relationship among the QTLs (Fig. 1). The digenic additive (additive epistatic effects ranged from -1.71 to 1.31 units (Table 2). For easy description, epistasis of Q_1 vs. Q_2, Q_1 vs. Q_3, Q_1 vs. Q_4, Q_2 vs. Q_3, Q_2

ance is then calculated as

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

The genotypic value G_k of k-th individual was obtained by summing additive and epistatic effects of all the QTLs according to the QTL genotypes of individuals generated in genotyping step. 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.

Various combinations of sample size (200, 400, and 600), trait heritability (0.20, 0.50, and 0.80), and significance level (0.05, 0.01, 0.005, and 0.001) were used to examine the properties of stepwise regression in identifying interaction markers under three QTL linkage relationships (genome A, B, C). For each case, simulations were replicated 200 times, and average results were presented.

3.2 Rationality of stepwise regression method in identifying interaction markers

Fig. 2 shows frequency distributions of identified interaction markers associated with the preset QTLs for a specific set of simulation conditions (heritability = 0.50, sample size = 200, and significance level = 0.005). Results from other combinations of simulation conditions showed the similar tendency (data not shown).

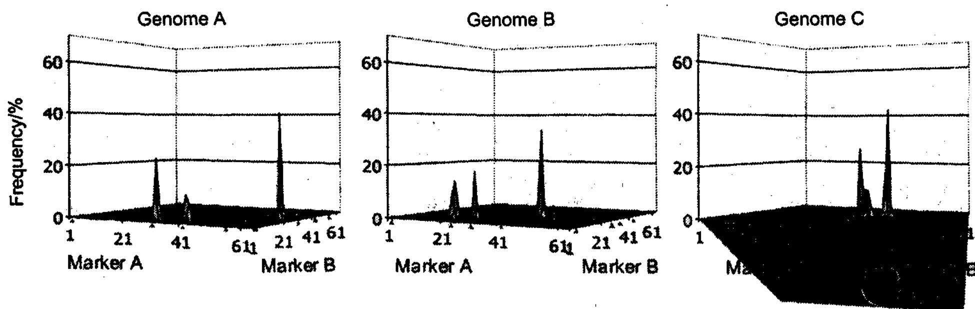

Fig. 2　Frequency distribution of interaction markers identified by stepwise regression along the whole genomes. For simplification, we did not separate the four chromosomes (markers 1-18 on chromosome Ⅰ; 19-33 on chromosome Ⅱ; 34-48 on chromosome Ⅲ; 49-64 on chromosome Ⅳ). Simulation conditions were: sample size = 200, heritability = 0.50, and significance level = 0.005.

Interaction markers associated with QTLs of large main and/or epistatic effects (relative contribution R^2_{QQ} to phenotypic variation >3%) were correctly identified by stepwise regression at reasonably high frequencies. In all the three genomes, the identified marker interactions were most often among four interactions between two pairs of markers flanking the intervals where the QTLs are located, and the marker pairs more closely linked to the QTLs were identified at higher frequencies. For example, in genome A, the interaction between M_{32} and M_{58} was detected 62.5% of the time by the four interactions between (M_{32}, M_{33}) and (M_{57}, M_{58}) that defined the two target intervals for QTLs #2 and #4. The power of detecting individual interacting QTL pairs depended on the distances between the markers and their linked QTLs in the relevant intervals.

Marker interactions representing small or zero epistatic effects (R^2_{QQ} < 3%) were almost unidentifiable by stepwise regression under the selected significance thresholds, and nor for the markers loosely linked to the epistatic QTLs. Close linkage between QTLs had negative impacts on identification of interaction markers. For example, interaction markers representing epistasis #1 and #2, could not be well resolved

in genome C. Therefore, it is not expected that stepwise regression can resolve tightly linked epistatic QTLs.

3.3 Comparison between two-way ANOVA and stepwise regression method

Table 2 shows the results of using stepwise regression and two-way ANOVA in identifying interaction markers. In two-way ANOVA, a wide range of marker pairs representing a single case of epistasis could be detected. Although the pair with the largest F value was the best representative for a specific epistasis, this largely reduced the resolution of identifying correct interaction markers. In every case, stepwise regression method was much more powerful than two-way ANOVA. On average, the frequency of detection by stepwise regression was nearly three times as that by two-way ANOVA in the simulations. This implies that using stepwise regression would possibly identify much more epistasis than using two-way ANOVA.

3.4 Factors affecting identification of interaction markers by stepwise regression

We also studied the impacts of heritabilities, sample sizes, and significance levels for F tests, as well as contribution of individual epistasis and linkage relations between QTLs, on power (or frequency) of identifying correct interaction markers using stepwise regression method (Fig. 3).

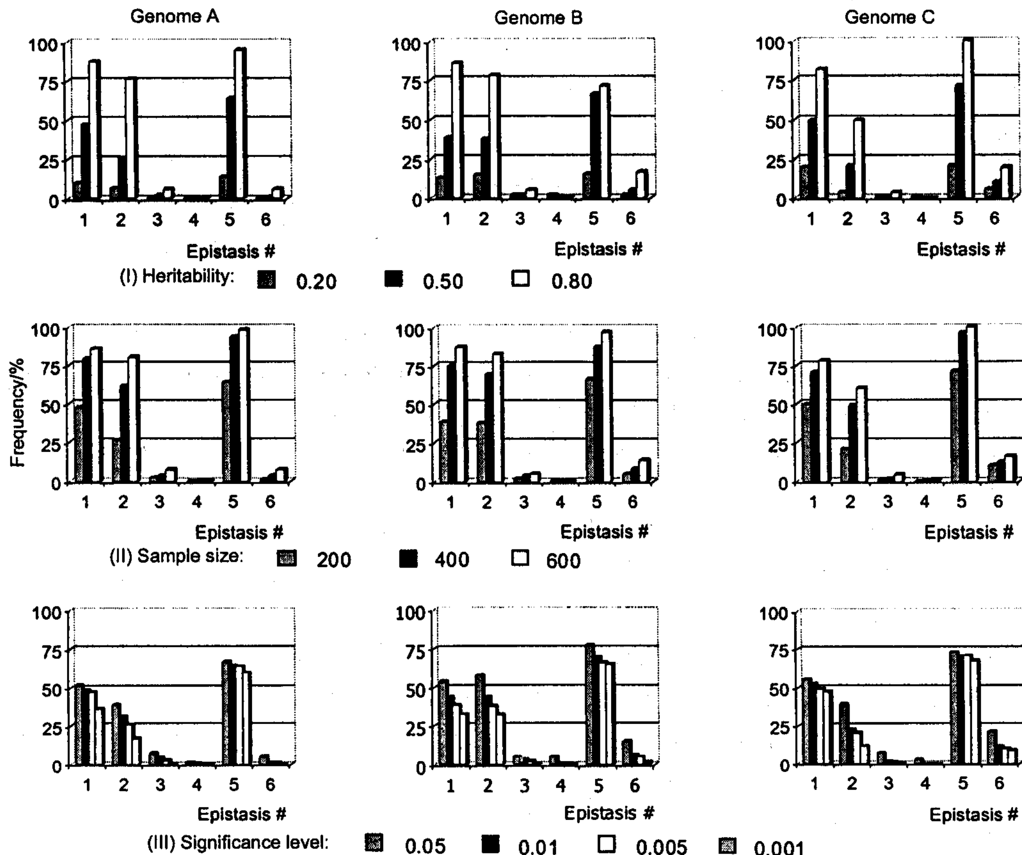

Fig. 3 Impacts of heritabilities (row Ⅰ), sample sizes (row Ⅱ), and significance levels (row Ⅲ), as well as linkages between QTLs and contributions of individual epistasis, on identification of interaction markers, Epistasis # is used here to indicate the marker interaction for each epistasis. The frequency for each case was obtained as a summation for four interactions between two pairs of markers constructing the intervals where the two TQLs involved in an epistasis were located.

Larger heritabilities and sample sizes invariably increased powers of identified interaction markers. The averaged frequencies across all the three genomes were 7.3%, 24.6%, and 43.7% for heritabilities of 0.20, 0.50, and 0.80, respectively, under a sample size of 200 and a significance level of 0.005. The averaged frequencies were 24.6%, 39.8%, and 46.5 % for sample sizes 200, 400, and 600, respectively, under a heritability of 0.50 and significance level of 0.005.

The impacts of heritabilities and sample sizes on detection power were mostly associated with relative contribution of individual QTL epistasis to phenotypic variation. Table 3 showed that larger epistatic effects (epistasis #1, #2, and #5 in all the three genomes) contributed more to phenotypic variation and so their corresponding marker interactions were identified at higher power, whereas small or zero epistatic effects (epistasis #3, #4 and #6) were almost unidentified.

Table 3 Partial regression coefficients and relative contributions of interaction markers identified by stepwise regression[1)]

Genome	Epist. #[2)]	QTL Epist.		Marker int.[4)]		Power[3)] (%)
		b_{QQ}	R^2_{QQ} (%)	$\hat{b}_{MM}$	$\hat{R}^2_{MM}$ (%)	
A	1	1.31	3.50	1.43	4.15	47.5
	2	1.08	2.38	1.40	3.98	26.0
	3	−0.38	0.29	−1.19	2.74	2.5
	4	0.00	0.00	−1.08	2.21	0.0
	5	−1.71	5.97	−1.64	5.60	64.0
	6	−0.41	0.34	−1.45	4.12	1.0
B	1	1.31	3.72	1.51	4.82	38.5
	2	1.08	2.53	1.48	4.60	38.0
	3	−0.38	0.31	−1.26	3.31	2.0
	4	0.00	0.00	1.08	2.46	0.5
	5	−1.71	6.34	−1.68	6.01	66.0
	6	−0.41	0.36	−1.19	2.95	5.0
C	1	1.31	4.19	1.70	5.73	49.5
	2	1.08	2.85	1.61	5.64	20.5
	3	−0.38	0.35	−1.43	3.92	1.0
	4	0.00	0.00	0.00	0.00	0.0
	5	−1.71	7.14	−1.74	7.00	71.0
	6	−0.41	0.41	−1.59	4.64	10.0

Note: 1), 2) and 3) are the same as notes 1), 2) and 3) in Table 2, respectively; 4) and for each epistasis were weighted averages of four pairs of marker interactions that best represent the epistasis. Simulation conditions are heritability = 0.50, sample size = 200, and significance level = 0.005.

Close linkages (here genome C) between epistatic QTLs could distort the power of identifying correct interaction markers (Fig. 3 and Table 3). Compared with independent inheritance case (genome A), some interaction markers (e.g. epistasis #5) were identified at higher frequency, some (e.g. epistasis #2) were detected at lower power, and others (e.g. epistasis #3) were identified at power similar to that for independent inheritance case. These were largely due to the mutual cancellation or enhancement of relevant epistatic effects resulting from close linkages between epistatic QTLs.

Larger values of significance levels could lead to higher frequencies of identified interaction markers. On average across the three genomes, the powers were 32.4%, 26.4%, 24.6%, and 21.2% for significance levels 0.05, 0.01, 0.005, and 0.001, respectively, under sample size = 200 and heritability = 0.50. However, higher power from larger value was always associated with lower correctness of identified interaction markers. For example, the averaged power of identification was increased by only 11.2% with significance levels changed from 0.001 to 0.05, but the correctness was decreased by 41.4%, according to the proportion of corrected marker interactions in all identified interactions in 200 simulations. Sample sizes and heritabilities had certain influences on such loss of correctness. The smaller the samples size and/or heritability was, the greater the loss would be for a given significance level, and vice versa. This implied that larger significance level could be taken for certain correctness under larger sample size and/or higher heritability.

3.5 Partial regression coefficients and relative contributions

Although partial regression coefficients ($\hat{b}_{MM_h}$) are generally considered as underesti-

mates of QTL epistatic effects, it would also be helpful to explore relationships between $\hat{b}_{MM_h}$ of interaction markers and the parameters of QTL epistatic effects, as well as associations between relative contributions ($\hat{R}^2_{MM}$) of interaction markers and contributions ($\hat{R}^2_{QQ}$) of QTL epistasis. Simulation results were summarized in Table 3 for simulation condition with heritability = 0.50, sample size = 200, and significance level = 0.005.

It was indicated that the estimates ($\hat{b}_{MM}$) of partial regression coefficients could partly reflect the magnitudes and directions of QTL epistatic effects, particularly for epistasis with relatively large R^2_{QQ} to phenotypic variation (such as epistasis #1, #2, and #5). But it was not true for small or zero epistasis (epistasis #3, #4 and #6). This was also the case for partial determination coefficient ($\hat{R}^2_{MM}$). However, $\hat{b}_{MM}$ and $\hat{R}^2_{MM}$ for identified marker interactions were not always underestimates of corresponding parameters of QTL epistasis as we usually expect, and on the contrary, they could be frequently overestimates, particularly for small-effect epistasis. This was because $\hat{b}_{MM}$ and $\hat{R}^2_{MM}$ were obtained based on a small portion of larger estimates that remained after the F tests among all possible estimates.

Linkages between epistatic QTLs also had influences on the relation between estimates ($\hat{b}_{MM}$ and $\hat{R}^2_{MM}$) of marker interactions and parameters of QTL epistasis, as shown in Table 3.

4 Discussion

With increased researches in QTL mapping, QTL epistasis is attracting more and more attentions of geneticists[14]. However, how to effectively analyze epistasis remains a problem to be solved. In this study, stepwise regression has been proven to be a powerful statistical method for identifying correct interaction markers that represent epistatic QTLs.

The results indicated that stepwise regression method was much better than two-way ANOVA, as was the case for identifying main-effect markers (data not shown). This higher power of stepwise regression originated from its effective control of background genetic variation caused by segregating QTLs other than the QTLs studied. If it were not controlled, this portion of variation would be pooled into random residuals, and would largely reduce the power and resolution in identifying correct interaction markers. Stepwise regression maximized the absorption of the effects of background QTLs by keeping all significant main-effect markers and interaction markers in the model, and thus realized effective control of background genetic variation when selecting new interaction markers.

Many factors affected the power of identifying correct marker interactions, as shown in the study. To increase the power, we could take a larger sample size and/or better control of environmental errors, but we can do nothing about the factors (linkage between QTLs, contributions of individual epistasis, etc.) related to genetic properties of epistasis. Significance level would be a difficult matter to be decided during implementation of stepwise regression. Larger values of significance level could result in more correct marker interactions identified, but also caused high chance of false positives. Smaller values could increase the reliability of identified interaction markers, but would also lose some power. To minimize such dilemma, it is suggested that several rounds of analyses could be conducted under different significance levels, and comparisons of results from the analyses be made. The repeatability of a specific identified marker interaction across all of the rounds would be measurement of the reliability (or correctness) of the marker interaction. On the other hand, less stringent significance level

could be taken for a larger sample size and/or higher heritability, since in these circumstances, random errors are better controlled.

Better understanding of QTL epistasis should be obtained from directly mapping epistatic QTLs. However, genome-wide search for epistasis would need tremendously large amount of computational work. Although correctly identified interaction markers usually do not locate exactly at the QTLs involved in the epistasis, they do provide an indication of the rough region where the epistatic QTLs are possibly located and the relative importance of the epistasis. This is important not only for genetic improvement of economically important quantitative traits, but also for further mapping epistatic QTLs. With identified interaction markers, we could focus on genomic regions revealed by the interaction markers so as to dramatically reduce the work needed for mapping epistatic QTLs along whole genome.

References:

[1] Paterson A H, *Molecular dissection of complex traits* [M]. Boca Raton, florida: CRC Press LLC, Florida, 1997.

[2] Lander E S, Botstein D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps[J]. *Genetics*, 1989, 121: 185-199.

[3] Jansen R C. Interval mapping of multiple quantitative trait loci[J]. *Genetics*, 1993, 135: 205-211.

[4] Zeng Z B. Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci[J]. *Proc Natl Acad Sci USA*, 1993, 90: 10972-10976.

[5] Zeng Z B. Precision mapping of quantitative trait loci[J]. *Genetics*, 1994, 136: 1457-1468.

[6] Sax K. The association of size differences with seed-coat pattern and pigmentation in Phaseolus vulgaris[J]. *Genetics*, 1923, 8: 552-560.

[7] Thoday J M. Location of polygenes[J]. *Nature*, 1961, 191: 368-370.

[8] Haley C S, Knott S A. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers[J]. *Heredity*, 1992, 69: 315-324.

[9] Spickett S G, Thoday J M. Regular response to selection 3. Interaction between located polygenes[J]. *Genet Res*, 1966, 7: 96-121

[10] Falconer D S. *Introduction to Quantitative Genetics*[M], 2nd ed. New York: Longman Press, 1981.

[11] Mather K, Jinks J L. *Biometrical genetics*[M], 3rd ed. London: Chapman and Hall, 1982.

[12] Pooni H S, Coombs D J. Jinks P S. Detection of epistatsis and linkage of interacting genes in the presence of reciprocal differences[J]. *Heredity*, 1987, 58: 257-266.

[13] Allard R W. Future directions in plant population genetics, evolution and breeding [A]. Brown AHD, Clegg MT, Kahler AL, *et al. Plant Population Genetics and Germplasm Resources* [C], Sunderland, MA: Sinauer Associates Inc, 1988. 1-19.

[14] Li Z K. Molecular analysis of epistasis [A], Paterson AH. *Molecular Dissection of Complex Traits*[C], Boca Raton, Florida: CRC Press LLC, 1997. 119-130.

[15] Xiao J, Li J, Yuan L, *et al.* Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers[J]. *Genetics*, 1995, 140: 745-754.

[16] Wu P, Zhang G, Huang N, *et al.* Non-additive interaction conditioning spikelet sterility in an F_2 population of *indica/japonica* cross in rice [J]. *Theor Appl Genet*, 1995, 91: 825-829.

[17] Li Z K, Pinson S R M, Park W D, *et al.* Epistasis for three grain yield components in rice (*Oryza sativa* L.) [J]. *Genetics*, 1997, 145: 453-465.

[18] Li Z K, Pinson S R M, Paterson A H. Genetics of hybrid sterility and hybrid breakdown in an inter-subspecific rice (*Oryza sativa* L.) population[J]. *Genetics*, 1997, 145: 1139-1148.

[19] Yu S B, Li J X, Xu C G, *et al.* Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid[J]. *Proc Natl Acad Sci USA*, 1997, 94: 9226-9231.

[20] Hackett C A, Van Ooijen J W, Jansen J. Selection of markers linked to quantitative trait loci by regression techniques[A], Van Ooijen J W, Jansen J. *Biometrics in plant breeding: applications of molecular markers* [C], the Netherlands: Wageningen, 1994. 99-106.

[21] Haldane J B S. The combination of linkage values and the calculation of distance between the loci of linked factors [J]. *J Genet*, 1919, 8: 299-309.

(责任编辑　杜玲玲)