JIPB Journal of Integrative Plant Biology

Journal of Integrative Plant Biology 2012, **54** (4): 270–279

Research Article

The Statistical Power of Inclusive Composite Interval Mapping in Detecting Digenic Epistasis Showing Common F_2 Segregation Ratios

Luyan Zhang, Huihui Li and Jiankang Wang[1*]

Institute of Crop Sciences, The National Key Facility for Crop Gene Resources and Genetic Improvement, and CIMMYT China, Chinese Academy of Agricultural Sciences, Beijing 100081, China

[*] Corresponding author

Tel: +86 10 8210 5846; Fax: +86 10 8210 8572; Email: wangjk@caas.net.cn

Articles can be viewed online without a subscription.

Available online on 16 February 2012 at www.jipb.net and www.wileyonlinelibrary.com/journal/jipb

doi: 10.1111/j.1744-7909.2012.01110.x

Abstract

Epistasis is a commonly observed genetic phenomenon and an important source of variation of complex traits, which could maintain additive variance and therefore assure the long-term genetic gain in breeding. Inclusive composite interval mapping (ICIM) is able to identify epistatic quantitative trait loci (QTLs) no matter whether the two interacting QTLs have any additive effects. In this article, we conducted a simulation study to evaluate detection power and false discovery rate (FDR) of ICIM epistatic mapping, by considering F_2 and doubled haploid (DH) populations, different F_2 segregation ratios and population sizes. Results indicated that estimations of QTL locations and effects were unbiased, and the detection power of epistatic mapping was largely affected by population size, heritability of epistasis, and the amount and distribution of genetic effects. When the same likelihood of odd (LOD) threshold was used, detection power of QTL was higher in F_2 population than power in DH population; meanwhile FDR in F_2 was also higher than that in DH. The increase of marker density from 10 cM to 5 cM led to similar detection power but higher FDR. In simulated populations, ICIM achieved better mapping results than multiple interval mapping (MIM) in estimation of QTL positions and effect. At the end, we gave epistatic mapping results of ICIM in one actual population in rice (*Oryza sativa* L.).

Keywords: Epistasis; false discovery rate; inclusive composite interval mapping; power analysis; simulation study.

Zhang L, Li H, Wang J (2012) The statistical power of inclusive composite interval mapping in detecting digenic epistasis showing common F_2 segregation ratios. *J. Integr. Plant Biol.* **54**(4), 270–279.

Introduction

Epistasis refers to interaction between two or more non-allelic genes. A number of experiments indicated the significant role of epistasis and its importance for complex traits (Lark et al. 1995; Holland et al. 1997; Yu et al. 1997; Ohno et al. 2000; Nagase et al. 2001; Boer et al. 2002; Luo et al. 2009). In evolution and breeding, epistasis can also maintain the additive variance, which may be the basis of long-term genetic gain (Bernardo 2002; Janick 2004; Zhai and Wang 2007). However, the principle of how interacting genes influenced the quantitative traits largely remains unclear until now. It is generally agreed that the identification and estimation of epistasis are much more difficult and complicated, compared with additive and dominance.

Inclusive Composite Interval Mapping (ICIM) is a critical step forward that highlights the importance of model selection and interval testing in quantitative trait loci (QTLs) linkage mapping (Li et al. 2007; Zhang et al. 2008; Wang 2009). In addition, ICIM can be readily extended to epistasis mapping by simultaneously considering marker variables and marker-pair multiplications in a linear model (Li et al. 2008). Two steps of stepwise regression were adopted to identify the most significant markers and marker-pair multiplications. Then a two dimensional interval mapping was conducted to detect epistatic QTL using the adjusted phenotypic values based

© 2012 Institute of Botany, Chinese Academy of Sciences

on the best multiple regression models. Epistatic QTL can be identified by ICIM, no matter whether the two interacting QTLs have any additive effects. Li et al. (2010) investigated the statistical properties of ICIM for additive mapping through simulation. Results indicated that the increase in marker density helps ICIM identify independent QTLs explaining more than 5% of phenotypic variance. However, only large-size populations can take advantage of densely distributed markers. When population size is greater than 200, ICIM achieves unbiased estimations of QTL position and effect. However, the property of ICIM for epistatic mapping is still unclear.

Our objectives in this study were (i) to evaluate the efficiency of ICIM epistatic mapping in F_2 and DH populations through computer simulation and an actual population, and (ii) to evaluate the effects of population size on the detection power of different types of epistasis represented by F_2 segregation ratios.

Results

QTL effects and genetic variance in F_2 and DH populations

Fourteen commonly observed F_2 segregation ratios and their genotypic values were shown in **Table 1**, and genetic effects of QTL for the 14 F_2 segregation ratios were calculated in **Table 2**. Most ratios had non-zero additive and dominance effects and epistatic effects except two segregation ratios. When the segregation ratio was 9:3:3:1, the two QTLs had no epistatic effects. When the segregation ratio was 10:6, QTLs had no additive and dominance effects.

Total genetic variance was decomposed into additive, dominance and epistatic variance components to demonstrate the importance of epistasis in each segregation ratio (**Table 3**). The proportion of additive, dominance and epistatic variance varied significantly among the 14 ratios (**Table 3**). Taking segregation ratio 9:3:3:1 and F_2 population for example, the genetic variance (V_G) was 0.94 including additive variance (V_A) 0.63 and dominance variance (V_D) 0.31. Epistatic variance was 0, indicating epistasis had no contribution to the genetic variance in the 9:3:3:1 ratio. For the 6:9:1 ratio, $V_G = 0.35$, where $V_A = 0.02$, $V_D = 0.01$, and $V_I = 0.32$, indicating epistasis had great contribution to the genetic variance. The heritability of epistasis was denoted as H_I whose value was equal to $V_I/(V_G + V_\varepsilon)$, where V_ε is the random error. We assumed heritability of traits was 0.6 in simulation, and then H_I can be calculated (**Table 3**). However, the same segregation ratio may lead to different epistatic variance in F_2 and DH populations (**Table 3**). The largest H_I of value 0.6 was in DH population for segregation ratio 10:6.

Detection power of epistasis mapping in F_2 and DH populations

From the mapping results on simulated populations with marker density 10 cM (**Figure1A** for F_2; **Figure1B** for DH), it can be seen that the detection power of epistasis mapping increased with the increase in population size for both F_2 and DH populations. Most QTLs had high detection power when population size was larger than 300. Taking segregation ratio 9:7 for example, detection powers of QTL in F_2 population were 59, 97 and 100% for population sizes 100, 200, and 300, respectively. Powers in DH population were 5, 30, 53,

Table 1. Genotypic values for 14 segregation ratios commonly observed in F_2 populations

	Genotype and genotypic values									
Segregation ratio in F_2	AABB	AABb	AAbb	AaBB	AaBb	Aabb	aaBB	aaBb	aabb	Segregation ratio in DH
9:3:3:1	3	3	2	3	3	2	1	1	0	1:1:1:1
9:3:4	2	2	1	2	2	1	0	0	0	1:1:2
12:1:3	2	2	2	2	2	2	0	0	1	2:1:1
3:9:4	1	1	2	1	1	2	0	0	0	1:1:2
12:3:1	2	2	2	2	2	2	1	1	0	2:1:1
9:7	1	1	0	1	1	0	0	0	0	1:3
3:13	0	0	1	0	0	1	0	0	0	3:1
9:4:3	2	2	1	2	2	1	0	0	1	1:2:1
9:1:6	2	2	0	2	2	0	0	0	1	1:2:1
10:3:3	2	2	1	2	2	1	0	0	2	2:1:1
15:1	1	1	1	1	1	1	1	1	0	3:1
3:12:1	1	1	2	1	1	2	1	1	0	2:1:1
10:6	1	1	0	1	1	0	0	0	1	1:1
6:9:1	1	1	2	1	1	2	2	2	0	1:2:1

The last column is the ratio in a DH population if the same genotypic values applied.

82 and 90% for population sizes 100, 200, 300, 400, and 500, respectively.

For the same population size, QTLs with larger proportion of epistatic variance had higher detection power. Taking population size 300 for example, H_I were 0.05, 0.06 and 0.09 in F_2 for segregation ratios 3:9:4, 12:3:1 and 9:7, and their detection powers were 54, 59 and 100%, respectively; H_I were 0.2, 0.3 and 0.6 in DH for segregation ratio 15:1, 3:12:1 and 10:6, and their detection powers were 51, 74 and 100%, respectively.

The amount and distribution of genetic effects may also have an effect on the detection power, if the heritability is fixed. For example, H_I was equal to 0.14 for ratios 3:13 and 9:4:3 in F_2, but the corresponding powers were different especially when population size was lower than 200. When population size was 100, detection powers were 56 and 76% for ratios 3:13 and 9:4:3; when population size was 200, powers were 94 and 99%, respectively. The reason may be QTLs had only dominance effects for ratio 9:4:3, while both additive and dominance effects were present for ratio 3:13. Fewer effect parameters may simplify the genetic model and result in more precision in QTL mapping. Epistasis effects for segregation ratio 9:4:3 were larger than those for ratio 3:13, resulting in higher probability of QTL detection for ratio 9:4:3.

Table 2. Genetic effects of two interacting quantitative trait loci (QTLs) for the 14 segregation ratios

		Genetic effects of QTL							
Segregation ratio in F_2	Mean (*m*)	a_1	d_1	a_2	d_2	*aa*	*ad*	*da*	*dd*
9:3:3:1	1.5	1.00	1.00	0.50	0.50	0.00	0.00	0.00	0.00
9:3:4	0.75	0.75	0.75	0.25	0.25	0.25	0.25	0.25	0.25
12:1:3	1.25	0.75	0.75	−0.25	−0.25	0.25	0.25	0.25	0.25
3:9:4	0.75	0.75	0.75	−0.25	−0.25	−0.25	−0.25	−0.25	−0.25
12:3:1	1.25	0.75	0.75	0.25	0.25	−0.25	−0.25	−0.25	−0.25
9:7	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
3:13	0.25	0.25	0.25	−0.25	−0.25	−0.25	−0.25	−0.25	−0.25
9:4:3	1.00	0.50	0.50	0.00	0.00	0.50	0.50	0.50	0.50
9:1:6	0.75	0.25	0.25	0.25	0.25	0.75	0.75	0.75	0.75
10:3:3	1.25	0.25	0.25	−0.25	−0.25	0.75	0.75	0.75	0.75
15:1	0.75	0.25	0.25	0.25	0.25	−0.25	−0.25	−0.25	−0.25
3:12:1	1.00	0.50	0.50	0.00	0.00	−0.50	−0.50	−0.50	−0.50
10:6	0.50	0.00	0.00	0.00	0.00	0.50	0.50	0.50	0.50
6:9:1	1.25	0.25	0.25	0.25	0.25	−0.75	−0.75	−0.75	−0.75

Table 3. Genetic variance components for the 14 segregation ratios

	F_2						DH				
Segregation ratio in F_2	V_A	V_D	V_I	V_G	[a]V_ε	[b]H_I	V_A	V_I	V_G	V_ε	H_I
9:3:3:1	0.63	0.31	0.00	0.94	0.63	0.00	1.25	0.00	1.25	0.83	0.00
9:3:4	0.45	0.23	0.04	0.72	0.48	0.03	0.63	0.06	0.69	0.46	0.05
12:1:3	0.39	0.20	0.04	0.63	0.41	0.03	0.63	0.06	0.69	0.46	0.05
3:9:4	0.27	0.13	0.04	0.44	0.29	0.05	0.63	0.06	0.69	0.46	0.05
12:3:1	0.20	0.10	0.04	0.34	0.23	0.06	0.63	0.06	0.69	0.46	0.05
9:7	0.14	0.07	0.04	0.25	0.16	0.09	0.13	0.06	0.19	0.13	0.20
3:13	0.08	0.04	0.04	0.16	0.10	0.14	0.13	0.06	0.19	0.13	0.20
9:4:3	0.31	0.16	0.14	0.61	0.41	0.14	0.25	0.25	0.50	0.33	0.30
9:1:6	0.39	0.20	0.32	0.91	0.60	0.21	0.13	0.56	0.69	0.46	0.49
10:3:3	0.20	0.10	0.32	0.62	0.41	0.31	0.13	0.56	0.69	0.46	0.49
15:1	0.02	0.01	0.04	0.07	0.04	0.36	0.13	0.06	0.19	0.13	0.20
3:12:1	0.06	0.03	0.14	0.23	0.16	0.36	0.25	0.25	0.50	0.33	0.30
10:6	0.06	0.03	0.14	0.23	0.16	0.36	0.00	0.25	0.25	0.17	0.60
6:9:1	0.02	0.01	0.32	0.35	0.23	0.56	0.13	0.56	0.69	0.46	0.49

[a] The heritability is assumed to be at 0.60. V_ε is random error variance.

[b] H_I is the heritability of epistasis, whose value is $V_I/(V_G+V_\varepsilon)$.

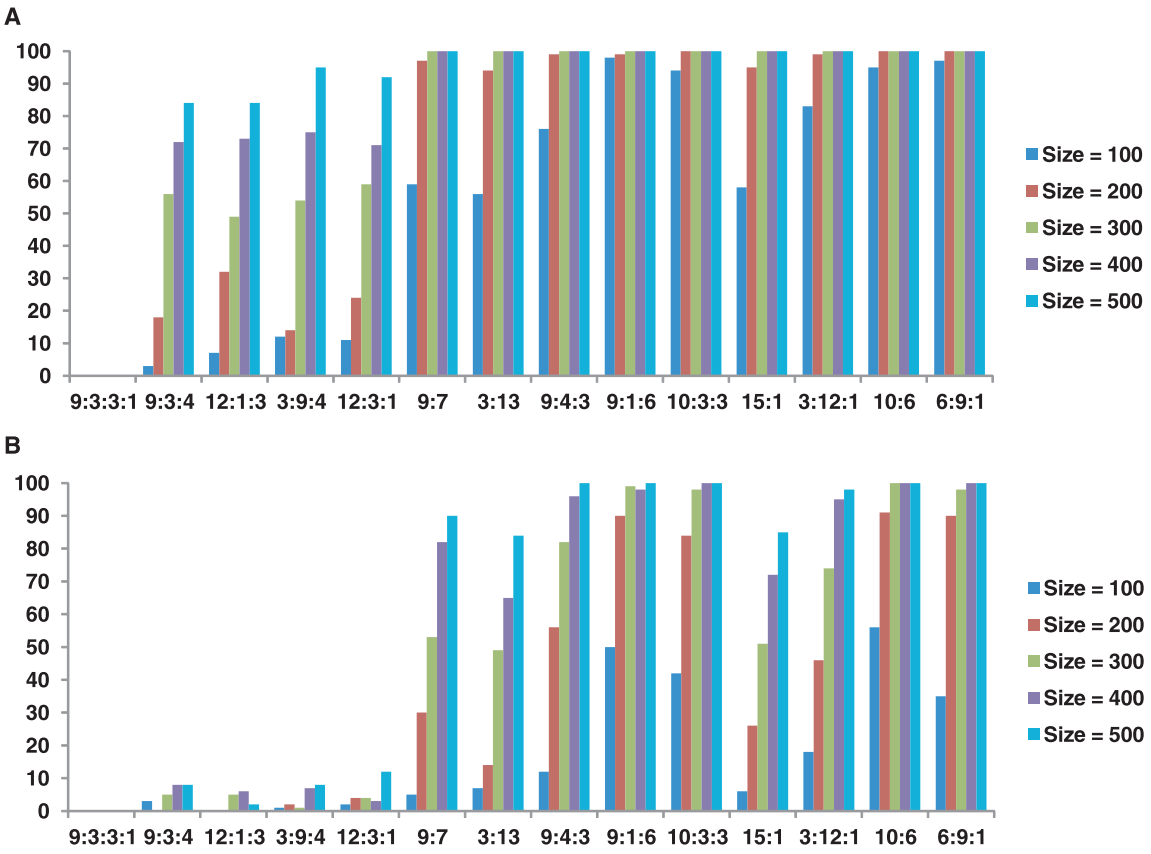

Figure 1. Power to detect interacting quantitative trait loci (QTLs) in F_2 and DH populations.

In conclusion, the detection power of interacting QTLs in epistatic mapping depends on population size, heritability of epistasis, and the amount and distribution of genetic effects. Compared with additive and dominance QTLs, epistatic mapping requires larger populations, especially for QTLs with smaller epistatic effects.

Estimated locations and genetic effects

To further understand the property of epistatic mapping, the estimated locations and genetic effects of interacting QTLs for population size 200 ares shown in **Table 4** (for F_2) and **Table 5** (for DH). For QTLs with power 0, no locations and effects were estimated, i.e. QTLs for segregation ratio 9:3:3:1, 9:3:4 and 12:1:3 in DH. For QTLs with small detection power, the estimated locations and effects were biased. For example in the DH population (**Table 5**), the power of QTLs was 2% for segregation ratio 3:9:4. The estimated positions of two interacting were 20.00 and 52.50, corresponding to the true positions 25 and 55. Estimated additive effects of two interacting QTLs and additive by additive effect were 0.06, −0.32 and −0.63, corresponding to the true effects 0.75, −0.25 and −0.25.

For QTLs whose detection power was not too small, the estimated locations and effects were almost unbiased. For example, in F_2 population (**Table 4**), power of QTL was 96% for ratio 3:13. The estimated positions of two interacting were 25.99 and 54.69, corresponding to the true positions 25 and 55. Estimated values of genetic effects a_1, d_1, a_2, d_2, *aa*, *ad*, *da* and *dd* were 0.23, 0.22, −0.22, −0.19, −0.24, −0.25, −0.25 and −0.24, corresponding to the true effects 0.25, 0.25, −0.25, −0.25, −0.25, −0.25, −0.25 and −0.25. Similar results were observed in the DH population (**Table 5**).

False discovery rate of epistasis mapping in F_2 and DH populations

False discovery rate (FDR) of the simulation study was shown in **Figure2** (**Figure2A** for F_2; **Figure2B** for DH). Generally speaking, FDR decreased with the increase in population size. For example, FDRs for segregation ratio 12:3:1 were 77.94, 52.38, 40.95, 34.21 and 28.13% for population sizes 100 to 500 in F_2; 83.33, 63.64, 33.33, 50 and 14.29% for population sizes 100 to 500 in DH, respectively. There were a few exceptions in our simulation study. For example, false discovery rates for segregation ratio 9:1:6 were 38.96, 32.89, 35.06, 27.54 and 40.83% for population sizes 100 to 500 in F_2, which may be caused by random errors in simulation.

Table 4. Estimated positions and genetic effects of quantitative trait loci (QTLs) in 100 simulated F_2 populations when population size is 200

Segregation ratio in F_2	Power (%)	[a]LOD	QTL1 position	QTL2 position	a_1	d_1	a_2	d_2	*aa*	*ad*	*da*	*dd*
9:3:3:1	1	5.44	15.00	70.00	−0.06	−0.42	0.11	−0.39	−0.6	0.32	−0.08	0.75
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
9:3:4	27	6.88	23.52	54.63	0.36	0.20	0.12	0.04	0.28	0.32	0.35	0.4
		(1.54)	(11.12)	(10.18)	(0.42)	(0.40)	(0.22)	(0.32)	(0.14)	(0.25)	(0.19)	(0.42)
12:1:3	35	6.89	23.43	56.57	0.53	0.39	−0.31	−0.27	0.33	0.32	0.35	0.36
		(1.59)	(7.15)	(8.76)	(0.32)	(0.34)	(0.09)	(0.21)	(0.11)	(0.19)	(0.17)	(0.32)
3:9:4	17	6.78	27.35	57.06	0.56	0.52	−0.04	−0.11	−0.33	−0.29	−0.28	−0.29
		(1.31)	(8.07)	(12.61)	(0.25)	(0.41)	(0.21)	(0.19)	(0.10)	(0.21)	(0.19)	(0.31)
12:3:1	30	7.31	28.00	53.50	0.59	0.62	0.25	0.35	−0.28	−0.3	−0.29	−0.5
		(1.59)	(10.69)	(8.86)	(0.26)	(0.28)	(0.13)	(0.21)	(0.10)	(0.18)	(0.23)	(0.32)
9:7	97	14.09	24.64	54.69	0.21	0.15	0.23	0.19	0.25	0.26	0.26	0.25
		(3.27)	(4.20)	(3.65)	(0.11)	(0.14)	(0.09)	(0.15)	(0.06)	(0.10)	(0.10)	(0.14)
3:13	96	14.31	25.99	54.69	0.23	0.22	−0.22	−0.19	−0.24	−0.25	−0.25	−0.24
		(3.42)	(6.44)	(4.13)	(0.06)	(0.09)	(0.11)	(0.15)	(0.07)	(0.09)	(0.10)	(0.13)
9:4:3	99	18.94	24.80	55.81	0.47	0.39	−0.04	−0.07	0.48	0.48	0.51	0.51
		(4.35)	(3.69)	(4.91)	(0.18)	(0.26)	(0.12)	(0.15)	(0.09)	(0.14)	(0.15)	(0.20)
9:1:6	100	26.89	25.00	54.95	0.20	0.17	0.23	0.17	0.72	0.74	0.73	0.72
		(5.54)	(2.83)	(2.96)	(0.17)	(0.26)	(0.15)	(0.24)	(0.12)	(0.15)	(0.16)	(0.22)
10:3:3	100	27.08	25.00	55.05	0.22	0.19	−0.25	−0.25	0.73	0.74	0.72	0.73
		(5.51)	(2.65)	(3.12)	(0.17)	(0.24)	(0.12)	(0.14)	(0.11)	(0.17)	(0.18)	(0.20)
15:1	96	14.84	25.47	55.47	0.22	0.23	0.23	0.23	−0.25	−0.23	−0.25	−0.26
		(3.64)	(5.95)	(6.81)	(0.08)	(0.09)	(0.08)	(0.09)	(0.07)	(0.1	(0.09	(0.14
3:12:1	100	19.70	25.85	55.50	0.451	0.47	0.03	0.04	−0.49	−0.5	−0.5	−0.51
		(4.19)	(5.57)	(3.04)	(0.12)	(0.15)	(0.12)	(0.14)	(0.10)	(0.16)	(0.12)	(0.21)
10:6	100	30.72	25.20	54.70	−0.03	−0.03	−0.02	−0.04	0.47	0.47	0.48	0.49
		(6.13)	(1.86)	(3.73)	(0.11)	(0.12)	(0.11)	(0.12)	(0.07)	(0.11)	(0.09)	(0.14)
6:9:1	100	28.01	24.90	55.05	0.26	0.22	0.23	0.27	−0.73	−0.74	−0.7	−0.74
		(5.18)	(2.83)	(2.18)	(0.12)	(0.14)	(0.11)	(0.16)	(0.11)	(0.18)	(0.15)	(0.23)

[a] The value in the brackets is *SD*.

It can be seen from Figures 1 and 2 that detection power of QTL was higher in the F_2 population than the power in the DH population; meanwhile, FDR in F_2 was also higher than FDR in DH. Taking segregation ratio 15:1 as an example, detection powers were 58, 95, 100, 100, and 100% in F_2 for population sizes 100 to 500; powers were 6, 26, 51, 72 and 85% in DH. The corresponding FDR were 64.41, 44.51, 42.53, 32.43 and 35.90% in F_2, and 66.67, 10.34, 5.56, 1.37 and 2.30% in DH. The reason may be that QTLs in F_2 have more parameters of genetic effects than DH, which lead to higher LOD score for the same segregation ratio. Then if we used the same LOD threshold for F_2 and DH, the peaks may be regarded as significant QTLs more easily in F_2 than DH. This may cause two possible results. One is that true QTLs are detected, which increases the detection power; the other is that false positive occurs, which increases FDR.

Comparison of ICIM with multiple interval mapping using simulated populations

For the first simulated DH population for each ratio, interacting QTLs were detected by ICIM in eight of the 14 ratios, while interacting QTLs were detected by MIM in only two ratios (Table 6). The estimated positions of multiple interval mapping (MIM) were more biased than ICIM. Taking segregation ratio 15:1 for example, the positions of interacting QTLs from ICIM were at 25 cM on chromosome 1 and 55 on chromosome 2, which were the same as the true positions. Meanwhile, the positions of interacting QTLs from MIM were at 30 cM on chromosome 1 and 50 on chromosome 2. The estimated effects of both ICIM and MIM were close to true values. In addition, MIM could detect interacting QTLs, which also had main effects, but ICIM could also detect interacting QTLs, which only had epistatic effects. For example, QTLs in segregation ratio 10:6

Table 5. Estimated positions and genetic effects of quantitative trait loci (QTLs) in 100 simulated DH populations when population size is 200

Segregation ratio in F_2	Power (%)	[a]LOD	QTL1 position	QTL2 position	a_1	a_2	aa
9:3:3:1	0						
9:3:4	0						
12:1:3	0						
3:9:4	2	6.87	20.00	52.50	0.06	−0.32	−0.63
		(0.50)	(10.00)	(2.50)	(0.02)	(0.06)	(0.03)
12:3:1	4	5.69	20.00	53.75	0.33	0.17	−0.59
		(0.41)	(9.35)	(4.15)	(0.33)	(0.09)	(0.02)
9:7	30	6.88	24.17	56.67	0.16	0.14	0.35
		(1.58)	(7.20)	(8.88)	(0.10)	(0.15)	(0.04)
3:13	14	6.41	23.57	56.43	0.16	−0.13	−0.34
		(1.46)	(11.09)	(6.66)	(0.11)	(0.14)	(0.03)
9:4:3	56	7.21	24.02	55.63	0.37	−0.01	0.60
		(1.74)	(7.03)	(7.74)	(0.19)	(0.08)	(0.07)
9:1:6	90	7.94	25.11	54.61	0.21	0.20	0.76
		(2.03)	(5.87)	(5.39)	(0.14)	(0.15)	(0.11)
10:3:3	84	8.05	24.64	55.00	0.23	−0.20	0.77
		(2.16)	(5.16)	(5.29)	(0.14)	(0.14)	(0.11)
15:1	26	6.76	24.62	55.19	0.17	0.18	−0.34
		(2.09)	(7.46)	(11.89)	(0.12)	(0.11)	(0.05)
3:12:1	46	6.88	25.87	53.26	0.38	0.00	−0.59
		(1.58)	(6.94)	(5.83)	(0.24)	(0.10)	(0.08)
10:6	91	8.69	23.74	54.40	0.00	−0.01	0.49
		(2.51)	(5.06)	(4.05)	(0.06)	(0.08)	(0.08)
6:9:1	90	8.06	24.17	55.28	0.19	0.19	−0.76
		(2.15)	(6.72)	(5.18)	(0.14)	(0.12)	(0.11)

[a] The values in the brackets stand for the *SD*.

could only be detected by ICIM because they had no main effects (Table 6). Obviously, the detection power of ICIM was higher than MIM, and the estimated positions of MIM were more biased than ICIM. And ICIM can detect more kinds of interacting QTL than MIM, especially for QTLs that only have epistatic effects.

Mapping results for the actual populations

Inclusive composite interval mapping detected five pairs of epistatic QTLs affecting grain length for the rice recombinant inbred line (RIL) population (Figure 3). The interacting QTL at 145 cM on chromosome 1 and 20 cM on chromosome 5 were detected in all three environments with LOD scores 6.39, 6.98 and 9.32, respectively. Two other pairs of interacting QTLs were identified only in environment GLHN2003. One was at 125 cM on chromosome 1 and 90 cM on chromosome 11. The other was at 80 cM on chromosome 4 and 80 cM on chromosome 11.

Discussion

More and more epistasis has been recognized in QTL mapping nowadays. Using DH as an example, genotypic means for QTL genotypes AABB, AAbb, aaBB and aabb were denoted as μ_{AABB}, μ_{AAbb}, μ_{aaBB} and μ_{aabb}. Then $\mu_{AABB} = m + a_1 + a_2 + aa$, $\mu_{AAbb} = m + a_1 - a_2 - aa$, $\mu_{aaBB} = m - a_1 + a_2 - aa$, $\mu_{aabb} = m - a_1 - a_2 + aa$, where m was the mean of QTL genotypic values; a1, a2 were additive effects of interacting QTL; aa was epistatic effects. Let m = 0, a1 = 1, a2 = 2, and aa = −1.5, we have $\mu_{AABB} = 1.5$, $\mu_{AAbb} = 0.5$, $\mu_{aaBB} = 2.5$, and $\mu_{aabb} = -1.5$. If episatsis is ignored, we have $\mu_{AABB} = 3$, $\mu_{AAbb} = -1$, $\mu_{aaBB} = 1$, and $\mu_{aabb} = -3$. Assuming additive effect can be unbiased estimated from 1D mapping, and the epistatsis is ignored, then the genotype with highest value will be $\mu_{AABB} = 3$. But actually, the best genotype should be aaBB. So the best genotype cannot be corrected picked up by 1D mapping, if epistasis exists. We also have such examples in the 14 segregation ratios. For example, $\mu_{AABB} = 1$, $\mu_{AAbb} = 0$,

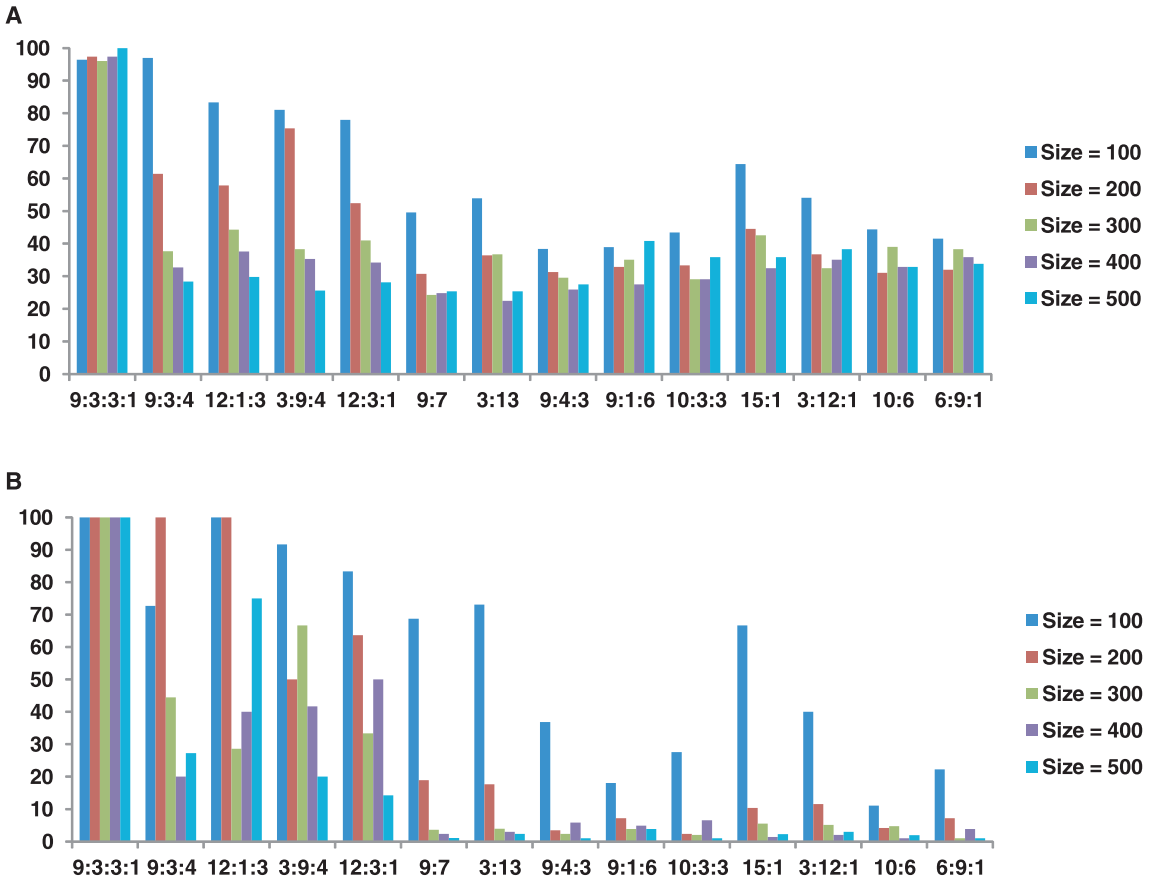

Figure 2. False discovery rates in detecting interacting quantitative trait loci (QTLs) in F_2 population and DH populations.

$\mu_{aaBB} = 0$, and $\mu_{aabb} = 1$ in DH for segregation ratio 10:6; and a1 = 0, a2 = 0, aa = 0.50, and m = 0.50. Then the genotypes with highest values were AABB and aabb. If the epistasis is ignored, μ_{AABB}, μ_{AAbb}, μ_{aaBB} and μ_{aabb} were all equal to 0.50, and the genotypes having highest values cannot be properly determined.

However, the detection of epistasis is more difficult and complex, compared with the detection of additive (and dominance) QTL. When epistatic effects are considered, the genetic model contains much more parameters compared with additive mapping. Therefore, it is more difficult to identify epistatic QTL and estimate the epistatic effects (Li et al. 2008). Simulation

Table 6. Digenic epistatic quantitative trait loci (QTLs) identified by inclusive composite interval mapping (ICIM) and multiple interval mapping (MIM) in the first simulated population of DH

		QTL1		QTL2						
Method	Segregation ratio	Chr.	Pos.	Chr.	Pos.	**LOD**	PVE (%)	a_1	a_2	aa
ICIM	9:7	1	25	2	55	6.19	9.58	0.3	0.05	0.32
	9:4:3	1	25	2	55	6.15	10.16	0.47	−0.11	0.48
	9:1:6	1	25	2	55	9.89	17.68	0.29	0.31	0.79
	10:3:3	1	20	2	60	8.58	11.78	0.17	−0.35	0.64
	15:1	1	25	2	55	7.25	10.68	0.17	0.24	−0.34
	3:12:1	1	25	2	55	8.21	11.28	0.53	0.07	−0.54
	10:6	1	30	2	50	10.03	14.25	−0.01	−0.02	0.42
	6:9:1	1	25	2	60	7.91	11.95	0.19	0.27	−0.62
MIM	3:13	1	30	2	50	2.69	3.7	0.20	−0.25	−0.18
	15:1	1	30	2	50	4.56	6.6	0.21	0.24	−0.25

LOD, likelihood of odd; PVE, phenotypic variance explained.

results in this study indicated that ICIM is an efficient statistical method for epistatic mapping, which has a relatively high detection power and higher precision than MIM. Estimations of QTL locations and effects from ICIM were unbiased, and the detection power was largely affected by population size, heritability of epistasis, and the amount and distribution of genetic effects. Two interacting QTLs had a high detection power if the heritability of epistasis was high and the population size was large enough.

Stepwise regression in ICIM was applied to identify the most significant markers and marker−pair multiplications (Li et al. 2008). Assume there were m markers in the genome, which consisted of c chromosomes. For DH, RIL and so on, which had only two marker types at each locus, there were two steps of regression. One regression was for marker variables, which had m variables, and the other was for marker−pair multiplications, which had $m(m-1)/2$ variables. When the number of markers was large, the difficulty of variable selection was obvious. However, for F_2, F_3 and so on, which had three marker types in each locus, it was proved that the dominance effect of QTL could cause the interaction between markers. So the number of variables for fixing additive and dominance effects of QTL was $4m-2c$ (Zhang et al. 2008). The number of variables for fixing additive, dominance and four kinds of epistatic effects was about $8m^2 - 26m + 25$. For example, when there were 120 markers, the regression model for epistatic mapping needed to consider about 112 105 variables at the same time. This exceeded the ability of most regression methods and might be solved by a method of reduction of dimension.

For simplicity, we used one step of regression, which was for marker variables in F_2, F_3 and so on, which controlled the main effects outside the scanning interval. This might lead to biased estimation of QTL positions and effects. But the two-dimensional scanning in ICIM could modify the bias to a great extent. In the future, we will consider the improvement of epistatic mapping in F_2 population, which can control both main effects and epistatic effects of background QTL at the same time. This will improve the detection power and accuracy of epistatic mapping in F_2.

We simulated populations with marker density 5 cM. From the power analysis, the increase of marker density 10 cM to 5 cM led to similar powers and higher FDR in QTL detection. This is because of the complex genetic model of epistatic mapping. When marker density was 5 cM, there were more variables in the regression model than marker density 10 cM. As we know, the regression model for epistatic mapping is much more complex than additive mapping owing to the marker-pair multiplications. Too many variables in the model will add the difficulty of fixing the model accurately. So for marker density 5 cM, the model may not be accurate, which leads to similar power and higher FDR compared with marker density 10 cM.

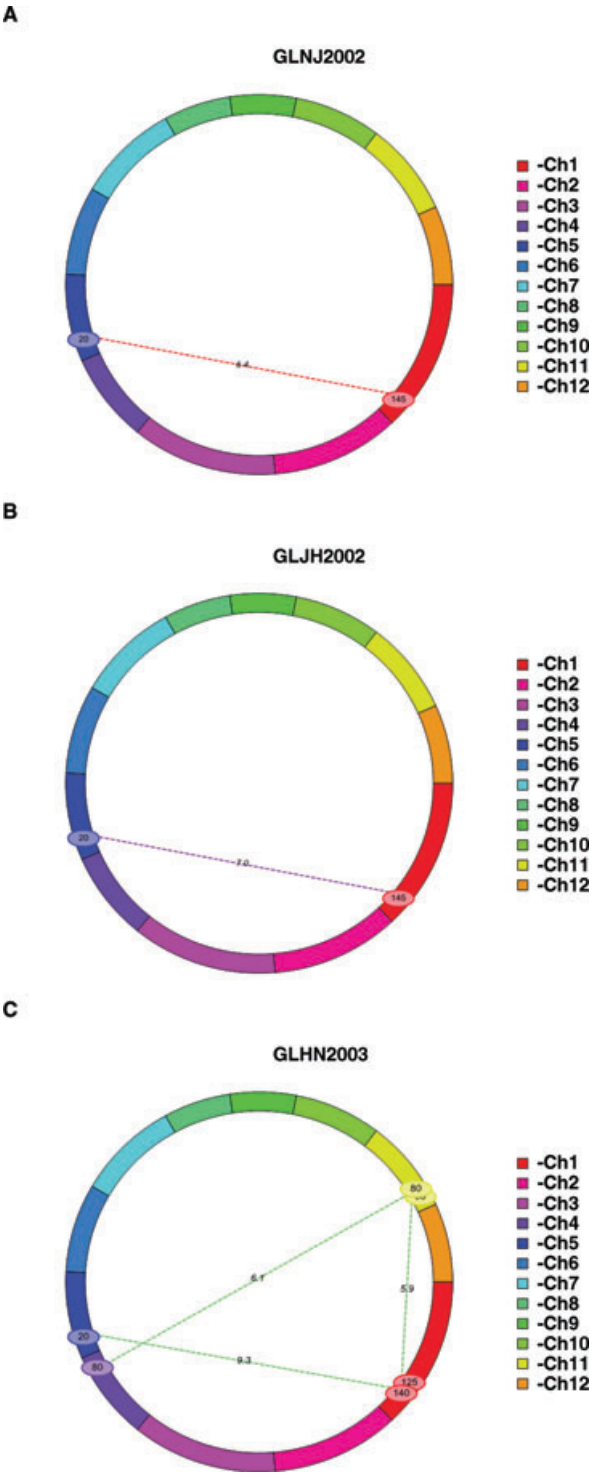

Figure 3. Chromosomal positions of epistatic quantitative trait loci (QTLs) affecting grain length in the rice recombinant inbred line (RIL) population.

One pair of QTLs was detected in all three environments. Its estimated epistatic effects were 0.29, 0.29 and 0.25 in the three environments, respectively.

Material and Methods

Segregation ratios in simulation

Assuming there are two independent loci (i.e. no linkage), represented by A-a and B-b, allele A is dominant to allele a, and allele B is dominant to allele b. When there is no interaction between the two loci, there will be four phenotypic classes having the segregation ratio 9:3:3:1 in an F_2 population derived from two homozygous parents AABB and aabb. When there are some interactions between the two loci, fewer phenotypic classes will be observed. In addition to ratio 9:3:3:1, other 13 commonly observed ratios in F_2 population considered in this study (Bernardo 2002; Janick 2004; Zhai and Wang 2007) were shown in **Table 1**, including the corresponding segregation ratios in DH population.

Quantitative trait loci in F_2 had two kinds of main effects, i.e. additive and dominance effect, and four kinds of epistatic effects, i.e. additive by additive, additive by dominance, dominance by additive and dominance by dominance effects. QTLs in DH had only additive effects and additive by additive effects. In the F_2 population, the relationship between genotypic value and genetic effects of QTL was as follows (Falconer and Mackay 1996),

$$\begin{pmatrix} \mu_{AABB} \\ \mu_{AABb} \\ \mu_{AAbb} \\ \mu_{AaBB} \\ \mu_{AaBb} \\ \mu_{Aabb} \\ \mu_{aaBB} \\ \mu_{aaBb} \\ \mu_{aabb} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & -1 & 0 & 0 & 0 & -1 & 0 \\ 1 & -1 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} m \\ a_1 \\ d_1 \\ a_2 \\ d_2 \\ aa \\ ad \\ da \\ dd \end{pmatrix},$$

where μ_{AABB}, μ_{AABb}, μ_{AAbb}, μ_{AaBB}, μ_{AaBb}, μ_{Aabb}, μ_{aaBB}, μ_{aaBb} and μ_{aabb} were genotypic means for QTL genotypes AABB, AABb, AAbb, AaBB, AaBb, Aabb, aaBB, aaBb, and aabb; m was the mean of QTL genotypic values; a1, a2 were additive effects of interacting QTL; d1, d2 were dominance effects of interacting QTL; aa, ad, da, dd were four epistatic effects, i.e. additive by additive, additive by dominance, dominance by additive and dominance by dominance effects. By resolving the above linear equations, genetic effects of QTL can be calculated as:

$$\begin{pmatrix} m \\ a_1 \\ d_1 \\ a_2 \\ d_2 \\ aa \\ ad \\ da \\ dd \end{pmatrix} = \begin{pmatrix} 0.25 & 0 & 0.25 & 0 & 0 & 0 & 0.25 & 0 & 0.25 \\ 0.25 & 0 & 0.25 & 0 & 0 & 0 & -0.25 & 0 & -0.25 \\ -0.25 & 0 & -0.25 & 0.5 & 0 & 0.5 & -0.25 & 0 & -0.25 \\ 0.25 & 0 & -0.25 & 0 & 0 & 0 & 0.25 & 0 & -0.25 \\ -0.25 & 0.5 & -0.25 & 0 & 0 & 0 & -0.25 & 0.5 & -0.25 \\ 0.25 & 0 & -0.25 & 0 & 0 & 0 & -0.25 & 0 & 0.25 \\ -0.25 & 0.5 & -0.25 & 0 & 0 & 0 & 0.25 & -0.5 & 0.25 \\ -0.25 & 0 & 0.25 & 0.5 & 0 & -0.5 & -0.25 & 0 & 0.25 \\ 0.25 & -0.5 & 0.25 & -0.5 & 1 & -0.5 & 0.25 & -0.5 & 0.25 \end{pmatrix} \begin{pmatrix} \mu_{AABB} \\ \mu_{AABb} \\ \mu_{AAbb} \\ \mu_{AaBB} \\ \mu_{AaBb} \\ \mu_{Aabb} \\ \mu_{aaBB} \\ \mu_{aaBb} \\ \mu_{aabb} \end{pmatrix}.$$

QTL position and marker density in simulation

We assume a genome consisting of four chromosomes; each chromosome is of 140 cM, evenly distributed with 15 codominant markers. In other words, the marker density is 10 cM in this genome. The two interacting QTLs were located at 25 cM on chromosome 1 and 55 cM on chromosome 2. Each trait was assumed to fit one of the 14 F_2 segregation ratios in simulation. The heritability was set at 0.6 for all traits (or equally all segregation ratios).

Size of F_2 and DH populations in simulation

Five levels of population sizes, i.e. 100, 200, 300, 400 and 500 were simulated for each population type and segregation ratios. Using the genetic models above, the populations were generated by QuLine (Wang et al. 2003, 2004). Thus, there were a total of 10 simulation scenarios (i.e. two population types by five population sizes). QTL mapping of epistatic QTLs was conducted by the software QTL IciMapping using ICIM. The threshold LOD score was set at 5.0 for F_2 and DH. The two probabilities for entering and removing variables were set at 0.001 and 0.002. And the length of support interval was 60 cM.

Power calculation and position and effect estimation

One hundred F_2 and DH populations were generated for each simulation scenario. In each simulated population, QTL epistatic mapping was conducted for the 14 segregation ratios. Epistatic mapping is not a point estimation, including estimations for positions of two interacting QTLs. Each predefined QTL was assigned to a support interval of given length centered at the true QTL location. The power for the two support intervals was estimated at the same time. For one simulation, if each of the two interacting QTLs was detected in the corresponding support interval, the power of QTL detection was added by one. QTLs identified in other intervals were viewed as false positives. If pairs of multiple peaks occurred in the two support intervals, only the highest one was counted. Mean QTL positions and effects were estimated according to the peaks in the support intervals whose LOD scores were higher than the given LOD threshold.

Estimation of FDR

In simulation studies, the FDR was defined as the proportion of false positives to the total number of significant discoveries (i.e. true positives plus false positives, Benjamini and Hochberg 1995). An efficient mapping method should not only have high detection power, but also have low FDR. So both powers and FDR have to be considered for evaluating the efficiency of QTL mapping methods. Owing to the complex genetic model in epistatic mapping, FDR might be higher than additive mapping.

Comparison of ICIM with MIM using simulated populations

To demonstrate the efficiency of ICIM in epistatic mapping compared with other methods, we applied both ICIM and MIM (Kao et al. 1999) on the first simulated DH population for each ratio, so as to compare the mapping results of both mapping methods. The parameters in ICIM were the same as used in simulations, and the parameters in MIM were used as default.

One actual RIL population in rice

We used one rice (*Oryza sativa* L.) RIL population for ICIM epistatic mapping. The population consisted of 71 individuals and grain length was investigated in three different environments (denoted as GLNJ2002, GLJH2002 and GLHN2003). A total of 250 relatively evenly distributed markers covered a whole genome of 1301.8 cM on the 12 rice chromosomes, and the average marker distance was 5.2 cM. The same mapping parameters were used as for the simulated populations.

Acknowledgements

This work was supported by the HarvestPlus Challenge Program of CGIAR, the Special Funds for EU Collaboration from the Ministry of Science and Technology of China (Project no. 1113), and the Seventh Framework Programme of European Commission (Project no. 266045).

Received 14 Dec. 2011 Accepted 13 Feb. 2012

References

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Statist. Soc. B* **57,** 289–300.

Bernardo R (2002) *Breeding for Quantitative Traits in Plants*. Stemma Press, Woodbury, Minnesota.

Boer M, Braak C, Jasen R (2002) A penalized likelihood method for mapping epistatic quantitative trait loci with one-dimensional genome searches. *Genetics* **162,** 951–960.

Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics*. 4 edn. Longman, Essenx.

Holland JB, Moser HS, O'Donoughue LS, Lee M (1997) QTLs and epistasis associated with vernalization responses in oat. *Crop Sci.* **37,** 1306–1316.

Janick J (2004) *Plant Breeding Reviews, Volume 24, Part 1: Long-term Selection: Maize*. John Wiley & Sons, Inc., Hoboken, New Jersey.

Kao CH, Zeng ZB, Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. *Genetics* **152,** 1203–1206.

Lark KG, Chase K, Adler F, Mansur LM, Orf JH (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 H, Ye G, Wang J (2007) A modified algorithm for the improvement of composite interval mapping. *Genetics* **175,** 361–374.

Li H, Ribaut JM, Li Z, Wang J (2008) Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations. *Theor. Appl. Genet.* **116,** 243–260.

Li H, Hearne S, Bänziger M, Li Z, Wang J (2010) Statistical properties of QTL linkage mapping in biparental genetic populations. *Heredity* **105,** 257–267.

Luo X, Fu Y, Zhang P, Wu S, Tian F, Liu J, Zhou Z, Yang J, Sun C (2009) Additive and over-dominant effects resulting from epistatic loci are the primary genetic basis of heterosis in rice. *J. Integr. Plant Biol.* **51,** 393–408.

Nagase H, Mao JH, de Koning JP, Minami T, Balmain A (2001) Epistatic interactions between skin tumor modifier loci in interspecific (spretus/musculus) backcross mice. *Cancer Res.* **61,** 1305–1308.

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.

Wang J, van Ginkel M, Podlich D, Ye G, Trethowan R, Pfeiffer W, DeLacy IH, Cooper M, Rajaram S (2003) Comparison of two breeding strategies by computer simulation. *Crop Sci.* **43,** 1764–1773.

Wang J, Van Ginkel M, Trethowan R, Ye G, Delacy I, Podlich D, Cooper M (2004) Simulating the effects of dominance and epistasis on selection response in the CIMMYT wheat breeding program using QuCim. *Crop Sci.* **44,** 2006–2018.

Wang J (2009) Inclusive composite interval mapping of quantitative trait genes. *Acta Agron. Sin.* **35,** 239–245.

Yu SB, Li JX, Xu CG, Tan YF, Gao YJ, Li XH, 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.

Zhang L, Li H, Li Z, Wang J (2008) Interactions between markers can be caused by the dominance effect of quantitative trait loci. *Genetics* **180,** 1177–1190.

Zhai H, Wang J (2007) *Applied Quantitative Genetics*. China Agricultural Science and Technology Press (in Chinese), Beijing.

(Co-Editor: Martin A. J. Parry)