Factors influencing QTL mapping accuracy under complicated genetic models by computer simulation

C.F. Su[1]*, W. Wang[2]*, S.L. Gong[3], J.H. Zuo[1] and S.J. Li[1]

[1]Department of Life Sciences, Liupanshui Normal University,
Liupanshui, Guizhou, China
[2]Agricultural Science Institute of Coastal Region of Jiangsu,
Yancheng, Jiangsu, China
[3]Liupanshui Academy of Agricultural Sciences, Liupanshui, Guizhou, China

*These authors contributed equally to this study.
Corresponding author: C.F. Su
E-mail: chfsu2008@163.com

ABSTRACT. The accuracy of quantitative trait loci (QTLs) identified using different sample sizes and marker densities was evaluated in different genetic models. Model I assumed one additive QTL; Model II assumed three additive QTLs plus one pair of epistatic QTLs; and Model III assumed two additive QTLs with opposite genetic effects plus two pairs of epistatic QTLs. Recombinant inbred lines (RILs) (50-1500 samples) were simulated according to the Models to study the influence of different sample sizes under different genetic models on QTL mapping accuracy. RILs with 10-100 target chromosome markers were simulated according to Models I and II to evaluate the influence of marker density on QTL mapping accuracy. Different marker densities

did not significantly influence accurate estimation of genetic effects with simple additive models, but influenced QTL mapping accuracy in the additive and epistatic models. The optimum marker density was approximately 20 markers when the recombination fraction between two adjacent markers was 0.056 in the additive and epistatic models. A sample size of 150 was sufficient for detecting simple additive QTLs. Thus, a sample size of approximately 450 is needed to detect QTLs with additive and epistatic models. Sample size must be approximately 750 to detect QTLs with additive, epistatic, and combined effects between QTLs. The sample size should be increased to >750 if the genetic models of the data set become more complicated than Model III. Our results provide a theoretical basis for marker-assisted selection breeding and molecular design breeding.

INTRODUCTION

An increasing number of quantitative trait loci (QTL) have been identified in various species because of the development of different QTL mapping methods (Edwards et al., 1987; Lander and Botstein, 1989; Zeng, 1994; Kao et al., 1999) and quantitative genetics. QTL identification methods are based on the construction of genetic linkage maps. The first step of the QTL mapping process is the development of a genetically segregated population and to determine the genetic value data and phenotypic value data of quantitative traits. The second step is determination of the relationship between the genetic value and phenotypic value using specific statistical methods. Finally, the third step is to identify putative QTLs on different groups of linkage maps and evaluate the genetic effects of the putative QTLs. Since Lander and Botstein (1989) reported the first QTL analysis software (Mapmaker/QTL), many other QTL mapping software programs have been developed and reported, including QTL Cartographer (Basten et al., 1994), PLABQTL (Utz and Melchinger, 1996), Map Manager (Manly et al., 2001), QGene (Nelson, 1997), MapQTL (Van Ooijen and Maliepaard, 1996), PGRI (Lu and Liu, 1995), QTLMAPPER (Wang et al., 2003), IciMapping (Li et al., 2007), and QTL network (Yang et al., 2007, 2008).

QTL mapping is a useful method for identifying genetic regions responsible for important phenotype variations (Takuno et al., 2012). QTL mapping has been widely used in plants and animals for several years and large numbers of QTL have been detected in various species. For example, a total of 8646 and 1747 QTLs have been identified in rice (*Oryza sativa*) and maize (*Zea mays* subsp *mays*), respectively (Gramene QTL database). A key point in QTL mapping is the accuracy of the QTL. Real QTLs can be used for marker-assisted breeding or map-based cloning, whereas false-positive QTLs can mislead mapping application results. Generally, it is easier to detect the existence of a QTL than to confirm the accurate position of the QTL and accurately estimate all of its genetic effects. Therefore, it is critical to accurately map target QTLs for marker-assisted breeding and map-based cloning.

The primary factors influencing QTL mapping accuracy are the sample size, marker density, QTL position on the chromosome, heritability of the objective traits, and QTL mapping

procedures. A previous study evaluated the influence of QTL mapping procedures on accuracy (Su et al., 2013). Among the factors influencing accuracy, QTL position and trait heritability were relatively fixed, whereas sample size and marker density could be controlled to some extent. Theoretically, QTL mapping accuracy increases concomitantly with increasing marker density; however, QTL mapping accuracy was shown to decrease with increasing marker density in a previous study (He et al., 2000). In theory, recombination events increase with increasing sample size; therefore, QTL mapping accuracy may increase if a greater amount of recombination information is supplied for the QTL analysis.

In this study, recombinant inbred line (RIL) populations with different types of genetic models were produced by computer simulation. The inclusive composite interval mapping (ICIM) mapping procedure in IciMapping Version 4.1 (released by a quantitative genetic research group from the Chinese Academy of Agricultural Sciences on January 2016, China) was used for QTL analysis of the simulated data. Our goal was to determine sample sizes and marker densities suitable for RILs with different genetic models and to determine QTL mapping strategies for unknown data from different genetic models by comparing the degree of coincidence of the QTL mapping results obtained from the RILs with different sample sizes and marker densities.

MATERIAL AND METHODS

RIL population simulation

Simulation of genotype and QTLs

The computer simulation method proposed by He et al. (2001) was used as a reference in this study. The method assumed that there were two adjacent markers named M_k and M_{k+1} on a chromosome with a recombination frequency r. The M_k and M_{k+1} genotypes were recorded as x_k and x_{k+1}. Values of '0' and '1' assigned to x_k and x_{k+1} indicated that the DNA of the two loci were from two homozygous parents P1 and P2, respectively. When the value of x_k is known and confirmed as '1', the genotype of x_{k+1} can be inferred and confirmed from the known x_k genotype. It is assumed that there are n_1 individuals whose genotypes are all x_k with values of '1' in n individuals. A total of n_1 uniformly distributed random numbers between 0 and 1 were generated continuously and recorded as 'rnd'. The value of x_{k+1} was set to '1' when $rnd \leq (1-r)^2$. The value of x_{k+1} was set to '0' when $rnd > (1-r)^2$. The values of x_{k+2}, x_{k+3}, etc. can be inferred and confirmed according to the known x_{k+1} value. When the value of x_k is known and confirmed as '0', determination of the values of x_{k+1}, x_{k+2}, x_{k+3}, etc. is the same as when the value of x_k is '1'. For confirmation of the marker genotypes on the starting position of the simulated linkage group, the probability of any x_k value being determined as 0 or 1 is 0.50. A total of n random numbers between 0 and 1 were generated continuously such that $x_k = 0$ when $rnd \leq 0.5$ and $x_k = 1$ when $rnd > 0.50$. The same method with the confirmation of marker genotypes was applied to confirm the QTL genotype and was recorded as x_Q.

A total of 38 groups of RIL data were produced by computer simulation and used for QTL detection. Based on Model I, seven groups of RIL data named as M-ss1-M-ss7 with different sample sizes (50, 150, 250, 350, 450, 550, and 1000 lines) were constructed. In each simulated RIL, one linkage group named as chr1 and containing 20 markers was generated and used for QTL analysis. The 20 markers were named as *M1-M20*. The related QTLs Q_{set1}

and *M10* were considered to be in complete linkage. Recombination fractions between two adjacent markers were randomly generated to be between 0 and 0.1 (**Table S1**). Based on Model I, six groups of RIL data named as M-sd1-M-sd6 with different marker densities [10 (*M1-M10*), 20 (*M1-M20*), 30 (*M1-M30*), 40 (*M1-M40*), 50 (*M1-M50*), and 100 (*M1-M100*) markers] on the linkage group were constructed by computer simulation. The sample sizes of M-sd1-M-sd7 were all 350. In each simulated RIL, one linkage group named as chr1 was generated and used for QTL analysis. The related QTLs Q_{set1} and *M4* in M-sd1, Q_{set1} and *M8* in M-sd2, Q_{set1} and *M11* in M-sd3, Q_{set1} and *M14* in M-sd4, Q_{set1} and *M16* in M-sd5, Q_{set1} and *M32* in M-sd6 and *M10* were considered to be in complete linkage. The recombination fractions between two adjacent markers of M-sd1-M-sd6 were 0.14, 0.078, 0.017, 0.056, 0.041, and 0.035, respectively.

Based on Model II, eight groups of RIL data named as M-cs1-M-cs8 were constructed with different sample sizes (50, 150, 250, 350, 450, 550, 750, and 1000 lines). In each simulated RIL of M-cs1-M-cs8, a total three linkage groups named as chr1, chr2, and chr3 containing 20 markers each (*M1-M20*, *M21-M40*, and *M41-M60*) were generated and used for QTL analysis. The related QTLs Q_{set1}, Q_{set2}, and Q_{set3} and *M10*, *M32*, and *M54* were considered to be in complete linkage, respectively. The recombination fractions between two adjacent markers were randomly generated to be between 0 and 0.1 (**Table S2**). Based on Model II, six groups of RIL data named as M-cd1-M-cd6 with different marker densities (10, 20, 30, 40, 50, and 100 markers) on each of three linkage groups were constructed by computer simulation. The sample sizes of M-cd1-M-cd6 were all 350. The detailed parameters applied for QTL mapping analysis of M-cd1-M-cd6 are listed in **Table S3**.

Based on Model III, eleven groups of RIL data named as M-tc1-M-tc11 were constructed with different sample sizes (50, 150, 250, 350, 450, 550, 650, 750, 850, 1000, and 1500 lines). In each simulated RIL of M-tc1-M-tc11, a total three linkage groups named as chr1, chr2, and chr3 containing 30 markers each (*M1-M30*, *M31-M60*, and *M61-M90*) were generated and used for QTL analysis. The related QTLs Q_{set1}, Q_{set2}, Q_{set4}, Q_{set5}, and *M8*, *M10*, M41, and M76 were considered to be in complete linkage, respectively. Q_{set1} and Q_{set2} were two adjacent QTLs located on chr1. Q_{set4} and Q_{set5} had no additive effects, but there are interaction effects for $Q_{set1} \times Q_{set4}$ and $Q_{set2} \times Q_{set5}$. The recombination fractions between two adjacent markers were set to 0.056 in M-tc1-M-tc11.

Confirmation of phenotypic value of quantity trait

The phenotypic values were set up according to the 44 Models (M-ss1-M-ss7, M-sd1-M-sd6, M-cs1-M-cs8, M-cd1-M-cd6, and M-tc1-M-tc11) as described above. The genetic parameters in each model were set up such that the genetic effect of a_1 was 2, a_2 was -2, a_3 was 3, a_4 was 0, a_5 was 0, aa_{12} was 2.5, aa_{14} was 2.5, and aa_{25} was 2. The effects of a_1, a_2, a_3, a_4, and a_5 were the additive effects of Q_{set1}, Q_{set2}, Q_{set3}, Q_{set4}, and Q_{set5}, respectively, whereas aa_{12}, aa_{14}, and aa_{25} were the epistatic effect of $Q_{set1 \times set2}$, $Q_{set1 \times set4}$, and $Q_{set2 \times set5}$ interactions. The same heritability of '30%' was set up in all 44 models.

Mapping procedures

ICIM in IciMapping Version 4.1 was used for QTL analysis of the simulated data. Analysis of 1000 permutations was performed to determine the logarithm of odds ratio threshold.

RESULTS

The ICIM mapping procedure was applied to analyze the QTL effects in each genetic model (M-ss1-M-ss7, M-sd1-M-sd6, M-cs1-M-cs8, M-cd1-M-cd6, M-ts1-M-ts6, and M-tc1-M-tc11) with 200 repetitions. Several parameters [i.e., average, accuracy, minimum, maximum, and coefficient of variation (CV)] were obtained by calculating and analyzing 200 values from 200 rounds of simulation. Related statistical analysis results of the QTL mapping power, variation range of the QTL position, and numbers of false-positive QTLs are shown in Tables 1-5.

Table 1. Estimates of genetic parameters of models M-ss1 to M-ss7 under different sample size conditions (50, 150, 250, 350, 450, 550, and 1000) based on 200 repetitions.

Sample size	Effect	TRUE value	Average	Accuracy	Minimum	Maximum	CV	RanQTL (cM)	Power (%)	False QTL
50 individuals	a1	2	2.07	0.967	1.26	3.27	0.186	83	94.5	2
	PVE (%)	30	31.67	0.944	15.42	56.84	0.012			
150 individuals	a1	2	2.03	0.987	1.31	2.64	0.123	7	100	3
	PVE (%)	30	31.43	0.952	12.41	47.58	0.008			
250 individuals	a1	2	2.01	0.994	1.48	2.62	0.094	4	100	3
	PVE (%)	30	30.88	0.971	17.13	43.9	0.006			
350 individuals	a1	2	2.01	0.996	1.56	2.36	0.076	5	100	8
	PVE (%)	30	30.54	0.982	20.77	43.05	0.050			
450 individuals	a1	2	2.06	0.971	1.67	5.81	0.150	4	100	17
	PVE (%)	30	32.31	0.923	20.28	232.28	0.010			
550 individuals	a1	2	2.06	0.969	1.65	4.32	0.105	3	100	6
	PVE (%)	30	32.13	0.929	22.13	141	0.007			
1000 individuals	a1	2	2.03	0.985	1.79	3.97	0.083	1	100	7
	PVE (%)	30	31.17	0.960	25.23	112.81	0.005			

Accuracy is determined as (1 - |average value - true value| / |true value|), CV = coefficient of variation, RanQTL = range of QTL location; the same abbreviations are used in the remaining tables.

Table 2. Estimates of genetic parameters of models M-sd1 to M-sd6 under different marker density conditions based on 200 repetitions.

No. of markers	r	Effect	TRUE value	Average	Accuracy	Minimum	Maximum	CV	RanQTL (cM)	Power (%)	False QTL
10 markers	0.017	a1	2	2.07	0.967	1.47	2.59	0.09	7	100	9
		PVE (%)	30	32.16	0.928	17.63	47.1	0.006			
20 markers	0.035	a1	2	2.05	0.973	1.66	2.51	0.077	7	100	6
		PVE (%)	30	31.68	0.944	21.41	43.56	0.005			
30 markers	0.041	a1	2	2.03	0.985	1.49	2.47	0.088	5	100	6
		PVE (%)	30	30.94	0.969	19.37	41.81	0.006			
40 markers	0.056	a1	2	2.03	0.987	1.63	2.63	0.085	4	100	13
		PVE (%)	30	30.59	0.98	23.17	57.77	0.006			
50 markers	0.078	a1	2	2.03	0.984	1.31	2.51	0.088	7	100	6
		PVE (%)	30	31.27	0.958	13.39	46.56	0.006			
100 markers	0.14	a1	2	2.02	0.991	1.53	2.52	0.082	5	100	8
		PVE (%)	30	30.71	0.976	16.61	42.36	0.005			

QTL mapping power

Variation in the detection power in models M-ss1-M-ss7 is shown in Figure 1. The detection power of a_1 in M-ss1 with a 50 line sample size was 94.5%. The detection power of a_1 increased to 100% when the RIL sample size was increased from 150 to 1000 in the M-ss1-M-ss7 models (Table 1). A sample size of 150 RILs was sufficient for the single additive QTL genetic model. In models M-sd1-M-sd6, the detection power was 100% when the marker

Table 3. Estimates of genetic parameters of M-cs1 to M-cs8 under different sample size conditions (50, 150, 250, 350, 450, 550, and 1000) based on 200 repetitions.

Sample size	Effect	TRUE value	Average	Accuracy	Minimum	Maximum	CV	RanQTL (cM)	Power (%)	False QTL
50 individuals	a1	2	4.52	0.205	3.80	5.42	0.105	42	7.0	0
	a2	-2	-4.54	0.213	-5.95	-3.61	0.129	55	13.0	0
	a3	3	4.44	0.521	3.22	5.75	0.129	63	12.0	0
	aa12	2.5	6.48	0.373	4.98	7.51	0.206		1.5	2
150 individuals	a1	2	2.83	0.586	1.96	3.85	0.145	60	65.0	0
	a2	-2	-2.76	0.618	-4.30	-1.81	0.142	70	74.5	0
	a3	3	3.04	0.988	1.82	4.92	0.167	40	93.5	2
	aa12	2.5	3.62	0.552	2.88	4.62	0.109		17.0	0
250 individuals	a1	2	2.35	0.813	1.66	3.47	0.179	40	85.5	2
	a2	-2	-2.23	0.884	-9.99	-1.57	0.348	27	71.0	8
	a3	3	2.88	0.960	1.76	4.22	0.160	16	99.5	2
	aa12	2.5	2.89	0.846	2.19	4.41	0.133		65.5	0
350 individuals	a1	2	2.14	0.931	1.43	3.33	0.166	12	93.5	1
	a2	-2	-2.16	0.921	-4.78	-1.36	0.195	26	97.5	2
	a3	3	3.17	0.942	1.75	4.05	0.128	9	100.0	2
	aa12	2.5	2.69	0.926	1.86	3.84	0.140		85.0	7
450 individuals	a1	2	2.12	0.940	1.18	3.13	0.163	19	98.0	2
	a2	-2	-2.00	1.000	-3.45	-1.19	0.173	13	99.5	2
	a3	3	3.11	0.964	2.33	4.87	0.121	8	100.0	5
	aa12	2.5	2.58	0.967	1.73	3.45	0.143		95.0	1
550 individuals	a1	2	2.01	0.997	1.24	2.86	0.166	16	100.0	3
	a2	-2	-2.04	0.979	-2.90	-1.24	0.163	14	99.5	4
	a3	3	3.15	0.948	2.29	4.22	0.106	2	100.0	6
	aa12	2.5	2.62	0.953	1.76	3.65	0.127		99.0	0
750 individuals	a1	2	2.09	0.957	1.18	2.94	0.145	16	100.0	1
	a2	-2	-2.05	0.973	-2.77	-1.46	0.130	23	100.0	0
	a3	3	3.03	0.990	2.35	3.69	0.090	16	100.0	0
	aa12	2.5	2.51	0.997	1.73	3.22	0.113		100.0	1
1000 individuals	a1	2	2.09	0.954	1.23	2.85	0.129	8	100.0	5
	a2	-2	-2.05	0.977	-5.10	-1.51	0.158	20	100.0	1
	a3	3	3.08	0.975	2.24	3.79	0.089	12	98.5	9
	aa12	2.5	2.58	0.970	1.79	3.37	0.094		100.0	1

Table 4. Estimates of genetic parameters of M-cd1 to M-cd6 under different marker density conditions based on 200 repetitions.

| Marker density	r	Effect	TRUE value	Average	Accuracy	Minimum	Maximum	CV	RanQTL (cM)	Power (%)	False QTL
10 markers	0.14	a1	2	2.29	0.855	1.39	3.51	0.199	51	96.5	2
		a2	-2	-2.35	0.826	-3.71	-1.34	0.199	23	98.0	2
		a3	3	3.09	0.970	1.50	4.14	0.142	13	100.0	5
		aa12	2.5	2.81	0.878	2.09	3.99	0.142		95.5	0
20 markers	0.078	a1	2	2.26	0.870	1.39	3.13	0.174	18	97.5	3
		a2	-2	-2.36	0.818	-3.54	-1.51	0.180	86	98.0	1
		a3	3	3.04	0.988	2.09	4.23	0.143	8	100.0	2
		aa12	2.5	2.75	0.898	1.95	3.85	0.137		97.0	0
30 markers	0.056	a1	2	2.31	0.843	1.42	3.47	0.169	15	99.0	3
		a2	-2	-2.20	0.900	-3.87	-1.52	0.176	55	97.0	4
		a3	3	3.06	0.978	1.73	4.16	0.143	6	100.0	4
		aa12	2.5	2.72	0.912	1.95	3.99	0.133		94.0	2
40 markers	0.041	a1	2	2.07	0.964	1.41	3.25	0.166	37	87.0	2
		a2	-2	-2.37	0.817	-4.52	-1.43	0.184	17	98.5	4
		a3	3	3.17	0.944	2.08	4.79	0.133	10	100.0	2
		aa12	2.5	2.71	0.915	1.93	3.85	0.139		92.0	0
50 markers	0.035	a1	2	1.98	0.992	1.44	3.90	0.177	31	85.5	9
		a2	-2	-2.16	0.921	-5.07	-1.40	0.199	53	92.0	3
		a3	3	3.15	0.948	2.15	4.73	0.119	17	100.0	6
		aa12	2.5	2.66	0.937	1.99	3.52	0.122		89.0	1
100 markers	0.017	a1	2	2.08	0.960	1.43	5.12	0.212	35	92.5	6
		a2	-2	-2.17	0.913	-5.89	-1.47	0.201	40	93.5	5
		a3	3	2.96	0.987	2.07	5.95	0.158	14	100.0	9
		aa12	2.5	2.46	0.983	1.86	3.34	0.119		87.0	1

Table 5. Estimates of genetic parameters of M-tc1 to M-tc11 under different sample size conditions (50, 150, 250, 350, 450, 550, 650, 750, 850, 1000, and 1500) based on 200 repetitions.

Sample size	Effect	TRUE value	Average	Accuracy	Minimum	Maximum	CV	RanQTL (cM)	Power (%)	False QTL
50 individuals	a1	2	4.16	0.073	4.16	4.16	-	0	0.5	15A+2AA
	a2	-2	-3.54	0.231	-3.54	-3.54	-	0	0.5	
	aa13	2.5	5.28	0.101	4.23	6.99	0.156		4.0	
	aa24	2	4.84	0.297	4.84	4.84	-		0.5	
150 individuals	a1	2	2.11	0.948	2.11	2.11	-	0	0.5	3A
	a2	-2	-2.26	0.871	-2.60	-2.00	0.084	17	5.5	
	aa13	2.5	3.45	0.618	2.69	4.96	0.117		43.5	
	aa24	2	3.30	0.350	2.88	3.62	0.080		3.5	
250 individuals	a1	2	2.70	0.651	1.75	3.83	0.263	2	6.5	6A +1AA
	a2	-2	-2.32	0.839	-3.28	-1.62	0.250	36	5.0	
	aa13	2.5	2.88	0.850	2.28	3.86	0.118		84.0	
	aa24	2	2.71	0.643	2.26	3.15	0.083		35.5	
350 individuals	a1	2	2.39	0.805	1.37	3.57	0.269	27	34.0	14A+2AA
	a2	-2	-2.38	0.808	-3.55	-1.48	0.194	9	20.5	
	aa13	2.5	2.62	0.951	1.88	3.38	0.134		96.5	
	aa24	2	2.44	0.782	1.86	3.33	0.116		66.5	
450 individuals	a1	2	2.23	0.886	1.19	3.36	0.235	51	44.5	14A+1AA
	a2	-2	-2.34	0.830	-4.20	-1.30	0.261	29	46.5	
	aa13	2.5	2.87	0.853	2.00	3.82	0.107		100.0	
	aa24	2	2.43	0.784	1.79	3.56	0.144		95.0	
550 individuals	a1	2	2.34	0.829	1.13	3.23	0.216	30	34.0	11A+2AA
	a2	-2	-2.15	0.925	-2.88	-1.16	0.220	11	30.0	
	aa13	2.5	2.86	0.855	2.12	3.57	0.103		100.0	
	aa24	2	2.24	0.881	1.61	2.96	0.128		98.0	
650 individuals	a1	2	2.08	0.959	1.09	3.19	0.192	6	48.5	11A
	a2	-2	-2.25	0.873	-3.02	-1.13	0.157	37	48.5	
	aa13	2.5	2.65	0.938	2.00	3.35	0.098		100.0	
	aa24	2	2.23	0.886	1.61	3.02	0.127		99.5	
750 individuals	a1	2	2.41	0.797	1.47	3.27	0.132	7	98.0	19A
	a2	-2	-2.39	0.805	-3.07	-1.32	0.121	4	98.0	
	aa13	2.5	2.51	0.994	1.79	3.16	0.098		100.0	
	aa24	2	1.97	0.984	1.36	2.65	0.127		100.0	
850 individuals	a1	2	2.12	0.942	1.28	2.90	0.146	14	93.0	9A+1AA
	a2	-2	-2.32	0.841	-3.17	-1.12	0.122	4	92.5	
	aa13	2.5	2.76	0.895	2.21	3.51	0.092		100.0	
	aa24	2	2.21	0.897	1.43	3.01	0.112		100.0	
1000 individuals	a1	2	2.19	0.907	1.19	2.84	0.116	5	96.5	9A
	a2	-2	-2.17	0.913	-2.87	-1.45	0.115	6	96.5	
	aa13	2.5	2.67	0.930	2.08	3.42	0.085		100.0	
	aa24	2	2.16	0.918	1.53	2.83	0.103		100.0	
1500 individuals	a1	2	1.90	0.951	1.24	2.36	0.105	4	99.5	8A
	a2	-2	-2.02	0.988	-2.75	-1.55	0.105	4	99.5	
	aa13	2.5	2.59	0.963	2.15	3.10	0.070		100.0	
	aa24	2	2.09	0.955	1.62	2.77	0.087		100.0	

The recombination fraction between two adjacent markers is 0.056. Accuracy is determined as (1 - |average value - true value| / |true value|), CV = coefficient of variation, A = additive QTL, AA = epistatic. QTL pair, RanQTL = range of QTL location.

densities were increased from 10 to 100 markers on chr1 (Table 2). Therefore, marker density appears to have no influence on single additive QTL genetic effect detection. In models M-cs1-M-cs8, the power to detect all genetic effects of a_1, a_2, a_3, and aa_{12} exhibited a continuously increasing trend when the RIL sample size was increased from 50 to 1000 (Figure 2). When the RIL sample size was 50 in M-cs1, the detection powers of a_1, a_2, a_3, an aa_{12} were 7, 13, 12, and 1.5%, respectively. The detection powers of a_1, a_2, a_3, and aa_{12} increased to more than 95% when the sample size was increased to more than 450 (Table 3). Therefore, the RIL sample size should be at least 450 to accurately estimate all genetic effects in Model II. In models M-cd1-M-cd6, all detection powers of a_1, a_2, a_3, and aa_{12} were greater than 85% as marker densities increased from 10 to 100 markers on each of the three chromosomes (Table 4). The highest detection power of a_1 was 99% when the marker density was 30 markers, the

highest detection power of a_2 was 98.5% when the marker density was 40 markers, and the highest detection power of aa_{12} was 97% when the marker density was 20 markers on the group. The detection power of a_3 was 100% in models M-cd1-M-cd6 with different marker densities (Table 4 and Figure 3). Overall, to accurately estimate all genetic effects in Model II, the optimum marker density is approximately 20 markers when the recombination fraction between two adjacent markers is 0.056. In models M-tc1-M-tc11, Q_{set1} and Q_{set2} represented two adjacent QTLs located on the same linkage group with the same additive effects; however, one effect was positive while the other was negative. When the RIL sample size was 50 in M-tc1, the detection powers of a_1, a_2, aa_{14}, and aa_{25} were 0.5, 0.5, 4, and 0.5%, respectively. When the RIL sample size was more than 450, the detection powers of aa_{14} and aa_{25} both increased to more than 95%; however, the detection powers of a_1 and a_2 were only 44.5 and 46.5%, respectively. When the sample size was increased to more than 750, the detection power of all genetic effects of a_1, a_2, aa_{14}, and aa_{25} increased to more than 90% (Table 5 and Figure 4). We concluded that the RIL sample size should be at least 750 to accurately estimate all genetic effects in Model III. If only the aa_{14} and aa_{25} effects must be estimated, a sample size of 450 is sufficient.
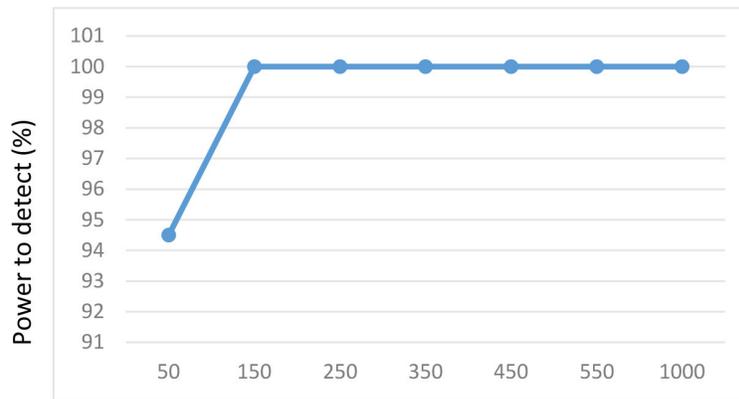
Figure 1. Variation tendency of detecting power of a1 in M-ss1-M-ss7.
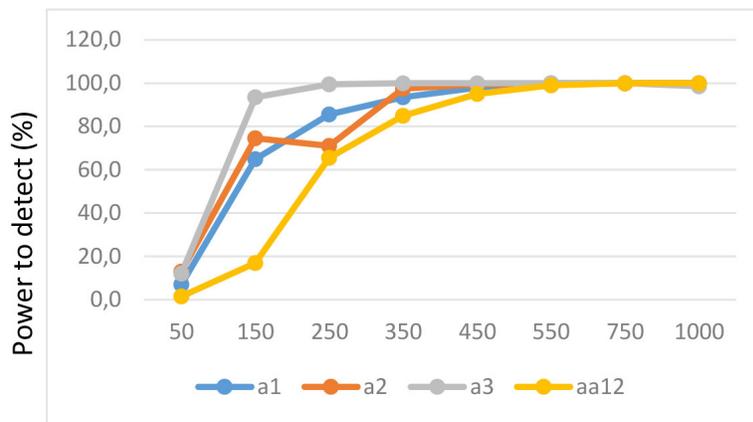
Figure 2. Variation tendency of detecting power of genetic effects in M-cs1-M-cs8.
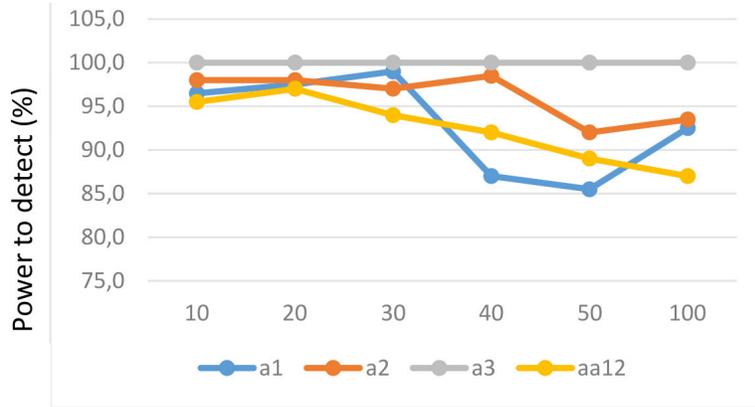
Figure 3. Variation tendency of detecting power of genetic effects in M-cd1-M-cd6.
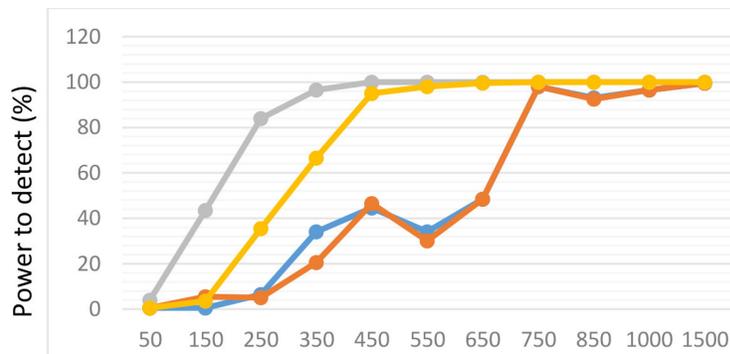
Figure 4. Variation tendency of detecting power of genetic effects in M-tc1-M-tc11.

Coincidence analysis between the actual value and preset value

The genetic effects, accuracy, and CV parameters of the same QTL in 200 rounds of simulation fluctuated to some degree. In models M-ss1-M-ss7, there was no significant difference in the accuracy of a_1 among the seven models. In models M-cs1-M-cs8, the accuracy of all genetic effects of a_1, a_2, a_3, and aa_{12} in M-cs1 to M-cs8 exhibited a continuously increasing trend when the RIL sample sizes were increased from 50 to 1000 (**Figure S1**). When the RIL sample size in M-cs1 was 50, the accuracies of a_1, a_2, a_3, and aa_{12} were only 0.205, 0.213, 0.521, and 0.373, respectively. The accuracies of a_1, a_2, a_3, and aa_{12} increased to more than 90% when the sample size increased to more than 350. The CVs of a_1, a_2, a_3, and aa_{12} became stable and decreased slowly when the sample size was increased from 350 to 1000 (**Figure S2**). There was no significant difference in the accuracy of a_1 among models M-sd1-M-sd6 (Table 2). The accuracy of all genetic effects of a_1, a_2, a_3, and aa_{12} were over 80% in models M-cd1 to M-cd6. It can be infer there was no significant difference in the accuracy of all genetic effects of a_1, a_2, a_3, and aa_{12} among models M-cd1-M-cd6 (**Figure S3**). The lowest variation ranges of the a_1, a_2, and a_3 actual values were 1.74, 2.03, and 2.14, respectively, in

M-cd2 when the marker density was 20 markers on the linkage group. The lowest variation range in the aa_{12} actual value was 1.47 in M-cd6. The variation range fluctuations of the a_1, a_2, and a_3 values gradually increased after increasing the markers to more than 20 on the simulated linkage group (**Figure S4**). In models M-tc1-M-tc11, the accuracy of all genetic effects of a_1, a_2, a_3, and aa_{12} exhibited a continuously increasing trend when the RIL sample sizes were increased from 50 to 1500 (**Figure S5**). When the RIL sample size was 50 in M-tc1, the accuracies of a_1, a_2, aa_{13}, and aa_{24} were 0.073, 0.231, 0.101, and 0.297, respectively. When the sample size was increased to more than 750, the accuracies of a_1, a_2, aa_{13}, and aa_{24} increased to more than 90%. The CVs of a_1, a_2, aa_{13}, and aa_{24} became stable and decreased slowly when the sample size was increased from 750 to 1500 (**Figure S6**). The false-positive QTLs detected in the different models resulted in irregular phenomena.

In summary, for Model I, a sample size of 150 was sufficient for detecting genetic effects; additionally, different marker densities had no significant influence on the accurate estimation of genetic effects. For model II, the sample size should be approximately 450 to accurately estimate all genetic effects; overall, for the parameters of detection power, accuracy, and variation ranges of all genetic effects, the optimum marker density was approximately 20 markers when the recombination fraction between two adjacent markers was 0.056 in M-cd1-M-cd6. For model III, the sample size should be at least 750 to accurately estimate all genetic effects. The different marker densities did not significantly influence the accurate estimation of genetic effects with simple additive models, but influenced QTL mapping accuracy for the additive and epistatic models. The optimum marker density for obtaining a relatively high QTL mapping accuracy was approximately 20 markers when the recombination fraction between two adjacent markers was 0.056 in the additive and epistatic models. Sample sizes have a large influence on QTL mapping; for Model I, a sample size of 150 was sufficient, while for model III, the sample size was increased to 750. In practical QTL mapping, the genetic model of mapping population is unknown; a larger sample size is better because more recombination events occur. Developing a mapping population with large sample size is time-consuming and costly, particularly for RIL populations. A sample size of 150 was sufficient for detecting simple additive QTLs without concerning epistatic effects and combined effects between two adjacent QTLs with short genetic distance. Our results suggest that a sample size of approximately 450 is needed to detect QTLs with genetic models of additive and epistatic effects. The sample size must be increased to approximately 750 to detect QTLs with additive, epistatic, and combined effects between two adjacent QTLs. The sample size should be increased continually to more than 750 if the genetic models of the data set become more complicated than Model III.

DISCUSSION

In practical QTL mapping, the heritability of the target QTL is fixed in a specific segregated population because heritability is a feature that belongs to the specific quantitative trait itself. Although heritability has a large influence on QTL mapping (Su et al., 2015), heritability determined by the specific quantitative trait itself cannot be altered; thus, researchers cannot increase QTL mapping accuracy by increasing heritability. The marker density of the genetic linkage map is currently thought to be one of most important factors affecting QTL mapping accuracy. The results of this study and those of previous studies (Xiong and Guo, 1997; Piepho, 2000) showed that maintaining a specific marker density was necessary for

further QTL analysis, although the QTL mapping accuracy did not increase after marker density reached a particular limit. Darvasi et al. (1993) found no significant differences in the detection power, confidence interval, and estimated genetic effects after the distance between two adjacent markers was decreased from 20 to 10 cM. Similarly, Piepho (2000) showed that there were no significant differences in the detection power and accuracy of estimating the genetic effects when the distance between two adjacent markers was increased to more than 10 cM. This may be because recombination events could not be increased by increasing marker densities, and no new recombination information could be supplied for the QTL analysis; thus, QTL mapping accuracy would not be increased. The results obtained in this study show some agreement and disagreement with those of previous studies. For simple additive models, our results and those of some previous studies indicated that different marker densities do not significantly influence the accurate estimation of genetic effects; however, for additive and epistatic models, there was an influence on QTL mapping accuracy. The optimum marker density to achieve a relatively high QTL mapping accuracy was found to be approximately 20 markers when the recombination fraction between two adjacent markers is 0.056 in the additive and epistatic models.

Sample size is one of the most important factors influencing QTL mapping. Many studies of the effect of sample size on QTL mapping have been conducted, but most of these studies only focused on simple additive effects without evaluating epistatic effects. Additionally, QTL mapping for agronomic traits revealed that QTLs controlling the same phenotype are often closely linked (Monna et al., 2002; Miyamoto et al., 2004; Ashikari et al., 2005; Thomson et al., 2006; Asano et al., 2011; Shen et al., 2011). Two linked QTLs that act in opposite directions are likely difficult to detect in a population with relatively few recombination sites, such as model IV in this study. In some species (i.e., rice), many QTLs co-localize in specific genomic regions and form QTL clusters (Yonemaru et al., 2010). Therefore, it is necessary to increase the sample size of the segregated population to accurately estimate all genetic effects. Recombination events can be increased by increasing the sample size of the segregated population. Few studies have evaluated additive and epistatic models with combined effects between two adjacent QTLs located on same chromosomes with short genetic distance. This study focused on QTL mapping accuracy under different genetic models from simple to complex. The results showed that the optimum sample size differed between genetic models; as the genetic model became more complicated, the sample size was required to be increased. For model I in this study, a sample size of 150 was sufficient for conducting QTL analysis. In Models II and III, the detection power and QTL mapping accuracy can be aided by increasing the sample sizes to 450 and 750, respectively. The sample size should be increased to more than 750 if the genetic models of data set become more complicated than Model III. Thus, to increase the detection power, recombination events should be increased by enlarging the sample size or efficiently applying limited recombination events.

CONCLUSIONS

Different marker densities had no significant influence on the accurate estimation of the genetic effects with simple additive models, but did influence QTL mapping accuracy to some degree with additive and epistatic models. The optimum marker density that may lead to a relatively high QTL mapping accuracy is approximately 20 markers when the recombination fraction between two adjacent markers is 0.056 in the additive and epistatic models.

The sample size should be increased as the numbers of different genetic effects increase to accurately estimate all genetic effects of the model. Sample sizes have a large influence on QTL mapping. In practical QTL mapping, the genetic model of the mapping population is unknown, and a larger sample size is preferential. A sample size of 150 was sufficient for detecting simple additive QTLs. A sample size of approximately 450 should be used to detect QTLs with genetic models of additive and epistatic effects. The sample size must be increased to approximately 750 to detect QTLs with additive, epistatic, and combined effects between two adjacent QTLs. The sample size should be increased continuously to more than 750 if the genetic models of the data set become more complicated than Model III.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

REFERENCES

Asano K, Yamasaki M, Takuno S, Miura K, et al. (2011). Artificial selection for a green revolution gene during japonica rice domestication. *Proc. Natl. Acad. Sci. USA* 108: 11034-11039. http://dx.doi.org/10.1073/pnas.1019490108

Ashikari M, Sakakibara H, Lin S, Yamamoto T, et al. (2005). Cytokinin oxidase regulates rice grain production. *Science* 309: 741-745. http://dx.doi.org/10.1126/science.1113373

Basten C, Weir B and Zeng Z (1994). Zmap-a QTL cartographer. Computing strategies and softwear. Proc 5th Congr on Genetics Applied to Livestock Production, Guelph, Ontario.

Darvasi A, Weinreb A, Minke V, Weller JI, et al. (1993). Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* 134: 943-951.

Edwards MD, Stuber CW and 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.

He X, Xu C, Kuai J, Gu S, et al. (2000). Principal factors affecting the power of detection and accuracy of QTL mapping. *Zuo Wu Xue Bao* 27: 469-475.

He XH, Xu CW, Kuai JM, Gu SL, et al. (2001). Principal factors affecting the power of detection and accuracy of QTL mapping. *Acta Agron. Sin.* 27: 469-475.

Kao CH, Zeng ZB and Teasdale RD (1999). Multiple interval mapping for quantitative trait loci. *Genetics* 152: 1203-1216.

Lander ES and Botstein D (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185-199.

Li H, Ye G and Wang J (2007). A modified algorithm for the improvement of composite interval mapping. *Genetics* 175: 361-374. http://dx.doi.org/10.1534/genetics.106.066811

Lu Y and Liu B (1995). A new computer package for genomic research: PGRI (plant genome research initiative). In: Abstract of plant genome III conference, San Diego, CA. http://www.intl-pag. org/3/abstracts/201pg3.html.

Manly KF, Cudmore RH, Jr. and Meer JM (2001). Map Manager QTX, cross-platform software for genetic mapping. *Mamm. Genome* 12: 930-932. http://dx.doi.org/10.1007/s00335-001-1016-3

Miyamoto N, Goto Y, Matsui M, Ukai Y, et al. (2004). Quantitative trait loci for phyllochron and tillering in rice. *Theor. Appl. Genet.* 109: 700-706. http://dx.doi.org/10.1007/s00122-004-1690-0

Monna L, Lin X, Kojima S, Sasaki T, et al. (2002). Genetic dissection of a genomic region for a quantitative trait locus, Hd3, into two loci, Hd3a and Hd3b, controlling heading date in rice. *Theor. Appl. Genet.* 104: 772-778. http://dx.doi.org/10.1007/s00122-001-0813-0

Nelson JC (1997). QGENE: software for marker-based genomic analysis and breeding. *Mol. Breed.* 3: 239-245. http://dx.doi.org/10.1023/A:1009604312050

Piepho HP (2000). Optimal marker density for interval mapping in a backcross population. *Heredity (Edinb)* 84: 437-440. http://dx.doi.org/10.1046/j.1365-2540.2000.00678.x

Shen B, Yu WD, Du JH, Fan YY, et al. (2011). Validation and dissection of quantitative trait loci for leaf traits in interval RM4923-RM402 on the short arm of rice chromosome 6. *J. Genet.* 90: 39-44. http://dx.doi.org/10.1007/s12041-011-0019-4

Su C, Qiu X and Ji Z (2013). Study of strategies for selecting quantitative trait locus mapping procedures by computer simulation. *Mol. Breed.* 31: 947-956. http://dx.doi.org/10.1007/s11032-013-9848-6

Su CF, Liu ZB and Li YG (2015). Influence of various quantitative trait loci (QTL) mapping methods on the mapping accuracy under varying heritability levels. *Genet. Mol. Res.* 14: 13003-13012. http://dx.doi.org/10.4238/2015.October.21.21

Takuno S, Terauchi R and Innan H (2012). The power of QTL mapping with RILs. *PLoS One* 7: e46545. http://dx.doi.org/10.1371/journal.pone.0046545

Thomson MJ, Edwards JD, Septiningsih EM, Harrington SE, et al. (2006). Substitution mapping of dth1.1, a flowering-time quantitative trait locus (QTL) associated with transgressive variation in rice, reveals multiple sub-QTL. *Genetics* 172: 2501-2514. http://dx.doi.org/10.1534/genetics.105.050500

Utz H and Melchinger A (1996). PLABQTL: a program for composite interval mapping of QTL. *J. Quant. Trait Loci* 2: 1-5.

Van Ooijen J and Maliepaard C (1996). MAPQTL version 3.0: software for the calculation of QTL positions on genetic maps. Plant Research International, Wageningen

Wang D, Zhu J, Li Z and Paterson A (2003). User manual for QTLMapper version 1.6-A computer software for mapping quantitative trait loci (QTLs) with main effects, epistatic effects and QTL× environment interactions. Department of Agronomy, Zhejiang University, Hangzhou, China. Department of Agronomy, Zhejiang University, Hangzhou, China (http://ibi.zju.edu cn/software/qtlmapper/).

Xiong M and Guo SW (1997). Fine-scale mapping of quantitative trait loci using historical recombinations. *Genetics* 145: 1201-1218.

Yang J, Hu C, Hu H, Yu R, et al. (2008). QTLNetwork: mapping and visualizing genetic architecture of complex traits in experimental populations. *Bioinformatics* 24: 721-723. http://dx.doi.org/10.1093/bioinformatics/btm494

Yang J, Zhu J and Williams RW (2007). Mapping the genetic architecture of complex traits in experimental populations. *Bioinformatics* 23: 1527-1536. http://dx.doi.org/10.1093/bioinformatics/btm143

Yonemaru JI, Yamamoto T, Fukuoka S, Uga Y, et al. (2010). Q-TARO: QTL annotation rice online database. *Rice (N. Y.)* 3: 194-203. http://dx.doi.org/10.1007/s12284-010-9041-z

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

Supplementary material

Table S1. Recombination fractions between two adjacent markers of M-ss1-M-ss7 for different sample sizes (50,150, 250, 350, 450, 550, and 1000) were randomly generated between 0 and 0.1.

Table S2. Recombination fractions between two adjacent markers of M-cs1-M-cs8 under different sample size conditions (50,150, 250, 350, 450, 550, and 1000) were randomly generated between 0 and 0.1.

Table S3. Correlated parameters of M-cd1-M-cd6.

Figure S1. Variation tendency of accuracy of genetic effects in M-cs1-M-cs8.

Figure S2. Variation tendency of CV of different genetic effects in M-cs1-M-cs8.

Figure S3. Variation tendency of accuracy of genetic effects in M-cd1-M-cd6.

Figure S4. Variation tendency of accuracy of genetic effects in M-tc1-M-tc11.

Figure S5. Variation tendency of CV of different genetic effects in M-tc1-M-tc11.

Figure S6. Variation tendency of CV of different genetic effects in M-tc1 to M-tc11 (A total three linkage groups with 30 markers each were simulated. Two additive QTLs located on one linkage group plus two pairs of epistetic QTLs were simulated. Sample sizes of RILs were all 350. The x axis denotes sample sizes while y axis denotes CV).