Epistasis plays an important role as genetic basis of heterosis in rice*

YU Sibin (余四斌), LI Jianxiong (李建雄), XU Caiguo (徐才国), TAN Yifang (谈移芳), GAO Youjun (高友军), LI Xianghua (李香花), ZHANG Qifa (张启发)

(State Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China)

and M. A. Saghai Maroof

(Department of Crop and Soil Environmental Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA)

Received November 20, 1997

Abstract The genetic basis of heterosis in rice was studied in a cross between Zhenshan 97 and Minghui 63, the parents of the best hybrid widely used in rice production in China. Field data for yield and yield components traits were collected over two years from 240 $F_{2:3}$ families of this cross planted in replicated field trials. These families were assayed with 151 marker loci that were polymorphic between the parents and a linkage map was constructed with Mapmaker analysis. A total of 32 quantitative trait loci (QTLs) were identified for yield and the three component traits; 12 QTLs were detected in both years and the remaining 20 QTLs were observed in only one year. A search for the entire genome by using two-way analyses of variance with all possible two-locus combinations detected a very large number of significant digenic interactions involving both the QTLs resolved by single-locus analysis and loci that did not show significant effects by single-locus analysis. Many of the interactions were simultaneously detected in both years. Partitioning of the epistatic interactions recovered all three types of interactions, i. e. additive by additive, additive by dominance and dominance by dominance, for all the traits. Some of the epistatic interactions even showed pleiotropic effects by simultaneously affecting two or more traits. The results clearly indicate that epistasis plays a significant role in the inheritance of yield traits as well as in the genetic basis of heterosis.

Keywords: hybrid rice, molecular marker, quantitative trait loci, interaction between loci.

The genetic basis of heterosis is still a debating issue. Two hypotheses, the dominance hypothesis and the overdominance hypothesis, both proposed in 1908[1—3], have competed for most part of this century. Although many researchers prefer one hypothesis to the other, experimental data allowing for critical assessment of the hypotheses remained largely unavailable until very recently with the advent of molecular markers and high density molecular linkage maps. Variable results have been obtained by single locus analyses using molecular markers. Evidences favoring both dominance[4] and overdominance hypothesis[5,6] have been presented in the literature.

Wright (1968)[7] visualized a 'net-like' structure of population genotypes such that the variations of most characters are affected by many loci and that each gene replacement may have effects on many characters. According to this perspective, epistasis should be one of the most important genetic components in the inheritance of quantitative traits as well as the genetic basis of heterosis. Results from recent studies showed that epistasis plays an important role in adaptation

* Project supported by the National Natural Science Foundation of China (Grant No. 39525014) and the Rockefeller Foundation.

and speciation[8,9]. The influence of epistasis on grain yield components has also been reported recently[10]. To assess the importance of epistasis as the genetic basis of heterosis, we conducted a molecular marker-based dissection of the genetic components involved in grains per panicle, a key element of yield, in an elite rice hybrid Shanyou 63. The results provided strong evidence that epistasis plays a significant role as the genetic basis of heterosis[11].

In this paper, we presented the results of further analyses of genetic basis of yield and yield component traits. We showed that epistasis also plays an important role as the genetic basis of heterosis in these traits.

1 Materials and methods

1.1 Experiment materials and field planting

The genetic materials were 240 F_3 families, each derived from bagged seeds of a single F_2 plant from a cross between Zhenshan 97 and Minghui 63, the parents of Shanyou 63 which is the most widely grown rice hybrid in China. The $F_{2:3}$ families together with two parents and the F_1 were transplanted to the field of the experimental farm of Huazhong Agricultural University at Wuhan in the 1994 and 1995 rice-growing seasons. The field planting followed a randomized complete block design with three replications. The plants were laid out at a distance of 17 cm between plants within a row and the rows were 27 cm apart. The field management followed essentially the normal agricultural practice. Each plant was harvested individually at maturity to prevent loss from over-ripening. Only the 15 plants in the middle of each row were used for traits scoring. The traits measured included yield, tillers per plant, grains per panicle and 1 000-seed weight.

1.2 Molecular marker assay

Approximately equal amounts of leaf tissues from 15—20 plants of each F_3 family were harvested and bulked for DNA extraction. Two classes of markers, RFLP (restriction fragment length polymorphisms) and SSR (simple sequence repeats) were used for surveying parental polymorphisms. RFLP analysis including digestion, Southern blotting, hybridization followed the method described by Liu et al. (1997)[12]. SSR analysis followed the methods by Wu and Tanksley (1993)[13]. The polymorphic markers detected between the parents were used to assay the entire population of 240 families, based on which the genotypes of F_2 individuals were deduced.

1.3 Data analysis

The estimates of means and variances for each trait were based on F_3 families. Genotype by year interaction was analyzed using a random model. Heritability (h^2) was estimated with the formula $h^2 = \sigma_g^2/(\sigma_g^2 + \sigma_{gl}^2/n + \sigma_e^2/nr)$, where σ_g^2 is the genotypic variance, σ_e^2 is the error variance, σ_{gl}^2 is the variance due to genotype by year interaction, r is the number of replications, and n is the number of years. The 90% confidence intervals on h^2 were calculated according to Knapp et al. (1985)[14].

The molecular linkage map was constructed using Mapmaker 3.0[15]. Marker-trait associations were assessed with one-way analysis of variance using Mapmanager QTb7[16]. QTLs (quantitative trait loci) were detected using Mapmaker/ QTL 1.1[17].

The entire genome was searched for digenic interactions for each trait with two-way analysis

of variance using all possible two-locus combinations of marker genotypes on the basis of unweighted cell means. The sums of squares were multiplied by the harmonic means of the cell sizes to form the test criteria[18]. There are eight degrees of freedom among the nine genotypes formed of two codominant loci (e.g. A and B), which can be partitioned into individual components: two degrees of freedom for the additive and dominance effects within each locus, four degrees of freedom for interaction (epistasis) between the two loci. The interaction can be further partitioned into four terms each specified by a single degree of freedom: additive (A locus) × additive (B locus) (AA), additive × dominance (AD), dominance × additive (DA), and dominance × dominance (DD). Statistical significance for each term was assessed using an orthogonal contrast test with the statistical package Statistica[19].

2 Results

2.1 Linkage map

A survey of 537 RFLP probes from two high density maps[20,21] and 54 SSR primer pairs resulted in a total of 151 polymorphic loci between the parents. Mapmaker analysis placed the 151 polymorphic loci into 14 linkage groups covering all 12 chromosomes. A map (fig. 1) was con-

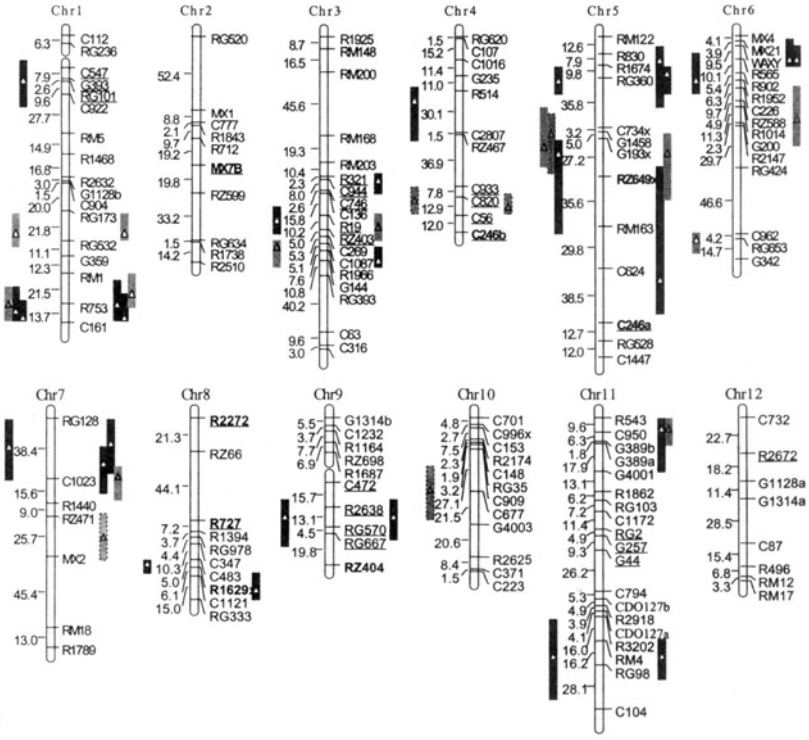

Fig. 1. The distribution of QTLs in the molecular marker linkage map constructed on the basis of F_2 individuals of Zhenshan 97/Minghui 63. Numbers on the left of each chromosome are map distance between adjacent markers. Bars indicate 1-LOD support intervals of the QTLs, which on the left and right of each chromosome represent the results of 1994 and 1995. Triangle in the bars indicates the maximum LOD position. Markers underlined indicate allelic frequencies deviated from 1:1. Markers printed in bold face indicate genotypes distorted from proportion of binomial square. ■, Yield; ▨, grains per panicle; [ ], tiller per plant; ▩, 1 000-seed weight.

structed which spanned 1 841.9 centi-Morgans (cM) in length with average interval of 12.1 cM between adjacent markers. This map well integrated the markers from the two high density RFLP linkage maps of rice.

One hundred and thirty-one of 151 loci showed codominant segregation, and the remaining 20 loci showed dominant type of segregation. Chi-square tests revealed that 24 marker loci (16%), mapped to 9 of the 12 chromosomes, deviated significantly from the expected 1:1 allelic frequencies. Eight of the 151 loci deviated from the theoretical binomial square proportions (p^2 :$2pq$:q^2), of which 5 loci showed distorted allelic frequencies.

2.2 The measurements of the traits

F_1 heterosis, measured as the percentage of deviation of the F_1 from the parental mean, was large for yield and grains per panicle. There was still large residual heterosis in the F_3 generation. Genotype variances among the $F_{2:3}$ families were highly significant ($P<0.01$) for all traits (table 1). The heritability estimates varied from one trait to another; it was the highest for seed weight (87.4%) and lowest for tillers per plant (61.7%). Genotype by year interactions was significant for some of the traits, but the interaction effects were small compared to the main effects of genotypes or years.

Table 1 Means, heterosis, variances and heritability estimated for the yield and yield component traits

Item	Yield	Tillers/plant	Grains/panicle	Seeds weight
Means				
Minghui 63	28.9(27.8)[a)]	10.7(11.0)	101.2(105.7)	27.7(25.6)
Zhenshan 97	21.5(19.9)	11.4(9.5)	81.8(89.5)	25.1(24.8)
F_1	37.9(39.6)	12.8(14.3)	114.9(111.9)	27.9(26.5)
F_3	32.9(29.4)	11.2(10.5)	112.9(110.4)	26.3(25.4)
Heterosis	50.4(65.7)	15.3(38.8)	25.6(14.7)	5.7(5.2)
Variances				
σ_g^2	14.40	0.66	179.08	4.45
σ_{gl}^2	6.05	0.12	64.60	0.44
σ_e^2	23.54	2.11	134.34	2.50
h^2	67.4	61.7	76.6	87.4
C.I.(90%)on h^2	58.9—74.5	52.8—69.4	70.5—81.4	84.2—90.1

a) Numbers out- and in-side the parentheses represent the data of 1994 and 1995, respectively.

2.3 QTLs for yield and yield component traits

Similar results were obtained by using one-way analysis of variance and interval mapping. QTLs resolved by interval mapping with LOD (logarithm of odds) threshold 2.4 were presented together with the 1-LOD support confidential intervals (figure 1).

The QTLs identified for yield, tillers per plant, grains per panicle, and 1 000-seed weight were 5, 3, 5 and 7 respectively in 1994, and 6, 2, 7 and 9 for these traits in 1995. Twelve QTLs were detected in both years as judged by their map locations and by the confidence intervals. The remaining 20 QTLs were detected only in one year or the other.

There were two noticeable features among the QTLs: (i) The QTLs showed various levels of dominance effects ranging from partial dominance, complete dominance to overdominance. Overdominance was observed at 6 of the 10 QTLs for yield, but at only a few QTLs for the three com-

ponent traits; (ii) at most of the QTLs, alleles from the higher parent, Minghui 63, contributed to the increase of the phenotypic values. However, there were also quite a few QTLs at which alleles from lower parent, Zhenshan 97, were in the direction of increasing the trait scores.

2.4 Digenic interactions (epistasis)

2.4.1 Number of locus pairs showing significant interactions. Table 2 presents the number of locus pairs that displayed significant interactions resulting from testing all possible two-locus combinations in the entire genome by two-way analyses of variance. The total numbers of valid tests were 7 585 for 1994 and 7 681 for 1995, respectively. It is highly likely that many of the epistatic interactions that attained the specified level of significance occurred by chance (false positive interactions). To delineate such false positive interactions, we calculated 99% confidence intervals for the expected numbers of spurious interactions which were 75.9 ± 20.0 for 1994 and 76.9 ± 20.1 for 1995. As can be seen from table 2, the numbers of interactions detected in both years for all the traits exceeded the expectations based on spurious interactions, indicating the existence of real interactions in the genome of this population.

Table 2 Numbers of significant digenic interactions ($P<0.01$) for yield and yield component traits

Trait	Significant interaction	1994	1995	Common	Locus type
Yield	AA	60	91	9	
	AD(DA)	51	73	9	
	DD	4	18	0	
	total types	115	182	18	
	positive pairwise loci	105	165	17	SS(1), SN(1), NN(15)
Tillers/plant	AA	79	105	20	
	AD(DA)	28	42	1	
	DD	10	6	0	
	total types	117	153	21	
	positive pairwise loci	105	141	20	SS(0), SN(1), NN(19)
Grains/panicle	AA	52	80	19	
	AD(DA)	56	74	15	
	DD	4	16	0	
	total types	112	170	34	
	positive pairwise loci	99	160	30	SS(2), SN(13), NN(15)
Seeds weight	AA	84	102	26	
	AD(DA)	47	71	15	
	DD	15	16	7	
	total types	146	189	48	
	positive pairwise loci	125	164	44	SS(4), SN(15), NN(25)
	total possible pairs tested	7585	7681		

AA, AD(DA) and DD indicate additive × additive, additive × dominance (dominance × additive) and dominance × dominance, respectively. SS, Interactions between QTLs detected by single locus analysis; SN, interactions between a QTL and a locus that did not show significant effects based on single locus analysis; NN, interaction between loci that did not show significant effects by single locus analysis.

Significant interactions were detected simultaneously in both years in many two-locus combinations, which we referred to as common interactions for ease of description. The numbers of two-locus combinations showing common interactions for yield, tillers per plant, grains per pani-

cle and 1 000-seed weight were 17, 20, 30 and 44, respectively (table 2). Theoretically, if two loci do not interact with each other, the probability for detecting the same interaction at $\alpha = 0.01$ in both years is less than 0.01%. Obviously, none of the two-locus interactions that were simultaneously observed in both years should be regarded as a chance event. These common interactions can be taken as the minimum of statistically significant interactions.

Table 2 also presents the numbers for various types of interactions resulting from the orthogonal contrast tests of the significant interactions. Three features were evident among the common interactions of the four traits: (i) Effect of each interaction term was generally small, and explaining on the average 3% of genotype variance; (ii) multiple interactions were detected in many two-locus combinations (e.g. AA and AD were present in the same two-locus combination); (iii) the total numbers of interacting loci were large and distributed in all 12 chromosomes.

2.4.2 The patterns of interactions. A few examples are given in fig. 2 to illustrate the patterns of the interactions.

(1) Additive by additive interaction. Since AA was the most common type of interactions in tillers per plant, we used the locus pair (Waxy × G342) affecting tillers per plant to illustrate the AA interaction. As can be seen from fig. 2 (a) and (b), the performance of the two homozygotes of the first locus (Waxy) was appreciably influenced by the homozygotes of the second locus (G342). The difference of tiller numbers between the two homozygotes of the Waxy locus was small, when G342 was homozygous for the Zhenshan 97 allele (22). The difference increased greatly when G342 was homozygous for the Minghui 63 allele (11), resulting in Waxy(11) /G342(22) as the favored genotype for tillers per plant.

(2) Additive by dominance interaction. Additive by dominance interactions was most common in grains per panicle, and we used the R1440 × RG236 interaction for grains per panicle to demonstrate AD interaction (fig. 2(c) and (d)). The phenotype of heterozygote of the locus R1440 was largely influenced by the homozygotes of the locus RG236. Overdominance at R1440 was detected only when RG236 was homozygous for the Minghui 63 allele (11). Conversely, the difference between the two homozygotes of the locus R1440 was enlarged, when RG236 was heterozygote compared with the two homozygotes.

(3) Dominance by dominance interaction. Seed weight showed the largest proportion of DD interaction, we thus used C347 × G4001 combination to illustrate DD interaction (fig. 2(e) and (f)). The basic characteristics of the DD interaction are a drastic reduction of the seed weight in the double heterozygote.

2.4.3 Pleiotropic effect of epistasis. There were a number of two-locus combinations each having effects on two or more traits. For examples, locus pairs RG236 × R1440 and R830 × RZ404 affected yield and grains per panicle; G342 × Waxy interaction influenced tillers per plant and seed weight; C1447 × C677 combination had effects simultaneously on grains per panicle and seed weight.

We used locus pair G342 × Waxy again to demonstrate the pleiotropic effects. The effect of this interaction on tillers per plant was described in the previous paragraph (fig. 2(a) and (b)). This two-locus interaction also had effect on seed weight (fig. 3(e) and (f)); the difference between the two homozygotes at the Waxy locus was dependent on the genotypes at the G342 locus. The seed weight difference was small when G342 was homozygous for the Zhenshan 97 allele, and

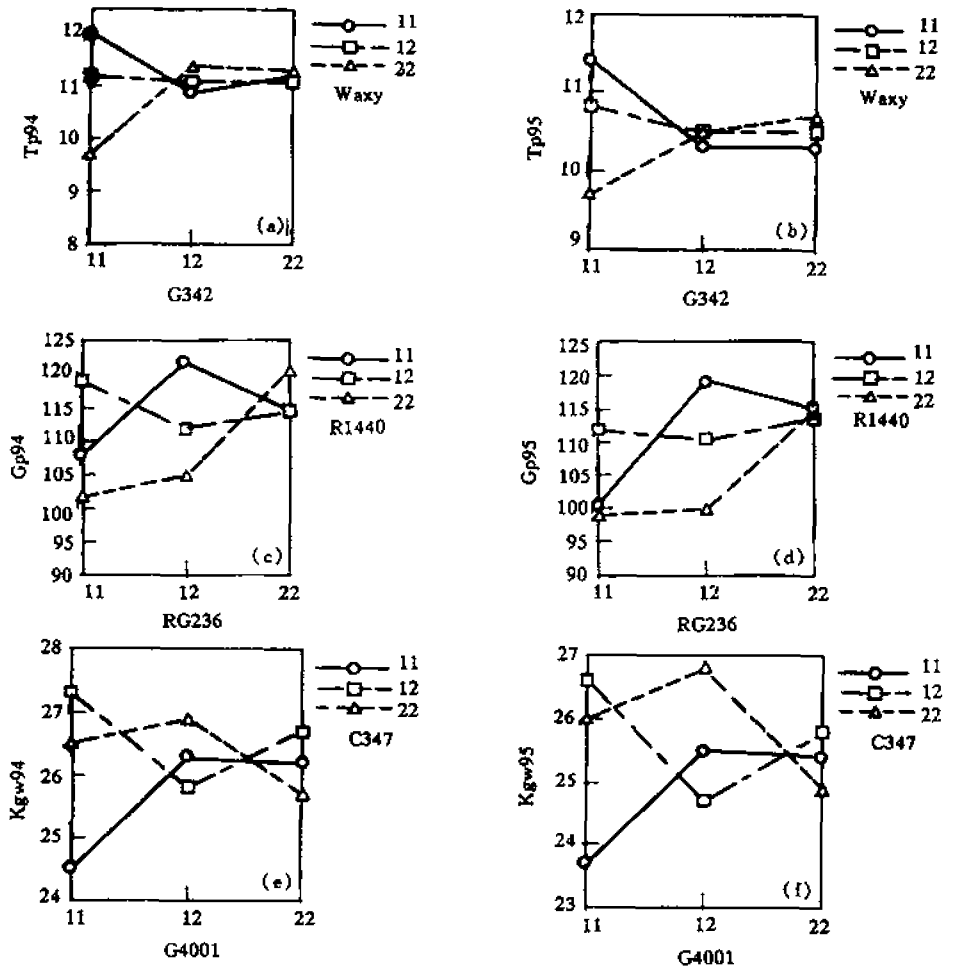

Fig. 2. Patterns of epistasis. (a), (c), (e) represent 1994 results; (b), (d), (f) represent 1995 results. 11, 12, 22 indicate the genotypes of each locus; 11, homozygote for Minghui 63 allele; 22, homozygote for Zhenshan 97 allele; and 12, heterozygote. Tp, tillers per plant; Gp, grains per panicle; Kgw, 1 000-seed weight. P_{AA}, P_{AD}, P_{DA} and P_{DD} are the probabilities for AA, AD, DA and DD interactions. (a) $P_{AA}=0.000$, (b) $P_{AA}=0.002$; (c) $P_{AD}=0.009$; $P_{DA}=0.022$; (d) $P_{AD}=0.012$; $P_{DA}=0.002$; (e) $P_{DD}=0.001$; (f) $P_{DD}=0.000$.

was greatly increased when G342 was homozygous for the Minghui 63 allele. Consequently, the two-locus combination Waxy(22)/G342(11) was the most favored genotype for seed weight.

Comparison of fig. 2(a) and (b) with fig. 3(e) and (f) clearly revealed that the same AA interaction had effects simultaneously on tillers per plant and seed weight, but in opposite directions. Waxy(22)/G342(11) combination was favorable for seed weight but unfavorable for tillers per plant, suggesting that pleiotropic effect of epistatic interactions may be an important course for genetic correlations between traits.

2.5 Influence of epistasis on single locus effect

The majority of the QTLs detected by single locus analysis interacted with at least one other locus. We used the locus pair RG173 × RM203 as an example to illustrate the influence of epistat-

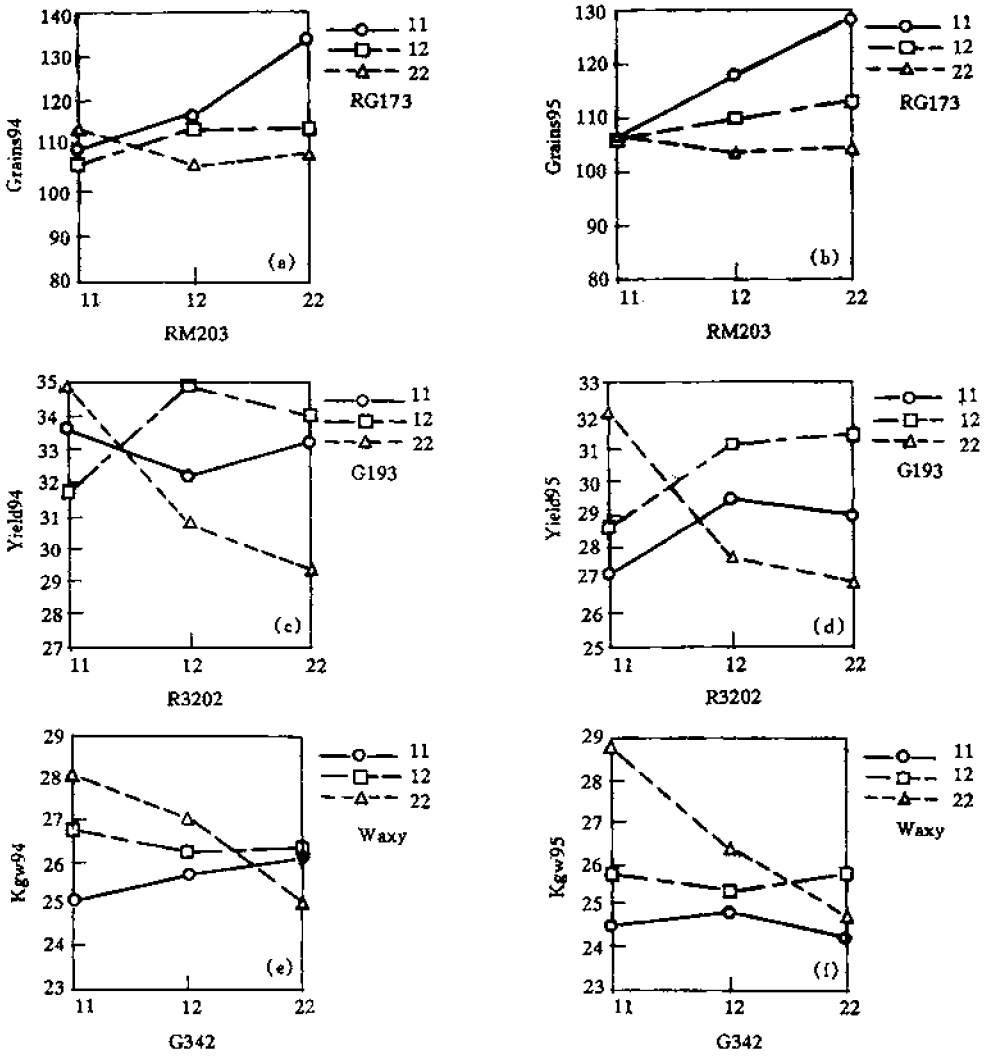

Fig. 3. Influence of epistasis on effect of single locus. See legend of fig. 2 for explanation. (a) $P_{AA}=0.002$; (b) $P_{AA}=0.002$; (c) $P_{DA}=0.004$; $P_{AA}=0.039$; (d) $P_{DA}=0.018$; $P_{AA}=0.005$; (e) $P_{AA}=0.003$; (f) $P_{AA}=0.003$.

ic interaction on the effect detected by single locus QTLs (fig. 3(a) and (b)). QTL analysis showed that RG173 is the closest marker to the QTL *gplb*, which had significant effect on grains per panicle such that alleles from Minghui 63 increased grains per panicle. However, as can be seen from fig. 3(a) and (b), the effect of RG173 alleles depended on the genotypes at the RM203 locus. The difference among the three genotypes of RG173 was very small when RM203 was homozygous for Minghui 63 allele, and increased markedly when RM203 was homozygous for the Zhenshan 97 allele.

A further example was given, using the locus pair G193 × R3202, to demonstrate the influence of interaction on overdominance of QTL (fig. 3(c) and (d)). Single locus analysis detected large overdominance effect for yield at the G193 locus. However, such overdominance was clearly dependent on genotypes at the R3202 locus. The G193 locus expressed overdominance when

R3202 was heterozygous or homozygous for the Zhenshan 97 allele, but not when R3202 was homozygous for the Minghui 63 allele.

3 Discussion

The most noticeable finding of the present studies is the prevalence of epistasis in the rice genome. Even if only the interactions simultaneously detected in both years for yield and yield component traits were considered, the significant interactions still involve a very large number of loci that are located on all 12 rice chromosomes. The number of loci involved in epistasis is much larger than those of QTLs detected by single-locus analysis.

While the dominant types of interactions (DD) may be the most relevant to F_1 heterosis, the analysis showed that AA seems to be more common than AD and DD. The deficiency of dominant types of interactions may be partly ascribed to the one generation lag that the marker data were collected on the basis of F_2 individuals and field data were obtained from F_3 families, in which dominance was reduced by half compared to F_2 individuals. Consequently, all the dominant types of effects were underestimated in the analysis. It is therefore necessary to design experiments to provide more precise estimates for the extent and significance of dominant types of interactions.

Dominance and overdominance at the single locus level were detected at many of the QTLs. However, most of the QTLs interacted with at least one other locus. Thus it may not be appropriated to interpret the single locus marginal effects without specifying the genotypes of the counterpart. Additionally, lack of correlation between heterozygosity in F_2 individuals and trait expression of F_3 families (Yu, Ph. D. dissertation 1997) implies that the effects of dominance and overdominance at single-locus level have limited contribution to heterosis observed in the population.

In summary, the results clearly indicate that epistasis plays a significant role in the inheritance of quantitative traits as well as in the genetic basis of heterosis. The genetic basis of heterosis is much more complex than it has commonly been expected on the basis of dominance and overdominance hypotheses.

Acknowledgement The authors thank the Cornell Group and the Japanese Rice Genome Project for kindly providing the RFLP probes.

References

1 Bruce, A. B., The Mendelian theory of heredity and augmentation of vigor, *Science*, 1910, 32: 627.
2 Shull, G. H., The composition of a field of maize, *Am. Breed Assoc.*, 1908, 4: 298.
3 East, E. M., Heterosis, *Genetics*, 1936, 21: 375.
4 Stuber, C. W., Lincoln, S. E., Wolff, D. W. et al., Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers, *Genetics*, 1992, 132: 823.
5 Mitchell-Olds, T., Interval mapping of viability loci causing heterosis in Arabidopsis, *Genetics*, 1995, 140: 1105.
6 Xiao, J., Li, J., Yuan, L. et al., Dominance is the major genetic basis in rice as revealed by QTL analysis using molecular markers, *Genetics*, 1995, 140: 745.
7 Wright, S., *Evolution and Genetics of Populations*, Vol. 1, Chicago: Univ. Chicago Press, 1968.
8 Allard, R. W., Genetic basis of the evolution of adaptedness in plants, *Euphytica*, 1996, 92: 1.
9 Rieseberg, L. H., Sinervo, B., Linder, C. R. et al., Role of gene interactions in hybrid speciation: evidence from ancient and experimental hybrids, *Science*, 1996, 272: 741.

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

11 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, *Proc. Natl. Acad. Sci. USA*, 1997, 94: 9226.

12 Liu, K. D., Wang, J., Li, H. B. et al., A genome-wide analysis of wide compatibility in rice and the precise location of the S5 locus in the molecular map, *Theor. Appl. Genet.*, 1997, 95: 809.

13 Wu, K. S., Tanksley, S. D., Abundance polymorphism and genetic mapping of microsatellites in rice, *Mol. Gen. Genet.*, 1993, 241: 225.

14 Knapp, S. J., Stroup, W. W., Ross, W. M., Exact confidence intervals for heritability on a progeny mean basis, *Crop Sci.*, 1985, 25: 192.

15 Lincoln, S., Daly, M., Lander, E., *Constructing Genetics Maps With MAPMAKER/EXP*3.0, Cambridge: Whitehead Institute Technical Report, 1992a.

16 Manly, K., Codmore, R., Kohler, G., *A Program for Genetic Mapping of Mendelian and Quantitative Trait Loci*, Springer-Verlag: New York Inc., 1994.

17 Lincoln, S., Daly, M., Lander, E., *Mapping Genes Controlling Quantitative Traits With MAPMAKER/QTL*1.1, Cambridge: Whitehead Institute Technical Report, 1992b.

18 Snedecor, G. W., Cochran, W. G., *Statistical Methods*, Ames: Iowa State Univ. Press, 1980.

19 StatSoft Inc., *Statistica*, Tusla: StatSoft Inc., 1991.

20 Causse, M. A., Fulton, T. M., Cho, Y. G. et al., Saturated molecular map of the rice genome based on an interspecific backcross population, *Genetics*, 1994, 138: 1251.

21 Kurata, N., Nagamura, Y., Yamamoto, K. et al., A 300 kilobase interval genetic map of rice including 883 expressed sequences, *Nature Genetics*, 1994, 8: 365.