Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid

Jinping Hua[†], Yongzhong Xing[†], Weiren Wu[‡], Caiguo Xu[†], Xinli Sun[†], Sibin Yu[†], and Qifa Zhang[†‡§]

[†]National Key Laboratory of Crop Improvement and National Center of Crop Molecular Breeding, Huazhong Agricultural University, Wuhan 430070, China; and [‡]Department of Agronomy, Fujian Agricultural University, Fuzhou 350000, China

Communicated by Gurdev S. Khush, Davis, CA, December 24, 2002 (received for review August 6, 2002)

The genetic basis of heterosis of an elite rice hybrid was investigated by using an "immortalized F_2" population produced by randomly permutated intermating of 240 recombinant inbred lines from a cross between the parents of Shanyou 63, the most widely cultivated hybrid in China. Measurements of heterosis for crosses in the immortalized F_2 population were obtained from replicated field trials over 2 years by inter-planting the hybrids with the parental recombinant inbred lines. The analyses were conducted making use of a linkage map comprising 231 segregating molecular marker loci covering the entire rice genome. Heterotic effects were detected at 33 loci for the four traits with modified composite interval mapping. The heterotic loci showed little overlap with quantitative trait loci for trait performance, suggesting that heterosis and trait performance may be conditioned by different sets of loci. Large numbers of digenic interactions were resolved by using two-way ANOVA and confirmed by randomization tests. All kinds of genetic effects, including partial-, full-, and overdominance at single-locus level and all three forms of digenic interactions (additive by additive, additive by dominance, and dominance by dominance), contributed to heterosis in the immortalized F_2 population, indicating that these genetic components were not mutually exclusive in the genetic basis of heterosis. Heterotic effects at the single-locus level, in combination with the marginal advantages of double heterozygotes caused by dominance by dominance interaction at the two-locus level could adequately explain the genetic basis of heterosis in Shanyou 63. These results may help reconcile the century-long debate concerning the genetic basis of heterosis.

hybrid vigor | "immortalized F_2" population | molecular marker | heterotic loci | epistasis

Utilization of heterosis has become a major strategy for increasing productivity of plants and animals. For crop plants, hybrid varieties have contributed greatly worldwide to the production of many crop species, including the most important food crops such as maize and rice (1, 2).

There has also been considerable interest in the genetic basis of heterosis. Two hypotheses, the dominance hypothesis (3) and the overdominance hypothesis (4, 5), were proposed early last century to explain the genetic basis of heterosis. Although many investigators favored one hypothesis over the other (6), data allowing for critical assessment of the hypotheses remained largely unavailable until very recently with the advent of molecular marker technology and high-density molecular linkage maps. Genetic analyses of heterosis based on molecular marker linkage maps have been reported recently in maize and rice. Stuber *et al.* (7), who analyzed the genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines, showed that heterozygotes of almost all quantitative trait loci (QTL) for yield had higher phenotypic values than the respective homozygotes. They suggested that both overdominance and QTL detected by single locus analysis play a significant role in heterosis. On the other hand, Xiao *et al.* (8), who conducted a genetic study of quantitative traits in an intersubspecific cross of rice, suggested that dominance was the genetic basis of heterosis in rice. Both the dominance and overdominance hypotheses are based only on single locus theory.

Wright (9) 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. Based on this perspective, epistasis should be one of the most important genetic components in the inheritance of quantitative characters. Epistasis was also speculated to contribute to the genetic basis of heterosis (6). More recently, Yu *et al.* (10) analyzed the genetic components underlying yield and its component traits by using an $F_{2:3}$ population derived from a highly heterotic rice cross, and detected a large number of digenic interactions involving loci distributed throughout the entire rice genome. They suggested that epistasis plays an important role as the genetic basis of heterosis. Li *et al.* (11) and Luo *et al.* (12) also suggested that epistasis and overdominance are the primary genetic basis of inbreeding depression and heterosis.

It is critical to have the appropriate experimental design and materials for the genetic analysis of heterosis. It is well known that the F_2 generation provides theoretically the most complete and most informative population for many genetic analyses (13). However, there are disadvantages associated with using the F_2 for genetic analysis of quantitative traits. For example, each genotype in an F_2 population is represented by only one individual, which makes it difficult to obtain replicated measurements of the same genotype. Also, the population is in a transient state, so the experiment cannot be repeated. Although genetic analyses using F_3 populations can produce useful information regarding the genetic components underlying heterosis (10, 14), such analyses suffer from several shortcomings that are inherent with the data generated from the populations (15). Populations derived by backcrossing recombinant inbred lines (RILs) with the parents have been used for genetic analyses of heterosis (7, 8, 11, 12), as variants of the design III (16). For QTL analyses, however, such populations are incomplete in terms of genotypic composition and thus not able to provide estimates for some of the genetic components at both single- and multilocus levels that are necessary for unraveling the genetic basis of heterosis.

Moreover, all previous molecular marker-based genetic analyses of heterosis, except one case using the backcross type of populations (11, 12), were based on the performance measurement of the trait rather than heterosis, and the genetic basis of heterosis was inferred from genetic components estimated for the trait performance. Although heterosis and trait performance are closely related, they are nonetheless distinct in many respects both statistically and biologically. Thus, to have a real picture of the genetic components underlying heterosis, it is necessary to use the measurements of heterosis as the data input in the analyses.

Abbreviations: QTL, quantitative trait locus; HL, heterotic loci; AA, additive by additive interaction; AD, additive by dominance interaction; DA, dominance by additive interaction; DD, dominance by dominance interaction; RIL, recombinant inbred line.

[§]To whom correspondence should be addressed. E-mail: qifazh@mail.hzau.edu.cn.

 www.pnas.org/cgi/doi/10.1073/pnas.0437907100

Rice is the staple food for a large segment of the world population. The success of hybrid rice breeding (2), together with the relatively small genome size (17), saturated molecular linkage maps (18, 19), and rapid advances in genome sequencing (20, 21), have provided a rare opportunity for dissecting the genetic basis of heterosis. In the study reported in this paper, we investigated the genetic components conditioning heterosis of yield and yield component traits in an elite rice hybrid, using an experimental design that produces an "immortalized F_2" population based on a RIL population (22). The main objective was to characterize the genetic components underlying heterosis in this hybrid, which may also have general implications in understanding the genetic basis of heterosis.

Materials and Methods

Design and Construction of the Immortalized F_2 Population. A population of 240 F_9 RILs, derived by single seed descent from a cross between two elite indica lines, Zhenshan 97 and Minghui 63, was intermated following a design for constructing an immortalized F_2 population. These two lines were the parents of Shanyou 63, the most widely cultivated hybrid with a planting area of ≈6.7 million hectares per year during its peak period in the late 1980s and early 1990s, accounting for ≈25% of the rice production during that period in China.

In this design, crosses were made between the RILs chosen by random permutations of the 240 RILs. In each round of permutation, the 240 RILs were randomly divided into two groups, and the lines in the two groups were paired up at random to provide parents for 120 crosses. Each of the 240 RILs was used only once in each round of pairing and crossing. This procedure was repeated three times, resulting in a population consisting of 360 crosses. This population resembles an F_2 population in that the compositions and frequencies of single- and multilocus genotypes are similar to those of an F_2 population.

Because there is a wide range difference in heading dates between the lines that were assigned as the parents for crossing, all of the RILs were planted in the nursery at intervals of 7–10 days, and the seedlings were transplanted in pairs according to the mating design. The planting and crossing were carried out in the four consecutive growing seasons of the summer (Wuhan, China) and winter (Hainan, China) of 1997–1998. At least 200 hybrid seeds per cross were produced by hand emasculation and hand pollination.

Field Planting and Examination. Field tests of the immortalized F_2 population were conducted in the rice growing seasons of 1998 and 1999. Adequate seeds were obtained for 324 crosses for the 1998 test and 358 crosses for the 1999 test. The two parents, Zhenshan 97 and Minghui 63, and their F_1 hybrid, or Shanyou 63, were also included in the field test. The field planting followed a randomized complete block design with two replications. Each plot consisted of four rows with 10 hills each: two rows of the F_1 resulting from the crossing and one row for each of the respective parents. Seedlings ≈35 days old for all of the experimental materials were transplanted to a bird-net-equipped field, with a layout of 26.5 cm between plants within a row and 33.3 cm between the rows, on the experimental farm of Huazhong Agricultural University in Wuhan, China. The field management followed essentially the normal agricultural practice.

True hybrid plants were determined by careful comparison of morphological characters with the parents throughout the growing season. In case such field examination failed to distinguish between the hybrids and the parents, simple sequence repeat markers were used to determine the hybrid plants.

Each row was harvested individually at its maturity to prevent loss from over-ripening. Only the eight plants in the middle of each row were used for scoring. Traits examined included yield per plant measured as the weight of all filled grains of the plant, which was converted to metric tons per hectare, tillers per plant scored as the number of seed-setting tillers per plant, grains per panicle scored as the total number of filled grains per plant divided by the number of reproductive tillers, and grain weight as the weight (g) of 1,000 seeds.

Molecular Markers and Linkage Maps. The molecular marker data for the RIL population were essentially as described (22), except that additional simple sequence repeat markers were added in certain regions to reduce gaps. The genotype for each cross in the immortalized F_2 population was deduced on the basis of the RILs that were used as the parents for the cross. Molecular marker linkage maps were constructed by using MAPMAKER (23).

Detection of Heterotic Loci (HL). Two data sets were generated in this experiment. The first consisted of direct measurements of the traits that were referred to as performance, of which the analysis was reported previously (15). As the second data set, mid-parent heterosis of each cross was calculated as $H = F_1 - (P_1 + P_2)/2$ (where H is the amount of heterosis, F_1 is the trait measurement of the hybrid, P_1 and P_2 are the measurements of the parents), and used as the input data for analyzing the genetic basis of heterosis. In detecting HL by using a single locus model, we assumed that the two homozygotes of each locus did not contribute to heterosis. A locus showing significant difference in heterosis between the heterozygote and the mean of the two homozygotes was considered to be a HL. This idea for detecting HL was incorporated in a composite interval mapping model: $h = d^*z^* + \Sigma_i d_i z_i + \varepsilon$, where d is the difference in heterosis between the heterozygote and the mean of the two homozygotes, z is indicator variable taking values 1 and 0 for heterozygote and homozygotes respectively, the asterisk indicates the putative HL to be tested, the subscript i indicates the ith cofactor marker, and ε is residual. Cofactors were selected by stepwise regression. Thresholds for logarithms of odds were estimated by permutation tests (24).

Detection of Digenic Interactions. The entire genome was searched for the effects of digenic interactions on heterosis for each trait with two-way ANOVA using all possible two-locus combinations of marker genotypes. The calculation was based on unweighted cell means (25) and the sums of squares were multiplied by the harmonic means of the cell sizes to form the test criteria. There would be >20,000 possible two-locus combinations for data of 231 marker loci, and for tests at $P \leq 0.001$, ≈0.1% of the F values were expected to exceed the threshold $F_{0.001}$ value by chance alone (false positive interaction). To further assess the likelihood for each of the significant interactions to be a chance event, a randomization test was conducted to identify the interactions that are more likely to be "really significant" than others. In conducting such a test, the entry order of heterosis data in the analysis was randomly permutated and the F statistic for the digenic interaction was recalculated by using the same two-locus marker data. This procedure was repeated 1,000 times, and the resulting 1,000 F values were compared with the F from the real data. If no more than one F value from the random permutations was larger than the F from the real data, the digenic interaction was regarded as significant at $P \leq 0.001$.

Each of the significant digenic interactions was further partitioned into four terms each specified by a single degree of freedom: additive (first locus) × additive (second locus) (AA), additive × dominance (AD), dominance × additive (DA) and dominance × dominance (DD). Statistical significance for each of the terms was assessed by using an orthogonal contrast test provided by the statistical package STATISTICA (26).

Results

The Amounts of Heterosis for Yield and Yield Component Traits. The values of midparent heterosis for yield and yield component traits of the immortalized F_2 population are given in Table 1. For the hybrid Shanyou 63, yield showed the largest amount of heterosis in

Table 1. Midparent heterosis of yield and yield component traits in Shanyou 63 and the immortalized F_2 population

	Shanyou 63		Immortalized F_2			
Trait*	Heterosis	%	Mean	%	Range	%
1998						
Yield	3.7	126.4	1.8	57.3	−2.4 to 4.2	−68.9 to 217.2
Ti/pl	2.3	17.8	4.0	31.1	−6.0 to 12.0	−44.0 to 108.9
Gr/pa	63.6	81.9	11.8	14.4	−34.6 to 58.6	−36.6 to 88.7
Gr wt	2.1	8.5	1.2	5.1	−3.5 to 4.9	−13.8 to 23.0
1999						
Yield	2.9	120.5	1.3	43.7	−1.4 to 4.7	−33.5 to 225.7
Ti/pl	4.5	51.3	2.4	22.5	−3.5 to 9.2	−27.4 to 116.3
Gr/pa	46.4	53.8	15.1	15.6	−40.1 to 73.7	−30.5 to 90.1
Gr wt	2.8	11.8	0.7	3.0	−2.9 to 5.4	−10.3 to 43.6

*Yield, measured in tons per ha; Ti/pl, tillers per plant; Gr/pa, grains per panicle; Gr wt, weight (g) of 1,000 grains.

both years. Among the component traits, grains per panicle exhibited the highest heterosis followed by tillers per plant, and grain weight expressed the lowest heterosis. The amounts of heterosis varied widely for all of the traits in the immortalized F_2 population, from highly negative to highly positive. Again, yield showed the highest heterosis on the average in both years. Among the component traits, however, tillers per plant showed the highest heterosis followed by grains per panicle.

Molecular Marker Linkage Maps. Molecular marker linkage maps consisting of 231 polymorphic loci were constructed for both the RIL and immortalized F_2 populations. The map constructed for the immortalized F_2 population, using the deduced marker genotypes based on RILs, spanned a total of 2,646.1 centimorgans (cM), which was longer than the map of 2,007.3 cM based on the RIL population (not shown). This is understandable because the map construction using RILs took into consideration of the multiple crossovers in RILs, whereas the software for map construction of the immortalized F_2 did not consider multiple crossovers. We thus used the map based on the RIL data for further analyses.

HL for Yield and Yield Component Traits. Two thousand random permutations for each of the four traits revealed that, with an experiment-wise error rate 0.10, the threshold for logarithm of odd ranged between 2.4 and 2.5 for the four traits in both years. We thus used 2.4 as the threshold for claiming HL, which identified a total of 33 HL in the 2 years for heterosis of the four traits (Table 2). Only two of the HL were detected in both years, and the remaining 31 HL were detected only in one year.

For yield, four and five HL were detected in 1998 and 1999, respectively. Two HL, *hyd9a* and *hyd12*, located on chromosomes 9 and 12, respectively, showed negative effects such that heterozygotes had lower heterosis values than the means of the two homozygotes. The other seven HL showed positive effects; the heterozygotes had higher heterosis values than the respective means of two homozygotes.

For tillers per plant, four and two HL were resolved in 1998 and 1999, respectively. However, positive heterotic effects were detected at three HL, whereas negative heterotic effects were detected at the other three HL.

Five and four HL were detected for grains per panicle in 1998 and 1999, respectively. One of the HL detected in 1998 and another detected in 1999 showed negative effects.

The largest number of HL was resolved for grain weight, though the smallest amount of heterosis was observed for this trait. Seven and five HL were detected in 1998 and 1999, respectively. Interestingly two HL that showed negative heterotic effects were consistently detected in both years.

We compared the HL with the QTL that were resolved in the trait analysis of the same immortalized F_2 population (15). Ten of the 33 HL were also detected in the QTL analysis by the composite interval mapping with about the same level of threshold (Table 2).

Table 2. Putative HL identified for yield and yield components traits in the immortalized F_2 population

Trait*	HL	Flanking markers	LOD†	h‡	A§	D§	d/a¶
1998							
Yield	*yd1*	C567-C2340	4.1	0.62	0.03	0.05	2.0
	yd3	RG393-G144	3.1	0.39	0.02	0.07	4.5
	yd6‖	RG653-G342	2.4	0.65	0.02	0.22	11.7
	yd8	G2132-R727	2.7	0.49	−0.16	0.08	0.5
Ti/pl	*tp6b*	RM225-C1496	3.0	−1.80	−0.24	−0.16	−0.7
	tp6c‖	RZ588-P	3.3	−1.12	0.72	−0.85	−1.2
	tp9	RM215-R1952b	2.6	1.32	−0.14	0.23	1.7
	tp12	C909B-RM17	2.9	1.57	−0.44	0.49	1.1
Gr/pa	*gp1*	RM237-C922	2.6	5.70	−3.56	6.02	1.7
	gp4	G235-R78	2.7	5.55	0.15	−0.48	−3.1
	gp8	RG333-C1121	2.6	8.43	−1.00	1.16	1.2
	gp10	C153A-RM222	2.6	−5.66	0.86	0.33	0.4
	gp11	RM20a-C104	2.4	6.17	−1.96	4.33	2.2
Gr wt	*gw3*‖	RZ403-C1087	3.9	−0.70	1.47	−0.47	−0.3
	gw4	G102-RM255	2.4	0.50	0.01	0.45	40.2
	gw6b‖	Y4073L-C751A	3.8	−0.76	0.11	0.11	1.0
	gw6c	C962-RZ242	3.3	0.64	0.30	−0.23	−0.7
	gw11a‖	RM209-RM229	3.0	1.01	0.16	0.05	0.3
	gw11b	RG2-RM21	3.0	−1.00	0.05	0.21	4.1
	gw12	C966-G1128a	4.5	−0.79	−0.57	0.08	0.1
1999							
Yield	*yd5*	R830-R3166	2.4	0.04	0.04	0.13	3.0
	yd7‖	C1023-R1440	5.1	0.55	0.31	0.31	1.0
	yd9a	C1232-R1164	3.4	−0.42	0.04	−0.12	−3.2
	yd9b	RM219-RZ698	3.1	0.56	−0.03	−0.02	−0.8
	yd12	C909B-RM17	2.9	−0.50	0.09	−0.17	−2.0
Ti/pl	*tp6a*	Wx-RM204	2.9	−0.68	0.14	−0.21	−1.5
	tp11	G257-RM209	2.9	1.21	−0.32	0.45	1.4
Gr/pa	*gp6*	RZ667-RG424	2.8	−10.40	−2.89	−2.42	−0.8
	gp7‖	C1023-R1440	5.5	13.20	8.04	5.84	0.7
	gp12	C732-RM20b	2.6	4.08	−1.40	3.81	2.7
Gr wt	*gw1*‖	C86-RG236	5.9	0.86	0.39	0.31	0.8
	gw3‖	RZ403-C1087	2.8	−0.41	1.39	−0.35	0.3
	gw6a	Y4073L-C751A	3.1	−2.44	−0.03	0.11	3.8
	gw6b‖	R2869-C474	3.2	0.64	−0.17	0.30	1.8
	gw10	C153A-RM222	3.2	0.45	−0.24	0.30	1.3

*See footnotes of Table 1 for the abbreviations of the traits.
†LOD, logarithm of odds.
‡Heterotic effect of the HL defined as the difference in heterosis between the heterozygote and the means of the two homozygotes.
§Additive and dominance effects calculated using the performance data at the exact genomic location of the HL.
¶Ratio of dominance to the absolute value of additive effect.
‖QTL were detected at these loci in trait analysis (15).

The Amounts of Dominance at the HL. To further characterize the genetic effects of the loci showing significant heterotic effects, we calculated the additive and dominance effects by using the performance data at the exact genomic locations of the HL (Table 2) by using composite interval mapping (27), even though QTL were not detected at most of the HL (15). Theoretically, a ratio of estimated dominance to the absolute value of additive effect (d/a) larger than unity is regarded as exhibiting overdominance, and a ratio between 0 and 1 represents partial dominance. The d/a at each of the HL is listed in Table 2, from which it can be seen that degrees of dominance at the HL varied from partial dominance, full dominance, to overdominance, with a number of the HL showing large overdominance.

Digenic Interactions Across the Entire Genome. Numbers of two-locus combinations that showed significant interactions as detected by

Table 3. Numbers of significant interactions for heterosis of yield and yield component traits identified by searching all possible two-locus combinations and confirmed by randomization tests

	Whole genome searching*			Randomization test†		
Trait	1998	1999	Common	1998	1999	Common
Yield	74	95	6	38	61	6
Tiller/plant	95	109	2	60	65	2
Grains/panicle	77	94	6	43	64	5
Grain weight	104	149	3	69	89	3

*Significant at $P \leq 0.001$ identified by the whole genome search.
†Interactions that were identified by the randomization test in one year and significant at $P \leq 0.001$ identified by the whole genome search in the other year are also listed.

using two-way ANOVA at $P \leq 0.001$ are listed in Table 3. Only two-locus data sets with all of the marker genotypic classes containing 5 or more crosses in the immortalized F_2 population were included in the calculation, resulting in 23,791 tests for 1998 and 24,259 tests for 1999. At a false-positive rate of 0.001 for individual tests, the expected numbers of interactions due to chance would be ≈24 in each year. Although the numbers of significant interactions detected were much larger than this, it is noted that some of the significant interactions declared at $P \leq 0.001$ by the whole genome search may be false positives.

To further assess the likelihood of the interactions identified above as chance events, each of the declared significant interactions was subjected to randomization test as described in *Materials and Methods*. The numbers of significant interactions so identified are also listed in Table 3, from which it can be seen that the numbers of significant interactions reduced by the randomization tests were much larger than the numbers of spurious interactions expected by chance events, indicating that the randomization test was highly stringent in identifying the digenic interactions.

A number of interactions simultaneously survived the randomization tests in both years for each of the traits (Table 3), and a larger number of significant interactions were detected in 1999 than 1998. The results of the two years are consistent in that the largest number of significant interactions was detected for grain weight followed by tillers per plant, with the smallest numbers of interactions detected for yield.

Table 4. Summary of significant interactions for the heterosis of yield and yield component traits based on the randomization test

Trait	Interaction*	1998	1999	Common
Yield	Positive pairs	38	61	6
	AA	38	49	6
	AD (DA)	4	33	0
	DD	2	11	0
Tiller/plant	Positive pairs	60	65	2
	AA	36	56	0
	AD (DA)	33	29	0
	DD	7	6	0
Grains/panicle	Positive pairs	43	64	5
	AA	38	55	3
	AD (DA)	23	26	4
	DD	2	6	0
Grain weight	Positive pairs	69	89	3
	AA	60	82	3
	AD (DA)	19	28	0
	DD	7	6	0

*The numbers of positive pairs are the same as the numbers under randomization test in Table 3. The cutoff point for the interaction terms (AA, AD/DA and DD) is at $P \leq 0.01$.

The numbers of three different types of interaction terms (AA, AD/DA, and DD) partitioned by using the orthogonal contrast tests for the two-locus combinations that were confirmed by the randomization tests are given in Table 4. In all four traits, AA occurred at predominantly high frequencies, followed by AD/DA, with DD being the least frequent.

Effects of Interactions. According to the coefficients used in the orthogonal contrasts (28), the test for an AA in reality provides a comparison for the four homozygotes of the two loci involved. The test for an AD compares the relative performance of the hetero-

Table 5. Comparative advantages in the amount of heterosis for grain per panicle of the best homozygote in each of the two-locus combinations showing significant AA

			Best homozygote		
Locus 1*	Locus 2*	Var % by AA	Genotype†	Over midparent‡	Over Minghui 63§
R753 (1)	RG128 (7)	5.00	22/22	2.08	4.17
RM259 (1)	C909B (12)	4.22	11/11	2.33	0.00
RM243 (1)	C87 (12)	4.00	22/22	0.55	1.10
RM243 (1)	R496 (12)	6.19	11/11	1.28	0.00
RM243 (1)	C909B (12)	7.00	11/11	1.02	0.00
RG173 (1)	RZ471 (7)	4.04	22/11	17.71¶	15.83¶
C112 (1)	C153B (9)	3.51	22/22	0.22	0.45
C112 (1)	Y2668LA (11)	3.43	22/22	2.88	5.77
C112 (1)	G181 (11)	3.68	22/22	3.85	7.69**
RG634 (2)‖	C347 (8)	1.90	11/22	9.76**	6.27
RZ386 (2)	P (6)	2.82	11/22	22.12¶	19.72¶
RM240 (2)	G102 (4)	4.90	22/11	15.80¶	22.12¶
RM213 (2)	RM20b (12)	3.21	22/22	4.05	8.09**
RM232 (3)	RM26 (5)	3.34	11/11	6.30	0.00
RM200 (3)	RZ66 (8)	4.70	22/11	12.30††	13.49††
C820 (4)	C1447 (5)	3.67	11/11	2.73	0.00
C820 (4)	R2549 (6)	4.35	22/22	4.83	9.67‖
C820 (4)	RM26 (5)	3.37	11/11	3.07	0.00
C933 (4)	C1447 (5)	4.42	11/11	4.42	0.00
C933 (4)	R2549 (6)	5.81	22/22	5.79	11.57††
R3166 (5)	C732 (12)	3.77	11/11	1.50	0.00
C246 (5)	C751B (4)	3.33	11/11	3.00	0.00
C1447 (5)	C751B (4)	3.80	11/11	4.40	0.00
RM31 (5)	P (6)	3.24	11/22	23.78¶	21.99¶
R2869 (6)	RM234 (7)	2.75	22/11	22.53¶	24.44¶
C474 (6)	RM234 (7)	3.01	22/11	22.73¶	21.60††
C764 (6)	C347 (8)	4.35	22/22	6.83	13.65††
C764 (6)	RG978 (8)	4.19	22/22	7.15	14.30¶
RZ398 (6)	C347 (8)	3.23	22/22	7.15	14.30††
R1014 (6)	RM12 (12)	2.78	22/11	13.73††	11.88††
P (6)¶	RM12 (12)	2.72	22/11	13.10††	6.87
R2147 (6)	C153A (10)	1.84	22/11	16.10¶	11.23††
Y4073L (6)	C153A (10)	1.98	22/11	16.27¶	10.49††
C751B (4)	RM26 (5)	3.68	11/11	3.19	0.00
RG528 (7)	G1149 (8)	3.72	11/11	0.73	0.00
RM26 (5)	C477 (9)	3.21	11/22	25.47¶	27.68¶
R1952b (9)	C87 (12)	3.25	22/22	6.56	13.13††
RZ404 (9)	C87 (12)	2.55	22/22	6.31	12.61††

The digenic interactions are identified by randomization tests and the cut-off for AA is at $P \leq 0.01$.
*The numbers in parentheses indicate the chromosomal locations of the markers.
†Genotype of the first locus/genotype of the second locus: 11, homozygous for the Minghui 63 allele; 22, homozygous for Zhenshan 97 allele.
‡Advantage of the best homozygote over the mean of the two parental genotypes.
§Advantage of the best homozygote over the Minghui 63 genotype.
¶Significantly different from zero at $P = 0.001$.
‖DD was also detected for this two-locus pair.
**Significant different from zero at $P = 0.05$.
††Significantly different from zero at $P = 0.01$.

GENETICS

Table 6. Comparative advantage of the double heterozygote in the amount of heterosis in each of the two-locus combinations showing significant DD for grain per panicle in 1998

			Double heterozygote					
Locus 1*	Locus 2*	Var % by DD	Heterosis	Over midparent‡	Over best homozygote	Over best genotype	Best homozygote†	Best genotype†
RM243 (1)	RM234 (7)	2.11	11.24	9.39§	−2.74	−6.67¶	11/22	12/11
RG634 (2)‖	C347 (8)	3.38	10.43	5.97	−3.79	−5.90¶	11/22	22/12
RM31 (5)	RM201 (9)	3.76	14.05	3.17	−5.00	−5.00	11/22	11/22
P (6)	R496 (12)	2.13	10.29	4.71	−7.31¶	−13.00**	22/11	22/12
P (6)‖	RM12 (12)	3.55	11.54	11.49**	−1.61	−14.02**	22/11	22/12

For illustration, the digenic interactions are declared at $P \leq 0.001$ level by the whole genome search strategy and the cut-off for DD is at $P \leq 0.01$.
*The numbers in parentheses indicate the chromosomal locations of the markers.
†Genotype of the first locus/second locus; 11, homozygous for the Minghui 63 allele; 22, homozygous for Zhenshan 97 allele; 12, heterozygote.
‡Mean of the two parental genotypes.
§Significantly different from zero at $P = 0.01$.
¶Significantly diferent from zero at $P = 0.05$.
‖AA was also detected for this two-locus pair.
**Significantly different from zero at $P = 0.001$.

zygote against the two homozygotes at one locus under the backgrounds of the two homozygotes of the other locus, or vice versa. The test for a DD provides a measurement for the performance of the heterozygote relative to the two homozygotes at one locus against the performance of the heterozygote relative to the two homozygotes at the other locus.

To illustrate the effects of digenic interactions on heterosis, we listed in Table 5 the comparative advantages for the best homozygotes of the two locus-pairs showing significant AA, and in Table 6 we listed advantages of the double heterozygotes for two-locus pairs showing significant DD for grains per panicle detected in 1998. Several points can be made from these tables.

In two-locus combinations showing significant AA, the parental two-locus homozygotes (11/11 or 22/22) in many cases showed marginal advantages, as compared with the means of the two parental genotypes (11/11 and 22/22) (Table 5). However, the complementary two-locus homozygotes (11/22 or 22/11) frequently appeared to be the best genotypes by showing large deviations from the means of the two parental genotypes as well as the Minghui 63 genotypes (Table 5). In two-locus combinations showing significant DD, all of the double heterozygotes (12/12) appeared to be advantageous compared with the means of the two parental genotypes (Table 6). However, the best two-locus genotypes were frequently those that were homozygous at one locus and heterozygous at the other locus (11/12, 12/11, 22/12 or 12/22), or single heterozygotes.

We also examined the heterotic values of two-locus combinations showing AD/DA interactions (data not shown). The general trend is similar. Namely, the parental two-locus homozygotes (11/11 or 22/22) had marginal advantages in some cases; the complementary two-locus homozygotes (11/22 or 22/11) frequently showed large effects on heterosis; single heterozygotes (11/12, 12/11, 22/12, and 12/22) can also be the best two-locus genotypes in a few cases. It should be noted that in no case did the double heterozygote show the highest level of heterosis.

Discussion

The Usefulness of the Immortalized F_2 Population for Heterosis Study. This study demonstrated the use of the immortalized F_2 population in genetic analyses of heterosis. For study of heterosis, such a population possesses several distinct advantages. First, the genotypes and their proportions in this population are similar to those of an F_2 population (i.e., 1:2:1 for single locus genotypes, 1:2:1:2:4:2:1:2:1 for two locus combinations, etc.). Thus such a population is clearly as informative as an F_2 population. However, instead of only one individual per genotype represented in an F_2 population, each genotype in this population is represented by as many plants as one wishes, thus permitting replicated trials in multiple environments. The whole population can be recreated when needed, either in exactly the same way, or by different permutation schemes. Second, plants used for measuring heterosis are hybrids rather than progenies of self-fertilization. This may provide particular value for heterosis analyses, if epigenetic changes are involved in heterosis or inbreeding depression. Third, molecular marker data need be collected from only the 240 RILs no matter how many crosses are included in the population. Most importantly, as demonstrated in the present study, it provides opportunities for mapping and genetic analysis of heterosis per se, rather than analyses based on performance measurements of the trait.

Detection of HL. Detection of HL by using the heterosis measurements enabled by the immortalized F_2 population represents another novel feature of the study. The detection by the modified composite interval mapping not only nullified the noise effected by the homozygotes of each locus, but also enabled the determination of the precise locations of the HL. Such analysis effectively separated the single-locus effects that cause heterosis from the QTL conditioning the trait performance as analyzed in all reported QTL studies. Making use of the immortalized F_2 population, we detected and mapped a total of 33 HL on 11 of the 12 rice chromosomes for yield and three other traits that are components of yield. An important feature revealed in the analysis is the high degree of overdominance in many HL when calculated by using the performance data, although many of the overdominance effects could not be detected by QTL analysis with the given threshold. Thus, the results imply that many of the HL identified in this analysis represent overdominance effects when the heterotic effects are interpreted in terms of single-locus genetic effects.

The fact that only 10 of the 33 HL identified in this analysis were detected by QTL analysis using the data of the same population collected from the same experiment that identified a total of 40 putative QTL for the four traits seems to indicate that trait performance and heterosis are conditioned by different sets of loci. Such results also demonstrated the unique usefulness of the data from heterosis measurements for the detection of HL, as well as for the characterization of the genetic basis of heterosis.

It seems surprising that only two of the 33 HL were simultaneously detected in both years. The main reason for this small number is that the heterotic effect defined in the analysis was a derived statistic that was converted twice in our analysis, first as the difference between the hybrid and the mean of the parental lines, and second as the difference between the heterozygote and the mean of the homozygotes of the locus. Thus, each measurement depends on many other genotypes, which made the heterosis

measurements as well as the HL highly sensitive to environmental influence.

The Genetic Basis of Heterosis. The analyses clearly demonstrated that the experimental design and analytical methods could resolve all of the genetic elements that are possible in any of the hypotheses concerning the genetic basis of heterosis. The challenge now is how to put the pieces together to frame a picture for comprehending the genetic basis of heterosis. For such a purpose, it is necessary to make a clear distinction between the heterosis observed in the immortalized F_2 population, and the heterosis expressed in the F_1 hybrid that we intended to elucidate in this study.

Because of the complexity of yield as a trait, which makes it difficult to follow the effects of genetic components resolved for this trait, we chose grains per panicle as the trait for detailed analysis, as this trait is less complex and the results may intuitively be more comprehensible. Other advantages of using this trait include the high level of heterosis and also high heritability that have been repeatedly observed in previous studies (10, 15, 29, 30).

The analyses of the immortalized F_2 population showed that all kinds of genetic effects, including heterotic effects (because of partial-, full-, and overdominance) at single locus level, and all three types of interactions (AA, AD/DA, and DD) at the two-locus level, were involved in the genetic basis of heterosis. The AA effects may partly be ascribed to the residual effects of the performance of the traits that were not completely removed by subtracting the mid-parent values from the F_1s in the immortalized F_2 population, as some of them were also detected in the trait analyses (15). However, the results clearly suggest that, at the population level, all kinds of genetic effects can be contributors to the genetic basis of heterosis. Thus, the effects of dominance, overdominance, and epistasis of various forms, are not mutually exclusive in the genetic basis of heterosis, as opposed to what was previously debated in favor of the different hypotheses (31); all of these components have a role to play depending on the genetic architecture of the population.

For the hybrid Shanyou 63, however, only genetic components associated with heterozygotes are relevant, given the conditions that all of the analyses were based on the loci that were polymorphic between the two parents, Minghui 63 and Zhenshan 97. Hence, only single-locus heterotic effects (caused by partial-, full-, and overdominance) and DD interactions were pertinent to the interpretation of heterosis in Shanyou 63. Assuming complete independence of the single-locus heterotic effects and digenic interactions involving DD detected in the analyses, which might be violated because of linkage, the genetic basis of heterosis for grains per panicle in the F_1 hybrid may tentatively be sketched as the following. Summing up the heterotic effects over the HL, as can be seen from Table 2, would produce a total of 20.19 grains per panicle; summation of the deviations of double heterozygotes from the means of the two parental genotypes over the two-locus combinations showing DD yielded a total of 34.73 grains per panicle (Table 6). Thus, together, the effects detected at the single-locus and two-locus levels could account for 54.92 (86%) of the 63.60 seeds that was measured as the amount of heterosis in the F_1 (Table 1). This seems to be an excellent approximation considering that the analysis failed to reveal many of the single-locus heterotic effects and DD because of the statistical thresholds imposed, and that the heritability for this trait, although high, is far from unity. A similar trend was also present for this trait in 1999. These results strongly suggest that heterotic effects (caused by partial-, full-, and overdominance) at the single-locus level, in combination with advantageous effects of double heterozygotes (caused by DD) at two-locus level, can adequately account for the genetic basis of heterosis in Shanyou 63, the most widely cultivated hybrid rice in China.

In summary, it is concluded that all kinds of genetic effects, including single-locus heterotic effects (caused by partial-, full-, and overdominance) and all three forms of digenic interactions (AA, AD/DA, and DD), contributed to heterosis in the immortalized F_2 population. Heterotic effects at the single-locus level, in combination with the advantageous effects of double heterozygotes caused by DD, can adequately explain the genetic basis of heterosis in Shanyou 63. These results may also help reconcile the century-long debate concerning the roles of dominance, overdominance, and epistasis in the genetic basis of heterosis (3–5, 7, 8, 10–12).

We thank the Cornell Group and the Japanese Rice Genome Project for kindly providing the RFLP probes. This research was supported by grants from the National Natural Science Foundation of China, the National Program on the Development of Basic Research, and the Rockefeller Foundation.

1. Stuber, C. W. (1994) *Plant Breed. Rev.* **12,** 227–251.
2. Yuan, L. P. (1998) in *Advances in Hybrid Rice Technology,* eds. Virmani, S. S., Siddiq, E. A. & Muralidharan, K. (International Rice Research Institute, Los Baños, Philippines), pp. 27–33.
3. Davenport, C. B. (1908) *Science* **28,** 454–455.
4. East, E. M. (1908) *Report of the Connecticut Agricultural Experiment Station for Years 1907–1908*, 419–428.
5. Shull, G. H. (1908) *Am. Breed. Assoc.* **4,** 296–301.
6. Hallauer, A. R. & Miranda, J. B. (1981) *Quantitative Genetics in Maize Breeding* (Iowa State Univ. Press, Ames).
7. Stuber, C. W., Lincoln, S. W., Wolff, D. W., Helentjaris, T. & Lander, E. S. (1992) *Genetics* **132,** 823–839.
8. Xiao, J. H., Li, J. M., Yuan, L. P. & Tanksley, S. D. (1995) *Genetics* **140,** 745–754.
9. Wright, S. (1968) *Evolution and Genetics of Populations* (Univ. of Chicago Press, Chicago), Vol. 1.
10. Yu, S. B., Li, J. X., Xu, C. G., Tan, Y. F., Gao, Y. J., Zhang, Q. & Saghai Maroof, M. A. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 9226–9231.
11. Li, Z. K., Luo, L. J., Mei, H. W., Wang, D. L., Shu, Q. Y., Tabien, R., Zhong, D. B., Ying, C. S., Stansel, J. W., Khush, G. S., *et al.* (2001) *Genetics* **158,** 1737–1753.
12. Luo, L. J., Li, Z. K., Mei, H. W., Shu, Q. Y., Tabien, R., Zhong, D. B., Ying, C. S., Stansel, J. W., Khush, G. S. & Paterson, A. H. (2001) *Genetics* **158,** 1755–1771.
13. Allard, R. W. (1956) *Hilgardia* **24,** 235–278.
14. Edwards, M. D., Stuber C. W. & Wendel, J. F. (1987) *Genetics* **116,** 113–125.
15. Hua, J. P., Xing, Y. Z., Xu, C. G., Sun, X. L., Yu, S. B. & Zhang, Q. (2002) *Genetics* **162,** 1885–1895.
16. Comstock, R. E. & Robison, H. F. (1948) *Biometrics* **4,** 254–266.
17. Arumuganathan, K. & Earle, E. D. (1991) *Plant Mol. Biol. Rep.* **9,** 208–218.
18. Causse, M. A., Fulton, T. M., Cho, Y. G., Ahn, S. N., Chunwongse, J., Wu, K., Xiao, J., Yu, Z. H., Ronald, P. C., Harrington, S. E., *et al.* (1994) *Genetics* **138,** 1251–1274.
19. Harushima, Y., Yano, M., Shomura, A., Sato, M., Shimano, T., Kuboki, Y., Yamamoto, T., Lin, S. Y., Antonio, B. A., Parco, A., *et al.* (1998) *Genetics* **148,** 479–494.
20. Yu, J., Hu, S., Wang, J., Wong, G. K.-S., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., *et al.* (2002) *Science* **296,** 79–92.
21. Goff, S. A., Ricke, D., Lan, T.-H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., *et al.* (2002) *Science* **296,** 92–100.
22. Xing, Y. Z., Tan, Y. F., Hua, J. P., Sun, X. L., Xu, C. G. & Zhang, Q. (2002) *Theor. Appl. Genet.* **105,** 248–257.
23. Lincoln, S., Daly, M. & Lander, E. (1992) *Constructing Genetics Maps with* MAPMAKER/EXP 3.0 (Whitehead Institute, Cambridge, MA), Whitehead Institute Technical Report.
24. Churchill, G. A. & Doerge, R. W. (1994) *Genetics* **138,** 963–971.
25. Snedecor, G. W. & Cochram, W. G. (1980) *Statistical Methods* (Iowa State Univ. Press, Ames).
26. StatSoft (1997) STATISTICA (StatSoft, Tulsa, OK).
27. Zeng, Z. B. (1994) *Genetics* **136,** 1457–1468.
28. Steel, R. G. D. & Torrie, J. H. (1980) *Principles and Procedures of Statistics* (McGraw–Hill, New York), 2nd Ed.
29. Zhang, Q., Gao, Y. J., Yang, S., Ragab, R., Saghai Maroof, M. A. & Li, Z. B. (1994) *Theor. Appl. Genet.* **89,** 185–192.
30. Zhang, Q., Gao, Y. J., Saghai Maroof, M. A., Yang, S. H. & Li, J. X. (1995) *Mol. Breed.* **1,** 133–142.
31. Allard, R. W. (1960) *Principles of Plant Breeding* (Wiley, New York).