ORIGINAL PAPER

Genetic basis of grain yield heterosis in an "immortalized F_2" maize population

**Tingting Guo · Ning Yang · Hao Tong · Qingchun Pan ·
Xiaohong Yang · Jihua Tang · Jiankang Wang ·
Jiansheng Li · Jianbing Yan**

Abstract

Key message **Genetic basis of grain yield heterosis
relies on the cumulative effects of dominance, overdom-
inance, and epistasis in maize hybrid Yuyu22.**

Abstract Heterosis, i.e., when F_1 hybrid phenotypes are
superior to those of the parents, continues to play a criti-
cal role in boosting global grain yield. Notwithstanding
our limited insight into the genetic and molecular basis
of heterosis, it has been exploited extensively using dif-
ferent breeding approaches. In this study, we investigated
the genetic underpinnings of grain yield and its compo-
nents using "immortalized F_2" and recombinant inbred
line populations derived from the elite hybrid Yuyu22. A
high-density linkage map consisting of 3,184 bins was used
to assess (1) the additive and additive-by-additive effects
determined using recombinant inbred lines; (2) the domi-
nance and dominance-by-dominance effects from a mid-
parent heterosis dataset; and (3) the various genetic effects
in the "immortalized F_2" population. Compared with a
low-density simple sequence repeat map, the bin map iden-
tified more quantitative trait loci, with higher LOD scores
and better accuracy of detecting quantitative trait loci. The
bin map showed that, among all traits, dominance was
more important to heterosis than other genetic effects. The
importance of overdominance/pseudo-overdominance was
proportional to the amount of heterosis. In addition, epista-
sis contributed to heterosis as well. Phenotypic variances
explained by the QTLs detected were close to the broad-
sense heritabilities of the observed traits. Comparison of
the analyzed results obtained for the "immortalized F_2"
population with those for the mid-parent heterosis dataset
indicated identical genetic modes of action for mid-parent
heterosis and grain yield performance of the hybrid.

Communicated by Thomas Lubberstedt.

T. Guo · X. Yang · J. Li
National Maize Improvement Center, China Agricultural
University, Beijing 100193, China
e-mail: lijiansheng@cau.edu.cn

T. Guo · J. Wang
Institute of Crop Science, The National Key Facility for Crop
Gene Resources and Genetic Improvement, and CIMMYT
China, Chinese Academy of Agricultural Sciences, No. 12
Zhongguancun South Street, Beijing 100081, China

N. Yang · H. Tong · Q. Pan · J. Yan (✉)
National Key Laboratory of Crop Genetic Improvement,
Huazhong Agricultural University, Wuhan 430070, China
e-mail: yjianbing@mail.hzau.edu.cn

J. Tang
College of Agronomy, Henan Agricultural University,
Zhengzhou 450002, China

Introduction

Heterosis, the phenomenon in which F_1 hybrids exhibit
phenotypes superior to their parents (Shull 1908; East
1908), has been exploited extensively in many field crops
and animals to increase agricultural yields throughout the
world. Extensive research on the genetic basis of heterosis
has been conducted for more than a century, but the molec-
ular underpinnings of the phenomenon remain conjectural
(Falconer 1981; Stuber 1994; Duvick 1999).

Accumulated evidence suggests that heterosis can be
accounted for by partial to complete dominance (Davenport
1908; Jones 1917; Hallauer et al. 1988). Overdominance

was considered to be important in pioneering studies (Hull 1945; Crow 1948; Stuber et al. 1992). Nonetheless, several rounds of recombination and re-analysis of the data from quantitative trait loci (QTL) mapping populations indicated that pseudo-overdominance caused by repulsion-phase linkages is more likely to account for heterosis than true overdominance (Crow et al. 1952; Gardner and Lonquist 1959; Moll et al. 1964; Cockerham and Zeng 1996; Graham et al. 1997). Larièpe et al. (2012) conducted a QTL mapping study using three recombinant inbred populations and the North Carolina Design III approach (Comstock and Robinson 1948, 1952). The results showed that linked genes with small individual effects might often appear as a single major QTL, particularly in chromosomal regions with high gene density relative to recombination. Epistasis might also contribute to the expression of heterosis in elite crosses (Richey 1942; Williams 1959; Minvielle 1987). Although a large number of genetic studies have been performed and each usually favors one of several potential hypotheses for each individual analysis, no universal conclusion has emerged regarding whether heterosis is mediated by a single mechanism or multiple genetic mechanisms across species. It was recently suggested that a combination of different genetic principles might best explain the manifestation of heterosis (Swanson-Wagner et al. 2006; Lippman and Zamir 2007). The dominance and overdominance hypotheses are certainly not mutually exclusive, and there are likely to be additional explanations for heterosis (Schnable and Springer 2013).

The use of molecular markers to analyze the genetic basis of quantitative traits in many crops has made it possible to identify and characterize the loci responsible for heterosis (Melchinger et al. 2007; Larièpe et al. 2012; Goff and Zhang 2013). However, low-density markers do not allow for the dissection of the multiple linked genes that control complex traits. Large-scale analysis of markers that cover the entire genome and can account for almost all potential recombinant events in a population might improve our understanding of the modes of allele action as related to heterosis. The source of genetic material and experimental design directly affect the ease with which the molecular bases of heterotic effects can be identified. Primary populations (e.g., members of generation F_2 or F_3) have been used widely to study heterosis (Yan et al. 2006; Yu et al. 1997). However, an obvious limitation on their use is the lack of an unlimited supply of genetically identical seeds for replication of experiments.

Hua et al. (2002) introduced an experimental design developed from pair crosses of recombinant inbred lines (RILs), which they called an "immortalized F_2" (or IF_2) population. This population resembles an F_2 population in the sense that the compositions and frequencies of single and multi-locus genotypes are the same as those of an F_2 population, but an IF_2 can be regrown repeatedly. The use of rice IF_2 populations to determine the genetic basis of heterosis has been proposed (Hua et al. 2003; Zhou et al. 2012). Here, we used an IF_2 population of maize generated from the elite heterotic hybrid Yuyu22 (Zong3 × 87-1), which has been widely planted in China over the past 2 decades. The mean performance of Yuyu22 in the report of Tang et al. (2010) was 10.63 t/ha for grain yield, 21.08 cm for ear length, 16.51 for ear row number, and 35.46 g for 100-kernel weight.

Although several heterotic loci for grain yield in maize have been identified in the IF_2 population (Tang et al. 2010), many questions surrounding their effects remain unresolved owing to the low density of the markers used and the limited methods available for their analysis. To address these deficiencies, we re-genotyped the materials using the maize SNP50 array (Ganal et al. 2011) and constructed a high-density bin map with 3,184 bins. SNPs cosegregating in two contiguous block borders were lumped as a bin (Xie et al. 2010). Our results showed that a bin map can more completely characterize heterosis compared with a low-density simple sequence repeat (SSR) linkage map. The QTL mapping, which included an assessment of both additive/dominance and epistasis factors, was conducted using inclusive composite interval mapping (ICIM) (Li et al. 2007; Zhang et al. 2008) but not CIM and two-way analysis of variance as was done by Tang et al. (2010). Marker variables were considered in a linear model in ICIM for additive/dominance mapping, and both marker variables and marker-pair multiplications were simultaneously considered for epistasis mapping.

In this study, we analyzed the genetic effects of all the bins in the IF_2 population including additive, dominance and two-locus interactions, and assessed their relative contributions to heterosis in the F_1 hybrid. We also analyzed a simulated IF_2 population and mid-parent heterosis (MPH) dataset to explore modes of allele action for MPH and per se performance of the hybrid.

Materials and methods

Plant materials and SNP bin map

We used a set of 294 F_8 RILs derived using the single-seed descent method from Yuyu22. Similar to the procedure for generating the IF_2 population described previously (Hua et al. 2002, 2003; Tang et al. 2010), the 294 RILs were randomly divided into two groups of 147 RILs. Then, single crosses were randomly made between the two groups without replacement. This procedure was repeated three times to produce 441 (147 × 3) single crosses, forming the IF_2 population (Fig. 1). The RIL and IF_2 populations were planted in 2003 and 2004 in Beijing (north of China, average daily
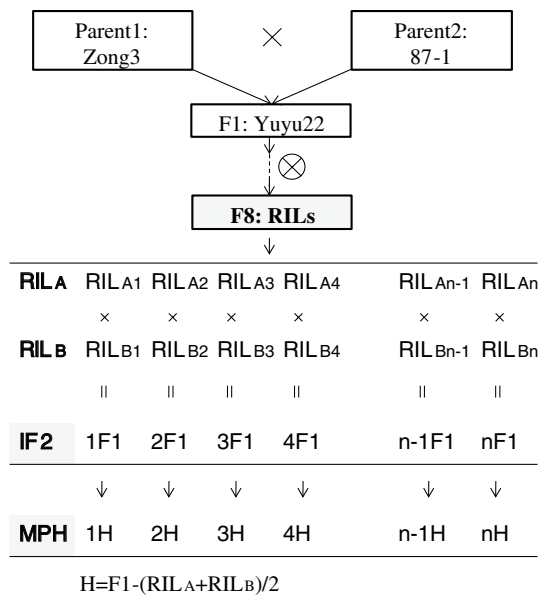
Fig. 1 Schematic representation of the development scheme for the RIL and IF$_2$ populations. The RILs were divided into two groups equally and randomly: RIL$_A$ and RIL$_B$. This procedure was repeated three times, and finally 3 ×n F$_1$ were produced, forming the IF$_2$ population

temperature of 11.8 °C, and average annual rainfall of 585 mm) and Xunxian, Henan Province (center of the North China Plain, average daily temperature 14.2 °C, and average annual rainfall 784 mm). At two locations, populations of RIL and IF$_2$ were in neighbored blocks, each planted in a randomized complete block design with three replications. Each plot included one row, 4 m long, with 0.67 m between rows. Population density was 45,000 plants per ha. After maturity, ten ears from consecutive plants in each plot were harvested by hand and air-dried until the grain moisture reached 13 %. The four traits evaluated were grain yield, ear length, 100-kernel weight, and row number. The details of the mating scheme, the genetic characteristics of the population, and the field experiments have been described previously (Tang et al. 2010; Guo et al. 2012).

All of the 294 RILs were genotyped with 261 polymorphic SSR markers covering the whole genome. The genotypes of 441 hybrids in the IF$_2$ population were deduced in terms of segments derived from each parent. A low-density linkage map was constructed using the SSR genotypes of the 294 RILs (Tang et al. 2010). In this study, a subset of 190 RILs was genotyped using a Maize SNP50 genotyping chip (Ganal et al. 2011). A bin map consisting of 3,184 bins was constructed with 18,840 polymorphic markers between Zong3 and 87-1, using an own developed algorithm. In addition, 175 RILs were genotyped with 1,536 SNPs (Yan et al. 2010). Of the two panels of RILs, 109 lines overlapped, from which the genotype datasets were derived from both

3,184 bins and 1,536 SNPs. Bins (3,184) from the parents (Zong3 and 87-1) were imputed onto the 175 RILs by first determining the physical position of the markers via BLAST to the genome B73 RefGen_v2. For each bin, its value for a RIL was assigned based on the SNP value of the RIL parents and on the genotype of the 1,536 SNPs in that RIL. Thus, the 109 overlapping lines had two panels of high-density genotypic information: from SNP genotyping chips and from imputation. The correlation coefficient between genotyped and imputed bins of 109 RILs was greater than 0.96, which underscored the accuracy of the imputed bin genotypes of the remaining 66 RILs. Consequently, the RIL population based on the bin map consisted of 256 individuals—190 genotyped and 66 imputed. Data for 312 crosses were obtained based on the 256 RILs from the dataset of 441 crosses. The bin genotypes of each cross in the IF$_2$ population were deduced from both of its RIL parents.

Data analysis and QTL mapping

Phenotypic data were analyzed for sources of variation using the following model:

$$y_{ijk} = \mu + e_i + r_j + g_k + ge_{k \times i} + \varepsilon_{ijk},$$

where y_{ijk} is the trait value of the ith environment, jth replication and kth entry; μ is the overall mean; e_i is the ith environment effect; r_j is the jth replication effect; g_k is the kth entry effect; $ge_{k \times i}$ is the interaction effect of the kth entry by the ith environment; and ε_{ijk} is the residual effect. All the variables except μ are considered as random effects. Broad-sense heritability (H^2) is equal to genetic variance (V_G) divided by summation of genetic variance (V_G), $G \times E$ variance ($V_{G \times E}$) and error variance (V_ε). The best linear unbiased predictors (BLUPs) for each line in RILs and each hybrid in IF$_2$ were calculated with a mixed linear model accounting for effects of environment, replication, genotype, and genotype by environment. The estimated genetic variance and error variance by analysis of variance were for BLUP estimators. The BLUPs for each line and hybrid from four locations were analyzed via QTL mapping. The MPH of each F$_1$ in the IF$_2$ population was calculated as follows: MPH = F$_1$ − MP, where F$_1$ is the BLUP genotypic value of the F$_1$, and MP is the mid-parental value of the corresponding parents. There was only dominance effect in the MPH dataset under the genetic model of additive and dominance effect, where the QTL effects were estimated by the differences between heterozygotes and the mean of the parental homozygotes.

QTL mapping was conducted separately with three different genotypic datasets (RIL, IF$_2$, and MPH) based on the low-density SSR linkage map and high-density bin map. The ICIM of QTLs was done using QTL IciMapping software (http://www.isbreeding.net). The largest P value for

entering variables in stepwise regression of phenotype on marker variables was 0.001 for additive/dominance QTLs and 0.0005 for epistatic QTLs. The largest *P*-value for removing variables was assumed to be twice the *P*-value for entering variables. The step size in the ICIM was 0.1 cM for additive/dominance QTL, which was lower than the average genetic distance between flanking markers (0.3 cM). For epistatic QTL, the step size was set to 2 cM. A LOD score threshold of 2.35 corresponds to an approximate nominal significance level of $P = 0.001$ per test with one degree of freedom. When QTL detection was for additive/dominance and epistasis, the LOD score thresholds of 3.0 and 4.0 were equivalent to a significance level of approximately 0.001, with degrees of freedom = 2 and 4 (Lander and Botstein 1989). Using the 1,000 permutation test, the LOD score threshold on the bin map was almost 0.5 higher than the value on the SSR map regardless of the population type. As there was not much difference in LOD score between the low- and high-density genetic maps, an empirical threshold of 2.5 was used to declare the existence of a QTL with an additive and/or dominance effect, and a threshold of 5.0 was used to declare the existence of an epistatic QTL. We also used single-marker analysis (SMA) to identify QTLs based on the difference between the mean phenotypic values of groups of individuals differing in marker genotypes.

Simulated phenotypes and genotypes of RIL and IF$_2$ populations

Phenotypes of individuals in these populations were simulated, considering that the trait of interest was controlled by 20 QTLs. The linkage map built from the actual maize RIL population was used, and two unlinked QTLs were distributed on each chromosome. Under the additive and dominance model in quantitative genetics, we generated 100 sets of random additive (*a*) and dominance (*d*) effects for the 20 QTLs, in which the genetic effects followed the uniform distribution from 0 to 1. According to the effects of QTLs, the ratio of the dominance variance to additive variance (V_D/V_A) was estimated as $\frac{1}{2}\sum_{m=1}^{20} d^2 \Big/ \sum_{m=1}^{20} a^2$, equivalent to a random-mating population in which there is no linkage equilibrium and the allele frequency is 0.5. Among the 100 sets of QTL effects, four were chosen to represent four types of genetic architectures whose respective V_D/V_A values were closest to 0.5, 1, 2, and 3. For each level of variance ratio, 294 F$_8$ RILs were simulated using the genetics and breeding simulation tool of QuLine, formerly called QuCim (Wang et al. 2003, 2004). An IF$_2$ population consisting of 441 hybrids was generated from the 294 RILs. The simulation was repeated 40 times, with 10 sets of RIL and IF$_2$ populations constructed for each type of genetic architecture. Similar to the actual populations, MPH data

were acquired from the RIL and IF$_2$ populations. The procedure of QTL mapping in the simulation was the same as in the real dataset.

Results

Construction of a high-density bin map

The 18,840 polymorphic markers were mapped and scored into 3,184 bins, accounting for total genetic and physical distances of 2,657.9 cM and 2,046.3 Mb, respectively. The average genetic distance between flanking bins was 0.84 cM, and the average physical size of a bin was 0.64 Mb. The genotypes of 312 F$_1$ hybrids were deduced from their corresponding RIL parents to provide a bin map for the crosses (Figure S1). There were three genotypes in each bin: homozygous genotype from Zong3 (*MM*), homozygous genotype from 87-1 (*mm*), and heterozygous genotype (*Mm*). For each cross, the average proportion of *MM*, *mm*, and *Mm* was 23.6, 26.8, and 49.5 %, respectively. Therefore, the composition of genotypes in IF$_2$ was similar to that in an F$_2$ population. This population could therefore be used to detect QTLs with the same analytical method used for the F$_2$ population.

Identification of major QTLs in an IF$_2$ population

Based on the bin map, we dissected the genetic underpinnings of grain yield and yield components within the IF$_2$ population. The bin locations and genetic effects of significant QTLs were detected and estimated using ICIM (Table S1). The degree of dominance was defined for each QTL as the ratio of the dominance effect to the additive effect (*d/a*). If the value of |*d/a*| was greater than 1.26, the QTL was considered as the overdominance type; otherwise, the QTL was considered as the dominance type (Falconer and Mackay 1996). A total of 11 QTLs that affected grain yield were mapped at a given value of LOD score, and each accounted for an average phenotypic variance of 4.9 %. Six QTLs with |*d/a*| ≤ 1.26 exhibited dominance, whereas the remaining five QTLs exhibited overdominance (Table S1). Sixteen QTLs were identified for ear length, five of which QTLs had |*d/a*| > 1.26, displaying overdominance, the others were of the dominance type. The 21 QTLs controlling 100-kernel weight included 16 dominant QTLs and 5 overdominant QTLs. All of them were distributed across the entire genome except chromosome 4, and each QTL accounted for an average phenotypic variance of 3.8 %. Only 2 of the 15 QTLs for row number were of the overdominance type. Across the four traits, the number of dominant QTL exceeded the number of overdominant QTL, especially for yield component traits.

Fig. 2 Phenotypic variance explained by the QTL effects based on the low- and high-density linkage maps in the RIL, IF_2, and MPH datasets for yield and three yield component traits. *PVE* phenotypic variance explained, *Epi* epistasis
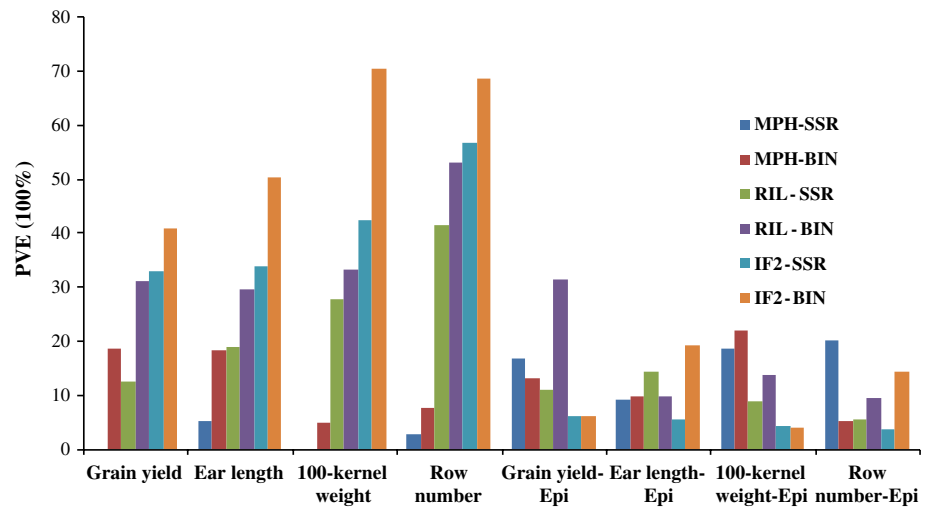
A bin map can identify more QTLs with higher resolution than a SSR linkage map

There were many differences between the SSR and bin maps, especially with respect to the relative saturation of recombination events by markers (Table S2). The bin map consisted of 3,184 markers across the whole genome, and the number of markers per chromosome ranged from 221 to 522, whereas the SSR map was composed of 261 markers across the genome, and the number of markers per chromosome ranged from 11 (chromosome 9) to 36 (chromosome 1). The maximum genetic distance between flanking markers was 38.1 cM on the SSR linkage map—much larger than 7.9 cM on the bin map. Particularly, the average genetic distance between flanking markers on the bin map was 0.84 cM—much shorter than 9.41 cM on the SSR map. These data indicated that the genome wide marker density was higher for the bin map than for the SSR map.

The high-density bin map represented almost all crossovers in the bi-parental population studied, and this was expected to improve the power and resolution of QTL detection. With ICIM, a high-density bin map was shown to have many advantages for QTL mapping compared with a low-density SSR linkage map (Table S3 and Figures S2). First, the high-density bin map could detect a larger number of significant loci: for all traits, 49 QTLs were identified using the low-density map compared with 63 QTLs identified using the high-density map in the IF_2 population. Second, the QTLs already detected using the low-density SSR map had higher LOD scores when the high-density bin map was used. Taking grain yield in the RIL population as an example, LOD scores of the QTL on chromosomes 5 and 7 were 4.0 and 5.8, respectively, for the low-density SSR map, but increased to 7.2 and 9.4 at approximately the same locations on the high-density bin map (Figure S2). Third, the high-density bin map improved the resolution

of QTL mapping from an average of 10.82 cM to an average of 0.88 cM (Table S3), which allowed us to investigate the candidate genes present in the QTL between flanking markers. Fourth, the total phenotypic variation explained (PVE) by the QTLs detected using the high-density bin map was larger than the value using the low-density SSR map (Figure 2). When using the additive and dominance genetic model, the PVE for grain yield using the low-density SSR map was 12.6 and 32.9 for the RILs and the IF_2 population, respectively, whereas these values increased to 31.0 and 40.8 using the high-density map.

Cumulative effects of dominance account for the majority of heterosis

To determine the primary mode by which heterosis occurs, we tested additive and dominance effects for each bin using SMA on the whole genome. By calculating the degree of dominance (*d/a*), we categorized the bins into two types—dominance and overdominance—each of which included negative and positive values. The ratio *d/a* was calculated in two ways: both additive and dominance genetic effects (*a* and *d*) were estimated from the IF_2 population; additive effect (*a*) was estimated from RILs and dominance effect (*d*) from MPH dataset. For both the IF_2 and the RIL/MPH datasets, the number of bins showing dominance was larger than the number showing overdominance for all traits (Fig. 3). Further, stronger trait heterosis was associated with more bins showing overdominance and fewer bins showing dominance. Grain yield exhibited the strongest heterosis, followed by ear length, 100-kernel weight, and row number (Table S4). The number of "dominance" bins in both the IF_2 and RIL/MPH datasets accounted for 58 and 55 % for grain yield, 59 and 70 % for ear length, 68 and 85 % for 100-kernel weight, 75 and 85 % for row number, respectively (Fig. 3). Positive and negative values for *d/a*
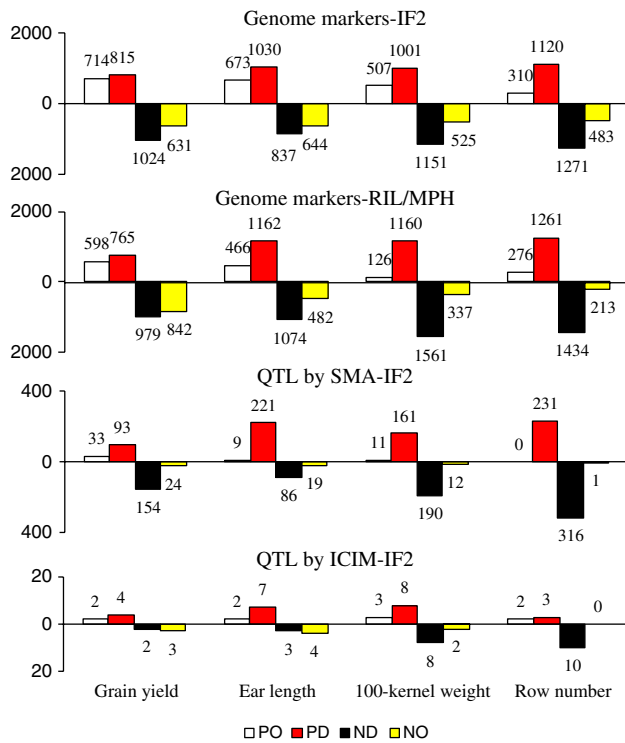
Fig. 3 Number of bins showing dominance and overdominance in the whole genome and number of significant loci for grain yield and three component traits. *SMA* single-marker analysis, *ICIM* inclusive composite interval mapping, *PO* positive overdominance, *PD* positive dominance, *ND* negative dominance type, *NO* negative overdominance. Both the additive and dominance effects were estimated from the IF$_2$ population; for the RIL population, only the additive effects were estimated, and for the MPH dataset only the dominance effects were estimated

were randomly distributed throughout the genome for all traits.

Considering that it was unlikely that most bins would associate with yield and its components, we excluded bins that did not exceed the given LOD score threshold. The SMA results revealed that most of the bins that were significantly associated with traits were of the dominance type for all traits: 81 % for grain yield and >90 % for yield component traits (Fig. 3). The results obtained by ICIM were very similar to those obtained by SMA. In conclusion, both ICIM and SMA demonstrated that QTLs showing dominance constituted the primary mode by which heterosis occurs; however, overdominance also contributed, especially for grain yield.

Epistasis also contributes to grain yield heterosis

Decades of research on epistasis have suggested the importance of multi-gene interactions in controlling heterosis for grain yield in plants (Yu et al. 1997; Li et al. 1997a, b). Here we observed that the major QTLs did not explain

all the phenotypic variance, which led us to assess the contribution of epistasis in our dataset. The IF$_2$ population is ideal for identifying the genetic components of heterosis, as it permits detailed dissection of additive genetic effects (*a*), dominant genetic effects (*d*), and the four types of two-locus interactions (*AA*, *AD*, *DA*, *DD*). A total of 101 two-locus interactions for the four traits were identified based on the bin map (Figure S3). The interactions could be partitioned into four types: *AA*-oriented, *AD*-oriented, *DA*-oriented, and *DD*-oriented according to the interactive effect. Within the IF$_2$ population, the *DD*-oriented QTL had the highest frequency.

We also calculated the proportion of PVE by analyzing of epistatic QTLs (Fig. 2). Compared with PVE by single-locus QTLs, epistatic QTLs explained less of the phenotypic variance for all traits except for grain yield in the RIL population and the 100-kernel weight in the MPH dataset (Fig. 2). In the IF$_2$ population, epistasis explained 28 % of the phenotypic variance for grain yield, followed by 19 % for ear length, 14 % for row number, and 4 % for 100-kernel weight. The PVE values determined by single-locus QTLs were 41 % for grain yield, 50 % for ear length, 71 % for 100-kernel weight, and 69 % for row number. Given that PVE values based on single-locus QTLs and epistasis together were close to the heritability of each trait, the bin map permitted us to account for almost all phenotypic variation for the four traits.

Identical modes of gene action for MPH and per se performance of the hybrid

Heterotic effects have been determined based on MPH, and several QTLs that control MPH (heterotic loci) have been identified both in rice and maize (Hua et al. 2003; Tang et al. 2010). Given that most of the QTLs controlling grain yield performance and MPH do not overlap, it appears that these two phenotypes are controlled by different genetic factors (Hua et al. 2003; Tang et al. 2010). Regardless of whether heterotic loci were independent or whether they were part of loci for per se performance of the hybrid, these alternatives are difficult to distinguish in actual populations. Estimates of the additive effect of the same QTL differed between the RIL and IF$_2$ populations, and estimates of the dominance effect of the same QTL differed between the IF$_2$ population and the MPH dataset, which could be caused by differences in the complexity of statistical models, or in different population size, among other possible factors. It is also difficult to determine which estimation method is more accurate using actual populations. However, all of the issues can be addressed in simulation studies provided that the true QTLs are known.

In our simulation studies, more true-positive QTLs were detected in the IF$_2$ population, followed by the

Table 1 QTL identification among various genetic compositions in RILs, IF$_2$ populations, and the MPH dataset in simulation studies

Genetic architecture	No. of QTL with additive effect				No. of QTL with dominance effect			
$V_D/V_A = 0.5$	RIL	14.1 ± 2.23	IF$_2$	16.7 ± 1.49	MPH	12.9 ± 1.45	IF$_2$	16.7 ± 1.49
	RIL-L [a]	2.7 ± 1.4	IF$_2$-L	5.9 ± 1.66	MPH-L	12.3 ± 1.34	IF$_2$-L	7.9 ± 2.18
	RIL-E [b]	10.4 ± 1.96	IF$_2$-E	8.3 ± 1.83	MPH-E	0.8 ± 0.79	IF$_2$-E	7.8 ± 2.74
	RIL-H [c]	1 ± 1.05	IF$_2$-H	2.5 ± 1.65	MPH-H	0 ± 0	IF$_2$-H	1 ± 1.33
	RIL-F [d]	0.6 ± 0.84	IF$_2$-F	3.1 ± 1.79	MPH-F	0.5 ± 0.71	IF$_2$-F	3.1 ± 1.79
$V_D/V_A = 1$	RIL	14.1 ± 2.23	IF$_2$	16.6 ± 1.52	MPH	13.1 ± 2.28	IF$_2$	16.6 ± 1.52
	RIL-L	2.7 ± 1.42	IF$_2$-L	6.3 ± 1.83	MPH-L	12.5 ± 2.32	IF$_2$-L	9.3 ± 1.83
	RIL-E	10.4 ± 1.96	IF$_2$-E	7.6 ± 2.55	MPH-E	0.6 ± 0.7	IF$_2$-E	5.9 ± 1.52
	RIL-H	1 ± 1.05	IF$_2$-H	2.8 ± 1.48	MPH-H	0 ± 0	IF$_2$-H	1.5 ± 1.08
	RIL-F	0.6 ± 0.84	IF$_2$-F	2.2 ± 1.4	MPH-F	0.5 ± 0.53	IF$_2$-F	2.2 ± 1.4
$V_D/V_A = 2$	RIL	14.1 ± 2.23	IF$_2$	16.1 ± 1.45	MPH	12.6 ± 1.71	IF$_2$	16.1 ± 1.45
	RIL-L	2.7 ± 1.4	IF$_2$-L	6.4 ± 1.26	MPH-L	12.1 ± 1.79	IF$_2$-L	8.5 ± 0.97
	RIL-E	10.4 ± 1.96	IF$_2$-E	6.5 ± 2.17	MPH-E	0.5 ± 0.71	IF$_2$-E	4.3 ± 1.57
	RIL-H	1 ± 1.05	IF$_2$-H	3.2 ± 1.87	MPH-H	0 ± 0	IF$_2$-H	3.3 ± 1.42
	RIL-F	0.6 ± 0.84	IF$_2$-F	2.2 ± 1.48	MPH-F	0.3 ± 0.67	IF$_2$-F	2.2 ± 1.48
$V_D/V_A = 3$	RIL	14.1 ± 2.23	IF$_2$	15.7 ± 1.7	MPH	12 ± 2.11	IF$_2$	15.7 ± 1.7
	RIL-L	2.7 ± 1.4	IF$_2$-L	6.2 ± 1.55	MPH-L	11.5 ± 2.17	IF$_2$-L	8.4 ± 1.9
	RIL-E	10.4 ± 1.96	IF$_2$-E	6.6 ± 2.41	MPH-E	0.1 ± 0.32	IF$_2$-E	4.2 ± 1.75
	RIL-H	1 ± 1.05	IF$_2$-H	2.9 ± 1.97	MPH-H	0.4 ± 1.26	IF$_2$-H	3.1 ± 2.02
	RIL-F	0.6 ± 0.84	IF$_2$-F	2.4 ± 1.26	MPH-F	0.5 ± 0.53	IF$_2$-F	2.4 ± 1.26

[a] *L* QTL effect was lower than the true value, [b] *E* QTL effect was equal to the true value with 1 error range allowed, [c] *H* QTL effect was the higher than true value, [d] *F* false-positive QTL

RIL, and finally the MPH dataset (Table 1). However, false-positive QTLs were also detected more frequently in the IF$_2$ population than in the RIL and MPH dataset regardless of the genetic architecture. Taking the genetic architecture $V_D/V_A = 0.5$ as an example, an average of 16.7 true-positives among 20 QTL was detected in the IF$_2$ population, followed by an average of 14.1 in the RILs and 12.9 in the MPH dataset. The average number of false-positive QTL was 3.1 in the IF$_2$ population compared with values of 0.6 and 0.5 for the RILs and the MPH dataset, respectively. The same situation was also observed for other genetic architectures, i.e., $V_D/V_A = 1$, 2, or 3. For the comparison between MPH and IF$_2$, almost all the QTLs in the MPH dataset was also detected in the IF$_2$ population, and the majority of QTLs in the MPH dataset displayed overdominance including mild ($1.75 > d/a > 1.26$), moderate ($2.25 > d/a > 1.76$) and extreme ($d/a > 2.26$) (Fig. 4). For the genetic architecture $V_D/V_A = 0.5$, an average of 11.9 of the 12.9 true-positive QTL in the MPH dataset was also detected in the IF$_2$ population, in which 7.3 of 11.9 QTL exhibited overdominance. When $V_D/V_A = 1$, 2, or 3, a range of 93.1 to 96.7 % of true-positive QTL in the MPH dataset were also detected in the IF$_2$ populations, in which 68.9–90.5 % of the QTL were of the overdominance type. Our simulation results also demonstrated that estimating the
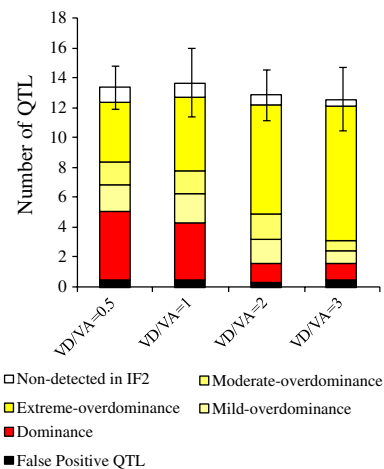

Fig. 4 QTL identified in the MPH dataset showing dominance as well as mild, moderate, and extreme overdominance in the IF$_2$ population among varying genetic architectures

additive effects was more accurate for the RIL population than for the IF$_2$ population, whereas the dominance effects were underestimated in the IF$_2$ population and substantially less abundant in the MPH dataset than in the IF$_2$ population (results not shown). In summary, the simulation analysis indicated identical modes of allele action for MPH and per se performance of the hybrid.

Discussion

The bin map and IF$_2$ population provide an effective basis for complete genetic characterization of heterosis

We used a high-density bin map and an IF$_2$ population to investigate the genetic basis of heterotic enhancement of grain yield. The usefulness of IF$_2$ populations of rice, maize, oilseed rape, and cotton has been discussed (Hua et al. 2002, 2003; Chen et al. 2007; Tang et al. 2010; Liu et al. 2011). The distinct advantages of high-density bin maps have been well documented (Xie et al. 2010; Pan et al. 2012; Zhou et al. 2012). In maize, the genetic components of heterotic performance have not been fully characterized using the low-density SSR map employed previously (Tang et al. 2010). This SSR map includes 261 markers, and the average genetic distance between flanking markers is 9.41 cM, which might leave out quite a number of recombination events. In contrast, the high-density bin map we constructed contains 3,184 bins, each bin consisting of 6 SNPs on average. The average genetic distance between flanking markers was 0.84 cM, which permitted identification of 14 more QTLs than the SSR map in the IF$_2$ population for traits combined. Compared with each QTL on the bin map relative to the QTLs on the SSR map, 23 QTLs were detected on both maps within the allowed range of error (Table S1). From the viewpoint of the degree of dominance, only 6 of 23 QTLs performed differently, displaying overdominance on the SSR map and dominance on the bin map, or vice versa. Consequently, the highly saturated map allowed more QTL to be detected and improved the resolution of QTL mapping, but it did not impact the relative assessment of genetic modes of action from differential detection. Through identification of 13 heterotic loci, Tang et al. (2010) confirmed the importance of dominance effects with respect to the impact of heterosis on grain yield. In our study, the use of an IF$_2$ population and a high-density bin map allowed us to demonstrate that the relative contributions of the various genetic effects to heterosis for grain yield and yield components were trait dependent.

The mechanism responsible for heterosis seems be consistent among maize and rice

Our findings are consistent with those of Zhou et al. (2012), who concluded that the cumulative effects of various genetic effects, including dominance, overdominance/pseudo-overdominance, and epistasis may adequately explain the genetic basis of heterosis. Further, partial dominance was the primary contributor to heterosis for almost all traits, and the contribution of overdominance was proportional to the level of MPH (Fig. 3). Molecular analyses have revealed that the percentage of expressed genes that exhibit partial dominance substantially exceeds the number

that exhibits overdominance (Swanson-Wagner et al. 2006; Stupar et al. 2008; Paschold et al. 2012). Notwithstanding examples of heterotic yield QTLs caused by pseudo-overdominance (Coors and Pandey 1999), no known true overdominance QTLs based on single-gene effects have been reported for rice or maize (Lippman and Zamir 2007; Xing and Zhang 2010). Therefore, we propose that the mechanism responsible for heterosis is consistent among different crop species—including rice and maize, which are typical self-pollinating and out-crossing species, respectively. That is, although various genetic effects contribute to heterosis, dominance (including overdominance) is the primary factor. Taken together, our heterosis results and those of others suggest that, although the genetic basis of heterosis is known, the molecular basis remains elusive but certainly involves multiple genes that differ among crosses with respect to the relative contribution of dominance, overdominance, and epistatic effects. Much work needs to be done to describe the diverse molecular mechanisms that contribute to overall hybrid performance.

Heterotic loci are not independent from loci controlling per se performance of the hybrid in the IF$_2$ population

Heterotic genes responsible for increasing crop yields are now being sought using genomics, particularly transcriptomics (Lippman and Zamir, 2007). Gene expression profiling studies have suggested that specific genes are involved in heterosis for a number of traits in maize, rice, and tomato (Hoecker et al. 2008; Wei et al. 2009; Guo et al. 2010; Krieger et al. 2010). However, whether such genes actually contribute to the general molecular mechanisms underlying the formation of hybrid vigor remains to be determined (Paschold et al. 2012). Previous studies have shown that heterotic loci are independent of the QTL that control yield-related traits (Hua et al. 2003; Tang et al. 2010). In our present study, we found that heterotic loci were not independent and that they comprised a subset of loci controlling per se performance of the hybrid. Our results show that three of five detected QTL for both grain yield and ear length in the MPH dataset overlapped with QTL in the IF$_2$ population, and all of them displayed overdominance in the IF$_2$ population (Table S5). For 100-kernel weight and row number, only one and two QTL, respectively, were detected in the MPH dataset that were also observed in the IF$_2$ population that exhibited overdominance (Table S5). In our simulated analysis, the relationship between heterotic loci and per se performance loci was less confusing. As shown in Table 1, the frequency of QTL detected in the MPH dataset (range 92.2–96.7 %) was observed in the IF$_2$ population, in which 61.3–90.5 % of QTL exhibited overdominance. The results from both of actual and simulated IF$_2$ populations provide evidence for our conclusion—identical genetic

modes of action for MPH and per se performance of the hybrid. In fact, the genetic basis of grain yield and heterosis should not be discussed separately, i.e., it is difficult to elucidate the genetic mechanism underlying grain yield without referring to heterosis, and vice versa.

A high-density bin map is very useful for positional cloning of genes

The availability of high-density markers throughout the genome has revolutionized our ability to dissect the genetic bases of complex traits. High-density markers enable positional cloning of genes underlying QTLs (Krieger et al. 2010). QTL mapping based on our bin map lowered the average genetic distance of QTL between flanking markers and narrowed this distance by more than 12-fold to less than 1 cM within the IF_2 population. This relatively small distance generally allows the identification of a limited number of candidate genes. There were 39 candidate genes annotated in the 30 QTLs for grain yield and its components (Table S6). Most of these genes were associated with photosynthesis, cell growth, differentiation/development, and the accumulation of sucrose, starch, and cellulose. Five linked genes on chromosome 4 provide candidates for further analysis; these include genes encoding glutathione S-transferase (*GST26*), small auxin-up RNA (*SAUR31*), ferredoxin-1, sugary-1 isoamylase (*SU1*), and C2H2 zinc-finger protein fragment (McClure et al. 1989; Hase et al. 1991; Marrs 1996; Rahman et al. 1998; Takatsuji 1998) (Figure S4). Not only the five genes located in a significant single bin (40.1–41.4 Mb on chromosome 4), but also two genes encoding calmodulin binding protein and chloroplastic quinine-oxidoreductase in a single bin (136.1–137.2 Mb on chromosome 7), might cause pseudo-overdominance, which we could not distinguish from overdominance in this study. Cloning the genes responsible for quantitative traits will resolve the issue of whether these traits are controlled by pseudo-overdominance or overdominance and contribute further to our understanding of the molecular basis of heterosis.

References

Chen W, Zhang Y, Liu X, Chen B, Tu J, Fu T (2007) Detection of QTL for six yield-related traits in oilseed rape (*Brassica napus*) using DH and immortalized F2 populations. Theor Appl Genet 115:849–858

Cockerham CC, Zeng ZB (1996) Design III with marker loci. Genetics 143:1437–1456

Comstock RE, Robinson HF (1948) The components of genetic variance in populations. Biometrics 4:254–266

Comstock RE, Robinson HF (1952) Estimation of average dominance of genes. In: Gowen JW (ed) Heterosis. ISU Press, Ames, pp 494–516

Coors JG, Pandey S (1999) Genetics and exploitation of heterosis in crops. ASA, CSSA, SSSA, Madison

Crow JF (1948) Alternative hypotheses of hybrid vigor. Genetics 33:477

Crow JF, Coors JG, Pandey S (1952) Heterosis. In: Gowen JW (ed) Dominance and overdominance. Iowa State College Press, Ames, pp 282–297

Davenport CB (1908) Degeneration, albinism and inbreeding. Science 28:454–455

Duvick DN (1999) The genetics and exploitation of heterosis in crops. In: Coors JG, Pandey S (eds) Crop Science Society of America, Madison, pp 19–29

East E (1908) Inbreeding in corn in Reports of the Connecticut Agricultural Experiment Station for Years 1907–1908. Connecticut Agricultural Experiment Station, New Haven, pp 419–428

Falconer DS (1981) Introduction to quantitative genetics, 2nd edn. Longman, London

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Longman, Essex

Ganal M, Durstewitz G, Polley A, Berard A, Buckler ES, Charcosset A, Clarke JD, Graner E, Hansen M, Joets J et al (2011) A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping and genetic mapping to compare with the B73 reference genome. PLoS ONE 6:e28334

Gardner CO, Lonquist JH (1959) Linkage and the degree of dominance of genes controlling quantitative characters in maize. Agron J 51:524–528

Goff SA, Zhang Q (2013) Heterosis in elite hybrid rice: speculation on the genetic and biochemical mechanisms. Curr Opin Plant Biol 16:221–227

Graham GI, Wolff DW, Stuber CW (1997) Characterization of a yield quantitative trait locus on chromosome five of maize by fine mapping. Crop Sci 37:1601–1610

Guo M, Rupe MA, Dieter JA, Zou JJ, Spielbauer D, Duncan KE, Howard RJ, Hou ZL, Simmons CR (2010) Cell Number Regulator 1 affects plant and organ size in maize: implications for crop yield enhancement and heterosis. Plant Cell 22:1057–1073

Guo T, Li H, Yan J, Tang J, Li J, Zhang Z, Zhang L, Wang J (2012) Performance prediction of F_1 hybrids between recombinant inbred lines derived from two elite maize inbred lines. Theor Appl Genet 126:189–201

Hallauer AR, Russell WA, Lamkey KR (1988) Corn breeding. In: Sprague GF, Dudley JW (eds) Corn and corn improvement, 3rd edn. ASA, CSSA, and SSSA, Madison, pp 469–564

Hase T, Kimata Y, Matsumura T, Sakakibara H (1991) Molecular cloning and differential expression of the maize ferredoxin family. Plant Physiol 96:77–83

Hoecker N, Keller B, Muthreich N, Chollet D, Descombes P, Piepho HP, Hochholdinger F (2008) Comparison of maize (*Zea mays* L.)

F₁-hybrid and parental inbred line primary root transcriptomes suggests organspecific patterns of nonadditive gene expression and conserved expression trends. Genetics 179:1275–1283

Hua J, Xing Y, Xu C, Sun X, Yu S, Zhang Q (2002) Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantageous for performance. Genetics 162:1885–1895

Hua J, Xing Y, Wu W, Xu C, Sun X et al (2003) Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. Proc Natl Acad Sci USA 100:2574–2579

Hull FH (1945) Recurrent selection for specific combining ability in corn. J Am Soc Agron 37:134–145

Jones DF (1917) Dominance of linked factors as a means of accounting for heterosis. Genetics 2:466–479

Krieger U, Lippman ZB, Zamir D (2010) The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. Nat Genet 42:459–463

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative trait loci using RFLP linkage maps. Genetics 121:185–199

Larièpe A, Mangin B, Jasson S, Combes V, Dumas F et al (2012) The genetic basis of heterosis: multiparental quantitative trait loci mapping reveals contrasted levels of apparent overdominance among traits of agronomical interest in maize (*Zea mays* L.). Genetics 190:795–811

Li Z, Pinson SRM, Paterson AH, Park WD, Stansel JW (1997a) Epistasis for three grain yield components in rice (*Oryza sativa* L.). Genetics 145:453–465

Li Z, Pinson SRM, Paterson AH, Park WD, Stansel JW (1997b) Genetics of hybrid sterility and hybrid breakdown in an inter-subspecific rice (*Oryza sativa* L.) population. Genetics 145:1139–1148

Li H, Ye G, Wang J (2007) A modified algorithm for the improvement of composite interval mapping. Genetics 175:361–374

Lippman ZB, Zamir D (2007) Heterosis: revisiting the magic. Trends Genet 23:60–66

Liu R, Wang B, Guo W, Wang L, Zhang T (2011) Differential gene expression and associated QTL mapping for cotton yield based on a cDNA-AFLP transcriptome map in an immortalized F2. Theor Appl Genet 123:439–454

Marrs KA (1996) The functions and regulation of glutathione S-transferases in plants. Annu Rev Plant Physiol Plant Mol Biol 47:127–158

McClure BA, Hagen G, Brown CS, Gee MA, Guilfoyle TJ (1989) Transcription, organization, and sequence of an auxin regulated gene cluster in soybean. Plant Cell 1:229–239

Melchinger AE, Utz HF, Piepho HP, Zeng Z, Schön CC (2007) The role of epistasis in the manifestation of heterosis: a systems-oriented approach. Genetics 177:1815–1825

Minvielle F (1987) Dominance is not necessary for heterosis: a two-locus model. Genet Res 49:245–247

Moll RH, Lindsey MF, Robinson HF (1964) Estimates of genetic variances and level of dominance in maize. Genetics 49:411–423

Pan Q, Ali F, Yang X, Li J, Yan J (2012) Exploring the genetic characteristics of two recombinant inbred line populations via high-density SNP markers in maize. PLoS ONE 7(12):e52777

Paschold A, Jia Y, Marcon C, Lund S, Larson NB et al (2012) Complementation contributes to transcriptome complexity in maize (*Zea mays* L.) hybrids relative to their inbred parents. Genome Res 22:2445–2454

Rahman A, Wong KS, Jane JL, Myers AM, James MG (1998) Characterization of SU1 isoamylase, a determinant of storage starch structure in maize. Plant Physiol 117:425–435

Richey FD (1942) Mock-dominance and hybrid vigor. Science 96:280

Schnable PS, Springer NM (2013) Progress toward understanding heterosis in crop plants. Ann Rev Plant Biol 64:71–88

Shull GH (1908) The composition of a field of maize. Am Breed Assoc Rep 4:296–301

Stuber CW (1994) Heterosis in plant breeding. Plant Breed Rev 12:227–251

Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander ES (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132:823–839

Stupar RM, Gardiner JM, Oldre AG, Haun WJ, Chandler VL, Springer NM (2008) Gene expression analysis in maize hybrids and hybrids with varying levels of heterosis. BMC Plant Biol 8:33

Swanson-Wagner RA, Jia Y, Decook R, Borsuk LA, Nettleton D et al (2006) All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. Proc Natl Acad Sci USA 103:6805–6810

Takatsuji H (1998) Zinc–finger transcription factors in plants. Cell Mol Life Sci 54:582–596

Tang J, Yan J, Ma X, Teng W, Wu W et al (2010) Dissection of the genetic basis of heterosis in an elite maize hybrid by QTL mapping in an immortalized F2 population. Theor Appl Genet 120:333–340

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

Wang J, van Ginkel M, Trethowan R, Ye G, Delacy I, Podlich D, Cooper M (2004) Simulating the effects of dominance and epistasis on selection response in the CIMMYT Wheat Breeding Program using QuCim. Crop Sci 44:2006–2018

Wei G, Tao Y, Liu GZ, Chen C, Luo RY, Xia HA, Gan Q, Zeng HP, Lu ZK, Han YN et al (2009) A transcriptomic analysis of super hybrid rice LYP9 and its parents. Proc Natl Acad Sci USA 106:7695–7701

Williams W (1959) Heterosis and the genetics of complex characters. Nature 184:527–530

Xie W, Feng Q, Yu H, Huang X, Zhao Q et al (2010) Parent-independent genotyping for constructing an ultra high-density linkage map based on population sequencing. Proc Natl Acad Sci USA 107:10578–10583

Xing Y, Zhang Q (2010) Genetic and molecular bases of rice yield. Annu Rev Plant Biol 61:421–442

Yan J, Tang H, Huang Y, Zheng Y, Li J (2006) QTL mapping and epistatic analysis for yield and yield components using molecular markers with an elite maize hybrid. Euphytica 149:121–131

Yan J, Yang X, Shah T, Sanchez-Villeda H, Li J, Warburton M, Zhou Y, Crouch JH, Xu Y (2010) High-throughput SNP genotyping with the GoldenGate assay in maize. Mol Breed 25:441–451

Yu SB, Li JX, Xu CG, Tan YF, Gao YJ et al (1997) Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. Proc Natl Acad Sci USA 94:9226–9231

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

Zhou G, Chen Y, Yao W, Zhang C, Xie W, Hua J, Xing Y, Xiao J, Zhang Q (2012) Genetic composition of yield heterosis in an elite rice hybrid Proc. Proc Natl Acad Sci USA 109:15847–15852