Genetic basis of grain yield heterosis in an “immortalized F\textsubscript{2}” maize population

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

Received: 7 November 2013 / Accepted: 19 July 2014 © Springer-Verlag Berlin Heidelberg 2014

Abstract

Heterosis, i.e., when F\textsubscript{1} hybrid phenotypes are superior to those of the parents, continues to play a critical role in boosting global grain yield. Notwithstanding our limited insight into the genetic and molecular basis of heterosis, it has been exploited extensively using different breeding approaches. In this study, we investigated the genetic underpinnings of grain yield and its components using “immortalized F\textsubscript{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 dominance and dominance-by-dominance effects from a mid-parent heterosis dataset; and (3) the various genetic effects in the “immortalized F\textsubscript{2}” population. Compared with a low-density simple sequence repeat map, the bin map identified 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, epistasis 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\textsubscript{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.

Introduction

Heterosis, the phenomenon in which F\textsubscript{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 molecular 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). Larrièpe 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 (Snähble 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; Larrièpe 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 F2 or F3) 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 F2” (or IF2) population. This population resembles an F2 population in the sense that the compositions and frequencies of single and multi-locus genotypes are the same as those of an F2 population, but an IF2 can be regrown repeatedly. The use of rice IF2 populations to determine the genetic basis of heterosis has been proposed (Hua et al. 2003; Zhou et al. 2012). Here, we used an IF2 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 IF2 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 IF2 population including additive, dominance and two-locus interactions, and assessed their relative contributions to heterosis in the F1 hybrid. We also analyzed a simulated IF2 population and mid-parent heterosis (MPH) dataset to explore modes of allele action for MPH and performance of the hybrid.

Materials and methods

Plant materials and SNP bin map

We used a set of 294 F8 RILs derived using the single-seed descent method from Yuyu22. Similar to the procedure for generating the IF2 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 IF2 population (Fig. 1). The RIL and IF2 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 IF2 populations. The RILs were divided into two groups equally and randomly: RILA and RILB. This procedure was repeated three times, and finally $3 \times n$ F1 were produced, forming the IF2 population.

Theor Appl Genet
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 \( \frac{1}{2} \sum_{m=1}^{20} d^2 / \sum_{m=1}^{20} a^2 \), 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 LOID 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.
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 IF2 population. 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.

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 (\(d/\alpha\)), we categorized the bins into two types—dominance and overdominance—each of which included negative and positive values. The ratio \(d/\alpha\) was calculated in two ways: both additive and dominance genetic effects (\(\alpha\) and \(d\)) were estimated from the IF2 population; additive effect (\(\alpha\)) was estimated from RILs and dominance effect (\(d\)) from MPH dataset. For both the IF2 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 IF2 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 \(d/\alpha\)
were estimated, and for the MPH dataset only the dominance effects occurred; 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
V and extreme (d/a > 2.26) (Fig. 4). For the genetic architecture V = 1.26), moderate (2.25 > d/a > 1.76) and mild (1.75 > d/a > 1.25), 14.1 ± 2.23) in the IF2 population, followed by an average of 14.1 in the IF2, almost all the QTls in the MPH dataset was also detected in the IF2 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 = 0.5, an average of 11.9 of the 12.9 true-positive QTls in the MPH dataset was also detected in the IF2 population, in which 7.3 of 11.9 QTl exhibited overdominance. When V = 1, 2, or 3, a range of 93.1 to 96.7 % of true-positive QTls was detected in the MPH dataset. Simulation results also demonstrated that estimating the additive effects was more accurate for the RIL population than for the IF2 population, whereas the dominance effects were underestimated in the IF2 population and substantially less abundant in the MPH dataset than in the IF2 population (results not shown). In summary, the simulation analysis indicated identical modes of allele action for MPH and per se performance of the hybrid.

**Table 1** QTL identification among various genetic compositions in RILs, IF2 populations, and the MPH dataset in simulation studies

<table>
<thead>
<tr>
<th>Genetic architecture</th>
<th>No. of QTL with additive effect</th>
<th>No. of QTL with dominance effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>V_V_A = 0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RIL</td>
<td>14.1 ± 2.23</td>
<td>16.7 ± 1.49</td>
</tr>
<tr>
<td>RIL-L^a</td>
<td>2.7 ± 1.4</td>
<td>5.9 ± 1.66</td>
</tr>
<tr>
<td>RIL-E^b</td>
<td>10.4 ± 1.96</td>
<td>8.3 ± 1.83</td>
</tr>
<tr>
<td>RIL-H^c</td>
<td>1 ± 1.05</td>
<td>2.5 ± 1.65</td>
</tr>
<tr>
<td>RIL-F^d</td>
<td>0.6 ± 0.84</td>
<td>3.1 ± 1.79</td>
</tr>
<tr>
<td>V_V_A = 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RIL</td>
<td>14.1 ± 2.23</td>
<td>16.6 ± 1.52</td>
</tr>
<tr>
<td>RIL-L^a</td>
<td>2.7 ± 1.4</td>
<td>6.3 ± 1.83</td>
</tr>
<tr>
<td>RIL-E^b</td>
<td>10.4 ± 1.96</td>
<td>7.6 ± 2.55</td>
</tr>
<tr>
<td>RIL-H^c</td>
<td>1 ± 1.05</td>
<td>2.8 ± 1.48</td>
</tr>
<tr>
<td>RIL-F^d</td>
<td>0.6 ± 0.84</td>
<td>2.2 ± 1.4</td>
</tr>
<tr>
<td>V_V_A = 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RIL</td>
<td>14.1 ± 2.23</td>
<td>16.1 ± 1.45</td>
</tr>
<tr>
<td>RIL-L^a</td>
<td>2.7 ± 1.4</td>
<td>6.4 ± 1.26</td>
</tr>
<tr>
<td>RIL-E^b</td>
<td>10.4 ± 1.96</td>
<td>6.5 ± 2.17</td>
</tr>
<tr>
<td>RIL-H^c</td>
<td>1 ± 1.05</td>
<td>3.2 ± 1.87</td>
</tr>
<tr>
<td>RIL-F^d</td>
<td>0.6 ± 0.84</td>
<td>2.2 ± 1.48</td>
</tr>
<tr>
<td>V_V_A = 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RIL</td>
<td>14.1 ± 2.23</td>
<td>15.7 ± 1.7</td>
</tr>
<tr>
<td>RIL-L^a</td>
<td>2.7 ± 1.4</td>
<td>6.2 ± 1.55</td>
</tr>
<tr>
<td>RIL-E^b</td>
<td>10.4 ± 1.96</td>
<td>6.6 ± 2.41</td>
</tr>
<tr>
<td>RIL-H^c</td>
<td>1 ± 1.05</td>
<td>2.9 ± 1.97</td>
</tr>
<tr>
<td>RIL-F^d</td>
<td>0.6 ± 0.84</td>
<td>2.4 ± 1.26</td>
</tr>
</tbody>
</table>

^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.
Discussion

The bin map and IF₂ population provide an effective basis for complete genetic characterization of heterosis

We used a high-density bin map and an IF₂ population to investigate the genetic basis of heterotic enhancement of grain yield. The usefulness of IF₂ 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 IF₂ between flanking markers was 0.84 cM, which permitted identification of 14 more QTls than the SSR map in the IF₂ population of 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 IF₂ 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 to 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 IF₂ 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 IF₂ population, and all of them displayed overdominance in the IF₂ 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 IF₂ 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 IF₂ population, in which 61.3–90.5 % of QTL exhibited overdominance. The results from both of actual and simulated IF₂ 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 IF2 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 isomylase (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.

Author contributions J.Y. and J.L. designed and supervised the study. X.Y., J.H., and J.Y. performed the experiments. T.G., N.Y., Q.P., and H.T. analyzed the data. J. W. contributed to the simulation experimental design and analysis. T.G. and J.Y. prepared the manuscript, and all the authors read and approved the manuscript.

Acknowledgments The authors gratefully thank the editor Dr. Thomas Lubberstedt and two anonymous reviewers for their valuable suggestions. This research was supported by the National Hi-Tech Research and Development Program of China (2012AA10A307) and the National Basic Research Program of China.

Conflict of interest The authors declare that they have no conflict of interest.

References


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


F1-hybrid and parental inbred line primary root transcriptomes suggests organspecific patterns of nonadditive gene expression and conserved expression trends. Genetics 179:1275–1283
Jones DF (1917) Dominance of linked factors as a means of accounting for heterosis. Genetics 2:466–479
Richey FD (1942) Mock-dominance and hybrid vigor. Science 96:280
Shull GH (1908) The composition of a field of maize. Am Breed Assoc Rep 4:296–301