DOI: 10.1534/genetics.107.082867

Quantitative Trait Loci Mapping and The Genetic Basis of Heterosis in Maize and Rice

Antonio Augusto Franco Garcia,* Shengchu Wang,† Albrecht E. Melchinger‡ and Zhao-Bang Zeng†,§,1

**Departamento de Genética, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo CP 83, 13400-970, Piracicaba, SP, Brazil, †Department of Statistics and Bioinformatics Research Center and §Department of Genetics, North Carolina State University, Raleigh, North Carolina 27695-7566 and ‡Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, 70599 Stuttgart, Germany*

Manuscript received October 4, 2007
Accepted for publication September 8, 2008

ABSTRACT

Despite its importance to agriculture, the genetic basis of heterosis is still not well understood. The main competing hypotheses include dominance, overdominance, and epistasis. NC design III is an experimental design that has been used for estimating the average degree of dominance of quantitative trait loci (QTL) and also for studying heterosis. In this study, we first develop a multiple-interval mapping (MIM) model for design III that provides a platform to estimate the number, genomic positions, augmented additive and dominance effects, and epistatic interactions of QTL. The model can be used for parents with any generation of selfing. We apply the method to two data sets, one for maize and one for rice. Our results show that heterosis in maize is mainly due to dominant gene action, although overdominance of individual QTL could not completely be ruled out due to the mapping resolution and limitations of NC design III. For rice, the estimated QTL dominant effects could not explain the observed heterosis. There is evidence that additive × additive epistatic effects of QTL could be the main cause for the heterosis in rice. The difference in the genetic basis of heterosis seems to be related to open or self pollination of the two species. The MIM model for NC design III is implemented in Windows QTL Cartographer, a freely distributed software.

HETEROSIS (or hybrid vigor) is a phenomenon in which an F_1 hybrid has superior performance over its parents. It has been observed in many plant and animal species. The utilization of heterosis is responsible for the commercial success of plant breeding in many species and leads to the widespread use of hybrids in several crops and horticultural species. In maize, the most notable example, heterosis is the primary reason for the success of commercial industry (Stuber *et al.* 1992). In China, hybrid rice varieties showed ~20% yield advantage over inbred varieties (Yuan 1992) and made a tremendous impact on rice production around the world.

Despite its importance, the genetic basis of heterosis has been debated for almost one century and is still not explained satisfactorily. The *dominance* hypothesis (Davenport 1908; Bruce 1910; Keeble and Pellew 1910; Jones 1917) suggests that the alleles from one parent are dominant over the alleles from the other parent, and due to the cancelation of deleterious effects at multiple loci, the F_1 hybrid is superior to the parents. The *overdominance* hypothesis (East 1908; Shull 1908) assumes that the loci with heterozygous genotypes are superior to both homozygous parents. Epistasis is also frequently mentioned as a possible cause of heterosis.

NC design III, or design III (Comstock and Robinson 1948, 1952), is an experimental design for estimating genetic variances and the average degree of dominance for quantitative trait loci (QTL) and has being used to study heterosis. Random F_2 individuals are taken from a population that originated by crossing two inbred lines. These individuals are backcrossed to both parental lines and a quantitative trait is measured in the progeny. An analysis of variance of the progenies gives estimates of the average degree of dominance, which can be used to infer the genetic basis of quantitative traits and study heterosis. Cockerham and Zeng (1996) extended the analysis of design III to include linkage, two-locus epistasis, and also the use of F_3 parents. Considering that the F_2 (or F_3) parents could be genotyped with molecular markers, they presented a statistical methodology based on four orthogonal contrasts for single-marker analysis of design III, allowing the study of the effects of QTL on both backcrosses simultaneously. Melchinger *et al.* (2007) studied the role of epistasis on the manifestation of heterosis in design III populations. They defined new types of heterotic genetic effects, the augmented additive and dominance effects

"Design III with marker loci" was the last article published by C. Clark Cockerham. This article is dedicated to his memory.

[1] *Corresponding author:* Bioinformatics Research Center, Department of Statistics, North Carolina State University, Raleigh, NC 27695-7566. E-mail: zeng@stat.ncsu.edu

of QTL, since the main effects also contain epistasis that could not be removed or estimated separately.

Stuber *et al.* (1992) used design III with marker loci to study the genetic basis of heterosis in maize. They conducted separate interval mapping analyses (Lander and Botstein 1989) in each backcross and concluded that overdominance (or pseudo-overdominance) is the major cause of heterosis. However, a combined analysis of both backcrosses showed that dominance is probably more likely to be a major cause of heterosis (Cockerham and Zeng 1996), although overdominance and epistasis were also present. In rice, design III using F_7 parents was used by Xiao *et al.* (1995) and the data were analyzed in the same way as that of Stuber *et al.* (1992). They concluded that dominance is the major genetic cause of heterosis in this species. Later, Z.-B. Zeng (unpublished results) analyzed this data set using the method of Cockerham and Zeng and concluded that epistasis is more likely to be a major cause of heterosis in rice.

The statistical analysis proposed by Cockerham and Zeng has several advantages. It allows estimates of both additive and dominance effects and has two contrasts for testing the presence of epistasis. However, it is based on single-marker analysis and was not developed for QTL mapping. The method has several limitations: the contrasts are biased due to the recombination fraction between marker and QTL, it is not possible to separate the additive and dominance effects of several QTL linked to the same marker, the contrasts for epistasis detect only a small portion of the interactions between QTL that are linked to the same marker, and it has low statistical power.

In this article, we first extend the method of Cockerham and Zeng in the framework of multiple-interval mapping (MIM) (Kao and Zeng 1997; Kao *et al.* 1999), which provides a sound basis for QTL mapping. Our MIM model for design III combines information from multiple markers and takes epistatic effects into account. By analyzing both backcrosses simultaneously, it provides estimates of augmented additive and dominance effects. The model can be used for parents with any number of generations in selfing. Then, we apply the model to the data of Stuber *et al.* (1992) and Xiao *et al.* (1995) to study the genetic basis of yield heterosis in maize and rice.

DESIGN III WITH MARKER LOCI

Before presenting the new model for design III, we first outline some important results for design III from Comstock and Robinson (1952) and Cockerham and Zeng (1996), adapting the notation when necessary. The genetic effects of QTL Q_r with genotypes Q_rQ_r, Q_rq_r, and q_rq_r are defined as $a_r - d_r/2$, $d_r/2$, and $-a_r - d_r/2$, respectively (using the F_2 model, see Zeng *et al.* 2005), where a_r and d_r are additive and dominance effects. The two-way epistatic interactions between QTL Q_r and Q_s are denoted as aa_{rs} for additive × additive ($a_r \times a_s$), ad_{rs} for additive × dominance ($a_r \times d_s$), da_{rs} for dominance × additive ($d_r \times a_s$), and dd_{rs} for dominance × dominance ($d_r \times d_s$) interaction.

On the basis of an analysis of variance for progenies of F_2 parents in the backcrosses in design III, Comstock and Robinson developed a theory for estimating genetic variances among F_2 parents (σ^2_{p}) and due to interactions of F_2 and inbred parents ($\sigma^2_{\mathrm{p}_j}$). They showed that, under the assumption of no epistasis for m independent loci, the genetic constitutions of these variances are $\sigma^2_{\mathrm{p}} = \sum_{r=1}^{m} a_r^2/8$ and $\sigma^2_{\mathrm{p}_j} = \sum_{r=1}^{m} d_r^2/4$. Cockerham and Zeng expanded these ideas to include F_3 parents, showing that in this case $\sigma^2_{\mathrm{p}} = 3\sum_{r=1}^{m} a_r^2/16$ and $\sigma^2_{\mathrm{p}_j} = 3\sum_{r=1}^{m} d_r^2/8$. For F_2 (and F_3) parents, the average degree of dominance for a quantitative trait can be inferred through the ratio $\bar{D} = \sqrt{\sigma^2_{\mathrm{p}_j}/(2\sigma^2_{\mathrm{p}})}$. When two-locus epistasis is considered, the additive effects include *ad* and *da*, and the dominance effects include *aa*, regardless of linkage. The variances are also affected: σ^2_{p} contains a and $aa + dd$; $\sigma^2_{\mathrm{p}_j}$ contains d and $ad + da$. However, the coefficients of epistatic effects on the variances are usually small.

Considering that information from molecular markers could be available, Cockerham and Zeng presented a statistical method to analyze design III in the framework of single-marker analysis. For a single-marker locus M with genotypes MM, Mm, and mm for each parent (F_2 or F_3), four orthogonal contrasts C_k ($k = 1, \ldots, 4$) can be used for testing linear functions of effects of QTL. The four contrasts explore the 2 d.f. for differences among the means of marker genotypes (C_1 and C_3) and the 2 d.f. for interaction of the marker genotypes with the inbred lines (C_2 and C_4).

To obtain a MIM model for design III, we first extend the contrasts of Cockerham and Zeng still in the framework of marker analysis (not interval mapping), but considering simultaneously two marker loci (M_1 and M_2) observed for F_2 parents and two QTL (Q_1 and Q_2). Then, we generalize the results for any number of QTL in any genomic position and develop a MIM model for design III.

Assume that the loci are linked with the order $Q_1M_1M_2Q_2$. We denote ρ_1, ρ, ρ_2, and ρ_{12} as recombination fractions for the intervals between Q_1 and M_1, M_1 and M_2, M_2 and Q_2, and Q_1 and Q_2, respectively. We calculated the relative frequencies of QTL genotypes given the marker genotype in the F_2 parent for two loci (Table 1) and then derived the genotypic means of the progenies in both backcrosses (appendix a). These means were denoted as H^j_g, where j is the inbred line ($j = 2, 1$) and g is the genotype of the two markers in the F_2 parent.

It is possible to define 17 orthogonal contrasts for testing differences among H^j_g means (appendix b). These contrasts correspond to an orthogonal decomposition of the degrees of freedom available when two loci and two backcrosses are considered. There are 2 d.f. for differences for marker genotypes of M_1, 2 for marker genotypes of M_2, 4 for the interaction $M_1 \times M_2$, 2 for the

TABLE 1

Conditional frequency of the QTL gamete from F_2 given the marker genotype

			QTL gametic frequencies			
Marker	f	g	Q_1Q_2	Q_1q_2	q_1Q_2	q_1q_2
$M_1M_1M_2M_2$	$\frac{(1-\rho)^2}{4}$	22	$(1-\rho_1)(1-\rho_2)$	$(1-\rho_1)\rho_2$	$\rho_1(1-\rho_2)$	$\rho_1\rho_2$
$M_1M_1M_2m_2$	$\frac{\rho(1-\rho)}{2}$	21	$\frac{1}{2}(1-\rho_1)$	$\frac{1}{2}(1-\rho_1)$	$\frac{1}{2}\rho_1$	$\frac{1}{2}\rho_1$
$M_1M_1m_2m_2$	$\frac{\rho^2}{4}$	20	$(1-\rho_1)\rho_2$	$(1-\rho_1)(1-\rho_2)$	$\rho_1\rho_2$	$\rho_1(1-\rho_2)$
$M_1m_1M_2M_2$	$\frac{\rho(1-\rho)}{2}$	12	$\frac{1}{2}(1-\rho_2)$	$\frac{1}{2}\rho_2$	$\frac{1}{2}(1-\rho_2)$	$\frac{1}{2}\rho_2$
$M_1m_1M_2m_2$	$\frac{(1-\rho)^2}{2}+\frac{\rho^2}{2}$	11	$\frac{-1}{\zeta}[\rho_{12}(1-\rho_{12})$ $-\frac{1}{4}(1+\zeta)]$	$\frac{1}{\zeta}[\rho_{12}(1-\rho_{12})$ $-\frac{1}{2}\rho(1-\rho)]$	$\frac{1}{\zeta}[\rho_{12}(1-\rho_{12})$ $-\frac{1}{2}\rho(1-\rho)]$	$\frac{-1}{\zeta}[\rho_{12}(1-\rho_{12})$ $-\frac{1}{4}(1+\zeta)]$
$M_1m_1m_2m_2$	$\frac{\rho(1-\rho)}{2}$	10	$\frac{1}{2}\rho_2$	$\frac{1}{2}(1-\rho_2)$	$\frac{1}{2}\rho_2$	$\frac{1}{2}(1-\rho_2)$
$m_1m_1M_2M_2$	$\frac{\rho^2}{4}$	02	$\rho_1(1-\rho_2)$	$\rho_1\rho_2$	$(1-\rho_1)(1-\rho_2)$	$(1-\rho_1)\rho_2$
$m_1m_1M_2m_2$	$\frac{\rho(1-\rho)}{2}$	01	$\frac{1}{2}\rho_1$	$\frac{1}{2}\rho_1$	$\frac{1}{2}(1-\rho_1)$	$\frac{1}{2}(1-\rho_1)$
$m_1m_1m_2m_2$	$\frac{(1-\rho)^2}{4}$	00	$\rho_1\rho_2$	$\rho_1(1-\rho_2)$	$(1-\rho_1)\rho_2$	$(1-\rho_1)(1-\rho_2)$

f is frequency of marker genotype; g is a coded variable for marker genotypes; ρ_1, ρ, ρ_2, and ρ_{12} are the recombination fractions between M_1 and Q_1, M_1 and M_2, Q_2 and M_2, and Q_1 and Q_2, respectively; $\zeta = 1 - 2\rho + 2\rho^2$.

interaction of marker M_1 with the inbred lines, 2 for the interaction of M_2 with the inbred lines, 4 for the interaction $M_1 \times M_2$ with inbred lines, and 1 for the difference between inbred lines. Using the genotypic means of the progenies and following the definitions of genetic effects based on the F_2 genetic model according to Cockerham and Zeng (1996; Zeng *et al.* 2005), we derived the genetic expectation of these 17 contrasts (appendix b).

There are seven QTL genotypes present in a population that originated from design III when two QTL are considered. It is important to note that some QTL genotypes do not occur in the backcross populations. For example, marker genotypes in the F_2 parents include M_1m_2/M_1m_2, but there is no QTL genotype Q_1q_2/Q_1q_2 in the backcross populations. Also not present is q_1Q_2/q_1Q_2. Hence, for a pair of QTL, it is possible to define only six contrasts for the differences between genotypes, even though there are eight parameters to be estimated (a_1, a_2, d_1, d_2, aa, ad, da, and dd). As a consequence, it is not possible to estimate all genetic parameters separately. Also, some of the 17 contrasts do not provide useful information for the genetic effects, because the genetic expectations are based on the segregating QTL in the backcross populations, not on the F_2 marker genotypes. For example, contrasts c_6, c_7, c_{15}, and c_{16} have genetic expectations equal to zero. Contrasts c_2 and c_4 have the same expectation, which is $-\frac{1}{2}$ of c_8. The same happens to c_{11}, c_{13}, and c_{17}.

TABLE 2

Orthogonal contrasts for the analysis of design III

Contrast	H^2_{22}	H^2_{21}	H^2_{20}	H^2_{12}	H^2_{11}	H^2_{10}	H^2_{02}	H^2_{01}	H^2_{00}
$\tilde{C}_1$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	0	0	0	$\frac{-1}{6}$	$\frac{-1}{6}$	$\frac{-1}{6}$
$\tilde{C}_3$	$\frac{1}{6}$	0	$\frac{-1}{6}$	$\frac{1}{6}$	0	$\frac{-1}{6}$	$\frac{1}{6}$	0	$\frac{-1}{6}$
$\tilde{C}_5$	$\frac{5}{6}$	$\frac{1}{3}$	$\frac{-1}{6}$	$\frac{1}{3}$	$\frac{-8}{3}$	$\frac{1}{3}$	$\frac{-1}{6}$	$\frac{1}{3}$	$\frac{5}{6}$

H^j_g is the genotypic mean of the backcross progenies from F_2 parents with marker genotype g backcrossed to parental line j ($j = 2, 1$). Only coefficients of $\tilde{C}_1$, $\tilde{C}_3$, and $\tilde{C}_5$ are given for H^2_g means. The coefficients of $\tilde{C}_1$, $\tilde{C}_3$, and $\tilde{C}_5$ for H^1_g are the same as those for H^2_g. $\tilde{C}_2$, $\tilde{C}_4$, and $\tilde{C}_6$ have the same coefficients as $\tilde{C}_1$, $\tilde{C}_3$, and $\tilde{C}_5$ for H^1_g; but for H^2_g, the coefficients have opposite signs.

Taking these into account, a new set of six orthogonal contrasts that provide useful information about the genetic parameters was defined (Table 2). Let $\tilde{C}_1 = c_1/6$, $\tilde{C}_2 = c_{10}/6$, $\tilde{C}_3 = c_3/6$, $\tilde{C}_4 = c_{12}/6$, $\tilde{C}_5 = c_5/2 + (c_2 + c_4 - c_8)/3$, and $\tilde{C}_6 = c_{14}/2 + (c_{11} + c_{13} - c_{17})/3$. The genetic expectations of these new contrasts are

$$
\begin{aligned}
E(\tilde{C}_1) &= (1-2\rho_1)a_1 - \frac{1}{2}(1-2\rho_1)da \\
E(\tilde{C}_2) &= (1-2\rho_1)d_1 - \frac{1}{2}(1-2\rho_1)aa \\
E(\tilde{C}_3) &= (1-2\rho_2)a_2 - \frac{1}{2}(1-2\rho_2)ad \\
E(\tilde{C}_4) &= (1-2\rho_2)d_2 - \frac{1}{2}(1-2\rho_2)aa \\
E(\tilde{C}_5) &= \frac{[(-16\rho+8)\rho_{12}+6\rho^2+2\rho-1](1-2\rho_1)(1-2\rho_2)}{3(1-2\rho+2\rho^2)} \\
&\quad \times (aa+dd) \\
E(\tilde{C}_6) &= \frac{[(-16\rho+8)\rho_{12}+6\rho^2+2\rho-1](1-2\rho_1)(1-2\rho_2)}{3(1-2\rho+2\rho^2)} \\
&\quad \times (ad+da).
\end{aligned}
$$

Contrasts $\tilde{C}_1$–$\tilde{C}_4$ are for additive and dominance effects and came directly from contrasts c_1, c_{10}, c_3, and c_{12}, respectively. They can be viewed as contrasts between marginal means of genotypic classes. Because we do not have all QTL genotypes, it is not possible in this case to define contrasts to test only the main effects (additive and dominance) without some bias due to epistatic effects. However, by considering contrasts for two QTL simultaneously, it is possible to test additive and dominance effects (plus epistatic effects) even if the two QTL are linked.

For epistasis, it is also not possible to separate *aa* from *dd* and *ad* from *da*. To test $aa + dd$, the contrast $c_5/2$ could be used. It is important to note that c_5 does not use the means from genotypes that are heterozygous for at least one marker locus. Thus, by using $c_5/2$, means H_{11}^2 and H_{11}^1 will not be used in the analysis. Also, contrasts c_2, c_4, and c_8, which could be used for estimating $aa + dd$, have the expectation zero if the markers are unlinked ($\rho = \frac{1}{2}$), which is an obvious disadvantage. Therefore, we suggest using a linear combination of contrasts (defined as $\tilde{C}_5$) that uses all H_g^j means. Note that if $\rho = \frac{1}{2}$, $E(\tilde{C}_5) = (1 - 2\rho_1)(1 - 2\rho_2)(aa + dd)$. The same argument applies to $\tilde{C}_6$, designed to test $ad + da$. Using u_{kgj} to denote the coefficients of contrasts in Table 2, the kth contrast is $\tilde{C}_k = \sum_g \sum_j u_{kgj} H_g^j$. The six new contrasts are orthogonal because $\sum_g \sum_j u_{kgj} u_{k'gj} = 0$ for any pair $\tilde{C}_k$ and $\tilde{C}_{k'}$ $(k \neq k')$.

The bias in the expectations of contrasts due to ρ_1 and ρ_2 can be removed by using multiple-interval mapping (next section). In MIM, we search and estimate the positions of QTL. Thus it is possible to test contrasts between putative QTL, not markers. This means that potentially $\rho_1 = 0$ and $\rho_2 = 0$; thus $E(\tilde{C}_1) = a_1 - \frac{1}{2}da$, $E(\tilde{C}_2) = d_1 - \frac{1}{2}aa$, $E(\tilde{C}_3) = a_2 - \frac{1}{2}ad$, and $E(\tilde{C}_4) = d_2 - \frac{1}{2}aa$. For epistasis, $E(\tilde{C}_5) = -((1 - 10\rho + 10\rho^2)/3(1 - 2\rho + 2\rho^2))(aa + dd)$ and $E(\tilde{C}_6) = -((1 - 10\rho + 10\rho^2)/3(1 - 2\rho + 2\rho^2))(ad + da)$. For unlinked QTL with $\rho = \frac{1}{2}$, $E(\tilde{C}_5) = (aa + dd)$ and $E(\tilde{C}_6) = (ad + da)$. This shows that given a correct identification of QTL model, the statistical analysis in the framework of MIM can minimize the bias in estimation and increase statistical power. Also, it is possible to test epistasis between any two QTL, not just QTL that are linked to a marker as in the approach of Cockerham and Zeng (1996).

In a study of the role of epistasis in the manifestation of heterosis, Melchinger *et al.* (2007) defined $a_r^* = [a_r - \frac{1}{2}\sum_{r \neq s} da_{rs}]$ as an augmented additive effect of QTL r and $d_r^* = [d_r - \frac{1}{2}\sum_{r \neq s} aa_{rs}]$ as an augmented dominance effect. These augmented effects are exactly the ones contained in contrasts $\tilde{C}_1$–$\tilde{C}_4$, if we generalize the expressions to multiple QTL. Therefore, in a statistical analysis by MIM, we estimate and test a_r^* and d_r^* as well as epistasis effects.

MIM MODEL FOR DESIGN III

The six new contrasts for two markers (Table 1) were used for the development of a MIM model for design III. Multiple-interval mapping (Kao and Zeng 1997; Kao *et al.* 1999; Zeng *et al.* 1999) is a procedure for mapping multiple QTL simultaneously with a model fitted with main and epistatic effects of multiple QTL. Combined with a search procedure, it tests and estimates the positions, effects, and interactions of multiple QTL.

Statistical model: The MIM model for design III is defined by generalizing the six contrasts for any number of putative QTL and level of inbreeding of the parents,

$$y_{ij} = \mu_j + \sum_{r=1}^{m} \alpha_r x_{ijr}^* + \sum_{r=1}^{m} \beta_r z_{ijr}^* + \sum_{r<s}^{t_1} \gamma_{rs} w_{ijrs}^* + \sum_{r<s}^{t_2} \delta_{rs} o_{ijrs}^* + \varepsilon_{ij}, \tag{1}$$

where y_{ij} is the phenotypic mean of the progenies of parent i ($i = 1, \ldots, n$) on the backcross with inbred line j ($j = 1, 2$). The parameters are the mean of backcross j (μ_j), the regression coefficients for augmented additive effect (a^*) and dominance (d^*) effect of QTL r (α_r and β_r, respectively), and the regression coefficients for epistatic interactions $aa + dd$ and $ad + da$ between QTL r and s (γ_{rs} and δ_{rs}, respectively). The residuals ε_{ij} are assumed to be $N(0, \sigma_j^2)$. The variables x_{ijr}^*, z_{ijr}^*, w_{ijrs}^*, and o_{ijrs}^* denote QTL genotypes corresponding to the main and epistatic effects specified by the six contrasts. They were coded as

$$x_{ijr}^* = \begin{cases} 1 & \text{if the genotype of } Q_r \text{ is } Q_rQ_r \\ 0 & \text{if the genotype of } Q_r \text{ is } Q_rq_r \\ -1 & \text{if the genotype of } Q_r \text{ is } q_rq_r \end{cases} \quad \text{for } j = 1, 2;$$

$$z_{ijr}^* = \begin{cases} x_{ijr}^* & \text{if } j = 1 \\ -x_{ijr}^* & \text{if } j = 2 \end{cases}$$

$$w_{ijrs}^* = \begin{cases} \frac{5}{6} & \text{if the QTL genotype is } Q_rQ_rQ_sQ_s \\ \frac{1}{6} & \text{if the QTL genotype is } Q_rQ_rQ_sq_s \\ \frac{-1}{6} & \text{if the QTL genotype is } Q_rQ_rq_sq_s \\ \frac{1}{6} & \text{if the QTL genotype is } Q_rq_rQ_sQ_s \\ \frac{-4}{6} & \text{if the QTL genotype is } Q_rq_rQ_sq_s \\ \frac{1}{6} & \text{if the QTL genotype is } Q_rq_rq_sq_s \\ \frac{-1}{6} & \text{if the QTL genotype is } q_rq_rQ_sQ_s \\ \frac{1}{6} & \text{if the QTL genotype is } q_rq_rQ_sq_s \\ \frac{5}{6} & \text{if the QTL genotype is } q_rq_rq_sq_s \end{cases} \quad \text{for } j = 1, 2;$$

$$o_{ijrs}^* = \begin{cases} w_{ijrs}^* & \text{if } j = 1 \\ -w_{ijrs}^* & \text{if } j = 2 \end{cases}.$$

The first two summations are over the m QTL currently fitted in the model, and the last ones are for significant t_1 and t_2 two-way epistatic interactions. The coefficients for the coded variables can be seen as a generalization of the orthogonal contrasts developed for two markers with some adaptations.

For design III from recombinant inbred lines (after continuing selfing from F_2 for a number of generations), the model can be further simplified. As a consequence of selfing, we note in Table 3 that the proportion of homozygous genotypes for at least one locus is becoming smaller in relation to the others. So, if the parents used in design III have several generations of selfing, the contrasts and the MIM model should be adapted to this situation. Details are presented in appendix c.

TABLE 3

Orthogonal contrasts for design III with two markers

Contrast	H^2_{22}	H^2_{21}	H^2_{20}	H^2_{12}	H^2_{11}	H^2_{10}	H^2_{02}	H^2_{01}	H^2_{00}
c_1	1	1	1	0	0	0	−1	−1	−1
c_2	1	1	1	−2	−2	−2	1	1	1
c_3	1	0	−1	1	0	−1	1	0	−1
c_4	1	−2	1	1	−2	1	1	−2	1

H^j_g is the genotypic mean of the backcross progenies from F_2 parents with marker genotype g (see APPENDIX A) backcrossed to parental line j ($j = 2, 1$). Only H^2_g means are presented, and the coefficients for H^1_g are the same as for H^2_g for c_1–c_4. Contrasts c_5–c_8 are $c_5 = c_1 \times c_3$, $c_6 = c_1 \times c_4$, $c_7 = c_2 \times c_3$, and $c_8 = c_2 \times c_4$. Contrast c_9 has $u_{9g1} = 1$ and $u_{9g2} = -1$. Contrasts c_{10}–c_{17} have the same coefficients as c_1–c_8 for H^1_g, respectively; for H^2_g the coefficients are the same but with opposite signs.

Likelihood and parameter estimation: As pointed out by Kao *et al.* (1999), MIM models contain missing data, since the QTL genotypes are not observed. Therefore, the likelihood function for the model, assuming that the y_{ij}'s are independent across observations and backcrosses, is

$$L(\mathbf{E}, \mu_j, \sigma^2_j \mid \mathbf{Y_j}, \mathbf{X}) = \prod_{i=1}^{n}\left[\sum_{g=1}^{3^m} p_{ig} \prod_{j=1}^{2} \phi(y_{ij} \mid \mu_j + \mathbf{D_{jg}E}, \sigma^2_j)\right],$$

where $\mathbf{Y_j}$ is a vector of phenotypic data for backcross j, $\mathbf{X}$ is a matrix with molecular data, g indicates the 3^m multiple-QTL genotypes, p_{ig} is the probability of each multilocus genotype conditional on marker data, $\phi(.)$ is a standard normal probability density function, $\mathbf{E}$ is a column vector with QTL parameters (α's, β's, γ's, and δ's), and $\mathbf{D_{jg}}$ is a row vector that specifies the configuration of x^*'s, z^*'s, w^*'s, and o^*'s associated with the parameters on $\mathbf{E}$ in each backcross (following the notation of Kao and Zeng 1997).

To obtain the maximum-likelihood estimates (MLEs), we adapted the general formulas of Kao and Zeng (1997) to the MIM model for design III, on the basis of the expectation-maximization (EM) algorithm (Dempster *et al.* 1977). The E and M steps are iterated until some convergence criteria are met and the converged values are the MLEs. Details are presented in APPENDIX D.

After the final model is selected, it is necessary to convert the estimates of the regression coefficients to the contrasts, which contain the desired genetic effects. This can be easily done on the basis of the genotypic expectations of the coefficients. For any type of selfing parents (F_2 to F_∞), for estimating augmented additive and dominance effects we simply multiply $\hat{\alpha}_r$ and $\hat{\beta}_r$ by 2, since $E(\hat{\alpha}_r) = \frac{1}{2}[a_r - \frac{1}{2}\sum_{r \neq s} da_{rs}] = \frac{1}{2}\, a^*_r$ and $E(\hat{\beta}_r) = \frac{1}{2}[d_r - \frac{1}{2}\sum_{r \neq s} aa_{rs}] = \frac{1}{2}\, d^*_r$. For epistasis between unlinked QTL, for F_2 (or F_3, etc.) parents $E(\hat{\gamma}_{rs}) = \frac{9}{31}(aa + dd)$ and $E(\hat{\delta}_{rs}) = \frac{9}{31}(ad + da)$. For homozygous parents (F_∞), the expectations are $E(\hat{\gamma}_{rs}) = \frac{1}{2}(aa + dd)$ and $E(\hat{\delta}_{rs}) = \frac{1}{2}(ad + da)$.

Melchinger *et al.* (2007) pointed out that a^*_r and d^*_r are the net contributions of QTL r to parental difference and midparent heterosis, respectively, considering simultaneously main effects and epistatic interactions with the genetic background. Therefore, by providing estimates of a^*_r, d^*_r, and epistasis, the MIM model for design III can be very useful for studying the genetic basis of heterosis.

Strategy for QTL mapping: The usual procedures for model selection in MIM can be used here and were discussed in detail by Kao *et al.* (1999) and Zeng *et al.* (1999). Briefly, forward, backward, and stepwise procedures can be applied, combined with selection criteria, such as Akaike information criteria (AIC) (Akaike 1974), the Bayesian information criterion (BIC) (Schwarz 1978), or the likelihood-ratio test. In stepwise selection, for a model with m QTL, the genome is scanned to find the best position of an $(m + 1)$th QTL. Then, all the QTL in the model are tested, one by one, to check if one of them should be removed. The process is repeated until no QTL was added or removed, and then the positions are refined. After finding the final model for main effects, the procedure can be repeated to identify significant epistatic effects.

ANALYSIS OF A MAIZE DATA SET

Experiment description: We applied our model to the maize data of Stuber *et al.* (1992), where detailed information about the experiment can be found. Briefly, starting from two inbred lines, *Mo17* (L_1) and *B73* (L_2), 264 F_3 lines were created and backcrossed to the two inbred lines. The backcross progenies of each of the F_3 parents were allocated in 22 sets of 12 parents and then evaluated in six locations or environments without further replication. Seven traits were measured on the backcross progenies, but we used just the adjusted means across locations for grain yield, calculated using the type III analysis of variance in the SAS general linear models procedure. Only 11 observations were missing. The F_3 parents were genotyped with RFLP and isozyme markers and a genetic map was built using the Kosambi map function to express distances in centimorgans. We used the same 73 markers analyzed by Cockerham and Zeng, obtaining multipoint estimates with MAPMAKER/EXP (Lander *et al.* 1987) for the distances not presented in their article.

Statistical analysis: *Interval mapping for design III:* First, we applied interval mapping (IM) for design III for the maize data. This corresponds to model (1) with only one QTL fitted in the model. This was done to (1) have comparisons with the results of Stuber *et al.* (using IM for each backcross separately) and Cockerham and

Zeng (using four contrasts for single-marker analysis of both backcrosses simultaneously) and (2) help on the selection of the final MIM model.

MIM for design III: To select number and map positions of putative QTL to be included in an initial model, a forward procedure was used on the basis of the ideas of Kao *et al.* (1999). Starting with a model with no QTL, a model with one QTL that resulted in the greatest increase in the likelihood was selected. The procedure was repeated for adding a second QTL and so on until no further QTL can be added with a model of, say, m QTL. The models with $m-1$ and m QTL were compared on the basis of BIC (Schwartz 1978). We also tried to add QTL on positions suggested by IM for design III, keeping them in the model if the effects were significant. When the QTL number of a model is changed, estimates of QTL positions were optimized. After a model with main effects and refined positions was established, a forward/backward procedure was applied to identify two-way epistasis between QTL. Every possible epistatic effect was tested and the one with the highest likelihood was selected. The procedure was repeated until no more effects could be added. We note that in using BIC few epistatic effects remain in the model. Since we are interested in estimating epistatic effects on heterosis, a less conservative criterion, AIC (Akaike 1974), was adopted. After epistatic effects were selected, all main and epistatic effects were tested for significance and the nonsignificant effects were removed. If the main effects of a QTL were not significant but it had some significant epistasis with at least one other QTL, it was kept in the model.

Results: *IM for design III:* The results for QTL mapping for grain yield are presented in Figure 1, A and B. In general, they are in close agreement with the previous analysis of Stuber *et al.* and Cockerham and Zeng, but provide more information and statistical power. Stuber *et al.* did the analysis on each backcross separately. A QTL was mapped if it had a significant effect in at least one backcross. We note that using IM for design III there are LOD peaks approximately in the same genomic regions previously identified, but the shape of the new curves is similar to the sum of the previous ones, with higher LOD scores. This is an indication of higher statistical power and results in more identifiable peaks in some regions, such as chromosomes 1 and 10. On the backcrosses to *B73* and *Mo17*, Stuber *et al.* found six and eight QTL, respectively, with LOD scores varying from 2.73 to 9.73. We also found evidence for QTL in the same regions, but with LOD scores between ~10 and 35. On chromosomes 8 and 10, the QTL that were barely detectable by the analysis on each backcross separately now have LOD scores ~10.

The separate analysis on each backcross can lead to difficult interpretation about QTL number. This can be alleviated by the new analysis. For example, on chromosome 10, IM for design (D)III shows a profile indicating that there is evidence for only one QTL in the middle of the chromosome, instead of two indicated before. However, IM for DIII still has some problems. For example, using an arbitrary LOD threshold of 3, it is difficult to precisely indicate how many QTL are on chromosomes 1, 2, 4, 5, 8, 9, and 10.

As pointed out by Cockerham and Zeng, by analyzing the backcrosses separately and estimating the genetic effects in terms of differences between heterozygous and homozygous, Stuber *et al.* actually estimated $d^* + a^*$ for the backcross to *Mo17* and $d^* - a^*$ for the backcross to *B73* ($d + a$ and $d - a$ in their notation). As a consequence, if a^* and d^* have the same magnitude, the QTL will not be identified in one backcross and its effect will be aggregated in the other. This seems to be the case for the QTL on chromosomes 3 and 4, where only one LOD curve is above the threshold. With IM for DIII, a^* and d^* can be estimated separately.

The Cockerham and Zeng approach does not provide LOD curves or an indication about QTL number, but their *P*-values can be used to identify genomic regions for the evidence of QTL. Their method is based on the analysis of both backcrosses simultaneously and also allows the estimation of a^* and d^* associated with markers. Marker analysis for all chromosomes has significant effects for at least one of the four contrasts. In general, there is correspondence between small *P*-values and LOD peaks for IM for design III, specially for d^* effects, which are the most significant ones. It is noted that d^* is positive in almost every position (with exceptions at the beginning of chromosomes 3 and 9) and is consistently larger in magnitude than a^*, whose sign varies from region to region. Few a^* effects were significant, mostly on chromosomes 3 and 4.

MIM for design III: We use this analysis to provide some detailed estimates and to provide some interpretation on the basis of these estimates (Figure 1, A and B; Tables 4–6). Compared to other methods, this analysis tends to provide better estimates on QTL number, positions, effects, and epistasis. Thirteen putative QTL were mapped in nine chromosomes with LOD score >5 (except for the closely linked QTL X and XI). All QTL together explain 74.90 and 78.23% of the phenotypic variation in backcrosses to *Mo17* and *B73*, respectively. These values are higher than the ones found by Stuber *et al.* (59.1 and 60.9%). The main effects of each QTL individually explained from 0.61 to 12.34% of the phenotypic variation.

The estimates of a^* are both positive and negative. However, the values of d^* are consistently positive and are generally higher than those of a^*. When a^* is positive, the favorable allele comes from *B73*, and when negative, it comes from *Mo17*. The magnitude of the effects varies from -5.48 to 6.28 for a^* and from 0.36 to 9.18 for d^*. These are generally consistent with Stuber *et al.*'s results. For example, they had estimates of $d^* + a^*$ for QTL IV and VI with values 11.57 and 10.55, respectively. In our

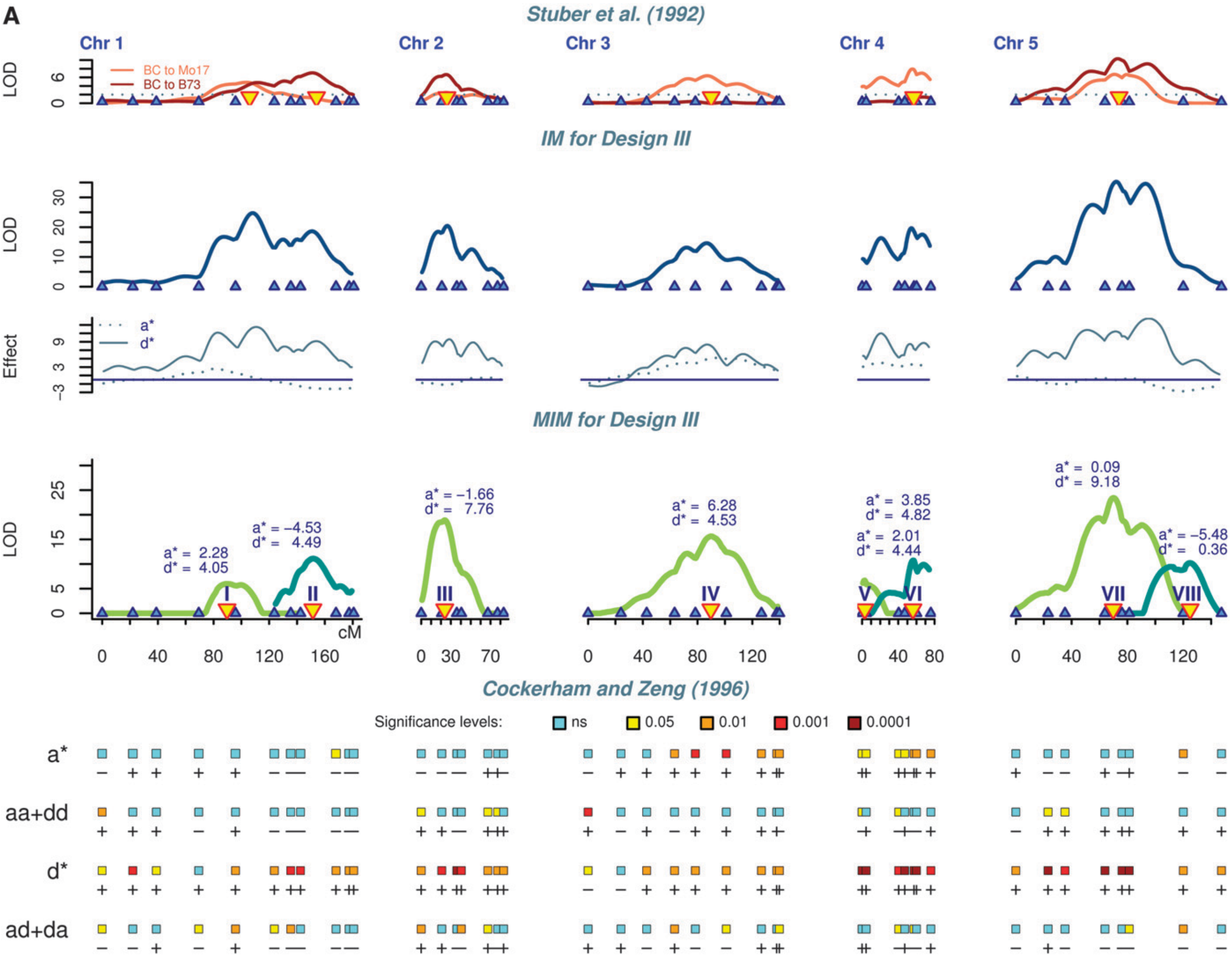

FIGURE 1.—Genetic mapping results of the maize data for grain yield (bushels/acre) (A) for chromosomes 1–5 and (B) for chromosomes 6–10. The results are shown for comparison by using four statistical methods: (1) interval mapping (IM) for each backcross (STUBER *et al.* 1992), with LOD threshold 2 (the identified QTL are indicated by yellow triangles); (2) interval mapping for design III showing augmented additive (a^*) and augmented dominance (d^*) effects; (3) multiple-interval mapping for design III indicating QTL number, effects, and positions; and (4) single-marker analysis of the four contrasts proposed by COCKERHAM and ZENG (1996). Each line corresponds to one contrast with effects indicated on the left. The rectangles correspond to the marker loci and their colors represent the P-values. Plus and minus signs indicate the direction of effects.

results, these estimates are 10.81 and 8.67. For $d^* - a^*$ for QTL II, they found 8.72; the MIM value is 9.02.

The QTL found on chromosomes 1, 2, 3, 7, and 9 are the same ones suggested by Stuber *et al.* The two QTL previously indicated on chromosome 10 are now estimated as a single one. We tried to fit a model with another QTL on this chromosome. There is not enough statistical evidence to support this model. For chromosomes 4, 5, and 8, there is evidence for three additional QTL: one near the beginning of chromosome 4, one at the end of chromosome 5, and one near the beginning of chromosome 8. The presence of QTL at the beginning of chromosome 4 was suggested by IM for design III and with more support from MIM. QTL VII on chromosome 5 has the largest LOD score (23.36) and explains 8.76 and 12.34% of the phenotypic variances in two backcrosses. This indicates the importance of this region and is in agreement with Stuber *et al.*'s results.

On chromosome 8 the two mapped QTL have a^* in opposite signs (repulsion linkage), making their identification difficult by using single-QTL models. QTL X and XI were barely detectable as a single one by Stuber *et al.* with LOD score 2.73. Cockerham and Zeng found P-values of 0.01 in this region only for the contrast for d^*. The two QTL also have smaller LOD scores using MIM for design III (2.48 and 0.89, respectively). However, they were retained in the model, since they were detected to have significant epistatic interaction with other QTL (Table 4).

For epistasis, the final selected model has 14 effects of $aa + dd$ and 8 effects of $ad + da$. Their LOD scores vary from 0.51 to 2.66, generally smaller than the ones for the main effects. Also, they explained individually only a

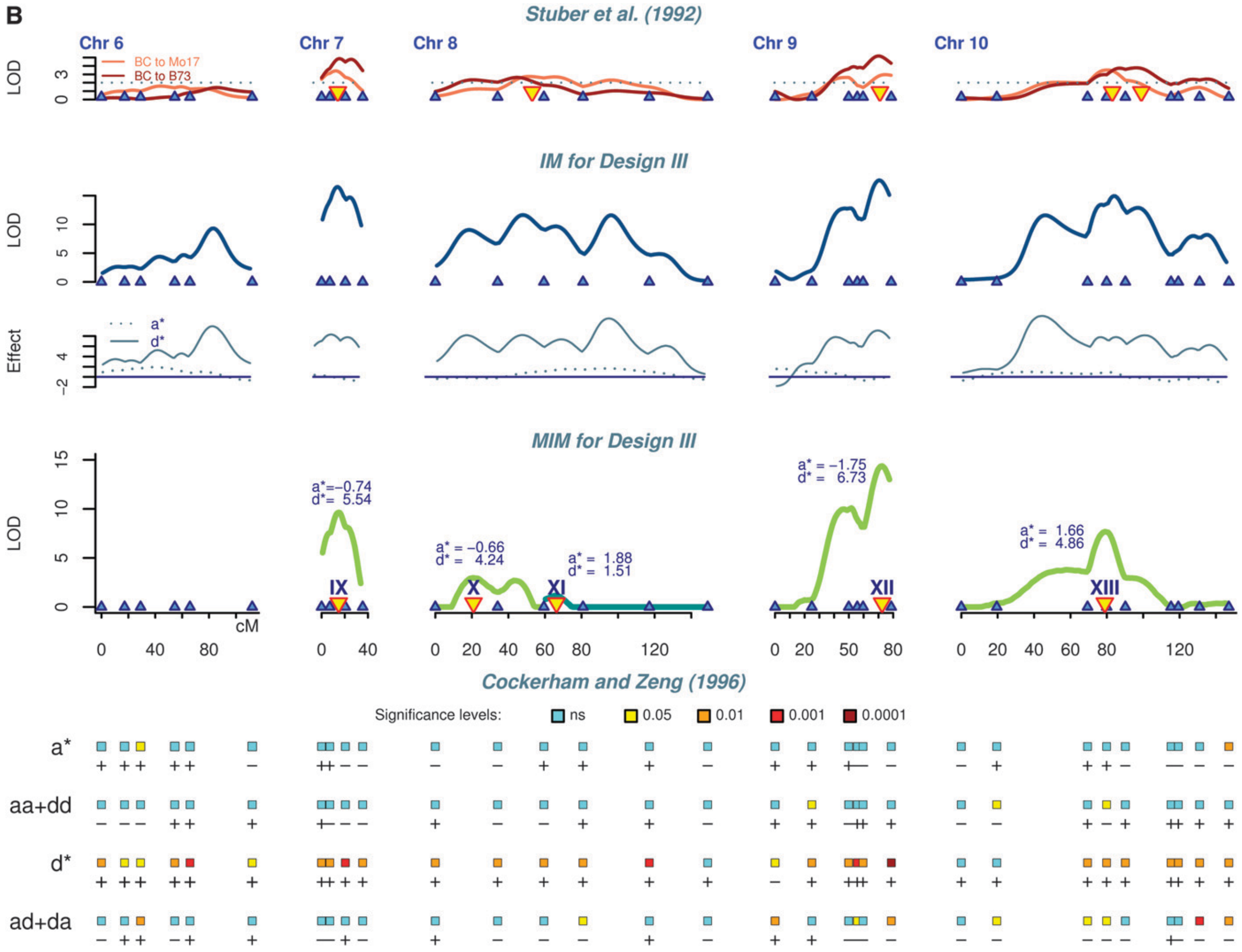

FIGURE 1.—*Continued.*

small fraction of the phenotypic variance (the highest R_j^2 was only 3.47% for $ad + da$ between QTL IX and XI in the backcross to *B73*). Because in design III it is impossible to estimate individual epistatic effects separately, the magnitude of the effects is generally higher than that for a^* and d^* separately, varying from −16.49 to 12.91.

A summary of the final results for the selected model is presented in Table 6. The means of the progenies for the backcross to *Mo17* and *B73* are 86.25 and 90.78 from Cockerham and Zeng, close to the model means 85.52 and 90.59 in Table 6. On the basis of the orthogonal principle for the genetic model used for this study, the difference between the means is an estimate of the sum of additive effects of all potential QTL (WANG and ZENG 2006). For the 13 QTL, $\sum_r a_r^* = 3.23$, which is somewhat close to the observed mean difference (4.53). From the estimates of genetic variance partition in the model, 21.02% is due to α, 59.71% to β, and 19.27% to epistasis (γ and δ).

Discussion: Since MIM for design III tends to provide more appropriate results as compared to other methods, the following discussion is based on this analysis. The signs of a^* effects vary from QTL to QTL, with seven positive (the plus allele from *B73*) and six negative (the plus allele from *Mo17*). The lines *B73* and *Mo17* are elite inbred lines for grain yield and produce a superior hybrid when crossed. These lines, or lines and cultivars derived from them, are widely used for commercial purposes (STUBER *et al.* 1992). We found favorable alleles evenly distributed between the inbred lines. Since the difference $\hat{\mu}_2 - \hat{\mu}_1$ is positive, one would also expect *B73* to have some advantage in terms of a^* effects, and our results corroborate this hypothesis, since $\sum_r a_r^* = 3.23$.

All mapped QTL have d^* with positive sign, meaning that the heterozygous genotype is always superior in the direction of the favorable allele, wherever it is. This is in line with the hypothesis of dominance of favorable alleles as the cause of heterosis in maize. The magnitude of d^* is >2.5 times greater than that of a^* for six QTL (III, VII, IX, X, XII, and XIII). Normally this would be interpreted as evidence of overdominance for these QTL (or some of them). For QTL VII on chromosome 5, further studies

TABLE 4

Estimates of QTL position, effect, LOD score, and coefficient of determination for the maize data using the MIM model for design III

	Position			Effect[a]					
QTL	Chromosome	cM	LOD	a^*	LOD	d^*	LOD	R_1^2 (%)[b]	R_2^2 (%)[b]
I	1	89.7	5.76	2.28	1.82	4.05	4.77	2.42	3.41
II	1	151.4	11.11	−4.53	6.51	4.49	4.82	4.40	6.20
III	2	23.8	18.80	−1.66	1.12	7.76	16.91	6.56	9.25
IV	3	89.7	15.60	6.28	12.45	4.53	6.25	6.22	8.77
V	4	2.9	6.61	2.01	1.56	4.44	5.93	2.47	3.49
VI	4	56.1	10.72	3.85	5.42	4.82	7.05	4.11	5.79
VII	5	69.8	23.36	0.09	0.01	9.18	23.16	8.76	12.34
VIII	5	124.9	10.21	−5.48	9.80	0.36	0.03	3.15	4.44
IX	7	14.8	9.48	−0.74	0.26	5.54	8.48	3.22	4.54
X	8	20.9	2.48	−0.66	0.05	4.24	2.28	1.93	2.73
XI	8	66.3	0.89	1.88	0.63	1.51	0.40	0.61	0.87
XII	9	72.5	14.33	−1.75	1.21	6.73	12.69	5.04	7.11
XIII	10	78.9	7.17	1.66	1.10	4.86	6.78	2.54	3.58

[a] Augmented additive (a^*) and dominance (d^*) effects in bushels/acre.

[b] $R_1^2(\%) = (\hat{\sigma}_r^2/\hat{\sigma}_{P_1}^2) \times 100$ and $R_2^2(\%) = (\hat{\sigma}_r^2/\hat{\sigma}_{P_2}^2) \times 100$ are the fraction of the phenotypic variance in backcrosses to *Mo17* ($\hat{\sigma}_{P_1}^2$) and *B73* ($\hat{\sigma}_{P_2}^2$), respectively, accounted for by each putative QTL *r*.

based on near isogenic lines dissected this QTL into at least two smaller ones, linked in repulsion to each other and with dominant gene action (Graham *et al.* 1997). Pseudo-overdominance, described first by Jones *et al.* (1917) as a possible cause of heterosis, is usually difficult to identify. Graham *et al.*'s result clearly indicates that QTL VII, which has the highest ratio $d^*/|a^*|$, might be due to pseudo-overdominance, rather than overdominance. Without further study it is difficult to know whether this might be also the case for QTL III, IX, X, XII, and XIII, although there is some weak indication for it as the estimates associated with a^* change in sign around those QTL regions by the analysis of Cockerham and Zeng and IM for design III. On the basis of a further study on F_7 parents from the same initial cross, LeDeaux *et al.* (2006) concluded that the genes act predominantly in a dominant manner (not overdominant). Further experiments with larger sample sizes may be required to check if some of those QTL have real overdominance.

Comstock and Robinson (1952) showed that, without epistasis, the average degree of dominance $\bar{D}$ is a weighted average for d effects over r loci with weights a_r^2. From MIM, the estimate of the augmented average degree of dominance is $\bar{D}^* = 3.60$. This value could be interpreted as evidence for overdominance. However, Melchinger *et al.* (2007) discussed in detail that $\bar{D}^*$ is not suitable to provide an accurate estimate of $\bar{D}$, because it is based on a ratio of quadratic forms due to d^* ($\varsigma_{d^*}^2$) and a^* ($\varsigma_{a^*}^2$) effects, being strongly affected by epistasis and the linkage disequilibrium between QTL. In our results, QTL pairs I–II, VII–VIII, and X–XI have a^* effects linked in repulsion, while for pair V–VI they are in coupling. In this situation, the contributions of linked QTL are likely to cancel in $\varsigma_{a^*}^2$. In contrast, $\varsigma_{d^*}^2$ is clearly overestimated since all d^* effects are positive. As a consequence, $\bar{D}^*$ is possibly overestimated.

It can be shown that the midparent heterosis h (considered only up to digenic epistasis) is $h = \sum_r d_r - \frac{1}{2}\sum_{r \neq s} aa_{rs} = \sum_r d_r^*$. Therefore, only negative aa epistasis increases h in addition to dominance effects. Unfortunately, in design III it is impossible to estimate aa effects separately from dd. Because we are estimating sums of $aa + dd$, if they have the same magnitude and opposite signs, the effects will cancel out and epistasis will not be detectable. With opposite signs, the effect can be detected only if one of them is much larger than the other. On the other hand, if they have the same sign, the effects will add up and the interaction can be more easily detected. So, if aa is important for heterosis and most of its effects are negative, one would expect the signs of $aa + dd$ estimates to be predominantly negative, because when dd is positive the effects tend to cancel out and would be more difficult to be detected. From the results, this does not seem to be the case, because there are seven positive and seven negative estimates of $aa + dd$. By these arguments, aa epistasis could be present, but is unlikely to contribute to the observed heterosis significantly in maize. Stuber *et al.* did not find evidence for epistasis, although they used an analysis with low statistical power. Cockerham and Zeng found some evidence for the presence of epistasis in their analysis. Their second and fourth contrasts estimate only a small fraction of linked $aa + dd$ and $ad + da$ epistasis. We found linked QTL on chromosomes 1, 4, 5, and 8, and for them the signs of the contrast for $aa + dd$ were both positive and negative. Therefore, unless most of the

TABLE 5

Estimated epistatic effects between QTL for the maize data

		Effect[a]			
QTL pair	LOD	*aa* + *dd*	*ad* + *da*	R_1^2 (%)[b]	R_2^2 (%)[b]
I, II	1.97	−7.20		0.53	0.74
I, V	1.12	−5.81		0.32	0.45
I, IX	2.66	9.57		0.90	1.27
I, XII	1.37	−6.54		0.38	0.53
II, III	1.36	7.65		0.52	0.74
II, IX	0.88	5.49		0.28	0.39
III, IV	1.50	−7.21		0.47	0.67
III, VI	1.13	−5.38		0.28	0.40
III, VIII	0.51	4.74		0.20	0.28
III, XIII	1.21	−5.72		0.31	0.43
IV, XII	1.28	7.09		0.44	0.62
V, VIII	0.91	−4.92		0.21	0.30
V, X	1.69	8.14		0.59	0.84
VIII, XIII	1.22	6.05		0.35	0.49
V, VIII	0.84		−6.59	0.38	0.54
VI, VII	1.22		−6.85	0.44	0.61
VI, VIII	1.88		8.33	0.70	0.99
VIII, XIII	1.05		6.12	0.36	0.50
IX, X	2.25		12.91	1.61	2.27
IX, XI	2.65		−16.49	2.46	3.47
IX, XII	0.80		4.97	0.24	0.33
X, XIII	0.92		5.63	0.30	0.43

[a] Epistatic effects in bushels/acre.

[b] $R_1^2(\%) = (\hat{\sigma}_r^2/\hat{\sigma}_{P_1}^2) \times 100$ and $R_2^2(\%) = (\hat{\sigma}_r^2/\hat{\sigma}_{P_2}^2) \times 100$ are the fraction of the phenotypic variance in backcrosses to *Mo17* ($\hat{\sigma}_{P_1}^2$) and *B73* ($\hat{\sigma}_{P_2}^2$), respectively, accounted for by each putative QTL epistatic interaction.

negative *aa* effects were canceled out by positive *dd* and not detected (which seems to be unlikely), epistasis is unlikely to be an important explanation for the heterosis in maize.

From the expression of midparent heterosis, the importance of having reliable estimates of d^* becomes evident. The augmented dominance effect d^* measures the net contribution of heterotic QTL to the midparent heterosis. On the basis of the results of QTL mapping, we have $\hat{h} = \sum_r \hat{d}_r^* = 62.51$ bushels/acre [3.92 tons/hectare (t/ha)]. Unfortunately, the inbred lines were not evaluated in the experiments used for the current analysis and so direct heterosis estimates for this data set are not available. James Holland (personal communication) provided some information about heterosis magnitude on the cross *Mo17* × *B73*. On the basis of means over evaluations in two locations near Lafayette, Indiana, in 2003, $\hat{h} = 5.25$ t/ha. The plant density used was 50,000 plants/ha, while Stuber *et al.* used from 36,000 to 50,000 plants/ha. Moreover, the growing conditions in Indiana are not necessarily similar to the ones used in Stuber *et al.*'s study, and some genotype × environment interaction might be expected. In any case, the estimate of heterosis based on MIM results seems to be comparable to the data provided by James Holland.

TABLE 6

Summary of parameter estimation of the MIM model for the maize data

	Backcross to	
	Mo17	*B73*
$\hat{\mu}_j$[a]	85.52	90.59
$\hat{\sigma}_j^2$[b]	44.59	27.44
$\hat{\sigma}_{P_j}^2$[b]	177.65	126.05
$\hat{\sigma}_G^2$[c]	113.20	
$\hat{\sigma}_\alpha^2$	23.80	
$\hat{\sigma}_\beta^2$	67.60	
$\hat{\sigma}_\gamma^2$	10.28	
$\hat{\sigma}_\delta^2$	11.53	
R_j^2 (%)[d]	74.90	78.23

[a] μ_j is mean of the model for backcross j (bushels/acre).

[b] σ_j^2 and $\sigma_{P_j}^2$ are residual and phenotypic variances in (bushels/acre)^2 for backcross j, respectively.

[c] σ_G^2 is variance in (bushels/acre)^2 due to the regression coefficients of the genetic effects in the model that is decomposed in parts due to α, β, γ, and δ.

[d] $R^2(\%) = 100 \times (\hat{\sigma}_{P_j}^2 - \hat{\sigma}_j^2)/\hat{\sigma}_{P_j}^2$ is coefficient of determination.

ANALYSIS OF A RICE DATA SET

Experiment description and statistical analysis: The rice data set was presented in detail in Xiao *et al.* (1995). Briefly, 194 F_7 parents were backcrossed to two elite homozygous lines, 9024 (L_1, *indica* parent) and LH422 (L_2, *japonica* parent). The backcross progenies were evaluated in a randomized complete block design with two replications. Twelve quantitative traits were measured, but we used just means over replications for grain yield (in tons/hectare). A genetic map for the recombinant inbred population was constructed with 141 RFLP markers and the genetic distances were expressed in centimorgans using the Kosambi map function.

To help in the selection of the final MIM model, the same procedures used for the maize data were applied. Initially, IM for design III was applied. Then, a MIM model for design III was selected. First a forward procedure was used until no more QTL could be added. Second, a forward/backward procedure was applied to find two-way epistasis between QTL. Models were compared using the BIC for the main effects and the AIC for epistatic effects. The positions were refined in every step of model updating. Finally, we also estimated the four contrasts proposed by Cockerham and Zeng for all markers. For epistasis, some markers did not have heterozygous genotypes and therefore the contrasts could not be estimated.

Results: *IM for design III:* The results for QTL mapping for grain yield are presented in Figure 2, A and B.

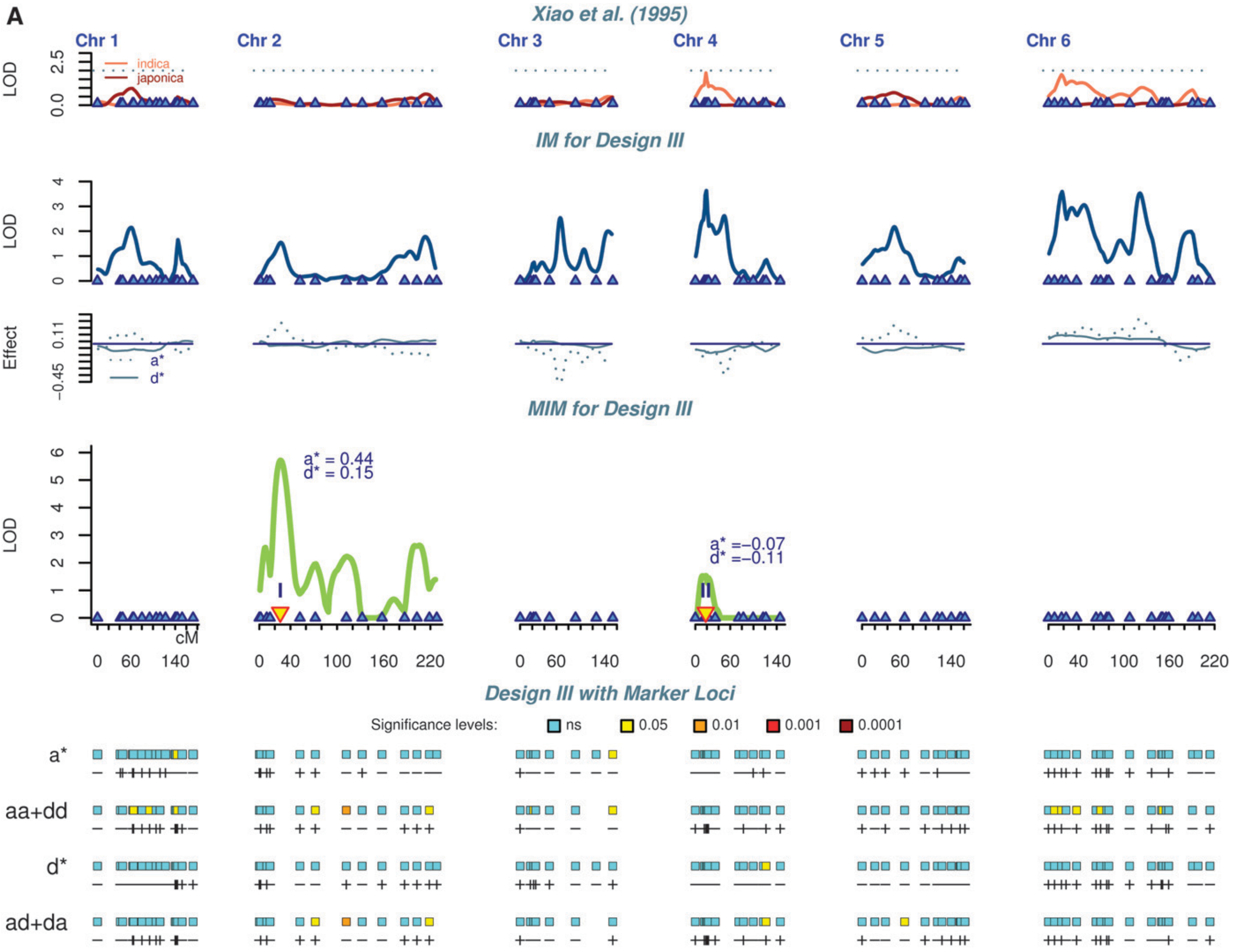

Figure 2.—Genetic mapping results of the rice data for grain yield (tons/hectare) (A) for chromosomes 1–6 and (B) for chromosomes 7–12. The results are shown for comparison by using four statistical methods: (1) interval mapping (IM) for each backcross (Xiao *et al.* 1995), with LOD threshold 2 (the identified QTL are indicated by yellow triangles); (2) interval mapping for design III showing augmented additive (a^*) and augmented dominance (d^*) effects; (3) multiple-interval mapping for design III indicating estimated QTL number, effect (tons/hectare), and position; and (4) single-marker analysis of the four contrasts proposed by Cockerham and Zeng (1996). Each line corresponds to one contrast whose effects are indicated on the left. The rectangles correspond to the marker loci with colors representing the *P*-values. Plus and minus signs indicate the direction of effects. Missing rectangles for epistasis are due to lack of heterozygous marker genotypes.

In the same way as for the maize data, they are in agreement with the analysis of Xiao *et al.*, but provide more information and statistical power. Xiao *et al.* did their analysis in a way similar to Stuber *et al.*, considering the backcrosses separately. They found only two QTL, one in the backcross to *japonica* on chromosome 8 (with LOD score 2.49), and another one in the backcross to *indica* on chromosome 11 (with LOD score 2.64). Using IM for design III there are LOD peaks in the same regions, but with higher LOD scores (∼4.5). Moreover, there is an indication of additional QTL in many other chromosomes.

In general, the LOD curves from Xiao *et al.* are flat and with small values. When the analysis is done for both backcrosses simultaneously, some peaks become more evident, such as on chromosomes 2, 3, 5, and 11. The QTL on chromosome 4, that had previously a LOD score <2 and thus was not selected, now has a more identifiable peak with LOD score ∼4. At the beginning of chromosome 11 there is strong evidence for the presence of a QTL, showing that the new analysis can significantly increase the ability for the identification of QTL. In fact, this QTL is the most important one in the MIM model (next section).

For the same reasons as discussed above for the maize data, Xiao *et al.* also estimated $d^* + a^*$ and $d^* - a^*$, leading to the identification of QTL in only one backcross if the effects are similar in magnitude. With the combined analysis, a^* and d^* could be estimated separately. The *P*-values for the contrasts of Cockerham and Zeng were not significant for all markers, with only few exceptions that are possibly false positives. None of

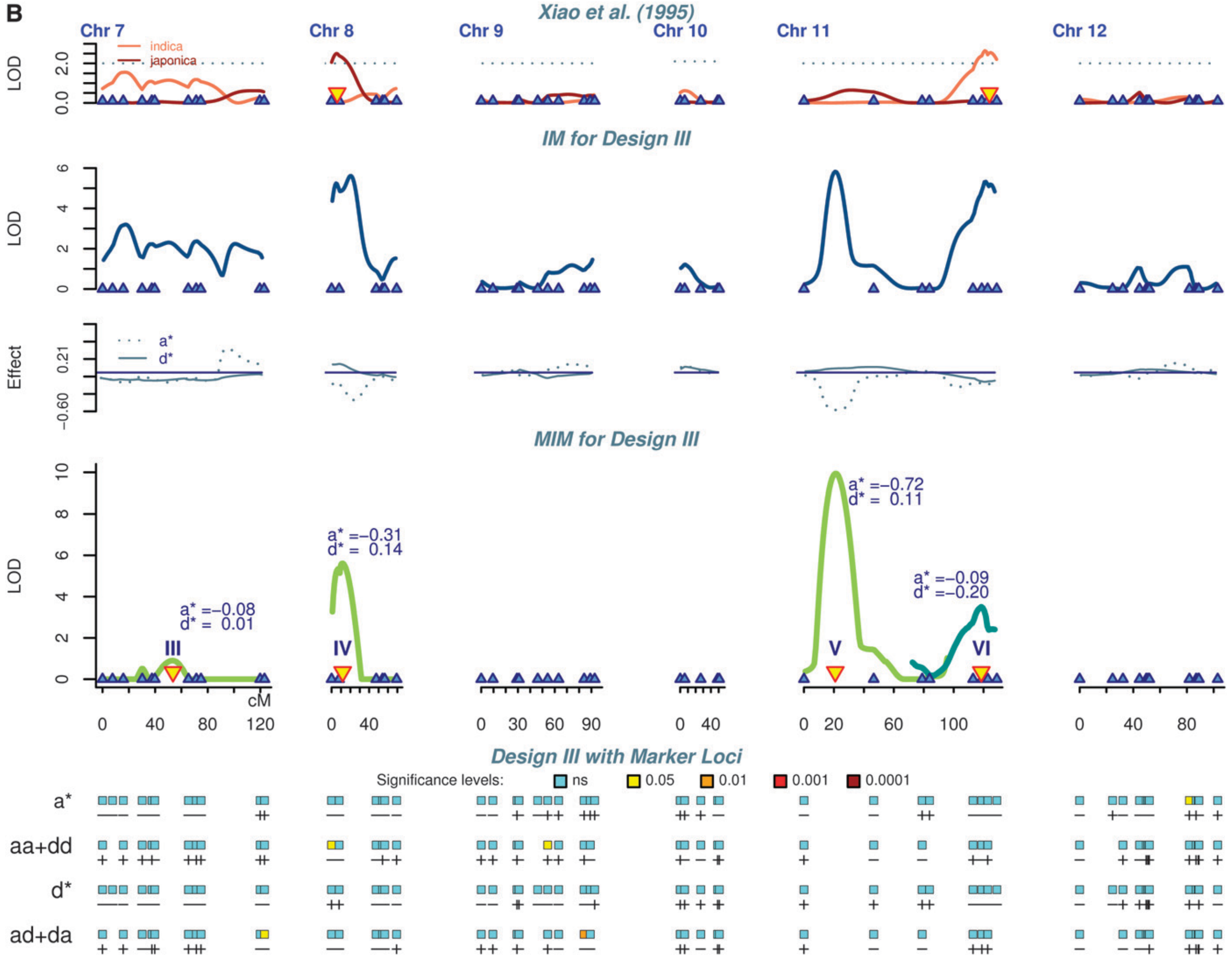

FIGURE 2.—*Contiuned.*

the *P*-values is <0.01. The signs of the contrasts are in agreement with the estimates from IM for design III. In contrast to the results for maize data, now *d** effects are positive and negative for approximately the same number of regions.

MIM for design III: Six QTL were mapped on chromosomes 2, 4, 7, 8, and 11, with LOD scores varying from 0.40 to 9.43 (Figure 2, A and B, Tables 7–9). QTL II and III were retained in the model because they had significant epistasis with another QTL. Not all putative QTL suggested by IM were kept in the final MIM model, since they were not significant. This is the case for putative QTL on chromosomes 1, 5, and 6 and also for the one near the end of chromosome 2. Only chromosome 11 has more than one QTL, but they are very far apart (>90 cM).

Surprisingly, QTL V at the beginning of chromosome 11 was not detected by Xiao *et al.*, having just a slight tendency for its presence in the backcross to *japonica*. However, it has the highest LOD and R^2 in our analysis. Its presence is also suggested by IM for design III. This is an indication that the analysis of the combined backcross has more statistical power and can lead to different results.

Together, all QTL explain 60.94 and 64.67% of the phenotypic variation in the backcrosses to *indica* and *japonica*, respectively. In their analysis, Xiao *et al.* found only two QTL (named IV and VI in our results), explaining 6.80 and 6.30% of the phenotypic variation. In our analysis, the main effects of QTL have R^2's varying from 0.34 to 31.13%. Four *aa* + *dd* and five *ad* + *da* epistasis effects were selected, with small LOD scores. For the estimated genetic variance, 74.29% is due to additive effects of QTL, 9.52% is due to dominance effects, and 16.19% is due to epistatic effects. In contrast to the maize results, *a** effects seem to be more important for rice.

The signs of *a** are negative for all QTL (except QTL I), showing that the favorable alleles are concentrated in *indica*. Their values vary from −0.723 to 0.442 (t/ha).

TABLE 7

Estimated QTL position, effect, LOD score, and variance component for the rice data using the MIM model for design III

	Position			Effect[a]					
QTL	Chromosome	cM	LOD	a^*	LOD	d^*	LOD	R_1^2 (%)[b]	R_2^2 (%)[b]
I	2	32.9	5.16	0.442	4.86	0.151	0.79	12.09	12.83
II	4	17.9	1.53	−0.067	0.22	−0.114	1.39	0.99	1.05
III	7	28.8	0.40	−0.081	0.34	0.011	0.01	0.34	0.36
IV	8	5.9	5.28	−0.312	3.58	0.141	1.52	5.69	6.04
V	11	24.9	9.43	−0.723	8.89	0.111	0.83	29.33	31.13
VI	11	115.7	3.29	−0.093	0.52	−0.196	2.96	2.63	2.79

[a] Augmented additive (a^*) and dominance (d^*) effects in tons/hectare.

[b] $R_1^2(\%) = (\hat{\sigma}_r^2/\hat{\sigma}_{\mathrm{P}_1}^2) \times 100$ and $R_2^2(\%) = (\hat{\sigma}_r^2/\hat{\sigma}_{\mathrm{P}_2}^2) \times 100$ are the fraction of the phenotypic variance in backcrosses to *indica* ($\hat{\sigma}_{\mathrm{P}_1}^2$) and *japonica* ($\hat{\sigma}_{\mathrm{P}_2}^2$), respectively, accounted for by each putative QTL.

Significantly different from maize, d^* effects are both positive (for four QTL) and negative (for two QTL) and are in general smaller than a^* in magnitude. No evidence for overdominance of any QTL is observed.

Discussion: Again, the following discussion is based on the results of MIM for design III. The a^* effect is positive for one QTL and negative for the other five, showing that the favorable alleles are distributed between the parents but with concentration in the *indica* parent. In contrast to maize, d^* estimates are now positive and negative, indicating that the heterozygote is not always superior in the direction of the favorable allele. This is not in line with the hypothesis that dominance is a major cause of heterosis in rice.

For rice, d^* effects are not significantly greater than a^* effects for any QTL. This can be interpreted as lack of overdominance (or pseudo-overdominance). Actually, from our results, $\hat{\bar{D}}^* = 0.12$, corroborating the importance of a^* effects for grain yield in rice. Even knowing that $\bar{D}^*$ can be strongly biased, one would expect this to occur in a smaller magnitude in this case, since there is no evidence for closely linked QTL (the only two QTL on the same chromosome are very far apart). Therefore, the bias due to *aa* and *da* effects contained in a^* and d^* and the overestimation that happened for $\bar{D}^*$ in maize is not expected here.

Xiao *et al.* concluded that dominance is the major genetic basis of heterosis in rice. In the same way as Stuber *et al.*, they used the difference between the phenotypic means of heterozygous and homozygous genotypes in each backcross as an estimate of the phenotypic effect of QTL. They found one positive and one negative result for these differences for the two QTL for grain yield. Since positive and negative signs indicate superior heterozygous and homozygous genotypes, respectively, they assumed lack of overdominance and concluded that dominance (or partial dominance) is the major contributor to F_1 heterosis. Probably, their conclusions were reinforced by the fact that they did not find significant epistasis. However, using differences on each backcross they were actually estimating $d^* + a^*$ and $d^* - a^*$ in the backcross to *indica* and *japonica*, respectively. Our estimates for $d^* + a^*$ and $d^* - a^*$ for QTL IV and V are, respectively, −0.171 and 0.834, with the same signs as the Xiao *et al.* estimates, showing that positive and negative estimates can appear, but are not necessarily evidence of dominance (or partial dominance) as a major cause for heterosis.

Since rice is a self-pollinated species, it is common to express heterosis also in terms of the difference between F_1 and the better parent (also called heterobeltiosis, H). Xiao *et al.* estimated heterobeltiosis $\hat{H} = 1.35$ t/ha. Melchinger *et al.* showed that $H = \sum_r (d_r^* - a_r^*)$. From the MIM results, $\hat{H} = 0.938$ t/ha, close to the observed heterosis. However, when considering the midparent heterosis h, we get from the MIM results $\hat{h} = \sum_r \hat{d}_r^* = 0.104$ t/ha, while Xiao *et al.*'s value is 1.605 t/ha, >15 times greater. One possible explanation for this difference is the presence of epistasis. As pointed out above, if *aa* is a cause for the midparent heterosis, its signs will be

TABLE 8

Estimated epistatic effect, LOD score, and variance component between QTL for the rice data

		Effect[a]			
QTL pair	LOD	*aa* + *dd*	*ad* + *da*	R_1^2 (%)[b]	R_2^2 (%)[b]
I, IV	1.14	−0.325		1.46	1.55
I, VI	0.74	−0.264		0.97	1.03
II, IV	1.53	−0.356		1.78	1.88
III, V	0.83	0.226		0.70	0.74
I, IV	1.04		−0.327	1.48	1.57
I, V	0.06		0.079	0.09	0.10
I, VI	0.86		0.267	0.99	1.05
III, IV	2.41		−0.358	1.80	1.90
IV, VI	0.88		0.207	0.61	0.64

[a] Epistatic effects in tons/hectare.

[b] $R_1^2(\%) = (\hat{\sigma}_r^2/\hat{\sigma}_{\mathrm{P}_1}^2) \times 100$ and $R_2^2(\%) = (\hat{\sigma}_r^2/\hat{\sigma}_{\mathrm{P}_2}^2) \times 100$ are the fraction of the phenotypic variance in backcrosses to *indica* ($\hat{\sigma}_{\mathrm{P}_1}^2$) and *japonica* ($\hat{\sigma}_{\mathrm{P}_2}^2$), respectively, accounted for by each QTL pair epistatic interaction.

TABLE 9

Parameter estimates of the MIM model for the rice data

	Backcross to		
	indica		*japonica*
$\hat{\mu}_j$[a]	6.17		6.31
$\hat{\sigma}_j^2$[b]	0.1738		0.1481
$\hat{\sigma}_{P_j}^2$[b]	0.4449		0.4192
$\hat{\sigma}_G^2$[c]		0.2711	
$\hat{\sigma}_\alpha^2$		0.2014	
$\hat{\sigma}_\beta^2$		0.0258	
$\hat{\sigma}_\gamma^2$		0.0218	
$\hat{\sigma}_\delta^2$		0.0221	
R^2 (%)[d]	60.94		64.67

[a] μ_j is mean of the model for backcross j (tons/hectare).

[b] σ_j^2 and $\sigma_{P_j}^2$ are residual and phenotypic variances in (tons/hectare)^2 for backcross j, respectively.

[c] σ_G^2 is variance in (tons/hectare)^2 explained by the regression coefficients of the genetic effects in the model and decomposed in parts due to α, β, γ, and δ.

[d] $R^2(\%) = 100 \times (\hat{\sigma}_{P_j}^2 - \hat{\sigma}_j^2)/\hat{\sigma}_{P_j}^2$ is coefficient of determination.

predominantly negative. But if *d* signs vary from locus to locus, d^* signs will tend to be positive and negative and therefore will tend to cancel each other out when added in *h*. Our estimates of $aa + dd$ showed three negative signs and one positive sign. This could be an indication of a tendency of *aa* to be predominantly negative and therefore potentially important as a cause for the midparent heterosis in rice. In addition to the facts that normally epistasis is difficult to detect and design III is also not suitable to estimate epistatic effects separately, the progeny data used in this research were evaluated in only one location and year, with few replications. So, it may be expected that the means used in the analysis were not estimated with good precision. Therefore, this tendency for the presence of negative *aa* epistasis as a cause for heterosis needs to be confirmed in further studies.

CONCLUSIONS

The objective of this research is to study the genetic basis of heterosis in maize and rice. Since maize and rice are economically important and are good examples of outcrossing and self-pollinating crops, we believe that the conclusions from this study may be useful for plant breeders and geneticists. To achieve this goal, we first extended the single-marker contrasts proposed by Cockerham and Zeng for the analysis of design III to two markers. On the basis of the genetic expectations of contrasts for the analysis of two markers simultaneously, we were able to propose a new model for a statistical analysis of design III, taking into account positions between markers. This leads to the MIM model for design III that provides a basis to estimate QTL number, positions, effects (a^* and d^*), and epistatic interactions ($aa + dd$ and $ad + da$) simultaneously. Our model can be used for parents with any number of generations in selfing.

After Stuber *et al.* and Cockerham and Zeng, a few authors also proposed methods for QTL mapping and analysis of design III, most of them based on the derivations of Cockerham and Zeng showing that the contrasts of heterozygous and homozygous genotypes on each backcross actually test $d^* + a^*$ and $d^* - a^*$. For example, Lu *et al.* (2003) and Ledeaux *et al.* (2006) proposed the utilization of composite-interval mapping (CIM) (Zeng 1994) on each backcross separately and, after QTL were mapped (in one or both backcrosses), a^* and d^* effects were estimated by a linear combination of the contrasts for each backcross. Although a^* and d^* effects can be estimated individually in this way, the results of QTL mapping are still based on the analysis of each backcross separately in a similar way to that of Stuber *et al.* Lu *et al.* proposed to test epistasis by fitting a two-locus linear regression model for the main effects and interaction between loci. If performed in this way, it is likely that epistasis will be rarely identified because the test tends to have relatively low statistical power and, even if identified, it is not clear how to interpret the results in a way to understand its influence on heterosis. In a different approach, Melchinger *et al.* (2007) suggested the use of CIM for the identification of genomic regions affecting heterosis. They defined two orthogonal single-marker contrasts based on progeny mean values for pair means and pair differences. These contrasts, which correspond to contrasts C_1 and C_3 of Cockerham and Zeng, and x^*_{ijr} and z^*_{ijr} in our MIM model, are used individually for CIM analysis of the combined backcrosses and the estimation of a^* and d^*. Although using information from both crosses simultaneously, their method is still based on CIM and does not capitalize on all the advantages of MIM models. To our knowledge, the proposed MIM model for design III is probably the most powerful statistical method for QTL mapping in this type of population currently. We developed a module of MIM for design III for Windows QTL Cartographer (Wang *et al.* 2007) specifically for its public use. The software can be freely downloaded from http://statgen.ncsu.edu/qtlcart/WQTLCart.htm.

We realize that by using AIC as a criterion for including epistasis in the MIM model, there is a risk that the final model may be overfitted. However, this was done mostly to study the sign of estimates for epistasis. Normally, epistasis is difficult to detect with statistical significance, and both Stuber *et al.* and Xiao *et al.* did not find evidence for it using statistical tests with relatively low statistical power. Since our model allows the inclusion of epistasis, it is possible to study its effects more clearly on maize and rice. The results showed that dominance is possibly a major cause of heterosis in maize, although overdominance (or pseudo-overdominance) of individual loci could not be ruled out. On the other hand, for rice there is evidence that additive × additive epistasis could be important for explaining heterosis. Maize and rice evolved from a common ancestor (Ahn and Tanksley 1993) but have different reproductive biology. As a consequence, maize is supposed to have more deleterious recessive alleles than rice, masked by their corresponding dominant counterparts. When inbreeding occurs, these unfavorable alleles are expressed in the homozygous loci, causing the inbreeding depression. In self-pollinating species, deleterious alleles are possibly eliminated by natural (and artificial) selection since the individuals are homozygous. Therefore, outcrossing species could be selected for true dominant loci to avoid the expression of these deleterious loci (causing the outbreeding advantage), whereas in self-pollinating species the selection for dominance is less important and, when an F_1 cross shows midparent heterosis, it is more likely due to epistatic interactions (*aa*) among loci.

Two important conferences about heterosis should be mentioned. In 1950, in Iowa, there was a 5-week conference (Gowen 1952). At that occasion, Comstock and Robinson (1952) proposed design III as a means to estimate the average degree of dominance and also presented some estimates, suggesting overdominance. Some authors proposed breeding schemes to exploit it. Since then, design III has been widely used in breeding programs over the years for understanding the genetic basis of many economically important traits and for developing breeding schemes. Crow (1999, p. 521) said that "1950 and the next few years was the zenith of overdominance," but in later years the importance of the dominance hypothesis increased. When comparing this conference with another one that took place in 1997 in Mexico City, Crow (1999) noted a change in emphasis, since in the second one many authors included epistasis in their presentations. We hope that the results presented here can make a contribution to this important discussion.

The authors thank Charles Stuber (North Carolina State University) and Steven Tanksley (Cornell University) for providing the maize and rice data, respectively. This research was done while A. A. F. Garcia was working with Z.-B.Z. as a visiting scientist (postdoc) at the Bioinformatics Research Center, North Carolina State University, with a fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil (grant no. 200345/2004-4). Z.-B.Z. was also partially supported by National Institutes of Health grant GM45344 and by the National Research Initiative of the U.S. Department of Agriculture Cooperative State Research, Education and Extension Service, grant no. 2005-00754. A.E.M. was supported by grants from the German Research Foundation (ME931/4-1 and ME937/4-2).

LITERATURE CITED

Ahn, S., and S. D. Tanksley, 1993 Comparative linkage maps of the rice and maize genomes. Proc. Natl. Acad. Sci. USA **90:** 7980–7984.

Akaike, H., 1974 A new look at the statistical model identification. IEEE Trans. Automat. Contr. **19:** 716–723.

Bruce, A. B., 1910 The Mendelian theory of heredity and the augmentation of vigor. Science **32:** 627–628.

Cockerham, C. C., and Z.-B. Zeng, 1996 Design III with marker loci. Genetics **143:** 1437–1456.

Comstock, R. H., and H. F. Robinson, 1948 The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance. Biometrics **4:** 254–266.

Comstock, R. H., and H. F. Robinson, 1952 Estimation of average dominance of genes, pp. 495–516 in *Heterosis*, edited by J. W. Gowen. Iowa State College Press, Ames, IA.

Crow, J. F., 1999 A symposium overview, pp. 521–524 in *The Genetic and Exploitation of Heterosis in Crops*, edited by J. G. Coors and S. Pandey. American Society of Agronomy, Madison, WI.

Davenport, C. B., 1908 Degeneration, albinism and inbreeding. Science **28:** 454–455.

Dempster, A. P., N. M. Laird and D. B. Rubin, 1977 Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. **39:** 1–38.

East, E. M., 1908 Inbreeding in corn. Rep. Conn. Agric. Exp. Stn. **1907:** 419–428.

Gowen, J. W., Editor, 1952 *Heterosis.* Iowa State College Press, Ames, IA.

Graham, G. I., D. W. Wolff and C. W. Stuber, 1997 Characterization of a yield quantitative trait locus on chromosome five of maize by fine mapping. Crop Sci. **37:** 1601–1610.

Jones, D. F., 1917 Dominance of linked factors as a means of accounting for heterosis. Genetics **2:** 466–479.

Kao, C.-H., and Z.-B. Zeng, 1997 General formulas for obtaining the MLEs and the asymptotic variance-covariance matrix in mapping quantitative trait loci when using the EM algorithm. Biometrics **53:** 653–665.

Kao, C.-H., Z.-B. Zeng and R. D. Teasdale, 1999 Multiple interval mapping for quantitative trait loci. Genetics **152:** 1203–1216.

Keeble, F., and C. Pellew, 1910 The mode of inheritance of stature and of time of flowering in peas (*Pisum sativum*). J. Genet. **1:** 47–56.

Lander, E. S., and D. Botstein, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121:** 185–199.

Lander, E. S., P. Green, J. Abrahamson, A. Barlow, M. J. Daly *et al.*, 1987 Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics **1:** 174–181.

LeDeaux, J. R., G. I. Graham and C. W. Stuber, 2006 Stability of QTL involved in heterosis in maize when mapped under several stress conditions. Maydica **51:** 151–167.

Lu, H., J. Romero-Severson and R. Bernardo, 2003 Genetic basis of heterosis explored by simple sequence repeat markers in a random-mated maize population. Theor. Appl. Genet. **107:** 494–502.

Melchinger, A. E., H. F. Utz, H. P. Piepho, Z.-B. Zeng and C. C. Schön, 2007 Quantitative genetic theory to elucidate the role of epistasis in the manifestation of heterosis. Genetics **117:** 1815–1825.

Schwarz, G., 1978 Estimating the dimension of a model. Ann. Stat. **6:** 461–464.

Shull, G. H., 1908 The composition of a field of maize. Am. Breeders Assoc. Rep. **4:** 296–301.

Stuber, G. W., S. E. Lincoln, D. W. Wolff, T. Helentjaris and E. S. Lander, 1992 Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics **132:** 823–839.

Wang, S., C. J. Basten and Z.-B. Zeng, 2007 *Windows QTL Cartographer 2.5.* Department of Statistics, North Carolina State University, Raleigh, NC. http://statgen.ncsu.edu/qtlcart/WQTLCart.htm.

Wang, T., and Z.-B. Zeng, 2006 Models and partition of variance for quantitative trait loci with epistasis and linkage disequilibrium. BMC Genet. **7:** 9.

Xiao, J., J. Li, L. Yuan and S. D. Tanksley, 1995 Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. Genetics **140:** 745–754.

Yuan, L. P., 1992 Development and prospects of hybrid rice breeding, pp. 97–105 in *Agricultural Biotechnology, Proceeding of Asian-Pacific Conference on Agricultural Biotechnology,* edited by C. B. You and Z. L. Chen. China Agriculture Press, Beijing.

Zeng, Z.-B., 1994 Precision mapping of quantitative trait loci. Genetics **136:** 1457–1468.

Zeng, Z.-B., C.-H. Kao and C. J. Basten, 1999 Estimating the genetic architecture of quantitative traits. Genet. Res. **74:** 279–289.

Zeng, Z.-B., T. Wang and W. Zou, 2005 Modeling quantitative trait loci and interpretation of models. Genetics **169:** 1711–1725.

Communicating editor: E. S. Buckler

APPENDIX A: GENOTYPIC CONSTITUTION OF THE PROGENIES FROM F_2 PARENTS

Here we expand the idea of Cockerham and Zeng (1996) and consider F_2 parents for two linked markers (M_1 and M_2) with recombination fraction ρ. The markers are linked to two QTL with the linkage order $Q_1M_1M_2Q_2$. The recombination fraction between Q_1 and M_1 is ρ_1, between M_2 and Q_2 is ρ_2, and between Q_1 and Q_2 is ρ_{12}. We assume no crossover interference, so $\rho_{12} = \rho_1(1-\rho)(1-\rho_2) + (1-\rho_1)\rho(1-\rho_2) + (1-\rho_1)(1-\rho)\rho_2 + \rho_1\rho\rho_2$. Assume that the inbred lines' genotypes are $L_2 = Q_1Q_1M_1M_1M_2M_2Q_2Q_2$ and $L_1 = q_1q_1m_1m_1m_2m_2q_2q_2$.

Denote F_1 gametes as

$$g' = \left[\, g'_{M_1M_2},\, g'_{M_1m_2},\, g'_{m_1M_2},\, g'_{m_1m_2} \right]$$

with

$$\begin{aligned}
g'_{M_1M_2} &= [\, Q_1M_1M_2Q_2, \quad Q_1M_1M_2q_2, \quad q_1M_1M_2Q_2, \quad q_1M_1M_2q_2 \,] \\
g'_{M_1m_2} &= [\, Q_1M_1m_2Q_2, \quad Q_1M_1m_2q_2, \quad q_1M_1m_2Q_2, \quad q_1M_1m_2q_2 \,] \\
g'_{m_1M_2} &= [\, Q_1m_1M_2Q_2, \quad Q_1m_1M_2q_2, \quad q_1m_1M_2Q_2, \quad q_1m_1M_2q_2 \,] \\
g'_{m_1m_2} &= [\, Q_1m_1m_2Q_2, \quad Q_1m_1m_2q_2, \quad q_1m_1m_2Q_2, \quad q_1m_1m_2q_2 \,].
\end{aligned}$$

The gametic frequencies are one-half of

$$\begin{aligned}
f'_{M_1M_2} &= [\,(1-\rho_1)(1-\rho)(1-\rho_2), \quad (1-\rho_1)(1-\rho)\rho_2, \quad \rho_1(1-\rho)(1-\rho_2), \quad \rho_1(1-\rho)\rho_2 \,] \\
f'_{M_1m_2} &= [\,(1-\rho_1)\rho\rho_2, \quad (1-\rho_1)\rho(1-\rho_2), \quad \rho_1\rho\rho_2, \quad \rho_1\rho(1-\rho_2) \,] \\
f'_{m_1M_2} &= [\,\rho_1\rho(1-\rho_2), \quad \rho_1\rho\rho_2, \quad (1-\rho_1)\rho(1-\rho_2), \quad (1-\rho_1)\rho\rho_2 \,] \\
f'_{m_1m_2} &= [\,\rho_1(1-\rho)\rho_2, \quad \rho_1(1-\rho)(1-\rho_2), \quad (1-\rho_1)(1-\rho)\rho_2, \quad (1-\rho_1)(1-\rho)(1-\rho_2) \,].
\end{aligned}$$

From these frequencies, it is easy to show the conditional frequencies of QTL gametes from F_2 with different marker genotypes (Table 1). These gametes are combined with the gametes Q_1Q_2 and q_1q_2 from inbred lines L_2 and L_1, respectively, to form two backcross populations.

Let H_g^j denote the genotypic means of backcross progenies with g marker genotype in the F_2 parent backcrossed to parental line j. There are 18 H_g^j values. They are weighted genotypic values of seven QTL genotypes (the nine possible genotypes at two loci of minor genotypes Q_1q_2/Q_1q_2 and q_1Q_2/q_1Q_2, which are not produced in the backcrosses) with weights given in Table 1.

APPENDIX B: ORTHOGONAL CONTRASTS WITH TWO MARKERS

When two markers are considered simultaneously in the two backcrosses of design III, it is possible to define a set of 17 orthogonal contrasts denoted as c_k ($k = 1, \ldots, 17$) (Table 3). Denoting the coefficients in Table 3 as u_{kgj}, the kth contrast is $c_k = \sum_g \sum_j u_{kgj} H_g^j$. All contrasts are orthogonal because $\sum_g \sum_j u_{kgj} u_{k'gj} = 0$ for any pair of contrasts c_k and c'_k $(k \neq k')$.

Contrasts c_1–c_4 are for marginal differences among means for marker genotypes of M_1 (c_1 and c_2) and M_2 (c_3 and c_4) and can be viewed as a direct expansion of the first and third contrasts of Cockerham and Zeng. Contrasts c_1 and c_3 are for differences between homozygous marker genotypes for M_1 and M_2, respectively, and c_2 and c_4 are for contrasts between heterozygous and homozygous marker genotypes. The contrasts c_5–c_8 are for interactions between c_1 and c_3, c_1 and c_4, c_2 and c_3, and c_2 and c_4, respectively. Contrast c_9 is for testing the difference between the inbred lines (not considered by Cockerham and Zeng) and c_{10}–c_{17} are for interactions of contrasts c_1–c_8 with the inbred lines (analogous to contrasts 2 and 4 of Cockerham and Zeng).

TABLE A1

Orthogonal contrasts for the analysis of design III with recombinant inbred lines

Contrast	H_{22}^2	H_{20}^2	H_{02}^2	H_{00}^2
$\ddot{C}_1$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{-1}{4}$	$\frac{-1}{4}$
$\ddot{C}_3$	$\frac{1}{4}$	$\frac{-1}{4}$	$\frac{1}{4}$	$\frac{-1}{4}$
$\ddot{C}_5$	$\frac{1}{2}$	$\frac{-1}{2}$	$\frac{-1}{2}$	$\frac{1}{2}$

H_g^j is the genotypic mean of the backcross progenies from F_∞ parents with marker genotype g backcrossed to parental line j ($j=2, 1$). Only H_g^2 means are presented, since the coefficients for H_g^1 are the same (for a given g) for $\ddot{C}_1$, $\ddot{C}_3$, and $\ddot{C}_5$. Contrasts $\ddot{C}_2$, $\ddot{C}_4$, and $\ddot{C}_6$ have the same coefficients as $\ddot{C}_1$, $\ddot{C}_3$, and $\ddot{C}_5$ for H_g^1, respectively; for H_g^2, the coefficients are the same but with opposite signs.

On the basis of the genotypic constitution of the progenies of F_2 parents (Table 1 and APPENDIX A) and substituting the genotypic values by the genetic effects based on the F_2 genetic model (COCKERHAM and ZENG 1996; ZENG *et al.* 2005), we derived the genetic expectation of the 17 contrasts:

$$\begin{aligned}
E(c_1) &= 6(1-2\rho_1)a_1 - 3(1-2\rho_1)da \\
E(c_2) &= E(c_4) = -\frac{1}{2}E(c_8) = -\frac{(1-2\rho_1)^2(1-2\rho)^2(1-2\rho_2)^2}{1-2\rho+2\rho^2}(aa+dd) \\
E(c_3) &= 6(1-2\rho_2)a_2 - 3(1-2\rho_2)ad \\
E(c_5) &= 2(1-2\rho_1)(1-2\rho_2)(aa+dd) \\
E(c_6) &= E(c_7) = E(c_{15}) = E(c_{16}) = 0 \\
E(c_9) &= -9(a_1+a_2) + \frac{(1-2\rho_1)^2(1-2\rho)^2(1-2\rho_2)^2}{2(1-2\rho+2\rho^2)}(ad+da) \\
E(c_{10}) &= 6(1-2\rho_1)d_1 - 3(1-2\rho_1)aa \\
E(c_{11}) &= E(c_{13}) = -\frac{1}{2}E(c_{17}) = -\frac{(1-2\rho_1)^2(1-2\rho)^2(1-2\rho_2)^2}{1-2\rho+2\rho^2}(ad+da) \\
E(c_{12}) &= 6(1-2\rho_2)d_2 - 3(1-2\rho_2)aa \\
E(c_{14}) &= 2(1-2\rho_1)(1-2\rho_2)(ad+da).
\end{aligned}$$

APPENDIX C: DESIGN III WITH RECOMBINANT INBRED LINES

If we continue selfing F_2 for a number of generations, it will lead to the development of recombinant inbred lines (F_∞) where heterozygote genotypes are eliminated. There are four homozygote genotypes for two loci in the recombinant inbred lines and eight genotypic means in the two backcrosses. The six contrasts can be further simplified from Table 2 and are presented in Table A1.

The genotypic expectations of the contrasts in the framework of MIM can be expressed for two QTL as

$$\begin{aligned}
E(\ddot{C}_1) &= a_1 - \frac{1}{2}da \\
E(\ddot{C}_2) &= d_1 - \frac{1}{2}aa \\
E(\ddot{C}_3) &= a_2 - \frac{1}{2}ad \\
E(\ddot{C}_4) &= d_2 - \frac{1}{2}aa \\
E(\ddot{C}_5) &= (aa+dd) \\
E(\ddot{C}_6) &= (ad+da).
\end{aligned}$$

The MIM model is then

$$y_{ij} = \mu_j + \sum_{r=1}^{m} \alpha_r x_{ijr}^{*} + \sum_{r=1}^{m} \beta_r z_{ijr}^{*} + \sum_{r<s}^{t_1} \gamma_{rs} w_{ijrs}^{*} + \sum_{r<s}^{t_2} \delta_{rs} o_{ijrs}^{*} + \varepsilon_{ij},$$

where y_{ij}, μ_j, α_r, β_r, γ_{rs}, δ_{rs}, and ε_{ij} have the same interpretation of the MIM model in the main text.

The indicator variables for the main and interaction effects are

$$x_{ijr}^{*} = \begin{cases} 1 & \text{if the genotype of } Q_r \text{ is } Q_r Q_r \\ & \\ -1 & \text{if the genotype of } Q_r \text{ is } q_r q_r \end{cases} \quad \text{for } j = 1,\ 2;$$

$$z_{ijr}^{*} = \begin{cases} x_{ijr}^{*} & \text{if } j = 1 \\ -x_{ijr}^{*} & \text{if } j = 2 \end{cases}$$

$$w_{ijrs}^{*} = \begin{cases} \frac{1}{2} & \text{if the QTL genotype is } Q_r Q_r Q_s Q_s \text{ or } q_r q_r q_s q_s \\ & \\ \frac{-1}{2} & \text{if the QTL genotype is } Q_r Q_r q_s q_s \text{ or } q_r q_r Q_s Q_s \end{cases} \quad \text{for } j = 1,\ 2;$$

$$o_{ijrs}^{*} = \begin{cases} w_{ijrs}^{*} & \text{if } j = 1 \\ -w_{ijrs}^{*} & \text{if } j = 2. \end{cases}$$

APPENDIX D: EM ALGORITHM

Adapting the general formulas of Kao and Zeng (1997) for the likelihood of our model, we present here the EM algorithm using matrix notation. (However, when coding the software, we took into consideration the problems for convergence presented by Zeng *et al.* 1999 and used a different notation; see Kao and Zeng 1997 for details). For the $[\tau + 1]$*th* iteration,

E step:

$$\pi_{ig}^{[\tau+1]} = \frac{p_{ig} \prod_{j=1}^{2} \phi(y_{ij} \mid \mu_j^{[\tau]} + \mathbf{D_{jg}} \mathbf{E}^{[\tau]}, \sigma_j^{2[\tau]})}{\sum_{g=1}^{3^m} \left[p_{ig} \prod_{j=1}^{2} \phi(y_{ij} \mid \mu_j^{[\tau]} + \mathbf{D_{jg}} \mathbf{E}^{[\tau]}, \sigma_j^{2[\tau]}) \right]}$$

M step:

$$\begin{aligned} \mathbf{E}^{[\tau+1]} &= \mathbf{r}^{[\tau]} - \mathbf{M}^{[\tau]} \mathbf{E}^{[\tau]} \\ \mu_j^{[\tau+1]} &= \left(\frac{1}{n}\right) \mathbf{1}'(\mathbf{Y_j} - \mathbf{\Pi}^{[\tau+1]} \mathbf{D_j} \mathbf{E}^{[\tau+1]}) \\ \sigma_j^{2[\tau+1]} &= \left(\frac{1}{n}\right) \Big[(\mathbf{Y_j} - \mathbf{1}\mu_j^{[\tau+1]})'(\mathbf{Y_j} - \mathbf{1}\mu_j^{[\tau+1]}) \\ &\qquad - 2(\mathbf{Y_j} - \mathbf{1}\mu_j^{[\tau+1]})' \mathbf{\Pi}^{[\tau+1]} \mathbf{D_j} \mathbf{E}^{[\tau+1]} \\ &\qquad + \mathbf{E}'^{[\tau+1]} \mathbf{V_j}^{[\tau]} \mathbf{E}^{[\tau+1]} \Big], \end{aligned}$$

where $\mathbf{1}$ is a column vector of ones, $\mathbf{\Pi} = \{\pi_{ig}\}_{n \times 3^m}$, $\mathbf{V_j} = \{\mathbf{1}'\mathbf{\Pi}(\mathbf{D_{jk}}\mathbf{D_{jl}})\}_{m(m+1) \times m(m+1)}$, $\mathbf{r} = \left\{ \left[\sum_j (1/\sigma_j^2)(\mathbf{Y_j} - \mathbf{1}\mu_j)' \mathbf{\Pi}\mathbf{D_{jk}} \right] \Big/ \left[\sum_j (1/\sigma_j^2) \mathbf{1}'\mathbf{\Pi}(\mathbf{D_{jk}}\mathbf{D_{jk}}) \right] \right\}_{m(m+1) \times 1}$, and $\mathbf{M} = \left\{ \left[\sum_j (1/\sigma_j^2) \mathbf{1}'\mathbf{\Pi}(\mathbf{D_{jk}}\mathbf{D_{jl}}) \right] \Big/ \left[\sum_j (1/\sigma_j^2) \mathbf{1}'\mathbf{\Pi}(\mathbf{D_{jk}}\mathbf{D_{jk}}) \right] \times \delta(k \neq l) \right\}_{m(m+1) \times m(m+1)}$. $\mathbf{D_{jk}}$ ($\mathbf{D_{jl}}$) is the *k*th (*l*th) column of the genetic design matrix $\mathbf{D_j}$, $\delta(k \neq l)$ is an indicator variable that assume values 1 if $k \neq l$ and 0 otherwise, and # denotes the Hadamard product. For details about genetic design matrices see Kao and Zeng (1997) and Kao *et al.* (1999).

To test the MLEs of the **E** vector, the likelihood-ratio test or the LOD score can be used. For example, for testing the effect E_r,

$$\text{LOD} = \log_{10} \frac{L(E_1 \neq 0, \ldots, E_{2m+t_1+t_2} \neq 0)}{L(E_1 \neq 0, \ldots, E_{r-1} \neq 0, E_r = 0, E_{r+1} \neq 0, \ldots, E_{2m+t_1+t_2} \neq 0)}.$$