

The Role of Epistasis in the Manifestation of Heterosis: A Systems-Oriented Approach

A. E. Melchinger,^{*,1} H. F. Utz,^{*} H.-P. Piepho,[†] Z.-B. Zeng[‡] and C. C. Schön[§]

^{*}*Institute of Plant Breeding, Seed Science and Population Genetics,* [†]*Bioinformatics Unit, Institute of Crop Production and Grassland Research and* [§]*State Plant Breeding Institute, University of Hohenheim, 70599 Stuttgart, Germany and* [‡]*Program in Statistical Genetics, Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695*

Manuscript received June 13, 2007
Accepted for publication August 28, 2007

ABSTRACT

Heterosis is widely used in breeding, but the genetic basis of this biological phenomenon has not been elucidated. We postulate that additive and dominance genetic effects as well as two-locus interactions estimated in classical QTL analyses are not sufficient for quantifying the contributions of QTL to heterosis. A general theoretical framework for determining the contributions of different types of genetic effects to heterosis was developed. Additive \times additive epistatic interactions of individual loci with the entire genetic background were identified as a major component of midparent heterosis. On the basis of these findings we defined a new type of heterotic effect denoted as augmented dominance effect d_i^* that comprises the dominance effect at each QTL minus half the sum of additive \times additive interactions with all other QTL. We demonstrate that genotypic expectations of QTL effects obtained from analyses with the design III using testcrosses of recombinant inbred lines and composite-interval mapping precisely equal genotypic expectations of midparent heterosis, thus identifying genomic regions relevant for expression of heterosis. The theory for QTL mapping of multiple traits is extended to the simultaneous mapping of newly defined genetic effects to improve the power of QTL detection and distinguish between dominance and overdominance.

THE concept of heterosis is widely used in plant and animal breeding. In many species, the controlled crossing of selected parental components, mainly inbred lines, is employed to maximize heterosis and thus agronomic performance of the resulting F_1 hybrids. However, the understanding of this biological phenomenon is limited and the genetic basis of heterosis has yet to be elucidated.

Quantitative genetics has contributed to our understanding of heterosis by (i) formulating genetic models explaining heterosis on the basis of different modes of gene action such as dominance, overdominance, and epistasis (for review see LAMKEY and EDWARDS 1999); (ii) devising the theory for the design and analysis of experiments investigating these types of gene action (for review see LYNCH and WALSH 1998); and (iii) gathering a plethora of experimental data supporting or questioning these theories. One of the experimental approaches proposed for investigating the relative importance of different types of gene action is the analysis of generation means (KEARSEY and POONI 1996). Six basic generations (the two parental lines and their F_1 and F_2 as well as backcrosses of the F_1 to each of the parents) are used to estimate the magnitude of additive, dominance, and epistatic effects affecting the quantitative trait un-

der study. However, these parameters capture the net contribution of gene effects summed over all loci and consequences of summation may be pronounced if positive and negative effects at individual loci cancel each other. An alternative approach is the partitioning of the genetic variance into independent components due to additive, dominance, and epistatic effects (FISHER 1918). However, as pointed out by LYNCH and WALSH (1998), unless information on gene frequencies of the reference population is available, variance components provide limited information on the relative importance of the different modes of gene action because dominance and epistasis can greatly affect additive or dominance components of variance. This limitation can be overcome by the use of populations derived from a cross between two inbred lines such as the North Carolina experiment III (design III) proposed by COMSTOCK and ROBINSON (1952). A random sample of F_2 individuals derived from a cross between two homozygous inbred lines is backcrossed to each of the parents, yielding a population with gene and genotype frequencies equivalent to an F_2 . A one-way analysis of variance (ANOVA) of phenotypic means and differences of the F_2 backcrosses yields estimates of the dominance and additive genetic variance with nearly equal precision and their ratio provides a weighted estimate of the squared degree of dominance.

A major step toward the analysis of the type of gene action at individual genetic loci affecting quantitative traits (QTL) was the advent of molecular marker technology.

¹*Corresponding author:* Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, Fruwirthstrasse 21, 70599 Stuttgart, Germany. E-mail: melchinger@uni-hohenheim.de

The first marker-aided study estimated the predominant type of gene action in segregating F_2 populations of maize (EDWARDS *et al.* 1987). Additive and dominance gene effects at individual QTL as well as digenic epistatic interactions were estimated on the basis of contrasts of marker genotype classes. However, these estimates are not sufficient for making inferences about the genetic basis of heterosis. The performance of an F_1 hybrid or its filial generations and midparent heterosis (MPH) have distinct genetic components. Assuming the presence of digenic epistasis, hybrid performance is a function of the sum of dominance effects and dominance \times dominance (*dd*) epistasis. To date, a general mathematical derivation of the contributions of different genetic effects to MPH is still lacking, but quantitative genetic expectations of MPH have been presented in a descriptive manner (*e.g.*, VAN DER VEEN 1959). While additive \times additive (*aa*) interactions contribute to MPH, *dd* interactions may not, depending on the metric used for describing genotypic values. In this study we show that QTL exhibiting significant dominance effects may not contribute to MPH if the sum of their *aa* epistatic interactions with the genetic background is positive. Furthermore, we show that significant heterotic QTL can result from *aa* epistatic interactions with the genetic background if positive and negative alleles are contributed with equal frequency by each parent. Consequently, we define a new genetic effect (d_i^*) that allows us to express MPH as the sum of individual QTL effects and demonstrate that design III is suitable for mapping of QTL with genotypic expectations equivalent to the augmented dominance effect d_i^* .

The objectives of this study were to (i) develop a general theoretical framework for determining the contributions of the different types of genetic effects to heterosis, (ii) dissect MPH into its underlying components using composite-interval mapping (CIM) and a variant of design III, (iii) give quantitative genetic expectations of heterotic QTL, and (iv) extend the theory for the joint analysis of multiple traits (JIANG and ZENG 1995) to the simultaneous mapping of different newly defined genetic effects to improve the power of QTL detection and distinguish between dominance and overdominance. Results from an experiment on heterosis in *Arabidopsis* for which the presented theory has formed the quantitative genetic basis are in an accompanying article (KUSTERER *et al.* 2007, this issue).

GENETIC EFFECTS CONTRIBUTING TO HETEROISIS

When assessing the type of gene action contributing to heterosis, we need to express heterosis as a function of additive, dominance, and epistatic gene effects. To date, no generalized mathematical derivation showing the respective contributions of the different genetic effects has been developed for an arbitrary number of QTL and all types of higher-order epistatic effects.

MPH for a quantitative trait is defined as the difference between the genotypic value of an F_1 hybrid (G_{F_1}) and the mean genotypic value of its two homozygous parents (G_{P_1} , G_{P_2}):

$$\text{MPH} = G_{F_1} - (G_{P_1} + G_{P_2})/2. \quad (1)$$

Let P_1 and P_2 differ at the loci set $Q = \{1, \dots, q\}$ affecting the quantitative trait of interest. Let v_i be an indicator variable for the genotype at QTL i taking values 0, 1, 2 if homozygous P_1 , heterozygous, or homozygous P_2 , respectively. The types of genetic effects contributing to G are described by the 3^q parameters α_{AD} , which define the genetic effects of type additive at the loci set A ($A \subset Q$) and of type dominance at the loci set $D \subset Q_A$ ($Q_A = Q \setminus A$, the complement of A in Q). For exemplification of α_{AD} for $q = 3$ see supplemental Table S1 at <http://www.genetics.org/supplemental/>.

The coefficients and meaning of parameters α_{AD} in the genotypic value G depend on the choice of the metric. In the quantitative genetic literature, two main metrics have been described for populations derived from a cross between two inbred lines: the F_2 metric and the F_∞ metric (VAN DER VEEN 1959; YANG 2004). The F_2 metric model, a special case of COCKERHAM'S (1954) model for partitioning the genetic variance into eight orthogonal contrasts due to additive, dominance, and epistatic effects, defines genetic effects as deviations from the mean of the F_2 population in linkage equilibrium. The F_∞ model defines genetic effects as contrasts between different genotypes without reference to any population (YANG 2004). Both models can be translated into each other by a linear transformation. In the presence of epistasis they differ with respect to interpretation of genetic effects and the structure of variance components. With the F_2 metric and an F_2 population in linkage equilibrium the genetic variance is partitioned into orthogonal components that represent sums of squared additive, dominance, and epistatic effects, providing insight into the relative importance of different types of gene action. Consequently, in the analysis of heterosis with design III, we prefer the F_2 metric. With the F_∞ metric, additional complexity is introduced, because besides the dominance effect at a QTL and its *aa* epistatic interactions, also *dd* epistatic interactions with the genetic background contribute to MPH (VAN DER VEEN 1959). Consequently, genetic expectations of individual heterotic QTL become unwieldy and more difficult to interpret with the F_∞ metric.

Under the F_2 metric, variables $x_{V,A} = \prod_{i \in A} x_{V,i}$ and $y_{V,D} = \prod_{i \in D} y_{V,i}$ determine the coefficient of α_{AD} in G with $x_{V,i} = -1, 0, 1$ and $y_{V,i} = -\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}$ when $v_i = 0, 1, 2$, respectively.

Thus, we can express the genotypic value of genotype $V = (v_1, \dots, v_q)$ as

$$G_V = \sum_{A \subset Q} \sum_{D \subset Q_A} x_{V,A} y_{V,D} \alpha_{AD}, \quad (2)$$

with $\sum_{A \subset Q}$ indicating summation over all possible subsets A within the set Q . For $A = \emptyset$ and $D = \emptyset$, $\alpha_{AD} = \mu$; for $D = \emptyset$, $\alpha_{AD} = \alpha_A$; for $A = \emptyset$, $\alpha_{AD} = \alpha_D$.

Let us assume an F_2 individual with $V = (v_i, v_j, v_k) = (2, 1, 0)$. Then, $x_{V,i} = 1$, $x_{V,j} = 0$, $x_{V,k} = -1$, $y_{V,i} = -\frac{1}{2}$, $y_{V,j} = \frac{1}{2}$, and $y_{V,k} = -\frac{1}{2}$. Thus,

$$G_V = \mu + a_i - a_k - \frac{1}{2}d_i + \frac{1}{2}d_j - \frac{1}{2}d_k - aa_{ik} + \frac{1}{2}ad_{ij} - \frac{1}{2}ad_{ik} + \frac{1}{2}da_{ik} - \frac{1}{2}da_{jk} - \frac{1}{4}dd_{ij} + \frac{1}{4}dd_{ik} - \frac{1}{4}dd_{jk} - \frac{1}{2}ada_{ijk} - \frac{1}{4}add_{ijk} + \frac{1}{4}dda_{ijk} + \frac{1}{8}ddd_{ijk}.$$

The parameter μ denotes the genotypic expectation of the F_2 generation in linkage equilibrium. In accordance with the definition of FALCONER and MACKAY (1996, p. 109) a_i denotes the additive or homozygous effect at QTL i and d_i the dominance or heterozygous effect. The additive effect at locus i is positive ($+a_i$) when the trait-increasing allele is contributed by P2 and negative ($-a_i$) when contributed by P1. The degree of dominance is expressed as d_i/a_i . Epistatic interactions between loci i, j , and k are denoted, adopting the notation of YANG (2004). Analogously, the genotypic values of the parental homozygous lines P1 and P2 and the F_1 hybrid are

$$G_{P1} = \sum_{A \subset Q} \sum_{D \subset Q_A} (-1)^{|A|} \left(-\frac{1}{2}\right)^{|D|} \alpha_{AD}$$

and

$$G_{P2} = \sum_{A \subset Q} \sum_{D \subset Q_A} (1)^{|A|} \left(-\frac{1}{2}\right)^{|D|} \alpha_{AD}$$

and

$$G_{F1} = \sum_{D \subset Q} \left(\frac{1}{2}\right)^{|D|} \alpha_D,$$

where $|A|$ and $|D|$ denote the number of elements in sets A and D , respectively. Then, MPH can be calculated as

$$\begin{aligned} MPH &= G_{F1} - (G_{P1} + G_{P2})/2 \\ &= \sum_{D \subset Q} \left(\frac{1}{2}\right)^{|D|} \alpha_D - \frac{1}{2} \sum_{A \subset Q} \sum_{D \subset Q_A} ((-1)^{|A|} + (1)^{|A|}) \left(-\frac{1}{2}\right)^{|D|} \alpha_{AD} \\ &= \sum_{\substack{D \subset Q \\ |D| \text{ odd}}} \left(\frac{1}{2}\right)^{|D|-1} \alpha_D - \sum_{\substack{A \subset Q \\ |A| \text{ even}}} \sum_{D \subset Q_A} \left(-\frac{1}{2}\right)^{|D|} \alpha_{AD}. \end{aligned} \tag{3}$$

This generalized derivation shows that under the F_2 metric the quantitative genetic expectation of MPH is affected by the dominance effects at QTL and by epistatic interactions including an odd number of dominance terms (e.g., ddd but not dd). In addition, epistatic effects including additive terms also contribute to MPH,

because MPH is based on the deviation of the F_1 hybrid from the mean of the two homozygous parental lines. However, only effects with an even number of additive terms (e.g., aa and aad but not ad and aaa) contribute to MPH.

Considering only digenic epistasis, Equation 3 can be written as

$$MPH = \sum_{i \in Q} \left(d_i - \frac{1}{2} \sum_{j \in Q_i} aa_{ij} \right)$$

with Q_i denoting the loci set Q excluding element i .

To express MPH as the sum of individual QTL effects we define a new type of heterotic genetic effect d_i^* that we denote as augmented dominance effect and that includes the dominance effect of QTL i (d_i) minus half the sum of its additive \times additive epistatic interactions (aa_{ij}) with all other QTL irrespective of linkage

$$d_i^* = d_i - \frac{1}{2} \sum_{j \in Q_i} aa_{ij}. \tag{4}$$

Then, MPH can be expressed as the sum of augmented dominance QTL effects

$$MPH = \sum_{i \in Q} d_i^*. \tag{5}$$

The effect of additive \times additive epistasis on MPH and d_i^* is demonstrated with a numerical example for four QTL in the supplemental information (supplemental Table S2 at <http://www.genetics.org/supplemental/>). Depending on gene dispersion and the magnitude of the aa epistatic effects, QTL with significant dominance effects d_i may not contribute to MPH if the sum of their aa epistatic interactions with the genetic background is positive [supplemental Table S2, $\mathbf{S} = (+, +, +, +)$]. Furthermore, significant heterotic QTL can result from aa epistatic interactions with the genetic background if positive and negative alleles are contributed with equal frequency by each parent, as would be expected for an elite \times elite cross [supplemental Table S2, $\mathbf{S} = (+, +, -, -)$].

The quantitative genetic expectation of the parental difference (PD) is found to be

$$PD = G_{P2} - G_{P1} = 2 \sum_{\substack{A \subset Q \\ |A| \text{ odd}}} \sum_{D \subset Q_A} \left(-\frac{1}{2}\right)^{|D|} \alpha_{AD}. \tag{6}$$

Note that under the F_2 metric the formulas for homozygous genotypes such as parents P1 and P2 involve also dominance terms (α_{AD}) besides purely additive terms (α_A). As a consequence, PD is affected by the additive effects at the QTL as well as by epistatic interactions including an odd number of additive and an arbitrary number of dominance terms (e.g., ad but not aa). Considering only digenic epistasis, PD reduces to

TABLE 1
Analysis of variance of design III evaluated in a randomized complete block design

Source of variation	d.f. ^a	Expected mean square ^b
Replications	$r - 1$	
Parental lines (P)	1	
Z_1	$n - 1$	$\sigma_c^2 + 2r\sigma_{Z_1(m)}^2$
Marker m	1	$\sigma_c^2 + 2r\sigma_{w_1(m)}^2 + r\bar{n}[\bar{\mathcal{E}}(Z_1(m))]^2$
Remainder $Z_1(m)$	$n_m - 2$	$\sigma_c^2 + 2r\sigma_{w_1(m)}^2$
Z_2	$n - 1$	$\sigma_c^2 + \frac{1}{2}r\sigma_{Z_2(m)}^2$
Marker m	1	$\sigma_c^2 + \frac{1}{2}r\sigma_{w_2(m)}^2 + \frac{1}{4}r\bar{n}[\bar{\mathcal{E}}(Z_2(m))]^2$
Remainder $Z_2(m)$	$n_m - 2$	$\sigma_c^2 + \frac{1}{2}r\sigma_{w_2(m)}^2$
Error	$(2n - 1)(r - 1)$	σ_c^2

Sums of squares for pair means Z_1 and pair differences Z_2 were subdivided into orthogonal contrasts allowing the comparison of genotypic classes at marker m .

^a r , number of replicates; n , number of RILs; n_m , number of RILs genotyped at marker m ,

$n_m = n_{m0} + n_{m2}$ with n_{m0} and n_{m2} referring to the number of RILs with marker genotypes 0 and 2, respectively.

^b $\bar{n} = 2/(1/n_{m0} + 1/n_{m2})$, $\sigma_{w_1(m)}^2$, and $\sigma_{w_2(m)}^2$ refer to the variance of residual genotypic effects in Z_1 and Z_2 not accounted for by the m th marker. σ_c^2 , error variance.

$$PD = \sum_{i \in Q} \left(2a_i - \sum_{j \in Q_i} da_{ij} \right) = \sum_{i \in Q} 2a_i^*,$$

where

$$a_i^* = a_i - \frac{1}{2} \sum_{j \in Q_i} da_{ij}. \quad (7)$$

In accordance with the term suggested for the effect d_i^* , a_i^* is denoted as augmented additive effect. It includes the additive effect for QTL i (a_i) minus half the sum of dominance \times additive epistatic interactions (da_{ij}) with all other QTL, corresponding exactly to the net contribution of QTL i to the parental difference.

IDENTIFICATION OF AUGMENTED QTL EFFECTS

On the basis of the results from the previous section it becomes obvious that we need to revise our approaches for the identification of genomic regions contributing to heterosis. Instead of identifying QTL with maximum dominance and dd interactions that increase F_1 performance we need to identify genomic regions that contribute to MPH, *i.e.*, that yield significant results in the QTL analysis due to the dominance effect at QTL i and its epistatic interaction with the genetic background. Thus, specific experimental designs are needed that identify QTL with genotypic expectations that precisely equal the augmented dominance effect d_i^* . To our knowledge this criterion is met only by genetic effects estimated with design III (COMSTOCK and ROBINSON 1952). The original design III comprised the analysis of F_2 individuals backcrossed to their parental inbred lines and was devised for estimating the average degree of dominance over all loci. We modified the statistical analysis of design

III to accommodate the analysis of testcrosses of recombinant inbred lines (RILs) because they are “immortal” test units that can be shared between research groups and can be repeatedly phenotyped.

Design III with RILs: Let us assume a random population of RILs derived from the cross between the two homozygous lines P1 and P2. Further, we assume that the RILs are backcrossed to their parental lines, yielding testcross progenies H_1 and H_2 . Gene effects of the design III testcross progenies are expressed with regard to the corresponding gene-orthogonal population, *i.e.*, the F_2 . The parental line exhibiting superior average testcross performance across both testers is denoted as P2. With testcrosses evaluated in a randomized complete block design we obtain phenotypic trait values Y_{tpk} of the testcross progeny H_t of RIL p ($p = 1, \dots, n$) crossed with tester t ($t = 1, 2$ for design III) in the k th block ($k = 1, \dots, r$). Following COMSTOCK and ROBINSON (1952), we perform two linear transformations Z_s ($s = 1, 2$) on the performance data Y_{tpk} of testcross progenies H_t with pair means $Z_{1pk} = (Y_{1pk} + Y_{2pk})/2$ and pair differences $Z_{2pk} = Y_{1pk} - Y_{2pk}$. Thus, Z_{spk} denotes the phenotypic value of transformation Z_s for RIL p grown in the k th block and Z_{sp} the progeny mean value of Z_s for RIL p . Expected mean squares from the ANOVA are given in Table 1.

Marker-based estimation of augmented effects a_i^* and d_i^* : The first to present the statistical theory for estimation of the type of gene action at individual QTL with design III were COCKERHAM and ZENG (1996). They presented genotypic expectations of marker contrasts using single-marker ANOVA and extended the analysis to test for two-locus interactions between linked QTL. They demonstrated that genotypic expectations of QTL mapped with design III were biased with epistasis. Dominance effects were confounded with aa epistasis, additive effects with da interactions regardless of linkage.

However, interactions of individual QTL with the entire genetic background were not accounted for and, most importantly, they did not make the connection to genetic expectations of MPH. While COCKERHAM and ZENG (1996) considered the confounding epistatic effects in their analysis a limitation, we claim that they are favorable for the identification of genomic regions contributing to heterosis. The following derivations show that genotypic expectations of QTL identified with design III precisely equal the augmented dominance effect d_i^* .

COCKERHAM and ZENG (1996) defined four orthogonal single-marker contrasts (C_1, C_2, C_3, C_4) among the means of testcross progenies from F_2 individuals or F_3 lines, estimating additive (C_1) and dominance effects (C_3) as well as digenic epistasis of linked QTL (C_2, C_4). Here, we extend their methods to the analysis of heterosis and to the analysis of RILs. We defined two orthogonal single-marker contrasts on the basis of progeny mean values Z_{sp} for pair means ($Z_1(m)$) and pair differences ($Z_2(m)$) (see the APPENDIX for a generalized derivation). While contrasts C_1 and C_3 in the notation of COCKERHAM and ZENG (1996) correspond to $Z_1(m)$ and $Z_2(m)$ in our notation, contrasts C_2 and C_4 , which provide tests for epistasis among linked QTL, cannot be calculated for RILs because they rely on comparisons of the heterozygous *vs.* homozygous marker classes.

With epistasis restricted to digenic effects, we obtain

$$\mathcal{E}(Z_1(m)) = 4 \sum_{i \in Q} D_{mi} \left[a_i - \frac{1}{2} \sum_{j \in Q_i} da_{ij} \right] = 4 \sum_{i \in Q} D_{mi} a_i^* \tag{8}$$

$$\mathcal{E}(Z_2(m)) = 8 \sum_{i \in Q} D_{mi} \left[d_i - \frac{1}{2} \sum_{j \in Q_i} aa_{ij} \right] = 8 \sum_{i \in Q} D_{mi} d_i^*, \tag{9}$$

D_{mi} being the linkage disequilibrium between QTL i and marker m (WEIR 1996). For RILs, D_{mi} can be calculated from the recombination frequency r_{mi} between QTL i and marker m as $D_{mi} = (1 - 2r_{mi})/4(1 + 2r_{mi})$ (for further details see the APPENDIX).

As can be seen from Equation 9, the expectation of $Z_2(m)$ is a multiple of the sum of the augmented dominance effects d_i^* for all QTL with $D_{mi} > 0$, weighted by their linkage disequilibrium D_{mi} to the marker m . Thus, estimation of dominance effects of QTL linked to marker m is confounded with digenic epistatic interactions of type additive \times additive with the entire genetic background. Analogously, the linear contrast $Z_1(m)$ is a multiple of the sum of augmented additive effects a_i^* for all QTL with $D_{mi} > 0$, weighted by their linkage disequilibrium D_{mi} to the marker.

Expected mean squares for an ANOVA that includes contrasts $Z_1(m)$ and $Z_2(m)$ for design III evaluated in a randomized complete block design are presented in Table 1. Appropriate statistical tests can be derived by

standard statistical theory. From expected mean squares, it becomes evident that the power of detecting significant QTL effects linked to marker m depends mainly on segregation variances $\sigma_{w_1(m)}^2$ and $\sigma_{w_2(m)}^2$ of the variables Z_1 and Z_2 with respect to QTL effects not accounted for by the marker m , *i.e.*, the genetic background variation.

With single-marker ANOVA the QTL position (here reflected by parameter D_{mi}) and QTL effects (here a_i^* or d_i^*) are confounded. Hence, the effects of linked QTL cannot be separated, and effects of other unlinked QTL, which cause background noise, are not accounted for. The above shortcomings have been overcome by CIM (ZENG 1994; JANSEN and STAM 1994), which allows separate estimation of the QTL position and the QTL effect as well as separation of linked QTL and control of genetic background variation. We have adopted this method for identification of genomic regions affecting heterosis. Using the model

$$Z_{sp} = b_{0s} + b_s^* x_p^* + \sum_l^h b_{ls} x_{lp} + \varepsilon_{sp}, \tag{10}$$

where Z_{sp} is the linear function Z_s calculated for the p th RIL, b_{0s} is the mean effect of the model for Z_s , b_s^* is the augmented effect of the putative QTL on Z_s , x_p^* is the conditional probability of the dummy variable θ_m given the observed genotypes at the marker loci m and $(m + 1)$ flanking the putative QTL (θ_m takes value 0 or 2 if the genotype of the RIL at the QTL is homozygous P1 or P2, respectively), b_{ls} are partial regression coefficients of Z_{sp} on the l th marker assuming h markers included in the model as cofactors, x_{lp} is an indicator variable taking value 0 or 2 depending on the genotype at marker l , and ε_{sp} is the residual effect on Z_s for the p th RIL. CIM analyses can be individually performed for the two linear functions Z_s from design III. Tests of significance are straightforward as described for CIM (ZENG 1994) with the null hypothesis $H_0: b_s^* = 0$ and the alternative hypothesis $H_A: b_s^* \neq 0$.

If we assume complete linkage between the marker m and one of the QTL ($r_{mi} = 0$), we obtain the following genotypic expectations for the contrast of the two homozygous marker classes at marker m :

$$\begin{aligned} \mathcal{E}(Z_1(m) | r_{mi} = 0) &= a_i - \frac{1}{2} \sum_{j \in Q_i} da_{ij} + \sum_{j \in Q_i} 4D_{ij} \left(a_j - \frac{1}{2} \sum_{k \in Q_j} da_{jk} \right) \\ &= a_i^* + \sum_{j \in Q_i} 4D_{ij} a_j^* \end{aligned} \tag{11}$$

$$\begin{aligned} \mathcal{E}(Z_2(m) | r_{mi} = 0) &= 2d_i - \sum_{j \in Q_i} aa_{ij} + \sum_{j \in Q_i} 4D_{ij} \left(2d_j - \sum_{k \in Q_j} aa_{jk} \right) \\ &= 2d_i^* + \sum_{j \in Q_i} 4D_{ij} 2d_j^*. \end{aligned} \tag{12}$$

The advantages of CIM in comparison to single-marker ANOVA become obvious immediately: (i) the position of the QTL i can be estimated and (ii) the effect of the

linked QTL j should be blocked by the use of cofactors (JANSEN and STAM 1994; ZENG 1994). Thus, it is possible to test contrasts for QTL (not markers) and the genotypic expectation for $Z_2(i)$, the contrast of Z_2 between the two unobservable homozygous genotype classes at QTL i , reduces to $\bar{X}(Z_2(i)) = 2d_i^*$. Consequently, a genome scan with Z_2 localizes genomic regions affecting MPH and $b_2^*/2 = d_i^*$. Accordingly, a_i^* , the contribution of QTL i to the parental difference, is equal to b_1^* . The advantage of design III is that genotypic expectations of QTL mapped with Z_2 precisely equal their net contribution to MPH. However, the contribution of the main dominance effect d_i cannot be assessed independently from the sum of *aa* epistatic interactions of QTL i with the genetic background, which is a limitation of design III.

Extension of heterotic QTL analyses: In addition to separate CIM of Z_1 and Z_2 we propose to apply the theory of a joint analysis for multiple traits as suggested by JIANG and ZENG (1995) to the simultaneous mapping of augmented QTL effects a_i^* and d_i^* . Depending on the correlation structure of data on multiple traits a joint analysis may increase the power of the likelihood-ratio (LR) tests for QTL detection and may allow us to distinguish between pleiotropy and close linkage of QTL for individual traits. Following JIANG and ZENG (1995), we regard Z_1 and Z_2 as two different traits and rewrite the model from Equation 10 in matrix notation,

$$\mathbf{Z} = \mathbf{x}^* \mathbf{b}^* + \mathbf{X}\mathbf{B} + \mathcal{E}, \quad (13)$$

where \mathbf{Z} is a matrix of Z_{sp} , \mathbf{x}^* is a column vector of x_p^* , \mathbf{b}^* is a row vector of b_s^* , \mathbf{X} and \mathbf{B} are two matrices controlling the genetic background variation, and \mathcal{E} is a matrix of ε_{sp} . The power of QTL detection using the model from Equation 13 is significantly increased by a joint LR test if the product of the effects b_1^* (corresponding to a_i^*) and b_2^* (corresponding to $2d_i^*$) and the correlation ρ_{12} between residuals ε_{1p} and ε_{2p} have different signs (JIANG and ZENG 1995). In the joint QTL analysis, the hypotheses to be tested are: $H_0: b_1^* = 0, b_2^* = 0$ vs. $H_A: b_1^* \neq 0$ or $b_2^* \neq 0$.

On the basis of experimental results from the analysis of complex traits it seems reasonable to assume that marker-trait associations detected with the model in Equation 10 will rarely explain >50% of the genotypic variance even with large sample sizes (SCHÖN *et al.* 2004). Thus, the residual component ε_{sp} must be subdivided into two components; *i.e.*, $\varepsilon_{sp} = g_{sp} + e_{sp}$, with g_{sp} reflecting the genetic component in ε_{sp} not accounted for by the putative QTL and the cofactors in the model, and e_{sp} being the experimental error. As a consequence, residual components ε_{1p} and ε_{2p} can be correlated within RILs but are independent among RILs. The correlation ρ_{12} between residuals ε_{1p} and ε_{2p} can be approximated by $\rho_{12} = \text{cov}(Z_1, Z_2) / \sqrt{\sigma_{Z_1}^2 \sigma_{Z_2}^2}$. Using the derivation of covariances presented in the APPENDIX and assuming digenic epistasis and absence of linkage,

$$\text{cov}(Z_1, Z_2) = \frac{1}{2} \sum_{i \in Q} a_i^* d_i^* + \frac{1}{8} \sum_{i \in Q} \sum_{j \in Q} ad_{ij} dd_{ij}. \quad (14)$$

Hence, it follows that ρ_{12} depends mainly on the variation in sign and magnitude of a_i^* and d_i^* . When P1 and P2 are elite inbred lines, we expect that about half of the trait-increasing alleles are contributed by P1 (*i.e.*, $a_i^* < 0$) and half by P2 (*i.e.*, $a_i^* > 0$). With directional dominance ($d_i^* > 0$), the correlation of residuals is in this case expected to be close to zero. Thus, the joint LR test statistic will approximate the sum of the individual LR test statistics and no increase in the power of QTL detection can be expected from a genome scan using the multivariate model from Equation 13. However, when analyzing progenies from crosses between an elite parent (P2) and an exotic donor line (P1) with a much smaller proportion of positive alleles than the elite parent the joint LR test statistic can be advantageous for finding heterotic QTL contributed by P1. If P2 contributes the majority of positive dominant alleles, *i.e.*, $a_i^* > 0$ and $d_i^* > 0$, then $\rho_{12} > 0$. In genome regions where a positive dominant allele originates from P1 ($a_i^* < 0$; $d_i^* > 0$), the product of effects a_i^* and d_i^* will be negative, while $\rho_{12} > 0$. Thus, the power of detecting heterotic QTL from exotic donor parents will be increased with the joint LR test.

In addition to increasing the power of QTL detection with multiple traits, the statistical method suggested by JIANG and ZENG (1995) also provides a test for separating pleiotropy from close linkage. In our case, the analogous test can be applied to distinguish between the hypothesis of dominance vs. overdominance in genomic regions with significant effects a_i^* and d_i^* . LR tests for the null hypothesis $H_0: z(1) = z(2)$ vs. $H_A: z(1) \neq z(2)$, where z is the position of the QTL affecting Z_1 and Z_2 , respectively, have been described by JIANG and ZENG (1995). Under the null hypothesis, the same genetic locus contributes to the augmented effects a_i^* and d_i^* . If both effects are of similar magnitude, acceptance of H_0 would imply that dominant gene action is prevalent. In contrast, acceptance of H_A would imply that one locus shows additive gene action ($a_i^* > 0$; $d_i^* = 0$) and at the second locus the heterozygote outperforms both homozygotes, thus exhibiting overdominance ($a_j^* = 0$; $d_j^* > 0$; $Y_{P1} = Y_{P2} < Y_{F1}$).

GENOTYPIC EXPECTATION OF BETTER PARENT HETEROISIS

When analyzing heterosis in self-pollinating crops such as wheat, rice, or tomato the reference base for the calculation of heterosis is often not the midparent value but the superiority over the better parent (*e.g.*, SEMEL *et al.* 2006). However, if one is interested in the genetic causes of heterosis, it is most plausible to compare the hybrid with the average performance of both parental lines (and not only the better parent) because the F_1 inherited half its nuclear genome from each parent. As can be seen from the following derivations, better parent

heterosis (BPH) can be expressed as a function of augmented dominance and additive effects. Consequently, the genetic causes of BPH are more complex than those of MPH and include the latter.

Defining P2 to be the better-performing parent, the genotypic expectation of BPH can be calculated as

$$\text{BPH} = \sum_{\substack{D \subset Q \\ |D| \text{ odd}}} \left(\frac{1}{2}\right)^{|D|-1} \alpha_D - \sum_{A \subset Q} \sum_{\substack{D \subset Q_A \\ |A| > 0}} \left(-\frac{1}{2}\right)^{|D|} \alpha_{AD}.$$

Thus, we obtain that BPH is affected by genetic effects with an odd number of dominance terms and effects with at least one additive and an arbitrary number of dominance terms. Considering only digenic epistasis,

$$\begin{aligned} \text{BPH} &= \sum_{i \in Q} \left(d_i - \frac{1}{2} \sum_{j \in Q} aa_{ij} \right) - \sum_{i \in Q} \left(a_i - \frac{1}{2} \sum_{j \in Q} da_{ij} \right) \\ &= \sum_{i \in Q} (d_i^* - a_i^*). \end{aligned}$$

BPH can be estimated using phenotypic data $-H_2$ (i.e., backcrosses of RILs to P2). If we assume complete linkage between the marker m and one of the QTL contributing to BPH ($r_{mi} = 0$), we obtain the following genotypic expectations for the contrast of the two homozygous genotypic marker classes at marker m :

$$\mathcal{E}(-H_2(m) | r_{mi} = 0) = d_i^* - a_i^* + \sum_{j \in Q_j} 4D_{ij}(d_j^* - a_j^*). \quad (15)$$

Using CIM for QTL mapping with $-H_2$, the position of the QTL i can be estimated and the effect of the linked QTL j is blocked by the use of cofactors. Hence, analogously to $Z_2(i)$ the genotypic expectation for $-H_2(i)$, the contrast of $-H_2$ between the two unobservable homozygous genotype classes at QTL i , reduces to $\mathcal{E}(-H_2(i)) = d_i^* - a_i^*$.

EXTENSION OF ANALYSES TO QUADRIGENIC EPISTASIS

Considering the special case of quadrigenic epistasis and using Equation 1, the genotypic expectations of PD and MPH can be calculated as

$$\begin{aligned} \text{PD} &= \sum_{i \in Q} 2a_i^* + \frac{1}{2} \sum_{i \in Q} \sum_{\{j,k,l\} \subset Q_j} add_{ijk} + \frac{2}{3} \sum_{i \in Q} \sum_{\{j,k\} \subset Q_j} aaa_{ijk} \\ &\quad - \sum_{i \in Q} \sum_{\{j,k,l\} \subset Q_j} daaa_{ijkl} - \frac{1}{4} \sum_{i \in Q} \sum_{\{j,k,l\} \subset Q_j} addd_{ijkl} \\ \text{MPH} &= \sum_{i \in Q} d_i^* + \frac{1}{2} \sum_{i \in Q} \sum_{\{j,k\} \subset Q_j} daa_{ijk} + \frac{1}{12} \sum_{i \in Q} \sum_{\{j,k\} \subset Q_j} ddd_{ijk} \\ &\quad - \frac{1}{4} \sum_{i \in Q} \sum_{\{j,k,l\} \subset Q_j} aaaa_{ijkl} - \frac{1}{8} \sum_{i \in Q} \sum_{\{j,k,l\} \subset Q_j} aadd_{ijkl}. \end{aligned}$$

It becomes obvious that effects of higher-order epistasis contribute to PD and MPH. Even if individual effects are small, their summed effects may be large. In the case of

quadrigenic interactions, the sum of effects $aaaa$ comprises $\binom{q}{4}$ individual effects $aaaa_{ijkl}$ and the sum of $aadd$ effects comprises $\binom{q}{2} \binom{q-2}{2}$ individual effects $aadd_{ijkl}$. When looking at expectations of the marker contrast at marker m for pair means Z_1 and pair differences Z_2 and assuming $r_{mi} = 0$ and linkage equilibrium between QTL, we obtain

$$\mathcal{E}(Z_1(m) | r_{mi} = 0) = a_i^* + \frac{1}{4} \sum_{\{j,k\} \subset Q_j} aaa_{ijk} - \frac{1}{8} \sum_{\{j,k,l\} \subset Q_j} daaa_{ijkl}$$

$$\mathcal{E}(Z_2(m) | r_{mi} = 0) = 2d_i^* + \frac{1}{2} \sum_{\{j,k\} \subset Q_j} daa_{ijk} - \frac{1}{4} \sum_{\{j,k,l\} \subset Q_j} aaaa_{ijkl}.$$

Thus, the estimate $b_2^*/2$ from CIM yields QTL effect estimates that account not only for d_i^* but also for the contribution of daa and $aaaa$ interactions of QTL i with the genetic background. However, compared with the contribution of digenic interactions to MPH, effects of type daa and $aaaa$ are only half accounted for and effects ddd and $aadd$ are not accounted for at all when estimating the heterotic effect at QTL i with CIM. Similar results are obtained for a comparison between genotypic expectations of PD and $\mathcal{E}(Z_1(m) | r_{mi} = 0)$ that differ mainly in the contribution of effects dda and $addd$. With arbitrary linkage disequilibrium among QTL the expectations above become rather unwieldy, but it can be shown that b_s^* explains a larger proportion of all types of higher-order epistatic effects contributing to midparent heterosis or to the parental difference if linkage between QTL is present.

GENETIC CONSTITUTION OF VARIANCES

Generalized derivations of the genetic constitution of variances $\sigma_{Z_1}^2$ and $\sigma_{Z_2}^2$ for two QTL are given in the APPENDIX. Contributions of genetic effects to variances of pair means ($\sigma_{Z_1}^2$) and pair differences ($\sigma_{Z_2}^2$) for RILs summed over all QTL assuming arbitrary linkage and digenic epistasis are presented in Table 2. If segregating QTL show intermediate linkage, the linkage disequilibrium coefficient D_{ij} and consequently genotypic expectations of $\sigma_{Z_1}^2$ and $\sigma_{Z_2}^2$ differ for RILs and double-haploid lines (DHLs), but deviations are small (see APPENDIX). For unlinked ($r_{ij} = 0.5$) and completely linked ($r_{ij} = 0$) loci, variances for RILs and DHLs are identical.

Using the definitions of a_i^* given in Equation 7 and d_i^* given in Equation 4 and defining the quadratic forms

$$s_{A^*}^2 = \sum_{i \in Q} a_i^{*2} + 4 \sum_{i \in Q} \sum_{j \in Q} D_{ij} a_i^* a_j^* \quad (16)$$

and

$$s_{D^*}^2 = \sum_{i \in Q} d_i^{*2} + 4 \sum_{i \in Q} \sum_{j \in Q} D_{ij} d_i^* d_j^*, \quad (17)$$

TABLE 2

Contribution of genetic effects to variances of pair means ($\sigma_{Z_1}^2$) and pair differences ($\sigma_{Z_2}^2$) of recombinant inbred lines (RILs) and F₂ progenies backcrossed to parental lines P1 and P2, assuming linkage and digenic epistasis

Parameter	Contribution of genetic effects
$\sigma_{Z_1}^2$ (RIL)	$\frac{1}{4} \sum_{i \in Q} \left[a_i - \frac{1}{2} \sum_{j \in Q} da_{ij} \right]^2 + \sum_{i \in Q} \sum_{j \in Q} D_{ij} \left[a_i - \frac{1}{2} \sum_{k \in Q} da_{ik} \right] \left[a_j - \frac{1}{2} \sum_{k \in Q} da_{jk} \right] + \frac{1}{32} \sum_{i \in Q} \sum_{j \in Q} \left(1 - 16D_{ij}^2 \right) [aa_{ij} + dd_{ij}]^2$ $= \frac{1}{4} s_{A^*}^2 + \frac{1}{32} \sum_{i \in Q} \sum_{j \in Q} \left(1 - 16D_{ij}^2 \right) [aa_{ij} + dd_{ij}]^2$
$\sigma_{Z_1}^2$ (F ₂)	$\frac{1}{8} s_{A^*}^2 + \frac{1}{128} \sum_{i \in Q} \sum_{j \in Q} \left[1 - 16D_{ij}^2 + 32D_{ij}^2 (1 - 4D_{ij}) \right] [aa_{ij} + dd_{ij}]^2$
$\sigma_{Z_2}^2$ (RIL)	$\sum_{i \in Q} \left(d_i - \frac{1}{2} \sum_{j \in Q} aa_{ij} \right)^2 + 4 \sum_{i \in Q} \sum_{j \in Q} D_{ij} \left(d_i - \frac{1}{2} \sum_{k \in Q} aa_{ik} \right) \left(d_j - \frac{1}{2} \sum_{k \in Q} aa_{jk} \right) + \frac{1}{8} \sum_{i \in Q} \sum_{j \in Q} \left(1 - 16D_{ij}^2 \right) [ad_{ij} + da_{ij}]^2$ $= s_{D^*}^2 + \frac{1}{8} \sum_{i \in Q} \sum_{j \in Q} \left(1 - 16D_{ij}^2 \right) [ad_{ij} + da_{ij}]^2$
$\sigma_{Z_2}^2$ (F ₂)	$\frac{1}{2} s_{D^*}^2 + \frac{1}{32} \sum_{i \in Q} \sum_{j \in Q} \left[1 - 16D_{ij}^2 + 32D_{ij}^2 (1 - 4D_{ij}) \right] [ad_{ij} + da_{ij}]^2$

the variances $\sigma_{Z_1}^2$ and $\sigma_{Z_2}^2$ can be expressed as shown in Table 2. For a direct comparison with the results obtained for F₂ progenies, Table 2 also presents variances $\sigma_{Z_1}^2$ (F₂) and $\sigma_{Z_2}^2$ (F₂), which can be readily transformed into those presented by COCKERHAM and ZENG (1996) with $\sigma_m^2 = \sigma_\alpha^2 = \sigma_{Z_1}^2$ (F₂) and $\sigma_{ml}^2 = \frac{1}{2} \sigma_\delta^2 = \frac{1}{2} \sigma_{Z_2}^2$ (F₂). Regardless of the type of population used for producing the testcrosses (*i.e.*, RIL or F₂), it can be seen that the quadratic forms $s_{A^*}^2$ and $s_{D^*}^2$ are the main components of variances $\sigma_{Z_1}^2$ and $\sigma_{Z_2}^2$, respectively. The bias due to digenic epistatic interactions of types *aa* and *dd* ($\sigma_{Z_1}^2$) as well as *ad* and *da* ($\sigma_{Z_2}^2$) is small, especially for F₂, as pointed out also by COCKERHAM and ZENG (1996). Thus, we postulate that $\sigma_{Z_1}^2$ (RIL) and $\sigma_{Z_2}^2$ (RIL) are reasonable approximations for $1/4 s_{A^*}^2$ and $s_{D^*}^2$, respectively. As a result, the estimate of $\sigma_{Z_2}^2$ obtained from the ANOVA of design III is a close approximation of the variance of heterotic effects at QTL segregating in the cross P1 \times P2.

AVERAGE DEGREE OF DOMINANCE

Originally, design III was devised to provide an estimate of the average degree of dominance over loci calculated from the ratio of dominance to additive variance (\bar{D}). The ratio $\sqrt{\sigma_{ml}^2/2\sigma_m^2}$ proposed by COMSTOCK and ROBINSON (1952) is equivalent to $\sqrt{\sigma_{Z_2}^2/4\sigma_{Z_1}^2}$ and is an approximation of $\sqrt{s_{D^*}^2/s_{A^*}^2}$, the ratio of quadratic forms $s_{D^*}^2$ and $s_{A^*}^2$ of augmented dominance (d_i^*) and additive (a_i^*) effects, respectively, rather than of the ratio of dominance and additive variance. Therefore, the ratio $\sqrt{\sigma_{ml}^2/2\sigma_m^2}$ should be denoted the augmented degree of dominance \bar{D}^* . Estimation of \bar{D}^* can be biased by linked QTL, if linkage equilibrium among them has not been reached. Genetic effects at linked QTL contributing to $s_{A^*}^2$ (*i.e.*, $4D_{ij}a_i^*a_j^*$) have the same sign when loci are in coupling, while signs are different when loci are in repulsion. If two elite parents are crossed and coupling and repulsion linkages occur with equal probabilities, the effects of linked loci are likely to cancel in $s_{A^*}^2$. However, estimates of $s_{D^*}^2$ are likely to be inflated because D_{ij} is

positive by definition and in hybrid breeding, loci with high positive augmented dominance effects (d_i^*) are favored in reciprocal recurrent selection. Thus, the contribution of genetic effects at linked loci to $s_{D^*}^2$ (*i.e.*, $4D_{ij}d_i^*d_j^*$) is generally positive irrespective of their linkage phase and \bar{D}^* will be strongly affected by the presence of epistasis and the magnitude of the linkage disequilibrium between QTL. The numerical example in supplemental Table S2 (<http://www.genetics.org/supplemental/>) clearly shows the effect of *aa* epistasis on estimates of \bar{D}^* for unlinked loci. Altogether, the ratio of variance components obtained with design III is not always a useful estimate of the average degree of dominance, thus questioning the interpretation of many experimental results on the relative importance of dominance, overdominance, and epistasis in the expression of quantitative traits.

DISCUSSION

Genotypic components of trait performance and heterosis: Elucidating the genetic basis of heterosis has been the aim of a number of studies making use of advances in molecular biology. Some studies compared specific molecular traits such as differential gene expression in the F₁ hybrid and the parental inbred lines (*e.g.*, GUO *et al.* 2006; SWANSON-WAGNER *et al.* 2006). A number of authors used testcross progenies of F₃ lines or RILs with the parental inbred lines as testers for QTL studies on heterosis (STUBER *et al.* 1992; XIAO *et al.* 1995; LI *et al.* 2001; LUO *et al.* 2001). Estimation of QTL effects was separately performed for the backcrosses to each parental line, except for a recent study with triple testcross progenies in maize, which used also 2Z₁ and Z₂ for QTL mapping (FRASCAROLI *et al.* 2007). For the backcrosses to the better-performing parent, this corresponds to the marker contrast $-H_2(m)$ defined in Equation 15. Thus, with interval mapping (LANDER and BOTSTEIN 1989) their analyses yielded estimates for QTL contributing to BPH ($d_i^* - a_i^*$) for the progenies backcrossed to one parent and for poorer parent heterosis ($d_i^* + a_i^*$) for

the progenies backcrossed to the other parent. Consequently, conclusions on gene action at individual QTL can be made only within this context.

SEMEL *et al.* (2006) used a set of tomato introgression lines to exclude the confounding effects of genetic background variation and epistasis from the analysis of heterosis. Parental genotypes and the F₁ hybrid differed exclusively in one defined chromosome segment while the entire genetic background originated from the elite parent. For a multitude of traits, the authors compared the phenotypic means of the elite parent, the homozygous introgression line, and the hybrid between the two. On the basis of these comparisons, the type of gene action at QTL was determined. However, as pointed out earlier, trait performance and heterosis have different genotypic expectations. Employing the quantitative genetic theory derived in this article, it can be shown in terms of the F₂ metric that genotypic expectations of QTL effects contributing to MPH and BPH comprise both main and epistatic effects, despite the fact that the entire genetic background originates from the elite parent (MELCHINGER *et al.* 2007).

The use of an immortalized F₂ population was proposed by HUA *et al.* (2003) for identifying genomic regions that contribute to MPH. RILs were derived from a heterotic rice cross and intermated for construction of an immortalized F₂ population. Heterosis was calculated as the phenotypic deviation of each immortal F₂ from the mean of its two RIL parents. Digenic interactions were estimated on the basis of the interaction effect of two marker loci. Genotypic expectations of the QTL main effects and interactions were not given by the authors. On the basis of our findings, we postulate that the immortalized F₂ design has great value for estimating the dominance effect d_i and certain types of digenic epistatic interactions but does not identify QTL with genotypic expectations that equal precisely d_i^* , their contribution to MPH (derivations not shown).

We conclude that none of the available experimental designs of quantitative genetics has the potential to separate the dominance effect d_i and half the sum of *aa* epistatic interactions confounded in the augmented dominance effect d_i^* of heterotic QTL. Design III identifies heterotic QTL but does not allow separation of dominance and *aa* epistasis. For the time being, the only solution to this problem is to identify QTL contributing to heterosis with design III and to estimate the augmented QTL effect d_i^* . In genomic regions exhibiting significant heterosis, the augmented dominance effect d_i^* can be dissected into its components (d_i and $-\frac{1}{2}\sum_{j \in Q} aa_{ij}$) by employing additionally the immortal F₂ design and estimating the magnitude of dominance effect d_i . With RILs, the same lines can be used for generating the progenies for design III and the immortal F₂ population, and marker data generated for the RILs can readily be employed in both types of analysis. The optimal dimensions of experimental studies applying our approach with re-

spect to population size and population type as well as marker density have yet to be determined. Epistatic interactions of QTL i with the genetic background are expected to vary considerably across RILs. Implications of this biological variation on estimates of d_i^* will be addressed in a separate study.

Epistasis and heterosis: As has been shown in this study, even small epistatic interactions can be important for the expression of heterosis because their contribution to the genotypic expectation of augmented QTL effects sums up over many effects. On the basis of results from classical quantitative genetic experiments, we have reason to assume that epistasis plays an important role in the inheritance of quantitative traits and heterosis (*e.g.*, JINKS and JONES 1958) but results from QTL studies on the importance of epistasis have been rather ambiguous (*e.g.*, STUBER *et al.* 1992; COCKERHAM and ZENG 1996; FRASCAROLI *et al.* 2007). In QTL analyses, the power of detecting QTL with epistatic effects is generally low mainly due to the problem of multiple testing in two- or multidimensional genome scans (LANDER and BOTSTEIN 1989) or due to the necessity of a *a priori* model selection with one-dimensional scans (KAO *et al.* 1999). We need to keep in mind that QTL with significant epistatic interaction effects might not be representative for the majority of QTL with small effects contributing to gene networks that control the expression of quantitative traits. Hence, we are likely to introduce an ascertainment bias as pointed out by KROYMANN and MITCHELL-OLDS (2005).

Type of epistasis: When estimating the most prominent type of epistatic interactions in the expression of quantitative traits and heterosis in rice, a preponderance of *aa* epistatic effects was identified compared with *ad* or *dd* interactions (YU *et al.* 1997; HUA *et al.* 2003). In self-pollinated crops, it is well known that coadapted gene complexes are favored by selection. As a consequence, half the sum of mainly positive *aa* interactions enters d_i^* with a negative sign, thus decreasing MPH and the power for detection of heterotic QTL. It is questionable if the results on heterosis from self-pollinated crops are directly applicable to cross-pollinated crops (FRASCAROLI *et al.* 2007). Economic seed production, however, requires the development of inbred lines with high grain yield. Combined with the management of separate heterotic pools, it is highly probable that coadapted gene complexes are selected during inbred line development. As a consequence, if the sum of *aa* epistatic interactions in the parents increases due to selection, MPH should decrease over time unless the sum of dominance effects at QTL influencing heterosis increases proportionally. A decrease in relative superiority of hybrids compared with their inbred lines has been described for maize (DUVICK 1999). This decrease in relative heterosis can be the result of the accumulation of favorable dominant alleles at individual QTL, but it can also be explained by overdominance in the presence of *aa* epistatic effects contributed by the parents. If this is the case, the out-

come of marker-assisted selection programs aiming at the transfer of QTL for maximization of heterosis will strongly depend on the presence of favorable epistatic interactions with the genetic background in the respective germplasm and will be difficult to predict.

In conclusion, the results presented here are important in two ways. First, they provide the quantitative genetic theory to express heterosis as the sum of individual QTL effects. Second, they allow the assessment of epistatic interactions of individual QTL with the entire genetic background, thus extending the concept of epistasis from single-gene to system-level interactions. We suggest the use of CIM and design III with RILs to identify QTL expressing maximum heterosis (*i.e.*, maximum d_i^*). All analyses can be performed with an extended version of the software PLABQTL (UTZ and MELCHINGER 1996; <http://www.uni-hohenheim.de/plantbreeding/software/index.html>). Permutation tests for determining the significance threshold (DOERGE and CHURCHILL 1996) and cross-validation for unbiased QTL estimation (UTZ *et al.* 2000) can be readily applied. Using the joint likelihood-ratio test for augmented effects a_i^* and d_i^* will improve identification of heterotic QTL from elite \times exotic crosses and provide a first test to distinguish between dominance and overdominance. Applying models accounting for multilocus epistasis and using molecular tools to finely dissect genomic regions contributing to heterosis will allow an assessment of the relative contribution of epistatic interactions in the manifestation of heterosis.

We thank two anonymous reviewers for their valuable contributions. This project was supported by the Deutsche Forschungsgemeinschaft (German Research Foundation) under the priority research program "Heterosis in Plants" (research grants ME931/4-1 and ME931/4-2, PI 377/7-1 and PI 377/7-2).

LITERATURE CITED

- COCKERHAM, C. C., 1954 An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* **39**: 859–882.
- COCKERHAM, C. C., and Z. B. ZENG, 1996 Design III with marker loci. *Genetics* **143**: 1437–1456.
- COMSTOCK, R. E., and H. F. ROBINSON, 1952 Estimation of average dominance of genes, pp. 494–516 in *Heterosis*, edited by J. W. GOWEN. Iowa State College Press, Ames, IA.
- DOERGE, R. W., and G. A. CHURCHILL, 1996 Permutation tests for multiple loci affecting a quantitative character. *Genetics* **142**: 285–294.
- DUVICK, D. N., 1999 Heterosis: feeding people and protecting natural resources, pp. 19–29 in *Genetics and Exploitation of Heterosis*, edited by J. G. COORS and S. PANDEY. American Society of Agronomy, Madison, WI.
- EDWARDS, M. D., C. W. STUBER and J. F. WENDEL, 1987 Molecular-marker-facilitated investigations of quantitative trait loci in maize. I. Numbers, genomic distribution and types of gene action. *Genetics* **116**: 113–125.
- FALCONER, D. S., and T. F. C. MACKAY, 1996 *Introduction to Quantitative Genetics*. Longman Science and Technology, Harlow, UK.
- FISHER, R. A., 1918 The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* **52**: 399–433.
- FRASCAROLI, E., M. A. CANÉ, P. LANDI, G. PEA, L. GIANFRANCESCHI *et al.*, 2007 Classical genetic and quantitative trait loci analyses of heterosis in a maize hybrid between two elite inbred lines. *Genetics* **176**: 625–644.
- FRISCH, M., and A. E. MELCHINGER, 2006 Marker-based prediction of the parental genome contribution to inbred lines derived from biparental crosses. *Genetics* **174**: 795–803.
- GUO, M., M. A. RUPE, X. YANG, O. CRASTA, C. ZINSELMEIER *et al.*, 2006 Genome-wide transcript analysis of maize hybrids: allelic additive gene expression and yield heterosis. *Theor. Appl. Genet.* **113**: 831–845.
- HUA, J., Y. XING, W. WU, C. XU, X. SUN *et al.*, 2003 Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. *Proc. Natl. Acad. Sci. USA* **100**: 2574–2579.
- JANSEN, R. C., and P. STAM, 1994 High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* **136**: 1447–1455.
- JIANG, C., and Z. B. ZENG, 1995 Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics* **140**: 1111–1127.
- JINKS, J. L., and R. M. JONES, 1958 Estimation of components of heterosis. *Genetics* **43**: 223–234.
- KAO, C. H., Z.-B. ZENG and R. D. TEASDALE, 1999 Multiple interval mapping for quantitative trait loci. *Genetics* **152**: 1203–1216.
- KEARSEY, M. J., and H. S. POONI, 1996 *The Genetical Analysis of Quantitative Traits*. Chapman & Hall, London.
- KROYMANN, J., and T. MITCHELL-OLDS, 2005 Epistasis and balanced polymorphism influencing complex trait variation. *Nature* **435**: 95–98.
- KUSTERER, B., H.-P. PIEPHO, H. F. UTZ, C. C. SCHÖN, J. MUMINOVIC *et al.*, 2007 Heterosis for biomass-related traits in Arabidopsis investigated by a quantitative trait loci analysis of the triple testcross design with recombinant inbred lines. *Genetics* **177**: 1839–1850.
- LAMKEY, K. R., and J. W. EDWARDS, 1999 The quantitative genetics of heterosis, pp. 31–48 in *The Genetics and Exploitation of Heterosis in Crops*, edited by J. G. COORS and S. PANDEY. Crop Science Society of America, American Society of America, and Soil Science Society of America, Madison, WI.
- LANDER, E. S., and D. BOTSTEIN, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185–199.
- LI, Z. K., L. J. LUO, H. W. MEI, D. L. WANG, Q. Y. SHU *et al.*, 2001 Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. *Genetics* **158**: 1737–1753.
- LUO, L. J., Z. K. LI, H. W. MEI, Q. Y. SHU, R. TABIEN *et al.*, 2001 Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. II. Grain yield components. *Genetics* **158**: 1755–1771.
- LYNCH, M., and B. WALSH, 1998 *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.
- MELCHINGER, A. E., H.-P. PIEPHO, H. F. UTZ, J. MUMINOVIC, T. WEGENAST *et al.*, 2007 Genetic basis of heterosis for growth-related traits in Arabidopsis investigated by testcross progenies of near-isogenic lines reveals a significant role of epistasis. *Genetics* **177**: 1827–1837.
- SCHÖN, C. C., F. H. UTZ, S. GROH, B. TRUBERG, S. OPENSHAW *et al.*, 2004 Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics* **167**: 485–498.
- SEMEL, Y., J. NISSENBAUM, N. MENDA, M. ZINDER, U. KRIEGER *et al.*, 2006 Overdominant quantitative trait loci for yield and fitness in tomato. *Proc. Natl. Acad. Sci. USA* **103**: 12981–12986.
- STUBER, C. 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.
- SWANSON-WAGNER, R. A., Y. JIA, R. DECOOK, L. A. BORSUK, D. NETTLETON *et al.*, 2006 All possible modes of gene action are observed in a global comparison of gene expression in a maize F₁ hybrid and its inbred parents. *Proc. Natl. Acad. Sci. USA* **103**: 6805–6810.
- UTZ, H. F., and A. E. MELCHINGER, 1996 PLABQTL: a program for composite interval mapping of QTL. *J. Agric. Genomics* **2**: 1–5.
- UTZ, H. F., A. E. MELCHINGER and C. C. SCHÖN, 2000 Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. *Genetics* **154**: 1839–1849.

VAN DER VEEN, J. H., 1959 Tests of non-allelic interaction and linkage for quantitative characters in generations derived from two diploid pure lines. *Genetica* **30**: 201–232.
 WEIR, B. S., 1996 *Genetic Data Analysis II*. Sinauer Associates, Sunderland, MA.
 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.
 YANG, R. C., 2004 Epistasis of quantitative trait loci under different gene action models. *Genetics* **167**: 1493–1505.
 YU, S. B., J. X. LI, C. G. XU, Y. F. TAN, Y. J. GAO *et al.*, 1997 Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proc. Natl. Acad. Sci. USA* **94**: 9226–9231.
 ZENG, Z. B., 1994 Precision mapping of quantitative trait loci. *Genetics* **136**: 1457–1468.

Communicating editor: J. B. WALSH

APPENDIX: GENERAL DERIVATION OF EXPECTATIONS, VARIANCES, AND COVARIANCES OF PAIR MEANS (Z_1) AND PAIR DIFFERENCES (Z_2) AS WELL AS MARKER CONTRASTS $Z_1(m)$ AND $Z_2(m)$

Let $\tau_{AD}(t)$ denote the coefficient of parameter α_{AD} in the conditional genotypic expectation of the testcross progeny H_t ($t = 1, 2$) of a RIL with genotype v_i at the i th QTL. Then, we obtain from Equation 1

$$\tau_{AD}(t) = \prod_{i \in A} x_{v,i}(t) \prod_{k \in D} y_{v,k}(t), \quad (A1)$$

with

$$x_{v,i}(t) = \frac{1}{2}v_i + t - 2$$

and

$$y_{v,i}(t) = 2v_i - \frac{1}{4}v_i^2 + v_it - t^2 + 4t - 3.5$$

Let \mathbf{E} denote the vector of genetic effects α_{AD} for two QTL i and j under digenic epistasis; *i.e.*, $\mathbf{E}' = (\mu, a_i, d_i, a_j, d_j, aa_{ij}, ad_{ij}, da_{ij}, dd_{ij})$ and $\mathbf{G}'_t = (\mathbf{G}_{22,t}, \mathbf{G}_{20,t}, \mathbf{G}_{02,t}, \mathbf{G}_{00,t})$ denote the vector of conditional genotypic expectations of testcross progeny H_t ($t = 1, 2$) with design III given the parental RIL has genotype v_iv_j (with $v_i = 0, 2$; $v_j = 0, 2$). Then, $\mathbf{G}_t = \mathbf{H}_t\mathbf{E}$ with elements of the matrices \mathbf{H}_t equal to coefficients calculated according to Equation A1.

By calculating $\mathbf{K}_s = \sum_{i=1}^2 p_{st}\mathbf{H}_t$ with p_{st} equal to the s th element of

$$\mathbf{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & -1 \end{bmatrix}$$

for design III, we get $\mathbf{Z}_s = \mathbf{K}_s\mathbf{E}$, where $\mathbf{Z}'_s = (Z_{22,s}, Z_{20,s}, Z_{02,s}, Z_{00,s})$ denotes the vector of conditional genotypic expectations of Z_s , given the genotype of the parental RIL.

To simplify formulas, to allow extension to multiple QTL, and to provide a generalized formula for RILs and DHLs, we use the parameter D_{ij} to quantify linkage disequilibrium between loci i and j . D_{ij} can be calculated from the recombination frequency r_{ij} between loci i and j by the formulas

TABLE A1

Frequencies ($f_{ij|m}$) of the four possible QTL genotypes of recombinant inbred line (RIL) parents at QTL i and j conditional on the marker genotype u at the marker locus m calculated using the linkage disequilibrium parameter D

Genotype at QTL		Genotype u at marker locus m : $f_{ij m}$	
i	j	MM ($u = 2$)	mm ($u = 0$)
2	2	$\frac{1}{4} + D_{mi} + D_{mj} + D_{ij}$	$\frac{1}{4} - D_{mi} - D_{mj} + D_{ij}$
2	0	$\frac{1}{4} + D_{mi} - D_{mj} - D_{ij}$	$\frac{1}{4} - D_{mi} + D_{mj} - D_{ij}$
0	2	$\frac{1}{4} - D_{mi} + D_{mj} - D_{ij}$	$\frac{1}{4} + D_{mi} - D_{mj} - D_{ij}$
0	0	$\frac{1}{4} - D_{mi} - D_{mj} + D_{ij}$	$\frac{1}{4} + D_{mi} + D_{mj} + D_{ij}$

$$D_{ij}(\text{RIL}) = \frac{1}{1 + 2r_{ij}} \frac{1 - 2r_{ij}}{4} (1 - r_{ij})^g$$

and

$$D_{ij}(\text{DHL}) = \frac{1 - 2r_{ij}}{4} (1 - r_{ij})^g,$$

with g being the number of random-mating generations prior to selfing for the development of RILs or production of DHLs (FRISCH and MELCHINGER 2006).

Under the assumption of Mendelian segregation we define

$$\mathbf{f}' = (f_{22}, f_{20}, f_{02}, f_{00})$$

$$= \frac{1}{4}(1 + 4D_{ij}, 1 - 4D_{ij}, 1 - 4D_{ij}, 1 + 4D_{ij})$$

and

$$\mathbf{F} = \text{diag}(f_{22}, f_{20}, f_{02}, f_{00}).$$

Expectations, variances, and covariances of Z_s are given by

$$\mathcal{E}(Z_s) = \mathbf{f}'\mathbf{K}_s\mathbf{E},$$

$$\sigma_{Z_s}^2 = \mathbf{E}'\mathbf{K}'_s(\mathbf{F} - \mathbf{f}\mathbf{f}')\mathbf{K}_s\mathbf{E},$$

and

$$\text{cov}(Z_s, Z_t) = \mathbf{E}'\mathbf{K}'_s(\mathbf{F} - \mathbf{f}\mathbf{f}')\mathbf{K}_t\mathbf{E}.$$

Expectations of marker contrasts $Z_1(m)$ and $Z_2(m)$ for the case of two QTL i and j linked to marker locus m with marker classes u ($u = 0, 2$) are given as follows. We define the parental genotypes as $0_m0_i0_j$ (P1) and $2_m2_i2_j$ (P2). Recombination frequencies between the three loci are denoted r_{mis} , r_{mj} , and r_{ij} , respectively. The frequencies ($f_{ij|m}$) of the four possible QTL genotypes ($ij = 22, 20, 02, 00$) conditional on the marker genotype u at the marker locus m are given for RILs and DHLs in Table A1. Thus, we obtain the vector $\mathbf{f}'_m = (f_{22|m}, f_{20|m}, f_{02|m}, f_{00|m})$ and the conditional expectations of linear functions Z_s can be calculated as $Z_{s|m} = \mathbf{f}'_m\mathbf{K}_s\mathbf{E}$. From this, we obtain the orthogonal marker contrasts $Z_s(m) = Z_{s|m}(u = 2) - Z_{s|m}(u = 0) = (\mathbf{f}'_m(u = 2) - \mathbf{f}'_m(u = 0))\mathbf{K}_s\mathbf{E}$.