New Concepts For Mapping Heterosis In Plants

C.C. Schön[*] and A.E. Melchinger[†]

Introduction

During an invited lecture in Göttingen, Germany, in 1914 the American geneticist G. H. Shull proposed the term "heterosis" for the phenomenon that "the physiological vigor of an organism […] is positively correlated with the degree of dissimilarity in the gametes […]. These differences need not be Mendelian in their inheritance […]"(Shull 1948). He had coined the term "heterosis" with the intention that it should be free from every hypothesis and stated: "I hold it to be absolutely necessary to distinguish sharply between the facts and the theory derived from them" (Shull 1948). Today, 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 performance of the resulting F_1 hybrids. However, the understanding of this biological key phenomenon is limited and the genetic basis of heterosis has yet to be elucidated.

Three main hypotheses have been put forward to explain the genetic causes underlying heterosis (for review see Lamkey and Edwards 1999). The dominance hypothesis attributes heterosis to the cumulative effect of favorable alleles with complete or partial dominance. However, repulsion phase linkage of such genes may lead to pseudo-overdominance. The overdominance hypothesis assumes superiority of heterozygous genotypes over both parental homozygous genotypes. The epistasis hypothesis explains heterosis on the basis of interactions among genes at different loci. Different models have been developed to distinguish between different types of gene action and to estimate the magnitude of genetic effects. Anderson and Kempthorne (1954) and Gamble (1962) proposed the estimation of additive, dominance and epistatic effects from first moment statistics, *i.e.*, generation means. These parameters reflect sums of gene effects over all loci and, consequently, positive and negative effects at individual loci may cancel each other. Cockerham (1954) proposed a general model for estimating the type of gene action from second moment statistics. He developed a set of orthogonal contrasts to partition the genetic variance into additive, dominance and epistatic components. However, 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.

The first marker-aided studies estimated the predominant type of gene action at individual QTL in segregating F_2 populations based on contrasts of marker genotype classes. However, these estimates are not sufficient for making inferences about the genetic basis of midparent heterosis (MPH). When investigating the genetic causes of MPH we need to take into account, that MPH is defined as the deviation of the performance of the F_1 hybrid from the mean of its parents. Thus, MPH cannot simply be considered as a function of the degree of dominance at individual QTL. In the presence of epistasis the genetic expectation of MPH also comprises the additive × additive epistatic interactions of individual QTL with the entire genetic background.

In plant breeding, different experimental designs can be employed for estimating the type of gene action at individual QTL. One of the most powerful designs of classical quantitative genetics is the North Carolina Experiment III (design III) proposed by Comstock and Robinson (1952). Originally it was devised for estimating the average degree of dominance over all loci. Under the design III a random sample of F_2 individuals derived from a cross between two inbred lines is backcrossed to each of the two homozygous parental lines. For each F_2 individual the phenotypic mean and the difference between its backcross progenies can be calculated. An

[*] Plant Breeding, Technische Universität München, 85350 Freising, Germany
[†] Applied Genetics and Plant Breeding, Universität Hohenheim, 70599 Stuttgart, Germany

analysis of variance of the means and the differences yields estimates of the additive and dominance genetic variance with nearly equal precision and their ratio provides a weighted estimate of the squared degree of dominance. The design can be modified by using different types of populations for producing the backcrosses. In this paper we give genetic expectations of QTL effects obtained with backcrosses of recombinant inbred lines (RILs). We demonstrate their relevance for elucidating the genetic basis of MPH and apply our theory to the analysis of three experimental design III studies.

Theory

Genetic effects contributing to heterosis

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}):

$$\text{MPH} = G_{F_1} - \left(G_{P_1} + G_{P_2}\right)/2$$

Let P_1 and P_2 differ at the loci set Q = {1,.., 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. We can express the genotypic value of genotype $V = (v_1, \ldots, v_q)$ as

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

The parameters α_{AD} define the genetic effects of type additive at the loci set A ($A \subset Q$) and of type dominance at the loci set D ($D \subset Q_A$, Q_A being the complement of A in Q), and $\sum_{A \subset Q}$ indicating summation over all possible subsets A within the set Q .

For populations derived from two inbred lines we apply the F_2 metric (van der Veen, 1959). Thus, 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.

Let us assume an F_2 individual and three QTL with $V = (v_i, v_j, v_k) = (2, 1, 0)$. Then, $x_{V,i} = 1$, $x_{V,j} = 0$, $x_{V,k} = -1$, $y_{V,i} = -½$, $y_{V,j} = ½$, and $y_{V,k} = -½$. Thus,

$$G_V = \mu + a_i - a_k - ½d_i + ½d_j - ½d_k - aa_{ik} + ½ad_{ij} - ½ad_{ik} + ½da_{ik} - ½da_{jk} - ¼dd_{ij} + ¼\, dd_{ik} - ¼dd_{jk} - ½ada_{ijk} - ¼add_{ijk} + ¼\, dda_{ijk} + ⅛ddd_{ijk}$$

The parameter μ denotes the genotypic expectation of the F_2 generation in linkage equilibrium, 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 P_2 and negative ($-a_i$) when contributed by P_1 and epistatic interactions between loci *i*, *j* and *k* are denoted accordingly. Analogously, the genotypic values of the parental homozygous lines P_1 and P_2 and the F_1 hybrid are

$$G_{P_1} = \sum_{A \subset Q} \sum_{D \subset Q_A} (-1)^{|A|} \left(-\frac{1}{2}\right)^{|D|} \alpha_{AD} \,,\quad G_{P_2} = \sum_{A \subset Q} \sum_{D \subset Q_A} (1)^{|A|} \left(-\frac{1}{2}\right)^{|D|} \alpha_{AD} \text{ , and } G_{F_1} = \sum_{D \subset Q} \left(\frac{1}{2}\right)^{|D|} \alpha_D \,,$$

where $|A|$ and $|D|$ denote the number of elements in set *A* and *D*, respectively. Then, MPH can be calculated as the deviation of the genotypic value of an F_1 hybrid and the mean genotypic value of its two homozygous parents

$$\text{MPH} = G_{F_1} - \left(G_{P_1} + G_{P_2}\right)/2 = \sum_{\substack{\text{D}\subset\text{Q} \\ |\text{D}|\text{odd}}} \left(\frac{1}{2}\right)^{|\text{D}|-1} \alpha_{\text{D}} - \sum_{\substack{\text{A}\subset\text{Q} \\ |\text{A}|\text{even}}} \sum_{\text{D}\subset\text{Q}_\text{A}} \left(-\frac{1}{2}\right)^{|\text{D}|} \alpha_{\text{AD}} \,.$$

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*, *aad* but not *ad* and *aaa*) contribute to MPH. Considering only digenic epistasis we obtain:

$$\text{MPH} = \sum_{i \in Q} \left(d_i - \frac{1}{2} \sum_{j \in Q_i} aa_{ij} \right) \text{ with } Q_i \text{ denoting the loci set } Q \text{ excluding element } i.$$

To express MPH as the sum of individual QTL effects, a new type of heterotic genetic effect d_i* denoted as augmented dominance effect is defined. It comprises the dominance effect of QTL *i* (d_i) minus half the sum of its additive × additive epistatic interactions (aa_{ij}) with all other QTL irrespective of linkage. The augmented dominance effect d_i* corresponds exactly to the net contribution of QTL *i* to MPH.

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

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

$$\text{MPH} = \sum_{i \in Q} d_i^* \,.$$

Analogously, the quantitative genetic expectation of the parental difference (PD) is found:

$$\text{PD} = G_{P_2} - G_{P_1} = 2 \sum_{\substack{A \subset Q \\ |A| \text{odd}}} \sum_{D \subset Q_A} \left(-\frac{1}{2}\right)^{|D|} \alpha_{AD} \,.$$

Considering only digenic epistasis, PD reduces to

$$\text{PD} = \sum_{i \in Q} \left(2a_i - \sum_{j \in Q_i} da_{ij} \right) = \sum_{i \in Q} 2a_i^* \text{ , where } a_i^* = a_i - \frac{1}{2} \sum_{j \in Q_i} da_{ij} \,.$$

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 × additive epistatic interactions (da_{ij}) with all other QTL, corresponding exactly to the net contribution of QTL *i* to the parental difference.

Genotypic expectations of QTL effects

The design III was originally devised for the analysis of backcross progenies of F_2 individuals. In this study we present genotypic expectations of QTL effects of a design III experiment conducted with fully homozygous inbred lines. Let us assume a random population of RILs from the cross between the two homozygous lines P_1 and P_2. Further, we assume that the RILs are backcrossed to their parental lines, yielding backcross (BC) progenies H_1 and H_2. The parental line exhibiting superior average testcross performance is denoted as P_2. For each RIL, the linear transformation Z_s (s = 1, 2) is computed as $Z_1 = (H_1 + H_2)/2$ (mean of its BC progenies) and $Z_2 = (H_1 - H_2)/2$ (half the difference between its BC progenies). Performing composite interval mapping with given genotypic data and progeny mean values of Z_s as phenotypic data input, it can be shown that in one-dimensional genome scans on Z_1 and Z_2 the contrast of the two (unobservable) homozygous genotype classes at QTL i equals the augmented additive (a_i*) and dominance (d_i*) effects (Melchinger et al. 2007):

$$\mathrm{E}\left(Z_1(i)\right) = a_i - \frac{1}{2}\sum_{j \in Q_i} da_{ij} = a_i^*$$

$$\mathrm{E}\left(Z_2(i)\right) = d_i - \frac{1}{2}\sum_{j \in Q_i} aa_{ij} = d_i^*$$

Consequently, a genome scan with Z_2 localizes genomic regions affecting MPH and genetic expectations of QTL effects equal precisely the net contribution of each QTL to MPH. Accordingly, genetic expectations of QTL effects obtained with a genome scan with Z_1 reflect the contribution of QTL i to the parental difference. The theoretical results given are not only valid for RILs but can readily be applied to design III experiments conducted with F_2, F_3 or intermated F_2 progeny derived from the cross of two homozygous inbred lines.

Experimental material

The analysis of QTL for MPH presented here is based on phenotypic data for grain yield (GY) and grain moisture (GM) as well as genotypic data generated in three earlier studies in maize by Stuber et al. (1992), Lu et al. (2003), and Frascaroli et al. (2007). Experimental details of these studies are summarized in Table 1. Stuber et al. (1992) selfed F_3 plants of hybrid B73 × Mo17 (Pop1) and simultaneously crossed each plant with both parent lines according to a design III mating scheme to develop 264 pairs of BC progenies. The BC progenies were field evaluated at six locations in the USA with one replication at each location. The genotype of each F_3 plant was determined from its F_4 line for 76 markers. In the Lu et al. (2003) study, a design III experiment was performed with hybrid LH200 × LH216 (Pop2). The F_2 generation was intermated three times to produce the F_2Syn3 population. Individual plants from F_2Syn3 were crossed to the two parental inbreds. The 351 pairs of BC progenies were field evaluated at five locations in the USA with one replication per location. For genotyping, the 351 F_2Syn3 parental plants were assayed with 160 markers. Frascaroli et al. (2007) based their study on hybrid B73 × H99 (Pop3). A total of 142 recombinant inbred lines (RILs), derived from this hybrid through 12 selfing generations, were crossed with both parents. The RILs were genotyped and 158 loci were mapped in the genetic linkage map. Phenotyping was conducted at three locations in Italy with two replications per location.

QTL analyses

For localization of QTL on the genetic map we used linkage maps constructed in the original publications. Due to a limited number of common markers, joint map construction for the three populations was not possible. Marker positions from individual studies were projected on the IBM2 2008 neighbors reference map obtained from MaizeGDB (http://www.maizegdb.org).

For the map of Pop2 (Lu et al. 2003) appropriate transformations were conducted to account for three generations of intermating.
For Pop1 and Pop2 QTL analyses were performed with linear transformations Z_1 (mean across backcross progenies) and Z_2 (half the difference between backcross progenies) employing composite interval mapping of Jansen and Stam (1994) and Zeng (1994). Marker cofactors were selected by stepwise regression using the Bayes information criterion (BIC) computed according to the method of Burnham and Anderson (2004). The proportion of total genotypic variance for Z_1 or Z_2 explained by all detected QTL in the model (p) was obtained as the adjusted multiple correlation coefficient in the simultaneous fit (Utz et al. 2000). Unbiased proportions of genotypic variance explained by QTL (p_{TS}) were estimated with 5-fold cross-validation using 2000 splits for each trait according to the procedure described by Utz et al. (2000). QTL results for Pop3 were obtained from the literature.
If QTL were found in the same or adjacent bins across Pop1 and Pop2 for a given trait and linear transformation (Z_1 or Z_2), the respective genomic regions were examined for presence of QTL in Pop3. Genomic regions where congruent QTL were identified across populations for Z_2 of GY were examined in more detail. In these genomic regions, map distances for Pop1 and Pop2 were re-scaled based on map distances calculated from the IBM2 2008 Neighbors map (http://www.maizegdb.org) and QTL positions were compared based on the re-scaled maps. QTL were declared as congruent when the re-scaling resulted in localization of LOD peaks within 20 cM distance from each other according to the concept of overlapping bins suggested by Tuberosa et al. (2002).

Table 1: Experimental details of three design III studies in maize

Population	P_1 x P_2	N*	E	M
Pop1 Stuber et al. (1992)	B73 x Mo17	264 F_3	6	76
Pop2 Lu et al. (2003)	LH200 x LH216	351 F_2Syn3	5	160
Pop3 Frascaroli et al. (2007)	B73 x H99	142 RILs	3	158

* N = population size, E = number of environments, M = number of markers

Results and discussion

Theoretical results presented by Melchinger et al. (2007) showed that additive and dominance genetic effects estimated in classical QTL analyses are not sufficient for quantifying the contributions of QTL to heterosis, because additive × additive epistatic interactions of individual loci with the entire genetic background are a major component of MPH. Consequently, the identification of QTL contributing to the manifestation of MPH can only be achieved with specific experimental designs. For the design III it has been shown that genetic expectations of QTL effects precisely equal their net effect to midparent heterosis (Melchinger et al. 2007).
Experimental results from QTL mapping with three populations tested under the design III (Pop1, Pop2, Pop3) revealed high power of QTL detection and substantial congruency of QTL positions across populations for genome scans with Z_2 for GY. The linear transformation Z_2 (half the difference of phenotypic trait values of each pair of BC progenies) identifies heterotic

QTL with genetic expectation d_i* comprising the dominance effect (d_i) minus half the sum of additive × additive interactions with all other QTL (aa_{ij}). Results from genome scans with Z_2 for GY and GM are presented in Table 2. For GY, the number of significant (P < 0.01) QTL ranged from 10 (Pop1) to 13 (Pop2 and Pop3). In a simultaneous fit of all putative GY QTL for Z_2 63 to 78% of the genotypic variance could be explained in the three populations. Cross-validated estimates of the genetic variance explained for Z_2 of GY in Pop1 and Pop2 were surprisingly high, *i.e*., 66 and 58%, respectively. For eight QTL in Pop1 and five QTL in Pop2 the partial correlation coefficient calculated from the simultaneous fit of all QTL exceeded 5% (Schön et al. 2010).

Table 2: Number of detected QTL (Q) and estimates of the proportion genotypic variance explained by QTL in the data set (p_{DS}) and averaged across 2000 test sets from 5-fold cross validation (p_{TS}) for Z_1 and Z_2 for grain yield and grain moisture.

Population	Q	p_{DS}	p_{TS}	Q	p_{DS}	p_{TS}
Grain yield		Z_1			Z_2	
Pop1	2	28.2	15.4	10	77.8	66.4
Pop2	2	23.7	10.8	13	77.0	58.3
Pop3	7	39.5		13	63.0	
Grain moisture		Z_1			Z_2	
Pop1	2	45.3	24.8	3	45.8	20.2
Pop2	4	30.7	13.3	6	38.4	12.0
Pop3	6	34.9		0		

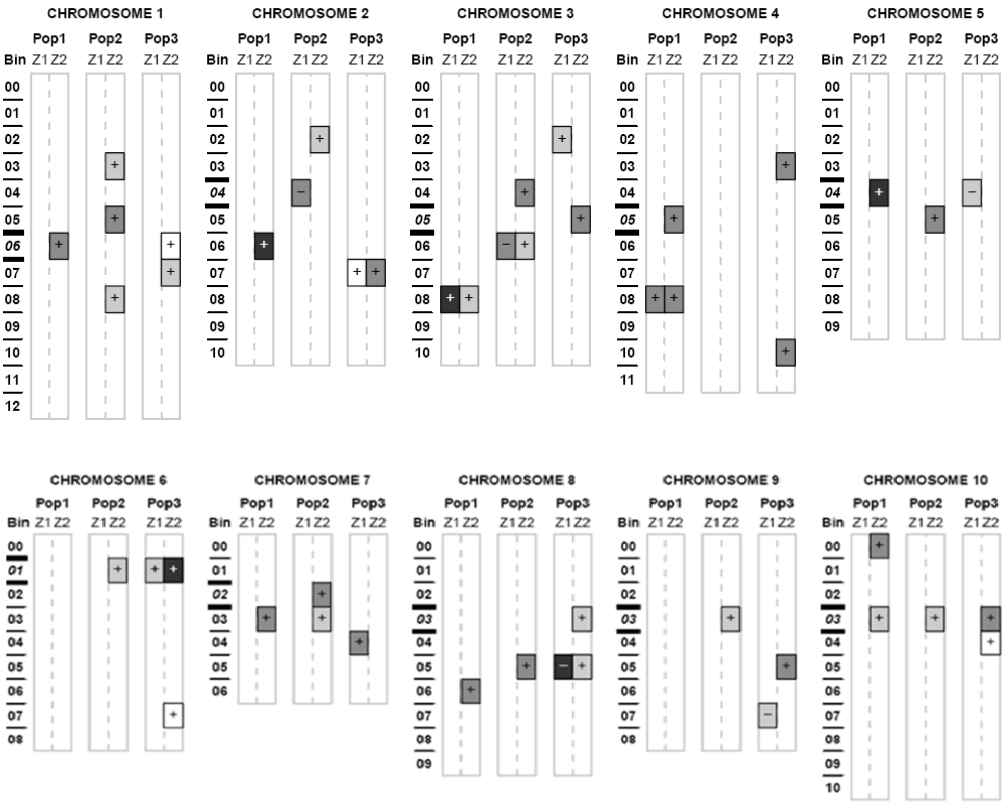
When comparing QTL across populations based on identical or adjacent bin positions, three genomic regions on Chromosome 1, 8 and 10 exhibited significant QTL for Z_2 of GY in all three populations (Figure 1). In two genomic regions on Chromosomes 5 and 7, matching QTL positions were also detected in Pop1 and Pop2.

The linear transformation Z_1 (mean of phenotypic trait values of each pair of BC progenies) identifies QTL with genetic expectation a_i* comprising the additive effect (a_i) minus half the sum of dominance × additive interactions with all other QTL (da_{ij}). Genome scans with Z_1 identified only few genomic regions significantly affecting GY with no apparent overlap across populations. After cross validation only between 11 and 16% of the genetic variance could be explained by QTL detected with genome scans with Z_1 and GY (Table 2).

In all three populations, the number of detected GY QTL differed substantially between Z_1 and Z_2. Most genomic regions exhibiting large d_i* effects did not show significant estimates of a_i*. In general, only few genomic regions with significant augmented additive effects were identified and detected QTL had small effects. We hypothesize that these findings point to pseudo-overdominance as a genetic explanation for heterosis. As a consequence of tight repulsion phase linkage of loci with dominant or partially dominant gene action, additive genetic effects of opposite sign cancel each other, thus, leading to subthreshold effects of QTL in genome scans with Z_1 and estimates for a_i* close to zero. On the other hand, due to high

selection pressure for GY MPH and divergent selection of heterotic pools specific alleles or allele clusters have been fixed within each heterotic group which in combination with the allele(s) from the opposite heterotic pool lead to a dramatic superiority of the heterozygote as compared to the parental mean, *i.e.* to high MPH for this trait. It is these QTL with large effects that can be readily detected in genome scans with Z_2. Due to the fixation of favorable alleles within heterotic groups and consequent "fixation" of allele combinations in inter-pool crosses, these genomic regions will not contribute to the variance of general and specific combining ability effects in segregating progenies derived from intra-pool crosses. Thus, when new inbred lines are developed from intra-pool crosses (as is the case in most maize breeding programs) these genomic regions will not contribute to genetic gain.

Figure 1: Chromosomal positions and magnitude of effects[a] of QTL detected for linear transformations Z_1 and Z_2 of GY in Pop1, Pop2 and Pop3 taken from Schön et al. (2010). The bin harboring the centromere is shown in italics in between bold lines. The +/- sign indicates the sign of a_i* (Z_1) and d_i* (Z_2) effects.



[a] Effect size: $p_i > 10\%$ $5\% < p_i \leq 10\%$ $2\% < p_i \leq 5\%$ $p_i \leq 2\%$

The hypothesis of pseudo-overdominance is further supported by the fact that in genomic regions congruent for Pop1 and Pop2 the ratio of $d_i^*/|a_i^*|$ was consistently smaller in Pop2 than in Pop1 presumably due to the three additional generations of intermating in Pop2 by which recombination should have at least partially broken up associations between genes linked in repulsion phase.

With the exception of chromosome 8, all congruent QTL were located in the same or an adjacent bin harboring the centromere. These genomic regions have been shown to have a high

density of genes and suppressed recombination (Gore et al. 2009). From the work of Hill and Robertson (1966) we know, that for linked QTL artificial selection in random-mating populations leads to an excess of repulsion phase gametes at fixation. This effect is expected to be stronger in regions of low recombination. Thus, we conclude that the localization of large, highly congruent QTL near the centromere as well as the strong directional augmented dominance effects and the low power of detecting augmented additive effects all point to pseudo-overdominance as a major cause for heterosis in this crop. It needs to be kept in mind however, that in the presence of epistasis, interpretation of $d_i^*/|a_i^*|$ with respect to partial, complete or overdominance is not straightforward. Different types of epistasis contribute to d_i^* and a_i^*. While it is mainly aa_{ij} epistasis contributing to d_i^*effects, it is mainly da_{ij} epistasis that contributes to a_i^*effects. Thus, the ratio $d_i^*/|a_i^*|$ is not an unbiased estimate of the degree of dominance.

For the second trait we investigated, grain moisture, few QTL were detected from genome scans with both linear transformations Z_1 and Z_2 (Table 2). After cross validation less than 25% of the genetic variance could be explained by the detected QTL. GM does generally not exhibit significant heterosis so results were according to expectations.

Conclusions

The genetic mapping of QTL for midparent heterosis requires specific experimental designs. Genotypic expectations of QTL effects must reflect the net contribution of individual QTL to MPH, a criterion which is only met by genetic effects estimated with the design III. Highly significant QTL for MPH can be identified in experimental populations of maize. QTL positions are highly congruent across populations due to high selection pressure on MPH for grain yield. The data from three experimental populations strongly support the hypothesis of pseudo-overdominance as a genetic explanation for the manifestation of midparent heterosis in maize.

References

Anderson, V.L., and Kempthorne, O. (1954) *Genetics* 39: 883-898.

Burnham, K.P., and Anderson, D.R. (2004) *Sociol. Methods Res.* 33: 261–304.

Cockerham, C.C. (1954) *Genetics* 39: 859-882.

Comstock, R.E., and Robinson, H.F. (1952) In *Heterosis*, ed. J.W. Gowen, pp. 494-516.

Frascaroli, E., Cane, M.A., Landi, P. *et al.* (2007) *Genetics* 176: 625–644.

Gamble, E.E. (1962) *Can J Plant Sci.* 42: 339–348.

Gore, M.A., Chia, J.M., Elshire, R.J. *et al.* (2009) *Science* 326: 1115–1117.

Hill, W.G., and Robertson, A. (1966) *Genet. Res. Camb*. 8:269-294.

Jansen, R.C., and Stam, P. (1994) *Genetics* 136: 1447–1455.

Lamkey, K.R., and Edwards J.W. (1999) In *The Genetics and Exploitation of Heterosis in Crops*, ed. J. G. Coors and S. Pandey. pp. 31-48.

Lu, H., Romero-Severson, J., and Bernardo, R. (2003) *Theor. Appl. Genet.* 107: 494–502.

Melchinger, A.E., Utz, H.F., Piepho, H.-P. *et al.* (2007) *Genetics* 117: 1815–1825.

Schön, C.C., Singh, B.S., Utz, H.F. *et al*. (2010) *Theor. Appl. Genet.* 120: 321-332.

Shull, G.H. (1948) *Genetics* 33: 439-446.

Stuber, C.W., Lincoln, S.E., Wolff, D.W. *et al.* (1992). *Genetics* 132: 823–839.

Tuberosa, R., Sanguineti, M.C., Landi, P. *et al.* (2002). *Plant Mol. Biol*. 48:697–712.

Utz, H. F., Melchinger, A.E., and Schön, C.C. (2000). *Genetics* 154: 1839–1849.

Van Der Veen, J.H. (1959) *Genetica* 30: 201–232.

Zeng, Z.-B. (1994) *Genetics* 136: 1457–1468.