Theor Appl Genet (2010) 120:311–319
DOI 10.1007/s00122-009-1193-0

ORIGINAL PAPER

Variation of the parental genome contribution in segregating populations derived from biparental crosses and its relationship with heterosis of their Design III progenies

Albrecht E. Melchinger · Baldev S. Dhillon · Xuefei Mi

Received: 4 August 2009 / Accepted: 20 October 2009 / Published online: 13 November 2009

Abstract The variation of the parental genome contribution (PGC) and its relationship with the genetic architecture of heterosis have received little attention. Our objectives were to (1) derive formulas for the variance of PGC in selfing, backcross (BC) or intermated generations produced from biparental crosses of homozygous parents, (2) investigate the correlation $r(Z_2, \Psi_M)$ of the PGC (Ψ_M) estimated by a set M of markers, with Z_2 (half the trait difference between each pair of BC progenies) in the Design III, and (3) interpret experimental results on this correlation with regard to the genetic basis of heterosis. Under all mating systems, the variance of PGC is smaller in species with a larger number and more uniform length of chromosomes. It decreases with intermating and backcrossing but increases under selfing. The ratio of variances of PGC in F_1DH (double haploids), F_2 and BC_1 populations is 4:2:1, but it is smaller in advanced selfing generations than expected for quantitative traits. Thus, altering the PGC by marker-assisted selection for the genetic background is more promising (i) in species with a smaller number and/or shorter chromosomes and (ii) in F_2 than in progenies of later selfing generations. The correlation $r(Z_2, \Psi_M)$ depends on the linkage relationships between M and the QTL influencing Z_2 as well as the augmented dominance effects d_i^* of the QTL, which include dominance and additive × additive effects with the genetic background, and sum up to mid-parent heterosis. From estimates of $r(Z_2, \Psi_M)$ as well as QTL studies, we conclude that heterosis for grain yield in maize is caused by the action of numerous QTL distributed across the entire genome with positive d_i^* effects.

We dedicate this article to Prof. Dr. H. Friedrich Utz on the occasion of his 70th birthday, in recognition of his numerous contributions to research on heterosis.

Communicated by A. Charcosset.

Contribution to the special issue "Heterosis in Plants".

A. E. Melchinger (✉) · B. S. Dhillon · X. Mi
Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, 70599 Stuttgart, Germany
e-mail: melchinger@uni-hohenheim.de

Introduction

The expectations of the parental genome contribution (PGC) in populations produced from biparental crosses are well known, but only limited information is available on the variation in PGC among genotypes in such populations. Hill (1993) derived the variance of the PGC among backcross (BC) individuals under the assumption of no interference in crossover formation. Adopting Hill's approach, Frisch and Melchinger (2007) derived the variance of PGC in homozygous lines developed by single seed descent (SSD) and double haploids (DH) produced from biparental crosses or backcrosses. They discussed the utility of the concept in plant variety protection and selection theory for marker-assisted background selection. However, formulas for the variance of PGC in intermated populations or not fully inbred generations developed by selfing or backcrossing have not been derived so far.

Investigations on the genetic and molecular basis of heterosis with modern genomics tools have recently received increasing attention in plant biology. Besides expression studies (cf. Paschold et al. 2010, Thiemann et al. 2010) and metabolite analyses (Lisec et al. 2009), quantitative genetic analyses with the North Carolina experiment III (Design III), devised by Comstock and

Robinson (1952), have played a prominent role in mapping quantitative trait loci (QTL) underlying heterosis in many plants (cf. Schön et al. 2010). Cockerham and Zeng (1996) developed statistical theory for simultaneous QTL mapping with both series of BC progenies in the Design III. Melchinger et al. (2007) developed general formulas showing the various types of genetic effects contributing to mid-parent heterosis (MPH). The QTL effects estimated by mapping with the linear contrast Z_2 (half the trait difference between each pair of BC progenies) are composed of the dominance effect as well as digenic epistatic effects with all other QTL in the genetic background and were denoted as augmented dominance effects d_i^*.

In a companion paper, Schön et al. (2010) used this new approach for QTL mapping with the Design III to dissect the genetic architecture of MPH in three crosses of maize. They detected several QTL with large d_i^* contributing to MPH for grain yield and reported a remarkable consistency of QTL positions across the crosses. Furthermore, they found in all three crosses a highly significant genotypic correlation ($r_g \approx 0.8$) of Z_2 with PGC in the genotypes used for producing the Design III progenies. However, they did not investigate the quantitative genetic causes underlying this correlation and possible inferences to be drawn from it with regard to the genetic basis of heterosis.

Our objectives were to (1) derive formulas for the variance of PGC in progenies of different selfing, backcross (BC) or intermated generations produced from biparental crosses of homozygous parents, (2) investigate the factors influencing the correlation $r(Z_2, \Psi_M)$ between PGC (Ψ_M), estimated by a set M of markers, and Z_2 in the Design III, and (3) interpret experimental results on this correlation reported in the literature with regard to inferences on the genetic architecture of heterosis.

Theory

Variance of parental genome contribution

We derive general formulas for the variance of PGC to genotypes in segregating populations generated without selection from biparental crosses of two homozygous parents P1 and P2. Adopting the notation of Cockerham and Weir (1973) and considering two loci i and j, we use P_{tu}^{rs} to denote the frequency of genotypes G_{tu}^{rs} formed by uniting maternal gametes $i_r j_s$ and paternal gametes $i_t j_u$. In order to trace the parental origin of the alleles at each locus, we assume that parent P1 carries allele 1 at each locus ($r = t = 1$ and $s = u = 1$) and P2 carries allele 2 at each locus ($r = t = 2$ and $s = u = 2$).

Let the random variable Ψ_i denote the PGC from parent P2 at locus i, i.e.,

$$\Psi_i = \begin{cases} 1 & \text{if } r = 2 \quad \text{and} \quad t = 2, \\ 0 & \text{if } r = 1 \quad \text{and} \quad t = 1 \text{ and} \\ 1/2 & \text{otherwise.} \end{cases}$$

With these definitions, we obtain for the covariance $\omega_{ij} = \mathrm{Cov}(\Psi_i, \Psi_j)$

$$\omega_{ij} = \sum_{r=1}^{2}\sum_{s=1}^{2}\sum_{t=1}^{2}\sum_{u=1}^{2} P_{tu}^{rs}\Psi_i\Psi_j - \left(\sum_{r=1}^{2}\sum_{t=1}^{2} P_{t.}^{r.}\Psi_i\right)\left(\sum_{s=1}^{2}\sum_{u=1}^{2} P_{.u}^{.s}\Psi_j\right), \quad (1)$$

where the dot notation is used to indicate summation over the corresponding index. By using and extending the results of Cockerham and Weir (1973) for the frequencies P_{tu}^{rs} in segregating populations derived from biparental crosses by selfing, backcrossing or random mating, we obtain for ω_{ij} the formulas in terms of the linkage value λ_{ij} (Schnell 1961) between loci i and j presented in Table 1. If we assume absence of interference (Stam 1979) of crossover formation, the formulas can be expressed in terms of the map position x and y of locus i and j, respectively, on the chromosome calculated by Haldane's (1919) mapping function.

For DH lines or recombinant inbred lines (RILs) developed by the SSD method, these formulas correspond to those of Frisch and Melchinger (2007), who showed that for homozygous genotypes ω_{ij} is identical to D_{ij}, the coefficient of linkage disequilibrium or synonymously the coefficient of gametic phase disequilibrium between loci i and j (Lynch and Walsh 1998, p. 94) in the segregating population considered. However, for F_3 and higher selfing generations F_t, ω_{ij} is no longer identical or proportional to D_{ij} but is a polynomial in terms of λ_{ij}. By setting $\lambda_{ij} = 1$,

Table 1 Covariance $\mathrm{Cov}(\Psi_i, \Psi_j)$ of the parental genome contribution (PGC) between two loci i and j in segregating populations of infinite size under various mating systems

Generation[a]	$\mathrm{Cov}(\Psi_i, \Psi_j)$ in terms of	
	Linkage value λ_{ij}	Map position x and y of loci i and j
F_1DH	$\frac{1}{4}\lambda_{ij}$	$\frac{e^{-2\lvert x-y\rvert}}{4}$
F_2	$\frac{1}{8}\lambda_{ij}$	$\frac{e^{-2\lvert x-y\rvert}}{8}$
F_3	$\frac{1}{8}\lambda_{ij} + \frac{1}{16}\lambda_{ij}^2$	$\frac{1}{8}\,e^{-2\lvert x-y\rvert} + \frac{1}{16}\,e^{-4\lvert x-y\rvert}$
F_{t+1}	$\frac{1}{4}\sum_{r=1}^{t}\left(\frac{\lambda_{ij}}{2}\right)^r$	$\sum_{r=1}^{t}\left(\frac{1}{2}\right)^{r+2} e^{-2r\lvert x-y\rvert}$
BC_t	$\left(\frac{1+\lambda_{ij}}{4}\right)^t - \left(\frac{1}{4}\right)^t$	$\left(\frac{1+e^{-2\lvert x-y\rvert}}{4}\right)^t - \left(\frac{1}{4}\right)^t$
$F_2\mathrm{Syn}_t$	$\frac{1}{8}\lambda_{ij}\left(\frac{1+\lambda_{ij}}{2}\right)^t$	$\frac{1}{8}e^{-2\lvert x-y\rvert}\left(\frac{1+e^{-2\lvert x-y\rvert}}{2}\right)t$

[a] t refers to the number of selfing, backcrossing (BC) and intermating (Syn) generations

we obtain $\mathrm{Var}(\Psi_i) = \omega_{ii}$, the variance of the PGC at one locus.

We now consider Ψ_M, the PGC of parent P2 with regard to a set M of $m = |M|$ selectively neutral marker loci located either on one chromosome or the entire genome. Then,

$$\Psi_M = \frac{1}{m}\sum_{i\in M}\Psi_i \tag{2}$$

and

$$\mathrm{Var}(\Psi_M) = \frac{1}{m^2}\sum_{i\in M}\sum_{j\in M}\mathrm{Cov}(\Psi_i,\Psi_j) = \frac{1}{m^2}\sum_{i\in M}\sum_{j\in M}\omega_{ij}. \tag{3}$$

Let Ψ_{c_b} denote the PGC of P2 regarding the entire chromosome C_b with length l_b (in Morgan units). Adopting the approach of Hill (1993) and Frisch and Melchinger (2007), we obtain the variance $\mathrm{Var}(\Psi_{c_b})$ as the expectation for the covariance of the PGC between two randomly sampled loci on that chromosome,

$$\mathrm{Var}(\Psi_{c_b}) = E\left[\mathrm{Cov}(\Psi_i,\Psi_j)\right] = \frac{1}{l_b^2}\int_0^{l_b}\int_0^{l_b} f(x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{4}$$

where $f(x, y)$ corresponds to ω_{ij} in terms of the chromosomal map positions x and y of locus i and j, respectively (Table 1). The integrals of $f(x, y)$ for the cases listed in Table 1 can be solved by making use of the detailed derivations given in the Appendix of Frisch and Melchinger (2007). The analytical results for $\mathrm{Var}(\Psi_{c_b})$ are summarized in Table 2 and the derivations for generations F_t are presented in the Appendix. For homozygous lines, they correspond to the results given by Frisch and Melchinger (2007). In order to check our derivations for not fully inbred genotypes, we compared the numerical solutions in Fig. 1 with the numerical results obtained from Eq. (2) by assuming a set M of m equidistantly located markers on chromosome C_b. The differences between the analytically determined variances and numerical approximations were negligible (<0.0001 for $m > 100{,}000$).

Our results on $\mathrm{Var}(\Psi_{c_b})$ for generation F_t in Table 2 can also be used to determine the segregation variances for (Ψ_{c_b}) among genotypes in advanced selfing generations, i.e., segregation among F_s individuals independently derived by SSD from an F_t genotype, by making use of the following relationship

Table 2 Variance $\mathrm{Var}(\Psi_{c_b})$ of the parental genome contribution (PGC) to a chromosome of length l_b under various mating systems

Generation[a]	$\mathrm{Var}(\Psi_{c_b})$
F_1DH	$\frac{1}{l_b^2}\cdot\frac{1}{4}\left[l_b - \frac{1}{2}\left(1-e^{-2l_b}\right)\right]$
F_2	$\frac{1}{l_b^2}\cdot\frac{1}{8}\left[l_b - \frac{1}{2}\left(1-e^{-2l_b}\right)\right]$
F_3	$\frac{1}{l_b^2}\cdot\frac{1}{128}\left(20l_b + 8e^{-2l_b} + e^{-4l_b} - 9\right)$
F_{t+1}	$\frac{1}{l_b^2}\sum_{r=1}^{t}\left(\frac{1}{2}\right)^{r+3}\cdot\frac{1}{r^2}\left[2rl_b - 1 + e^{-2rl_b}\right]$
BC_t	$\frac{1}{l_b^2}\cdot\frac{1}{4^{t+1}}\sum_{r=1}^{t}\binom{t}{r}\frac{1}{2r^2}\left[2rl_b - 1 + e^{-2rl_b}\right]$
$F_2\mathrm{Syn}_t$	$\frac{1}{l_b^2}\cdot\frac{1}{8^{t+1}}\left[l_b\left(2-\frac{1}{2^t}\right) - \frac{1}{2^{t+1}}\sum_{r=1}^{t+1}\binom{t+1}{r}\frac{1}{r}\left(1-e^{-2rl_b}\right)\right]$

[a] See footnote of Table 1

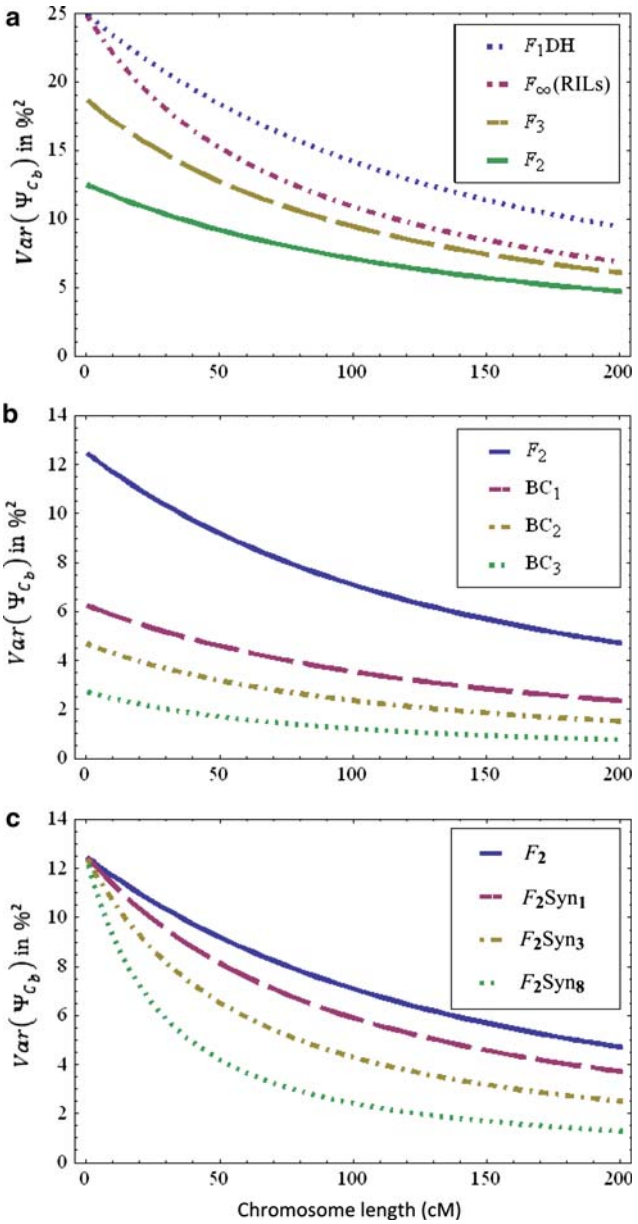

Fig. 1 Relationship between chromosome length l_b and $\mathrm{Var}(\Psi_{c_b})$ (variance of the parental genome contribution (PGC) as a function of l_b in cM units) under three mating systems: (**a**) selfing series, (**b**) backcross series, and (**c**) intermating series

$$\mathrm{Var}(\Psi_{c_b}|F_{t:s}) = \mathrm{Var}(\Psi_{c_b}|F_s) - \mathrm{Var}(\Psi_{c_b}|F_t) \quad (5)$$

where $\mathrm{Var}(\Psi_{c_b}|F_t)$ refers to $\mathrm{Var}(\Psi_{c_b})$ for generation F_t.

Meiotic recombination and transmission of genetic material from parents to offspring are stochastically independent for different chromosomes. Hence, if we denote by Ψ_c the PGC of P2 across all chromosomes

$$\Psi_c = \frac{1}{l}\sum_{b=1}^{B} l_b \Psi_{c_b} \quad (6)$$

where B is the number of chromosomes and $l = \sum_{b=1}^{B} l_b$ the length of the whole genome in Morgan units. Then we obtain for the variance of the PGC from P2 to progeny

$$\mathrm{Var}(\Psi_C) = \sum_{b=1}^{B}\left(\frac{l_b}{l}\right)^2 \mathrm{Var}(\Psi_{C_b}). \quad (7)$$

Variance of Z_2 in Design III studies

In the Design III, randomly chosen genotypes of a segregating population derived from a biparental cross are mated with each homozygous parent P1 and P2 to develop pairs of BC progenies, which are phenotyped. We use H_1 and H_2 to denote the phenotypic value of BC progenies of each genotype of the segregating population with parent P1 and P2, respectively, where P2 is the parent with the higher population mean of BC progenies. Melchinger et al. (2007) demonstrated that QTL mapping, performed with marker data of the genotypes in the segregating population and phenotypic data for the linear contrast $Z_2 = (H_1 - H_2)/2$ of the corresponding BC progenies, provides estimates of the augmented dominance effects $d_i^* = d_i - \frac{[aa_i.]}{2}$, where d_i is the dominance effect of QTL i, and $[aa_i.] = \sum_{j\in Q\setminus\{i\}} aa_{ij}$ is the sum of additive × additive (aa_{ij}) effects of QTL i with all other QTL j in the genetic background. They further demonstrated that the sum of d_i^* over the entire set Q of QTL is equal to MPH for the trait under study.

Let g_{vw} denote the genotype of a gamete with allele v at locus i and allele w at locus j. Using results on conditional expectations (e.g., Chung 1974, p. 300), we can calculate the conditional expectations of Z_2 with respect to the loci i and j by the following equation, given that the individual used for producing the BC progenies had genotype G_{tu}^{rs}

$$E\left(Z_2\middle|G_{tu}^{rs}\right) = \sum_{v=1}^{2}\sum_{w=1}^{2}\Pr\left(g_{vw}\middle|G_{tu}^{rs}\right)E(Z_2|g_{vw}), \quad (8)$$

where

$$E(Z_2|g_{vw}) = \Big[-\left(a_i + a_j\right) + (-1)^v d_i^* + (-1)^w d_j^* + (-1)^{v+w}\left(ad_{ij} + da_{ij}\right)\Big]/2, \quad (9)$$

and the conditional probability $\Pr\left(g_{vw}\middle|G_{tu}^{rs}\right)$ is equal to

$$\begin{cases} 1 & \text{if } [v=r=t \wedge w=s=u], \\ \frac{1}{2} & \text{if } [v=r=t \wedge s\neq u] \vee [w=s=u \wedge r\neq t], \\ \frac{1+\lambda_{ij}}{4} & \text{if } [r\neq t \wedge s\neq u] \wedge [[v=r \wedge w=s] \vee [v=t \wedge w=u]], \\ \frac{1+\lambda_{ij}}{4} & \text{if } [r\neq t \wedge s\neq u] \wedge [[v=r \wedge w=u] \vee [v=t \wedge w=s]] \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

Accounting for all types of digenic epistasis but ignoring epistasis of higher order as well as linkage between more than two loci, we can calculate the variance of Z_2 with the help of Eqs. (8) and (9) and the frequencies P_{tu}^{rs} of genotypes G_{tu}^{rs} and obtain

$$\mathrm{Var}(Z_2) = \sum_{i\in Q}\sum_{j\in Q}\omega_{ij}d_i^*d_j^* + \sum_{i\in Q}\sum_{j\in Q}\gamma_{ij}\left(ad_{ij} + da_{ij}\right)^2 \quad (10)$$

$$= F\sum_{i\in Q} d_i^{*2} + \sum_{i\in Q}\sum_{\substack{j\in Q \\ i\neq j}}\omega_{ij}d_i^*d_j^* + \sum_{i\in Q}\sum_{j\in Q}\gamma_{ij}\left(ad_{ij} + da_{ij}\right)^2. \quad (11)$$

Here, F denotes the inbreeding coefficient of the genotypes in the segregating population ($F = 0.5$ for F_2 and F_2Syn_t, $F = 0.75$ for F_3, $F = 1 - (0.5)^t$ for F_{t+1}, and $F = 1$ for RILs and DH lines), ω_{ij} is the covariance as defined in Eq. (1), and γ_{ij} is a complex function of the frequencies P_{tu}^{rs}, which simplifies in special cases as shown below. The reason for obtaining ω_{ij} as a coefficient of product $d_i^*d_j^*$ is that the coefficients of d_i^* in Eq. (8) are identical with the values for Ψ_i. Explicit formulas of γ_{ij} for F_2 and RIL populations in terms of the gametic phase disequilibrium D_{ij} were given by Melchinger et al. (2007). For F_2 and F_2Syn_t populations, we get $\gamma_{ij} = \frac{1}{128}(1 + \gamma_{ij}^2 - 32\lambda_{ij}D_{ij}^2)$ with $D_{ij} = \frac{\lambda_{ij}}{4}$ for F_2 and $D_{ij} = \left(\frac{1+\lambda_{ij}}{2}\right)^2\frac{\lambda_{ij}}{4}$ for F_2Syn_t. For RIL and DH populations, we have $\gamma_{ij} = \frac{1}{32}(1 - 16D_{ij}^2)$ with $D_{ij} = \frac{\lambda_{ij}}{4(2-\lambda_{ij})}$ for RILs and $D_{ij} = \frac{\lambda_{ij}}{4}$ for DH lines. For F_3 populations we have

$$\gamma_{ij} = \left(\frac{1-\lambda_{ij}}{32}\right)\left(\frac{9}{16} + \frac{9}{16}\lambda_{ij} + \frac{1}{2}\lambda_{ij}^2 + \frac{1}{4}\lambda_{ij}^3 + \frac{1}{8}\lambda_{ij}^4\right).$$

From these results, we obtain $\gamma_{ij} \leq 1/32$ for RILs and DH lines and $\gamma_{ij} \leq 9/512$ for F_3 populations, with the maximum values being assumed for $\lambda_{ij} = 0$, i.e., unlinked loci. For F_2 we have $\gamma_{ij} \leq 0.0081$ and the maximum is assumed for $\lambda_{ij} = \frac{1}{3}$. For F_2Syn_3 and F_2Syn_8, $\gamma_{ij} \leq 0.0098$ and $\gamma_{ij} \leq 0.0114$ and the maxima are assumed for $\lambda_{ij} = 0.63$ and $\lambda_{ij} = 0.76$, respectively. In view of the small value of the coefficients γ_{ij} compared with the inbreeding coefficient F, the last term in Eqs. (10) and (11) can safely be ignored unless the terms $(ad_{ij} + da_{ij})^2$ are large compared to d_i^{*2}. However, this seems rather unlikely, because additive × dominance effects do not contribute to

the genetic gain in hybrid breeding and, hence, are not selected for. Consequently, we can approximate $\mathrm{Var}(Z_2)$ by

$$\mathrm{Var}(Z_2) \cong \sum_{i\in Q}\sum_{j\in Q}\omega_{ij}d_i^*d_j^* = F\sum_{i\in Q}d_i^{*2} + \sum_{i\in Q}\sum_{\substack{j\in Q\\ i\neq j}}\omega_{ij}d_i^*d_j^*. \quad (12)$$

Covariance of Z_2 with the parental genome contribution

As above, we consider two QTL i and j and the PGC at a third locus k with arbitrary linkage among these loci. Making use of the formulas for the conditional expectations $E\left[(Z_2|G_{tu}^{rs})\right]$ derived in the previous section, and three-locus genotype frequencies P_{tuw}^{rsv}, which are obtained by straightforward extensions of the formulas for the two-locus genotype frequencies P_{tu}^{rs} given by Cockerham and Weir (1973), we obtain

$$\mathrm{Cov}(Z_2(i,j),\Psi_k) = \omega_{ik}d_i^* + \omega_{jk}d_j^* \quad (13)$$

and, consequently, we have

$$\mathrm{Cov}(Z_2,\Psi_M) = \frac{1}{|M|}\sum_{i\in Q}\sum_{j\in M}\omega_{ij}d_i^* \quad (14)$$

This result reveals that only augmented dominance effects d_i^*, i.e., dominance effects d_i and additive × additive effects with the genetic background $[aa_{i.}]$ contribute to $\mathrm{Cov}(Z_2,\Psi_M)$.

Correlation of Z_2 and Ψ_M

From Eqs. (3), (12), and (14), we obtain for the correlation between Z_2 and the PGC Ψ_M with respect to marker loci set M

$$r(Z_2,\Psi_M) = \frac{\sum_{i\in Q}\sum_{k\in M}\omega_{ik}d_i^*}{\left[\left(\sum_{i\in Q}\sum_{j\in Q}\omega_{ij}d_i^*d_j^*\right)\left(\sum_{k\in M}\sum_{l\in M}\omega_{kl}\right)\right]^{1/2}} \quad (15)$$

For loci $i \neq j$ located on different chromosomes, we have $\lambda_{ij} = 0$ and consequently $\omega_{ij} = 0$ (see Table 1). Thus, only linked pairs of loci located on the same chromosome contribute to $\mathrm{Var}(\Psi_M)$, $\mathrm{Var}(Z_2)$, $\mathrm{Cov}(Z_2,\Psi_M)$ and $r(Z_2,\Psi_M)$. Therefore, Eq. (15) can be rewritten as

$$r(Z_2,\Psi_M) = \frac{\sum_{b=1}^{B}\sum_{i\in Q_b}\sum_{k\in M_b}\omega_{ik}d_i^*}{\left[\left(\sum_{b=1}^{B}\sum_{i\in Q_b}\sum_{j\in Q_b}\omega_{ij}d_i^*d_j^*\right)\left(\sum_{b=1}^{B}\sum_{k\in M_b}\sum_{l\in M_b}\omega_{kl}\right)\right]^{1/2}}, \quad (16)$$

where Q_b and M_b refer to the set of QTL and markers, respectively, located on chromosome C_b.

Numerical and experimental results

Graphs for $\mathrm{Var}(\Psi_{c_b})$ under three mating systems (selfing, intermating and backcrossing) for chromosomal length l_b from 0 to 200 cM are presented in Fig. 1. $\mathrm{Var}(\Psi_{c_b})$ is larger in advanced selfing generations but smaller in advanced backcross generations. Under both mating systems the values decrease with an increasing chromosome length and the differences among generations narrow down as l_b increases. Irrespective of the chromosome length, the values of $\mathrm{Var}(\Psi_{c_b})$ for F_1DH lines are exactly twice as large as those for the F_2 generation as follows from the formulas in Table 2. The curve for F_3 is initially intermediate between F_1DH and F_2 and converges towards that for F_2 with larger l_b. Compared with the F_2, $\mathrm{Var}(\Psi_{c_b})$ is exactly half in BC_1. A similar reduction is observed in successive BC generations except that it is one-third from BC_1 to BC_2. Under intermating, $\mathrm{Var}(\Psi_{c_b})$ decreases and the differences among the various generations (F_2, F_2Syn_1 to F_2Syn_8) are more pronounced for intermediate chromosomal lengths (50–100 cM) than for extremes.

Based on published genetic maps of Arabidopsis, barley, maize, rice and wheat, $\mathrm{Var}(\Psi_c)$ was estimated for the three mating systems (Table 3). The rate of change of $\mathrm{Var}(\Psi_c)$ under selfing, backcrossing and intermating was similar in

Table 3 Variance $\mathrm{Var}(\Psi_c)$ of the parental genome contribution (PGC) for various crops under various mating systems

Generation	$\mathrm{Var}(\Psi_c)$ in %				
	Arabidopsis	Barley	Maize	Rice	Wheat
Selfing series					
F_1DH	306.72	157.52	119.26	95.66	61.80
F_2	153.36	78.76	59.63	47.83	30.90
F_3	206.01	102.47	78.00	62.34	40.64
F_4	225.72	110.86	84.56	67.50	44.16
F_∞	239.26	116.41	88.92	70.92	46.51
Backcross series					
BC_1	76.69	39.38	29.82	23.91	15.45
BC_2	51.50	25.62	19.50	15.59	10.16
BC_3	26.71	12.88	9.85	7.85	5.16
BC_4	12.68	5.94	4.56	3.62	2.40
Intermating series					
F_2Syn_1	129.32	63.09	48.18	38.43	25.19
F_2Syn_2	110.70	51.79	39.83	31.63	20.98
F_2Syn_3	96.06	43.46	33.61	26.60	17.81
F_2Syn_8	55.44	22.93	17.99	14.12	9.69

Based on the chromosome length (in Morgan units) in the maps of cereals in the database http://www.gramene.org (maize: nested association map, wheat: consensus SSR map, rice: Cornell SSR map, barley: Consensus 2006, Stein) and Arabidopsis in Törjék et al. (2006)

all crop species, but in Arabidopsis the increase in the variance with selfing was slightly higher.

In our companion paper (Schön et al. 2010), we re-analyzed two Design III studies in maize with regard to QTL mapping for Z_2 and presented results on the phenotypic and genotypic correlations of $r(Z_2, \Psi_M)$. Briefly, in the study by Stuber et al. (1992) 264 F_3 lines (Pop1) from the hybrid B73 × Mo17 were genotyped with 76 markers and their BC progenies were phenotyped for grain yield and other important agronomic traits. Likewise, in the study by Lu et al. (2003) 351 F_2Syn_3 plants (Pop2) from the hybrid LH200 × LH216 were genotyped with 158 markers and their BC progenies were phenotyped for grain yield and other traits. The average marker distance was 18 cM in Pop1 and 10 cM in Pop2. In both studies, the means of Ψ_M (50.2% in Pop1 and 49.1% in Pop2) were not significantly different from the expected value of 50%. Likewise, the values of $\mathrm{Var}(\Psi_M)$ (84.3% in Pop1 and 37.4% in Pop2) were in good agreement with those given in Table 3 for $\mathrm{Var}(\Psi_c)$ in maize. In both populations, Ψ_M and Z_2 for grain yield showed high correlations $r(Z_2, \Psi_M)$ at the phenotypic (r_p) as well as the genotypic (r_g) levels (Fig. 2).

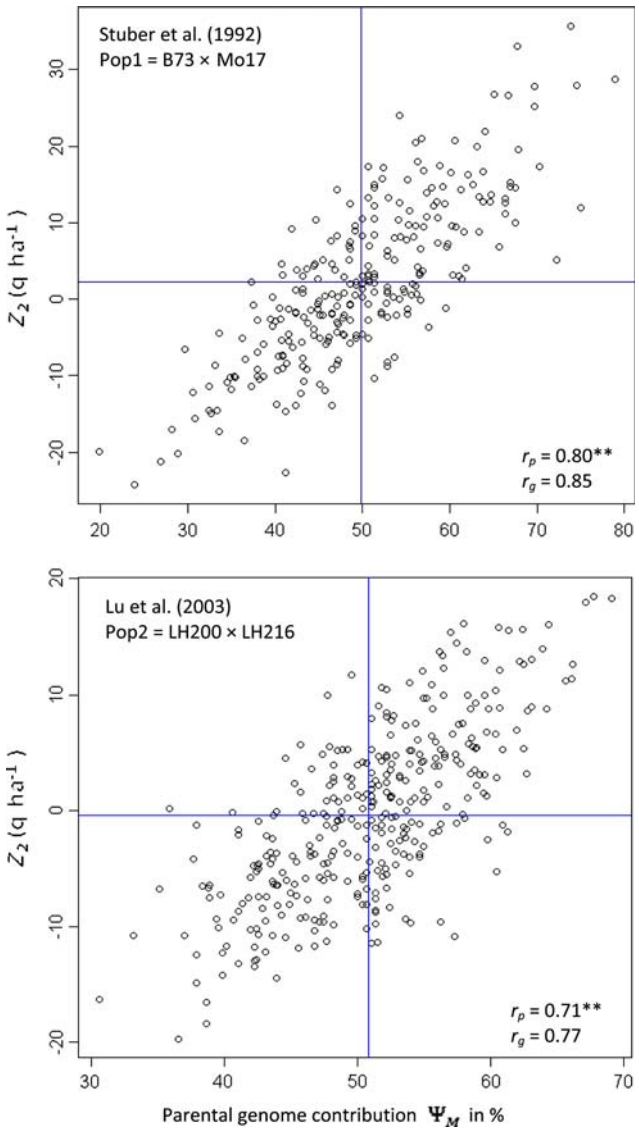

Fig. 2 Correlation between Ψ_M (parental genome contribution (PGC) with respect to set M) and Z_2 (half the difference between each pair of BC progenies in Design III) for grain yield in two populations of maize from the literature

Discussion

Variance of parental genome contribution

In the derivation of the formulas for $\mathrm{Var}(\Psi_{c_b})$, we assumed no interference in meiosis (Stam 1979). This is also a basic assumption underlying Haldane's (1919) mapping function. Consequently, for calculating $\mathrm{Var}(\Psi_{c_b})$ and $\mathrm{Var}(\Psi_c)$ according to the formulas in Eqs. (4) and (7), respectively, one should insert chromosome lengths l_b determined by Haldane's mapping function. When interference is assumed, which is the case with Kosambi's (1944) mapping function, $\mathrm{Var}(\Psi_{c_b})$ is expected to increase because $\mathrm{Cov}(\Psi_i, \Psi_j)$ is in this case larger than expected in the absence of the interference.

For a given mating system, $\mathrm{Var}(\Psi_c)$ depends on the number, average length, and the variation in the length of chromosomes. With similar length l_b of each chromosome, $\mathrm{Var}(\Psi_c)$ decreases according to Eq. (7) for species with a larger number of chromosomes. Given a fixed total genome length l, $\mathrm{Var}(\Psi_c)$ is greater for species with fewer but longer chromosomes (Eq. (7) and Fig. 1). Likewise, if both the total genome length and the chromosome number are fixed, $\mathrm{Var}(\Psi_c)$ is smaller for species with uniform length of chromosomes. The differences in $\mathrm{Var}(\Psi_c)$ among the five species (Table 3) can be explained on the basis of these rules.

As expected, the magnitude of $\mathrm{Var}(\Psi_{c_b})$ decreased with intermating and backcrossing but increased under selfing (Fig. 1). The variance decreased more rapidly with backcrossing than with intermating, because with backcrossing there is a systematic incorporation of PGC of the recurrent parent, which reduces the probability of recombination between the parental genomes. The trends in the relative magnitude of $\mathrm{Var}(\Psi_c)$ under selfing, intermating and backcrossing were similar for all five species except for a larger increase with selfing in Arabidopsis as a consequence of the smallest chromosome number, shortest genome length and largest variation in chromosome length (Table 3).

The ratio of $\mathrm{Var}(\Psi_{c_b})$ for F_1DH, F_2 and BC_1 populations is exactly 4:2:1, as applies also to the ratio of the additive genetic variance (σ_A^2) of quantitative traits in these

generations (Melchinger 1987). In addition to the formulas given in Table 2, this result follows also from simple arguments considering the variation and covariation of the PGC in the parental gametes of the individuals in these generations. However, instead of the 1:$\frac{3}{2}$:2 ratio expected for σ_A^2 in generations F_2, F_3 and F_∞ in the absence of linkage, the corresponding ratio of $\mathrm{Var}(\Psi_{c_b})$ for a chromosome of 200 cM length was 1.00:1.29:1.46. Thus, it follows from Eq. (5) that $\mathrm{Var}(\Psi_{c_b}|F_{t:s})$ arising from segregation in higher selfing generations is smaller than the values expected from $\mathrm{Var}(\Psi_{c_b})$ in F_2. An explanation is that with long chromosomes, there is a large number of heterozygous chromosome segments in F_2 genotypes, which behave in subsequent selfing generations like a large number of short and almost independently segregating chromosomes. Consequently, the variance for Ψ_c in subsequent selfing generations is smaller than expected from B chromosomes with length $l_b/2$. The same argument applies to the segregation of Ψ_{c_b} in advanced selfing generations and explains, why the curve for F_∞ was below the curve for F_1DH lines. These examples illustrate that the inverse of $\mathrm{Var}(\Psi_c)$ provides an indirect measure for the extent of genome reshuffling by recombination as a result of recurrent selfing or intermating.

Altogether, our findings on the effects of chromosome number B, length l_b, and variation in l_b on $\mathrm{Var}(\Psi_c)$ under a given mating system (Eq. (7), Fig. 1, Table 3) show that alterations in the PGC by marker-assisted selection (MAS) for the genetic background (Tanksley et al. 1989, Hospital and Charcosset 1997) are most promising in species with a small number and/or short length of chromosomes. In comparison to the F_2 generation, the prospects of MAS in advanced selfing generations are much lower than normally expected on the basis of the segregation ratios for quantitative traits. Once MAS has been practiced, our formulas are not applicable for the determination of its effectiveness in subsequent generations, because of the assumption of no selection.

Factors influencing the correlation of Z_2 with Ψ_M

The correlation $r(Z_2, \Psi_M)$ is influenced by (i) the linkage relationships between loci in Q and M, between loci within Q, and between loci within M and (ii) the variance of d_i^* effects. We analyzed the effect of each of these factors following the approach used by Charcosset and Essioux (1994) for investigating the relationship between genetic distance and MPH. Our Eqs. (3), (10) and (14) have the same formal structure as their Eq. (8) to (10), when ω_{ij} is replaced by D_{ij}^2 and epistasis is ignored.

Under the assumptions of (i) constant d_i^* for all QTL $i \in Q$ and (ii) $Q = M$, i.e., each QTL is tagged by one functional marker (Anderson and Lübberstedt 2003) and each marker tags one QTL ($\lambda_{ij} = 1$), we obtain $r(Z_2, \Psi_M) = 1$. If we relax the second assumption and assume that all marker loci in set M and all QTL in set Q are unlinked ($\lambda_{ij} = 0$) and only a subset $S = Q \cap M$ of QTL are tagged by markers, then

$$r^2(Z_2, \Psi_M) = \frac{|s|^2}{|Q| \cdot |M|} = \frac{|Q \cap M|}{|Q|} \cdot \frac{|Q \cap M|}{|M|} = QtM \times MtQ \tag{17}$$

where QtM is the proportion of QTL for Z_2 tagged by a functional marker, and MtQ is the proportion of markers that tag a QTL for Z_2. This illustrates that markers not associated with QTL for Z_2 will decrease the correlation $r(Z_2, \Psi_M)$ through MtQ. Likewise, QTL not tagged by a marker also reduce the correlation through QtM. If determination of Ψ_M is based on a uniform, dense marker coverage, each QTL is tightly linked to a marker and QtM should be close to 1. However, MtQ might be considerably smaller than 1, if numerous markers are not linked with QTL. Further, adding markers less tightly linked than the marker(s) tagging the QTL reduces the correlation, as follows from Eq. (17).

Assuming $Q = M$ and no linkage among the QTL, the correlation is inversely related to the variation of d_i^* across the QTL

$$r^2(Z_2, \Psi_M) = \frac{\left(\sum_{i \in Q} d_i^*\right)^2}{\left(\sum_{i \in Q} d_i^*\right) \cdot |M|} = \frac{1}{1 + \frac{\mathrm{Var}(d_i^*)}{(\mathrm{mean}\, d_i^*)^2}} \tag{18}$$

where $\mathrm{mean}\, d_i^*$ and $\mathrm{Var}(d_i^*)$ are the mean and variance of d_i^* across the entire set Q of QTL, respectively.

If all chromosomes have similar marker coverage, i.e., $\sum_{i \in M_b} \sum_{j \in M_b} \omega_{ij} = \kappa_b = \kappa$ for $b = 1,\ldots, B$, and R out of the B chromosomes have the same QTL/marker profile, i.e., $\sum_{i \in Q_b} \sum_{j \in M_b} \omega_{ij} d_i^* = \tau_b = \tau$ and $\sum_{i \in Q_b} \sum_{j \in Q_b} \omega_{ij} d_i^* d_j^* = \upsilon_b = \upsilon \neq 0$, while the remaining B-R chromosomes harbor no QTL, i.e., $\upsilon_b = 0$, then $r(Z_2, \Psi_M) = \sqrt{\frac{R}{B}}\xi$, with $\xi = \frac{T}{\sqrt{\upsilon\kappa}}$. Thus, if up to half of the chromosomes of equal length l_b with identical marker coverage harbor QTL for Z_2, then $r(Z_2, \Psi_M) \leq \sqrt{0.5}\xi$, where ξ refers to the association between markers and QTL on those chromosomes.

Correlation of Z_2 with Ψ_M in different types of segregating populations

As follows from Eq. (15) and the formulas for ω_{ij} given in Table 1, the expectation of $r(Z_2, \Psi_M)$ is identical for F_2 and F_1DH lines in the absence of epistatic effects other than aa_{ij}. Under identical marker densities, a lower value of $r(Z_2, \Psi_M)$ is expected in $F_2\mathrm{Syn}_t$ (e.g., $t = 3$) than in F_2, because recombination reduces ω_{ij} only for $i \neq j$ but not

for $i = j$, and, hence, $\mathrm{Cov}(Z_2, \Psi_M)$ is affected by recombination to a larger extent than $\mathrm{Var}(Z_2)$ and $\mathrm{Var}(\Psi_M)$. Based on same reasoning, we expect smaller values for $r(Z_2, \Psi_M)$ in RIL than in DH line populations. For dense marker coverage, the differences for $r(Z_2, \Psi_M)$ between F_2 and F_3 populations should be small, because in this case, $\lambda_{ij}\left(1 + \frac{\lambda_{ij}}{2}\right) \approx \frac{3}{2}\lambda_{ij}$.

Correlation of Z_2 and Ψ_M in maize and other species

A high genotypic correlation ($r_g(Z_2, \Psi_M) \approx 0.8$) was reported for grain yield in maize by Schön et al. (2010) for the two populations studied by us (Fig. 2) and one by Frascaroli et al. (2007). Several reasons may account for its deviation from unity. (i) Marker coverage of the genome in these studies had several gaps larger than 40 cM. Thus, some QTL for Z_2 may not have been closely tagged by markers. (ii) Chromosomal regions varied in their effect on Z_2, as indicated by the size of d_i^* effects in the QTL mapping results (Schön et al. 2010). As follows from Eq. (18), $\mathrm{Var}(d_i^*)$ reduces the correlations. (iii) Epistasis might have also weakened the correlation. While additive × additive epistatic effects (aa_{ij}) are fully accounted for by the estimates of d_i^*, certain types of higher order epistatic interactions contributing to MPH are only partly captured by d_i^* (Melchinger et al. 2007). As shown in the theory section, it is rather unlikely that $(ad_{ij} + da_{ij})^2$ effects contributed substantially to a reduction in the observed correlations, because they do not influence $\mathrm{Cov}(Z_2, \Psi_M)$ and their effect on an increased $\mathrm{Var}(Z_2)$ is expected to be negligible in view of the small coefficients of γ_{ij} compared with ω_{ij}.

From the estimates of $r_g(Z_2, \Psi_M) \approx 0.80$, we can draw several conclusions. First, at least two-thirds of the chromosomes must have harbored QTL for Z_2. This follows from the formula $r(Z_2, \Psi_M) = \sqrt{\frac{R}{B}\xi}$ derived in the previous section. Second, most chromosomes must have carried several QTL for Z_2, because otherwise ξ would be much smaller than one, resulting in a further reduction of the estimates of $r_g(Z_2, \Psi_M)$. Third, essentially all QTL for Z_2 must have had a positive sign for the d_i^* effects, because otherwise positive and negative terms would cancel each other in the sum of $\mathrm{Cov}(Z_2, \Psi_M)$ (Eq. 14). All these conclusions based on theoretical reasoning are corroborated by QTL analyses for Z_2 of grain yield (Schön et al. 2010), in which 11 QTL located on eight chromosomes for Pop1 and 15 QTL located on nine chromosomes for Pop2 were found, all with positive d_i^*. When inserting the map position of the markers and the detected QTL as well as the estimates of d_i^* from the study of Schön et al. (2010) into Eq. (15), we obtained correlations of 0.50 and 0.44 for Pop1 and Pop2, respectively. The discrepancies between these values and the observed estimates of $r_g(Z_2, \Psi_M)$ demonstrate that additional QTL must contribute to MPH for grain yield in each population. Thus, numerous QTL may have remained undetected in QTL mapping of Z_2, because their effects were too small. This conclusion is confirmed by the proportion of the genotypic variance of Z_2 explained by all detected QTL in cross validation, which was 66% in Pop1 and 58% in Pop2 (Schön et al. 2010).

The high values for $r(Z_2, \Psi_M)$ in maize are in contrast to smaller estimates of this correlation for grain yield in rice ($r_p = 0.13$ A. Augusto F. Garcia, personal communication) and biomass yield in Arabidopsis ($r_p = 0.06$, A. E. Melchinger, unpublished data). The lower correlations in rice and Arabidopsis could be attributable to (i) the lower level of MPH for grain and biomass yield in autogamous species compared with grain yield in maize (Becker 1993) and a different genetic architecture underlying MPH with fewer QTL displaying large d_i^* effects, some of which may even have a negative sign as reported in the corresponding QTL studies in rice (Garcia et al. 2008) and Arabidopsis (Kusterer et al. 2007).

In conclusion, the results of the present study on $r(Z_2, \Psi_M)$ together with those on QTL mapping (Schön et al. 2010) clearly support the hypothesis that MPH for grain yield in maize is of polygenic nature. Further, the majority of chromosomes carry QTL with positive d_i^* effects and these QTL include not only major QTL, but also a large number of QTL with small effects, distributed along the chromosomes. Thus, $r(Z_2, \Psi_M)$ allows making inferences beyond conclusions to be drawn from QTL mapping on the distribution of QTL for MPH across the genome and also on the variation in the magnitude of all underlying d_i^* effects.

Acknowledgments We are grateful to R. Bernardo, C.W. Stuber and Z.-B. Zeng for providing the raw data for Pop1 and Pop2 used in preparation of Fig. 2 and to C.C Schön for suggestions on the manuscript. This research was financed by the Deutsche Forschungsgemeinschaft (DFG) within the priority program SPP 1149 "Heterosis in Plants" (Research Grant ME931/4-3).

Appendix

Here, we derive $\mathrm{Var}(\Psi_{c_b})$ for generation F_{t+1} ($t \geq 1$). According to Table 2, we have for F_{t+1} $\mathrm{Cov}(\Psi_i, \Psi_j) = \sum_{r=1}^{t} \left(\frac{1}{2}\right)^{r+2} e^{-2r|x-y|}$. From Eq. [A3] of the Appendix of Frisch and Melchinger (2007), we obtain

$$\int_0^{l_b}\int_0^{l_b} e^{-2r(x-y)}\mathrm{d}x\,\mathrm{d}y = \frac{1}{2r^2}\left(2rl_b - 1 + e^{-2rl_b}\right)$$

and together with Eq. (4), we get

$$\begin{aligned}\mathrm{Var}(\Psi_{c_b}) &= \frac{1}{l_b^2}\int_0^{l_b}\int_0^{l_b}\mathrm{Cov}(\Psi_i,\Psi_j)\mathrm{d}x\,\mathrm{d}y \\ &= \frac{1}{l_b^2}\int_0^{l_b}\int_0^{l_b}\sum_{r=1}^{t}\left(\frac{1}{2}\right)^{r+2}e^{-2r|x-y|}\mathrm{d}x\,\mathrm{d}y \\ &= \frac{1}{l_b^2}\sum_{r=1}^{t}\left(\frac{1}{2}\right)^{r+3}\frac{1}{r^2}\left(2rl_b-1+e^{-2rl_b}\right)\end{aligned}$$

References

Anderson JR, Lübberstedt T (2003) Functional markers in plants. Trends Plant Sci 8:554–560

Becker HC (1993) Pflanzenzüchtung. Eugen Ulmer Verlag, Stuttgart

Charcosset A, Essioux L (1994) The effect of population structure on the relationship between heterosis and heterozygosity at marker loci. Theor Appl Genet 89:336–343

Chung KL (1974) A course in probability theory. Academic Press, New York

Cockerham CC, Weir BS (1973) Descent measures for two loci with some applications. Theor Pop Biol 4:300–330

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

Comstock RE, Robinson HF (1952) Estimation of average dominance of genes. In: Gowen JW (ed) Heterosis. Iowa State College Press, Ames, pp 495–516

Frascaroli E, Canè MA, Landi P, Pea G, Gianfranceschi L, Villa M, Morgante M, Pè ME (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, Melchinger AE (2007) Variance of the parental genome contribution to inbred lines derived from biparental crosses. Genetics 176:477–488

Garcia AAF, Wang S, Melchinger AE, Zeng Z-B (2008) Quantitative trait loci mapping and the genetic basis of heterosis in maize and rice. Genetics 180:1707–1724

Haldane JBS (1919) The combination of linkage values and the calculation of distance between the loci of linkage factors. J Genet 8:299–309

Hill WG (1993) Variation in genetic composition in backcrossing programs. J Hered 84:212–213

Hospital F, Charcosset A (1997) Marker assisted introgression of quantitative trait loci. Genetics 147:1469–1485

Kosambi DD (1944) The estimation of the map distance from recombination values. Ann Eugen 12:172–175

Kusterer B, Piepho H-P, Utz HF, Schön CC, Muminovic J, Meyer RC, Altman T, Melchinger AE (2007) Heterosis for biomass-related traits in Arabidopsis investigated by QTL analysis of the triple testcross design with recombinant inbred lines. Genetics 177:1839–1850

Lisec J, Steinfath M, Meyer RC, Selbig J, Melchinger AE, Willmitzer L, Altman T (2009) Identification of heterotic metabolite QTL in *Arabidopsis thaliana* RIL and IL populations. Plant J 59:777–788

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

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland

Melchinger AE (1987) Expectation of means and variances of testcrosses produced from F_2 and backcross individuals and their selfed progenies. Heredity 59:105–115

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

Paschold A, Marcon C, Hoecker N, Hochholdinger F (2010) Molecular dissection of heterosis manifestation during early maize root development. Theor Appl Genet (this volume)

Schnell FW (1961) Some general formulations of linkage effects in inbreeding. Genetics 46:947–957

Schön CC, Dhillon BS, Utz HF, Melchinger AE (2010) High congruency of QTL positions for heterosis of grain yield in three crosses of maize. Theor Appl Genet (this volume)

Stam P (1979) Interference in genetic crossing over and chromosome mapping. Genetics 92:573–594

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

Tanksley SD, Young ND, Patterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding: new tools for an old science. Bio/Technology 7:257–263

Thiemann A, Fu J, Schrag TA, Melchinger AE, Frisch M, Scholten S (2010) Correlation between parental transcriptome and field data for the characterization and prediction of heterosis in maize. Theor Appl Genet (this volume)

Törjék O, Witucka-Wall H, Meyer RC, von Korff M, Kusterer B, Rautengarten C, Altmann T (2006) Segregation distortion in *Arabidopsis* C24/Col-0 and Col-0/C24 recombinant inbred line populations is due to reduced fertility caused by epistatic interaction of two loci. Theor Appl Genet 113:1551–1561