Theor Appl Genet (1992) 83:628–634

THEORETICAL AND APPLIED GENETICS

Relationship between single-cross performance and molecular marker heterozygosity *

R. Bernardo

Limagrain Genetics, 737 County Rd., 1200N, Champaign, IL 61821, USA

Received June 1, 1991; Accepted July 26, 1991
Communicated by A. R. Hallauer

Summary. In studies involving isozymes or restriction fragment length polymorphisms (RFLPs), correlations of parental molecular marker diversity with grain yield of maize (*Zea mays* L.) single-crosses have been too low to be of any predictive value. The relationship of molecular marker heterozygosity (D_{ij}) with hybrid performance (μ_{ij}) and combining ability was examined. For a simple genetic model involving uncorrelated parental allele frequencies and complete coverage of quantitative trait loci (QTL) by molecular markers, the correlations between μ_{ij} and D_{ij} were ≤ 0.25. μ_{ij} and D_{ij} were partitioned into general and specific effects. The expected correlation between specific combining ability and specific molecular marker heterozygosity is high. Expected correlations between general combining ability and general molecular marker heterozygosity are either positive or negative, depending on allele frequencies in the tester lines. Computer simulation was used to investigate a more complex but more realistic genetic model involving incomplete coverage of QTL by molecular markers. All of the following conditions are necessary for effective prediction of hybrid performance based on molecular marker heterozygosity: (1) dominance effects are strong; (2) allele frequencies at individual loci in the parental inbreds are negatively correlated; (3) trait heritability is high; (4) average parental allele frequencies vary only within a narrow range; (5) at least 30–50% of the QTL are linked to molecular markers; and (6) not more than 20–30% of the molecular markers are randomly dispersed or unlinked to QTL.

Key words: Isozymes – RFLPs – Heterozygosity – Hybrid performance – Combining ability

* A contribution from Limagrain Genetics, a Groupe Limagrain company

Introduction

In maize (*Zea mays* L.) hybrid breeding programs, the identification of inbred lines with superior yield performance in single-cross combination is costly and time-consuming. Because of strong dominance effects for maize grain yield (Hallauer and Miranda 1981), hybrid performance cannot be predicted from inbred per se data (Smith 1986). Thus, it is necessary to make actual crosses between lines and evaluate the crosses in extensive yield trials.

The use of line per se molecular marker data has been suggested as a means of predicting hybrid performance prior to making and evaluating the actual crosses themselves. Considerable research is currently being conducted to determine the relationship of molecular marker diversity with hybrid performance and heterosis. In studies using isozymes (Hunter and Kannenberg 1971; Heidrich-Sobrinho and Cordiero 1975; Gonella and Peterson 1978; Price et al. 1986; Frei et al. 1986; Lamkey et al. 1987) and restriction fragment length polymorphisms or RFLPs (Lee et al. 1989; Godshalk et al. 1990; Melchinger et al. 1990; Dudley et al. 1991) the correlations of hybrid performance with molecular marker diversity between parents have been too low to be of any predictive value. Inadequate genome coverage, randomly dispersed (unlinked to quantitative trait loci or QTL) molecular markers, and different levels of dominance among hybrids have been suggested as possible reasons for the low correlations obtained between hybrid performance and marker diversity.

The objective of this study was to examine the theoretical relationship of molecular marker heterozygosity with hybrid performance and general and specific combining abilities. The effects of the following factors on the correlation between hybrid performance and molecu-

lar marker heterozygosity were investigated: (1) level of dominance; (2) covariance of parental allele frequencies; (3) amount of variation in parental allele frequencies; (4) proportion of QTL linked to molecular marker loci; and (5) proportion of randomly dispersed molecular marker loci.

Theory and computer simulation

Assume that a series of homozygous lines in Set A is crossed to a series of homozygous lines in Set B in a factorial mating scheme. The Set A and Set B lines may originate from the same population, but in applied maize breeding programs, the Set A and Set B lines typically belong to different heterotic groups (e.g., Iowa Stiff Stalk Synthetic- and Lancaster Sure Crop-type lines).

Let n ($k=1, \ldots, n$) be the number of quantitative trait loci (QTL) affecting a trait. When the ith Set A line (A_i) is crossed to the jth Set B line (B_j), the mean of the $A_i \times B_j$ cross at the kth QTL is

$$\mu_{ijk} = a_k (p_{ik} + r_{jk} - 1) + d_k (p_{ik} + r_{jk} - 2 p_{ik} r_{jk}),$$

where: p_{ik} = frequency (1 or 0) of the + allele at the kth locus in A_i; r_{jk} = frequency (1 or 0) of the + allele at the kth locus in B_j; a_k = half the difference between coded values of the ++ and −− homozygotes at the kth locus; and d_k = coded value of the +− genotype at the kth locus. In the absence of epistasis among QTL, the summed effect over n loci is

$$\mu_{ij.} = \left(\sum_k p_{ik} a_k + \sum_k r_{jk} a_k - \sum_k a_k\right) + \left(\sum_k p_{ik} d_k + \sum_k r_{jk} d_k - 2\sum_k p_{ik} r_{jk} d_k\right)$$

If a_k and d_k are uncorrelated with parental allele frequencies or if a_k and d_k are constant for each of n QTL, the average effect at a single QTL is obtained by dividing $\mu_{ij.}$ by n:

$$\mu_{ij} = a (p_i + r_j - 1) + d \left(p_i + r_j - 2\sum_k p_{ik} r_{jk}/n\right),$$

where: p_i = average frequency $\left(\sum_k p_{ik}/n\right)$ of the + allele over n loci in A_i; r_j = average frequency $\left(\sum_k r_{jk}/n\right)$ of the + allele over n loci in B_j; $a = \sum_k a_k/n$ and $d = \sum_k d_k/n$.

The term $\sum_k p_{ik} r_{jk}/n$ is equal to $\sigma_{pirj} + p_i r_j$, where σ_{pirj} is the covariance between frequencies of the + allele in A_i and B_j. If lines in Set A and B are chosen at random from the same random-mating source population, zero values of σ_{pirj} are expected. But if the Set A and Set B lines are from complementary heterotic groups, negative σ_{pirj} values are likely and p_{ik} and r_{jk} are negatively correlated; i.e., Set A lines tend to have the + allele at loci where Set B lines have the − allele and vice-versa.

Genetic model involving uncorrelated parental allele frequencies ($\sigma_{pirj}=0$)

With zero σ_{pirj} values, the mean (expressed in terms of the average at one QTL) of the $A_i \times B_j$ cross is

$$\mu_{ij} = a (p_i + r_j - 1) + d (p_i + r_j - 2 p_i r_j)$$

The effectiveness of molecular markers in predicting hybrid performance should increase as the amount of variation in quantitative trait values accounted for by molecular markers increases. For the zero σ_{pirj} model, tight linkage (no recombination) between each QTL and a corresponding molecular marker is assumed. Given this assumption, the results in this paper for the zero σ_{pirj} model reflect the maximum correlations possible between molecular marker heterozygosity and hybrid performance. If each QTL is linked to a single molecular marker, the average frequencies of molecular marker alleles linked to + alleles at QTL are equal to p_i in A_i and r_j in B_j. The average molecular marker heterozygosity expected in the $A_i \times B_j$ cross is

$$D_{ij} = (p_i + r_j - 2 p_i r_j)$$

μ_{ij} values can be partitioned into general and specific combining ability effects (Griffing 1956):

$$\mu_{ij} = \mu + GCA_i + GCA_j + SCA_{ij},$$

where: μ = population mean; GCA_i = general combining ability effect of A_i when crossed to a series of Set B lines; GCA_j = general combining ability effect of B_j when crossed to a series of Set A lines; and SCA_{ij} = specific combining ability effect of the $A_i \times B_j$ cross. Similarly, D_{ij} values can be partitioned into general (GD_i and GD_j) and specific distances (SD_{ij}) (Melchinger et al. 1990):

$$D_{ij} = \bar{D}_{ij} + GD_i + GD_j + SD_{ij}$$

Genetic model involving negatively correlated parental allele frequencies ($\sigma_{pirj}<0$)

The zero σ_{pirj} genetic model involves two simplifying assumptions which may be unrealistic: (1) heterozygosity in the $A_i \times B_j$ single-cross is determined solely by the average parental allele frequencies (p_i and r_j, respectively), and (2) an exact, one-to-one correspondence exists between the QTL and molecular marker loci. If σ_{pirj} is nonzero, single-cross performance is equal to

$$\mu_{ij} = a [p_i + r_j - 1] + d [p_i + r_j - 2 (\sigma_{pirj} + p_i r_j)] \quad (1)$$

The average heterozygosity at a QTL in the single-cross is equal to

$$H_{ij} = p_i + r_j - 2 (\sigma_{pirj} + p_i r_j)$$

H_{ij} will be greater if QTL allele frequencies in A_i and B_j are negatively correlated (negative σ_{pirj} values) than when parental allele frequencies are uncorrelated. Values of

σ_{pirj} are unknown but the minimum (<0) possible value of σ_{pirj} can be derived. σ_{pirj} is equal to $\sum_k p_{ik} r_{jk}/n - p_i r_j$. For specific values of p_i and r_j, σ_{pirj} is minimum if $\sum_k p_{ik} r_{jk}/n$ is minimum. Given a binomial distribution of allele frequencies in homozygous lines, the smallest possible value of $\sum_k p_{ik} r_{jk}/n$ is zero if $(p_i+r_j)\leq 1$. If $(p_i+r_j)>1$, some QTL in the single-cross will be homozygous for the + allele such that $\sum_k p_{ik} r_{jk}/n>0$. The smallest possible value of $\sum_k p_{ik} r_{jk}/n$ is equal to (p_i+r_j-1) if $(p_i+r_j)>1$. Thus, the minimum value of σ_{pirj} is min $(\sigma_{pirj})=-p_i r_j$ if $(p_i+r_j)\leq 1$. If $(p_i+r_j)>1$, the minimum value of σ_{pirj} is min $(\sigma_{pirj})=p_i+r_j-1-p_i r_j$.

On the assumption that σ_{pirj} is equal to some value between min (σ_{pirj}) and 0, computer simulation was used to investigate the effects of the following factors on the correlation between hybrid performance and molecular marker heterozygosity: (1) level of dominance; (2) amount of variation in parental allele frequencies; (3) proportion of QTL linked to molecular marker loci; and (4) proportion of randomly dispersed molecular marker loci. A simulation program was written in advanced BASIC and run on a PC-compatible 80386 microcomputer. Partial (value of heterozygote = half the value of + + homozygote) and complete dominance were considered. The range in average parental allele frequencies was either narrow $[0.4\leq (p_i, r_j)\leq 0.6]$ or wide $[0.2\leq (p_i, r_j)\leq 0.8]$. A uniform distribution of parental allele frequencies was assumed. Similarly, the simulated value of σ_{pirj} for each $A_i\times B_j$ cross was uniformly distributed and was equal to x [min (σ_{pirj})], where x is a uniformly distributed random variable ranging from 0 to 1.

A trait controlled by 100 QTL ($n=100$) was considered. Individual effects at the kth QTL (a_k and d_k) were assumed to be uncorrelated with parental allele frequencies. Given this assumption, no further assumption regarding the distribution of QTL effects (i.e., equal or unequal locus effects across n QTL) was necessary; μ_{ij} values were obtained using Eq. 1.

The proportion of QTL linked (no recombination) to molecular markers (c) varied from 0.1 to 1. The proportion of randomly dispersed (unlinked to QTL) molecular markers (y) varied from 0 to 0.8. Heterozygosity at molecular marker loci is a function of both heterozygosity at marker loci linked to QTL and heterozygosity at randomly dispersed marker loci. When a fixed number of QTL is sampled (through QTL-marker linkage), the total number of heterozygous marker loci in the sample follows a hypergeometric distribution. Average heterozygosity at marker loci linked to QTL is denoted $D_{ij'}$ and follows an approximate normal distribution with a mean equal to H_{ij} (i.e., heterozygosity at QTL) and a variance of $\sigma^2_{Dij'}=[H_{ij}(1-H_{ij})(1-c)]/c(n-1)$. If all QTL are marked ($c=1$), $\sigma^2_{Dij'}=0$ and $D_{ij'}=H_{ij}$. On the assumption of random heterozygosity at marker loci unlinked to QTL, simulated values of average heterozygosity at molecular marker loci were calculated as

$$D_{ij}=D_{ij'}(1-y)+x\,y,$$

where x is a uniformly distributed random variable ranging from 0 to 1. Thus, $D_{ij}=D_{ij'}$ if none of the markers is randomly dispersed ($y=0$).

Five thousand $A_i\times B_j$ crosses were simulated for each combination of level of dominance, range of parental allele frequencies, proportion of marked QTL, and proportion of randomly dispersed markers. Correlations between molecular marker heterozygosity and hybrid performance were calculated.

Results and discussion

Expectations of parameters for the zero σ_{pirj} model

The expectations for the zero σ_{pirj} model of single-cross mean performance (μ_{ij}) and molecular marker heterozygosity (D_{ij}) indicate that D_{ij} can account only for variation in μ_{ij} due to dominance effects (Table 1). Both D_{ij} and the portion of μ_{ij} due to dominance effects (d) are functions of $(p_i+r_j-2p_i r_j)$. If dominance is partial to complete ($d\leq a$), as accumulated data on maize grain yield indicate (Hallauer and Miranda 1981), μ_{ij} increases as both p_i and r_j approach 1. But D_{ij} is maximum when p_i approaches 1 and r_j approaches 0, or vice-versa. Because of the different combinations of parental allele frequencies required to attain maximum values of μ_{ij} and

Table 1. Expectations of single-cross ($A_i\times B_j$) performance and molecular marker (MM) heterozygosity and their components[a]

Quantitative trait	MM heterozygosity
$\mu_{ij}=a(p_i+r_j-1)+d(p_i+r_j-2p_ir_j)$[b]	$D_{ij}=(p_i+r_j-2p_ir_j)$
$GCA_i=(p_i-\bar{p})[a+d(1-2\bar{r})]$	$GD_i=(p_i-\bar{p})(1-2\bar{r})$
$GCA_j=(r_j-\bar{r})[a+d(1-2\bar{p})]$	$GD_j=(r_j-\bar{r})(1-2\bar{p})$
$TC_i=p_i[a+d(1-2r_j)]+[a(r_j-1)+r_jd]$	$TD_i=p_i(1-2r_j)+r_j$
$TC_j=r_j[a+d(1-2p_i)]+[a(p_i-1)+p_id]$	$TD_j=r_j(1-2p_i)+p_i$
$SCA_{ij}=2d[(p_i-\bar{p})(\bar{r}-r_j)]$ $=2d[(r_j-\bar{r})(\bar{p}-p_i)]$	$SD_{ij}=2[(p_i-\bar{p})(\bar{r}-r_j)]$ $=2[(r_j-\bar{r})(\bar{p}-p_i)]$

[a] μ_{ij}, Performance of $A_i\times B_j$ single-cross; D_{ij}, MM heterozygosity in $A_i\times B_j$; $GCA_{i(j)}$, general combining ability of A_i (B_j); $GD_{i(j)}$, general MM heterozygosity of A_i (B_j); $TC_{i(j)}$, testcross performance of A_i (B_j) lines when crossed to a single B_j (A_i) line; $TD_{i(j)}$, testcross MM heterozygosity of A_i (B_j); SCA_{ij}, specific combining ability of $A_i\times B_j$; and SD_{ij}, specific MM heterozygosity of $A_i\times B_j$

[b] p_i and r_j, Frequencies of + allele in A_i and B_j, respectively; $\bar{p}$ and $\bar{r}$, average frequencies of + allele in a series of A_i and B_j lines, respectively; a, half the difference between coded genetic values of + + and − − homozygotes; d, coded value of heterozygote

D_{ij}, a strong linear relationship between hybrid performance and molecular marker heterozygosity is not expected.

Midparent heterosis in the $A_i \times B_j$ cross is equal to d $(p_i + r_j - 2\, p_i\, r_j)$, i.e., the portion of μ_{ij} due to dominance effects. Expectations indicate that both specific combining ability (SCA_{ij}) and specific molecular marker heterozygosity (SD_{ij}) in the $A_i \times B_j$ cross are functions of $[(p_i - \bar{p})(\bar{r} - r_j)]$, which is also equal to $[(r_j - \bar{r})(\bar{p} - p_i)]$ (Table 1). Thus, a correlation of 1.0 is expected between midparent heterosis and D_{ij} and between SCA_{ij} and SD_{ij} regardless of the level of dominance (including overdominance). But hybrid combinations that exhibit large midparent heterosis or specific combining ability effects do not necessarily have the best hybrid performance, which is the breeder's main objective in varietal development. Also, it is assumed in the derivations for the zero σ_{pirj} model that each QTL is tightly linked to a molecular marker. If some molecular markers are randomly dispersed and are unlinked to QTL, the correlations of midparent heterosis with D_{ij} and of SCA_{ij} with SD_{ij} will be less than 1.0.

The expectation of general combining ability of A_i lines (GCA_i) when crossed to B_j lines is $(p_i - \bar{p})\,[a + d(1 - 2\,\bar{r})]$, where $\bar{p}$ and $\bar{r}$ are the average frequencies of the + allele among the lines in Set A and Set B, respectively (Table 1). Regardless of the average frequency of the + allele ($\bar{r}$) in the tester lines, the term $[a + d\,(1 - 2\,\bar{r})]$ will be positive if dominance is partial or complete ($d \leq a$). Thus, GCA_i increases as the frequency of + alleles (p_i) in A_i increases. The expectation of general molecular marker heterozygosity of A_i lines (GD_i) is $(p_i - \bar{p})\,(1 - 2\,\bar{r})$. GD_i increases with larger p_i values only if $\bar{r}$ is less than 0.5. If $\bar{r}$ is greater than 0.5, GD_i decreases with larger p_i values. When B_j lines are crossed to A_i lines, similar relationships exist between general combining ability (GCA_j) and general molecular marker distance (GD_j). Hence, a positive relationship between general combining ability and general molecular marker heterozygosity is expected if tester frequencies for the + allele are less than 0.5. But if tester frequencies for the + allele are greater than 0.5, general combining ability increases as general marker heterozygosity decreases.

Typically, elite inbred lines are available in applied hybrid development programs. A breeder may want to find an inbred line to complement an elite line in hybrid combination. Based on expectations (Table 1), the testcross performance of A_i lines (TC_i) when crossed to a single B_j line increases as the frequency of the + allele in A_i (p_i) increases. Similar to GCA_i, this positive relationship between p_i and TC_i holds true regardless of the frequency of the + allele in the fixed B_j tester (r_j) if dominance is partial to complete. But the testcross molecular marker distance of A_i lines (TD_i) increases with larger p_i values only if r_j is less than 0.5. TD_i decreases with larger p_i values if r_j is greater than 0.5, causing a negative relationship between TC_i and TD_i. When B_j lines are crossed to a single A_i line, the same results are obtained for testcross performance (TC_j) and testcross molecular marker distance (TD_j). Thus, for inbred testers with average (across QTL) frequencies of + alleles greater than 0.5, testcross performance is expected to increase as testcross molecular marker heterozygosity decreases.

Numerical results for the zero σ_{pirj} model

On the assumption of uncorrelated parental allele frequencies ($\sigma_{pirj} = 0$), values of hybrid performance (μ_{ij}) and molecular marker heterozygosity (D_{ij}) for the $A_i \times B_j$ cross were calculated assuming partial ($d = a/2$) or complete dominance and frequencies of the + allele of $0.3 \leq p_i \leq 0.8$ in A_i and $0.2 \leq r_j \leq 0.7$ in B_j. The numerical results illustrate the absence of a consistent relationship between μ_{ij} and D_{ij} (Table 2). μ_{ij} was largest when frequencies of the + allele were closest to 1 ($\mu_{ij} = 88$ when $p_i = 0.8$ and $r_j = 0.7$). However, D_{ij} was largest when the frequency of the + allele approached 1 in one parent and 0 in the other parent ($D_{ij} = 0.68$ when $p_i = 0.8$ and $r_j = 0.2$). Multiple μ_{ij} values were observed at a given level of molecular marker heterozygosity (Fig. 1). For example, μ_{ij} ranged from 20 to 80 when $D_{ij} = 50\%$. For the values of p_i and r_j considered in this paper, the correlation between μ_{ij} and D_{ij} was $r_{\mu ij\, Dij} = 0.25$. With partial dominance, the correlation between μ_{ij} and D_{ij} decreased to $r_{\mu ijDij} = 0.13$. In empirical studies on the relationship of grain yield performance of maize single-crosses with RFLP distances between parental lines, correlations of 0.09, 0.14, 0.32, and 0.46 were obtained by Godshalk et al. (1990), Dudley et al. (1991), Melchinger et al. (1990), and Lee et al. (1989), respectively. These correlations between hybrid performance and molecular marker distance are too small to be of any predictive value.

$r_{\mu ijDij}$ values were affected by different restrictions on parental allele frequencies. With complete dominance, $r_{\mu ijDij} = 0.95$ when both p_i and r_j ranged from 0.1 to 0.5. $r_{\mu ijDij} = 0.41$ when p_i ranged from 0.1 to 0.5 and r_j ranged from 0.5 to 0.9. However, a negative correlation of $r_{\mu ijDij} = -0.71$ was obtained when both p_i and r_j ranged from 0.5 to 0.9. Although these sets of restrictions on allele frequencies may be unrealistic, the results indicate that both the sign and size of the correlation between hybrid performance and molecular marker heterozygosity are highly dependent on QTL allele frequencies in the material being tested.

The relationship between general combining ability (GCA) and general molecular marker heterozygosity (GD) depends on average allele frequencies in the tester lines (Table 2). Average frequencies of the + allele were $\bar{p} = 0.55$ for the A_i lines and $\bar{r} = 0.45$ for B_j lines. GCA_i and

Table 2. Single-cross ($A_i \times B_j$) performance (μ_{ij}) and molecular marker heterozygosity (D_{ij}, in parentheses) with complete dominance and different allele frequencies[a]

p_i	r_j						r_{TCjTDj}	GCA_i	GD_i
	0.2	0.3	0.4	0.5	0.6	0.7			
0.3	−12[b]	2	16	30	44	58	1.0	−27.5	
	(0.38)	(0.42)	(0.46)	(0.50)	(0.54)	(0.58)			−0.025
0.4	4	16	28	40	52	64	1.0	−16.5	
	(0.44)	(0.46)	(0.48)	(0.50)	(0.52)	(0.54)			−0.015
0.5	20	30	40	50	60	70	–	−5.5	
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)			−0.005
0.6	36	44	52	60	68	76	−1.0	5.5	
	(0.56)	(0.54)	(0.52)	(0.50)	(0.48)	(0.46)			0.005
0.7	52	58	64	70	76	82	−1.0	16.5	
	(0.62)	(0.58)	(0.54)	(0.50)	(0.46)	(0.42)			0.015
0.8	68	72	76	80	84	88	−1.0	27.5	
	(0.68)	(0.62)	(0.56)	(0.50)	(0.44)	(0.38)			0.025
r_{TCiTDi}	1.0	1.0	1.0	–	−1.0	−1.0			
GCA_j	−22.5	−13.5	−4.5	4.5	13.5	22.5			
GD_j	0.025	0.015	0.005	−0.005	−0.015	−0.025			

[a] p_i, Frequency of + allele in A_i; r_j, frequency of + allele in B_j; r_{TCiTDi}, correlation between testcross performance and molecular marker heterozygosity when A_i lines are crossed to a single B_j line; r_{TCjTDj}, correlation between testcross performance and molecular marker heterozygosity when B_j lines are crossed to a single A_i line; $GCA_{i(j)}$, general combining ability of A_i (B_j) line; $GD_{i(j)}$, general molecular marker heterozygosity of A_i (B_j) line

[b] $\mu_{ij} = a(p_i + r_j - 1) + d(p_i + r_j - 2p_i r_j)$, where $a = d = 100$; $D_{ij} = (p_i + r_j - 2p_i r_j)$

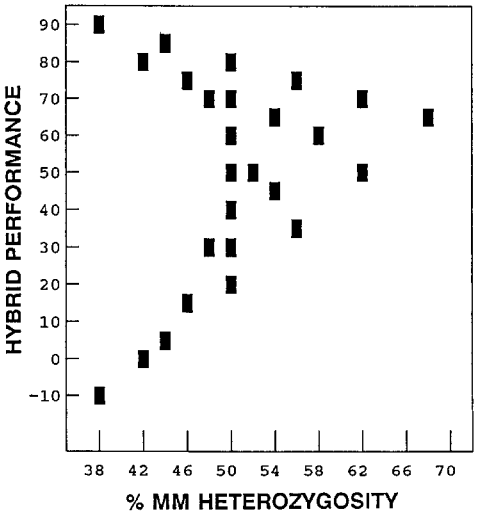

Fig. 1. Relationship between molecular marker (*MM*) heterozygosity and single-cross hybrid performance for a genetic model involving complete dominance and uncorrelated allele frequencies in the parental lines

GD_i were positively correlated ($r_{GCAiGDi} = 1.0$) because the A_i lines were tested against B_j lines wherein the average frequency of the + allele was less than 0.5. In contrast, GCA_j and GD_j were negatively correlated ($r_{GCAjGDj} = -1.0$) because the B_j lines were tested against A_i lines wherein the average frequency of the + allele was greater than 0.5. Similar results were obtained for testcross performance (TC) and testcross molecular marker heterozygosity (TD) when A_i or B_j lines were crossed to a single B_j or A_i line, respectively. r_{TCiTDi} and r_{TCjTDj} values were either 1.0 or −1.0, depending on the average frequency of + alleles in the inbred tester used. For the zero σ_{pirj} model, there is no consistent relationship between performance and molecular marker heterozygosity of lines in testcross combination.

As expected, the correlation between specific combining ability (SCA_{ij}) and specific molecular marker heterozygosity (SD_{ij}) in the $A_i \times B_j$ crosses was equal to 1.0 regardless of the level of dominance. In Lee et al.'s (1989) study involving an eight-parent diallel, the correlation between SCA_{ij} and SD_{ij} for maize grain yield was $r_{SCAijSDij} = 0.82$. However, this high correlation was obtained for crosses within (wherein low SCA_{ij} values were expected) and between (wherein high SCA_{ij} values were expected) heterotic groups. When the analysis is restricted to crosses between heterotic groups (Stiff Stalk × Lancaster lines), $r_{SCAijSDij}$ decreases from 0.82 to −0.17.

Simulation results for the genetic model involving negatively correlated parental allele frequencies ($\sigma_{pirj} < 0$)

With negatively correlated p_{ik} and r_{jk} values, correlations between hybrid performance (μ_{ij}) and molecular marker heterozygosity (D_{ij}) were larger with complete than with partial dominance (Table 3). For values of p_i and r_j ranging from 0.4 to 0.6, the $r_{\mu ijDij}$ values ranged from ≤ 0.10 to 0.84 if dominance was complete and from < 0.10 to 0.61 if dominance was partial. Similar to the zero σ_{pirj} model, the $r_{\mu ijDij}$ values were affected by the amount of

variation in parental allele frequencies. Compared with $r_{\mu ijDij}$ values obtained when p_i and r_j ranged from 0.4 to 0.6, the values of $r_{\mu ijDij}$ were reduced by roughly 50% when p_i and r_j ranged from 0.2 to 0.8 (results not shown). Variation in the amount of heterozygosity among crosses is due to differences in both σ_{pirj} and average parental allele frequencies. If average allele frequencies at QTL do not vary within sets of A_i and B_j lines, variation in μ_{ij} among $A_i \times B_j$ crosses is solely due to different σ_{pirj} values for each crosses. In this extreme situation, a high correlation between μ_{ij} and D_{ij} is expected because differences in hybrid performance are due to dominance-associated effects only. Thus, higher $r_{\mu ijDij}$ values are expected when variation in p_i and r_j among lines is small [$0.4 \leq (p_i, r_j) \leq 0.6$] than when variation in p_i and r_j is large [$0.2 \leq (p_i, r_j) \leq 0.8$].

As expected, the correlation between hybrid performance and molecular marker heterozygosity increased as the proportion of QTL linked to molecular markers (c) increased (Table 3). Regardless of the level of dominance, range of parental allele frequencies, or proportion of dispersed (unlinked to QTL) molecular markers, a system of diminishing returns was associated with increases in c. Increments in $r_{\mu ijDij}$ were small beyond 30–50% QTL coverage by marker loci. Hence, not all QTL need to be tagged with molecular markers to have good estimates of average QTL heterozygosity. Sampling (through linkage with molecular markers) 30–50 out of 100 QTL will provide adequate estimates of average heterozygosity across all 100 QTL.

Based on their effects on $r_{\mu ijDij}$ values, the proportion of dispersed molecular markers (y) is more important than the proportion of QTL linked to molecular marker loci (c). Differences in $r_{\mu ijDij}$ values due to different values of c were most evident when $y = 0$. The effects of c were small if y was greater than 30%. When 80% of the molecular markers were dispersed ($y = 80\%$), $r_{\mu ijDij}$ values were ≤ 0.10 and were not affected by c. Thus, arbitrarily chosen molecular marker loci will not provide good estimates of QTL heterozygosity. Rather, it is necessary to identify a set of molecular markers that are mostly linked to QTL. In determining QTL-marker linkage, a Type 1 error or false positive occurs if it is erroneously concluded that a QTL-marker linkage exists when in fact no such association exists. A Type 2 error occurs if it is erroneously concluded that no QTL-marker linkage exists when in fact such association does exist. Because $r_{\mu ijDij}$ values are affected to a greater extent by the proportion of dispersed markers than by the proportion of QTL covered by markers, Type 1 errors or false positives should be reduced even at the risk of having more Type 2 errors.

If the molecular markers used are mostly linked to QTL, molecular marker heterozygosity may not be the most effective predictor of hybrid performance. Rather,

Table 3. Correlation[a] between molecular marker heterozygosity and hybrid performance when parental allele frequencies (p_i and r_j) are negatively correlated ($\sigma_{pirj} < 0$) and $0.4 \leq (p_i, r_j) \leq 0.6$

Dominance	% QTL coverage	% Dispersed molecular marker loci[c]								
		0	10	20	30	40	50	60	70	80
Complete	10	0.60	0.59	0.55	0.49	0.40	0.32	0.24	0.14	0.09
	20	0.70	0.69	0.64	0.54	0.44	0.34	0.22	0.18	0.10
	30	0.76	0.73	0.66	0.57	0.45	0.34	0.26	0.15	0.08
	40	0.79	0.76	0.70	0.58	0.47	0.33	0.24	0.16	0.09
	50	0.80	0.77	0.71	0.59	0.45	0.32	0.26	0.18	0.09
	60	0.81	0.79	0.71	0.60	0.45	0.33	0.23	0.16	0.10
	70	0.83	0.80	0.72	0.59	0.45	0.34	0.24	0.13	0.08
	80	0.83	0.81	0.73	0.60	0.46	0.35	0.23	0.16	0.09
	90	0.84	0.81	0.73	0.60	0.47	0.33	0.24	0.14	0.09
	100	0.84	0.82	0.74	0.59	0.49	0.34	0.25	0.14	0.08
Partial[b]	10	0.44	0.44	0.40	0.36	0.30	0.24	0.18	0.10	0.07
	20	0.51	0.51	0.47	0.39	0.32	0.25	0.16	0.14	0.07
	30	0.56	0.53	0.48	0.42	0.33	0.25	0.20	0.11	0.05
	40	0.59	0.57	0.52	0.41	0.35	0.24	0.18	0.12	0.08
	50	0.59	0.56	0.52	0.44	0.32	0.23	0.20	0.14	0.07
	60	0.60	0.58	0.52	0.44	0.33	0.24	0.16	0.12	0.08
	70	0.61	0.60	0.53	0.43	0.33	0.25	0.18	0.10	0.06
	80	0.61	0.60	0.54	0.44	0.34	0.25	0.17	0.11	0.06
	90	0.61	0.60	0.54	0.44	0.34	0.24	0.18	0.10	0.07
	100	0.61	0.61	0.55	0.43	0.37	0.25	0.18	0.10	0.06

[a] Correlations were calculated for a data set of 5,000 simulated $A_i \times B_j$ crosses. All correlation coefficients were significant at $\alpha = 0.01$

[b] With partial dominance, the value of the heterozygote is half the value of the ++ homozygote

[c] % Dispersed molecular marker loci = % of molecular marker loci unlinked to QTL; % QTL coverage = % of QTL loci linked to molecular markers

the particular isozyme or RFLP variant at a given locus that is linked to the favorable allele at a QTL may be determined as part of the QTL-tagging procedure. Crosses can then be made between lines that are expected to result in the highest concentration of favorable alleles (either in homozygous and heterozygous state) in the single-cross. This approach might provide a better prediction of hybrid performance because, in contrast to molecular marker heterozygosity, variation in hybrid performance due to both homozygote- and dominance-associated effects is accounted for.

Some of the $r_{\mu ijDij}$ values in Table 3 may be high enough to allow preliminary identification of superior $A_i \times B_j$ crosses. $r_{\mu ijDij}$ is roughly ≥ 0.6 when dominance is complete, at least 30% of the QTL are linked to markers, and less than 30% of the markers are randomly dispersed. But the results in this study assume a heritability (h^2) of 1 for the quantitative trait so that the observed phenotypic value of a hybrid is the same as its genotypic value. $r_{\mu ijDij}$ will be severely reduced if h^2 is low. With $h^2 < 1$, the $r_{\mu ijDij}$ values will be reduced by a factor equal to the square root of h^2. The highest $r_{\mu ijDij}$ value in Table 3 is 0.84 (for complete dominance, 100% of the QTL linked to markers, and no randomly dispersed markers). This $r_{\mu ijDij}$ value decreases from 0.84 to 0.59 if $h^2 = 50\%$.

Because empirical data indicate partial to complete dominance for maize grain yield (Hallauer and Miranda 1981), the correlation between hybrid performance and molecular marker heterozygosity in the presence of overdominance has not been addressed thus far. However, pseudo-overdominance may result if two QTL, each exhibiting partial to complete dominance, are linked in repulsion phase. Computer simulation was done to obtain $r_{\mu ijDij}$ values with overdominance (value of heterozygote = 1.5 times the value of + + homozygote). $r_{\mu ijDij}$ values with overdominance (results not shown) ranged from less than 0.10 when 80% of the markers are randomly dispersed to 0.92 when none of the markers is dispersed and all QTL are linked to markers. Compared to values obtained with complete dominance (from Table 3, in parentheses), $r_{\mu ijDij}$ values with overdominance ranged from 0.62 (0.57) to 0.87 (0.80) when 30–50% of the QTL are linked to markers and $\leq 30\%$ of the markers are randomly dispersed. Thus, the association between hybrid performance and molecular marker heterozygosity becomes stronger as the level of dominance increases.

The results in this study indicate that the relationship between molecular marker heterozygosity and hybrid performance is highly dependent on the material being tested. Because a strong and/or consistent relationship does not exist, information on isozyme or RFLP heterozygosity is of limited value in identifying inbred lines with superior performance in hybrid combination. Several conditions are necessary for a high correlation between $A_i \times B_j$ hybrid performance and molecular marker heterozygosity. Molecular marker heterozygosity would be most useful for predicting hybrid performance in crop species wherein (1) dominance effects are strong (e.g., complete dominance or overdominance) and (2) heterotic groups are complementary and allele frequencies at individual loci in the parental inbreds are negatively correlated. These two requirements are fulfilled in a crop species such as maize. Furthermore, (3) trait heritability must be high, (4) average parental allele frequencies must vary only within a narrow range, (5) at least 30–50% of the QTL must be linked to molecular markers, and (6) not more than 20–30% of the molecular markers must be randomly dispersed or unlinked to QTL. If any of these conditions are not met, small correlations between hybrid performance and molecular marker heterozygosity are likely.

References

Dudley JW, Saghai-Maroof MA, Rufener GK (1991) Molecular markers and grouping of parents in maize breeding programs. Crop Sci 31:718–723

Frei OM, Stuber CW, Goodman MM (1986) Uses of allozymes as genetic markers for predicting performance in maize single cross hybrids. Crop Sci 26:37–42

Godshalk EB, Lee M, Lamkey KR (1990) Relationship of restriction fragment length polymorphisms to single-cross hybrid performance of maize. Theor Appl Genet 80:273–280

Gonella JA, Peterson PA (1978) Isozyme relatedness of inbred lines of maize and performance of their hybrids. Maydica 23:55–61

Griffing B (1956) Concept of general and specific combining ability in relation to diallel crossing systems. Aust J Biol Sci 9:463–493

Hallauer AR, Miranda JB Fo (1981) Quantitative genetics in maize breeding. Iowa State University Press, Ames, Iowa

Heidrich-Sobrino E, Cordiero AR (1975) Codominant isozymic alleles as markers of genetic diversity correlated with heterosis in maize (*Zea mays* L.). Theor Appl Genet 46:197–199

Hunter RB, Kannenberg LW (1971) Isozyme characterization of corn (*Zea mays* L.) inbreds and its relation to single cross hybrid performance. Can J Genet Cytol 13:649–655

Lamkey KR, Hallauer AR, Kahler AL (1987) Allelic differences at enzyme loci and hybrid performance in maize. J Hered 78:231–234

Lee M, Godshalk EB, Lamkey KR, Woodman WW (1989) Association of restriction fragment length polymorphisms among maize inbreds with agronomic performance of their crosses. Crop Sci 29:1067–1071

Melchinger AE, Lee M, Lamkey KR, Woodman WL (1990) Genetic diversity for restriction fragment length polymorphisms: Relation to estimated genetic effects in maize inbreds. Crop Sci 30:1033–1040

Price SC, Kahler AL, Hallauer AR, Charmley P, Giegel DA (1986) Relationships between performance and multilocus heterozygosity at enzyme loci in single-cross hybrids of maize. J Hered 77:341–344

Smith OS (1986) Covariance between line per se and testcross performance. Crop Sci 26:540–543