Copyright © 1993 by the Genetics Society of America

Efficiency of Generations for Estimating Marker-Associated QTL Effects by Multiple Regression

J. Moreno-Gonzalez

Centro de Investigaciones Agrarias de Mabegondo. Apartado 10, 15080 La Coruña, Spain

Manuscript received November 14, 1992
Accepted for publication May 14, 1993

ABSTRACT

Knowledge about the efficiency of generations for estimating marker-associated QTLs is needed for selection. The objective of this paper is to develop a theory to compare the efficiency of segregating generations and testcrosses from the cross of two inbred lines differing in value for a quantitative trait ($P_1 \times P_2$) for estimating additive, dominance and heterotic effects of QTLs by stepwise regression. An equation that predicts the smallest gene effect in genetic standard deviation units that can be detected with 50% chance at a significance level as a function of the heritability (h^2) and the recombination frequency (r) of markers was developed for the segregating generations and testcrosses. For estimating additive effects, the most efficient generation was the doubled-haploid (DH) lines; the most inefficient was the North Carolina Design III (NCD III), followed by selfed backcrosses (SB); the selfed families from F_2 individual plants ($F_{2:3}$ lines) are inferior to the recombinant inbreds (RI) for low r, but are better than RI for high h^2 and r. Dominance effects are less efficiently estimated than additive effects. The NCD III is better than the SB and the $F_{2:3}$ lines for detecting dominance effects. The RI and DH do not estimate dominance effects. The differential heterotic QTL effects of lines P_1 and P_2 when crossed with tester T can be estimated by evaluating testcrosses of individual F_2 plants (F2T), recombinant inbreds (RIT) and double-haploid lines (DHT). The DHT is superior to the other generations. The F2T is better than the RIT for $r \geq 0.20$, but inferior for $r \leq 0.1$ or low heritability.

THEORY on the identification and utilization of quantitative trait loci (QTL) associated with markers has been formulated (THODAY 1961; JAYAKAR 1970; MCMILLAN and ROBERTSON 1974; SOLLER and BECKMAN 1983; LANDER and BOTSTEIN 1989; KNAPP, BRIDGES and BIRKES 1990; LANDE and THOMPSON 1990). Several statistical methods that include contrasting marker means (SOLLER and BECKMAN 1983), multiple regression (COWEN 1989; MORENO-GONZALEZ 1992a) and maximum likelihood (LANDER and BOTSTEIN 1989) have been suggested to estimate the QTL effects. The use of individual markers and flanking markers associated with QTLs has also been studied. However, no definitive method of mapping QTLs is still available. These approaches all have advantages and shortcomings. Application of the theory to breeding programs needs to answer some important questions: Which type of segregating generation estimates most efficiently the additive and dominance effects of the QTLs? How do the number of progeny, mapping marker density, heritability and genetic variance of the trait affect the estimates? How can QTLs with specific heterotic effects be identified and selected for use in a hybrid breeding program? Different orthogonal generations from the cross of two inbreds have been proposed to estimate QTL effects linked to markers (SOLLER and BECKMAN 1983; COWEN 1988; KNAPP, BRIDGES and BIRKES 1990; MORENO-GONZALEZ 1992a). COWEN (1988) compared the power of some of these generations for estimating gene effects by contrasting phenotypic class means of individual markers associated with one QTL.

The objective of this paper is to develop a theory to compare the efficiency of (a) segregating generations from the cross of two parental inbred lines for estimating the additive and dominance effects of QTLs associated with markers and (b) their testcrosses for estimating the heterotic effects, using stepwise regression.

MATERIALS AND METHODS

Theory: The following segregating generations were studied: selfed families from individual F_2 plants ($F_{2:3}$ lines), selfed backcrosses (SB), recombinant inbreds (RI), doubled-haploid lines (DH) and North Carolina Design III (NCD III) from the cross of two inbred lines ($P_1 \times P_2$). The following model proposed by MORENO-GONZALEZ (1992a) was applied:

$$p_j = \mu_0 + \sum_i (a_i x'_i + d_i y'_i) + \epsilon_j, \quad (1)$$

where p_j is the phenotypic value of line j in generations $F_{2:3}$, SB, RI, DH or NCD III; μ_0 is the mean of QTL genotypes; a_i and d_i are the additive and dominance values of the QTL associated with the marker segment S_i ($i = 1, 2, \ldots n$); x'_i and y'_i are dummy variables associated with a_i and d_i, respectively; ϵ_j is the residual effect associated with line j; values of x'_i and y'_i for the marker classes of generations $F_{2:3}$, SB, RI and DH when $\rho_i = ½$ are derived according to

TABLE 1

Values of the dummy variables x' and y' for the marker classes of segregating generations and their testcrosses for a flanking marker model[a] when $\rho_i = ½$[b] and no epistasis is assumed

Marker class	Coded class	Segregating generations								Testcrosses of generations		
		$F_{2:3}$ lines		Recombinant inbred lines		Selfed backcross		Doubled-haploid lines		F_2	Recombinant inbred lines	Doubled-haploid lines
		x'	y'	x'	y'	x'	y'	x'	y'	x'	x'	x'
$M_iM_iM_{i+1}M_{i+1}$	1	1	0	D[d]	0	1	0	1	0	1	D[d]	1
$M_iM_iM_{i+1}m_{i+1}$	2	0.5	0.25			0.5	0.25			0.5		
$M_iM_im_{i+1}m_{i+1}$	3	0	0.25	0	0			0	0	0	0	0
$M_im_iM_{i+1}M_{i+1}$	4	0.5	0.25			0.5	0.25			0.5		
$M_im_iM_{i+1}m_{i+1}$	5	0	E[e]			0	0.5			0		
$M_im_iM_{i+1}m_{i+1}$	5′					0	0.5					
$M_im_im_{i+1}m_{i+1}$	6	−0.5	0.25			−0.5	0.25			−0.5		
$m_im_iM_{i+1}M_{i+1}$	7	0	0.25	0	0			0	0	0	0	0
$m_im_iM_{i+1}m_{i+1}$	8	−0.5	0.25			−0.5	0.25			−0.5		
$m_im_im_{i+1}m_{i+1}$	9	−1	0	$-D$	0	−1	0	−1	0	−1	$-D$	−1

[a] Model for segregating generations: $\hat{p}_j = \sum_i (a_i x'_i + d_i y'_i)$; model for testcrosses: $t_j = \sum_i h_i x'_i$.

[b] ρ_i is the ratio of the recombination frequencies between *QTL* i and its left-hand side flanking marker relative to that between the two flanking markers M_i and M_{i+1} (r_i).

[c] Marker classes 1, 2, 4 and 5 belong to the backcross to parent 1 and 5′, 6, 8 and 9 belong to the backcross to parent 2.

[d] $D = (1 - r_i^2)/(1 + r_i^2)$; r_i is the recombination frequency between flanking markers.

[e] $E = (2 - 4r_i + 3r_i^2)/2(2 - 4r_i + 4r_i^2)$.

MORENO-GONZALEZ (1992a) and shown in Table 1; ρ_i is the ratio of the recombination frequency between QTL i and a flanking marker to that between the two flanking markers (KNAPP, BRIDGES and BIRKES 1990).

Testcrosses of individual F_2 plants (F2T), recombinant inbreds (RIT) and doubled-haploid lines (DHT) with an inbred tester T were compared for estimating the differential heterotic effect of marker-associated QTLs between the two parental lines. The generation of selfed families (STW) from plants of the three-way cross $(P_1 \times P_2) \times T$ was also included in the same group of comparisons. The following model was used:

$$t_j = \mu_0 + \sum_i h_i x'_i + \epsilon_j, \qquad (2)$$

where t_j is the phenotypic value of the testcross F2T, RIT, DHT or STW; μ_0 is the mean; h_i is one-half the difference of the genotypic values of genotypes ${}^1Q_i{}^TQ_i$ and ${}^2Q_i{}^TQ_i$, where ${}^1Q_i$, ${}^2Q_i$, ${}^TQ_i$ are the alleles of the QTL i associated with the marker segment S_i ($i = 1, 2, \ldots n$) in the inbreds P_1, P_2 and T, respectively; x'_i are dummy variables associated with h_i; ϵ_j is the residual effect of testcross j; values of x'_i for the marker classes of generations F2T, RIT and DHT when $\rho_i = ½$ are shown in Table 1; The values of x'_i for generation STW are the same than those for selfed backcrosses in Table 1.

Stepwise multiple linear regression analysis (DRAPER and SMITH 1981) was applied to both models to estimate gene effects.

Gene effects (g_i) different from zero are detected at the α level with 50% chance (MORENO-GONZALEZ 1992b), provided that:

$$g_i > t_{k,\alpha}(S_e^2 c_{ii})^{½}, \qquad (3)$$

where $t_{k,\alpha}$ is the tabular t-value for the k degrees of freedom of the residual mean squares and the chosen α significance level; S_e^2 is the residual mean square and c_{ii} is a diagonal term of the matrix $[\mathbf{X'X}]^{-1}$; where $\mathbf{X}$ is the design matrix of the multiple regression analysis (DRAPER and SMITH 1981). The expected values of c_{ii} for additive and dominance effects in the selfed backcross generation have been derived (MORENO-GONZALEZ 1992b). Likewise, expected values of c_{ii} for $F_{2:3}$ lines, recombinant inbreds, doubled-haploid lines, NCD III and their testcrosses can also be derived using the same approach and are shown in Table 2. Derivation of c_{ii} values for the $F_{2:3}$ generation is shown in Appendix 1 as an illustrative example.

The residual mean square S_e^2 of Equation 3 has the following components:

$$S_e^2 = \sigma_E^2 + \sigma_{g'}^2 + \Phi^2, \qquad (4)$$

where σ_E^2 is the environmental error variance; $\sigma_{g'}^2$ is the part of the genetic variance (σ_g^2) accounted for by QTL not yet included in the model; $\sigma_g^2 = \sigma_{g'}^2 + \sigma_{g''}^2$; $\sigma_{g''}^2$ is the part of the genetic variance accounted for by QTL in the model; Φ^2 is a component due to the deviations of the assigned genotypic values to marker classes in the model from their real genotypic values. The expected value of Φ^2 for the selfed backcross generation was derived by MORENO-GONZALEZ (1992b). Likewise, expected values for other generations can also be derived using the same approach and are shown in Table 3. Derivation of Φ^2 for the $F_{2:3}$ generation is shown in Appendix 2 as an example.

The following expressions hold:

$$\sigma_{g''}^2 = p\sigma_g^2 \text{ and } \sigma_{g'}^2 = (1 - p)\sigma_g^2,$$

where p is the portion of the total genetic variance accounted for by the detected QTLs in the model. Taking into account the above expressions, the following is derived:

$$g_s > \sqrt{c_{ii}F_{1,k,\alpha}[\sigma_E^2 + (1 - p)\,\sigma_g^2 + pm\sigma_g^2]} \qquad (5)$$

$$g_s/\sigma_g > \sqrt{c_{ii}F_{1,k,\alpha}[1/h_b^2 - p(1 - m)]} \qquad (6)$$

where g_s is the smallest gene effect that can be detected in a group of estimated QTLs that account for a portion p of the total genetic variance σ_g^2; values of c_{ii} are computed from Table 2 and depend on the number of tested progeny, the generation involved, the recombination frequency between markers and whether the detected gene effect is for

TABLE 2

Expected values of the diagonal terms c_{ii} corresponding to additive, dominance and heterotic effects of *QTL* i in the inverse matrix used for estimating gene effects of independent *QTL*s with different generations by multiple linear regression

Generations	c_{ii} for additive effects[a]	c_{ii} for dominance effects[a]	c_{ii} for differential heterotic effects[a]
$F_{2:3}$ lines	$\frac{1}{N(1 - r_i)}$	$\frac{32}{N(2 - 4r_i + r_i^2)}$	
Recombinant inbreds	$\frac{(1 + 2r_i)(1 + r_i^2)^2}{N(1 - r_i^2)^2}$	[b]	
Doubled-haploid lines	$\frac{1}{N(1 - r_i)}$	[b]	
Selfed backcrosses	$\frac{4}{N(1 - r_i)}$	$\frac{16}{N(1 - r_i)}$	
NC design III	$\frac{8}{N(1 - r_i)}$	$\frac{8}{N(1 - r_i)}$	
Testcrosses			
F_2 plants			$\frac{2}{N(1 - r_i)}$
Recombinant inbreds			$\frac{(1 + 2r_i)(1 + r_i^2)^2}{N(1 - r_i^2)^2}$
Doubled-haploid lines			$\frac{1}{N(1 - r_i)}$
Selfed three-way crosses			$\frac{4}{N(1 - r_i)}$

[a] N refers to the number of testing progenies and r_i is the recombination frequency between flanking markers.
[b] No dominance effects are estimated with inbred generations.

additivity, dominance or the differential heterosis of the two parental lines crossed to a tester; m is the coefficient of $\sigma^2_{g'}$ or $\sigma^2_{h'}$ in Table 3 and can be computed for different generations and recombination frequencies; $F_{1,k,\alpha}$ is the tabular F-value in the F distribution for 1 and k (residual) degrees of freedom at the α significant level; h_b^2 is the broad-sense heritability.

To make fair comparisons among generations, the smallest gene effects of each generation should be referred to a common denominator (*e.g.*, the genetic standard deviation of the $F_{2:3}$ lines when studying the segregating generations and that of the F2T when studying their testcrosses). Then, Equation 6 will be modified as follows:

$$\mathbf{g}_s > \sqrt{c_{ii}F_{1,k,\alpha}[1/h_{F2:3}{}^2 - 1 + v - vp(1 - m)]}, \qquad (7)$$

where $\mathbf{g}_s$ (either additive $\mathbf{a}_s$, dominance $\mathbf{d}_s$ or heterotic $\mathbf{h}_s$ effect) is the smallest gene effect in $F_{2:3}$ or F2T genetic standard deviation units; $h^2_{F2:3}$ is the heritability of the $F_{2:3}$ lines; and v is the ratio of the genetic variance of the generation under study to that of the $F_{2:3}$ or F2T.

RESULTS AND DISCUSSION

Equation 6 predicts the smallest gene effect that can be detected with a 50% chance at the α significance level when a group of QTLs, which accounts for a portion p of the genetic variance, has already been estimated by the model using stepwise regression. The equation is valid for any distribution of gene effects. The stepwise regression incorporates QTLs into the model starting with the largest gene effect. Once a QTL is in the model, it accounts for a portion of the genetic variance and thus reduces the least significant value for incorporating the next QTL into the model. The process continues until the decreasing least significant value is larger than the QTL effect being tested (presumably, the smallest effects from the distribution). For the particular case of equal gene effects, once the first QTL enters in the model, the remaining QTLs will be easier detected. However, this unlikely situation of equal effects is the most unfavorable for incorporating the first QTL into the model (MORENO-GONZALEZ 1992b), because no gene effect is larger than the others. Equations 5–7 relate several parameters such as gene effects, genetic variance, heritability, marker density and number of progeny (included in the c_{ii} expressions of Table 2). Therefore, they can be managed to study different genetic situations under proper assumptions. Some situations will be studied below.

The expected value of the smallest additive effect in $F_{2:3}$ genetic standard deviation units ($\mathbf{a}_s$) that can be detected with a 50% chance at the 0.005 level as a function of the heritability of the trait was plotted for different segregating generations and values of the recombination frequencies (r) between markers in Figure 1, a to d, according to Equation 7. The prog-

TABLE 3

Expected values of the squared deviations (Φ^2) of the assigned genotypic values to marker classes from their real genotypic value for different generations

Generation	Φ^{2a}
Segregating	
$F_{2:3}$ lines	$r\left[1 + \left(1 - \frac{r}{2} - \frac{r^3}{4B}\right)\frac{\sigma_d^{2\prime\prime}}{\sigma_g^{2\prime\prime}}\right]\sigma_g^{2\prime\prime}$
Recombinant inbreds	$\frac{2r(1 + 2r + 2r^2 + r^4)}{(1 + 2r)(1 + r^2)^2}\sigma_g^{2\prime\prime}$
Doubled-haploid lines	$r\sigma_g^{2\prime\prime}$
Selfed backcrosses	$r\sigma_g^{2\prime\prime}$
NC design III	$r\sigma_g^{2\prime\prime}$
Testcrosses:	
F_2 plants	$r\sigma_h^{2\prime\prime}$
Recombinant inbreds	$\frac{2r(1 + 2r + 2r^2 + r^4)}{(1 + 2r)(1 + r^2)^2}\sigma_h^{2\prime\prime}$
Doubled-haploid lines	$r\sigma_h^{2\prime\prime}$
Selfed three-way crosses	$r\sigma_h^{2\prime\prime}$

[a] The whole expression that multiplies $\sigma_g^{2\prime\prime}$ or $\sigma_h^{2\prime\prime}$ corresponds to m in equations 5–7; $\sigma_g^{2\prime\prime}$, $\sigma_d^{2\prime\prime}$ and $\sigma_h^{2\prime\prime}$ refer to the genetic, dominance and heterotic-effect variances of the corresponding generation accounted for by *QTL* in the model, respectively.

$B = \frac{1}{2} - r + r^2$; r is the recombination frequency between two consecutive flanking markers.

eny size ($N = 500$) and the fraction ($P = 0.95$) of the genetic variance accounted for by the QTLs in the model were kept constant for all cases. Complete dominance was assumed for computing m (Table 3) and v; $v = 1$, 16/9, 16/9, 5/9 and 4/3 for $F_{2:3}$, RI, DH, SB and NCD III, respectively. The most efficient generation for detecting the smallest additive effects in all situations is the doubled-haploid lines. By the contrary, the most inefficient generation among the four plotted is the SB, except for high r and h^2 where RI is worst (Figure 1d). The NCD III (plot not shown) is even more inefficient than the SB, as it is easily seen by comparing their c_{ii} values from Table 2. Unfortunately, random doubled-haploid lines cannot be developed in most of the crops, thus their use is very limited. The $F_{2:3}$ lines are inferior to RI for low r (Figure 1, a and b), but are better than RI for high heritability and r (Figure 1, c and d). Comparison of RI and DH shows that RI is almost as efficient as DH for small recombination frequencies and any heritability value, *e.g.*, $r = 0.05$ (Figure 1a) and $r = 0.1$ (Figure 1b). However, RI lose their relative efficiency as the recombination frequency between markers increases. Therefore, random recombinant inbreds can be an appropriate option for detecting small additive effects when the marker density of the genome is high or the heritability of the trait is low. $F_{2:3}$ lines are preferable for a situation with high heritability and low or moderate marker density. The progeny size (N) affects the $\mathbf{a}_s$ value by the reciprocal of its square root [c_{ii} (Table 2) in Equation 7], but it does not alter the relative efficiency of generations. If no dominance was assumed, the efficiency of the $F_{2:3}$ lines relative to RI and DH will increase and will follow the same pattern than F2T relative to RIT and DHT in Figure 2, a to d.

Dominance effects can be detected with $F_{2:3}$ lines and SB. However, significant estimates of the smallest dominance effects ($\mathbf{d}_s$) are much larger than those of the smallest additive effects ($\mathbf{a}_s$) as is easily seen by looking at the values of c_{ii} from Table 2. By comparing the c_{ii} values for additive and dominance effects, the efficiency of significant estimates of $\mathbf{a}_s$ relative to $\mathbf{d}_s$ is approximately $2\sqrt{2}$ for the $F_{2:3}$ lines and 2 for SB. The NCD III is more efficient than $F_{2:3}$ and SB for detecting dominance effects, because of its smaller c_{ii} value (Table 2). The smallest significant dominance effects for SB are approximately $\sqrt{2}$ times larger than for NCD III.

The pedigree method is a common breeding strategy for developing inbred lines from a chosen cross ($P_1 \times P_2$) between two parental lines. In a hybrid program, new developed lines are further crossed to an inbred tester T to select superior hybrids. Evaluation of the testcrosses is an adequate strategy to develop inbred lines with a good tester-specific combining ability and to identify marker-associated QTLs with favorable heterotic effects. Application of Equation 2 to the testcrosses will identify the differential heterotic effects of the QTL alleles in P_1 and P_2. The favorable alleles can be further followed by their markers through the F_2, F_3 . . . or backcross generations for selecting inbred lines with superior specific combining ability. The choice of the initial $P_1 \times P_2$ cross could be based on the theory of transfer of alleles (DUDLEY 1984, 1987).

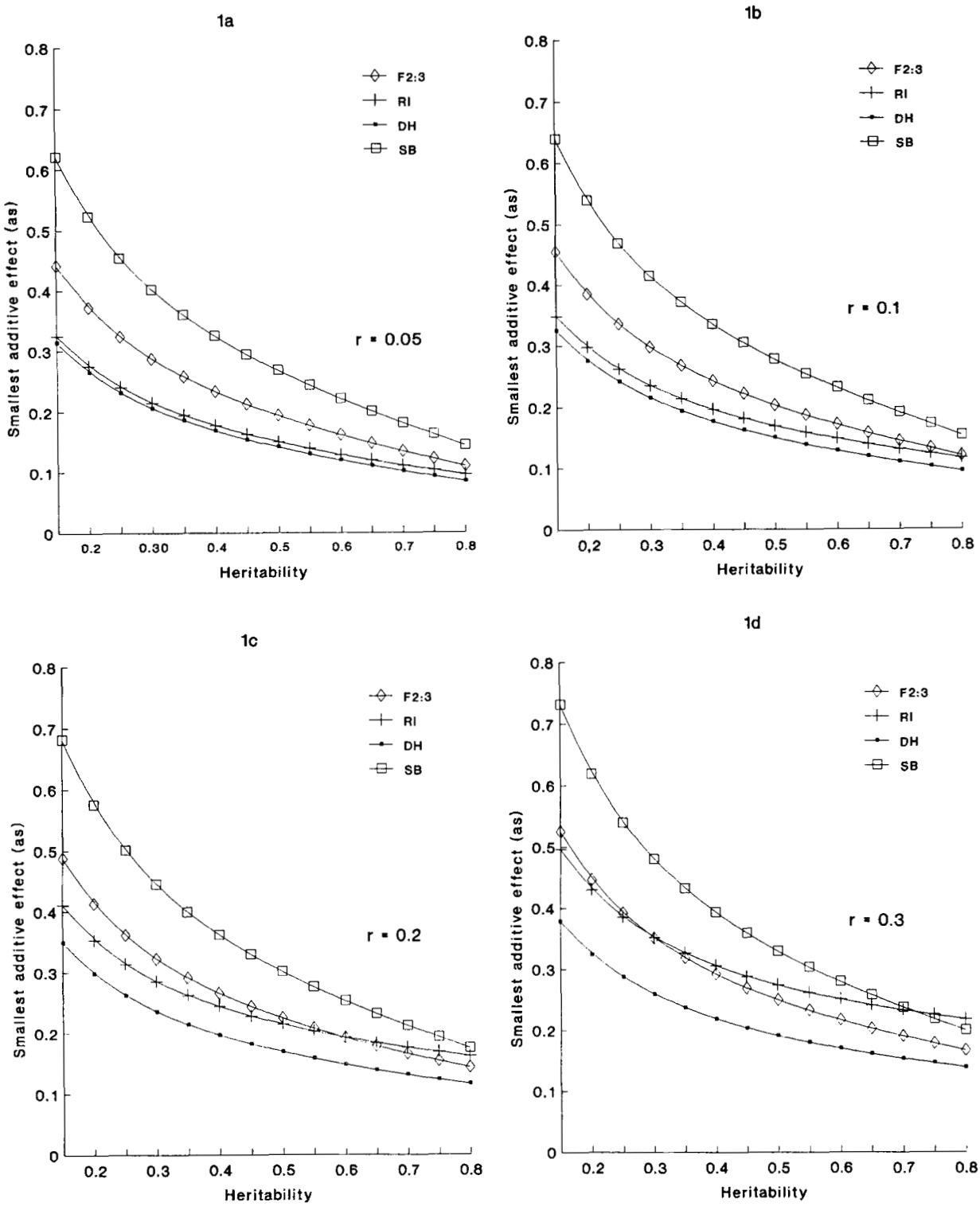

FIGURE 1.—Smallest additive effects (**as**) in $F_{2:3}$ genetic standard deviation units as function of the heritability for generations $F_{2:3}$ lines ($F_{2:3}$), double-haploid lines (DH), recombinant inbreds (RI) and selfed backcrosses (SB), when the recombination frequency (r) between markers was assumed to be $r = 0.05$ (1a), $r = 0.1$ (1b), $r = 0.2$ (1c) and $r = 0.3$ (1d).

The expected value of the smallest differential heterotic effect in F2T genetic standard deviation units ($\mathbf{h}_s$) that can be detected with a 50% chance at the 0.005 level as a function of the heritability of the trait was plotted for different generations and recombination frequencies in Figure 2, a to d, according to Equation 7. A constant progeny size ($N = 500$) and $P = 0.95$ were assumed; v is 1, 2, 2 and 0.5 for F2T, RIT, DHT and STW, respectively. Testcrosses of the random doubled-haploid lines are the most efficient. However, use of this generation is limited to only few crops. STW is the most inefficient generation, except

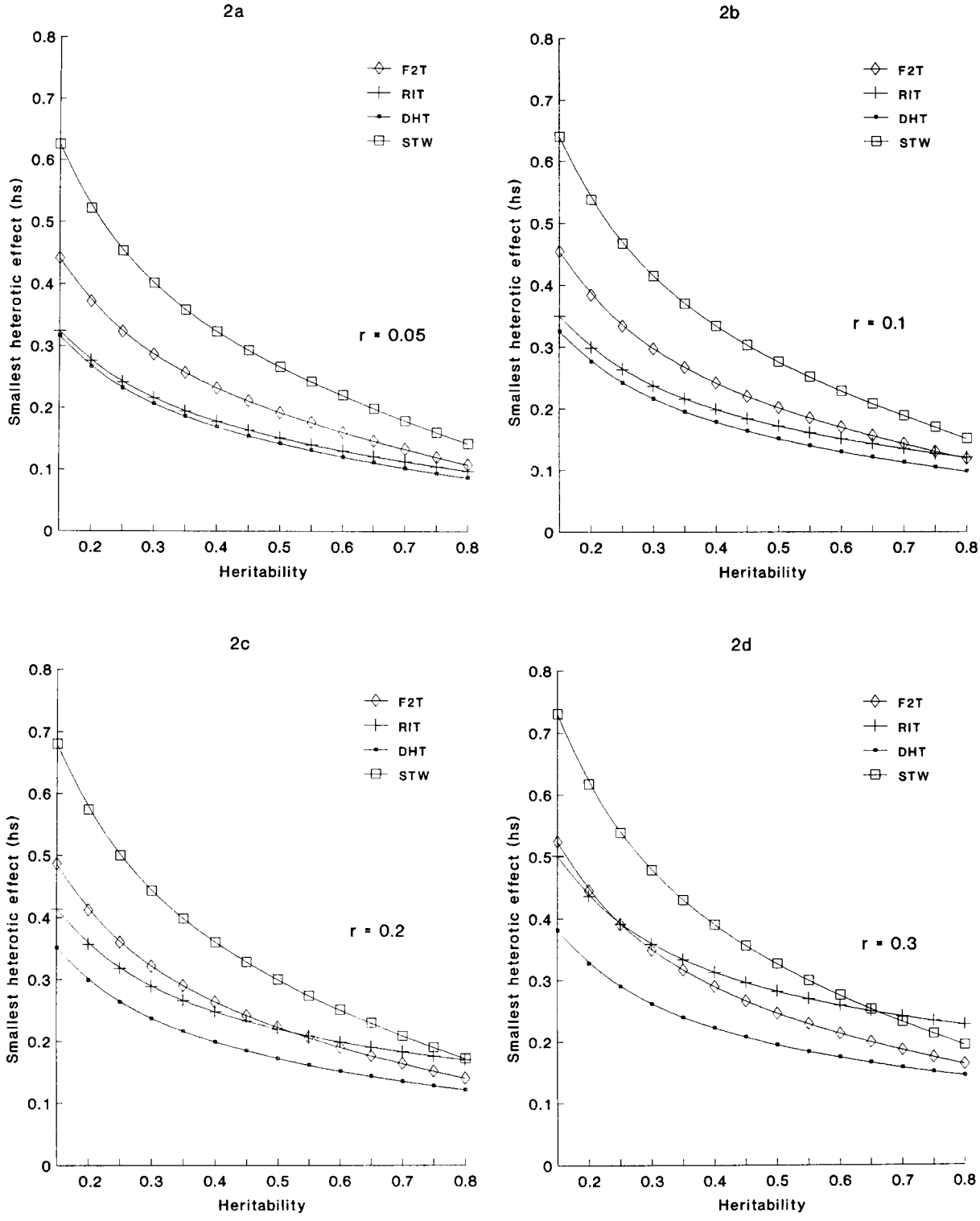

FIGURE 2.—Smallest heterotic effects (**hs**) in F2T genetic standard deviation units as function of the heritability for testcrosses with tester T from F_2 plants (F2T), double-haploid lines (DHT) and recombinant inbreds (RIT) and selfed families (STW) from the three-way cross (P_1 × P_2) × T, when the recombination frequency (r) between markers was assumed to be $r = 0.05$ (2a), $r = 0.1$ (2b), $r = 0.2$ (2c) and $r = 0.3$ (2d).

for $r = 0.3$ and high heritability where RIT is worst (Figure 2d). The F2T generation is more efficient than the RIT generation for low or moderate marker density ($r \geq 0.20$) and moderate or high heritability (Figure 2, c and d), but RIT is better than F2T for high marker density of the genome ($r \leq 0.10$; Figure 2, a and b). Therefore, F2T is an appropriate option for detecting small heterotic effects in many crops,

especially when the marker density of the genome is not high and the heritability is not low. The RIT will be preferable when a high marker density is feasible.

The differential heterotic effect ($\mathbf{h}_r$) was set up as one-half the difference between the genotypic values of ${}^1Q_i{}^TQ_i$ and ${}^2Q_i{}^TQ_i$ in Equation 2. If dominance of favorable alleles is assumed as the dominance hypothesis of heterosis requires, then differential heterotic effects exist provided that (a) ${}^1Q_i$ and ${}^2Q_i$ have different gene effects and (b) ${}^TQ_i$ is recessive for at least one of the two parental alleles. If a superior ${}^*P_1 \times$ T hybrid is sought by improving line P_1 through backcrossing, favorable heterotic effects can be identified in P_2 and later transferred to P_1.

Dominance effects cannot be efficiently estimated in the segregating generations of $P_1 \times P_2$, but the differential heterotic effects (a kind of dominance) of two QTL alleles when they are both in the heterozygous state with a common tester allele can be efficiently estimated by evaluation of appropriate testcrosses. Identification of these heterotic effects are especially important for hybrid breeding programs.

General assumptions of the model: The c_{ii} values to estimate the smallest gene effects (Equations 5–7, Table 2) were obtained for independent QTLs. If linked QTLs were assumed, the c_{ii} values would be larger than those expressed in Table 2 when the QTLs are incorporated into the model, because of the relationship among gene effects. Thus, the model is less efficient for linked than for independent QTLs. In addition, $\rho_i = 0.5$ was assumed for all QTLs in the model. However, true values of ρ_i different from 0.5 will produce only a slight bias in the estimates (MORENO-GONZALEZ 1992b).

This work was supported in part by the CICYT (Comisión Interministerial de Ciencia y Tecnología, Spain) grant AGR 90-0265

LITERATURE CITED

COWEN, N. M., 1988 The use of replicated progenies in marker-based mapping of QTLs. Theor. Appl. Genet. **75:** 857–862.

COWEN, N. M., 1989 Multiple linear regression analysis of RFLP data sets used in mapping QTLs, pp. 113–116 in *Development and Application of Molecular Markers to Problems in Plant Genetics*, edited by T. HELENTJARIS and B. BURR. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

DRAPER, N. R., and H. SMITH, 1981 *Applied Regression Analysis*, Ed. 2. John Wiley, New York.

DUDLEY, J. W., 1984 A method of identifying lines for use in improving parents of a single cross. Crop Sci. **24:** 355–357.

DUDLEY, J. W., 1987 Modification of methods for identifying inbred lines useful for improving parents of elite single crosses. Crop Sci. **27:** 944–947.

JAYAKAR, S. D., 1970 On the detection and estimation of linkage between a locus influencing a quantitative character and a marker locus. Biometrics **26:** 451–464.

KNAPP, S. J., W. C. BRIDGES, JR. and D. BIRKES, 1990 Mapping quantitative trait loci using molecular marker linkage maps. Theor. Appl. Genet. **79:** 583–592.

LANDE R., and R. THOMPSON, 1990 Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics **124:** 743–756.

LANDER, E. S., and D. BOTSTEIN, 1989 Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121:** 185–199.

MCMILLAN, I., and A. ROBERTSON, 1974 The power of methods for detection of major genes affecting quantitative characters. Heredity **32:** 349–356.

MORENO-GONZALEZ, J., 1992a Genetic models to estimate additive and non- additive effects of marker-associated QTL using multiple regression techniques. Theor. Appl. Genet. **85:** 435–444.

MORENO-GONZALEZ, J., 1992b Estimates of marker-associated QTL in Monte Carlo backcross generations using multiple regression. Theor. Appl. Genet. **85:** 423–434.

SOLLER, M., and J. S. BECKMAN, 1983 Genetic polymorphism in varietal identification and genetic improvement. Theor. Appl. Genet. **67:** 25–33.

THODAY, J. M., 1961 Location of polygenes. Nature **191:** 368–370.

Communicating editor: T. F. C. MCKAY

APPENDIX 1

The **X** design (data) matrix (Table 4) can be written for the $F_{2:3}$ generation in the model of Equation 1, where $B = \frac{1}{2} - r + r^2$; $E = (2 - 4r + 3r^2)/2(2 - 4r + 4r^2)$; N is the number of scored individuals in the $F_{2:3}$ generation; μ_0 has already been defined; x'_i, x'_{i+1}, . . . and y_i, y'_{i+1}, . . . are dummy variables corresponding to the additive and dominance effects of QTL i, $i + 1$ . . ., respectively; values of x' and y' are taken from Table 1; r is the recombination frequency between flanking markers. If the QTL are unlinked (independent), the expected $\mathbf{X'X}$ matrix ($E[\mathbf{X'X}]$) is obtained, as shown in Scheme 1, where column 1 corresponds to μ_0; columns 2 to $f + 1$ correspond to variables x'_i associated with additive effects and columns $f + 2$ to $f + h + 1$ correspond to variables y'_i associated with dominance effects; f and h are the

TABLE 4

***X* design matrix**

Individual	Marker class	Expected frequency of marker class	Values of variables in the data matrix X		
			$\mu_0 \ldots x'_i$	$x'_{i+1} \ldots y'_i$	$y'_{i+1} \ldots$
1	1	$\frac{1}{4}(1-r)^2$	1 . . . 1	1 0	0 . . . .
2	2	$\frac{1}{2}r(1-r)$	1 . . . ½	½ . . . . ¼	¼ . . . .
3	3	$\frac{1}{4}r^2$	1 . . . 0	0 . . . . ¼	¼ . . . .
4	4	$\frac{1}{2}r(1-r)$	1 . . . ½	½ . . . . ¼	¼ . . . .
5	5	B	1 . . . 0	0 . . . . E	E . . . .
6	6	$\frac{1}{2}r(1-r)$	1 . . . −½	−½ . . . . ¼	¼ . . . .
7	7	$\frac{1}{4}r^2$	1 . . . 0	0 . . . . ¼	¼ . . . .
8	8	$\frac{1}{2}r(1-r)$	1 . . . −½	−½ . . . . ¼	¼ . . . .
9	9	$\frac{1}{4}(1-r)^2$	1 . . . −1	−1 0	0 . . . .
.	.	.	. . . . .	. . . . .	. . . . .
.	.	.	. . . . .	. . . . .	. . . . .
N	.	.	1	. . . . .	. . . . .

TABLE 5

Deviations of the real genotypic values of the *QTL* genotypes from the genotypic values assigned to the marker classes for the $F_{2:3}$ generation

Marker class	Expected frequency	Real situation of *QTL* genotypes			Deviations of genotypic values of *QTL* genotypes from the assigned genotypic values to marker classes (when ρ = 0.5)
		F_2 genotype	Expected frequency	Genotypic value ($F_{2:3}$)	
2	$\frac{1}{2}r(1-r)$	QQ	$1-\rho$	a	$\frac{1}{2}a-\frac{1}{4}d$
		Qq	ρ	$\frac{1}{2}d$	$-\frac{1}{2}a+\frac{1}{4}d$
3	$\frac{1}{4}r^2$	QQ	$(1-\rho)^2$	a	$a-\frac{1}{4}d$
		Qq	$2\rho(1-\rho)$	$\frac{1}{2}d$	$\frac{1}{4}d$
		qq	ρ^2	$-a$	$-a-\frac{1}{4}d$
4	$\frac{1}{2}r(1-r)$	QQ	ρ	a	$\frac{1}{2}a-\frac{1}{4}d$
		Qq	$1-\rho$	$\frac{1}{2}d$	$-\frac{1}{2}a+\frac{1}{4}d$
5	$B=\frac{1}{2}-r+r^2$	QQ	$\rho(1-\rho)r^2/2B$	a	$a-\frac{1}{2}d(1-r^2/4B)$
		Qq	$1-\rho(1-\rho)r^2/B$	$\frac{1}{2}d$	$dr^2/8B$
		qq	$\rho(1-\rho)r^2/2B$	$-a$	$-a-\frac{1}{2}d(1-r^2/4B)$
6	$\frac{1}{2}r(1-r)$	qq	ρ	$-a$	$-\frac{1}{2}a-\frac{1}{4}d$
		Qq	$1-\rho$	$\frac{1}{2}d$	$\frac{1}{2}a+\frac{1}{4}d$
7	$\frac{1}{4}r^2$	QQ	ρ^2	a	$a-\frac{1}{4}d$
		Qq	$2\rho(1-\rho)$	$\frac{1}{2}d$	$\frac{1}{4}d$
		qq	$(1-\rho)^2$	$-a$	$-a-\frac{1}{4}d$
8	$\frac{1}{2}r(1-r)$	qq	$1-\rho$	$-a$	$-\frac{1}{2}a-\frac{1}{4}d$
		Qq	ρ	$\frac{1}{2}d$	$\frac{1}{2}a+\frac{1}{4}d$

$$E[X'X]=\frac{N}{2}\begin{bmatrix} 2 & 0 & 0 & \cdots & \frac{1}{2} & \frac{1}{2} & \cdots & \frac{1}{2} \\ 0 & 1-r & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1-r & \cdots & 0 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & 0 \\ \frac{1}{2} & 0 & 0 & \cdots & \frac{1-r}{4}+\frac{r^2}{16} & \frac{1}{8} & \cdots & \frac{1}{8} \\ \frac{1}{2} & 0 & 0 & \cdots & \frac{1}{8} & \frac{1-r}{4}+\frac{r^2}{16} & \cdots & \frac{1}{8} \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \\ \frac{1}{2} & 0 & 0 & \cdots & \frac{1}{8} & \frac{1}{8} & \cdots & \frac{1-r}{4}+\frac{r^2}{16} \end{bmatrix}$$

SCHEME 1

$$E([X'X]^{-1})=\frac{2}{N(2-4r+r^2)}\begin{bmatrix} \frac{2-4r+r^2+h}{2} & 0 & 0 & \cdots & -4 & -4 & \cdots & -4 \\ 0 & \frac{2-4r+r^2}{1-r} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \frac{2-4r+r^2}{1-r} & \cdots & 0 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ -4 & 0 & 0 & \cdots & 16 & 0 & \cdots & 0 \\ -4 & 0 & 0 & \cdots & 0 & 16 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ -4 & 0 & 0 & \cdots & 0 & 0 & \cdots & 16 \end{bmatrix}$$

SCHEME 2

number of variables with additive and dominance effects in the model, respectively. The expected inverse matrix was computed and is shown in Scheme 2.

Therefore, the expected values of the diagonal terms c_{ii} for the additive and dominance effects from the independent QTL i are $2/N(1-r)$ and $32/N(2-4r+r^2)$, respectively.

APPENDIX 2

The deviations of the real genotypic values of QTL genotypes from the genotypic values assigned to the

marker classes 2–8 in the model for the $F_{2:3}$ generation are shown in Table 5. The deviations for the marker classes 1 and 9 are zero. The sum of squares of these deviations multiplied by their corresponding frequencies yields the expected value of Φ^2. Avoiding cumbersome algebra and assuming $\rho = \frac{1}{2}$, the following result is obtained from Table 5:

$$\Phi^2 = \tfrac{1}{2}r(a^2 + \tfrac{1}{8}d^2) + (1/16)rd^2(1 - \tfrac{1}{2}r - r^3/4B),$$

where a and d are additive and dominance effects, respectively; r is the recombination frequency between markers; and $B = \frac{1}{2} - r + r^2$. Taking into account that $\frac{1}{2}(a^2 + \frac{1}{8}d^2)$ and $d^2/16$ are the genetic ($\sigma^2_g{}''$) and dominance ($\sigma^2_d{}''$) variances of the $F_{2:3}$ lines for the QTL in the model, respectively, then

$$\begin{aligned}\Phi^2 &= r[\sigma^2_g{}'' + \sigma^2_d{}''(1 - \tfrac{1}{2}r - r^3/4B)] \\ &= r[1 + (1 - \tfrac{1}{2}r - r^3/4B)\sigma^2_d{}''/\sigma^2_g{}'']\sigma^2_g{}''.\end{aligned}$$

The ratio σ^2_d/σ^2_g is 1/9 for complete dominance and lower for incomplete dominance.