

SAMPLE SIZE REQUIRED TO DETECT LINKAGE BETWEEN A MARKER AND QTL

Claudio Lopes de Souza Jr.

ABSTRACT

The sample sizes required to detect linkage between a marker and QTL for an unreplicated F₂ population, and for replicated F_{3:2}, F_{4:3} and recombinant inbred lines were investigated. The F test with P ≤ 0.01 in the analysis of variance was used to detect linkage. The expectations of the mean squares were manipulated to express the F ratio in terms of the distance between the marker and the QTL (r), the heritability (h²) and the level of the dominance of the trait, the sample size, and the ratio $\phi = \sigma_{Aq}^2 / \sigma_A^2$. The sample sizes were estimated from the F test. As r decreases, and h² and ϕ increase, the sample size decreases; and as r increases and h² and ϕ decrease, the sample size increases.

INTRODUCTION

Isozymes and restriction fragment length polymorphisms (RFLPs) have been used as genetic markers in an attempt to elucidate the control of quantitative traits in several crop species (Hamrick and Allard, 1975; Ledig *et al.*, 1983; Helentjaris *et al.*, 1986; Edwards *et al.*, 1987; Weller *et al.*, 1988; Keim *et al.*, 1990).

Statistical analyses and types of genotypes for detecting linkage between a marker and QTL (QTL mapping) have been reported (MacMillan and Robertson, 1974; Soller *et al.*, 1976; Asins and Carbonell, 1988; Cowen, 1988; Knapp and Bridges, 1990; Soller and Beckmann, 1990).

Sample sizes for QTL mapping for several types of genotypes have also been reported (Soller *et al.*, 1976; Asins and Carbonell, 1988; Soller and Beckmann, 1990). These authors showed the importance of the heritabilities of the traits and of the recombination fraction between the marker and the QTL(s) in determining the sample

size required to detect significant differences between the means of the homozygote markers genotypes, and/or to increase the power of the test. However, the sample sizes used for QTL mapping have varied. Helentjaris *et al.* (1986), Edwards *et al.* (1987), Keim *et al.* (1990), Abler *et al.* (1991), and Paterson *et al.* (1991) used 46, 1776, 60, 504, and 350 F₂ plants, respectively; whereas Burr *et al.* (1988) used 48 and 41 recombinant inbred lines from two maize populations.

Analysis of variance also can be used to detect linkage between a marker and QTL(s) (Edwards *et al.*, 1987; Keim *et al.*, 1990; Knapp and Bridges, 1990). We determined the sample size required for QTL mapping by using the F test in the analysis of variance to detect linkage between a marker and a QTL for unreplicated and replicated genotypes.

METHODS

The breeding population

An F₂ population from the cross of two highly inbred lines (L₁ and L₂) was the reference population. The two inbred lines were polymorphic at the marker locus (M) and at the linked QTL (B). The inbred lines L₁ and L₂ had the M_iB_k/M_iB_k and M_jB_l/M_jB_l genotypes, respectively, with the marker alleles (M_i and M_j) being codominant and inherited in a mendelian fashion. The recombination frequency between the marker and the QTL was designated by r. The gametes of the F₁ plants were M_iB_k, M_iB_l, M_jB_k, and M_jB_l, with frequencies (1-r)/2, r/2, r/2, and (1-r)/2, respectively. The random combination of these gametes gave rise to the F₂ population. The F_{3:2} and F_{4:3} lines were produced by self-pollinating the F₂ and F₃ plants, respectively; and RI lines are recombinant inbred lines produced by at least eight generations of self-pollinating the source population. The frequencies of recombinant and non-recombinant inbred lines were R and (1-R), respectively, with $R=2r/(1+2r)$ (Haldane and Waddington, 1931). Evaluation of the F₂ population was considered to be on a single plant basis (unreplicated genotypes), whereas the F_{3:2}, F_{4:3}, and RI lines were considered to be evaluated in a randomized complete block design (replicated genotypes). The plants and the lines were assumed to be random samples of the source populations.

The genetic model

The additive and dominance genetic variances of an F₂ population are $\sigma_A^2 = (1/2)a^2$ and $\sigma_D^2 = (1/4)d^2$, where a and d are half of the difference of the homozygotes

genotypic values and the heterozygote genotypic value, respectively (Mather and Jinks, 1982; Falconer, 1989). Let $k = d/a$ be the level of dominance, and the dominance variance becomes $\sigma_D^2 = (k^2/2)\sigma_A^2$. The levels of dominance used were complete dominance ($k=1.0$) and partial dominance ($k=0.5$). The heritabilities of the F_2 population and of the $F_{3:2}$, $F_{4:3}$, and RI lines are, respectively: $h_{F_2}^2 = \sigma_A^2/(\sigma_A^2 + \sigma_D^2 + \sigma_E^2)$, $h_{F_{3:2}}^2 = \sigma_A^2/\sigma_{PhF_{3:2}}^2$, $h_{F_{4:3}}^2 = (1.5)\sigma_A^2/\sigma_{PhF_{4:3}}^2$, and $h_{RI}^2 = 2\sigma_A^2/\sigma_{PhRI}^2$; where $\sigma_{PhF_{3:2}}^2 = \sigma_A^2 + (1/4)\sigma_D^2 + [(1/2)(\sigma_A^2 + \sigma_D^2) + \sigma_{cw}^2]/pn + \sigma_c^2/n$, $\sigma_{PhF_{4:3}}^2 = (3/2)\sigma_A^2 + (3/16)\sigma_D^2 + [(1/4)(\sigma_A^2 + \sigma_D^2) + \sigma_{cw}^2]/pn + \sigma_c^2/n$, and $\sigma_{PhRI}^2 = 2\sigma_A^2 + \sigma_{cw}^2/pn + \sigma_c^2/p$. Here, σ_E^2 , σ_c^2 , σ_{cw}^2 , p , and n , are residual variance, the plot-to-plot residual variance, within plots residual variance, number of replications, and number of plants per plot, respectively.

The genetic variances of a marker of the types of genotypes considered in this paper can be expressed as (Souza Jr., 1992) $VM = (1-2r)^2c\sigma_A^2 + [1-8r(1-3r + 4r^2-2r^3)]t\sigma_D^2$, where c and t refer to coefficients that define the amount of σ_A^2 and σ_D^2 , respectively, of the genetic variances among plants in the inbred populations, and among progenies of inbred lines; and $c=t=1$ for F_2 populations, $c=1$ and $t=1/4$ for $F_{3:2}$ lines, $c=3/2$ and $t=3/16$ for $F_{4:3}$ lines, and $c=2$ and $t=0$ for RI lines (Mather and Jinks, 1982).

Heritability values of the types of populations used in this paper were simulated with given values for the genetic and residual variances. The values used for the simulation were $\sigma_A^2 = 100.0$, $\sigma_D^2 = 50.0$ ($k=1.0$), $\sigma_E^2 = 800.0$, and $\sigma_{cw}^2 = \sigma_c^2 = 400.0$, with $h_{F_2}^2 = 10.53\%$; $\sigma_A^2 = 150.0$, $\sigma_D^2 = 18.75$ ($k=0.5$), $\sigma_E^2 = 565.0$, and $\sigma_{cw}^2 = \sigma_c^2 = 282.50$, and $h_{F_2}^2 = 20.44\%$; $\sigma_A^2 = 150.0$, $\sigma_D^2 = 18.75$ ($k=0.5$), $\sigma_E^2 = 330.0$, and $\sigma_{cw}^2 = \sigma_c^2 = 165.0$, with $h_{F_2}^2 = 30.08\%$. These values were used to simulate three traits with different levels of heritabilities. Estimates of the heritabilities of the $F_{3:2}$, $F_{4:3}$, and RI lines were then obtained for $n = 10$ plants per plot, and $p = 1$ to 10 replications, for each of the three simulated traits.

The F test

The F test ($P \leq 0.01$) in the analysis of variance of each marker was used for the detection of linkage between a marker and a QTL. The F test was estimated by the ratio $F = MS \text{ Markers}/MS \text{ Residual}$ for the unreplicated F_2 genotypes (Table I), and for the replicated genotypes $F = MS \text{ Markers}/(MS \text{ Lines}/\text{Markers})$ (Table II), where MS refers to mean squares. The values of the F ratios were compared to the F values in the probability table for $P \leq 0.01$, to test the significance.

The sample size

The segregation of the marker genotypes ($M_iM_i:M_iM_j:M_jM_j$) of the F_2 population and of the $F_{3:2}$ lines is $1/4:1/2:1/4$; the segregation of the $F_{4:3}$ lines markers is $3/8:1/4:3/8$, and of the RI lines markers is $1/2:0:1/2$. Hence, the data will be unbalanced for the F_2 population, and for the $F_{3:2}$ and $F_{4:3}$ lines. The weighted mean number of replications per marker genotype is $z = [L - (\sum_{i=1} l_i^2)/L]/(q-1)$, where L is the number of plants or lines evaluated, l_i is the number of plants or lines of the i^{th} marker genotype, and q is the number of marker genotypes ($q=3$ for F_2 , $F_{3:2}$, and $F_{4:3}$ lines, and $q=2$ for RI) (Knapp and Bridges, 1990). The z value multiplies the variance among markers in the expectations of the markers mean squares in the analysis of variance (Tables I and II). The number of genotypes required to detect linkage between a marker and a QTL were estimated from the F test after manipulating the expectations of the mean squares in order to express the F ratio as a function of h^2 , r , ϕ , k , and z , where $\phi = \sigma_{Aq}^2/\sigma_A^2$ is the ratio of the additive variance of the QTL (σ_{Aq}^2) to the total additive variance. The proper h^2 , r , ϕ , k , and L values were supplied with a computer program. Then, the z values were estimated, and the first L value that provided a significant F test ($P \leq 0.01$) was considered as the sample size required to detect linkage between the marker and QTL.

RESULTS AND DISCUSSION

The analysis of variance of unreplicated genotypes in F_2 populations, with the expectations of the mean squares is given in Table I. The F ratio was estimated by $F = MS \text{ Markers} / MS \text{ Residual}$, where MS refers to mean squares. The expectations of the mean squares are $F = (\sigma^2 + zVM)/\sigma^2$ or $F = 1 + zVM/\sigma^2$. The expectation of the residual mean square is $\sigma^2 = \sigma_E^2 + \sigma_{GW}^2$, and $\sigma_{GW}^2 = \sigma_G^2 - VM$, where σ_G^2 , σ_{GW}^2 , VM , and σ_E^2 are the genetic variance ($\sigma_G^2 = \sigma_A^2 + \sigma_D^2$), the genetic variance within markers or the genetic variance unexplained by the marker, the genetic variance of the marker, and the residual variance, respectively. σ_E^2 , σ_{GW}^2 , and σ^2 , can be expressed as $\sigma_E^2 = [(1/h_f^2) - (2+k^2)/2]\sigma_A^2$, $\sigma_{GW}^2 = [(1-VM/\sigma_G^2)(2+k^2)/2]\sigma_A^2$, and $\sigma^2 = [(1/h_f^2) - (VM/\sigma_G^2)(2+k^2)/2]\sigma_A^2$.

The ratio VM/σ_G^2 is $2\psi_1\phi/(2+k^2)$, where $\phi = \sigma_{Aq}^2/\sigma_A^2$, and $\psi_1 = (1-2r)^2 + [1-8r(1-3r+4r^2-2r^3)](k^2/2)$. σ_{Aq}^2 is the additive variance of the QTL linked to the marker, and σ_A^2 is the total additive variance of the trait. Thus, the F test to detect linkage between

a marker and QTL for unreplicated F₂ genotypes can be expressed as $F=1 + z\psi_1\phi/[(1/h^2_{f_2})-\phi\psi_1]$.

The analysis of variance for replicated genotypes in a randomized complete block design for F_{3:2}, F_{4:3}, and RI lines, with the expectations of the mean squares is given in Table II. The F ratio for the markers is $F=MS\text{ Markers}/(MS\text{ Lines}/\text{Markers})$. The expectations of the mean squares are $F=1 + pzVM/(\sigma^2 + p\sigma^2_{L/M})$, where p and z are the number of replications per line, and the weighted mean number of lines per marker genotype, respectively. The expectation of the residual mean square is $\sigma^2 = \sigma^2_e + (\sigma^2_{GW} + \sigma^2_{ew})/n$, and $\sigma^2_{GW} = \sigma^2_G - \sigma^2_L$, where σ^2_L , σ^2_{GW} , $\sigma^2_{L/M}$, σ^2_e , σ^2_{ew} , and n, are the genetic variance of the lines, the genetic variance within lines, the genetic variance of lines within markers, the residual plot-to-plot variance, the within plot residual variance, and the number of plants per plot, respectively.

Table I - Analysis of variance of a completely randomized experiment for unreplicated genotypes.

Source of variation	Degrees of freedom *	Mean squares	E(Ms) **	F
Markers	2	MS1	$\sigma^2 + zVM$	MS1/MS2
Residual	N-3	MS2	σ^2	

*N is the number of plants evaluated.

** z is the weighted number of marker replications.

For the F_{3:2} lines, σ^2 and $\sigma^2_{L/M}$ can be expressed as: $\sigma^2 = p[(1/h^2_{f_3:2}) - (8+k^2)/8]\sigma^2_{\lambda}$, and $\sigma^2_{L/M} = [(1-VM/\sigma^2_L) (8+k^2)/8]\sigma^2_{\lambda}$. Thus, $\sigma^2 + p\sigma^2_{L/M} = p[(1/h^2_{f_3:2}) - (VM/\sigma^2_L) (8+k^2)/8]\sigma^2_{\lambda}$. For the F_{4:3} lines we have $\sigma^2 = p[(3/2) (1/h^2_{f_4:3}) - (48 + 3k^2)/32]\sigma^2_{\lambda}$, and $\sigma^2_{L/M} = [(1 - VM/\sigma^2_L) (48 + 3k^2)/32]\sigma^2_{\lambda}$, and the denominator of the F test is $\sigma^2 + p\sigma^2_{L/M} = p[(3/2) (1/h^2_{f_4:3}) - (VM/\sigma^2_L) (48 + 3k^2)/32]\sigma^2_{\lambda}$. For the RI lines $\sigma^2 = 2p[(1/h^2_{rI}) - 1]\sigma^2_{\lambda}$, and $\sigma^2_{L/M} = 2(1-VM/\sigma^2_L)\sigma^2_{\lambda}$, and $\sigma^2 + p\sigma^2_{L/M} = 2p[(1/h^2_{rI} - (VM/\sigma^2_L)]\sigma^2_{\lambda}$.

The ratio $VM/\sigma^2_L = \psi_2c\phi$, and $\psi_2 = (1-2r)^2c + [1-8r(1-3r+4r^2-2r^3)] (k^2/t)$, with $c=1, 3/2, 2$, and $t=1/8, 1/16$, and 0, for F_{3:2}, F_{4:3}, and RI lines, respectively. Notice that the recombination fraction (r) changes for F_{4:3} because these lines have another recombination generation, and for RI lines r must be substituted by $R=2r/(1+2r)$ (Haldane

and Waddington, 1931). Thus, the F test can be expressed as $F=1 + \{(z\psi_2\phi)/[(1/h^2) - \psi_2\phi]\}$. There are no dominance effects for the RI lines, and, therefore, the F test can be expressed as $F=1 + \{[z(1-2R)^2\phi]/[1/h_{RI}^2 - (1-2R)^2\phi]\}$.

The F test to detect linkage depends on the heritability and on the level of the dominance of the trait, on the sample size, the distance between the marker and the QTL, and on the amount of the total additive genetic variance that can be explained by the additive variance of the QTL. Obviously, the F test of the RI lines does not depend on the level of the dominance of the trait.

Table II - Analysis of variance of a randomized complete block experiment for replicated genotypes.

Source of variation	Degrees of freedom *	Mean squares	(E(Ms)**	F
Replications	(p-1)	MS1	-----	
Lines	(l-1)	MS2	$\sigma^2 + p\sigma_L^2$	
Markers	(m-1)	MS3	$\sigma^2 + p\sigma_M^2 + pzVM$	MS3/MS4
Lines/Markers	(l-3)	MS4	$\sigma^2 + p\sigma_M^2$	
Residual	(p-1)(l-1)	MS5	σ^2	

* p and l are the numbers of replications and lines, respectively.

** z is the weighted number of marker replications.

Soller *et al.* (1976) reported that a sample size of 1,050 plants would be required to detect linkage between a marker and QTL in an F₂ population, by using a test for the contrast between means of the homozygote markers, irrespective of the dominance level. They assumed the marker controls 1.0% of the total phenotypic variance, and complete linkage between the marker and QTL ($r=0$). With incomplete linkage ($r \neq 0$) they suggested the sample size will increase by the fraction $1/(1-2r)^2$. Asins and Carbonell (1988) reported that the sample size needed to detect linkage between a marker and QTL by a test of the contrast between means of homozygote markers, is a function of the level of dominance, the strength of linkage, and the heritability of the trait. They showed some figures for the minimum sample sizes required to detect such linkage, but they assumed negligible environmental variance. Soller and Beckmann (1990) reported that the difference between homozygote genotypes at the marker locus for a quantitative trait is

a function of the heritability of the trait (h^2), the recombination fraction (r), and the number of line replications (n). They presented the sample sizes required to detect linkage for F_3 , F_4 , RI, doubled-haploid lines, and vegetative clones, relative to the sample size required for the F_2 population. For example, the number of F_3 lines required to detect linkage relative to the F_2 population is $h^2 + (1-h^2/2)/n$. Knapp and Bridges (1990) showed that the power of the test of the difference between means of homozygote markers increases with the increase in the number of replications per marker genotype, with the increase of the number of replications per line, and increases as the genetic variance among lines, which is not explained by the reduction in the QTL parameter.

The heritability values of the lines increase as the number of plants per plot (n) and replications (p) increase. The higher the heritability values the more powerful the F test, i.e., the F test should detect smaller effects with higher heritabilities. Knapp and Bridges (1990) showed that an increase of n and p decrease the standard error of a marker genotype increasing the power of the test of the contrast between the means of the homozygote markers.

The number of F_2 plants required to detect linkage between a marker and a QTL with the F test ($P \leq 0.01$) in the analysis of variance is in Table III. For $h^2_{F_2} = 0.10$ the sample size to be used is high even when the marker and QTL are tightly linked. For example, 813 plants would be necessary with $r=0.01$, and 2,718 plants with $r=0.20$; however with $h^2_{F_2} = 0.20$ and $r=0.01$ and $r=0.20$ the sample size is reduced to 402 plants and 1,353 plants, respectively. As heritability increases the sample size decreases. Therefore the F_2 population should be used for QTL mapping only for traits with high F_2 heritability ($h^2_{F_2} \geq 0.50$), because for these traits around 400 plants would be required to detect linkage up to $r=0.15$. Obviously, if the heritability and the genetic parameters of the trait in the population being analyzed are already known, one could estimate the sample size for QTL mapping with the expression of the F test provided in this paper.

The heritability of a trait in the F_2 population can be increased with the use of lines produced from F_2 or derivative populations evaluated in proper experimental designs with replications and several plants per plot. The F_2 heritabilities of three simulated traits and the heritabilities of these traits in $F_{3:2}$, $F_{4:3}$, and RI lines with 10 plants per plot and one to 10 replications per line are given in Figure 1. For example, the $h^2_{F_2} \approx 0.10$ will be increased roughly to 0.40, 0.50, and 0.60, in the $F_{3:2}$, $F_{4:3}$, and RI lines, respectively, evaluated with four replications and 10 plants per plot. Hence, the sample size required for a trait in the F_2 population should be compared with the $F_{3:2}$, $F_{4:3}$, and RI lines by using the proper heritability values, with a defined number of plants per plot and replications per line.

Table III - Number of plants required to detect linkage between a marker and QTL for F_2 populations with the F test ($P \leq 0.01$) in the analysis of variance.

r	$h_{F_2}^*$					
	0.10	0.20	0.30	0.50	0.70	0.90
0.01	813	402	267	209	149	115
0.05	1,008	503	337	255	181	139
0.10	1,362	682	452	332	234	182
0.15	1,889	949	630	439	315	243
0.20	2,718	1,353	908	612	434	340
0.25	4,112	2,049	1,363	898	638	494
0.30	6,696	3,344	2,225	1,409	1,005	788
0.35	12,320	6,148	4,097	2,535	1,810	1,406
0.40	28,395	14,189	9,455	5,755	4,108	3,195
0.45	115,260	57,630	38,450	23,140	16,530	12,870

* $h_{F_2}^2 \leq 0.30$ with $k=1.0$, and $h_{F_2}^2 \geq 0.50$ with $k=0.50$, k is the level of dominance, and $\phi = \sigma_{\lambda q}^2 / \sigma_{\lambda}^2 = 0.10$.

The sample size required for QTL mapping for the three simulated traits with different levels of heritability in the F_2 population is in Table IV for four replications and 10 plants per plot. With $r=0.10$ and $h_{F_2}^2 = 0.10$ the sample size required is 1,362, 413, 407, and 417, for F_2 , $F_{3:2}$, $F_{4:3}$, and RI lines, respectively. With $h_{F_2}^2 \approx 0.20$ we have $h_{F_{3:2}}^2 \approx 0.60$, $h_{F_{4:3}}^2 \approx 0.70$, and $h_{RI}^2 \approx 0.80$, with required sample sizes of 682, 275, 292, and 313; and for $h_{F_2}^2 \approx 0.30$ we have $h_{F_{3:2}}^2 \approx 0.70$, $h_{F_{4:3}}^2 \approx 0.80$, and $h_{RI}^2 \approx 0.90$, with required sample sizes of 452, 234, 254, and 277, respectively for F_2 , $F_{3:2}$, $F_{4:3}$, and RI genotypes, and $r=0.10$.

Thus, the sample size required for QTL mapping is reduced with the use of replicated lines, especially for the low heritability traits. However, there were no significant differences between the sample sizes of the $F_{4:3}$ and RI lines for $r < 0.20$. Notice that the $F_{3:2}$, $F_{4:3}$, and RI lines will be replicated, but solely the F_2 , F_3 , and the inbred plants that gave rise to the $F_{3:2}$, $F_{4:3}$, and RI lines, respectively, will be used for marker mapping. Obviously, a mixture of plants within lines can also be used.

Therefore, for QTL mapping the sample size depends mainly on the distance between the marker and the QTL (r), the heritability of the trait (h^2), and the amount of

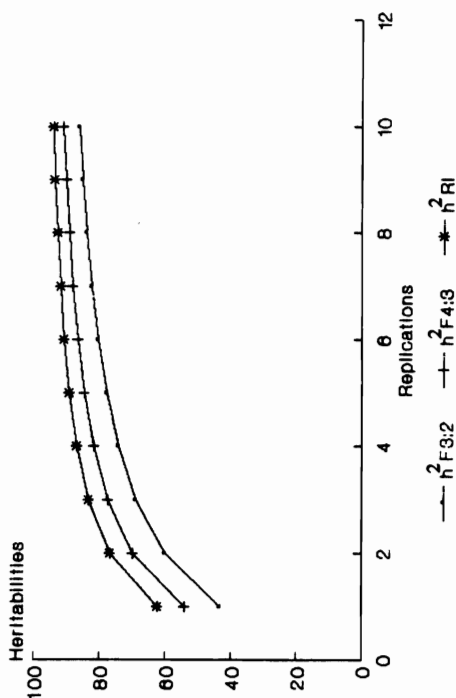
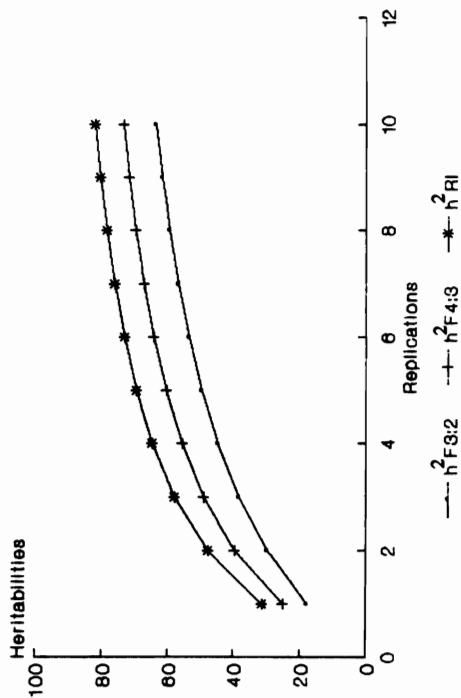
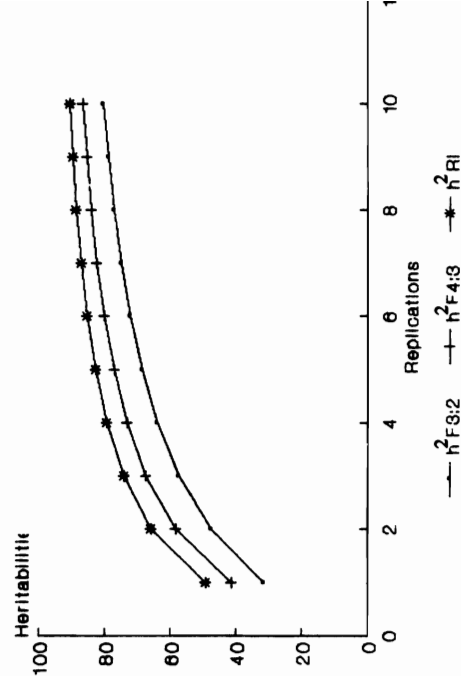


Figure 1 - Heritability values of $F_{3:2}$, $F_{4:3}$, and RI lines, for several replications, 10 plants per plot, and 10.53% (upper figure), 20.44% (intermediate figure), and 30.08% (lower figure) as the heritability values of an F_2 population.

total additive genetic variance that a QTL controls (ϕ). As r decreases and h^2 and ϕ increase the sample size decreases. Conversely, as r increases and h^2 and ϕ decrease the sample size increases. The distance between the QTLs and the markers can be reduced by increasing the number of markers; the heritabilities of the traits can be increased by using proper experimental designs with replicated genotypes, and experiments evaluated across environments could increase even further the heritabilities of the traits by reducing the genotype by environment interaction. The only parameter that can not be controlled is the ratio $\phi = \sigma_{Aq}^2 / \sigma_A^2$.

Experimental designs, number of replications, and the number of plants per plot are usually designed for the low heritability traits. For the intermediate heritability traits, the same experiment should be used, with a reduction in the number of replications and plants per plot.

Table IV - Number of lines required to detect linkage between a marker and QTL for $F_{3:2}$, $F_{4:3}$, and RI replicated lines with the F test ($P \leq 0.01$) in the analyses of variance, for three heritability levels of an F_2 population. *

r	$h_{F_2}^2 \approx 0.10$			$h_{F_2}^2 \approx 0.20$			$h_{F_2}^2 \approx 0.30$		
	$h_{F_{3:2}}^2$	$h_{F_{4:3}}^2$	h_{RI}^2	$h_{F_{3:2}}^2$	$h_{F_{4:3}}^2$	h_{RI}^2	$h_{F_{3:2}}^2$	$h_{F_{4:3}}^2$	h_{RI}^2
.01	265	217	199	174	153	147	149	133	130
.05	322	286	276	210	203	204	181	176	182
.10	413	407	417	275	292	313	234	254	277
.15	552	607	644	369	431	481	315	381	427
.20	767	945	1,017	508	672	766	434	587	680
.25	1,114	1,546	1,688	746	1,101	1,264	638	973	1,122
.30	1,763	2,791	3,009	1,172	1,991	2,254	1,003	1,741	2,002
.35	3,171	5,779	6,048	2,110	4,125	4,534	1,807	3,609	4,029
.40	7,195	15,299	15,272	4,792	10,925	11,452	4,105	9,559	10,178
.45	28,930	72,918	68,099	19,275	52,081	51,072	16,520	45,570	45,390

* h^2 values rounded. $\phi = \sigma_{Aq}^2 / \sigma_A^2 = 0.10$. $h_{F_{3:2}}^2$, $h_{F_{4:3}}^2$, and h_{RI}^2 estimated with four replications and 10 plants per plot.

The values presented for the sample sizes are for $\phi = 0.10$. For $\phi = 0.05$ the sample size required would be nearly twice that required for $\phi = 0.10$. It seems unlikely

that for $r > 0.20$ and $\phi < 0.10$ the linkage between a QTL and marker could be detected, because of the sample size required. Only a QTL linked to a marker was considered; however, if more QTLs are linked to the same marker the ratio $\phi = \sigma_{Aq}^2 / \sigma_A^2$ could increase and, thus, the sample size necessary for QTL mapping would decrease.

ACKNOWLEDGMENTS

This research was carried out during the author's study leave at Iowa State University, sponsored by the Interamerican Development Bank/Universidade de São Paulo (BID/USP) program.

Publication supported by FAPESP.

RESUMO

Os tamanhos de amostras requeridas para detectar ligação entre um marcador e QTL para populações F_2 não repetidas, e para genótipos em experimentos com repetições de $F_{3;2}$, $F_{4;3}$, e linhagens recombinantes, foram investigados. O teste F com $P \leq 0.01$ na análise de variância foi usado para detectar ligação. As esperanças dos quadrados médios foram manipulados e a razão de F foi expressa em termos da distância entre o marcador e o QTL (r), a herdabilidade (h^2) e o nível de dominância do caráter, o tamanho da amostra, e o parâmetro $\phi = \sigma_{Aq}^2 / \sigma_A^2$. Os tamanhos das amostras foram estimados através do teste F. O tamanho da amostra para o mapeamento dos QTLs depende dos fatores em que foi expresso o teste F. Quando r diminui, e h^2 e ϕ aumentam o tamanho da amostra diminui; e quando r aumenta e h^2 e ϕ diminuem o tamanho da amostra aumenta.

REFERENCES

- Abler, B.S.B., Edwards, M.D. and Stuber, C.W. (1991). Isoenzymatic identification of quantitative trait loci in crosses of elite maize inbreds. *Crop. Sci.* 31: 267-274.
- Asins, M.J. and Carbonell, E.A. (1988). Detection of linkage between restriction fragment length polymorphism marker and quantitative traits. *Theor. Appl. Genet.* 76: 623-626.
- Burr, B., Burr, F.A., Thompson, K.H., Albertson, M.C. and Stuber, C.W. (1988). Gene mapping with recombinant inbreds in maize. *Genetics* 118: 519-526.
- Cowen, N.M. (1988). The use of replicated progenies in marker-based mapping of QTLs. *Theor. Appl. Genet.* 75: 857-862.
- Edwards, M.D., Stuber, C.W. and Wendel, J.F. (1987). Molecular marker-facilitated investigations of quantitative trait loci in maize. I. Numbers, genomic distribution and types of gene action. *Genetics* 116: 113-125.
- Falconer, D.S. (1989). *Introduction to Quantitative Genetics*. 3rd Edition. John Wiley & Sons, New York, pp. 438.
- Haldane, J.B.S. and Waddington, C.H. (1931). Inbreeding and linkage. *Genetics* 16: 357-374.

- Hamrick, J.L. and Allard, R.W. (1975). Correlations between quantitative characters and enzyme genotype in *Avena barbata*. *Evolution* 29: 438-442.
- Helentjaris, R., Slocum, M., Wright, S., Shaeffer, A. and Nienhuis, J. (1986). Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. *Theor. Appl. Genet.* 72: 761-769.
- Keim, P., Diers, B.W. and Shoemaker, R.C. (1990). Genetic analysis of soybean hard seedness with molecular markers. *Theor. Appl. Genet.* 79: 465-469.
- Knapp, S.J. and Bridges, W.C. (1990). Using molecular markers to estimate quantitative trait locus parameters: power and genetic variances for unreplicated and replicated progeny. *Genetics* 126: 769-777.
- Ledig, F.T., Guries, R.P. and Bonefeld, B.A. (1983). The relation of growth to heterozygosity in pitch pine. *Evolution* 37: 1227-1238.
- Mather, K. and Jinks, J.L. (1982). *Biometrical Genetics*. Chapman and Hall, New York, pp. 396.
- McMillan, I. and Robertson, A. (1974). The power of methods for detection of major genes affecting quantitative characters. *Heredity* 32: 349-356.
- Paterson, A.H., Damon, S., Hewitt, J.D., Zamir, D., Rabinowitch, H.D., Lincoln, S.E., Lander, E.S. and Tanksley, S.D. (1991). Mendelian factors underlying quantitative traits in tomato: comparison across species, generations and environments. *Genetics* 127: 181-197.
- Soller, M. and Beckmann, J.S. (1990). Marker-based mapping of quantitative trait loci using replicated progenies. *Theor. Appl. Genet.* 80: 205-208.
- Soller, M., Genizi, A. and Brody, T. (1976). On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. *Theor. Appl. Genet.* 47: 35-39.
- Souza Jr., C.L. (1992). Genetic variances of molecular markers of F₂ and related populations. *Rev. Brasil. Genet.* 15: 913-926.
- Weller, J.I., Soller, M. and Brody, T. (1988). Linkage analysis of quantitative traits in an interspecific cross of tomato (*Lycopersicon esculentum* x *Lycopersicon pimpinellifolium*) by means of genetic markers. *Genetics* 118: 329-339.

(Received August 26, 1991)