Theor Appl Genet (1992) 83:582–588

Power of different sampling strategies to detect quantitative trait loci variance effects

J. I. Weller * and A. Wyler **

Institute of Animal Sciences, A. R. O., The Volcani Center, Bet Dagan 50250, Israel

Received February 11, 1991; Accepted July 18, 1991
Communicated by D. Van Vleck

Summary. Many studies have shown that segregating quantitative trait loci (QTL) can be detected via linkage to genetic markers. Power to detect a QTL effect on the trait mean as a function of the number of individuals genotyped for the marker is increased by selectively genotyping individuals with extreme values for the quantitative trait. Computer simulations were employed to study the effect of various sampling strategies on the statistical power to detect QTL variance effects. If only individuals with extreme phenotypes for the quantitative trait are selected for genotyping, then power to detect a variance effect is less than by random sampling. If 0.2 of the total number of individuals genotyped are selected from the center of the distribution, then power to detect a variance effect is equal to that obtained with random selection. Power to detect a variance effect was maximum when 0.2 to 0.5 of the individuals selected for genotyping were selected from the tails of the distribution and the remainder from the center.

Key words: Quantitative trait loci – Variance effects – Sampling strategies

Introduction

Numerous studies have shown that individual loci affecting quantitative traits (Henceforth QTL) can be detected via linkage to genetic markers (Sax 1923; reviewed by Soller 1990). Studies of this nature have been facilitated in the last decade by the development of methods to detect polymorphisms at the DNA level (Beckmann and Soller 1983; Jeffreys et al. 1985; Kashi et al. 1990; Litt and Luty 1989; Soller and Beckmann 1983; Weber and May 1989). A few studies have also estimated marker-linked QTL effects on the variance of traits (Edwards et al. 1987; Stuber et al. 1987; Weller 1987; Weller et al. 1988; Zhuchenko et al. 1979). Under the null hypothesis of equal variance, the ratio of the variances of two different marker genotypes will have a central *F*-distribution. Significant deviation from the expected ratio of unity is indicative of a marker-linked variance effect. All of these studies found significant effects.

For certain traits, variance may be more important economically than the mean. For crossbreeding plants, it is important that all individuals flower at the same time. For efficient mechanical harvesting, it is desirable that all fruit should ripen at the same time. Likewise it is desirable that all chicks should hatch after the same incubation period. Furthermore, as noted by Asins and Carbonell (1988), even if all of the QTL genotypes in an F_2 of a cross between two inbred lines have equal variance, the marker genotypes may have different variances due to dominance at the QTL and incomplete linkage between the QTL and the genetic marker. Thus, even if the QTL genotypes have equal variance, a marker-associated variance effect contributes information that can be utilized by maximum likelihood methodology to map the QTL (Weller 1986).

Genotype determination at the DNA level is still quite costly, both in terms of labor and laboratory requirements, on a per individual basis (Kashi et al. 1986). For many plant species the cost of growing and scoring individual plants for the traits of interest may be much less than the cost of the genotype determination. Furthermore, for certain animal species, especially dairy cattle, large data banks on traits of economic interest are available at virtually no cost (Weller et al. 1990). Rather

* To whom correspondence should be addressed
** *Permanent address:* Institut fuer Nutztierwissenschaften, TAN, ETH-Zentrum, CH-8092, Switzerland

than genotyping a random sample from the population scored for a quantitative trait, individuals should be selected so as to maximize power to detect segregating QTL. Several studies have shown that power to detect a QTL effect on the trait mean is increased by selecting those individuals with the most extreme phenotypes for the trait (Lander and Botstein 1989; Lebowitz et al. 1987; Soller 1990). However, these studies did not consider the effect of "selective genotyping" on power to detect segregating QTL variance effects.

If the population is a mixture of two genotypes with different variances but equal means for the quantiative trait, then the genotype with the higher variance will have a higher frequency at the extreme ends of the distribution, while the genotype with the lower variance will have a higher frequency near the mean. Therefore, if only individuals with extreme phenotypes are selected for genotyping, then most will be of the genotype with the higher variance. However, no difference is expected in the relative frequency of the two genotypes in the groups selected for low versus high trait value. Thus, both individuals with extreme phenotypes and individuals with phenotypes close to the mean should be genotyped to detect a marker-linked variance effect.

The objectives of this study were to determine: (1) the effect of selective genotyping on the power to detect variance effects, (2) the power of alternative sampling strategies to detect variance effects, as a function of total population size and number of individuals genotyped, and (3) the optimum sampling strategy to detect QTL variance effects.

Genetic and statistical models

A number of different experimental designs of QTL detection have been considered (Knapp et al., 1990). Among them are analyses of whole-sib pairs, backcrosses (BC), testcrosses (TC), F_2 progeny from a cross between inbred lines, doubled haploids, recombinant inbred lines, and various combinations thereof. Most of the early studies considered linkage between a QTL and a single genetic marker, while most of the more recent studies have assumed a QTL bracketed between two linked genetic markers (Knapp et al. 1990; Paterson et al. 1988, Tanksley et al. 1982; Weller 1987; Weller et al. 1988). For this study, the significant difference between these designs is the number of different QTL and marker genotypes. We chose the BC design with inbred lines and a single genetic marker because it is the most amenable to analysis. When homozygosity is assumed for a different allele at both the QTL and the genetic marker in each of the parental strains, BC progeny have only two genotypes for both the QTL and the genetic marker. Under these same conditions, F_2 progeny with a single segregating marker have three genotypes for each locus, and F_2 progeny for two linked markers have nine marker and three QTL genotypes. In addition we assumed:

1) complete linkage between the genetic marker and the QTL;
2) that the putative QTL affects only the variance, but not the mean of the quantitative trait;
3) a normal distribution for the quantitative trait, with a phenotypic variance of unity for the genotype with the lower variance;
4) a probability of 0.5 for either genotype at both loci in the BC population.

In the discussion we will consider the implications of removing some of these restrictions.

Description of the simulations

BC populations were simulated using the NORMAL and UNIFORM functions of SAS (SAS Institute Inc. 1985). The genotype of each individual was determined by random sampling from a uniform distribution between 0 and 1. If a value of less than 0.5 was obtained, then the individual was assigned the QTL and linked-marker genotype with the lower variance ($\alpha_1^2 = 1$). Otherwise, the individual was assigned the QTL with the higher variance ($\alpha_2^2 > 1$). The simulated ratio of the variances, α_2^2/α_1^2, was denoted G*. The trait value for each individual was determined by random sampling from a normal population with a mean of zero and a variance of either α_1^2 or α_2^2, as previously determined. The number of individuals per population (N) was varied from 500 to 4000 individuals.

Three stratified sampling strategies were tested; tails only, center only, and both tails and center. For the last sampling strategy, the ratio of individuals sampled from the tails to the total sample, R, was varied form 0.1 to 0.9. Three values were tested for p, the fraction of the population sampled: 0.05, 0.1, and 0.15. The standard normal integral was used to determine the appropriate truncation points for each sampling strategy, under the assumption that an equal mixture of two normal distributions with the same means and different variances will approximate a normal distribution. The variance of the total population will be equal to the mean of the variances of the two component distributions. This approximation was compared to the precise truncation points obtained by summing over the two separate normal integrals from $-\infty$ until the required probability was obtained.

Number of individuals of each genotype sampled, "observed" variances of the sampled individuals for each genotype from both the center and the tails of the distribution, and ratio of these variances were computed for each population sampled. Ratios of the observed variances were denoted G, as opposed to G*, the simulated ratio. Twenty replicates were simulated for each combination of parameters.

Estimation of statistical power

The expected distribution of the G-statistic under the null hypothesis of equal variance is not equivalent to the central F-distribution for two reasons. First, the stratified samples will have non-normal distributions. Furthermore, since the individuals were selected by truncation selection on mixtures of two distributions, the two samples being compared are not independent. Therefore, the expected distribution of the G-statistic under the null hypothesis was estimated by sampling from simulated populations of N = 2,000 by the same strategies given above, with p = 0.1 (number of individuals selected for genotyping, $n = 200$), and R varied from 0 two 1 by increments of 0.1. One thousand replicates were simulated for each value of R. The distributions of the G values for each group of 1,000 replicates were compared to the theoretical F-distribution for appropriate degrees of freedom. Deviation of the observed G distribution from the central F-distribution was tested by Chi-square (X^2).

Power to detect a variance effect with type I error of $\alpha = 0.05$ was estimated with R varying from 0 to 1 by increments of 0.1; N = 2,000, and $n = 200$, under the alternative hypothesis of G* = 1.3. One thousand populations were simulated for each R value. Power with stratified sampling was compared to power obtained with random sampling. With random sampling the ratio of the variances of the two genotypes will have a non-central F-distribution under the alternative hypothesis. Probabilities for the F-distribution were computed using the PROBF function of SAS (SAS Institute Inc 1985).

Results

Distribution of the G-statistic, under the null hypothesis of equal variances from stratified samples with p = 0.1, $n = 200$, G* = 1, and R varied from 0 to 1.0 by increments of 0.2 is shown in Table 1. The theoretical F-distribution for a sample of 1,000 populations and the X^2 values for deviation from this distribution are also given. At R < 0.4, the variance of G was greater than F; at R = 0.4, the distribution of the G-statistic was not significantly different from the F-distribution ($\alpha < 0.05$), and at R > 0.4, the variance was lower. Thus, assumption of an F-distribution for the null hypothesis resulted in overestimation of the significance level to detect a variance effect for R-values > 0.4 and underestimation of the significance level for R-values > 0.4. Even with R = 0.2, less than 1% of the samples had a G-value > 2.

The approximate truncation values used are compared to the precise truncation values in Fig. 1 for G* = 1.5, R = 1.0, and p varied from 0.01 to 0.1. At p = 0.08 the two truncation values were nearly equal. At lower p-values the precise truncation point was higher, while at p = 0.1, the precise truncation point was slightly lower. Thus at low p-values the expected sample size selected from the tails is slightly larger than that obained with the precise truncation values.

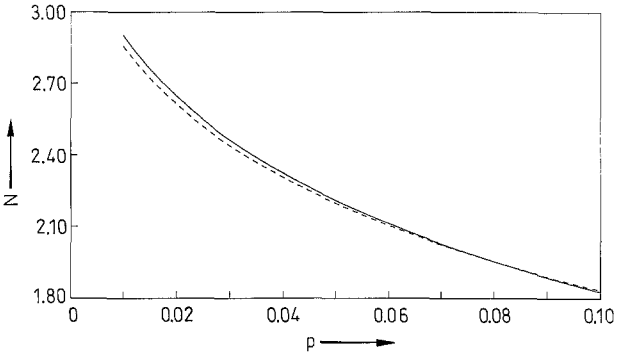

Fig. 1. Truncation points for a mixture of two normal distributions with equal means and a variance ratio of 1.5, as a function of p; the probability > Z where Z is the abscissa for the normal distribution with the lower variance in standard deviation units. — Precise truncation values, based on summing the normal integrals; – – approximate truncation values, based on assuming a normal distribution for the mixture

Table 1. Distribution of variance ratios (G) as a function of R = (sample from tails)/(total sample) for a simulated variance ratio, G* = 1.0[a]

G[b]	Frequencies (R)						$F_{100/100}$ distribution
	0	0.2	0.4	0.6	0.8	1.0	
<0.4	0	2	0	0	0	0	0.0
0.4–0.5	0	16	0	0	0	0	0.3
0.5–0.6	0	34	5	0	0	0	5.3
0.6–0.7	0	77	30	3	0	0	32.4
0.7–0.8	15	109	78	57	8	0	95.1
0.8–0.9	96	125	155	164	113	1	166.5
0.9–1.0	243	137	200	281	392	175	200.3
1.0–1.1	311	129	183	282	340	651	182.7
1.1–1.2	207	96	157	150	127	166	135.5
1.2–1.3	93	86	90	45	17	6	86.1
1.3–1.4	25	58	61	11	3	1	48.6
1.4–1.5	7	39	29	5	0	0	25.1
1.5–1.6	3	30	7	1	0	0	12.1
1.6–1.7	0	15	5	1	0	0	5.6
1.7–1.8	0	18	0	0	0	0	2.5
1.8–1.9	0	14	0	0	0	0	1.1
1.9–2.0	0	7	0	0	0	0	0.4
>2.0	0	8	0	0	0	0	0.2
χ^2[c]	281	846	16	218	600	1,676	

[a] One thousand populations were simulated for each sampling strategy; 200 individuals were selected from populations of 2,000
[b] Variance ratio. G values <0.6 and >1.6 were combined for the χ^2 test, thus $df = 12$
[c] χ^2 for deviation of the G-distribution from the F-distribution; 0.05 and 0.01 significance levels for χ^2 with 12 df are 21 and 26, respectively

Mean G-values as a function of R are presented in Table 2 for p = 0.1 and N = 2,000, with R varied from 0

Table 2. Observed mean variance ratios (G) and proportion of individuals selected from each genotype from a stratified sample as a function of the simulated variance ratio (G*) and R = (sample from tails)/(total sample)

G*	Statistic[a]	Fraction of sample from tails/total sample (R)							
		0.0	0.1	0.2	0.4	0.5	0.67	0.75	1.00
1.1	G	1.05	1.33	1.22	1.19	1.16	1.15	1.10	1.06
	n_1/n_2	1.09	1.04	0.99	0.96	0.88	0.91	0.91	0.87
1.2	G	0.98	2.87	2.06	1.53	1.38	1.23	1.17	1.06
	n_1/n_2	1.17	1.06	0.98	0.89	0.83	0.78	0.75	0.71
1.3	G	1.06	4.50	2.78	1.74	1.46	1.42	1.21	1.12
	n_1/n_2	1.15	1.04	0.97	0.85	0.81	0.74	0.72	0.67
1.4	G	1.95	29.56	3.61	2.32	1.87	1.51	1.41	1.14
	n_1/n_2	1.17	1.06	0.97	0.84	0.73	0.68	0.66	0.58
1.5	G	1.00	5.86	3.97	2.59	2.02	1.62	1.56	1.16
	n_1/n_2	1.35	1.03	0.90	0.80	0.70	0.65	0.58	0.51

n_1 = number of individuals with QTL genotype with the lower variance; n_2 = number of individuals with QTL genotype with the higher variance

[a] G = mean variance ratio obtained from 20 simulations, each consisting of 200 individuals genotyped from a population of 2000 individuals

to 1.0. With either all of the individuals selected from the tails or all of the individuals selected from the middle, observed G-values were smaller than the simulated G*. The critical central F-value for $\alpha = 0.05$ is 1.4. For R = 1, none of the G-values were greater than the critical F-value, and for R = 0, only G* = 1.4 was greater. Within these extremes, G-values increased with decreased R. With R = 0.1 (only 1% of the individuals from the tails), the G-values were more than double the G* values. Since the expected number of individuals selected from each tail was only 10, it is possible that all, or nearly all of the individuals from the tail could be of a single genotype. This would have resulted in an enormous ratio and could explain the value of 29.56 obtained for G* = 1.4, even though this value was the mean of 20 simulations. Similar results were more frequent with smaller N (data not shown).

For R-values from 0.67 to 0.75, G-values were similar to those obtained by random sampling, while the variance of the G-distribution under the null hypotheses was less than that of the F-distribution. As the fraction sampled from the tails increased, the number of individuals having the genotype with higher variance also increased. With R = 1.0, up to 65% of the individuals selected had the genotype with the greater variance. This difference is significant by X^2 with one degree of freedom ($\alpha = 0.01$).

The effects of varying N and n on mean G-value were tested separately. The results are shown in Table 3 for mean G-value from a stratified sample with R = 0.67, p = 0.1, and N varying from 500 to 2,000. The expected number of individuals selected for genotyping for each simulation were Np = n = 50, 100, and 200, respectively. In general, G values were greater than G*, although, as expected, there was more random variation for small samples.

The effects of varying N and p, with n = 200 and R = 0.67, are presented in Table 4. The fraction selected was varied from 0.05 to 0.15; thus N was varied from 4000 to 1333. Differences for G for the different values for p increased with the increase in G*, but, in general, as p decreased, G values increased only slightly. For example, for p = 0.05 and G* = 1.3, G was 1.51. The same G value was obtained for p = 0.1 and G* = 1.4, and was slightly less than the G value obtained for p = 0.15 and G* = 1.5.

Distributions of G with G* = 1.3, N = 2,000, n = 200, and R varied from 0 to 1 by increments of 0.2 are presented in Table 5. The non-central F distribution for an expected variance ratio of 1.3 is also given. Statistical power with type I error of $\alpha = 0.05$ and R varied by increments of 0.1 has been plotted in Fig. 2. A third-order polynomial curve was fitted to the data points. Although the power values are each the result of 1,000 simulations computed under the null and alternative hypothesis, random variation is still significant. For example, power at R = 1.0 is slightly greater than power at R = 0.9. Under the null hypothesis of equal variance and type I error of $\alpha = 0.05$, the critical values were 1.4 for the F-distribution and 1.7, 1.4, and 1.3 for the G-distributions with R of 0.2, 0.4, and 0.6, respectively. The corresponding critical values for $\alpha = 0.01$ were 1.6 for the F-distribution and 2.0, 1.5, and 1.4 for the G-distributions (Table 1). Critical

Table 3. Observed mean variance ratios (G) and proportion of individuals selected from each genotype from a stratified sample, as a function of the simulated variance ratio (G*) and number of individuals scored for the quantitative trait (N), with 0.033 of these individuals selected from each tail and 0.033 from the center of the distribution

G*	Statistic[a]	Number scored for trait (N)		
		500	1000	2000
1.10	G	1.18	1.09	1.15
	n_1/n_2	1.05	0.82	0.90
1.20	G	1.34	1.34	1.23
	n_1/n_2	0.85	0.86	0.78
1.30	G	1.40	1.29	1.42
	n_1/n_2	0.74	0.70	0.74
1.40	G	1.85	1.51	1.51
	n_1/n_2	0.70	0.60	0.68
1.50	G	1.79	1.20	1.62
	n_1/n_2	0.46	0.66	0.65

n_1 = number of individuals with QTL genotype with the lower variance; n_2 = number of individuals with QTL genotype with the higher variance. The expectation of $n_1 + n_2 = N(0.1)$

Table 4. Observed mean variance ratios (G) and proportion of individuals selected from each genotype from a stratified sample as a function of the simulated variance ratio (G*), and the fraction sampled for genotyping from the total population scored for the quantitative trait (p), with one-third selected from each tail and one-third from the center of the distribution[a]

G*	Statistic[b]	Proportion selected for genotyping (p)		
		0.05	0.10	0.15
1.10	G	1.11	1.15	1.20
	n_1/n_2	0.92	0.90	0.91
1.20	G	1.31	1.23	1.14
	n_1/n_2	0.74	0.78	0.83
1.30	G	1.51	1.42	1.37
	n_1/n_2	0.66	0.74	0.77
1.40	G	1.60	1.51	1.44
	n_1/n_2	0.64	0.68	0.75
1.50	G	1.91	1.62	1.57
	n_1/n_2	0.49	0.65	0.69

[a] The number of individuals scored for the trait (N) were 4000, 2000, and 1333 for p = 0.05, 0.10, and 0.15, respectively. N was set so that the number of individuals selected for genotyping n = Np = 200 for each value of p

[b] G = mean variance ratio obtained from 20 simulations; n_1 = number of individuals with QTL genotype with the lower variance; n_2 = number of individuals with QTL genotype with the higher variance

Table 5. Distribution of observed variance ratios (G) for a simulated variance ratio, G* = 1.3, as a function of R = (sample from tails)/(total sample)[a]

G	Frequencies (R)						$F_{100/100}$[b] distribution
	0.0	0.2	0.4	0.6	0.8	1.0	
<1.0	504	2	1	4	7	47	92.2
1.0–1.1	264	6	6	19	106	449	106.4
1.1–1.2	147	3	11	58	255	436	143.4
1.2–1.3	60	9	40	129	313	67	157.0
1.3–1.4	17	15	51	185	204	1	146.0
1.4–1.5	7	29	86	185	85	0	119.3
1.5–1.6	0	35	110	167	28	0	87.8
1.6–1.7	1	40	118	106	1	0	59.4
1.7–1.8	0	44	126	65	1	0	37.6
1.8–1.9	0	45	87	39	0	0	22.6
1.9–2.0	0	41	90	21	0	0	13.0
2.0–2.2	0	114	123	15	0	0	11.0
2.2–2.4	0	101	79	7	0	0	3.1
2.4–2.6	0	90	38	0	0	0	0.8
2.6–2.8	0	79	17	0	0	0	0.2
2.8–3.0	0	72	7	0	0	0	0.0
3.0–3.2	0	67	2	0	0	0	0.0
3.2–3.4	0	54	6	0	0	0	0.0
3.4–3.6	0	36	7	0	0	0	0.0
3.6–3.8	0	21	1	0	0	0	0.0
3.8–4.0	0	26	7	0	0	0	0.0
>4.0	0	74	1	0	0	0	0.0

[a] One thousand populations were simulated for each sampling strategy; 200 individuals were selected from populations of 2,000

[b] Expected frequency for the theoretical F distribution with 100 *df* in both numerator and denominator and a non-central parameter of 30, i.e., expectation of 1.3

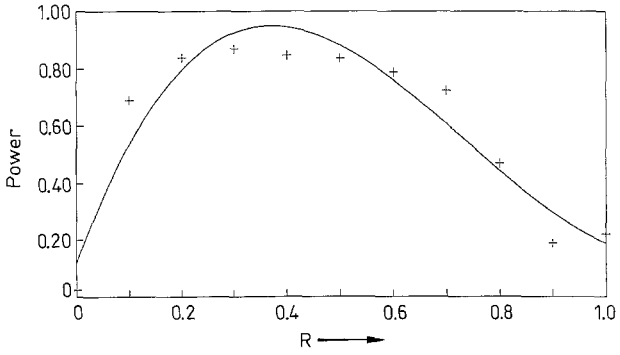

Fig. 2. Power to detect a variance effect, with Type 1 error, $\alpha = 0.05$, and an alternative hypothesis that the expected variance ratio is 1.3, as a function of R, the fraction of the sample selected from the tails of the distribution. Two hundred individuals were selected from populations of 2,000. One thousand replicates were simulated for each R-value for both the null and alternative hypothesis. A third-order polynomial was fitted to the values tested

values of 1.4 and 1.6 for the F-distribution correspond to power of 0.35 and 0.15, respectively. Power for the G-statistic was maximum for R between 0.2 and 0.5. In this range power was above 0.8 (Table 5, Fig. 2). Power for $\alpha = 0.01$ was 0.71, 0.77, and 0.51, for $R = 0.2$, 0.4 and 0.6, respectively. Thus for intermediate R values the G-statistic is significantly more powerful than the F-test on a random sample. For $R = 0.6$ and $\alpha = 0.01$ the median of the observed G-values was slightly greater than 1.4, as expected for power of 0.5.

Discussion

Power to detect a QTL variance effect with selective genotyping, as suggested by Lebowitz et al. (1986) and Lander and Botstein (1989), was smaller than with random sampling. However, if a sample of individuals was also selected from the middle of the distribution, then power to detect a variance effect in most of the cases tested was greater than that obtained by a random sample of the same size.

If the primary interest in the experiment is to estimate mean effects, but variance effects are also of interest, a reasonable sampling structure would be to maximize power to detect mean effects, under the restriction that power to detect variance effects should be no less than in a random sample. If 2% of the individuals are selected from the middle and 8% from the tails, then power is still slightly greater than that obtained with random sampling (0.4 versus 0.35 for $\alpha = 0.05$). With this sampling strategy, power to detect mean effects will be twice the power obtained by random sampling (Soller 1990).

Unlike the results for mean effects, with $R = 0.67$, little is gained by increasing N with a constant n, within the range of values tested in Table 4 ($N = 1{,}333$ through to 4,000). A rather surprising result was that although G values were low for both $R = 0$ and $R = 1$, the mean G value increased as R decreased even to 0.1. However, the variance of the expected distribution of G-values also increased as R decreased, thus maximum power as a function of R was obtained within the range of 0.2 to 0.5.

As R increased, the number of individuals selected with the genotype having the higher variance increased. For example, with $R = 1$ and $G^* = 1.5$, 65% of the individuals selected were from the genotype with the higher variance. However, a preponderance of one genotype in the selected sample is not conclusive evidence of a variance effect. Other factors could produce this result, such as unequal viability of the genotypes.

The results expected if some of the restrictions are relaxed will now be considered briefly. We assumed that the two marker genotypes were at equal frequency in the population. Statistical power for a given number of individuals sampled is maximum when the two groups being compared are at equal frequency. Thus, if the two marker genotypes are present in the population at unequal frequency, the statistical power will be decreased. We assumed that the quantitative trait had a normal distribution. If this is not the case, it will generally be possible to use a mathematical transformation to approximate normality and then analyze the transformed variable (Weller 1987). If the QTL affects the mean of the distribution, then one genotype will have a higher frequency in one tail, while the other genotype will have a higher frequency in the other tail (Lebowitz et al. 1986). However, if the variances of the genotypes are equal, they will be expected to have equal frequencies in the middle of the distribution, and the variance of the genotypes in the selected samples will be equal. If there are more than two alleles present at the QTL, then the marker genotype distributions will be a mixture of several QTL genotypes. In this case the variance of each distribution will also depend on the difference between the means of the QTL genotypes.

Asins and Carbonell (1988) showed that for an F_2 population with incomplete linkage between the QTL and the genetic marker, the variance of marker genotypes will depend on the dominance relationship between the QTL alleles. Based on this finding they further concluded that all marker-associated variance effects can be explained within this context. This conclusion is clearly inaccurate. Zhuchenko et al. (1979) found significant variance effects in backcross populations where the dominance relationship would have no effect on the variance ratio. Furthermore, when using maximum likelihood to estimate both the means and variances of the three QTL genotypes in an F_2 population, Weller (1986, 1987) found significant differences between the variances of the QTL genotypes, both for loci with complete and partial dominance.

Acknowledgements. We thank M. Soller, E. Ezra and Margaret Mackinnon for useful discussions. This research was supported by a grant by the Israel Cattle Breeders Association.

References

Asins MJ, Carbonell EA (1988) Detection of linkage between restriction fragment length polymorphisms and quantitative traits. Theor Appl Genet 76:623–626

Beckmann JS, Soller M (1983) Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. Theor Appl Genet 67:35–43

Edwards MD, Stuber CW, Wendel JF (1987) Molecular-marker-facilitated investigations of quantitative trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics 116:113–125

Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable "minisatellite" regions in human DNA. Nature 314:67–72

Kashi Y, Iraqi F, Tikochinsky Y, Ruzitsky B, Nave A, Beckmann JS, Friedmann A, Soller M, Gruenbaum Y (1990)

$(TG)_n$ uncovers a sex-specific hybridization pattern in cattle. Genomics 7:31–36

Kashi Y, Soller M, Hallerman EM, Beckmann JS (1986) Restriction fragment length polymorphisms in dairy cattle genetic improvement. Proc 3rd World Cong Genet Appl Livest Prod 12:57–63

Knapp SJ, Bridges WC, Birkes D (1990) Quasi-Mendelian analyses of quantitative trait loci using molecular marker linkage maps. Theor Appl Genet 79:583–592

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Lebowitz RJ, Soller M, Beckmann JS (1987) Trait-based analyses for the detection of linkage between marker loci and quantitative trait loci in crosses between inbred lines. Theor Appl Genet 73:556–562

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44:397–401

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantiative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721–726

SAS Institute Inc (1985) SAS users' guide: basics, version 5. SAS Institute Inc, Cary, N.C.

Sax K (1923) The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8:552–560

Soller M (1990) Genetic mapping of the bovine genome using DNA-level markers with particular attention to loci affecting quantitative traits of economic importance. J Dairy Sci 73:2628–2646

Soller M, Beckmann JS (1983) Genetic polymorphisms in varietal identification and genetic improvement. Theor Appl Genet 67:25–33

Stuber CW, Edwards MD, Wendel JF (1987) Molecular marker-facilitated investigations of quantitative trait loci. II. Factors influencing yield and its component traits. Crop Sci 27:639–648

Tanksley SD, Medina-Filho H, Rick CM (1982) Use of naturally-occurring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. Heredity 49:11–25

Weber JL, May PE (1989) Abundant class of Human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388–396

Weller JI (1986) Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. Biometrics 42:627–640

Weller JI (1987) Mapping and analysis of quantitative trait loci in *Lycopersicon*. Heredity 59:413–421

Weller JI, Soller M, Brody T (1988) Linkage analysis of quantitative traits in an Interspecific cross of tomato (*Lycopersicon esculentum* and *Lycopersicon pimpinellifolium*) by means of genetic markers. Genetics 118:329–339

Weller JI, Kashi Y, Soller M (1990) Estimation of sample size necessary for genetic mapping of quantitative traits in dairy cattle using genetic markers. J Dairy Sci 73:2525–2537

Zhuchenko AA, Samovol AP, Korol AB, Andryushchenko VK (1979) Linkage between loci of quantitative characters and marker loci. II. Influence of three tomato chromosomes on variability of five quantitative characters in backcross progenies. Genetika 15:672–683