Estimating variance effect of QTL: an important prospect to increase the resolution power of interval mapping

A. B. KOROL[1]*, Y. I. RONIN[1], Y. TADMOR[2], A. BAR-ZUR[2], V. M. KIRZHNER[1] AND E. NEVO[1]

[1] *Institute of Evolution, University of Haifa, Mount Carmel, Haifa 31905, Israel*
[2] *Corn Breeding Unit, Neve Yaar Research Center ARO, P.O.B. 90000, Haifa 31900, Israel*

(*Received 9 January 1995 and in revised form 6 September 1995*)

Summary

Equal variances within quantitative trait locus (QTL) groups in the segregating population are a usual simplifying assumption in QTL mapping. The objective of this paper is to demonstrate the advantages of taking into account potential variance effect of QTLs within the framework of standard interval mapping approach. Using backcross case as an example, we show that the resolution power of the analysis may be increased in the presence of variance effect, if the latter is allowed for in the model. For a putative QTL (say, A/a) one can compare two situations, (i) $\sigma_{Aa}^2 = \sigma_{aa}^2 = \sigma_0^2$ and (ii) $\sigma_{Aa}^2 \neq \sigma_{aa}^2$. It was found that, if the variance effect of A/a is large enough, then in spite of the necessity to evaluate an increased number of parameters, the more correctly specified model provides an increase in the resolution power, as compared to the situation (i). This is not unexpected, if either σ_{Aa}^2 or σ_{aa}^2 in (ii) is lower than σ_0^2 from (i). But our conclusion holds even if $\sigma_{Aa}^2 > \sigma_{aa}^2 = \sigma_0^2$ or $\sigma_{aa}^2 > \sigma_{Aa}^2 = \sigma_0^2$. These advantages are illustrated on sweet corn data (F_3 families of F_2 genotypes). In particular, the log-likelihood test statistics and the parameter estimates obtained for a QT locus in the distal region of chromosome 2 show that the allele enhancing the trait is recessive over the opposite allele simultaneously for the mean value and variance.

1. Introduction

The resolution capacity of marker analysis of quantitative traits (QTs) is the major factor affecting the practical importance of QT loci (QTL) mapping. A detailed discussion of the issues concerning the power of tests for detecting linkage can be found in many publications (e.g. Demenais *et al.* 1988; Lander & Botstein, 1989; Soller & Beckmann, 1990; Carbonell *et al.* 1993). The precision of the parameter estimates depends on the effect of the QTL in question relative to the total phenotypic variance of the trait in the mapping population. In other words, the higher the discrepancy between the distribution densities of the QTL groups (say, $f_{aa}(x)$ and $f_{Aa}(x)$, for a backcross), the better is the expected resolution. Among several possibilities to improve the precision of mapping, it is worth mentioning selective sampling (Lebowitz *et al.* 1987; Carey & Williamson, 1991; Darvasi & Soller,

1992), replicated progeny testing (Soller & Beckmann, 1990), sequential experimentation (Boehnke & Moll, 1989; Motro & Soller, 1993), multi-interval analysis (Jansen & Stam, 1994; Zeng, 1994) and multi-trait analysis (Korol *et al.* 1994, 1995; Ronin *et al.* 1995).

With few exceptions (Zhuchenko *et al.* 1979; Korol *et al.* 1981, 1994; Weller, 1986, 1987), variance effects of QTLs have not been considered previously. However, as rightly pointed out by Weller & Wyler (1992), the effect of a QTL on the variance is sometimes likely to be economically more critical than on the mean (e.g. for earliness, flowering time, ripening time under machine harvesting, time to hatching in chicken). The same applies to QTLs related to fitness traits in natural populations, e.g. seed dormancy, or flowering time. Clearly, this aspect of the problem might be very important in supporting climatic adaptive radiation into increasingly hostile conditions.

The objective of this paper is to demonstrate the advantages of taking into account potential variance effects of QT loci within the framework of the standard interval mapping approach. Using backcross case as an example, we show that the resolution power of the

* Corresponding author: A. B. Korol, Institute of Evolution, University of Haifa, Mount Carmel, Haifa 31905, Israel. Tel. (972)-4240-449, Fax: (972)-4246-554, E-mail: RABI307@UVM. Haifa.ac.il.

marker analysis of QTL may be increased in the presence of the variance effect, if the latter is allowed for by the model. This advantage is demonstrated on sweet corn data (F_3 families of F_2 genotypes) (Tadmor *et al.* 1994).

2. Interval analysis of QTL

(i) *Mixture-model formulation*

Assume that the putative QT locus A/a resides in an interval flanked by two marker loci, M_1/m_1 and M_2/m_2, with recombination rates r_1 and r_2 in intervals $M_1/m_1 - A/a$ and $A/a - M_2/m_2$. Many modes of interference in the interval could be considered, but we assume no interference, so that $r = r_1 + r_2 - 2r_1 r_2$, where r is the recombination rate between M_1/m_1 and M_2/m_2. Based on the marker scores and measurements of the QT of interest (say, x) for individuals from the mapping population, we should test whether or not variation of x indeed depends on the interval $M_1/m_1 - M_2/m_2$, and, if so, identify the corresponding locus A/a. For a backcross case, the expected distributions of the trait in each of the four marker groups, $U_{m1m2}(x) = U_1(x)$, $U_{M1m2}(x) = U_2(x)$, $U_{m1M2}(x) = U_3(x)$ and $U_{M1M2}(x) = U_4(x)$ can be written as:

$$U_i(x) = \pi_i f_{aa}(x) + (1 - \pi_i) f_{Aa}(x), \quad i = \overline{1,4}, \quad (1)$$

the proportions $\pi_i = \pi_i(r_1, r_2)$ depending on the unknown rates of recombination r_1 and r_2. With no interference,

$$\pi_1 = (1 - r_1)(1 - r_2)/(1 - r);$$

$$\pi_2 = r_1(1 - r_2)/r;$$

$$\pi_3 = 1 - \pi_2; \quad \text{and} \quad \pi_4 = 1 - \pi_1.$$

The specification of the densities $f_{aa}(x)$ and $f_{Aa}(x)$ depends on the assumptions made about the genetic control of the trait. Thus, if one anticipates that no other oligogenes affecting x are segregating or, by contrast, that the number of such loci is not too small, then normal density could be a good approximation, $f_{aa}(x) = N(x_{aa}, \sigma_{aa})$ and $f_{Aa}(x) = N(x_{Aa}, \sigma_{Aa})$, where x_{aa} and x_{Aa} are the expected mean values of x in groups aa and Aa, and σ_{aa} and σ_{Aa} are the standard deviations.

(ii) *Lod-score test and parameter estimation*

Assuming that locus A/a belongs to the interval $M_1/m_1 - M_2/m_2$, the log-likelihood for a sample of

measurements x_k in marker groups with sizes n_i ($i = \overline{1,4}$) can be written as:

$$\ln L(\Theta_{s1}) = \sum_{i=1}^{4} \sum_{k=1}^{n_i} \ln U_i(x_k)$$

$$= \sum_{i=1}^{4} \sum_{k=1}^{n_i} \ln [\pi_i f_{aa}(x_k) + (1 - \pi_i) f_{Aa}(x_k)],$$

where Θ_{s1} is the vector of s_1 unknown parameters, specifying recombination rates and distribution of the trait x in the QTL groups aa and Aa. The assumption of no effect of genes from the interval $M_1/m_1 - M_2/m_2$ on x can formally be presented by another set of parameters, $\Theta = \Theta_{s0}$ (the null hypothesis $\{H_0 : \Theta = \Theta_{s0}\}$) as contrasting to the alternative one $\{H_1 : \Theta = \Theta_{s1}\}$). According to the likelihood ratio test approach (Wilks, 1962), if H_0 is true, the statistic

$$\chi^2 = 2 \ln [\max_{\Theta_{s1} \in S_1} L(\Theta_{s1}) / \max_{\Theta_{s0} \in S_0} L(\Theta_{s0})] \quad (2)$$

is distributed asymptotically as chi-square with $s_1 - s_0$ degrees of freedom, where S_0 and S_1 are the parameter spaces corresponding to H_0 and H_1, respectively. The null hypothesis H_0 is rejected if χ^2 exceeds some critical value, corresponding to a preset level of significance. Then, the numerical values of the parameters that provide maximum to $L(\Theta_{s1})$ could be considered as maximum likelihood estimates characterizing our QT locus A/a (Knott & Haley, 1992). However, the suitability of the chi-square approximation for the above test statistic remains an open question (e.g. Churchill & Doerge, 1994). In the multi-interval formulation of the QTL mapping problem, the exact asymptotic distribution of the log-likelihood ratio is also unknown (Lander & Botstein, 1989; Zeng, 1994). In such a case, Monte Carlo simulation for a given H_0 allows us to obtain an empirical critical value of the test statistics (2) (Knott & Haley, 1992; Ooijen, 1992; Zeng, 1994).

3. Dependence of the likelihood ratio on variance effect of the QT-locus

Let us estimate the joint effect of non-equal variances and averages in the putative QT locus groups of a backcross on the resolution of the lod-score test. For the sake of simplicity we consider the case of a very small marker interval, so that an approximation to no-recombination between marker loci and the putative QTL is suitable ($r = 0$). We assume also that the estimates x_{aa}, σ_{aa}, x_{Aa}, and σ_{Aa} are replaced by their expectations. For such a simplified situation the likelihood ratio will be as follows:

$$\frac{\prod_{i=1}^{n} \left\{ \dfrac{1}{\sqrt{2\pi}\sigma_{aa}} \exp[-(x_{i1} - x_{aa})^2 / 2\sigma_{aa}^2] \right\} \prod_{i=1}^{n} \left\{ \dfrac{1}{\sqrt{2\pi}\sigma_{Aa}} \exp[-(x_{i2} - x_{Aa})^2 / 2\sigma_{Aa}^2] \right\}}{\prod_{\substack{k=1,2; \\ i=1,n}} \{ \exp[-(x_{ik} - 0 \cdot 5(x_{aa} + x_{Aa}))^2 / G] / \sqrt{\pi G} \}} = \mathbf{B/C},$$
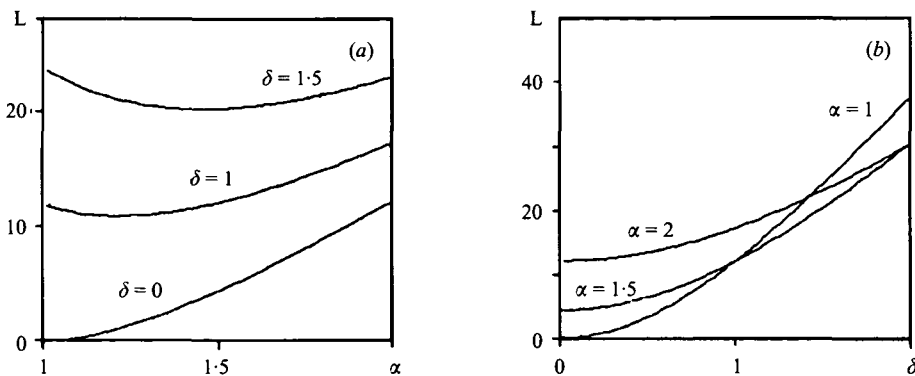
Fig. 1. The expected likelihood ratio (L) of QTL-marker analysis as a function of the variance effect $\alpha = (\sigma_{Aa}/\sigma_{aa})$ and mean value effect $\delta = d/\sigma_{aa}$ of the putative QT locus, assuming very close linkage between marker loci spanning the QTL, and 250 individuals in the sample: (a) dependence of L on α, given fixed value of δ; (b) dependence of L on δ, given fixed value of α.

where $d = x_{Aa} - x_{aa}$, $2n = N$ is the size of the mapping population, so that n is the expected size of the two marker subgroups, and $\mathbf{G} = (\sigma_{aa}^2 + \sigma_{Aa}^2 + 0.5d^2)$ is twice the phenotypic variance of the trait in the population. Let us now calculate the expected mean value of $\log B/C = \overline{\log B} - \overline{\log C} = L$. We will have separately:

$$\overline{\log B} = -\log(2\pi\sigma_{aa}\sigma_{Aa})$$
$$+ \Sigma[-(x_{i1}-x_{aa})^2/2\sigma_{aa}^2] + \Sigma[-(x_{i2}-x_{Aa})^2/2\sigma_{Aa}^2]$$
$$= -n\log(2\pi\sigma_{aa}\sigma_{Aa}) - n$$
$$= -n[1 + \log(2\pi\sigma_{aa}\sigma_{Aa})];$$

$$\overline{\log C} = -n\log\pi G$$
$$- \sum_{k-1,2;i-1,n} [x_{ik} - 0.5(x_{aa}+x_{Aa})]^2/G$$
$$= -n(1 + \log\pi G).$$

Finally,

$$L = \overline{\log(B/C)} = n\log[\tfrac{1}{2}(\alpha + \alpha^{-1}) + \tfrac{1}{4}\delta^2/\alpha] = L(\alpha, \delta),$$

where $\delta = d/\sigma_{aa}$ and $\alpha = \sigma_{Aa}/\sigma_{aa}$. This measure shows a monotonic increase with distance between the trait means of the QTL groups (δ) and a non-monotonic (in general) dependence on the variance ratio of the groups (α) (Fig. 1).

4. Simulation study

The above analysis, conducted for the case of a very close linkage between the marker loci spanning the putative QTL, revealed a surprising dependence of the test statistics (lod-score) on trait variance ratio. Namely, given any fixed value of the variance in one of the QTL groups, the test statistic may grow with increasing values of the variance in the other groups. It is of primary interest to see whether this effect holds also under looser linkage between the flanking markers. Such an examination was done by Monte Carlo simulations.

(i) Generating the data

Monte Carlo simulations were used to produce the 'observations'. For each situation studied, 200 repeated backcross populations were generated. Normal distribution was used for the trait groups aa and Aa. For comparative analysis of different situations we used, where possible, one and the same set of data. The composition of the marker groups (mixtures U_i, $i = \overline{1,4}$) were modelled as binomial distributions with proportions $\pi_i(r_1, r_2)$ and $1 - \pi_i(r_1, r_2)$. Clearly, this restriction is not principal. For most of the experiments, the parameter values used were in the range: $0.5 \leqslant d_x = x_{Aa} - x_{aa} \leqslant 0.75$, $\sigma_{aa} = 1$, $1 \leqslant \sigma_{Aa} \leqslant 2$, sample size $N = 250$; the length of the marker interval was 20 cM with the QT locus in the middle. No interference was assumed in our model (and Haldane's mapping function is suitable).

(ii) Obtaining numerical solutions

The target of this work was to estimate the gain in the test power and estimation accuracy when the variance effect of the putative QTL is taken into account. Therefore, we do not dwell in this study on problems of numerical procedures of multidimensional optimization. The main objective here was to check how the difference of the QTL groups with respect to trait variance affects the detection power of the likelihood ratio test and closeness of the optimal points (representing the estimate of the parameter vector Θ) to the true parameter set. For this specific goal, we do not have to search the solution starting from arbitrary points. The simplest way to obtain the necessary

Table 1. *Resolution power of the LOD score test and precision of the parameter estimates in internal mapping of a QTL with a variance effect (results of Monte Carlo simulations)*

		Situation								
		$\sigma_{aa}=1$ $\sigma_{Aa}=1$			$\sigma_{aa}=1$ $\sigma_{Aa}=\sqrt{2}$			$\sigma_{aa}=1$ $\sigma_{Aa}=2$		
Assumption	d	0·25	0·50	0·75	0·25	0·50	0·75	0·25	0·50	0·75
$\sigma_{aa}=\sigma_{Aa}$	LOD	0·98	2·90	5·90	0·76	2·05	4·11	0·59	1·37	2·64
	β	4	41	91	2	19	72	1	8	34
	r	0·088	0·092	0·093	0·085	0·090	0·092	0·087	0·088	0·090
	s_r	0·069	0·053	0·041	0·071	0·059	0·048	0·073	0·064	0·056
	v_r	78	58	44	84	66	52	84	73	62
	d	0·243	0·483	0·729	0·244	0·483	0·724	0·242	0·485	0·722
	s_d	0·140	0·136	0·138	0·178	0·167	0·169	0·239	0·219	0·217
	v_d	58	28	19	73	35	23	99	45	30
	σ	0·994	0·995	0·995	1·221	1·222	1·223	1·580	1·581	1·582
	s_σ	0·045	0·044	0·044	0·058	0·057	0·057	0·085	0·085	0·085
	v_σ	4·5	4·4	4·4	4·7	4·7	4·7	5·4	5·4	5·4
$\sigma_{aa}\neq\sigma_{Aa}$	LOD	1·23	3·13	6·13	3·92	5·16	7·21	11·32	12·04	13·29
	LOD_e	0·84	3·29	7·14	3·76	5·41	8·06	12·45	13·45	15·08
	β	2	38	88	52	77	96	100	100	100
	β^*	2	36	88	50	77	97	100	100	100
	r	0·091	0·092	0·093	0·095	0·095	0·094	0·094	0·094	0·093
	s_r	0·071	0·053	0·042	0·055	0·048	0·041	0·035	0·034	0·032
	v_r	78	58	45	58	51	44	37	36	34
	d	0·237	0·481	0·728	0·222	0·473	0·724	0·214	0·465	0·717
	s_d	0·143	0·137	0·139	0·179	0·177	0·174	0·220	0·220	0·220
	v_d	60	28	19	81	37	24	103	47	31
	σ_1	0·982	0·984	0·985	0·985	0·987	0·988	0·984	0·986	0·987
	$s_{\sigma 1}$	0·069	0·068	0·067	0·067	0·067	0·068	0·069	0·070	0·071
	$v_{\sigma 1}$	7·0	6·9	6·8	6·8	6·8	6·9	7·0	7·1	7·2
	σ_2	1·004	1·003	1·003	1·417	1·416	1·416	2·005	2·005	2·004
	$s_{\sigma 2}$	0·065	0·064	0·064	0·089	0·089	0·089	0·122	0·123	0·123
	$v_{\sigma 2}$	6·5	6·4	6·4	6·3	6·3	6·3	6·1	6·1	6·1

In simulation of backcross progeny (200 runs with 250 plants in each) the QT locus was assumed to reside in the middle of the marked interval (total length 20 cM); $d = x_{Aa} - x_{aa} \in [0, 1]$ is the QTL effect on mean value of the trait, $\alpha = \sigma_{Aa}/\sigma_{aa} \in [1, 2]$ is variance effect. LOD is the mean value of the maximum lod-score in the interval averaged over the runs, LOD_e is the predicted value of LOD for the case of very close linkage; s_Θ and $v_\Theta = (s_\Theta/\Theta)\,100\%$ are the standard deviation and coefficient of variation of the estimated parameter Θ ($\Theta = r, d, \sigma_{aa}$ or σ_{Aa}); $\beta(\%)$ is the power of the QTL detection test at the 0·1 % level of significance based on the asymptotic chi square distribution, while $\beta^*(\%)$ is the corresponding estimation based on Monte Carlo simulation of the distribution of the test statistic under H_0.

estimates is to use as an initial point in the optimization procedure the parameter values equal to the 'true' ones of the considered sample (e.g. Titterington *et al.* 1985). Based on numerical analysis, we found that for the studied combinations of the model parameters this initial point lies in the domain of the attraction of the global maximum of the ML-functional. Of course, it could not be true for small sample sizes (Titterington *et al.* 1985). In case of simulation of the test statistic distribution under H_0 (see below) different starting points were used to ensure convergence to the global maximum. As tools for local optimization we employed different modifications of the gradient and Newton methods.

(iii) *Estimation of the power of the test*

In order to estimate approximately the power of the log-likelihood ratio test we used the critical level of the test statistics (2) $\chi^2 = \chi^2_{critical}$ based on the asymptotic distribution (chi square with D.F. $= s_1 - s_0$), where s_1 is the number of parameters in the model corresponding to H_1 and s_0 the number of parameters for the model H_0 (Wilks, 1962; Knott & Haley, 1992). In order to check whether the real distribution of the test statistic under H_0 is approximated by the asymptotic one, Monte Carlo simulations were conducted allowing us to obtain an empirical significance threshold. The proportion of cases where the QT locus was revealed

when it really exists was measured using both the asymptotic and empirical critical values.

(iv) *Estimating the accuracy of obtained solutions*

Usually, standard errors of the estimates are employed as a means for accuracy comparison of the estimation procedures. However, in addition to random fluctuations around the mean, another possible source of disturbances, the bias of the estimates, should also be taken into account. Thus, one should simultaneously take care of the estimation variance and estimate bias. These two measures are reflected in the average and standard deviation of the estimates.

(v) *Resolution power as a function of the variance effect of QTL*

For the backcross case, we have simulated and analysed some situations when a QT locus (A/a) residing in a marked interval $(M_1/m_1 - M_2/m_2)$ affects both mean and variance of the QT in question. In order to show the advantage of including the variance effect into the QTL detection and estimation model, we compared the resulting characteristics (test power and precision of estimates) with those obtained under assumption of equal variances and with the results when the putative QTL does not affect the trait variance (Table 1).

Consider the situation when equal variance assumption $(\sigma^2_{aa} = \sigma^2_{Aa})$ is made. Then, as one could expect, an increase in one of the variances (say, σ^2_{Aa}) leads to reduction in the power of the LOD test and in estimation precision of the genetic parameters. But much less expected are the results obtained for the same variants when the fact that $\sigma^2_{aa} \neq \sigma^2_{Aa}$ is taken into account in the model. In such circumstances a substantial increase in resolution can be achieved, as compared both with the last result and with the situation when A/a does not affect the trait variance (compare the corresponding columns in Table 1).

It is worthwhile to recall the foregoing conclusion for the limiting case of no recombination between the putative QT locus and flanking markers (see Fig. 1). Here the prediction of an increase in the resolution power of the LOD test for QTL detection due to the variance effect of the QT locus in question is verified by simulations for the more general case of non-zero recombination. It is interesting that the expected values of LOD obtained for $r = 0$ (see the above expression $L(\alpha, \delta)$) are, in fact, a good representation of the corresponding situations with non-zero recombination. One could easily come to this conclusion by comparison of $\text{LOD}_e = L(\alpha, \delta)$ values in the last column of Table 1 with the LODs obtained from Monte Carlo simulations when the variance effect $(\sigma^2_{aa} \neq \sigma^2_{Aa})$ is included into the model. Thus, besides biological importance (see Introduction), variance effect may help to improve the sensitivity and precision

of the interval analysis, of course with an appropriate model. However, one may ask whether strong enough variance effects are possible in practice, to allow for their detection and a rise in the resolution power in spite of increased number of parameters in the model specifying the variance effect. An example of application of the proposed procedure to real data provided below shows that both questions can be answered positively.

5. An example of application

The following analysis of real data obtained in a study of sweet corn quantitative traits illustrates the points. A part of that study was devoted to reveal QTLs responsible for differences in sucrose content (Tadmor *et al.* 1994). Only an isolated example of Tadmor *et al.* (1994) results are presented here; the full analysis will be published elsewhere.

(i) *Experimental design and the estimation procedure*

Plants of an F_2 population resulting from a cross between two sweet corn inbreds (IL731a and W6786) were scored for 93 markers (mainly RFLPs) from all 10 chromosomes. In order to obtain reliable estimates of the quantitative traits in question, each of the 214 F_2 genotypes was represented by mean trait values of its F_3 family (no less than 30 plants per family). However, beside an increased resolution, such a design complicates to some degree the estimation procedure.

Indeed, let us consider some QT locus A/B from an interval $M_1/m_1 - M_2/m_2$. If one takes from F_2 a homozygote for this locus, either AA or BB, then for this locus the genotype of the entire F_3 family will be the same, with the expected variance σ^2_{AA}/p or σ^2_{BB}/p, respectively, where p is the family size. But for a heterozygote AB, the selfed F_3 progeny will involve a mixture of three classes. Assuming no interaction between A/B and loci from other chromosomes, the expected mean value of the QT in question is $x_{F3(F2=AB)} = 0.25x_{AA} + 0.5x_{AB} + 0.25x_{BB}$. The expected variance can then be presented as:

$$\sigma^2_{F3(F2=AB)} = \{[\sigma^2_{AA} + 2\sigma^2_{AB} + \sigma^2_{BB}]/2$$
$$+ [4(x_{AB}-x_{AA})^2 + 4(x_{AB}-x_{BB})^2 + (x_{BB}-x_{AA})^2]/32\}/p.$$

Thus, each of the nine marker groups for the interval $M_1/m_1 - M_2/m_2$ of our mapping F_2 population is a mixture of three components, $f_{AA}(x)$, $f_{AB}(x)$, and $f_{BB}(x)$, with means x_{AA}, $x_{F3(F2=AB)} = X(x_{AA}, x_{AB}, x_{BB})$, and x_{BB}, respectively, and variances σ^2_{AA}/p, $\sigma^2_{F3}(F2 = AB)$, σ^2_{BB}/p:

$$U_i(x) = \pi_{1i}f_{AA}(x) + \pi_{2i}f_{AB}(x)$$
$$+ \pi_{2i}f_{BB}(x), \quad i = \overline{1,9}. \quad (1\,a)$$

Here the proportions π_{ji} of the components depend on the position of A/B within the interval $M_1/m_1 - M_2/m_2$, and the chosen mapping function. Then

for any set of parameters Θ_{s1} specifying the recombination rates and distribution of the trait x in the QTL groups AA, AB, and BB, one can define the log-likelihood as

$$\log L(\Theta_{s1}) = \sum_{i=1}^{9} \sum_{k=1}^{n_i} \ln U_i(x_k),$$

and employ the log-likelihood ratio test (see eqn 2).

The above procedure was applied to our data on sucrose content in dry kernels (40 d after pollination). The importance of the variance effect is demonstrated in the behaviour of the respective statistics obtained for the interval from *umc36a* to *sel* which was scored phenotypically (La Bonte & Juvik, 1990; Tadmor *et al.* 1994). This interval was shown to explain a major portion of sucrose content variation in dry seeds and was suggested to include the *se* gene itself (Tadmor *et al.* 1994). Our numerical analysis showed that along the considered portion of the chromosome 2 the $\log L$ is a unimodal function of Θ. The extremum can easily be obtained by an algorithm which involves scanning for r within the *umc36a-sel* interval with gradient optimization for other parameters of the set Θ.

(ii) *Results*

We considered 12 versions of the QT-locus dominance effect specification derived from three possibilities for inter-allele relationships with respect to trait mean value and four possibilities for the variance effect (Table 2). In each of these formulations the H_0 hypothesis assumes 'no effect on the trait mean value and trait variance'. Several conclusions may be derived from the presented tests and estimates. Let us denote by A the allele of the putative QTL from the *umc36a-sel* interval corresponding to the lower trait value, and by B the enhancing allele.

Consider first the general model, with no constraints either for trait mean effect or for the variance effect (model 1, or M_1, Table 2). The obtained estimates allow one to assume that the trait value of heterozygotes lies between the homozygotes with a clear tendency of A to dominate. For the variance effect, the QTL group homozygote for the enhancing allele (BB) has higher variance than the alternative homozygote group AA, while the heterozygote group AB has the lowest variance. The last fact may reflect higher homeostatic ability of heterozygotes (Lerner, 1954). Comparison of the three cases presented in the first column of Table 2 allows one to conclude that with respect to the effect on the trait mean value the attenuating allele A is dominant over the enhancer B (the respective model 2 fits the data as well as the full model 1). The opposite assumption of dominance of B is rejected at high level of significance (χ^2 for M_1 as against M_2 is $81\cdot70 - 53\cdot93 = 27\cdot77$, D.F. $= 1$).

The next step was to consider the variance effect. As before, we can start with the full model for mean value

effect, i.e. to contrast different models of the first row. Comparison of M_1 and M_4 clearly demonstrates the importance of the variance effect: $\chi^2 = 81\cdot70 - 42\cdot83 = 38\cdot87$ (D.F. $= 2$). Thus, beside losing biologically important information, neglecting this effect results in a very serious reduction in the resolution power. In addition, some bias in the estimation of the QTL position within the interval is also possible. Further, comparison of the full model M_1 with M_7 and M_{10} leads to the conclusion that allele A tends to be dominant over B with respect to the variance effect. Taking into account also the above result concerning the mean trait effect, we can expect that the most economic description of the putative QT locus will be a model with allele A dominant for both trait mean and variance effects. And indeed, the results presented in Table 2 and Fig. 2 confirm this expectation (the model M_8 provides nearly the same LOD value as the full model M_1 employing a lower number of parameters).

Notably, the *se* phenotype was originally described as dependent on a recessive gene (Ferguson *et al.* 1978). In our LOD score based identification of the putative QTL (A/B) the interval *umc36a-sel* was the most distal one available on chromosome 2. The estimates obtained indicate that the revealed locus A/B lies within this interval close to *sel*. To come to a final decision whether or nor A/B indeed coincides with *sel*, one more interval, distal to *sel*, is required.

6. Discussion

Equal variances of the trait in question within the QT locus groups are a usual simplifying assumption in marker analysis of quantitative traits, both for single marker analysis and interval mapping. The dependence of resolution power of marker analysis of QTs on the assumption of 'no effect of the putative QT locus on trait variance' was considered recently for single marker cases (Weller & Wyler, 1992; Korol *et al.* 1994; Ronin *et al.* 1995). The conclusions we drew for such situations earlier, and confirmed here for interval mapping, seem to contradict the intuitive expectations. Indeed, consider a simple example of a QTL segregating in a backcross progeny. One can compare two situations, $\sigma_{Aa}^2 = \sigma_{aa}^2 = 1$ and $\sigma_{Aa}^2 \neq \sigma_{aa}^2 = 1$. Clearly, if the last inequality is $\sigma_{Aa}^2 < \sigma_{aa}^2$, and the inequality of the variances is allowed by the model, then an increase in resolution is expected (compared to the situation with $\sigma_{Aa}^2 = \sigma_{aa}^2 = 1$) and this definitely will be the case. But what will be the consequences of an opposite situation, $\sigma_{Aa}^2 > \sigma_{aa}^2 = 1$, if the variance effect is included in the model? Clearly, the situations $\sigma_{Aa}^2 < \sigma_{aa}^2 = 1$ and $\sigma_{Aa}^2 > \sigma_{aa}^2 = 1$ are non-equivalent: in the first one the relative distance of the means of the QTL groups is larger and in the second one smaller than in case of $\sigma_{Aa}^2 = \sigma_{aa}^2 = 1$. We found that if the effect of the putative QTL on trait variance is large, then allowing for $\sigma_{Aa}^2 \neq \sigma_{aa}^2$ in the model increases the

Table 2. *LOD score analysis of 12 models specifying mode of action of a putative QT locus (interval umc36a-sel of the chromosome 2) affecting sucrose content in dry kernels of sweet corn. The most suitable marginal models (M_2 and M_7) as well as the best final model (M_8) are boxed.*

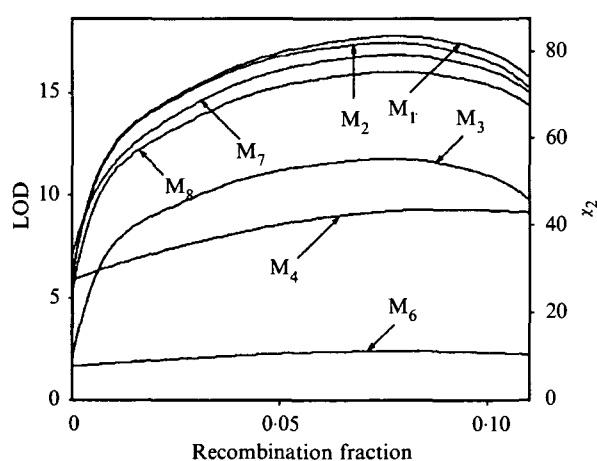| Assumption ⟶		$\sigma_{AA} \neq \sigma_{AB} \neq \sigma_{BB}$		$\sigma_{AA} = \sigma_{AB} = \sigma_{BB}$		$\sigma_{AA} = \sigma_{AB} \neq \sigma_{BB}$		$\sigma_{AA} \neq \sigma_{AB} = \sigma_{BB}$	
		M_1		M_4		M_7		M_{10}	
		Mean	σ	Mean	σ	Mean	σ	Mean	σ
	AA	35·7	10·8	35·4	11·8	34·9	8·3	35·0	8·3
	AB	38·4	7·4	39·6	11·8	38·7	8·3	39·9	12·7
	BB	57·4	14·4	55·1	11·8	57·4	17·5	54·7	12·7
No constraints	r	0·080		0·090		0·077		0·088	
	LOD	17·74		9·30		16·81		10·61	
	χ^2(D.F.)	81·7(5)		42·8(3)		77·4(4)		48·8(4)	
		M_2		M_5		M_8		M_{11}	
		Mean	σ	Mean	σ	Mean	σ	Mean	σ
	AA	38·0	11·4	38·4	11·8	37·6	8·5	37·9	8·9
	AB	38·0	7·3	38·4	11·8	37·6	8·5	37·9	12·6
	BB	57·3	17·4	55·6	11·8	57·4	17·4	55·1	12·6
Dominant A	r	0·076		0·086		0·078		0·083	
	LOD	17·41		8·69		15·98		9·38	
	χ^2(D.F.)	80·2(4)		40·0(2)		73·6(3)		43·2(3)	
		M_3		M_6		M_9		M_{12}	
		Mean	σ	Mean	σ	Mean	σ	Mean	σ
	AA	35·9	11·1	35·1	13·2	34·8	8·4	35·0	8·1
	AB	39·1	7·5	43·7	13·2	39·7	8·4	43·8	14·3
	BB	39·1	25·2	43·7	13·2	39·7	24·9	43·8	14·3
Dominant B	r	0·076		0·078		0·075		0·074	
	LOD	11·71		2·37		11·02		4·90	
	χ^2(D.F.)	53·9(4)		10·9(3)		50·8(3)		22·6(3)	

Fig. 2. Comparison of several models specifying the mode of action of a QTL affecting sucrose content in dry kernels. The log-likelihood test and the estimation procedure were applied to the most distal available interval (umc36a-*sel*) of the chromosome 2. For the description of the models see Table 2.

resolution also in the case $\sigma_{Aa}^2 > \sigma_{aa}^2 = 1$, as compared to $\sigma_{Aa}^2 = \sigma_{aa}^2 = 1$.

These results are in agreement with our previously suggested explanation of an analogous effect of non-equal variances $\sigma_{Aa}^2 \neq \sigma_{aa}^2$ in single marker analysis (Korol *et al.* 1994). It employs a notion of 'discrepancy of the QT locus group distributions', $D(f_{aa}(x), f_{Aa}(x))$, as a function of $d = x_{Aa} - x_{aa}$ and $\sigma_{Aa}^2/\sigma_{aa}^2$. We found that both $D(f_{aa}(x), f_{Aa}(x))$ and the resolution power may grow not only with increasing $d = x_{Aa} - x_{aa}$ (which is quite expected), but also with increasing $\sigma_{Aa}^2/\sigma_{aa}^2$, provided d is relatively small. As one could easily see from Fig. 1B, for any large enough $\alpha = \sigma_{Aa}/\sigma_{aa} > 1$, a range of values $\delta = d/\sigma_{aa}$ could be found, so that within this range the power of the test is higher than for the corresponding case with $\alpha = 1$ (in spite of the increased number of parameters). The last fact may help to understand better the seemingly paradoxical behaviour of the resolution power as a function of $\alpha = \sigma_{Aa}^2/\sigma_{aa}^2$. The putative QT locus could affect either the mean value of the trait x or its variance, or both. Thus, even if $d \approx 0$, one could try to identify the effect of $a- > A$ substitution on the trait variance. It is natural to expect that with small d the resolution should grow with increasing effect of the locus A/a on the trait variance, i.e. with increased deviation of α from unity, e.g. due to increase of σ_{Aa} given fixed σ_{aa}.

Using sweet corn data on sucrose content in dry

kernels, we demonstrated the possibility of extracting important biological information concerning the mode of QTL effect on trait mean and variance. In particular, the conducted log-likelihood tests and obtained parameter estimates show that the allele enhancing the trait is recessive over the opposite allele simultaneously for the mean value and variance. In spite of the necessity to evaluate an increased number of parameters, the more correctly specified model allowed us to achieve a serious increase in the power of the test and better precision of the resulting estimates.

It is likewise important that variance effect of QTLs may be of primary biological interest and the target of mapping efforts (Zhuchenko *et al.* 1979; Korol *et al.* 1981, 1994; Weller & Wyler, 1992). Of special interest may be variance effects of loci involved in control of fitness related quantitative traits in natural populations, like seed dormancy, or flowering time. Clearly, this aspect might be very important in supporting climatic adaptive radiation into increasingly hostile and unpredictive condition.

References

Boehnke, M. & Moll, P. (1989). Identifying pedigrees segregating at a major locus for a quantitative trait: an efficient strategy for linkage analysis. *American Journal of Human Genetics* **44**, 216–224.

Carbonell, E. A., Asins, M. J., Baselga, M., Balansard, E. & Gerig, T. M. (1993). Power studies in the estimation of genetic parameters and the localization of quantitative trait loci for backcross and doubled haploid populations. *Theoretical and Applied Genetics* **86**, 411–416.

Carey, G. & Williamson, J. (1991). Linkage analysis of quantitative traits: increased power by using selected samples. *American Journal of Human Genetics* **49**, 786–796.

Churchill, G. A. & Doegre, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963–971.

Darvasi, A. & Soller, M. (1992). Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. *Theoretical and Applied Genetics* **85**, 353–359.

Demenais, E., Lathrop, G. M. & Lalouel, J. M. (1988). Detection of linkage between a quantitative trait and marker locus by the lod scores method: sample size and sampling considerations. *Annals of Human Genetics* **52**, 237–246.

Ferguson, J. E., Rhoads, A. M. & Dickinson, D. B. (1978). The genetics of *sugary enhancer* (*se*) and independent modifier of sweet corn (*su*). *Journal of Heredity* **69**, 377–380.

Jansen, R. C. & Stam, P. (1994). High resolution of

quantitative traits into multiple loci via interval mapping. *Genetics* **136**, 1447–1455.

Knott, S. A. & Haley, C. S. (1992). Aspects of maximum likelihood methods for mapping of quantitative trait loci in line crosses. *Genetical Research* **60**, 139–151.

Korol, A. B., Preygel, I. A. & Preygel, S. I. (1994). *Recombination Variability and Evolution.* London: Chapman & Hall.

Korol, A. B., Ronin, Y. I. & Kirzhner, V. M. (1995). Interval mapping of quantitative trait loci employing correlated trait complexes. *Genetics* **140**, 1137–1147.

Korol, A. B., Zhuchenko, A. A. & Samovol, A. P. (1981). Linkage between quantitative trait loci and marker loci. 3. Biasing of estimates resulting from departures from initial assumptions. *Genetika* (USSR) **17**, 1234–1247.

Kullback, S. (1959). *Information Theory and Statistics.* New York: John Wiley & Sons.

La Bonte, D. R. & Juvik, J. A. (1990). Characterization of *sugary* (*su1*) and *sugary enhancer* (*se*) kernels in segregating maize (*Zea mays* L.) populations. *Journal of the American Society for Horticultural Science* **114**, 153–157.

Lander, E. S. & Botstein, D. (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199.

Lebowitz, B. J., Soller, M. & Beckmann, J. S. (1987). Trait-based analyses for the detection of linkage between marker loci and quantitative trait loci in crosses between inbred lines. *Theoretical and Applied Genetics* **73**, 556–562.

Lerner, I. M. (1954). *Genetic Homeostasis.* Edinburgh and London: Oliver & Boyd.

Motro, U. & Soller, M. (1993). Sequential sampling in determining linkage between marker loci and quantitative trait loci. *Theoretical and Applied Genetics* **85**, 658–664.

Ronin, Y. I., Kirzhner, V. M. & Korol, A. B. (1995). Linkage between loci of quantitative traits and marker loci: multi-trait analysis with a single marker. *Theoretical and Applied Genetics* **90**, 776–786.

Soller, M. & Beckmann, J. S. (1990). Marker-based mapping of quantitative trait loci using replicated progeny. *Theoretical and Applied Genetics* **80**, 205–208.

Tadmor, Y., Azanza, F., Han, T., Rocheford, T. R. & Juvik, J. A. (1995). RFLP mapping of the *sugary enhancer 1* gene in sweet corn. *Theoretical and Applied Genetics* **88** (in the Press).

Titterington, D. M., Smith, A. F. & Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions.* Wiley, Chichester, UK.

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

Weller, J. I. (1987). Mapping and analysis of quantitative trait loci in *Lycopersicon* (tomato) with the aid of genetic markers using approximate maximum likelihood methods. *Heredity* **59**, 413–421.

Weller, J. I. & Wyler, A. (1992). Power of different sampling strategies to detect quantitative trait loci variance effect. *Theoretical and Applied Genetics* **83**, 582–588.

Wilks, S. S. (1962). *Mathematical Statistics.* New York: Wiley.

Zeng, Z.-B. (1994). Precise mapping of quantitative trait loci. *Genetics* **136**, 1457–1468.

Zhuchenko, A. A., Samovol, A. P., Korol, A. B. & Andryuschenko, V. K. (1979). Linkage between quantitative trait loci and marker loci. 2. Effects of three tomato chromosomes on the variation for five quantitative characters in backcross generations. *Genetika* (USSR) **15**, 672–683.