Interval Mapping of Quantitative Trait Loci Employing Correlated Trait Complexes

Abraham B. Korol, Yefim I. Ronin and Valery M. Kirzhner

Institute of Evolution, University of Haifa, Haifa 31905, Israel

Manuscript received September 6, 1994
Accepted for publication April 11, 1995

ABSTRACT

An approach to increase the resolution power of interval mapping of quantitative trait (QT) loci is proposed, based on analysis of correlated trait complexes. For a given set of QTs, the broad sense heritability attributed to a QT locus (QTL) (say, A/a) is an increasing function of the number of traits. Thus, for some traits x and y, $H^2_{xy}(A/a) \geq H^2_x(A/a)$. The last inequality holds even if y does not depend on A/a at all, but x and y are correlated within the groups AA, Aa and aa due to nongenetic factors and segregation of genes from other chromosomes. A simple relationship connects H^2 (both in single trait and two-trait analysis) with the expected LOD value, ELOD $= -{}^1/_2 N \log(1 - H^2)$. Thus, situations could exist that from the inequality $H^2_{xy}(A/a) \geq H^2_x(A/a)$ a higher resolution is provided by the two-trait analysis as compared to the single-trait analysis, in spite of the increased number of parameters. Employing LOD-score procedure to simulated backcross data, we showed that the resolution power of the QTL mapping model can be elevated if correlation between QTs is taken into account. The method allows us to test numerous biologically important hypotheses concerning manifold effects of genomic segments on the defined trait complex (means, variances and correlations).

THE resolution of marker analysis of quantitative trait variation is a major factor affecting practical applications of quantitative trait locus (QTL) mapping. A detailed discussion of the issues concerning the power of tests for detecting linkage can be found in many publications (*e.g.*, SOLLER and GENIZI 1978; DEMENAIS *et al.* 1988; LANDER and BOTSTEIN 1989; SOLLER and BECKMANN 1990; WELLER and WYLER 1992; CARBONELL *et al.* 1993). The precision of the parameter estimation depends on the effect of the QT locus 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 QT locus groups [$f_{aa}(x)$ and $f_{Aa}(x)$ for a backcross], the better is the expected resolution. Several procedures have been proposed to improve the precision of mapping, including the multimarker (interval mapping) analysis (LANDER and BOTSTEIN 1989; KNOTT and HALEY 1992), selective sampling (LEBOWITZ *et al.*, 1987; CAREY and WILLIAMSON 1991; DARVASI and SOLLER 1992), replicated progeny testing (SOLLER and BECKMANN 1990), and sequential experimentation (BOEHNKE and MOLL 1989; MOTRO and SOLLER 1993).

Recently, much attention has been paid to improve the efficiency of marker analysis of QTL by taking into account the dependence of the quantitative trait on many QT loci (genomic segments) (JANSEN and STAM 1994; ZENG 1994). A situation when one QT locus (or a chromosome segment defined by a pair of flanking markers) affects simultaneously several QTs could also be considered. Such analysis may be of major importance in formulating marker-assisted breeding strategies, dissecting heterosis as a multilocus multitrait phenomenon, developing optimized programs for evaluation and bioconservation of genetic resources, revealing genetic architecture of fitness systems in natural populations, etc. Multivariate approach in segregation and linkage analysis of quantitative traits was recently used in human genetics (*e.g.*, AMES and LAING 1993) but has not yet been applied for QTL mapping.

In mapping QT loci, the experimental design usually includes simultaneous measurements of many QTs and subsequent treatment of these individual traits. Results obtained in many recent studies showed that some genomic segments indeed affect several traits (*e.g.*, EDWARDS *et al.* 1992; DOEBLEY and STEC 1993). Multimaker analysis allows multiple effects of any segment being estimated, if measurements of many traits are available. Nevertheless, in many current QTL mapping methods each trait is analyzed separately (but see JIANG and ZENG 1995). As will be shown in this paper, a substantial amount of genetic transmission information available in the data may be lost by this approach. Consequently, an increase in resolution power of QTL-marker linkage analysis can be achieved by accounting of correlations between the QTs, *i.e.*, when one considers joint distributions of several traits in the QT locus groups [$f_{aa}(x, y)$ and $f_{Aa}(x, y)$ rather than the marginal distributions $f_{aa}(x)$ and $f_{Aa}(x)$, or $f_{aa}(y)$ and $f_{Aa}(y)$]. Earlier we showed the increased efficiency of the

Corresponding author: A. B. Korol, Institute of Evolution, University of Haifa, Mount Carmel, Haifa 31905, Israel.
E-mail: rabi307@uvm.haifa.ac.il

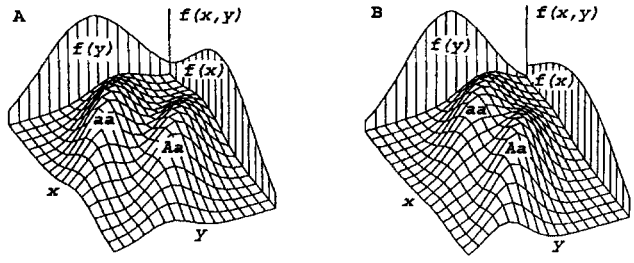

FIGURE 1.—Joint distribution $f(x, y)$ of two correlated traits, x and y, in backcross population. Even with the clear-cut bimodality of $f(x, y)$ [when the components $f_{aa}(x, y)$ and $f_{Aa}(x, y)$ are far enough and the correlation is high] the marginal distributions $f(x)$ and $f(y)$ are unimodal. (A) The QT locus affects both traits, x and y. (B) The QT locus affects only one of the traits, x.

multitrait analysis for different progeny types in estimating linkage between a QT locus and a single marker locus (KOROL *et al.* 1987, 1994; RONIN *et al.* 1995).

The objective of this paper is to demonstrate the advantages of the multitrait analysis within the framework of interval mapping of QTL (see also JIANG and ZENG 1995). Besides of increased power of the statistical test and higher precision of estimation of the genetic parameters, the proposed approach allows for an integral evaluation of the effects of genomic segments on defined trait complexes (mean values, variances and covariances). Because of the internal balance of the organism's systems (SCHMALHAUSEN 1942), such an approach for QTL mapping seems to be much more justified biologically than the usual trait-by-trait analysis.

GENERAL DESCRIPTION OF THE METHOD

The model: Consider a genomic segment carrying a QT locus (with alleles A and a) that affects two quantitative traits, x and y. We will confine our analysis to the backcross situation. Any other type of mapping population may be treated in a similar way. In the illustration presented in Figure 1, the marginal densities are strongly overlapped. Without accounting of any additional information, based only on the observed marginal distributions $f(x)$ and $f(y)$, one would hardly assume that the progeny is polymorphic for an oligogene (A/a). Nevertheless, the presence of an oligogene can easily be seen from the joint distribution $f(x, y)$.

In general, one may assume that the putative QT locus affects not only the mean values of the traits but also the trait variances and covariance. In such a case, the two-dimensional phenotype (x, y) of an arbitrary individual of a backcross progeny can be modeled as follows:

$$x = m_x + 0.5\,d_x g + e_x, \quad y = m_y + 0.5\,d_y g + e_y.$$

x and y are the individual's phenotype scores of the analyzed QTs, m_x and m_y are trait means, d_x and d_y are the effects of substitution $aa \rightarrow Aa$ with respect to mean values of x and y (*i.e.*, $d_x = \mu x_{Aa} - \mu x_{aa}$ and $d_y = \mu y_{Aa} - \mu y_{aa}$), g denotes the genotype at locus A/a ($g = -1$ for aa and 1 for Aa), e_x is a random variable with zero mean and variances σ^2_{1x} and σ^2_{2x} for $g = -1$ and $g = 1$; similarly, e_y is a random variable with mean zero and variances σ^2_{1y} and σ^2_{2y}. The variables e_x and e_y are assumed to be correlated with correlation coefficients R_{1xy} and R_{2xy} for $g = -1$ and $g = 1$, respectively. Correlation between the traits x and y within the QT locus groups may be due to other segregating QTLs or nongenetic correlation. Although we may assume that locus A/a can also affect trait variances and covariance, in this paper however we will deal mainly with the situation of equal variance-covariance matrix, Σ, in the groups aa and Aa, *i.e.*, $\Sigma_1 = \Sigma_2 = \Sigma$.

In one trait analysis, the power and precision of QTL mapping analysis depend on the proportion of the variance caused by segregation of the putative QTL in the total variance of the trait x in the population, H^2_x. Indeed, the expected LOD value for a backcross when the position of the closest marker coincides with that of the QTL is (see LANDER and BOTSTEIN 1989)

$$\text{ELOD} = {}^1/_2 N \log(1 + \sigma^2_{\text{exp}}/\sigma^2_{\text{res}}) = -{}^1/_2 N \log(1 - H^2_x), \quad (1)$$

where σ^2_{exp} is the trait variance associated with the putative QTL ($\sigma^2_{\text{exp}} = {}^1/_4 d^2$) and σ^2_{res} is the residual variance, so that $H^2_x = \sigma^2_{\text{exp}}/(\sigma^2_{\text{exp}} + \sigma^2_{\text{res}})$. As shown in APPENDIX A the same relationship holds for the bivariate case, with a properly defined H^2_{xy}, a two-dimensional analogue of H^2_x,

$$\text{ELOD} = -{}^1/_2 N \log(1 - H^2_{xy}), \quad (2)$$

where

$$H^2_{xy} = 1 - \frac{\sigma^2_x \sigma^2_y (1 - R^2_{xy})}{(\sigma^2_x + d^2_x/4)(\sigma^2_y + d^2_y/4) - \sigma^2_x \sigma^2_y [R_{xy} + d_x d_y/(4\sigma_x \sigma_y)]^2}. \quad (3)$$

The parameter H^2_{xy} obtained in APPENDIX A coincides with a natural bivariate analogue of H^2_x based on the standard multivariate generalization of variance as a measure of variability. Namely, in measuring variability, determinant of the variance-covariance matrix, $|\Sigma|$, is considered as such a generalization (SOKAL 1965). If so, then one may formulate the multivariate (*e.g.*, bivariate) analogue of H^2_x as the proportion of the determinant of the variance-covariance matrix caused by segregation of A/a (assuming $\Sigma_{aa} = \Sigma_{Aa} = \Sigma$), relative to that of the variance-covariance matrix of the total population (Σ_{BC}), *i.e.*, $H^2_{xy} = |\Sigma_{\text{exp}}|/|\Sigma_{BC}| = (|\Sigma_{BC}| - |\Sigma|)/|\Sigma_{BC}| = 1 - |\Sigma|/|\Sigma_{BC}|$. The important fact is that H^2_{xy} defined in this way coincides with H^2_{xy} from (3) derived from the ELOD (see APPENDIX A).

Due to (2), one could expect the resolution power of the analysis increases as H^2_{xy} does. Clearly, statistical aspects concerned with possible increase in the number of parameters to be estimated and changes in the degrees of freedom should be taken into account. Given fixed d_x/σ_x, or $H^2_x = {}^1/_4 d^2_x/({}^1/_4 d^2_x + \sigma^2_x)$, how could we increase the resolution by taking into account other traits? Clearly, if $R_{xy} = 0$, the effect of an additional trait is simply due to the increased euclidian distance between the (two-dimensional) centers of the groups aa and A/a. It is easy to see from (3), that if $d_y \neq 0$, $R_{xy} \neq 0$ and sign $(R_{xy} d_x d_y) < 0$, then $H^2_{xy} \geq H^2_x$ and one could expect a respective increase in resolution. Moreover, the inequality $H^2_{xy} > H^2_x$ holds even if $d_y = 0$, but $R_{xy} \neq 0$, no matter what sign of correlation we have.

It is easy to show (APPENDIX B) that (3) is invariant with respect to nondegenerate linear transformation of the variables. Consequently, one may assume that mapping problems with the same level of H^2_{xy} could be considered as formally equivalent with the only complication due to the possibility of different number of the degrees of freedom. For instance, it may be of interest to compare equivalent (*i.e.*, with the same value of H^2_{xy}, while with different values of other parameters) situations: $d_x \neq 0$, $d_y \neq 0$ and $R_{xy} \neq 0$ (locus A/a affects two correlated traits); $d_x \neq 0$, $d_y = 0$ and $R_{xy} \neq 0$ (locus A/a affects x but not y, while within each of the groups Aa and aa the traits x and y are correlated); $d_x \neq 0$, $d_y \neq 0$ and $R_{xy} = 0$ (locus A/a affects two noncorrelated traits) etc. The proximity of different situations with the same level of H^2_{xy}

will be demonstrated on different examples both for the power of the log-likelihood ratio test and for the precision of parameter estimation.

Mixture-model formulation for multitrait interval analysis: Interval mapping of QTL based on *multitrait analysis* can be conducted employing the same techniques that were developed for the single-trait analysis (*e.g.*, LANDER and BOTSTEIN 1989; JANSENS and STAM 1994). The only difference is the increased number of parameters to be estimated and tested. Our pilot analysis with available experimental data on sweet corn (Y. TADMOR, Y. RONIN, A. KOROL, A. BAR-ZUR, and E. NEVO, unpublished results) shows that in many situations improved resolution makes up the latter drawback.

Assume that a QT locus A/a resides in some 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 $A/a - M_2/m_2$. Different modes of exchange interference in the interval could be considered; we will confine ourselves to the no interference case, so that $r = r_1 + r_2 - 2r_1r_2$, where r is the rate of recombination between M_1/m_1 and M_2/m_2. Based on the marker scores and the measurements of traits of interest (x and y) for individuals from the mapping population, we should test whether or not variation of x and/or y indeed depends on the interval $M_1/m_1 - M_2/m_2$, and, if yes, identify the corresponding locus A/a. For a backcross case, the expected joint distributions of the traits x and y in each of the marker groups, $U_{m1m2}(x, y) = U_1(x, y)$, $U_{M1m2}(x, y) = U_2(x, y)$, $U_{m1M2}(x, y) = U_3(x, y)$ and $U_{M1M2}(x, y) = U_4(x, y)$, can be written as follows:

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

the proportions $\pi_i = \pi_i(r_1, r_2)$ being dependent of the unknown recombination rates r_1 and r_2. With no interference, $\pi_1 = (1 - r_1)(1 - r_2)/(1 - r)$, $\pi_2 = r_1(1 - r_2)/r$, $\pi_3 = 1 - \pi_2$, and $\pi_4 = 1 - \pi_1$. The specification of the densities $f_{aa}(x, y)$ and $f_{Aa}(x, y)$ depends on the assumptions made about the genetic control of the traits. Thus, if one assumes that no other oligogenes affecting x and/or y are segregating, then two-dimensional normal density could be a good approximation,

$$f_{aa}(x, y) = [2\pi\sigma_x\sigma_y(1 - R^2)]^{-1/2} \times \exp\left\{-\frac{1}{2(1 - R^2)}\left[\frac{(x - \mu x_1)^2}{\sigma_x^2} - 2R\frac{(x - \mu x_1)(y - \mu y_1)}{\sigma_x\sigma_y} + \frac{(y - \mu y_1)^2}{\sigma_y^2}\right]\right\},$$

$$f_{Aa}(x, y) = [2\pi\sigma_x\sigma_y(1 - R^2)]^{-1/2} \times \exp\left\{-\frac{1}{2(1 - R^2)}\left[\frac{(x - \mu x_2)^2}{\sigma_x^2} - 2R\frac{(x - \mu x_2)(y - \mu y_2)}{\sigma_x\sigma_y} + \frac{(y - \mu y_2)^2}{\sigma_y^2}\right]\right\},$$

here μx_i and μy_i ($i = 1, 2$) are the expected mean values of x and y in groups aa ($i = 1$) and Aa ($i = 2$), σ_x, σ_y and R are the standard deviations and correlation between x and y. The assumption of normality could also be suitable if several QTLs affecting the traits in question are segregating independently of A/a. To take into account possible deviations from normality caused by a strong gene on another chromosome we can represent each of the densities, $f_{aa}(x, y)$ and $f_{Aa}(x, y)$ as a sum of two bivariate normals (see also KNOTT and HALEY 1992).

LOD-score test and parameter estimation: Assuming that locus A/a is situated in the interval $M_1/m_1 - M_2/m_2$, the log-likelihood for a sample of two-dimensional measurements x_k, y_k in marker groups with sizes N_i ($i = 1, 4$) can be written as

$$\ln L(\theta_{n1}) = \sum_{i=1}^{4}\sum_{k=1}^{N_i} \ln U_i(x_k, y_k) = \sum_{i=1}^{4}\sum_{k=1}^{N_i} \ln [\pi_i f_{aa}(x_k, y_k) + (1 - \pi_i) f_{Aa}(x_k, y_k)].$$

In the general case of $d_x \neq 0$, $d_y \neq 0$, and $\Sigma_{aa} \neq \Sigma_{Aa}$, so that $\theta_{n1} = \{r_1, \mu x_1, \mu x_2, \mu y_1, \mu y_2, \sigma_{x1}, \sigma_{x2}, \sigma_{y1}, \sigma_{y2}, R_1, R_2\}$ is the vector of $n_1 = 11$ unknown parameters, specifying recombination rate and joint distributions of traits x and y in the QTL groups aa and Aa (in case of F_2, θ_{n1} could include up to 16 parameters). The assumption of no effect of genes from the interval $M_1/m_1 - M_2/m_2$ on the traits (x, y) can formally be presented by another set of parameters, $\theta = \theta_{n0} = \{\mu x, \mu y, \sigma_x, \sigma_y, R\}$ (the null hypothesis $\{H_0: \theta = \theta_{n0}\}$ as contrasted to the alternative one $\{H_1: \theta = \theta_{n1}\}$). According to the likelihood ratio test approach (WILKS 1962), if H_0 is true, the statistics

$$X^2 = 2\ln[\max_{\theta_{n1} \in S_1} L(\theta_{n1}) / \max_{\theta_{n0} \in S_0} L(\theta_{n0})] \quad (5)$$

is distributed asymptotically as chi square with $n_1 - n_0$ degrees of freedom, where S_0 and S_1 are the parameter spaces corresponding to H_0 and H_1, respectively (WILKS 1962). Thus, if X^2 exceeds some critical value, corresponding to a preset level of significance α, the null hypothesis can be rejected. In such a case, the numerical values providing maximum to $L(\theta_{n1})$ could be considered as maximum likelihood estimates of the parameters characterizing our QT locus A/a (KNOTT and HALEY 1992). However, in the multi-interval mapping the problem of the exact asymptotic distribution of the test statistics remains unsolved even in the single trait analysis (see ZENG 1994). If so, one could use extensive Monte-Carlo simulations to obtain an empirical critical value of the statistics for each considered situation. Two additional points are worth mentioning here.

1. Introduction of any (additional) parameter(s) specifying the QTL mapping model should be justified statistically by comparison to the corresponding reduced model. This is relevant to any complication of the QTL mapping model including the replacement of single-trait mapping analysis by its multitrait analogue. Thus, if one starts with the full formulation of H_1: $\{\theta = \theta_{n1}, n1 = 11\}$ specifying the putative QTL, then consequently reduced versions of H_1 should be tested, *e.g.*, those with $\Sigma_{aa} = \Sigma_{Aa}$ ($\theta_{n1} = \{r_1, \mu x_1, \mu x_2, \mu y_1, \mu y_2, \sigma_x, \sigma_y, R\}$, $n1 = 8$), etc. Parameters that do not affect the significance level should be removed from the mapping model.
2. An increase in the number of parameters in the two-trait mapping model does not necessary mean a substantial increase in the number of degrees of freedom of the test statistics. That is because the number of parameters specifying the null hypothesis H_0 = (no QTL in the considered interval) also increases.

SIMULATION PROCEDURE AND OPTIMIZATION

Generating the data: Monte-Carlo simulations were used to produce the observations. For each situation studied, 200 repeated mapping populations have been generated using pseudorandom numbers. Bivariate normal distribution was used for the trait groups *aa* and *Aa*. However, our numerical results show (see be-

low) that bivariate QTL mapping analysis is rather robust with respect to deviation from normality assumption caused by independent segregation of other QTLs. For comparative analysis of different methods and situations we used the same set of data. The composition of the marker groups (mixtures U_i, $i = \overline{1, 4}$) were modeled as binomial distributions with expected proportions $\pi_i(r_1, r_2)$ and $1 - \pi_i(r_1, r_2)$. For most of the experiments, parameter values used for simulations were in the following range: $0 \leq d_x = x_{Aa} - x_{aa} \leq 0.6$, $0 \leq d_y = y_{Aa} - y_{aa} \leq 0.6$, $\sigma_{aa} = 1$, $\sigma_{Aa} = 1$, $0 \leq |R| \leq 0.7$, $N = 250$. The length of the marker interval 20 cM with the QT locus in the middle. No interference was assumed in the data presented below (and HALDANE's mapping function is suitable), but this restriction is not essential and the proposed method of analysis can be conducted with any other mode of multiple exchanges.

Obtaining numerical solutions: The target of this work was to compare the above described approach with the single trait analysis, or to put it more exactly, to estimate the gain in accuracy when the correlation between the quantitative traits is taken into account. Therefore, we do not dwell enough in this study on problems of numerical procedures of multi-extremal multidimensional optimization. The main objective here was to check how the correlation between the considered traits 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 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 of the described functionals, 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). As tools for local optimization we employed different modifications of the gradient and Newton methods.

Estimation of the power of the test: To estimate the power of the log-likelihood ratio test, we used the critical level of the test statistics (5) $X^2 = X^2$ critical based on the asymptotic distribution (chi square with $df = n_1 - n_0$). The goodness of fit of the expected distribution was tested by simulations of the mapping population under H_0 (no QTL in the considered interval) using 5000 trials. The proportion of cases where the QT locus was revealed when it really exists was measured for different situations using critical values obtained in these simulations and those from the asymptotic (chi-square) distribution. These two estimates of the power happened to be very close. They were also complemented by an additional indicator (P), the proportion of cases where the highest value of the test statistics was achieved in the proper interval (not in the neighboring ones).

Estimating the accuracy of obtained solutions: Usually, variances or SE 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. Moreover, each of these two components of the deviation of the estimates from the true value could depend on the level of the parameters. To allow for possible differences in biases of the estimates, we employed the absolute error of the estimate, averaged over the repeated experiments:

$$\delta u = \frac{1}{n} \sum_{k=1}^{n} |\hat{u}_k - u|,$$

where $\hat{u}_k$ and u are, respectively, the estimated and expected values of the parameter u (*i.e.*, u can be any component of the vector θ, say r_1, μx_1, d_x, μy_1, d_y, σ_x^2, etc).

SIMULATION RESULTS

For the backcross case, we have simulated and analyzed a number of situations when a QT locus (A/a) residing in a marker interval ($M_1/m_1 - M_2/m_2$) affects simultaneously two correlated QTs, x and y. To show the advantages of the proposed approach, we compare the resulting characteristics (test power and precision of the estimates) with those obtained using single-trait analysis as well as two-trait analysis with no correlation between the QTs. Two versions of H_0 (no QTL in the interval in question) will be considered: no other QTL in the genome, so that the normal distribution could be used, and another QT locus with a strong effect segregates independently of the marked chromosome, preventing the applicability of the normal approximation. In this case, numerous modes of (epistatic) interaction between the two QTLs might, in principle, be considered in the framework of multitrait linkage analysis. However, we will restrict our attention here only to additive cases.

No other QT loci segregating in the mapping population: This means that first version of H_0 is suitable and bivariate normal approximation could be used. In this case, the log-likelihood ratio will be distributed asymptotically as chi square with $df = n_1 - n_0$ (difference between the number of parameters under H_1 and H_0 formulations). Based on simulations of 5000 backcross populations, we found that the distribution of X^2 from (5) when H_0 holds, is indeed close to chi square with the corresponding degree of freedom (not shown). Thus, $df = 11 - 5 = 6$ for full models H_1 and H_0 with $\theta_{n1} = \{r_1, \mu x_1, \mu x_2, \mu y_1, \mu y_2, \sigma_{x1}, \sigma_{x2}, \sigma_{y1}, \sigma_{y2}, R_1, R_2\}$ and $\theta_{n0} = \{\mu x, \mu y, \sigma_x, \sigma_y, R\}$, or $df = 8 - 5 = 3$, for the model with H_1 assuming $\Sigma_{aa} = \Sigma_{Aa}$ (*i.e.*, $\theta_{n1} = \{r_1,$

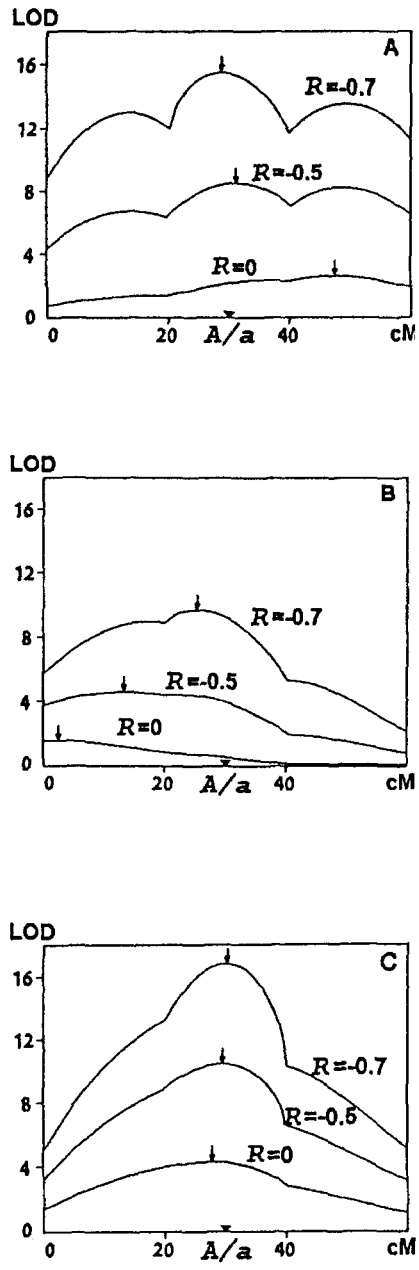

FIGURE 2.—Illustration of the effect of the within-group correlation (R) between the analyzed QTs on the resolution of bivariate interval mapping. Locus A/a has equal effects on both traits, $d_x = d_y = 0.25\sqrt{2}$ (so that its total effect is $d = 0.5$); $\sigma_x = \sigma_y = 1$ in each of the QT locus groups of the backcross. Total population size was $N = 250$. Three independent Monte-Carlo trials (A–C) with the above parameters are presented. Note, that with increasing R the maximum of the LOD also increases and the bias of the estimated position of the putative QTL decreases (for the same sequence of pseudo-random numbers used to produce the sample of size N).

μx_1, μx_2, μy_1, μy_2, σ_x, σ_y, $R\}$) and the same H_0: $\theta_{n0} = \{\mu x, \mu y, \sigma_x, \sigma_y, R\}$.

Figure 2 illustrates the behavior of the LOD score along the interval $M_1/m_1 - M_2/m_2$ and in neighboring intervals. Note, that with increasing $|R_{xy}|$ the maximum of the LOD also increases and the bias of the estimated position of the putative QTL decreases, provided sign $(d_x d_y R_{xy}) < 0$.

Averaged characteristics for a series of simulation experiments described above are presented in Table 1. These results obtained under the assumption of equal variance-covariance matrices in the QT locus groups Aa and aa ($\Sigma_{Aa} = \Sigma_{aa} = \Sigma$) clearly demonstrate the superiority of the bivariate linkage analysis, provided sign $(d_x d_y R_{xy}) < 0$ (Table 1). This is manifested in a considerable increase in the power of the log-likelihood ratio test and a decrease in deviations of the parameter estimates from the expected values. The closer the correlation between the involved traits the higher is H^2_{xy} (and the ELOD value) and the better the resolution, given $d_x d_y \neq 0$ and sign $(d_x d_y R_{xy}) < 0$.

Paradoxical on the first glance is the significant increase in resolution power of mapping of the QT locus affecting trait x, due to the information provided by a correlated trait y, when y does not depend at all on A/a, *i.e.*, when $d_y = 0$ (Table 1). Nevertheless, this is exactly what follows from the comparisons of H^2_{xy} [see (2) and (3)]. Indeed, in spite of the fact that $d_y = 0$, in all cases where the information supplemented by y results in increased resolution, we have an increased level of bivariate broad sense heritability attributed to A/a as compared to the univariate (or no correlation) case, *i.e.*, $H^2_{xy} \geq H^2_x$, or for any sign (R_{xy}) when either $d_x = 0$ or $d_y = 0$. Note, that in cases with no correlation, the power of the test is only 33–37% at the 0.1% significance level usually used when many intervals are treated simultaneously in a multi-chromosomal genome. The difference between the LOD values in the proper and adjacent intervals (ΔLOD) also increases several times when correlation is taken into account. This coincides with a decline in the proportion of cases where the maximum of the LOD function does not lie in the proper interval (equal to $100 - P$) (from 41–44% at $R = 0$ to 13–21% at $R = -0.7$).

Another QT locus is segregating in the population: Denote this additional locus as B/b. If the effects of B/b on x and/or y are strong enough, then the normality assumption is no longer suitable. The H_0 hypothesis "no QT locus in the considered interval" should be formulated, similarly to the univariate case (*e.g.*, KNOTT and HALEY 1992), as if all of the four marker groups have the same distribution $0.5\, f_{bb}(x, y) + 0.5\, f_{Bb}(x, y)$. The components f_{bb} and f_{Bb} of the last mixture may be bivariate normals or any other bivariate densities. We come to the bivariate analogue of the joint interval mapping and segregation analysis: testing for the presence of a QT locus (A/a) in some marked interval and estimating its effects, while allowing another QT locus (B/b) segregating independently of A/a. In fact, the accessibility of many dozens of molecular markers throughout the genome makes it reasonable to include them as cofactors in interval mapping models (JANSENS and STAM 1994; ZENG 1994). We believe that multitrait analogues of these new algorithms will pro-

TABLE 1

Effect of correlation (R) between the QTs on the resolution of interval mapping of the QT locus

Situation								$\beta = \beta(\alpha)$			
d_x	d_y	R	H^2_{xy}	LOD	ΔLOD	df	P	$\alpha = 1$	$\alpha = 0.1$	δr	δd_x
$\sqrt{2}/4$	$\sqrt{2}/4$	0	0.06	3.05 ± 0.10	0.44 ± 0.06	3	59	64[a]	33	45.1 ± 2.2	111.0 ± 6.1
								66[b]	35		
		−0.5	0.14	5.43 ± 0.14	0.87 ± 0.07	3	75	96	80	34.1 ± 1.9	110.9 ± 6.1
								98	82		
		−0.7	0.17	8.44 ± 0.17	1.46 ± 0.09	3	87	100	100	26.8 ± 1.5	109.0 ± 6.1
								100	100		
1/2	0	0	0.06	3.12 ± 0.11	0.48 ± 0.06	2	56	65	37	44.2 ± 2.2	111.3 ± 6.2
								71	37		
		−0.5	0.08	3.87 ± 0.12	0.60 ± 0.06	2	66	79	55	40.8 ± 2.1	110.3 ± 6.1
								82	56		
		−0.7	0.11	5.66 ± 0.15	0.98 ± 0.08	2	79	91	73	33.6 ± 1.8	114.0 ± 6.3
								97	85		

In simulated 200 replicates of backcross progeny (250 individuals in each), A/a locus resides in the middle of a marked interval (length 20 cM) with the effect on the traits $d = (d_x^2 + d_y^2)^{1/2} = 0.5$ or $d = d_x = 0.5$ ($d_y = 0$). In the first case $H_x^2 \approx 0.03$ and in the second one $H_x^2 \approx 0.06$, δr and δd_x are the mean absolute errors (multiplied by 1000) of the corresponding parameter estimates, LOD is the mean value of the maximum lod-score in the interval, ΔLOD is the mean excess of the maximum LOD in the true interval over that in the neighboring ones, $\beta(\%)$ is the power of the test at the significance level $\alpha(\%)$, and $P(\%)$ is the proportion of cases where the maximum of LOD-score resides in the true interval.

[a] Estimates obtained employing asymptotic distribution of the test statistics.

[b] Estimates obtained employing 5000 simulation runs under H_ϕ.

vide further increase in the efficiency of marker analysis of quantitative variation (see also JIANG and ZENG 1995).

Different situations could be considered, depending on the effects of A/a and B/b on the correlated traits: each of the loci affects both traits; A/a affects both traits, while only one of the traits depends on B/b; A/a affects one of traits and B/b the other one; etc. We will refer to the effects of A/a and B/b on the traits as d_x and d_y, and D_x and D_y, respectively. We have restricted our consideration here only to two situations: $d_x \neq 0$ and $d_y = 0$, and $D_x \neq 0$ and $D_y = 0$ (Table 2); and $d_x \neq 0$ and $d_y \neq 0$, and $D_x \neq 0$ and $D_y = 0$ (Table 3). In both cases, the results obtained by the proposed procedure of bivariate analysis obviously demonstrate the positive effect of correlation on the resolution power.

Consider the case, where both loci affect the same trait, *i.e.*, $d_x \neq 0$ and $d_y = 0$, and $D_x \neq 0$ and $D_y = 0$ (Table 2). Let us first compare the upper and lower parts of the table. Clearly, ignoring the dependence of x on locus B/b, *i.e.*, assuming bivariate normality of $f_{aa}(x, y)$ and $f_{Aa}(x, y)$, does not decrease seriously the resolution. This is manifested in proximity of precisions

TABLE 2

Effect of an independently segregating QT locus on the resolution power of bivariate interval analysis: the involved QT loci affect the same trait

				$\beta = \beta(\alpha)$				
R	LOD	ΔLOD	P	$\alpha = 1$	$\alpha = 0.1$	δr	δd_x	δD_x
				Ignoring B/b				
0	2.05 ± 0.08	0.37 ± 0.04	60	45	18	46.6 ± 2.2	128.3 ± 7.3	—
−0.7	2.87 ± 0.10	0.51 ± 0.05	67	73	40	42.5 ± 2.0	104.6 ± 6.0	—
				Including B/b into the model				
0	2.09 ± 0.08	0.39 ± 0.04	67	45	18	47.8 ± 2.2	126.1 ± 7.0	290.0 ± 20.8
−0.7	3.15 ± 0.10	0.57 ± 0.05	72	80	51	43.5 ± 2.0	97.9 ± 5.6	119.8 ± 9.5

In these simulations the effects of A/a locus on the traits x and y were $d_x = 0.5$ and $d_y = 0$, respectively, and the effects of B/b on x and y were $D_x = 1.5$ and $D_x = 0$. In all subgroups (*aabb*, *Aabb*, etc.) $\sigma_x = \sigma_y = 1$. All other characteristics are as described in Table 1. In the first case with $R = 0$, the vector of genetic parameters for H_1 is $\Theta_{n1} = \{r_1, \mu x_1, \mu x_2, \mu y, \sigma_x, \sigma_y\}$ while for H_0 $\Theta_{n0} = \{\mu x, \mu y, \sigma_x, \sigma_y\}$, *i.e.*, $df = 6\text{-}4 = 2$, with $R \neq 0$, $\Theta_{n1} = \{r_1, \mu x_1, \mu x_2, \mu y, \sigma_x, \sigma_y, R\}$ and $\Theta_{n0} = \{\mu x, \mu y, \sigma_x, \sigma_y, R\}$, so that $df = 2$ again. In the second case with $R = 0$, $\Theta_{n1} = \{r_1, \mu x_1, \mu x_2, \mu y, \sigma_x, \sigma_y, D_x\}$ and $\Theta_{n0} = \{\mu x, \mu y, \sigma_x, \sigma_y, D_x\}$, thus $df = 2$, clearly, with $R \neq 0$, $df = 2$ as well.

TABLE 3

Effect of an independently segregating QT locus on the resolution power of bivariate interval analysis: the putative QTL (A/a) affects both traits (x and y), while the independently segregating locus (B/b) affects only one of the traits (x)

				$\beta = \beta(\alpha)$					
R	LOD	ΔLOD	P	$\alpha = 1$	$\alpha = 0.1$	δr	δd_x	δd_y	δD
					Ignoring B/b				
0	2.83 ± 0.10	0.46 ± 0.05	61	52	27	47.0 ± 1.8	186.0 ± 9.9	164.5 ± 8.3	—
−0.7	5.50 ± 0.14	0.96 ± 0.07	81	96	86	33.6 ± 1.7	183.1 ± 8.6	166.3 ± 9.5	—
					Including B/b into the model				
0	2.84 ± 0.10	0.47 ± 0.05	66	53	28	47.1 ± 2.0	183.2 ± 9.6	164.6 ± 8.4	283.9 ± 20.8
−0.7	5.86 ± 0.14	1.06 ± 0.05	84	97	89	34.6 ± 1.9	179.3 ± 9.0	166.3 ± 8.5	119.6 ± 9.2

In these simulations the effect of A/a locus on each of the traits, x and y, was $^1/_4\sqrt{2}$, and the effect of B/b on x was 1.5. In all subgroups (*aabb*, *Aabb*, etc.) $\sigma_x = \sigma_y = 1$. All other characteristics are as described in Table 1. The number of degrees of freedom is determined in the same way as it is shown in Table 2. For example, in the second case with $R \neq 0$, $\Theta_{n1} = \{r_1, \mu x_1, \mu x_2, \mu y_1, \mu y_2, D_x, \sigma_x, \sigma_y, R\}$ and $\Theta_{n0} = \{\mu x, \mu y, D_x, \sigma_x, \sigma_y, R\}$, so that $df = 9\text{-}6 = 3$.

of estimates, LOD values, as well as the proportion of the repeats where the maximum of the LOD function was achieved in the proper interval (P). Such a proximity coincides with the earlier claimed robustness of the interval mapping procedures to disturbance of the normality assumption. The criteria presented in Table 2 demonstrate a tendency for an increased power of the bivariate model, in spite of an increased number of parameters to be estimated and rather small effect of the putative QTL.

Basically, the same results and conclusion about usefulness of the information provided by a covariate trait are obtained in a qualitatively different situations, *e.g.*, in the second situation where A/a affects both traits and B/b only one of the traits (Table 3). For example, as the correlation increases from 0 to 0.7, the mean LOD score increases from 2.8 to 5.9. The power of the test at the 0.1% significance level increases from 28 to 89%. Similarly, the proportion of cases with wrong interval location diminishes from 34 to 16%.

An additional point, common to both considered situations of combined interval mapping and segregation analysis, should be mentioned here. Quite unexpectedly there seems to be an apparent reduction (though a rather small one) in the power β of the test for existence of A/a when the effect of B/b on one of the traits is taken into account. This reduction is characteristic only to situations with small correlation (*i.e.*, at lower resolution). We could assume, that in such cases, neglecting the effect of B/b differentially affects the log-likelihood for H_1 and H_0, resulting in upward bias of the LOD score. Seemingly, the nominator of the test statistics (5) corresponding to H_1 is more robust to inadequate specification of the model, than the denominator (which is, in fact, the likelihood of bivariate normal distribution of the observations).

Resolution power of bivariate interval analysis when, in addition to trait means, the putative QT locus affects also correlation: The problem of identification QTL effects on trait variances has already been discussed in the case of single quantitative trait analysis (*e.g.*, ZHUCHENKO *et al.* 1979; WELLER and WYLER 1992). A similar question about the dependence of resolution (precision of parameter estimates) on the assumption of equal variance-covariance matrices was considered in the multitrait analysis, but with single marker (KOROL *et al.* 1994; RONIN *et al.* 1995). For the sake of simplicity, one may assume that no such effects are presented in the data and put in the model $\Sigma_{aa} = \Sigma_{Aa}$, or $\sigma^2_{aa} = \sigma^2_{Aa}$ for both traits, x and y, as well as $R_{aa} = R_{Aa}$. However, in both single- and multitrait analysis, such kind of simplifications lead to a considerable loss in the test power and precision of the estimates, if these assumptions do not fit the data. On the other hand, the resolution could be improved significantly if indeed $\Sigma_{aa} \neq \Sigma_{Aa}$ and this fact is taken into account (KOROL *et al.* 1994; RONIN *et al.* 1995).

Consider an example of one-trait analysis. One can compare two situations, $\sigma^2_{Aa} = \sigma^2_{aa} = \sigma^2$ and $\sigma^2_{Aa} > \sigma^2_{aa} = \sigma^2$, with all other parameters remaining the same. Clearly, a reduction in the resolution power is expected in the second case, and this indeed will be the case, if the fact $\sigma^2_{Aa} \neq \sigma^2_{aa}$ is ignored in the model. Let the effect of the putative QTL on trait variance be large enough. For instance, one may think of a QTL with linear effect $\mu x_{Aa} = c_1 \mu x_{aa} + c_2$ (KOROL *et al.* 1994) with opposite additive (c_2) and multiplicative ($(c_1 - 1)\mu x_{aa}$) effects. Given $c_1 > 1$ and $c_2 < 0$, the ratio $|d|/\sigma_{aa} = |\mu x_{Aa} - \mu x_{aa}|/\sigma_{aa} = |(c_1 - 1)\mu x_{aa} + c_2|/\sigma_{aa}$ may be relatively small as compared to $c_1 = \sigma_{Aa}/\sigma_{aa}$ [and such situations are not unrealistic, *e.g.*, ZHUCHENKO *et al.* (1979)]. Then, allowing for $\sigma^2_{Aa} \neq \sigma^2_{aa}$ in the model seriously *increases* the resolution when $\sigma^2_{Aa} > \sigma^2_{aa} = \sigma^2$ as compared to that when $\sigma^2_{Aa} = \sigma^2_{aa} = \sigma^2$ (KOROL *et al.* 1994; RONIN *et al.* 1995).

We showed above the positive effect of correlation on the power of the QTL detection test in interval analysis provided $\Sigma_{aa} = \Sigma_{Aa}$ (*e.g.*, Table 1). What will happen

TABLE 4

Effect of nonequal correlation on the efficiency of bivariate interval mapping

Situation		Assumption							
		$R_{aa} = R_{Aa}$				$R_{aa} \neq R_{Aa}$			
			$\beta = \beta(\alpha)$				$\beta = \beta(\alpha)$		
R_{aa}	R_{Aa}	LOD	$\alpha = 1$	$\alpha = 0.1$	δr	LOD	$\alpha = 1$	$\alpha = 0.1$	δr
0	0	1.68 ± 0.04	19	6	56.8 ± 2.3	2.42 ± 0.06	16	4	62.2 ± 2.2
−0.4	−0.4	2.41 ± 0.06	42	19	50.0 ± 2.2	3.14 ± 0.08	34	12	56.3 ± 2.3
−0.4	0	1.94 ± 0.05	27	10	54.3 ± 2.3	4.66 ± 0.10	69	38	48.5 ± 2.1
−0.7	−0.7	4.21 ± 0.09	84	57	39.1 ± 2.0	4.92 ± 0.10	70	45	40.9 ± 2.1
−0.7	0	2.26 ± 0.06	38	15	50.9 ± 2.3	12.05 ± 0.16	100	100	32.7 ± 1.4

In the simulation experiments the traits of interest were assumed to depend only on the QT locus A/a, with $d_x = d_y = (3\sqrt{2})^{-1} \approx 0.23$. For other details see Table 1.

to the power if the correlation in one of the QT locus groups is *reduced* (*e.g.*, $|R_{Aa}| < |R_{aa}|$) with all other parameters remaining the same? The results presented in Table 4 demonstrate that this reduction, as expected, seriously diminishes the resolution when the inequality $R_{Aa} \neq R_{aa}$ (*i.e.*, $\Sigma_{aa} \neq \Sigma_{Aa}$) is ignored. However, what would be less expected is that the reduction in correlation, if included into the model, could considerably increase the resolution power as compared to the initial situation with $R_{Aa} = R_{aa}$. Thus, in all cases presented in Table 4, the resolution power of interval analysis increases across the situations as follows: $\beta\{R_{aa} = 0, R_{Aa} = 0\} < \beta\{R_{aa} = R, R_{Aa} = R\} < \beta\{R_{aa} = R, R_{Aa} = 0\}$.

These results are in agreement with our previously suggested explanation of an analogous effect of nonequal variances $\sigma^2_{Aa} \neq \sigma^2_{aa}$ in single trait 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 $\sigma^2_{Aa}/\sigma^2_{aa}$. We found that both $D(f_{aa}(x), f_{Aa}(x))$ and resolution power may grow not only with increasing $d = x_{Aa} - x_{aa}$, but also with deviation of $\sigma^2_{Aa}/\sigma^2_{aa}$ from unity as well (provided d is relatively small). This consideration could be extended on multitrait analysis. Namely, it seems reasonable to assume that the resolution capacity of the marker analysis in the case of two correlated QTs depends on the discrepancy between the bivariate distributions $f_{aa}(x, y)$ and $f_{Aa}(x, y)$, $D(f_{aa}(x, y), f_{Aa}(x, y))$. While we did not calculate the effect of changes of different parameters of $f_{Aa}(x, y)$ in reducing $|\Sigma_{Aa}|$ [*e.g.*, increased $\sigma^2_{Aa}(x)$ or reduced $R_{Aa}(x, y)$] on $D(f_{aa}, f_{Aa})$, this assumption seems to be a reasonable explanation for the obtained results (see also KOROL *et al.* 1994).

DISCUSSION

Usually, the effect of QTLs on the trait mean values is the target of mapping efforts. In a such a case, differ-

TABLE 5

Invariance of the resolution capacity of interval QTL mapping with respect to bivariate heritability (H^2_{xy}) and comparison with corresponding univariate analysis

Situation									$\beta = \beta(\alpha)$	
d_x	d_y	R	$\frac{H^2_x}{H^2_y}$	H^2_{xy}	QTs included into the model	df	LOD	δr	$\alpha = 1$	$\alpha = 0.1$
0.3	0.3	−0.6	$\frac{0.022}{0.022}$	0.1	x, y	3	4.96 ± 0.13	35.8 ± 1.9	94	73
					x	2	1.27 ± 0.07	57.9 ± 2.2	19	7
0.53	0	−0.6	$\frac{0.066}{0}$	0.1	x, y*	3	4.91 ± 0.10	36.0 ± 2.0	94	74
					x***	2	3.21 ± 0.11	41.6 ± 2.1	78	50
0.53	0	−0.6	$\frac{0.066}{0}$	0.1	x, y**	2	4.63 ± 0.13	35.0 ± 1.9	95	79
0.6	0.6	0.6	$\frac{0.083}{0.083}$	0.1	x, y	3	4.98 ± 0.11	35.9 ± 1.9	94	77
					x	2	3.98 ± 0.12	38.3 ± 1.9	88	71

The results presented in * and ** illustrate the reduction in the test power due to unjustifiable increase in the number of parameters. In the *, $\Theta_{n1} = \{r_1, \mu x_1, \mu x_2, \mu y_1, \mu y_2, \sigma_x, \sigma_y, R\}$ and $\Theta_{n0} = \{\mu x, \mu y, \sigma_x, \sigma_y, R\}$, so that $df = 3$. In the **, $\Theta_{n1} = \{r_1, \mu x_1, \mu x_2, \mu y, \sigma_x, \sigma_y, R\}$ and $\Theta_{n0} = \{\mu x, \mu y, \sigma_x, \sigma_y, R\}$, so that $df = 2$. The value of X^2 for the effect of A/a on the trait y is 1.3 ($df = 1$), which is nonsignificant. Therefore, ** is the true model that allowed to increase the power of the test for detection of A/a from 50% *** up to 79% due to the information provided by measurements of the correlated trait y.

ence measures like $(x_{AA} - x_{aa})/\sigma_x$ or equivalently proportion of the trait variance caused by segregation of the QT locus in question (H_x^2), are of primary interest when discussing the resolution power of linkage analysis [see above (1)]. Among several possibilities suggested to increase the efficiency of marker analysis of QTs, simultaneous analysis of many loci (genomic segments), either linked or unlinked (JANSENS and STAM 1994; ZENG 1994), seems to be especially promising. Complementary to this approach is our suggestion to employ joint distribution of a set of correlated QTs to achieve a further increase in resolution power of QTL mapping. Earlier, this idea was proven to work in single-marker analysis (using as examples backcross, F_2 and recombinant inbred lines) (KOROL *et al.* 1994; RONIN *et al.* 1995). Here we demonstrated the efficiency of the multitrait approach in interval QTL mapping, using simulated backcross data (see also JIANG and ZENG 1995). Clearly, it may be applied to much more complicated structures of mapping populations, *e.g.*, those arising in genetics of trees or animals as well as in human genetics.

For a given set of QTs, the broad sense heritability attributed to a QTL (*e.g.*, A/a) is an increasing function of the number of considered traits. Thus, for some traits x and y, $H_{xy}^2(A/a) \geq H_x^2(A/a)$. The last inequality holds even if y does not depend on A/a at all, but x and y are correlated within the QTL groups (AA, Aa and aa) due to nongenetic factors and segregation of genes from other chromosomes (see also GINSBURG 1983). According to the equality (2), the higher the portion of bivariate variability in the mapping population attributed to the QT locus in question, the better will be the resolution. And indeed, the results presented above (Table 1) qualitatively confirm this expectation. Moreover, one could further assume that the increment in $H_{xy}^2(A/a)$, as compared to $H_x^2(A/a)$, even quantitatively determines the increase in resolution (in spite of complications due to certain statistical nonequivalence) no matter how this increment in $H_{xy}^2(A/a)$ was produced either (i) due to the pleiotropic effect of A/a on x and y, (ii) due to correlation between x and y within the Aa and aa groups caused by nongenetic effects or segregation of unlinked genes, or (iii) due to combined effect of both factors (i) and (ii). The parameter $H_{xy}^2(A/a)$ could be considered as a kind of an invariant in determining the resolution capacity of bivariate interval mapping (at least with an additional assumption of $\Sigma_{aa} = \Sigma_{Aa}$). This is illustrated numerically in Table 5. It is easy to see that the three rather different situations with the same level of H_{xy}^2 (= 0.1) are very close with respect to the LOD value, power of the test for the presence of a QT locus in the marker interval, and precision of estimates.

As we could see from Table 5, with relatively high correlations between QTs (say, $R = -0.6$) a good test power is possible even if the effect of the putative QT locus on either of the traits is too small to allow the detection of the QTL by each single-trait interval analysis (in spite of the increased number of parameters). This may be the case in many practically important situations: (1) when manifold consequences of a segment transfer from donor to recipient genotype should be taken into account, (2) in predicting the best genotypes (lines) in marker assisted selection (see also LANDE and THOMPSON 1990), (3) in dissecting heterosis as a multilocus and multitrait phenomenon, (4) in estimating effects of individual segments on genetic correlations of multitrait complexes and their role in transgressions for trait combinations (DE VINSENTE and TANKSLEY 1993; KOROL *et al.* 1994), etc. Due to the high cost of molecular marker typing, many QTs are usually measured within one experiment. Genetic dissection of quantitative variation of genome expression with respect to the amounts of many individual proteins (DAMERVAL *et al.* 1994) seems to be among the most appealing applications of the proposed multitrait version of interval analysis.

In conclusion, we would like to stress that higher statistical resolution provided by the proposed mapping strategy is not the only advantage of the method for mapping QTL and, probably, not the most important advantage. This approach allows us to test numerous biologically important hypotheses concerning manifold effects of genomic segments on the defined trait complex (means, variances and correlations). Because of the internal balance of the organism's systems (SCHMALHAUSEN 1942), multiple trait analysis seems to be much more justified biologically than the usual trait-by-trait analysis.

We are grateful to Dr. Z.-B. ZENG who made available to us his manuscript on multiple trait analysis before publication. We appreciate his helpful comments and suggestions as well as comments from an anonymous referee. We acknowledge the encouragement of Prof. E. NEVO and the support of the Israeli Ministry of Science (grant no. 3675-1-91), Wolfson Family Charitable Trust and Institut Alain de Rothschild.

LITERATURE CITED

AMES, C. I., and A. E. LAING, 1993 A comparison of univariate and multivariate tests for genetic linkage. Genetic Epidemiol. **10:** 671–676.

BOEHNKE, M., and P. MOLL, 1989 Identifying pedigrees segregating at a major locus for a quantitative trait: an efficient strategy for linkage analysis. Am. J. Hum. Genet. **44:** 216–224.

CARBONELL, E. A., M. J. ASINS, M. BASELGA, E. BALANSARD and T. M. GERIG, 1993 Power studies in the estimation of genetic parameters and the localization of quantitative trait loci for backcross and doubled haploid populations. Theor. Appl. Genet. **86:** 411–416.

CAREY, G., and J. WILLIAMSON, 1991 Linkage analysis of quantitative traits: increased power by using selected samples. Am. J. Hum. Genet. **49:** 786–796.

DAMERVAL, C., A. MAURICE, J. M. JOSSE and D. DE VIENNE, 1994 Quantitative trait loci underlying gene product variation: a novel perspective for analyzing regulation of genome expression. Genetics **137:** 289–301.

DARVASI, A., and M. SOLLER, 1992 Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. Theor. Appl. Genet. **85:** 353–359.

DARVASI, A., A. WEINREB, V. MINKE, J. I. WELLER and M. SOLLER, 1993 Detection marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics **134:** 943–951.

DEMENAIS, E., G. M. LATHROP and J. M. LALOUEL, 1988 Detection of linkage between a quantitative trait and marker locus by the lod scores method: Sample size and sampling considerations. Ann. Hum. Genet. **52:** 237–246.

DEVICENTE, M. C., and S. D. TANKSLEY, 1993 QTL analysis of transgressive segregation in an interspecific tomato cross. Genetics **134:** 585–596.

DOEBLEY, J., and A. STEC, 1993 Inheritance of the morphological differences between maize and teosinte: comparison of results for two F_2 populations. Genetics **134:** 559–570.

EDWARDS, M. D., T. HELENTJARIS, S. WRIGHT and C. W. STUBER, 1992 Molecular marker-facilitated investigations of quantitative trait loci in maize. Theor. Appl. Genet. **83:** 765–774.

GINSBURG, E. KH., 1983 Analysis of the method for background corrections. 1. Phenomenological construction of connection functions. Genetika (USSR) **19:** 761–769.

JANSEN, R. C., and P. STAM, 1994 High resolution of quantitative traits into multiple loci via interval mapping. Genetics **136:** 1447–1455.

JIANG, C., and Z.-B. ZENG, 1995 Multiple trait analysis and genetic mapping for quantitative trait loci. Genetics (in press).

KLEIJNEN, J. P., 1974 *Statistical Techniques in Simulation.* Marcel Dekker, New York.

KNOTT, S. A., and C. S. HALEY, 1992 Aspects of maximum likelihood methods for mapping of quantitative trait loci in line crosses. Genet. Res. **60:** 139–151.

KOROL, A. B., I. A. PREYGEL and N. BOCHARNIKOVA, 1987 Linkage between loci of quantitative traits and marker loci. 5. Simultaneous analysis of a set of marker and quantitative traits. Genetika (USSR) **23:** 1421–1431.

KOROL, A. B., I. A. PREYGEL and S. I. PREYGEL, 1994 *Recombination Variability and Evolution.* Chapman & Hall, London.

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

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

LEBOWITZ, B. J., M. SOLLER and J. S. BECKMANN, 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.

MOTRO, U., and M. SOLLER, 1993 Sequential sampling in determining linkage between marker loci and quantitative trait loci. Theor. Appl. Genet. **85:** 658–664.

RONIN, Y. I., V. M. KIRZHNER and A. B. KOROL, 1995 Linkage between loci of quantitative traits and marker loci: multitrait analysis with a single marker. Theor. Appl. Genet. (in press).

SCHMALHAUSEN, I. I., 1942 *Organism as a Whole Entity in the Individual and Historic Development,* USSR Academy of Sciences, Moscow-Leningrad (in Russian).

SOKAL, R. R., 1965 Statistical methods in systematics. Biol. Rev. Camb. Phil. Soc. **40:** 337–391.

SOLLER, M., and J. S. BECKMANN, 1990 Marker-based mapping of quantitative trait loci using replicated progeny. Theor. Appl. Genet. **80:** 205–208.

SOLLER, M., and A. GENIZI, 1978 The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. Biometrics **34:** 47–55.

TITTERINGTON, D. M., A. F. SMITH, and U. E. MAKOV, 1985 *Statistical Analysis of Finite Mixture Distributions.* Wiley, Chichester, UK.

WELLER, J. I., Y. KASHI and M. SOLLER, 1990 Estimation of sample size necessary for genetic mapping of quantitative traits in dairy cattle using genetic markers. J. Dairy Sci. **73:** 2525–2537.

WELLER, J. I., and A. WYLER, 1992 Power of different sampling strategies to detect quantitative trait loci variance effect. Theor. Appl. Genet. **83:** 582–588.

WILKS, S. S., 1962 *Mathematical Statistics.* Wiley, New York.

ZENG, Z.-B., 1994 Precise mapping of quantitative trait loci. Genetics **136:** 1457–1468.

ZHUCHENKO, A. A., A. SAMOVOL, A. B. KOROL and V. ANDRYUSCHENKO, 1979 Linkage between quantitative trait loci and marker loci. 2. Effects of three tomato chromosomes on the variation for five quantitative characters in backcross populations. Genetika (USSR) **15:** 672–683.

Communicating editor: B. S. WEIR

APPENDIX A

Expected lod score as a function of H^2_{xy}

For the case of a backcross population let us make the following simplifying assumptions: (1) A saturated map of markers is available, so that we can consider a marker located exactly at the position of the putative QT locus (see also LANDER and BOTSTEIN 1989; JIANG and ZENG 1995). (2) The expected value of LOD score (ELOD) is evaluated by substitution the statistics in the likelihood functions by their respective parameters. (3) The joint distribution of the traits within the QTL groups (*aa* and *Aa*) is bivariate normal.

With these assumptions we will show that

$$\mathrm{ELOD} = -{}^1\!/_2 N \log(1 - H^2_{xy}),$$

where E is a symbol of mathematical expectation. In the above consideration H^2_{xy} was proposed as a natural bivariate analogue of a single trait measure of the proportion of variance attributed to the putative QT locus (H^2_x). This was done for the case of no effect of the QTL (A/a) on the variance-covariance matrix of the traits x and y, *i.e.*, $\Sigma_{Aa} = \Sigma_{aa}$. Here we will obtain the expression for ELOD without the last constraint. This allows us to evaluate the upper bound of the expected resolution power (*i.e.*, at absolute linkage between the marker and QTL) and to estimate how it depends on the effects of the QTL not only on trait(s) mean value(s) but also on the variances and correlation. Moreover, this will allow us to propose H^2_{xy} as an analogue of H^2_x for this more general case.

Because of the assumptions (1–3) the expected LOD score can be presented as

$$\mathrm{ELOD} = E\{\log(B/C)\},$$

$$B = \bigcap_{\substack{i=1 \\ k=1,2}}^{N/2} 1/[2\pi\sigma_{xk}\sigma_{yk}\sqrt{(1 - R^2_k)}] \times \exp\left\{ -\frac{1}{2(1 - R^2_k)} \left[\frac{(x_{ik} - \mu_{xk})^2}{\sigma^2_{xk}} - 2R_k \frac{(x_{ik} - \mu_{xk})(y_{ik} - \mu_{yk})}{\sigma_{xk}\sigma_{yk}} + \frac{(y_{ik} - \mu_{yk})^2}{\sigma^2_{yk}} \right] \right\},$$

$$C = \bigcap_{\substack{i=1 \\ k=1,2}}^{N/2} 1/[2\pi\sigma'_x\sigma'_y\sqrt{(1 - R'^2)}] \times \exp\left\{ -\frac{1}{2(1 - R'^2)} \left[\frac{(x_{ik} - \mu'_x)^2}{\sigma'^2_x} - 2R' \frac{(x_{ik} - \mu'_x)(y_{ik} - \mu'_y)}{\sigma'_x\sigma'_y} + \frac{(y_{ik} - \mu'_y)^2}{\sigma'^2_y} \right] \right\}; \quad \text{(A1)}$$

and $k = 1$ and 2 stands for QTL groups *aa* and *Aa*, respectively. It is easy to show that

$$\mu'_x = {}^1/_2(\mu_{x1} + \mu_{x2}), \quad \mu'_y = {}^1/_2(\mu_{y1} + \mu_{y2}),$$
$$\sigma'^2_x = {}^1/_2(\sigma^2_{x1} + \sigma^2_{x2} + {}^1/_2 d^2_x),$$
$$\sigma'^2_y = {}^1/_2(\sigma^2_{y1} + \sigma^2_{y2} + {}^1/_2 d^2_y),$$
$$R' = (\sigma_{x1}\sigma_{y1}R_1 + \sigma_{x2}\sigma_{y2}R_2 + {}^1/_2 d_x d_y)/(2\sigma'_x\sigma'_y). \quad \text{(A2)}$$

Then, $E\{\log B\} = -{}^1/_2 N(2 + \log\{4\pi^2\sigma_{x1}\sigma_{x2}\sigma_{y1}\sigma_{y2}\sqrt{[(1 - R_1^2)(1 - R_2^2)]}\})$. Let us assume, for a moment, that the putative QT locus has no effect on the variance-covariance matrix: $\sigma_{x1} = \sigma_{x2} = \sigma_x$, $\sigma_{y1} = \sigma_{y2} = \sigma_y$, $R_1 = R_2 = R$. In such a case

$$E\{\log B\} = -N\{1 + \log[2\pi\sigma_x\sigma_y\sqrt{(1 - R^2)}]\};$$
$$E\{\log C\} = -N\{1 + \log[2\pi\sigma'_x\sigma'_y\sqrt{(1 - R'^2)}]\},$$

so that

ELOD

$$= {}^1/_2 N \log\{\sigma'_x\sigma'_y\sqrt{(1 - R'^2)}/[\sigma_x\sigma_y\sqrt{(1 - R^2)}]\}.$$

Using (A2), we obtain

$$\text{ELOD} = {}^1/_2 N \log \frac{\sigma^2_x\sigma^2_y(1 - R^2) + {}^1/_4\sigma^2_x d^2_y + {}^1/_4\sigma^2_y d^2_x - {}^1/_2 d_x d_y R\sigma_x\sigma_y}{\sigma^2_x\sigma^2_y(1 - R^2)}, \quad \text{(A3)}$$

or

$$\text{ELOD} = {}^1/_2 N \log \frac{(\sigma^2_x + {}^1/_4 d^2_x)(\sigma^2_y + {}^1/_4 d^2_y) - \sigma^2_x\sigma^2_y[R + d_x d_y/(4\sigma_x\sigma_y)]^2}{\sigma^2_x\sigma^2_y(1 - R^2)},$$

which is equivalent to $-{}^1/_2 N \log(1 - H^2_{xy})$.

If $\Sigma_{Aa} \neq \Sigma_{aa}$, *i.e.*, at least one of the equations $\sigma_{x1} = \sigma_{x2}$, $\sigma_{y1} = \sigma_{y2}$ or $R_1 = R_2$ is violated, then the following will be true:

$$\text{ELOD} = {}^1/_2 N \log[(S_1 + S_2 + S_3 - S_4)/S_1],$$

where

$$S_1 = \sigma_{x1}\sigma_{x2}\sigma_{y1}\sigma_{y2}\sqrt{[(1 - R_1^2)(1 - R_2^2)]},$$
$$S_2 = {}^1/_4[\sigma_{x1}\sigma_{y1}\sqrt{(1 - R_1^2)} - \sigma_{x2}\sigma_{y2}\sqrt{(1 - R_2^2)}]^2 + \sigma_{x1}\sigma_{y1}\sigma_{x2}\sigma_{y2}\cdot\{(1 - R_1R_2) - \sqrt{[(1 - R_1^2)(1 - R_2^2)]}\} + {}^1/_4(\sigma_{x1}\sigma_{y2} - \sigma_{x2}\sigma_{y1})^2,$$
$$S_3 = {}^1/_2(\sigma^2_{x1} + \sigma^2_{x2})\,{}^1/_4 d^2_y + {}^1/_2(\sigma^2_{y1} + \sigma^2_{y2})\,{}^1/_4 d^2_x,$$
$$S_4 = {}^1/_4 d_x d_y(\sigma_{x1}\sigma_{y1}R_1 + \sigma_{x2}\sigma_{y2}R_2).$$

Thus, from (A3) we have the following expression for bivariate H^2:

$$H^2_{xy}(A/a) = 1 - S_1/(S_1 + S_2 + S_3 - S_4).$$

The last expression could be proposed as a natural bivariate analog of H^2_x attributed to the putative QTL, covering the case of variance-covariance effect of the QTL.

APPENDIX B

Invariance of H^2_{xy} with respect to linear transformation of the variables

Consider the case of equal variance-covariance matrices, *i.e.*, $\sigma_{x1} = \sigma_{x2} = \sigma_x$, $\sigma_{y1} = \sigma_{y2} = \sigma_y$, $R_1 = R_2 = R$. Consider an arbitrary nondegenerative linear transformation,

$$x = \alpha_1 u + \alpha_2 v + \alpha_3; \quad y = \beta_1 u + \beta_2 v + \beta_3$$
$$(\text{Det} = \alpha_1\beta_2 - \alpha_2\beta_1 \neq 0).$$

Clearly,

$$\sigma^2_x = \alpha^2_1\sigma^2_u + \alpha^2_2\sigma^2_v + 2\alpha_1\alpha_2\sigma_u\sigma_v R_{uv},$$
$$\sigma^2_y = \beta^2_1\sigma^2_u + \beta^2_2\sigma^2_v + 2\beta_1\beta_2\sigma_u\sigma_v R_{uv},$$
$$\sigma_x\sigma_y R_{xy} = \alpha_1\beta_1\sigma^2_u + \alpha_2\beta_2\sigma^2_v + (\alpha_1\beta_2 - \alpha_2\beta_1)\sigma_u\sigma_v R_{uv}.$$
$$d_x = \alpha_1 d_u + \alpha_2 d_v, \quad dy = \beta_1 d_u + \beta_2 d_v.$$

Substitution of these relationships into (3) gives for the nominator,

$$\sigma^2_x\sigma^2_y(1 - R^2_{xy}) = (\alpha_1\beta_2 - \alpha_2\beta_1)\sigma^2_u\sigma^2_v(1 - R^2_{uv}),$$

for the denominator,

$$(\sigma^2_x + d^2_x/4)(\sigma^2_y + d^2_y/4) - \sigma^2_x\sigma^2_y[R_{xy} + d_x d_y/(4\sigma_x\sigma_y)]^2 = (\alpha_1\beta_2 - \alpha_2\beta_1)\{(\sigma^2_u + d^2_u/4)(\sigma^2_v + d^2_v/4) - \sigma^2_u\sigma^2_v[R_{uv} + d_u d_v/(4\sigma_u\sigma_v)]^2\}.$$

This proves the statement.