Theor Appl Genet (1995) 90:776–786

Y. I. Ronin · V. M. Kirzhner · A. B. Korol

Linkage between loci of quantititative traits and marker loci: multi-trait analysis with a single marker

Received: 7 September 1994 / Accepted: 8 September 1994

Abstract An efficient approach to increase the resolution power of linkage analysis between a quantitative trait locus (QTL) and a marker is described in this paper. It is based on a counting of the correlations between the QTs of interest. Such correlations may be caused by the segregation of other genes, environmental effects and physiological limitations. Let a QT locus A/a affect two correlated traits, x and y. Then, within the framework of mixture models, the accuracy of the parameter estimates may be seriously increased, if bivariate densities $f_{aa}(x, y)$, $f_{Aa}(x, y)$ and $f_{AA}(x, y)$ rather than the marginals are considered as the basis for mixture decomposition. The efficiency of the proposed method was demonstrated employing Monte-Carlo simulations. Several types of progeny were considered, including backcross, F_2 and recombinant inbred lines. It was shown that provided the correlation between the traits involved was high enough, a good resolution to the problem is possible even if the QTL groups are strongly overlapping for their marginal densities.

Key words ML-estimation · QTL · Mixture model · Multitrait complexes

Introduction

The resolution power of marker-based analysis of quantitative traits is the major factor affecting the practical importance of quantitative trait locus (QTL) mapping. A detailed discussion of the issues concerning the power of tests for detecting linkage and designing experiments 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 estimates depends on the effect of the QTL in question relative to the total phenotype variance of the trait. Thus, with a QTL (A/a) segregating in a backcross progeny, in the case of equal variances in the groups Aa and aa ($\sigma^2_{Aa} = \sigma^2_{aa} = \sigma^2$), the ratio $(x_{Aa} - x_{aa})/\sigma$ is of primary interest when discussing the power of tests for marker-QTL linkage or the estimation precision of the QT locus effect $d = x_{Aa} - x_{aa}$ and recombination rate between A/a and a linked marker. Therefore, it is expected that lowering the trait variance, e.g. by progeny testing, could increase the resolution power (Thoday 1967; Soller and Beckmann 1990). What can happen if the variability in one of the groups, say in Aa, grows? Intuitively, one would expect a decrease in resolution proportional to the increase in σ^2_{Aa}. However, we found that if the distance $d = x_{Aa} - x_{aa}$ between the means x_{Aa} and x_{aa} is small, then the precision of the estimates increases with increasing σ^2_{Aa}, if $\sigma^2_{Aa} \neq \sigma^2_{aa}$ is taken into account (Korol et al. 1994).

Among other possibilities for improving the precision of mapping, it is worth mentioning the multimarker (interval) approach (Lander and Botstein 1989; Haley and Knott 1992; Martinez and Curnow 1992), selective sampling (Lebowitz et al. 1987; Carey and Williamson 1991; Darvasi and Soller 1992), sequential analysis (Morton 1955; Boehnke and Moll 1989; Motro and Soller 1993; Korol et al. 1994b) and simultaneous analysis of several QTL (including linked ones) affecting the target trait (Jansen 1993; Jansen and Stam, 1994; Zeng 1994).

Under certain conditions, one can expect to increase the discrepancy between the QTL groups (Aa and aa, for a backcross case) based on the analysis of joint distribution of several traits [say $f(x, y)$] even if the groups are strongly overlapping in their marginal distributions $f(x)$ and $f(y)$. An increased discrepancy between the component distributions of the QT in the segregating progeny may result in a higher accuracy of linkage estimation between the marker and QTL. Indeed, earlier we have

Communicated by F. Salamini

Y. I. Ronin · V. M. Kirzhner · A. B. Korol (✉)
Institute of Evolution, University of Haifa, Mount Carmel, Haifa 31905, Israel

shown that with high enough correlations between quantitative traits, a good resolution is possible, in principle, even if the QTL groups are strongly overlapping for their marginal distributions (Korol et al. 1987, 1994b). It should be stressed here that the basic idea of using correlations to increase the performance of analysis is rather common in genetics. In this connection, such situations as genotype-environment interaction (Falconer 1981), divergence of populations or cultivars (Arunachalam, 1981), index selection (Lin 1978), marker-assisted breeding (Lande and Thompson 1990), the use of other markers as co-factors in interval mapping of QTL (Jansen 1993; Jansen and Stam 1994; Zeng 1994) are worth mentioning as examples.

The objective of this paper is to demonstrate the advantage of the multi-trait analysis for different progeny types in estimating linkage between a QTL and a marker. Note, that with dense enough map of markers, single-marker analysis has approximately the same resolution power as interval mapping (of course, only if one QTL from the chromsome in question segregates in the mapping population) (Knott and Haley 1992; Darvasi et al. 1993).

General description of the method

Consider first the simplest case of backcross progeny with a QTL (A/a) that influences two correlated traits, x and y. In the example presented in Fig. 1, the marginal distributions are strongly overlapping. Without any additional information and based only on the observed marginal distributions $f(x)$ and $f(y)$, one would hardly assume that the progeny is polymorphic for an oligogene. Nevertheless, the presence of an oligogene can easily be seen from the joint distribution $f(x, y)$.

Joint analysis of several traits may be no less important in situations where the trait of interest (say x) is dependent on the locus A/a and is strongly correlated with another trait (y), while the latter one is independent of A/a. Even in this case, the additional information provided by measurements of y can dramatically increase the mapping precision of the gene A/a. Such correlations may be caused by the segregation of other genes, environmental effects and physiological limitations. Figure 2 shows schematically some simple cases where two-trait analysis may give a resolution power that could not be achieved in a marker analysis of a single quantitative trait (with the same effect of the putative QTL and sample size). In a sense, the basic idea here is close to the classification methods based on the main component methods. Our previous Monte-Carlo study confirmed this expectation of an increase in resolution capacity of marker-QTL linkage analysis when the estimation procedure takes into account the correlations between the quantitative traits in marker groups (Korol et al. 1994).

Technically, the algorithms of multi-trait analysis are similar to those of the single-trait situation (Titterington et al. 1985; Darvasi and Weller 1992; Korol et al. 1994). No special difficulties are expected in developing multi-trait analogues of mapping based on the maximum likelihood (ML) method. The only difference is in the increased number of parameters to be estimated. We think that an improved resolution compensates for the latter drawback. In order to obtain a ML solution to the problem, we should specify the joint distribution of the considered correlated traits in QTL groups (say Aa and aa, in a backcross; AA, Aa and aa, in F_2; AA and aa, in the case of recombinant inbred lines, RILs).

In this paper we restrict our analysis to single-marker situations, while the proposed approach is no less effective within the framework of interval mapping of QTL (Korol et al., in preparation). Let A/a be the QTL affecting several quantitative traits (x, y, etc.), and M/m be the linked marker locus. Consider the case of two traits (x, y) and let (for the sake of certainty) the joint distributions $f_{aa}(x, y) = f_1(x, y)$, $f_{Aa}(x, y) = f_2(x, y)$ and $f_{AA}(x, y) = f_3(x, y)$ be the two-dimensional normal. Then, the expected distributions in the three-marker groups, $S_{mm}(x, y) = S_1(x, y)$, $S_{Mm}(x, y) = S_2(x, y)$, and $S_{MM}(x, y) = S_3(x, y)$ can be written as:

$$S_i(x, y) = \sum_{j=1}^{3} \pi_{ij}(r) f_j(x, y), \tag{1}$$

with the proportions $\pi_{ij}(r)$ being dependent on the unknown exchange rate r between A/a and M/m. The form of $\pi_{ij}(r)$ should be specified for each type of the progeny. In case of bivariate normality,

$$f_1(x, y) = [2\pi\sigma_{1x}\sigma_{1y}(1 - R_1^2)]^{-1/2} \tag{2}$$

$$\times \exp\left\{-\frac{1}{2(1-R_1^2)}\left[\frac{(x-\bar{x}_1)^2}{\sigma_{1x}^2} - 2R_1\frac{(x-\bar{x}_1)(y-\bar{y}_1)}{\sigma_{1x}\sigma_{1y}} + \frac{(y-\bar{y}_1)^2}{\sigma_{1y}^2}\right]\right\},$$

$$f_2(x, y) = [2\pi\sigma_{2x}\sigma_{2y}(1 - R_2^2)]^{-1/2}$$

$$\times \exp\left\{-\frac{1}{2(1-R_2^2)}\left[\frac{(x-\bar{x}_2)^2}{\sigma_{2x}^2} - 2R_2\frac{(x-\bar{x}_2)(y-\bar{y}_2)}{\sigma_{2x}\sigma_{2y}} + \frac{(y-\bar{y}_2)^2}{\sigma_{2y}^2}\right]\right\},$$

$$f_3(x, y) = [2\pi\sigma_{3x}\sigma_{3y}(1 - R_3^2)]^{-1/2}$$

$$\times \exp\left\{-\frac{1}{2(1-R_3^2)}\left[\frac{(x-\bar{x}_3)^2}{\sigma_{3x}^2} - 2R_3\frac{(x-\bar{x}_3)(y-\bar{y}_3)}{\sigma_{3x}\sigma_{3y}} + \frac{(y-\bar{y}_3)^2}{\sigma_{3y}^2}\right]\right\},$$

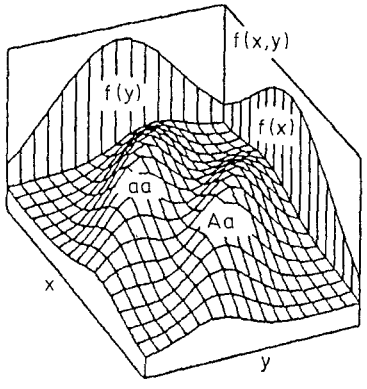


Fig. 1 Join distribution $f(x, y)$ of two correlated traits, x and y, in the backcross population. Even with the clear-cut bimodality of $f(x, y)$ [when the components $f_{aa}(x, y)$ and $f_{Aa}(x, y)$ are far enough apart and the correlation is high], the marginal distributions $f(x)$ and $f(y)$ are unimodal

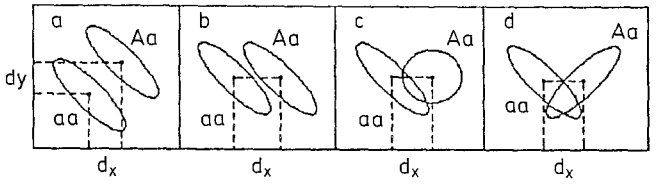


Fig. 2a–d Some possible situations in two-trait analysis. An example is taken when the correlations between x and y within the groups are either the same and negative (**a** and **b**) or different: (**c**) negative and zero, (**d**) of opposite signs. The effects of the locus A/a are: (**a**) increase in both traits, (**b–d**) increase in x and no effect in y

where $\bar{x}_i$, $\bar{y}_i$, σ_{ix}, σ_{iy} $(i=1,2,3)$ are the mean values and standard deviations of x and y in groups aa $(i=1)$, Aa $(i=2)$, and AA $(i=3)$; R_i are the correlations between x and y in the groups.

Log-likelihood of a sample with the sizes N_i of the marker groups can be written as:

$$\ln L(\Theta) = \sum_{i=1}^{N_1} \ln S_1(x_i, y_j) + \sum_{j=1}^{N_2} \ln S_2(x_j, y_j) + \sum_{k=1}^{N_3} \ln S_3(x_k, y_k) + \text{const}, \quad (3)$$

where Θ is the vector of unknown parameters, e.g. for F_2 $\Theta = \{r, \bar{x}_1, \bar{y}_1, \bar{x}_2, \bar{y}_2, \bar{x}_3, \bar{y}_3, \sigma_{1x}, \sigma_{1y}, \sigma_{2x}, \sigma_{2y}, \sigma_{3x}, \sigma_{3y}, R_1, R_2, R_3, \}$, x and y with indices are the trait values of the corresponding genotypes. In some cases only two marker groups will be presented in the progeny and then only two sums will be presented in $\ln L(\Theta)$.

Several types of progeny are considered below, including backcross, F_2 and recombinant inbred lines.

Simulation procedure and ML-estimation

Generating the data

Monte-Carlo simulations were used to produce the "observations". For each situation and each combination of parameters studied 100–200 repeated mapping populations have been generated using pseudo-random numbers. Normal distribution was used for the trait groups aa, Aa and AA, while any other density could be considered as well. To obtain normally distributed variables the inverse transformation to the normal distribution function has been employed. For comparative analysis of different methods and situations we used, where possible, one and the same set of data. In order to get a bivariate normal distribution with a preset correlation coefficient R, the approach described in Kleijnen (1974) was employed: x and y will have the desired joint distribution if y is simulated according to the following relationship:

$$y - \bar{y} = R\sigma_y/\sigma_x(x - \bar{x}) + (1 - R^2)\sigma_y z,$$

where z is a standard normal variable $N(0, 1)$. The composition of the marker groups (mixtures S_i, $i=1,2,3$) were modelled as trinomial distributions with proportions $\pi_{ij}(r)$. For most of the experiments, parameter values used for simulations were in the range: $0.1 \le r \le 0.4$ $0.5 \le d_x = x_{Aa} - x_{aa} \le 2, 0 \le d_y = y_{Aa} - y_{aa} \le 1, \sigma_{aa} = 1, 0 \le |R| \le 0.95$, $N = 1000$.

Obtaining numerical solutions

The target of this work was to compare the estimation approach described above 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. As in our previous study concerned with the comparsion of different statistical approaches for accuracy of QTL-marker linkage estimation (Korol et al. 1994), we do not dwell enough in this study on the problems of numerical procedures of multiextremal, multidimensional optimization. The main objective here was to check how the correlation between the considered traits affects the closeness of the optimal points (representing the estimate of the paramter vector Θ) to the true parameter set used for simulation and/or to the sets corresponding to each simulated sample. For this specific goal, we do not have to search for the solution starting from arbitrary points. The simplest way to obtain the necessary estimates is to use as 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 combinations of the model parameters studied 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.

Estimating the accuracy of the obtained solutions

Usually, variances or standard errors of the estimates are employed as a means for efficiency comparison of the estimation procedures. However, in addition to random fluctuations around the mean, another possible source of distrubance, 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 values could depend on the level of the parameters. In our Monte-Carlo experiments, the simulated parameter values are fluctuating stochastically around their chosen expectations, while the estimates may, possibly, be biased. Thus, we need an integral indicator of the accuracy of the respective estimates. In order to allow for possible differences in biases of the estimates, we employed the absolute error of the estimate, averaged over the repeated experiments:

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

where $\tilde{u}_k$ and $\hat{u}_k$ are, respectively, the simulated ('occurred') and estimated values of the parameter u (i.e. u can be any component of the vector Θ, say r, d, x_{aa}, σ^2_{aa} etc).

Results

Backcross progeny

Bivariate distribution of traits x and y in each of the marker groups, mm and Mm, is a mixture of two components, $f_{aa}(x, y) = f_1(x, y)$ and $f_{Aa}(x, y) = f_2(x, y)$:

$$S_{mm}(x, y) = S_1(x, y) = (1 - r)f_1(x, y) + rf_2(x, y),$$
$$S_{Mm}(x, y) = S_2(x, y) = rf_1(x, y) + (1 - r)f_2(x, y), \quad (4)$$

where the densities $f_1(x, y)$ and $f_2(x, y)$ are from Eq. 2. The expressions of Eqs. 4 are then used to get the likelihood functional like Eq. 3. In this formulation, the parameter space is 11-dimensional.

Most of the results described below are concerned with the situation presented in Fig. 2b, where the trait x depends on the putative QTL (A/a) linked to the marker M/m. The second trait, y, is assumed to be independent of the locus A/a while correlated with x within each of the QT locus groups, aa and Aa (usually, but not necessarily, with one and the same correlation coefficient). Figure 3 shows the distribution of the ML estimates of r as dependent on the correlation R_{xy} within the QTL groups. It is easy to see that with increased R_{xy} the precision of the estimates monotonically increases. This is true also for all other parameters (see examples in Fig. 4).

Even small quantitative trait effects could be estimated satisfactory if R_{xy} is high enough. Let, for example, the effect of $aa \rightarrow Aa$ substitution on the trait x be $d = 0.5$ and $\sigma_{1x} = \sigma_{2x} = 1$, so that segregation at locus

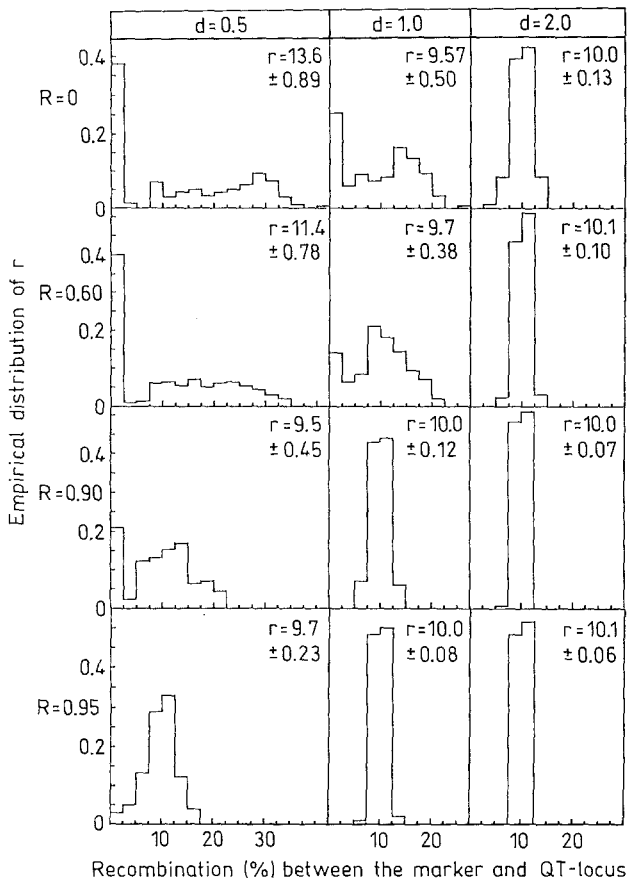


Fig. 3 The effect of correlation between traits (R) on the distribution of the estimates of recombination rate between the marker and QTL. For the case of a backcross progeny, 200 replicates, each of size $n = 1000$, were simulated at a recombination rate $r = 10\%$, with the distance between QTL groups $d_x = x_{Aa} - x_{aa} = \{0.5, 1, 2\}$ and $d_y = y_{Aa} - y_{aa} = 0$, $\sigma x_{Aa} = \sigma x_{aa} = \sigma y_{Aa} = \sigma y_{aa} = 1$. In the ML-functional all σ's were assumed to be different, as were R_{Aa} and R_{aa}.

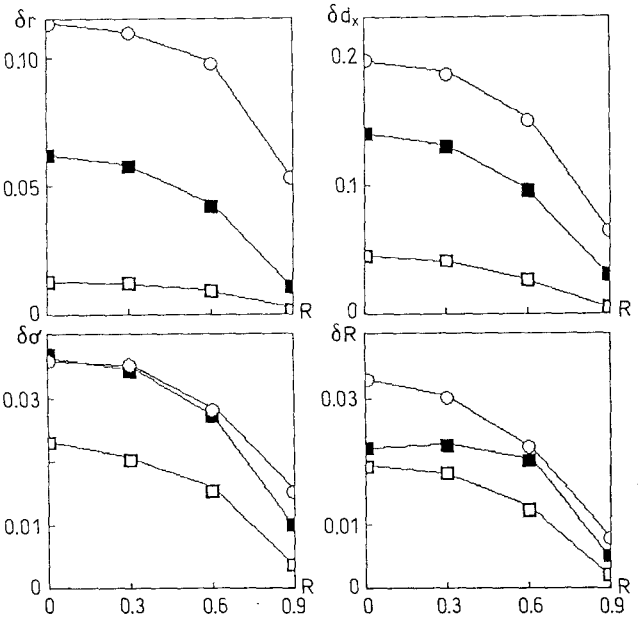


Fig. 4 The dependence of mean absolute errors (δ) of the parameter estimates in the two-trait analysis on correlation (R) between the traits in case of a backcross progeny. Two hundred replicates, each of size $n = 1000$, were simulated at recombination rate $r = 0.1$, with the distance between QTL groups $d_x = x_{Aa} - x_{aa} = \{0.5, 1, 2\}$ and $d_y = y_{Aa} - y_{aa} = 0$, $\sigma x_{Aa} = \sigma x_{aa} = \sigma y_{Aa} = \sigma y_{aa} = 1$. In the ML-functional all σ's were assumed to be different, as were R_{Aa} and R_{aa}. *Open circles*, *black squares* and *open squares* correspond to $d = 0.5$, $d = 1$, and $d = 2$, respectively.

A/a is responsible for 6% of the total phenotypic variance of x in the backcross progeny. Then, the mean absolute error of the ML estimate of d is $\delta d = 0.204 \pm 0.012$, i.e. the relative error is about 40%. However, if an additional trait, y, independent of the locus A/a but closely correlated with x ($R_{xy} = 0.9$), is taken into account, then the absolute error of d is reduced to $\delta d = 0.066 \pm 0.0037$, i.e. three-fold. The corresponding estimates δr are 0.120 ± 0.0045 and 0.053 ± 0.0025 (recall that the value $r = 0.1$ was used in this series of simulations). At $R_{xy} = 0.95$ we obtain $\delta d = 0.042 \pm 0.0024$ and $\delta r = 0.024 \pm 0.0015$. With $R_{xy} = \pm 0.95$, an acceptable precision is obtained even for such a small effect as $d = 0.25$ (segregation at A/a is responsible for less than 1.54% of the total phenotypic variance of x in the backcross): in this case $\delta d = 0.054 \pm 0.0044$ and $\delta r = 0.066 \pm 0.0036$.

The growth in resolution due to an accounting of the correlation between traits is also expected when both of the traits depend on the QTL in question, like in case a in Fig. 2. Our simulations have shown that this is really the case. Some examples are presented in Table 1. Especially important is the fact of a considerable decrease of δr, δd_x and δd_y with increased correlation between the traits controlled by QTL, no matter what the realtive rating of the considered variants, $d_x = 0.5$ and $d_y = 0$ and $d_x = 0.5$ and $d_y = -0.5$, at each specific level of R_{xy}. The same conclusion is also valid for other parameters (standard variations of x and y within the QTL groups aa and Aa and correlations R_{xy} in these groups), which could be of interest in some formulations of the QTL mapping problem.

One can assume that a further increase in the efficiency of the estimation procedure is possible with the accounting of additional quantitative traits correlated with the trait in question. Indeed, in simulation experiments with three traits, we found that such a tendency is manifested by all of the parameters, at least when the correlations between the traits are high enough (Table 2). In the case $d_x = x_{Aa} - x_{aa} = 0.5$ of the example shown in Table 2, the mean absolute error of the estimates of r is reduced twofold when a second trait, y, correlated to x with $R_{xy} = 0.9$ is also taken into account (as compared to the case $R = 0$). A further reduction of δr, threefold as compared to the case $R = 0$, is observed when a third trait correlated to the x trait, z, is included into the model ($R_{xz} = 0.9$).

It is well-known that with increased distance between the QTL and marker, the estimation accuracy of the genetic parameters falls down, the effect being especially pronounced for the estimate of r (Luo and Woolliams

Table 1 The dependence of mean absolute errors (δ)[a] of the parameter estimates in the two-trait analysis on correlation (R) between the traits in the case of a backcross progeny. Two hundred replicates, each of size $n = 1000$, were simulated at recombination rate $r = 0.1$, distance between QT locus groups $d_x = x_{Aa} - x_{aa} = 0.5$ and $d_y = y_{Aa} - y_{aa} \in \{0, -0.5\}$, $\sigma x_{Aa} = \sigma x_{aa} = \sigma y_{Aa} = \sigma y_{aa} = 1$. In the ML-functional all σ's were assumed to be different, as well as R_{Aa} and R_{aa}

	Variable parameter values used in simulations					
$R_{Aa} = R_{aa} = R$	0			0.6		0.9
d_x	0.5	0.5	0.5	0.5	0.5	0.5
d_y	0.0	−0.5	0.0	−0.5	0.0	−0.5
	Mean absolute errors of the parameter estimates					
δr	114 ± 4.4	97 ± 3.1	97 ± 3.8	54 ± 2.5	53 ± 2.5	10 ± 0.6
δd_x	196 ± 12	133 ± 7.8	149 ± 9.6	71 ± 3.4	68 ± 3.2	27 ± 1.5
δd_y	55 ± 5.8	137 ± 8.0	49 ± 4.0	67 ± 3.4	38 ± 2.3	27 ± 1.4
$\delta\sigma x_{aa}$	36 ± 2.8	24 ± 1.6	28 ± 2.2	15 ± 0.9	15 ± 0.9	8 ± 0.5
$\delta\sigma y_{aa}$	20 ± 1.4	25 ± 1.8	16 ± 1.1	14 ± 0.9	11 ± 0.7	7 ± 0.4
δR_{aa}	34 ± 2.7	45 ± 2.9	23 ± 1.5	25 ± 1.2	7 ± 0.4	4 ± 0.3
$\delta\sigma x_{Aa}$	38 ± 2.9	24 ± 1.8	29 ± 2.1	15 ± 0.9	14 ± 0.8	8 ± 0.4
$\delta\sigma y_{aa}$	20 ± 1.6	26 ± 2.0	16 ± 1.2	15 ± 0.8	12 ± 0.7	9 ± 0.5
δR_{Aa}	29 ± 2.2	46 ± 3.1	21 ± 1.5	26 ± 1.4	7 ± 0.4	5 ± 0.3

[a] The mean absolute errors of the estimates are multiplied by 1000

Table 2 The effect of the number of correlated traits (k) on the accuracy of parameter estimates in case of a backcross progeny. Two hundred replicates, each of size $n = 1000$, were simulated at recombination rate $r = 0.1$, distance between QTL groups $d_x = x_{Aa} - x_{aa} \in \{0.5, 1, 2\}$, $d_y = y_{Aa} - y_{aa} = 0$, $\sigma x_{Aa} = \sigma x_{aa} = \sigma y_{Aa} = \sigma y_{aa} = 1$. In the three-dimensional case, in addition to the above, the conditions $d_z = z_{Aa} - z_{aa} = 0$ and $\sigma z_{Aa} = \sigma z_{aa} = 1$ were used in the simulations. In the ML-functional all σ's were assumed to be different, as well as R_{Aa} and R_{aa}

k		Variable parameter values used in simulations								
	$R_{Aa} = R_{aa} = R$	0.0			0.6			0.9		
	d_x	0.5	1	2	0.5	1	2	0.5	1	2
		Mean absolute errors[a] of the parameter estimates								
2	δr	114.4 ± 4.4	62.0 ± 2.3	12.6 ± 0.8	97.4 ± 3.8	43.0 ± 2.3	8.7 ± 0.5	52.6 ± 2.5	10.4 ± 0.6	1.8 ± 0.1
	δd_x	196.1 ± 12.0	140.1 ± 6.2	45.6 ± 2.6	149.0 ± 9.6	96.2 ± 5.2	26.7 ± 1.6	66.1 ± 3.7	30.3 ± 1.7	4.5 ± 0.4
3	δr	119.6 ± 4.5	62.2 ± 2.4	12.9 ± 0.8	100.0 ± 3.5	37.1 ± 2.1	7.0 ± 0.4	39.8 ± 2.2	7.3 ± 0.4	1.2 ± 0.1
	δd_x	203.8 ± 11.9	144.1 ± 6.2	46.9 ± 2.8	144.1 ± 8.2	82.4 ± 4.2	21.6 ± 1.3	45.9 ± 3.0	18.9 ± 1.2	1.4 ± 0.1

[a] The mean absolute errors of the estimates are multiplied by 1000

1993; Korol et al. 1994). An important advantage of the multi-trait analysis is demonstrated by Fig. 5: a much slower growth of the absolute errors with r. The higher the correlation, the stronger the effect. The remarkable fact is that with high enough R_{xy}, δr does not increase with r.

The last question we would like to consider here is the effect of inequality of bivariate distributions in the QTL groups, *Aa* and *aa*. As mentioned in the Introduction, taking into account the fact that $\sigma^2_{aa} \neq \sigma^2_{Aa}$ may result in a serious increase in the resolution power in the single-trait analysis (for more details see Korol et al. 1994). On the contrary, if for example $\sigma^2_{Aa} \neq \sigma^2_{aa}$ and this fact is ignored, a dramatic reduction in the accuracy will be obtained. We found the same effect in the two-dimensional case. The analogue of variance in the multi-dimensional case is the norm of the variance-covariance matrix. Therefore, we have checked whether or not an increase in the norm of this matrix for the group *Aa*, $|\Sigma_{Aa}|$, may result in a higher resolution. Three possibilities exist to increase $|\Sigma_{Aa}|$: (1) increase in $\sigma^2 x_{Aa}$; (2) increase in $\sigma^2 y_{Aa}$; (3) decrease in Rxy_{Aa}. Starting from the case $\sigma^2 x_{Aa} = \sigma^2 x_{aa} = 1$, $\sigma^2 y_{Aa} = \sigma^2 y_{aa} = 1$ and $Rxy_{Aa} = Rxy_{aa} = 0.7$ ($|\Sigma_{Aa}| = |\Sigma_{aa}| = 0.51$), we considered also the above three cases with increased $|\Sigma_{Aa}|$ (as compared to the case $|\Sigma_{Aa}| = 0.51$) and unchanged $|\Sigma_{aa}|$ (Table 3).

The results presented in Table 3 demonstrate unequivocally that an increase in the level of variation in the second group (increased $|\Sigma_{Aa}|$) results in a better estimation accuracy of the parameters characterizing the QTL effect and position, no matter what the cause of the increased variation: increased variance of the trait x controlled by A/a, increased variance of the correlated

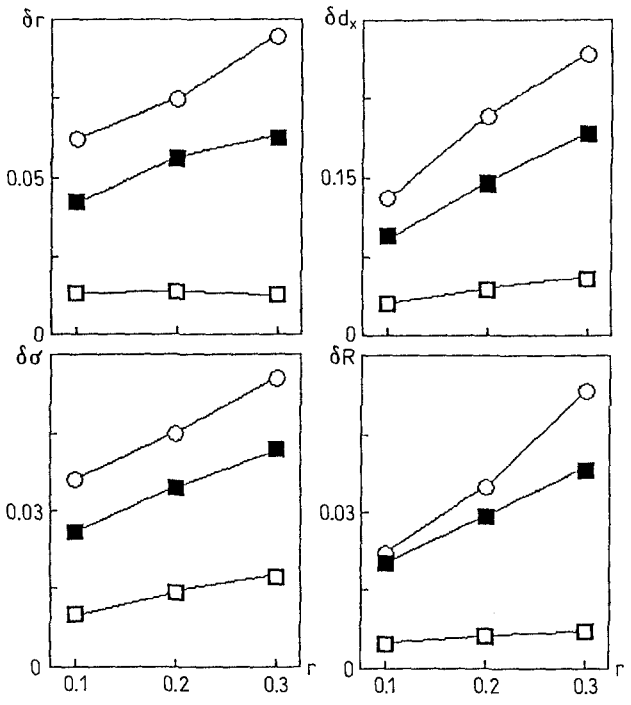


Fig. 5 The effect of correlation between traits (R) on the dependence of the estimation accuracy of the main genetic parameters in the two-trait analysis on the recombination rate (r) between the marker and QTL. Two hundered replicates, each of size $n = 1000$, were simulated at distance between QTL groups $d_x = x_{Aa} - x_{aa} = 1$ and $d_y = y_{Aa} - y_{aa} = 0$, $\sigma x_{Aa} = \sigma x_{aa} = \sigma y_{Aa} = \sigma y_{aa} = 1$. In the ML-functional all σ's were assumed to be different, as were R_{Aa} and R_{aa}. *Open circles*, *black squares* and *open squares* correspond to $R = 0$, $R = 0.6$, and $R = 0.9$, respectively

with x trait y (which does not depend on locus a/a) or a reduced correlation Rxy in the group Aa) (compare the last three columns with the second one). However, the three variants of increased $|\Sigma_{Aa}|$ are not quantitatively equivalent with respect to the level of accuracy of the estimates of the main parameters, r and d_x. The highest benefit in terms of a decrease in δr and δd_x, given the same $|\Sigma_{Aa}| > |\Sigma_{aa}|$, was obtained with a changed Rxy_{Aa} ($Rxy_{Aa} < Rxy_{aa}$): $\delta r = 0.018 \pm 0.0010$ and $\delta d_x = 0.043 \pm 0.0022$, as compared to $\delta r = 0.034 \pm 0.0020$ and $\delta d_x = 0.076 \pm 0.0041$ in the case of $|\Sigma_{Aa}| = |\Sigma_{aa}|$). The lowest decrease of δr and δd_x was obtained with $\sigma^2 x_{Aa} > \sigma^2 x_{aa}$: $\delta r = 0.025 \pm 0.0014$ and $\delta d_x = 0.059 \pm 0.0031$.

Recombinant inbred lines (RILs)

Here we have two marker groups, *mm* and *MM*, each consisting of genotypes *aa* and *AA* in the following proportions:

$$S_{mm}(x, y) = S_1(x, y) = [1 - \pi(r)] f_1(x, y) + \pi(r) f_3(x, y),$$

$$S_{MM}(x, y) = S_3(x, y) = \pi(r) f_1(x, y) + [1 - \pi(r)] f_3(x, y), \quad (5)$$

where the densities $f_i(x, y)$ and $f_3(x, y)$ are from Eq. 2, $\pi(r) = 4r/(1 + 6r)$ when brother-sister mating was practiced and $\pi(r) = 2r/(1 + 2r)$ with selfing (e.g. Simpson 1989). One important advantage of RILs in QTL-marker linkage analysis is the possibility to increase the resolution power by a repeated progeny testing for the quantitative trait in question (Soller and Beckmann 1990). The multi-trait approach provides an additional and still unemployed possibility to increase the accuracy of the estimates. From the viewpoint of parameter estimation, the case of RILs is equivalent to the backcross, but instead of r we have here $\pi(r)$ as a mixture proportion.

An important distinction of the proposed multi-trait analysis when applied to RILs is the necessity to take into account the effect of progeny size per line (k) on the within-QTL-groups (AA and aa) variation. Namely,

Table 3 The effect of an increase in the variance-covariance matrix $|\Sigma_{Aa}|$ on mean absolute errors (δ)[a] of the parameter estimates in the two-trait analysis of a backcross progeny. Two hundred replicates, each of size $n = 1000$, were simulated at recombination rate $r = 0.1$, distance between QTL groups $d_x = x_{Aa} - x_{aa} = 1$ and $d_y = y_{Aa} - y_{aa} = 0$

	Variable parameter values used in simulations				
$\|\Sigma_{aa}\|$	1.0	0.51	0.51	0.51	0.51
$\|\Sigma_{Aa}\|$	1.0	0.51	1.0	1.0	1.0
R_{aa}	0.0	0.7	0.7	0.7	0.7
R_{Aa}	0.0	0.7	0.0	0.7	0.7
σx_{aa}	1.0	1.0	1.0	1.0	1.0
σy_{aa}	1.0	1.0	1.0	1.0	1.0
σx_{Aa}	1.0	1.0	1.0	$1/\sqrt{0.51}$	1.0
σy_{Aa}	1.0	1.0	1.0	1.0	$1/\sqrt{0.51}$
	Mean absolute errors of the parameter estimates				
δr	62.0 ± 2.3	33.8 ± 2.0	18.3 ± 1.0	25.4 ± 1.4	19.8 ± 1.1
δd_x	140.0 ± 6.2	76.4 ± 4.1	43.1 ± 2.2	59.2 ± 3.1	51.7 ± 2.8
δd_y	36.7 ± 1.9	36.8 ± 2.1	32.9 ± 1.8	35.6 ± 1.8	40.1 ± 2.3
$\delta\sigma x_{aa}$	36.0 ± 1.9	20.2 ± 1.2	11.9 ± 0.7	17.5 ± 1.0	17.2 ± 1.0
$\delta\sigma y_{aa}$	13.9 ± 0.7	11.8 ± 0.7	10.8 ± 0.6	11.9 ± 0.7	14.1 ± 0.8
δR_{aa}	22.3 ± 1.3	16.2 ± 0.9	12.6 ± 0.6	12.2 ± 0.6	11.5 ± 0.6
$\delta\sigma x_{Aa}$	38.8 ± 1.9	21.7 ± 1.2	15.5 ± 0.8	15.3 ± 0.9	13.7 ± 0.8
$\delta\sigma y_{Aa}$	13.6 ± 0.7	11.5 ± 0.7	10.6 ± 0.6	11.6 ± 0.7	10.4 ± 0.6
δR_{Aa}	21.9 ± 1.2	14.8 ± 0.9	17.1 ± 0.9	10.4 ± 0.6	11.1 ± 0.6

[a] The mean absolute errors of the estimates are multiplied by 1000

with increased k, the contribution of non-genetic components to the variance of each of the traits, x and y, and to covariance between them will decrease at a rate dependent on the heritability coefficients h_x^2 and h_y^2. Let first $k = 1$. If r_g and r_e are the coefficients of genetic and non-genetic correlation, respectively, between x and y, the same for both groups AA and aa, then the within-group phenotypic correlation, R_{xy}, will be:

$$R_{xy} = r_g h_x h_y + r_e \sqrt{(1 - h_x^2)(1 - h_y^2)}.$$

With $k > 1$, $R_{xy} = r_g h_x h_y + r_e \sqrt{(1 - h_x^2)(1 - h_y^2)}$,

where $h_x'^2 = kh_x^2/[(k-1)h_x^2 + 1]$, $h_y'^2 = kh_y^2/[(k-1)h_y^2 + 1]$.

The new values of the within-group variances can be written as

$$\sigma_x'^2 x = \sigma^2 x[h_x^2 + (1 - h_x^2)/k;$$

$$\sigma_y'^2 y = \sigma^2 y[h_y^2 + (1 - h_y^2)/k].$$

Some numerical examples of the effect of parameter values on the efficiency of two-trait analysis are shown in Table 4.

Several conclusions follow from the presented material: (1) with increased family size (k), the estimates are more accurate; (2) lower h^2 for the within-group variances ($\sigma^2 AA$ and $\sigma^2 aa$) caused by genes of other chromosomes results in lower errors of the estimates. This seemingly paradoxical result can easily be understood if we recall that the reducing effect of increased family size k on variation between families is higher when h^2 is small; (3) for any combination of parameters, an increase in R_{xy} leads to a higher resolution, and provided any fixed h_x^2, h_y^2 and R_{xy}, the higher the r_g the better is the accuracy of estimation with increased k.

F_2 progeny

Two situations will be considered here, depending on dominance relations in the QTL.

1. *Full dominance at the QTL*

$$S_{mm}(x, y) = S_1(x, y) = (1 - r)^2 f_1(x, y) + [1 - (1 - r)^2] f_2(x, y),$$

Table 4 The dependence of mean absolute errors (δ) of the parameter estimates in two-trait analysis on correlation (R) between the traits in the case of recombinant inbred lines. Simulated were 100 replicates, each with $n = 200$ lines and k individuals per line; distances between QTL groups were $d_x = x_{AA} - x_{aa} = 1$ and $d_y = y_{AA} - y_{aa} = 0$, recombination rate between the QTL and marker $r = 0.1$. Two levels of heritability h^2 for the "within QTL group variation" were considered to be the same for both traits (x and y), with $h^2 = h^2 \in \{0.25, 0.5\}$ at $k = 1$; in this case the values of other parameters were as follows: $\sigma x_{AA} = \sigma x_{aa} = \sigma y_{AA} = \sigma y_{aa} = 1$, $R_{AA} = R_{aa} = R \in \{0, 0.3, 0.6\}$, the genotypic correlation coefficient $rg_{AA} = rg_{aa} \in \{0, 0.6, 0.9\}$. With $k > 1$ (e.g. 5 or 20) $R'_{xy} = r_g h_x h_y + r_e \sqrt{(1 - h_x^2)(1 - h_y^2)}$, where $h_x'^2 = kh_x^2/[(k-1)h_x^2 + 1]$, and $h_y'^2 = kh_y^2/[(k-1)h_y^2 + 1]$. The new values of the within QTL groups phenotypic variances will be $\sigma'^2 x = \sigma^2 x[h_x^2 + (1 - h_x^2)/k]$ and $\sigma_y'^2 = \sigma_y^2[h_y^2 + (1 - h_y^2)/k]$

h^2	k		Variable parameter values used in simulations				
		R	0.0	0.0	0.3	0.6	0.6
		r_g	0.0	0.6	0.6	0.6	0.9
			Mean absolute errors of the parameter estimates				
0.25	1	δr	73.4 ± 4.7	73.4 ± 4.7	70.5 ± 4.7	53.6 ± 3.8	53.6 ± 3.8
		δd_x	348 ± 18.1	348 ± 18.1	340 ± 19.3	287 ± 19.9	287 ± 19.9
		$\delta\sigma x_{AA}$	95.9 ± 6.5	95.9 ± 6.5	95.8 ± 6.3	79.8 ± 5.9	79.8 ± 5.9
		$\delta\sigma y_{AA}$	57.4 ± 5.9	57.4 ± 5.9	56.3 ± 4.4	53.0 ± 4.9	53.0 ± 4.9
	5	δr	40.5 ± 3.5	34.6 ± 3.2	28.0 ± 2.7	20.6 ± 1.7	13.5 ± 1.1
	20		25.3 ± 2.5	16.6 ± 1.5	15.0 ± 1.4	13.9 ± 1.4	7.2 ± 0.5
	5	δd_x	171 ± 15.7	150 ± 13.9	135 ± 12.3	102 ± 9.5	71.9 ± 6.8
	20		93 ± 10.9	61.6 ± 5.7	61.8 ± 6.7	58.5 ± 6.5	24.5 ± 2.1
	5	$\delta\sigma x_{AA}$	64.4 ± 4.9	55.3 ± 4.5	48.8 ± 4.0	38.1 ± 3.4	29.1 ± 2.8
	20		41.8 ± 4.0	28.3 ± 2.3	28.2 ± 2.5	26.9 ± 2.5	12.8 ± 1.1
	5	$\delta\sigma y_{AA}$	27.2 ± 2.6	24.7 ± 2.2	24.8 ± 2.2	22.7 ± 2.0	17.2 ± 1.8
	20		19.4 ± 2.1	16.5 ± 1.5	16.4 ± 1.5	15.5 ± 1.4	6.5 ± 0.6
0.5	5	δr	55.4 ± 4.0	50.0 ± 3.9	46.7 ± 3.7	42.5 ± 3.3	32.6 ± 2.7
	20		52.2 ± 3.9	38.8 ± 3.9	36.3 ± 3.0	34.6 ± 2.9	13.1 ± 1.1
	5	δd_x	244 ± 16.5	235 ± 17.6	228 ± 17.3	214 ± 16.4	183 ± 15.4
	20		225 ± 16.8	194 ± 15.9	182 ± 15.5	176 ± 15.0	79.4 ± 7.5
	5	$\delta\sigma x_{AA}$	77.8 ± 5.0	72.2 ± 5.0	67.8 ± 5.1	63.9 ± 5.0	54.3 ± 4.7
	20		74.2 ± 4.9	59.6 ± 4.9	56.3 ± 4.8	54.2 ± 4.6	28.9 ± 3.0
	5	$\delta\sigma y_{AA}$	38.1 ± 3.6	36.8 ± 3.4	36.7 ± 3.5	35.1 ± 3.0	31.5 ± 2.7
	20		34.8 ± 3.2	31.3 ± 2.7	31.3 ± 2.6	31.1 ± 2.6	17.8 ± 2.0

$$S_{Mm}(x, y) = S_2(x, y) = r(1-r)f_1(x, y) + [1 - r(1-r)]f_2(x, y),$$

$$S_{MM}(x, y) = S_3(x, y) = r^2 f_1(x, y) + (1 - r^2)]f_2(x, y),$$

where $f_1(x, y) = f_{aa}(x, y)$ and $f_2(x, y) = f_A(x, y)$.

2. General case: all three genotypes are different at the QTL

Then:

$$S_{mm}(x, y) = S_1(x, y) = (1-r)^2 f_1(x, y) + 2\alpha(r)f_2(x, y) + r^2 f_3(x, y),$$

$$S_{Mm}(x, y) = S_2(x, y) = \alpha(r)f_1(x, y) + [(1 - 2\alpha(r)]f_2(x, y) + \alpha(r)f_3(x, y),$$

$$S_{MM}(x, y) = S_3(x, y) = r^2 f_1(x, y) + 2\alpha(r)f_2(x, y) + (1-r)^2 f_3(x, y),$$

where $\alpha(r) = r(1-r)$. The effect of correlation on the efficiency of the estimation procedure in the above situations was studied with the following parameter values: $r = 0.1$, $d_x = x_{AA} - x_{aa} = \{1 \text{ and } 2\}$, $d_y = 0$, $\sigma_{1x} = \sigma_{2x} = \sigma_{3x} = 1$, $\sigma_{1y} = \sigma_{2y} = \sigma_{3y} = 1$, $R_{1xy} = R_{2xy} = R_{3xy} = \{0, 0.6 \text{ and } 0.9\}$. The data obtained are presented in (Table 5). Note, that the data generated according to dominance assumption $f_{AA}(x, y) = f_{Aa}(x, y)$ were analysed by the full dominance model (1) as well as by full model (2).

Different comparisons could be done on this data set, and for all of the parameters the main tendency within each of the models can readily be seen: an increase in resolution of marker-QTL analysis resulting from the measurements of an additional trait (y) correlated with the main trait (x).

Noteworthy also are some other effects. Dominance at the QTL ($h_x = 0$) leads to much lower deviations of the parameter estimates from their expectations [provided the assumption $h_x = 0$ is taken into account, i.e. model (1) is applied], with the size of the effect being dependent on d_x/σ_x and R_{xy}. The reason for this is clear. When $f_{Aa}(x, y)$ coincides with either $f_{aa}(x, y)$ or $f_{AA}(x, y)$ then the resolution is proportional to the distance between the densities $f_{aa}(x, y)$ and $f_{AA}(x, y)$, all other things being equal. When $f_{Aa}(x, y)$ is intermediate between $f_{aa}(x, y)$ and $f_{AA}(x, y)$ then it "dilutes" the initial difference [between $f_{aa}(x, y)$ and $f_{AA}(x, y)$]. Thus, it can easily be seen from Table 5 that in every case with QTL dominance and $d = x_{AA} - x_{aa}$ the precision of the parameter estimates is approximately the same as in the corresponding non-dominance case with only half of the above difference d (compare the cases with $d_x = 1$ and $h_x = 0$ and those with $d_x = 2$ and $h_x = 0.5d_x = 1$). Note also a reduction in the precision if in the case of full dominance ($hx = 0$) model (2) instead of (1) is used in the analysis (which will try to make a resolution into three components instead of two).

Discussion

It seems reasonable to execpt an increase in resolution of the QTL-marker linkage estimation procedures with increased discrepancy between the QTL groups. Usually, the effect of the quantitative trait gene (s) on the mean value of the trait in question is the target of such an analysis. Therefore, difference measures like $(x_{AA} - x_{aa})/$

Table 5 The dependence of mean absolute errors (δ)[a] of the parameter estimates in the two-trait analysis on correlation (R) between the traits in the case of F_2 progeny. Two hundred replicates, each of size $n = 1000$, were simulated at recombination rate $r = 0.1$, distance between QTL groups $d_x = x_{AA} - x_{aa} \in \{1, 2\}$, $h_x = x_{Aa} - x_{aa} \in \{0.5, 1\}$, $d_y = y_{AA} - y_{aa} = 0$, $h_y = y_{Aa} - y_{aa} = 0$, $\sigma x_{AA} = \sigma x_{Aa} = \sigma x_{aa} = 1$, $\sigma y_{AA} = \sigma y_{Aa} = \sigma y_{aa} = 1$. In the ML-functional all σ's were assumed to be different, as were R_{AA}, R_{Aa} and R_{aa}. The variant designation is as following: F the condition $h_x = 0.5d_x$ was used in data simulation (i.e. additive action of A/a is assumed) while in the ML-functional h_x was independent on d_x, Q $h_x = 0$ in the data simulation, and in the ML-functional h_x was an independent parameter, Q_d the same as Q but with h_x known

$R_{AA} = R_{Aa} = R_{aa} = R$		Variable parameter values used in simulations					
		0		0.6		0.9	
d_x		1	2	1	2	1	2
		Mean absolute errors of the parameter estimates					
F	δr	99.7 ± 4.1	61.0 ± 2.5	90.7 ± 3.4	39.6 ± 2.1	36.8 ± 2.3	8.2 ± 0.5
	δh_x	180 ± 11.7	166 ± 7.6	145 ± 9.6	104 ± 5.3	61.6 ± 3.6	44.5 ± 2.3
	δd_x	319 ± 18.6	295 ± 12.4	256 ± 14.1	169 ± 9.1	83.6 ± 5.1	57.9 ± 3.2
Q	δr	60.9 ± 2.8	15.4 ± 0.8	44.5 ± 2.3	9.7 ± 0.5	11.4 ± 0.6	6.1 ± 0.3
	δh_x	97.7 ± 5.6	69.3 ± 3.7	73.4 ± 4.1	65.9 ± 3.3	60.5 ± 3.2	56.1 ± 3.0
	δd_x	148 ± 6.7	79.0 ± 4.1	108 ± 5.1	64.1 ± 3.3	52.6 ± 2.9	46.1 ± 2.8
Q_d	δr	49.9 ± 2.5	13.8 ± 0.7	35.1 ± 2.0	8.0 ± 0.4	9.1 ± 0.5	2.5 ± 0.2
	δd_x	149 ± 7.7	65.1 ± 3.7	101 ± 4.9	38.9 ± 2.0	34.9 ± 2.3	9.4 ± 0.5

[a] The mean absolute errors of the estimates are multiplied by 1000

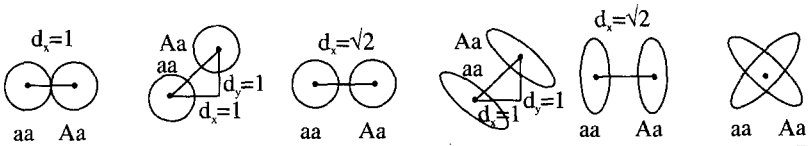
σ_x are of interest when the resolution power of tests for QTL detecting and procedures of parameter estimation are considered. We suggested earlier that a strong increment in a discrepancy between QTL groups in segregating populations could be achieved when employing joint distribution of a set of correlated QTs (Korol et al. 1987, 1994). Consequently, a serious gain in performance of QTL-marker analysis is expected on this basis.

As we have seen above, with high correlations between quantitative traits, a good resolution is possible even if the QTL groups (say, *aa* and *Aa*) are strongly overlapping for their marginal distributions. The advantage of the multi-trait analysis may be especially attractive when the QT factor (locus) in question influences several traits simultaneously. This corresponds to situations of pleiotropy. Another possible class of situations where this analysis is relevant is close linkage between different loci affecting related traits. It may be especially important when a block of economically important genes is going to be transferred from one species to another and the recombination within the block become suppressed (e.g. Rick 1972; Zhuchenko and Korol 1985; Haley et al. 1993).

What are the reasons for the increased resolution when correlation between the traits affected by the target QTL is taken into account. They can easily be understood from the following example of backcross situations presented in Table 6. Here the locus *A/a* affects both traits, x and y, with $d_x = d_y = d$; in both groups, *Aa* and *aa*, $\sigma_x = \sigma_y = \sigma$ and $R_{xy} = R \leq 0$. It is clear that even in the case of zero correlation R, the resolution will be better if the second trait is taken into account (compare columns 1 and 2). The reason is the increased distance between the group centres [e.g. $\sqrt{d_x^2 + d_y^2} = \sqrt{2}d$ in the example presented in column 2]. Therefore, we can expect in this case that two-trait analysis with zero correlation is equivalent to the signle-trait situation with the distance between the group means multiplied by $\sqrt{2}$ and the variances unchanged. These assumptions were tested using simulated data. The results shown in Table 6 confirm the expectations (compare δr in columns 2 and 3). Let us discuss now the effect of correlation between x and y. Geometrical considerations allow us to assume that the resolution of our backcross bivariate mixtures $S_{mm}(x, y)$ and $S_{Mm}(x, y)$ into components $f_{aa}(x, y)$ and $f_{Aa}(x, y)$ can be reduced to a single-trait resolution problem, with a new trait, x', being the first main component of the system (x, y) (provided that the respective ellipses in *Aa* and *aa* are parellel). This formulation will be equivalent to the initial one if we put $d_x = \sqrt{d_x^2 + d_y^2} = \sqrt{2}d$, and for both groups, *Aa* and *aa* $\sigma_x = \sigma_x\sqrt{1+R}$. The closeness of columns 4 and 5 in Table 6 with respect to δr indicates that this expectation holds as well.

A reduction of the within-group variance seems to be the most important factor causing the increase in accuracy when correlated traits are involved in the analysis. Nevertheless, as was shown in Table 3, a decrease in correlation between the traits in one of the groups, say *Aa*, does not necessarily reduce the accuracy of the parameter estimates. On the contrary, an increase in variation within the *Aa* group manifested in an increased norm of the covariance matrix $|\Sigma_{Aa}| > |\Sigma_{aa}|$ may result in an *increased* precision, no matter what factor caused the increase in $|\Sigma_{Aa}|$: reduced correlation Rxy_{Aa} or increased variances $\sigma^2 x_{Aa}$ or $\sigma^2 y_{Aa}$. Thus, it is reasonable to assume, as we did in single-trait analysis (Korol et al. 1994), that the resolution capacity of the marker analysis in the case of two correlated quantitative traits depends on the discrepancy between the bivariate distributions $f_{aa}(x, y)$ and $f_{Aa}(x, y)$, $D(f_{aa}(x, y), f_{Aa}(x, y)$.

Table 6 Demonstration of the equivalence, with respect to the resolution power of the marker-QTL linkage analysis, of the effect of correlation between the traits within the QTL groups and the effect of reduced within-group variance (a backcross progeny case is considered). Two hundred replicates, each of size $n = 1000$, were simulated at recombination rate $r = 0.1$. According to the explanation given in the text, one will expect the same δr values for the pairs of columns: 2–3, and 4–5

	Variable parameter values used in simulations					
R_{aa}	0	0	0	−0.8	0	−0.8
R_{Aa}	0	0	0	−0.8	0	0.8
d_x	1	1	$\sqrt{2}$	1	$\sqrt{2}$	0
d_y	0	1	0	1	0	0
$\sigma x_{Aa} = \sigma x_a$	1	1	1	1	$\sqrt{0.2}$	1
$\sigma y_{Aa} = \sigma y_a$	1	1	1	1	$\sqrt{1.8}$	1
Presentation of the case						
	Mean absolute errors of the parameter estimates					
δr	62.0 ± 2.3	33.0 ± 2.0	35.5 ± 1.8	4.8 ± 0.3	5.2 ± 0.3	9.7 ± 0.5
δd_x	140 ± 6.2	75.0 ± 4.1	103 ± 5.0	20.1 ± 1.0	7.6 ± 0.4	14.4 ± 0.7
δd_y	36.7 ± 1.9	74.8 ± 4.2	35.6 ± 1.8	20.5 ± 0.9	24.4 ± 1.3	14.3 ± 0.8
$\delta\sigma x_{aa}$	36.0 ± 1.9	21.3 ± 1.2	36.5 ± 1.9	6.8 ± 0.4	7.6 ± 0.4	5.0 ± 0.3
$\delta\sigma y_{aa}$	13.9 ± 0.7	20.9 ± 1.2	12.0 ± 0.5	7.3 ± 0.4	7.8 ± 0.4	4.6 ± 0.3
δR_{aa}	22.4 ± 1.3	35.7 ± 2.1	19.6 ± 1.0	5.9 ± 0.3	14.9 ± 0.8	7.0 ± 0.4
$\delta\sigma x_{Aa}$	38.8 ± 1.9	19.9 ± 1.2	36.6 ± 1.8	7.2 ± 0.4	7.7 ± 0.4	5.0 ± 0.3
$\delta\sigma y_{Aa}$	13.6 ± 0.7	22.5 ± 1.2	12.0 ± 0.6	7.2 ± 0.4	7.9 ± 0.4	4.6 ± 0.3
δR_{Aa}	21.9 ± 1.2	38.4 ± 2.1	19.8 ± 1.1	5.8 ± 0.3	15.6 ± 0.8	6.7 ± 0.4

While we did not calculate the effect of increased $|\Sigma_{Aa}|$ on $D(f_{aa}, f_{Aa})$, this assumption seems to be a reasnable explanation of the estimates presented in Table 3. This means, that the analogy with the method of main components is rather approximate here, which can easily be demonstrated by the formal example shown in the last column of Table 6. Here the mixtures $S_{mm}(x, y)$ and $S_{Mm}(x, y)$ are resolved into the component densities f_{aa} and f_{Aa}, which differ from each other only due to differences in Rxy.

The multi-trait approach could be no less important in situations where the trait of interest (say x) is dependent on the QTL in question (A/a) and is strongly correlated with another trait (y), the latter being independent of A/a. As we could see, in this case too the additional information provided by measurements of y increases dramatically the mapping precision of the locus A/a, no matter what type of mapping population was considered (backcross, F_2 etc).

Such correlations may be caused by the segregation of other genes, environmental effects and physiological limitations. As examples, trait pairs like "grain weight – protein level" or "milk production – fat content" could be mentioned. Due to the high cost of molecular marker typing, many quantitative traits are usually measured within one experiment, so that the needed structure of data for the proposed QTL mapping approach is not an exception.

Another application is studies of reaction norms of the trait of interest to different environmental conditions. The rationality of such an approach is due to the well-known dependence of the quantitative trait expression on environment. The estimates of QTL effects are usually based on the identification of segregating progeny under some ecological conditions and may be entirely different from those obtained under other conditions. The phenomenon of genotype-environment interaction can be analysed based on marker approach. The appropriate analysis may reveal QTL affecting: (1) the 'developmental potential' of a trait under optimal environmental conditions; (2) the stability of the trait in the face of limiting conditions (with little or no effect under optimum environment); (3) both the 'developmental potential' and stability simultaneously (Korol et al., 1994). In addition to such a formulation, if a set of genotypes (e.g., RILs or vegetative clones of an F_2) could be tested under different conditions, then the resulting measurements of one trait in these conditions can be considered as a multi-trait set (Falconer, 1981) and treated by our procedures. This will result in a higher efficiency of information extraction from the data than the usual single-trait analysis can provide.

It is clear that the multi-trait approach of marker-QTL linkage analysis could be also used within the framework of interval mapping of QTL. Our results (Korol et al., in preparation) show that this is indeed a promising way to elevate the resolution power of the interval procedures. It is worth mentioning that multivariate analysis has been successfully used to increase the precision of QTL mapping (e.g. Jansen and Stam 1994; Zeng 1994). However, our multi-trait approach opens an additional and yet unexploited possibility for further increase in resolution.

Acknowledgements The authors are grateful to the unknown referee for useful comments and suggestions. We acknowledge with thanks the support of the Israeli Ministry of Science (grant No 3675-1-91) and Wolfson Family Charitable Trust.

References

Arunachalam V (1981) Genetic distance in plant breeding. Indian J Genet Plant Breed 41:226–236

Boehnke M, Moll P (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 EA, Asians MJ, Baselga M, Balansard E, Gerig TM (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, Williamson J (1991) Linkage analysis of quantitative traits: increased power by using selected samples. Am J Hum Genet 49:786–796

Darvasi A, Soller M (1992) Selective genotyping for determination of linkage between a marked locus and a quantitative trait locus. Theor Appl Genet 85:353–359

Darvasi A, Weinreb A, Minke V, Weller JI, Soller M (1993) Detection marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics 134:943–951

Demenais E, Lathrop GM, Lalouel JM (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

Falconer DS (1981) Introduction to quantitative genetics. Longman, London

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315–324

Haley SD, Miklas PN, Stavely JR, Byrum J, Kelly J (1993) Identification of RAPD markers linked to a major rust resistance gene block in common bean. Theor Appl Genet 86:505–512

Jansen RC (1993) Interval mapping of multiple quantitative trait loci. Genetics 135:205–211

Jansen RC, Stam P (1994) High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136:1447–1455

Kleijnen JPC (1974) Statistical techniques in simulation. Marcel Dekker, New York

Knott SA, Haley CS (1992) Aspects of maximum likelihood methods for mapping of quantitative trait loci in line crosses. Genet Res 60:139–151

Korol AB, Preygel IA, Preygel SI (1994) Recombination variability and evolution. Chapman & Hall, London

Lande R, Thompson R (1990) Efficiency of marker assisted selection in the improvement of quantitative traits. Genetics 124:743–756

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

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

Lin CY (1978) Index selection for genetic improvement of quantitative characters. Theor Appl Genet 52:49–56

Luo ZW, Woolliams JA (1993) Estimation of genetic parameters using linkage between a marker gene and a locus underlying a quantitative character in F_2 populations. Heredity 70:245–253

Martinez O, Curnow, RN (1992) Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theor Appl Genet 85:480–488

Morton NE (1955) Sequential tests for the detection of linkage. Am J Hum Genet 7:277–318
Motro U, Soller M (1993) Sequential sampling in determining linkage between marker loci and quantitative trait loci. Theor Appl Genet 85:658–664
Rick CM (1972) Further studies on segregation and recombination in backcross derivatives of a tomato species hybrid. Biol Zentralbl 91:209–200
Simpson P (1989) Detection of linkage between quantitative trait loci and restriction length polymorphisms using inbred lines. Theor Appl Genet 77:815–819
Soller M, Beckmann JS (1990) Marker-based mapping on quantitative trait loci using replicated progeny. Thoer Appl Genet 80:205–208
Soller M, Genizi A (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
Thoday JM (1967) Genes in the study of continuous variation. Ciencia é cultura 19:54–63
Titterington DM, Smith AFM, Makov UE (1985) Statistical analysis of finite mixture distributions. Wiley, Chichester
Weller JI, Wyler A (1992) Power of different sampling strategies to detect quantitative trait loci variance effect. Theor Appl Genet 83:582–588
Zeng Z-B (1984) Precise mapping of quantitative trait loci. Genetics 136:1457–1468
Zhuchenko AA, Korol AB (1985) Recombination in evolution and breeding. Nauka Press, Moscow