

Multi-interval mapping of correlated trait complexes

A. B. KOROL*†, Y. I. RONIN†, E. NEVO† & P. M. HAYES‡

†*Institute of Evolution, University of Haifa, Mount Carmel, Haifa 31905, Israel* and ‡*Crop Science Department, Oregon State University, Corvallis, OR 97331, U.S.A.*

Employing Monte Carlo simulations, we demonstrate the advantages of multitrait analysis in detection of linked QTL effects within the framework of mixture models. In spite of an increased number of parameters to be estimated, compared to the single-trait formulation, the proposed method allows for an improvement of detection power and estimation precision of linked QTLs in both adjacent or nonadjacent intervals, with coupling and repulsion effects. The results obtained are illustrated by examples based on data of the North American Barley Genome Mapping Project.

Keywords: correlation, linked QTLs, LOD-score, maximum likelihood, multiple traits.

Introduction

Many efforts have been devoted in the last decade to increase the efficiency of marker analysis of quantitative traits. Among the most effective approaches are interval analysis (Lander & Botstein, 1989; Knott & Haley, 1992), selective sampling (Lebowitz *et al.*, 1987; Darvasi & Soller, 1992), replicated progeny testing (Soller & Beckmann, 1990) and sequential experimentation (Motro & Soller, 1993). Especially encouraging are the recent successful attempts to improve the efficiency of QTL mapping by taking into account simultaneous segregation at many genomic segments affecting the trait in question (Jansen & Stam, 1994; Zeng, 1994; Jansen, 1996). A complementary situation, when one QTL (or a chromosome segment) affects several traits simultaneously, can also be considered, resulting in increased resolution power (Korol *et al.*, 1987, 1994, 1995; Preygel & Korol, 1989; Jiang & Zeng, 1995; Ronin *et al.*, 1995). Such analysis may be of major importance in formulating marker-assisted breeding strategies, dissecting heterosis as a multilocus multitrait phenomenon, obtaining unbiased parameter estimates of QTL effects in selective genotyping of correlated traits, developing an optimized programme for evaluation and conservation of genetic resources, revealing the genetic architecture of fitness systems in natural populations, etc. The combination of the multi-interval and multitrait mapping strategies may help to cope with a difficult

problem arising when the chromosome under consideration contains several QTLs (e.g. Jiang & Zeng, 1995). It is known that if one attempts to fit a single-locus model to such a case, a ghost QTL can be found in an interval which has no effect on the trait (Knott & Haley, 1992; Martinez & Curnow, 1992). It is especially difficult to recognize situations with *trans* effects of linked QTLs (Haley & Knott, 1992; Luo & Kearsey, 1992).

Employing Monte Carlo simulations, we demonstrate here the advantages of multitrait analysis in detection of linked QTL effects within the framework of mixture mapping models. In spite of an increased number of parameters to be estimated, compared to the single-trait formulation, the proposed method allows for an improvement of detection power and estimation accuracy of linked QTLs. The results are illustrated by examples based on data of the North American Barley Genome Mapping Project.

Bivariate mixture model with two linked QTLs

Formulation of the model

Consider a situation of two linked loci, A/a and B/b , residing in two marker intervals, $M_{11}/m_{11}-M_{12}/m_{12}$ and $M_{21}/m_{21}-M_{22}/m_{22}$, and affecting correlated quantitative traits, x and y . We will demonstrate that joint treatment of correlated traits may provide a better power of detection and higher precision of parameter estimation for linked QTLs than the usual

*Correspondence. E-mail: korol@esti.haifa.ac.il

single-trait analysis. The consideration will be confined to a dihaploid or backcross situation, although other types of mapping populations can be treated in a similar way. We assume additive gene effects across loci and traits but this restriction can easily be omitted. Generally, our analysis is free of the standard simplifying assumption of equal variances (covariances) in the QTL groups. As was shown elsewhere (Korol *et al.*, 1996), this last assumption can reduce the resolution power of the QTL mapping analysis. This may confer certain limits to the employment of the usual regression mapping models based on the 'no variance effect' assumption.

For the case of two linked QTLs in a dihaploid mapping population, four bivariate density distribution functions should be specified, for each allelic combination at the linked QTLs, $f_{AB}(x,y)$, $f_{Ab}(x,y)$, $f_{aB}(x,y)$ and $f_{aa}(x,y)$. The expected mean values of the traits and variance-covariance matrices will be denoted by μ_{uAB} , μ_{uAb} , μ_{uAB} and μ_{uab} ($u = x$ or y), and Σ_{AB} , Σ_{Ab} , Σ_{aB} , Σ_{ab} , respectively. Each matrix Σ_i specifies the residual variances and covariance of the pair x , y : σ_x^2 , σ_y^2 and $\text{COV}_{xy} = R_i \sigma_x^2 \sigma_y^2$, caused by segregation of genes from other chromosomes and nongenetic factors (e.g. environmental heterogeneity). Based on marker scores and measurements of the traits x and y , we should contrast, for the considered chromosome, the hypotheses that one (H_1) or two (H_2) intervals affect the observed variation of x and y and compare both with H_0 ('no effect of the chromosome tested'). Clearly, H_1 is a complex hypothesis: for any pair of intervals one can assume that either the first or the second interval has no effect on the traits of interest. Actually, the situation is even more complicated, because a series of 'partial' hypotheses should be considered ranging from 'full' H_2 (both traits depend on both QTLs) to H_0 (no effect of the QTLs on either of the two traits). For the dihaploid case, the two-interval consideration results in 16 marker groups. The expected joint distribution of the traits x and y in each is a mixture of four densities, $f_{AB}(x,y)$, $f_{Ab}(x,y)$, $f_{aB}(x,y)$ and $f_{aa}(x,y)$:

$$h_i(x,y) = \pi_{i1}f_{AB} + \pi_{i2}f_{Ab} + \pi_{i3}f_{aB} + \pi_{i4}f_{aa}, \quad \sum_j \pi_{ij} = 1, i = \overline{1,16},$$

where the size of the group i and mixture proportion π_{ij} depend on relative positions of the putative QTLs with respect to marker loci, recombination rates, and interference mode and level (Jiang & Zeng, 1995). Some reasonable assumptions can be made leading to a simplification of the analysis arising from a reduction in the number of marker groups and/or f -components within groups (see below). For

an arbitrary individual of the mapping population we present its bivariate phenotype (x,y) as

$$x = \mu_x + 0.5da_x g_a + 0.5db_x g_b + e_x$$

$$y = \mu_y + 0.5da_y g_a + 0.5db_y g_b + e_y, \quad (1)$$

where x and y are the individual's phenotype scores of the analysed traits, μ_x and μ_y are trait means, da_x and da_y are the effects of substitution at the A/a locus with respect to mean values of x and y (i.e. $da_x = \mu_{x_{AA}} - \mu_{x_{aa}}$ and $da_y = \mu_{y_{AA}} - \mu_{y_{aa}}$), g_a denotes the genotype at locus A/a ($g_a = -1$ for aa and 1 for AA); the same symbol usage is applied to locus B/b . In general, one may assume that the putative QTLs affect not only the mean values of the traits but also the trait variances and covariance. In such a case, e_x is a random variable with zero mean and variances σ_{11u}^2 , σ_{12u}^2 , σ_{21u}^2 and σ_{22u}^2 for $\{(g_a, g_b)\} = \{(-1, -1), (-1, 1), (1, -1), (1, 1)\}$, $u = x$ or y . The variables e_x and e_y are assumed to be correlated with correlation coefficients R_{11xy} , R_{12xy} , R_{21xy} and R_{22xy} for $\{(g_a, g_b)\} = \{(-1, -1), (-1, 1), (1, -1), (1, 1)\}$, respectively. Correlation between the traits x and y within the QTL groups may be caused by other segregating QTLs or nongenetic correlation. Although we may assume that loci A/a and B/b can also affect trait variances and covariance, in most cases we will deal mainly with the situation of equal variance-covariance matrices in the QTL groups, $\Sigma_{AB} = \Sigma_{Ab} = \Sigma_{aB} = \Sigma_{ab} = \Sigma$.

LOD-score test and parameter estimation

The log-likelihood for a sample of two-dimensional measurements x_k, y_k in marker groups with sizes N_i ($i = \overline{1,16}$) can be written as:

$$\ln L(\Theta_{n_2}) = \sum_{i=1}^{16} \sum_{k=1}^{N_i} \ln h_i(x_k, y_k), \quad (2)$$

where Θ_{n_2} is the vector of genetic parameters characterizing the effects and positions of the putative QTLs. In the general case, $da_u \neq 0$, $db_u \neq 0$ ($u = x$ or y), and all σ_{iju}^2 are different as well as all R_{jxy} , so that $\Theta_{n_2} = \{r_1, r_2, \mu_x, \mu_y, da_x, da_y, db_x, db_y, \sigma_{11x}^2, \sigma_{12x}^2, \sigma_{21x}^2, \sigma_{22x}^2, \sigma_{11y}^2, \sigma_{12y}^2, \sigma_{21y}^2, \sigma_{22y}^2, R_{11}, R_{12}, R_{21}, R_{22}\}$ is the vector of $n_2 = 20$ unknown parameters, specifying recombination rates in the pair of trial intervals and joint distributions of traits x and y in the QTL groups. The assumption of no effect of genes from the intervals $M_{11}/m_{11}-M_{12}/m_{12}$ and $M_{21}/m_{21}-M_{22}/m_{22}$ on the traits (x,y) can be presented by another set of parameters, $\Theta = \Theta_{n_0} = \{\mu_x, \mu_y, \sigma_x, \sigma_y, R\}$ (the null hypothesis $\{H_0: \Theta = \Theta_{n_0}\}$) as contrasted to the foregoing 'full' $H_2: \Theta = \Theta_{n_2}$, or to any of the 'partial' alterna-

tives $\{H_1: \Theta = \Theta_{n1}\}$. According to the likelihood ratio test approach (Wilks, 1962), if H_0 is true, the statistic

$$\chi^2 = 2\ln[\max_{\Theta_{n1} \in S_1} L(\Theta_{n1}) / \max_{\Theta_{n0} \in S_0} L(\Theta_{n0})] \quad (3)$$

is distributed asymptotically as chi-squared 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). The same method is applied to compare H_2 and H_1 . If χ^2 exceeds some critical value, corresponding to a preset level α , then the H_1 hypothesis can be rejected. In such a case, the numerical values providing the maximum to $L(\Theta_{n2})$ could be considered as ML-estimates of the parameters characterizing our putative QTL loci A/a and B/b . However, in multi-interval mapping the problem of the exact asymptotic distribution of the test statistic remains unsolved even for the single-trait analysis (see Zeng, 1994). This is especially true when linked QTLs are considered (Lander & Botstein, 1989). If so, one could use extensive Monte Carlo simulations to obtain an empirical critical value of the statistics for each situation.

One more comment on multi-interval mapping models of correlated trait complexes is worth mentioning (see also Korol *et al.*, 1995). Introduction of additional parameters specifying the QTL mapping model should be justified statistically by comparison to the corresponding 'reduced' model. This is relevant to any complication of the mapping model, the replacement both of single-trait mapping analysis by its multitrait analogue and of a single-interval model by a two-interval one. Parameters which do not affect the significance level should be removed from the model.

Monte Carlo simulations

Generating the data

For each situation studied, 200 repeated mapping populations were generated using pseudorandom numbers. A bivariate normal distribution was used for the trait groups $AABB$, $aaBB$, $AAbb$ and $aabb$. The compositions of the marker groups (mixtures h_i , $i = \overline{1,16}$) were modelled as four-component distributions. The length of the marker interval was 20 cM with the QTLs in the middle of their intervals. No double exchanges were assumed within the intervals in the data presented below (hence Morgan's mapping function is suitable). The simulated chromosomes consisted of eight intervals each, with QTLs residing in intervals 3 and 4 or 3 and 5.

Obtaining numerical solutions

Optimization was by modified gradient and Newton methods. The possibility of multiple maxima was tested by optimizing a few test cases using various starting points. In all cases only a single maximum was found. Thus, for all Monte Carlo experiments the simulated parameter sets were used as starts (Titterton *et al.*, 1985). When dealing with real data on barley, multiple random initial points were used to provide the unique solution.

Estimation of the power of the test

To estimate the power of the log-likelihood ratio test we used the critical level of the statistics (eqn 3) $\chi^2 = \chi^2_{\text{critical}}$ based on the asymptotic distribution (chi-squared with d.f. = $n_1 - n_0$ when H_0 vs. H_1 is tested, or d.f. = $n_2 - n_1$ for H_1 vs. H_2 comparisons). The goodness of fit of the expected distribution was tested by simulations using 5000 trials. Provided H_1 is true, the proportion of cases where the second QTL is revealed when it really exists was measured for different situations using critical values obtained in these simulations. A similar approach was also proposed by Doerge & Churchill (1996) on the basis of a permutation test. We found that the asymptotic and simulated distributions result in close estimates of power.

Simulation results

For the dihaploid case, we have simulated and analysed several situations when two QTLs (A/a and B/b) residing in adjacent or nonadjacent intervals affect two correlated traits, x and y . In order to show the advantages of joint analysis of correlated traits, we compare power of the test for detection of both QTLs and accuracy of parameter estimates with those obtained for situations with no correlation between the involved quantitative traits.

Four basic configurations were analysed, with the two QTLs residing in adjacent (**AD**) and nonadjacent (**NA**) intervals and acting in the same direction (coupling phase, **CP**) and in opposite ones (repulsion phase, **RP**), correspondingly. Following are the results obtained for several possible combinations of the traits and QTLs involved.

(i) One of the traits, x , depends on both QTLs, A/a and B/b , whereas the correlated trait y is independent of these QTLs (see also Korol *et al.*, 1994, 1995; Ronin *et al.*, 1995). The results of scanning along possible pairs of intervals for all four configurations (adjacent and nonadjacent locations of the QTLs, each at coupling and repulsion phase)

are presented in Table 1. Based on these data the following conclusions can be made. The maximum of the average LOD scores of the two-QTLs model is attained at the true pair of the intervals. Moreover, the modal class of the bivariate distribution of the individual LODs also corresponds to the true pair of intervals (see the part of Table 1 which lies below the diagonal). Although the correlated trait y does not depend on either of the two QTLs, the additional information provided by y allows, as expected, the LOD values to increase. Moreover, the differences between LODs corresponding to the true and neighbouring positions of the QTLs are also increased when the correlated trait y is taken into account. These results hold for both adjacent and nonadjacent configurations, independently of the phase (coupling or repulsion).

How are these effects reflected in the test power and estimation precision? Table 2 illustrates the gains of the two-trait analysis compared to the single-trait analysis. These are manifested in: (a) increase in power of detection of any QTL activity in the marked chromosome, as reflected in the differences between the mean LOD values for H_2 and H_0 ; (b) higher power of discrimination between H_2 and H_1 . The benefit of two-trait analysis is higher for NA configurations (nonadjacent location of the QTLs), and for the repulsion phase as compared to the coupling one; (c) reduced biases of parameter estimates, manifested mainly for the coordinates of the QTLs within the intervals; and (d) lower variances in all parameter estimates.

(ii) A/a affects both of the traits whereas B/b affects only one trait, x ; within each of the four QTL groups x and y are correlated because of nongenetic mechanisms and segregation of genes from other chromosomes. Comparisons of the test power and estimation accuracy enable us to conclude that bivariate mapping analysis is superior to the single-trait one (Table 3). This is manifested in: (a) a higher power of QTL detection (compare LODs for H_2 vs. H_0) and discrimination between 'two-QTLs' and 'single-QTL' hypotheses, i.e. H_2 vs. H_1 ; and (b) lower variances of parameter estimates. Notably, the reduction in the variance for recombination distances is more pronounced for the QTL affecting both traits as compared to that affecting only one of the traits. Also, a slight reduction is found in biases of the estimates of QTL positions resulting from the bivariate analysis. But in contrast to the case (i), the bivariate model has no advantages with respect to the biases of the QTL effects (Table 3).

(iii) Both A/a and B/b have pleiotropic effects on both traits, x and y . All conclusions reached for the

previous case hold also here (see the results in Fig. 1 and Table 4). For instance, correlation between the analysed traits leads to a pronounced increase in the proportion of the modal class which corresponds to the true pair of intervals (compare (b) vs. (a) and (d) vs. (c) in Fig. 1. In addition, we compare here two estimates of the power of the test ' H_2 vs. H_1 ' (the assumption of the presence of two QTLs contrasted to that of one). These estimates were obtained when the critical value of the test was calculated from the asymptotic chi-squared approximation (eqn 3), and from 5000 Monte Carlo simulations of the situation when H_1 is true. It appeared that the resulting two estimates of the power are close (compare β_t and β_{mk} in the last column of Table 4).

Examples from the barley 'Steptoe \times Morex' dataset

As an example of the application of multi-interval two-trait analysis, we have chosen two economically important traits — malt extract (%) and alpha amylase activity — from the Steptoe \times Morex barley mapping data set (Hayes *et al.*, 1996). The Steptoe \times Morex population is an extensively described mapping and QT reference population (Hayes *et al.*, 1993, 1994). Malting quality is determined by a number of component traits. Malting is a carefully controlled germination process in which complex proteolytic pathways are manipulated to develop an ideal substrate for subsequent fermentation. Kernel carbohydrates are hydrolysed by α -amylases. Malt extract percentage is a measure of soluble sugars and proteins and thus expresses the overall efficiency of the malting process. Thus, α -amylase may be a component of malt extract, and coincident QTLs for the two traits could be attributed to pleiotropic effects of α -amylase. This is probably the case on chromosome 1, where malt extract and α -amylase QTLs were detected in the vicinity of the *Amy2* locus (Hayes *et al.*, 1993).

In Table 5 some examples of two-trait mapping analysis of α -amylase and malt extract are presented. For the measurements in environment 4, previous analysis revealed a QTL in interval no. 12 of chromosome 2 (*ABG14-His3C*) (Hayes *et al.*, 1994). Two-interval analysis of the same trait allows the detection of a second significant QTL in interval no. 2 (*ABG703-CHS1B*). Although two-trait two-interval analysis results in the same power of detection of a second interval affecting α -amylase, it gives a slightly different location for both QTLs. The same conclusions were reached using data from

Table 1 Variation of bivariate LOD scores across interval pairs in the case of two linked QTLs affecting a quantitative trait when another trait correlated with the target one and independent of the chromosome under consideration is involved

$j \downarrow$	R	QTL in adjacent intervals 3 and 4						QTL in nonadjacent intervals 3 and 5							
		$i \rightarrow 1$	2	3	4	5	6	1	2	3	4	5	6	7	
1	CP	0.0	4.5	11.9	21.1	21.7	14.7	8.1	3.3	8.6	14.7	17.2	17.0	12.5	7.4
		0.7	6.2	17.3	32.1	32.9	21.3	11.3	4.6	13.3	22.0	25.8	25.5	17.7	9.7
	RP	0.0	0.9	1.2	1.4	2.0	2.1	1.4	1.1	2.0	2.4	2.5	3.8	3.7	2.3
		0.7	1.2	1.7	2.0	2.3	3.4	2.2	1.6	3.7	4.2	3.9	6.1	5.9	3.5
2	CP	0.0	0	11.8	21.3	22.8	18.6	14.1	0	8.3	14.8	18.1	19.2	16.0	11.7
		0.7	0	17.2	32.6	35.1	27.8	20.7	0	13.0	22.1	28.3	30.2	24.7	17.4
	RP	0.0	1	1.4	2.1	3.4	3.4	2.2	0	2.1	2.4	4.4	5.9	5.4	3.5
		0.7	0	2.0	3.5	5.8	5.7	3.5	0	3.8	4.3	7.8	10.8	9.9	6.4
3	CP	0.0	2	0	21.1	23.2	22.9	21.7	0	0	14.4	18.6	20.6	19.3	16.5
		0.7	0	0	31.9	36.1	35.2	32.9	0	0	21.6	29.3	32.6	29.7	24.6
	RP	0.0	1	3	1.3	3.8	3.4	2.1	0	0	2.1	5.9	7.5	6.1	3.7
		0.7	0	0	1.9	6.9	5.9	3.4	0	0	3.9	10.1	13.5	11.3	6.9
4	CP	0.0	1	24	37	21.2	21.5	21.3	0	5	0	16.3	19.1	18.7	17.4
		0.7	0	9	75	31.9	32.9	32.7	0	1	2	24.3	29.3	28.2	25.6
	RP	0.0	1	17	58	1.3	2.0	1.3	0	0	5	1.0	6.1	4.8	2.5
		0.7	0	10	77	1.9	3.4	2.0	0	0	0	1.8	10.4	8.4	4.4
5	CP	0.0	0	0	28	2	12.6	12.7	0	10	75	5	15.4	15.7	15.8
		0.7	0	0	15	0	17.9	32.2	0	5	95	0	22.9	27.5	23.2
	RP	0.0	0	6	9	0	1.4	1.1	0	0	85	0	2.1	2.4	2.4
		0.7	1	2	7	0	2.0	1.6	0	0	100	0	3.5	4.1	3.8
6	CP	0.0	0	0	3	3	0	4.9	0	0	5	0	0	10.0	10.4
		0.7	0	0	0	1	0	6.4	0	0	0	0	0	14.0	14.4
	RP	0.0	0	3	0	0	1	0.9	0	0	10	0	0	2.1	2.0
		0.7	0	3	0	0	0	1.2	0	0	0	0	0	3.5	3.4

The results of Monte Carlo simulations (100 runs) of a doubled haploid mapping population with 200 genotypes. A single chromosome was modelled with two QTLs (A/a and B/b) of equal additive effects in coupling (CP) and repulsion (RP) phases. The distance between consecutive markers was 20 cM, QTLs were located in the middle of adjacent or nonadjacent intervals (nos 3 and 4, or nos 3 and 5). The effects of each of A/a and B/b were $da_x = db_x = 0.5$; $da_y = db_y = 0$; the residual standard deviations $\sigma_{jx} = \sigma_{jy} = 0.5$; and residual phenotypic correlation coefficient $R = \{0, 0.7\}$. The data under the diagonal are the observed frequencies of outcomes in which the maximum LOD appeared in the indicated pair of intervals; the data above the diagonal are the mean values of the two-interval LODs. The cases $i = j$ represent the results of application of a 'single-QTL model'.

environment no. 14. However, in this case the detectability of the second QTL for α -amylase by single-trait analysis is rather poor ($P \approx 0.27$) whereas the significance for the second QTL in the two-trait analysis is $P \approx 0.037$. The next example in Table 5

concerns the pair ' α -amylase–kernel weight' in environments 9, 10 and 13, for chromosome 1. Usual single-trait interval analysis (see Hayes *et al.*, 1994) reveals a QTL for α -amylase at interval 10 (*Brz-ABC156D*) or 11 (*ABC156D-ABG22A*) mani-

Table 2 Efficiency of bivariate two-interval mapping of two linked QTLs affecting a quantitative trait assisted by another trait correlated with the target one and independent of the chromosome under consideration

	R	LOD		da_x	db_x	r_1	r_2	β_t	
		$H_2 H_0$	$H_1 H_0$						$H_2 H_1$
CP	0.0	23.31	21.78	1.53	0.53	0.49	42.91	79.98	9
					0.21	0.22	14.34	21.65	
	0.7	36.16	32.74	3.42	0.52	0.49	47.63	74.30	55
					0.14	0.14	7.94	9.71	
AD	0.0	3.96	2.04	1.92	0.46	-0.45	42.03	72.51	20
					0.22	0.23	13.35	14.08	
	0.7	6.99	2.84	4.1	0.48	-0.48	45.93	73.45	73
					0.15	0.14	10.06	10.67	
CP	0.0	20.20	16.53	3.67	0.48	0.53	46.40	89.98	59
					0.09	0.10	10.55	7.49	
	0.7	31.79	24.44	7.35	0.49	0.53	48.85	90.49	97
					0.07	0.06	5.85	2.06	
NA	0.0	8.23	3.33	4.90	0.47	-0.46	48.86	91.02	85
					0.10	0.10	3.61	11.05	
	0.7	14.17	5.01	9.16	0.48	-0.47	50.07	89.79	100
					0.08	0.06	2.85	4.03	

The results of simulations experiments are described in Table 1. **AD** and **NA** indicate 'adjacent' and 'nonadjacent' intervals, respectively. LOD here denotes mean value of the maximum lods across all pairs of intervals averaged over 100 runs. All these lods were obtained for three pairs of contrasted hypotheses: $H_2 | H_0$ (two QTLs vs. no QTL), $H_2 | H_1$ (two QTLs vs. a single QTL), and $H_1 | H_0$ (a single QTL vs. no QTL). The standard deviations of the parameters are presented under the mean values: da_x and s_{dax} , db_x and s_{dbx} , etc., r_1 and r_2 are the recombination distances of the QTLs (in cM), starting from the first marker position; β_t is the test power for detecting a second QTL ($H_2 | H_1$) based on chi-squared approximation for the calculation of the critical value of test statistics under H_1 (here and in Tables 3 and 4 the chosen significance level was 0.1 per cent).

festing in environments 9 and 10. In environment 13, in addition, a factor from interval 15 (*ABC455-Amy2*) was also detected (Hayes *et al.*, 1994). Two-interval single-trait analysis shows a significant ($P = 0.007$) additional effect of the interval no. 21 (*ABG461-Cat3*) in environment 9; the putative QTL is in repulsion phase with that of interval 10. Incorporation of the second trait, kernel weight, into the mapping model increases the significance ($P = 0.002$). Especially pronounced enhancement in detection power of the test for the presence of the QTL for α -amylase in interval 21 was found for environment 10: from $P = 0.045$ in single trait analysis to $P = 0.003$ in bivariate analysis.

Concerning the detection of the effect of interval 21 on α -amylase, the situation for environment 13 is similar to those in environments 9 and 10. However, in the last case the two-interval analysis is complicated by the presence of an additional QTL at interval 15. It is noteworthy that when the interval 10 (or 11) was ignored, its effect was ‘absorbed’ by the interval 15, so that the two-interval analysis for the pair of intervals 15 and 21 results in a biased (upward) estimation of the effect of interval 15. This can be seen when interval pair 10 and 15 is considered (the last row of Table 5).

Coincident QTLs may result from tight linkage or pleiotropy. In these terms, there are three possible

Table 3 Efficiency of bivariate two-interval mapping of linked QTLs, when the first of the correlated QTs depends on both QTLs (*A/a* and *B/b*) whereas the second trait depends only on one of the QTLs (*A/a*)

	<i>R</i>	LOD			<i>da_x</i>		<i>da_y</i>		<i>db_x</i>		<i>r₁</i>		<i>r₂</i>		β_i
		<i>H₂ H₀</i>	<i>H₁ H₀</i>	<i>H₂ H₁</i>	<i>s_{dax}</i>	<i>s_{dax}</i>	<i>s_{dbx}</i>	<i>s_{r1}</i>	<i>s_{r1}</i>	<i>s_{r2}</i>	<i>s_{r2}</i>				
CP	0.0	17.32	15.78	1.54	0.34	0.36	0.37	46.24	80.45	6					
					0.23	0.09	0.22	12.35	15.64						
AD	0.7	36.57	33.65	2.92	0.34	0.35	0.37	49.38	74.60	43					
					0.15	0.07	0.14	6.22	11.35						
RP	0.0	6.93	5.16	1.77	0.32	0.36	-0.32	48.25	79.50	16					
					0.20	0.09	0.19	13.03	17.04						
CP	0.7	13.30	10.28	3.02	0.32	0.36	-0.33	49.35	78.09	46					
					0.15	0.07	0.14	6.83	13.14						
NA	0.0	15.66	12.81	2.85	0.32	0.37	0.40	47.65	93.66	40					
					0.09	0.10	0.09	10.17	13.88						
RP	0.7	33.14	27.57	5.57	0.31	0.37	0.40	51.21	87.77	90					
					0.07	0.06	0.05	3.03	7.21						
RP	0.0	9.18	5.84	3.34	0.33	0.37	-0.32	51.21	98.13	54					
					0.10	0.10	0.11	9.90	11.35						
	0.7	18.94	13.02	5.92	0.32	0.37	-0.31	50.72	91.15	89					
					0.07	0.06	0.06	2.83	7.65						

All details are the same as described in Tables 1 and 2. The only difference is in the QTL effects. The effects of *A/a* and *B/b* were $da_x = db_x = da_y = 0.375$ and $db_y = 0$.

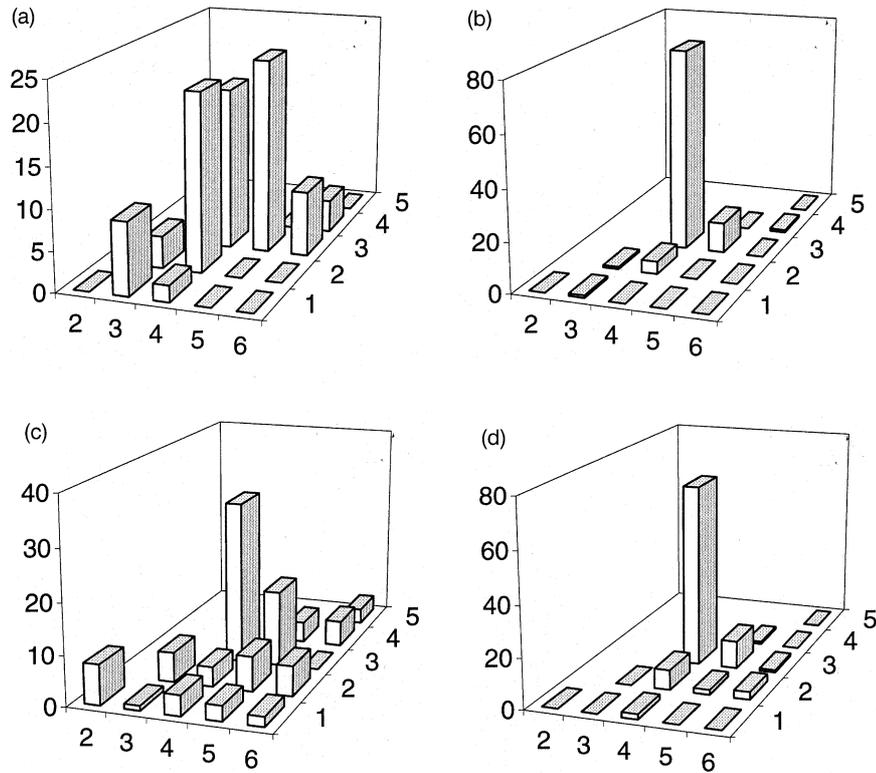


Fig. 1 Variation of LOD scores across interval pairs in the case when both of the correlated quantitative traits depend on two linked QTLs (residing in adjacent intervals nos 3 and 4). For most of the parameter values used in simulations see Table 4; (a) $R = 0$, coupling phase; (b) $R = 0.7$, coupling phase; (c) $R = 0$, repulsion phase; (d) $R = 0.7$, repulsion phase.

scenarios involving malt extract and α -amylase: (i) coincident malt extract and α -amylase QTLs, where the latter is a determinant of the former, as on chromosome 1; (ii) malt extract QTL without accompanying α -amylase QTL, where, by default, the observed extract is attributed to a factor other than α -amylase, such as another enzyme or starch structure/composition; or (iii) an α -amylase QTL with an accompanying malt extract QTL.

Based on single-interval mapping, chromosome 2 appears to be an example of scenarios (ii) and (iii) as described above: α -amylase QTL at LOD >2.0 is seen near the centromere, and malt extract QTLs with LOD >2.0 are seen on the short arm (Hayes *et al.*, 1993). However, with a two-trait analysis, the malt extract QTLs are observed to coincide with α -amylase QTLs (see Table 5). This provides an additional and potentially valuable insight into the basis of the observed malt extract effect, transforming it from a type (ii) to a type (i) scenario. These results provide new tools for studying the biochemistry of the malting process, as barley α -amylases are of two types (the *Amy1* and *Amy2* groups) and loci determining the isoforms map to chromosomes 6 and 1, respectively. No α -amylase loci have been mapped to chromosome 2, so the

observed α -amylase QTLs may be attributable to genes that somehow regulate or modulate α -amylase expression.

Discussion

An approach to increase the resolution power of interval mapping of QTLs was proposed earlier based on analysis of correlated trait complexes (Korol *et al.*, 1987, 1994, 1995; Preygel & Korol, 1989; Jiang & Zeng, 1995; Ronin *et al.*, 1995). It is well known that, in an attempt to fit a single-locus mapping model to a case *with several QTLs*, a QTL can be found in an interval which actually does not affect the considered trait (Haley & Knott, 1992; Martinez & Curnow, 1992). As a result, the estimated effect of this ghost locus could be much higher than that of any of the real QTLs in the chromosome. An opposite and even more difficult situation could be when the chromosome in question contains a couple of linked QTLs in repulsion phase. Then, a conclusion of 'no effect' on the considered trait may result from single-interval mapping analysis. That *trans*-association of QTLs could be a rather common phenomenon even in interspecific crosses has been demonstrated by DeVicente & Tanksley

(1993) in tomato: they found that up to 36 per cent of the detected QTLs had alleles with effects opposite to the direction expected from the parental differences.

The usual way of dealing with several linked QTLs is multiple regression analysis or mixture-model interval analysis which include markers as regression-derived cofactors to account for segregation of QTLs on the same chromosome (Jansen & Stam, 1994; Jiang & Zeng, 1995). Employing Monte

Carlo simulations, we have demonstrated here the advantages of multitrait analysis in the detection of linked QTLs within the framework of mixture models. In spite of an increased number of parameters to be estimated, the proposed method allows for an improvement in the correct detection and estimation of linked QTLs. Although our model deals with bivariate trait distributions, multivariate situations can also be considered without a necessity for a further increase in the number of parameters.

Table 4 Efficiency of bivariate two-interval mapping of linked QTLs, when both of the correlated QTs depend on both QTLs (*A/a* and *B/b*)

<i>R</i>	LOD			<i>da_x</i>	<i>da_y</i>	<i>db_x</i>	<i>db_y</i>	<i>r₁</i>	<i>r₂</i>	<i>β_t</i>	
	<i>H₂ H₀</i>	<i>H₁ H₀</i>	<i>H₂ H₁</i>	<i>s_{dax}</i>	<i>s_{dax}</i>	<i>s_{dbx}</i>	<i>s_{dbx}</i>	<i>s_{r1}</i>	<i>s_{r2}</i>	<i>β_{mk}</i>	
CP	0.0	16.09	14.56	1.53	0.26	0.27	0.27	0.26	40.13	78.37	2
					0.22	0.21	0.23	0.21	17.24	18.61	2
	0.7	35.90	32.11	3.79	0.24	0.29	0.29	0.24	48.73	72.71	50
					0.16	0.13	0.15	0.13	6.13	10.07	46
AD	0.0	3.51	1.77	1.74	0.22	0.22	-0.22	-0.22	40.64	79.30	8
					0.25	0.29	0.25	0.26	21.07	22.57	3
	0.7	7.33	2.81	4.52	0.24	0.27	-0.24	-0.27	46.58	74.83	69
					0.15	0.15	0.14	0.14	10.58	11.61	61
NA	0.0	13.86	11.48	2.38	0.27	0.28	0.26	0.25	48.70	90.82	15
					0.17	0.13	0.18	0.13	15.86	16.19	12
	0.7	31.55	24.21	7.34	0.26	0.27	0.28	0.26	49.49	89.35	95
					0.09	0.09	0.09	0.06	5.84	4.72	97
RP	0.0	6.09	2.83	3.26	0.28	0.27	-0.28	-0.27	49.38	91.86	42
					0.17	0.20	0.18	0.18	15.48	15.68	37
	0.7	14.34	5.10	9.24	0.26	0.27	-0.26	-0.28	49.76	89.87	100
					0.10	0.09	0.09	0.09	5.85	5.62	100

Most details are the same as described in Tables 1 and 2. The QTL effects were equal: $da_x = db_x = da_y = db_y = 0.265$. Two estimates of the power of the test $H_2 | H_1$ (two QTLs vs. a single QTL) were compared here: (i) β_t , based on the chi-squared approximation for calculation of the critical value of the test statistics under H_1 , and (ii) β_{mk} , based on the critical value derived from the distribution of the test statistics when the two-QTLs model was applied to 5000 Monte Carlo experiments with one segregating QTL (i.e. H_1 was true).

Table 5 Some examples of two-interval mapping analysis on barley data of NABGMP (Steptoe/Morex dihaploid mapping population)

	Int	LOD	da_x	db_x	σ_x	da_y	db_y	σ_y	R	$P_{H2 H1}$	(d.f.)		
Single-trait analysis: α -amylase													
Example 1	12	4.96	—	2.9	3.4								
Chr2–Env4	2,12	7.77	2.0	2.8	3.3					0.0015	(2)		
	3,11	7.06	—	2.8	3.4	0.5	—	1.4	0.62				
	3,11	9.28	1.9	2.4	3.3	0.9	—	1.4	0.61	0.0012	(1)		
Single-trait analysis: α -amylase													
Example 2	12	4.19	—	3.4	4.1								
Chr2–Env14	6,12	4.68	1.4	2.6	4.1					0.270	(2)		
	5,12	6.09	—	3.1	4.1	0.9	—	1.4	0.27				
	5,12	7.03	1.3	2.7	4.0	1.0	—	1.3	0.31	0.037	(1)		
Single-trait analysis: α -amylase													
Example 3	10	5.26	3.5	—	4.2								
Chr1–Env9	10,21	7.41	3.7	–2.5	4.0					0.007	(2)		
	10,21	6.21	3.5	—	4.2	–1.0	–0.4	2.4	–0.13				
	10,21	8.25	3.6	–2.3	4.0	–0.4	–0.8	2.4	–0.13	0.002	(1)		
Single-trait analysis: α -amylase													
Example 4	10	6.79	4.1	—	4.2								
Chr1–Env10	11,21	8.14	4.5	–2.2	4.0					0.045	(2)		
	11,21	7.23	4.3	—	4.2	0.2	–0.6	3.1	–0.23				
	11,21	8.99	4.5	–2.3	4.0	0.2	–0.5	3.1	–0.24	0.003	(1)		
	Int	LOD	da_x	db_x	dc_x	σ_x	da_y	db_y	dc_y	σ_y	R	$P_{H2 H1}$	(d.f.)
Single-trait analysis: α -amylase													
Example 5	11	7.95	4.4	—	—	3.9							
Chr1–Env13	11,21	9.33	4.5	—	–1.7	3.8						0.038	(2)
	11,21	10.83	4.5	—	—	3.9	–1.8	—	–2.7	4.2	–0.39		
	11,21	11.98	4.7	—	–1.6	3.8	–1.8	—	–2.0	4.2	–0.39	0.021	(1)

Table 5 Continued

Int	LOD	da_x	db_x	σ_x	da_y	db_y	σ_y	R	$P_{H_2 H_1}$	(d.f.)
Single-trait analysis: α -amylase										
15	6.99	—	4.0	—	4.0			0.010	(2)	
15,21	8.98	—	4.3	-2.0	3.9					
Two-trait analysis: α -amylase – kernel weight										
15,21	9.82	—	4.0	—	4.0	—	-1.2	-2.7	4.2	-0.42
15,21	11.54	—	4.2	-2.0	3.9	—	-1.4	0.006	(1)	
10,15	8.85	2.4	2.6	—	-3.9	-1.2	-0.9	-1.7	4.2	-0.41
								—	4.2	-0.31

Most designations are the same as in Table 2; **Chri-Envj** indicate the chromosome number and environmental number according to Hayes *et al.* (1994); **Int** refers to the interval number(s) which appeared to manifest local maximum of the LOD score; a , b and c refer to QTLs from interval 10 (or 11), 15 and 21, respectively. For each case of a single-trait analysis, we compare the results of a single-interval and two-interval analysis. The analogous comparison for the two-trait analysis was conducted in the following way. The fourth rows in the Table represent the full model with QTL effects of both intervals. Then, in the full model the effect of one of the intervals on the first trait was assumed to be equal to zero, resulting in some change of the optimal values of the remaining parameters and the LODs. The two models obtained (those in rows 3 and 4) were compared to evaluate the significance of the tested QTL effect.

This seems to be possible if for any pair of intervals one can use the first two principal components of the multivariate complex (Korol *et al.*, 1994; Ronin *et al.*, 1995; Weller *et al.*, 1996).

Acknowledgements

We acknowledge with thanks the comments and suggestions of two anonymous referees. The work was supported by the Israeli Ministry of Absorption and the Ancell-Teicher Research foundation for Genetics and Molecular Evolution.

References

- DARVASI, A. AND SOLLER, M. 1992. Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. *Theor. Appl. Genet.*, **85**, 353–359.
- DEVICENTE, M. C. AND TANKSLEY, S. D. 1993. QTL analysis of transgressive segregation in an interspecific tomato cross. *Genetics*, **134**, 585–596.
- DOERGE, R. W. AND CHURCHILL, G. A. 1996. Permutation tests for multiple loci affecting a quantitative character. *Genetics*, **142**, 285–294.
- HALEY, C. S. AND KNOTT, S. A. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, **69**, 315–324.
- HAYES, P. M., LIU, B., KNAPP, S. J., CHEN, F., JONES, B., BLAKE, T. *ET AL.* 1993. Quantitative trait locus effects and environmental interaction in a sample of North American barley germplasm. *Theor. Appl. Genet.*, **87**, 392–401.
- HAYES, P. M., IYAMABO, I. AND THE NABGMP. 1994. Summary of QTL effects in the Steptoe \times Morex population. *Barley Genet. Newsl.*, **23**, 98–143.
- HAYES, P. M., BRICENO, G. AND MATTHEWS, D. 1996. The Steptoe \times Morex barley mapping population. <http://wheat.pw.usda.gov/graingenes.html>
- JANSEN, R. C. 1996. A general Monte Carlo method for mapping multiple quantitative trait loci. *Genetics*, **142**, 305–311.
- JANSEN, R. C. AND STAM, P. 1994. High resolution of quantitative traits into multiple loci via interval mapping. *Genetics*, **136**, 1447–1455.
- JIANG, C. AND ZENG, Z.-B. 1995. Multiple trait analysis and genetic mapping for quantitative trait loci. *Genetics*, **140**, 1111–1127.
- KNOTT, S. A. AND HALEY, C. S. 1992. Aspects of maximum likelihood methods for mapping of quantitative trait loci in line crosses. *Genet. Res.*, **60**, 139–151.
- KOROL, A. B., PREYGEL, I. A. AND BOCHARNIKOVA, N. I. 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., PREYGEL, I. A. AND PREYGEL, S. I. 1994. *Recombination Variability and Evolution*. Chapman and Hall, London.
- KOROL, A. B., RONIN, Y. I. AND KIRZHNER, V. M. 1995. Interval mapping of quantitative trait loci employing correlated trait complexes. *Genetics*, **140**, 1137–1147.
- KOROL, A. B., RONIN, Y. I., TADMOR, Y., BAR-ZUR, A., KIRZHNER, V. M. AND NEVO, E. 1996. Estimating variance effect of QTL: an important prospect to increase the resolution power of interval mapping. *Genet. Res.*, **67**, 187–194.

- LANDER, E. S. AND BOTSTEIN, D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, **121**, 185–199.
- LEBOWITZ, B. J., SOLLER, M. AND BECKMANN, J. S. 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.
- LUO, Z. W. AND KEARSEY, M. J. 1992. Interval mapping of quantitative trait loci in an F₂ population. *Heredity*, **69**, 236–242.
- MOTRO, U. AND SOLLER, M. 1993. Sequential sampling in determining linkage between marker loci and quantitative trait loci. *Theor. Appl. Genet.*, **85**, 658–664.
- PREYGEL, I. A. AND KOROL, A. B. 1989. Marker analysis of quantitative characters. *Uspekhi Sovremennoy Genetiki (Advances in Modern Genet.) (USSR)*, **16**, 82–95 (in Russian).
- RONIN, Y. I., KIRZHNER, V. M. AND KOROL, A. B. 1995. Linkage between loci of quantitative traits and marker loci: multi-trait analysis with a single marker. *Theor. Appl. Genet.*, **90**, 776–786.
- SOLLER, M. AND BECKMANN, J. S. 1990. Marker-based mapping of quantitative trait loci using replicated progeny. *Theor. Appl. Genet.*, **80**, 205–208.
- TITTERINGTON, D. M., SMITH, A. F. AND MAKOV, U. 1985. *Statistical Analysis of Finite Mixture Distributions*. Wiley, Chichester.
- WELLER, J. I., WIGGANS, G. R., VAN RADEN, P. M. AND RON, M. 1996. Application of a canonical transformation to detection of quantitative trait loci with the aid of genetic markers in a multi-trait experiment. *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.