Theor Appl Genet (1998) 97: 1169–1178

Y. I. Ronin · A. B. Korol · J. I. Weller

Selective genotyping to detect quantitative trait loci affecting multiple traits: interval mapping analysis

Received: 23 February 1998 / Accepted: 15 May 1998

Abstract Segregating quantitative trait loci can be detected via linkage to genetic markers. By selectively genotyping individuals with extreme phenotypes for the quantitative trait, the power per individual genotyped is increased at the expense of the power per individual phenotyped, but linear-model estimates of the quantitative-locus effect will be biased. The properties of single- and multiple-trait maximum-likelihood estimates of quantitative-loci parameters derived from selectively genotyped samples were investigated using Monte-Carlo simulations of backcross populations. All individuals with trait records were included in the analyses. All quantitative-locus parameters and the residual correlation were unbiasedly estimated by multiple-trait maximum-likelihood methodology. With single-trait maximum-likelihood, unbiased estimates for quantitative-locus effect and location, and the residual variance, were obtained for the trait under selection, but biased estimates were derived for a correlated trait that was analyzed separately. When an effect of the QTL was simulated only on the trait under selection, a "ghost" effect was also found for the correlated trait. Furthermore, if an effect was simulated only for the correlated trait, then the statistical power was less than that obtained with a random sample of equal size. With multiple-trait analyses, the power of quantitative-trait locus detection was always greater with selective genotyping.

Communicated by J. W. Snape

Y. I. Ronin · A. B. Korol (✉)
Institute of Evolution, University of Haifa,
Mt Carmel, Haifa 31905, Israel
Tel.: (972)-48240-449
Fax: (972)-48246-554
E-mail: korol@esti.haifa.ac.il

J. I. Weller
Animal Science Institute, ARO, Volcani Center, Israel

Key words Selective genotyping · Multiple traits · QTL · Interval mapping · Maximum likelihood

Introduction

Many studies have shown that the effects of individual quantitative trait loci (QTLs) can be isolated and estimated with aid of linked genetic markers if there is linkage disequilibrium between the genetic markers and the QTLs (Payne 1918; Sax 1923) (reviewed by Soller 1990, 1991, 1994; Korol et al. 1994). Numerous experimental designs have been proposed to generate linkage disequilibrium (reviewed by Weller 1992, 1996). Genotyping for genetic markers is often much more expensive than obtaining data on the quantitative traits of interest, particularly if existing data banks are analyzed. This is generally the case for humans and large animals (Weller et al. 1990). Therefore, several methods have been proposed to increase the statistical power to detect a segregating QTL per individual genotyped at the expense of the power per individual phenotyped for the quantitative trait. These methods include replicate progeny, selective genotyping, sequential sampling, and sample pooling (Lebowitz et al. 1987; Lander and Botstein 1989; Darvasi and Soller 1992, 1994; Motro and Soller 1993; Plotsky et al. 1993). Some of these methods, such as replicate progeny, are independent of the number of traits analyzed, but most are trait-dependent. Thus, their effectiveness decreases as the number of traits analyzed increases.

Selective genotyping (SG), first proposed by Lebowitz et al. (1987), exploits the fact that most of the information for QTL effects is in the "tails" of the quantitative trait distribution. They therefore suggested to first score individuals for the quantitative trait, and then select individuals with extreme values for marker genotyping. SG can increase the statistical power per genotype five-fold as compared to random

genotyping, provided that the number of individuals phenotyped is increased nearly four-fold (Darvasi and Soller 1992). Even if the cost of obtaining trait records is minor with respect to genotyping costs, SG still has significant drawbacks. The estimate of the QTL effect derived from a linear-model analysis will be biased. Darvasi and Soller (1992) derived a formula to obtain an unbiased estimate of the QTL effect, provided that SG is applied to both tails of the distribution. However, this will usually not be the case.

Generally, populations with linked QTLs are the most challenging for analysis. Recently, Lin and Ritland (1996) evaluated the bias in estimating the recombination frequency between linked QTLs based on an analysis of recombination frequency after SG as compared to the recombination frequency before SG. They found that SG can bias estimates of the recombination frequency between linked QTLs–upward when the QTLs are in repulsion phase, and downward when the QTLs are in coupling phase. The bias increased with an increase in the QTL effects.

Nearly all analyses of QTL effects have been based on single-trait analyses, even if multiple traits were analyzed (e.g. Weller et al. 1988; Georges et al. 1995). If multiple correlated traits are analyzed by single-trait analyses, and effects are associated with the same markers on more than a single trait, it cannot be determined whether the effect is due to two different QTLs or to pleiotropic effects of a single locus. Furthermore, the calculation of an experiment-wise type-I error is not straightforward. Weller et al. (1996) suggested that these problems could be solved by a canonical transformation of the quantitative traits; the QTL analysis is then performed on the set of uncorrelated canonical variables. Although this method can be readily applied to any number of correlated traits, it has several significant disadvantages (Weller et al. 1996). Application of a canonical transformation with respect to the phenotypic data does not guarantee the independence of the canonical variables within the QTL groups. This will restrict the application of this approach to SG. Furthermore, the effects of SG for a single trait on QTL-parameter estimates for other correlated traits has not been studied.

As first demonstrated by Weller (1986 a), maximum-likelihood (ML) methodology can be used to derive an unbiased estimate of the QTL effect and the recombination frequency between the QTL and the genetic marker. Surprisingly, for single-trait analyses, ML is not noticeably more powerful than linear-model analyses, whether single- or multiple-linked markers are employed (Simpson 1989, 1992; Haley and Knott 1992; Darvasi et al. 1993). However, Korol et al. (1995) showed that the power to detect a QTL with an ML bi-variate analysis of two correlated traits can be greater than the power obtained with two single-trait analyses, depending on the residual trait correlations and the correlations between the QTL effects (see also Korol et al. 1987, 1994, 1996; Preygel and Korol 1989; Jiang and Zeng 1995; Ronin et al. 1995).

Using ML, and considering all individuals with trait records even if they are not genotyped, it should be possible to obtain asymptotically unbiased (i.e. consistent) estimates of QTL effects with SG (Lander and Botstein 1989). Interval-mapping results from ML analyses of SG populations when more than one QTL controls the target trait have not been previously presented. Furthermore, it is not clear how a single-trait ML will perform if individuals are selected for genotyping based on a single trait, but are analyzed for other, correlated, traits.

It has been previously demonstrated that the power of QTL detection is increased if the QTL also has an effect on the residual variance, provided that the analysis model is appropriate (Korol et al. 1996). Similarly, an effect of the QTL on the residual covariance in a two-trait analysis also increases the detection power, if included into the model (Korol et al. 1995; Ronin et al. 1995). Weller and Wyler (1992) have shown that SG may be an efficient strategy to detect a QTL with a variance effect if some of the individuals with trait values close to the mean are also genotyped. Therefore, it should be of interest to evaluate the efficiency of two-trait analysis combined with SG when the QTL affects both the trait means and variances.

The objectives of the present study were to answer these questions for single- and multi-trait ML based on Monte-Carlo simulations of backcross populations between inbred lines within the framework of standard interval mapping.

Materials and methods

Mixture-model two-trait interval analysis for selective genotyping design

Two-trait interval analysis for SG backcross populations was performed as described by Korol et al. (1995), with the likelihood function modified to include phenotypic data on individuals that were not genotyped

A QTL Q/q flanked by two marker loci, M_1/m_1 and M_2/m_2, with recombination rates r_1 and r_2 in M_1/m_1–Q/q *and* $Q/q - M_2/m_2$ was simulated. For the sake of simplicity, the study was restricted to the 'no interference' case, and the recombination rate between the markers, r, was assumed to be known without error. Thus: $r = r_1 + r_2 - 2r_1r_2$, and r_2 can be computed as a function of r_1 and r. Before selection, the trait values for an individual with genotype g at the QTL ($g = -1$ for qq and $g = 1$ for Qq) was simulated as follows:

$$\begin{aligned} x &= \mu_x + 0.5a_x g + e_x \\ y &= \mu_y + 0.5a_y g + e_y, \end{aligned} \qquad (1)$$

where μ_x and μ_y are the means for traits x and y respectively, a_x and a_y are the substitution effects at the Q/q locus for x and y (i.e. $a_x = \mu_{xQq} - \mu_{xqq}$ and $a_y = \mu_{yQq} - \mu_{yqq}$). The expected joint distributions of the traits x and y in each of the four marker groups of the genotyped part of the population are $U_{m_1m_2}(x, y) = U_1(x, y)$,

$U_{M_1m_2}(x, y) = U_2(x, y)$, $U_{m_1M_2}(x, y) = U_3(x, y)$ and $U_{M_1M_2}(x, y) = U_4(x, y)$, and can be written as:

$$U_i(x, y) = \pi_i f_{qq}(x, y) + (1 - \pi_i) f_{Qq}(x, y), \quad i = 1, \ldots, 4, \tag{2}$$

where $f_{qq}(x, y)$ and $f_{Qq}(x, y)$ are the bivariate normal densities for individuals with a QTL genotype qq and Qq, respectively, and π_i is the probability of the QTL genotype qq conditional on marker genotype i, which is dependent on recombination rates r_1 and r_2. With no interference, $\pi_1 = (1 - r_1)(1 - r_2)/(1 - r)$; $\pi_2 = r_1(1 - r_2)/r$; $\pi_3 = 1 - \pi_2$; and $\pi_4 = 1 - \pi_1$. For the remaining, non-genotyped, individuals the expected distribution will be

$$U_0(x, y) = 0.5 f_{qq}(x, y) + 0.5 f_{Qq}(x, y). \tag{2'}$$

Two-dimensional normal densities were used in our Monte-Carlo mapping experiments to specify the densities $f_{qq}(x, y)$ and $f_{Qq}(x, y)$. In general, $f_{qq}(x, y)$ and $f_{Qq}(x, y)$ may differ not only for bivariate mean values, but also for the residual variances for both traits and for residual correlation coefficients. Hence, for a backcross, the vector of parameters can be presented as $\Theta_{n_1} = (r_1, \mu_{x_1}, \mu_{x_2}, \mu_{y_1}, \mu_{y_2}, \sigma_{x_1}, \sigma_{x_2}, \sigma_{y_1}, \sigma_{y_2}, R_1, R_2)$, where σ_{x_1}, σ_{x_2}, σ_{y_1}, and σ_{y_2} are the residual standard deviations of the traits x and y for genotypes qq and Qq, and R_1 and R_2 are the residual correlation coefficients for these genotypes. The residual correlation between the traits could be caused by genetic factors from other chromosomes (e.g. due to their pleiotropic action on both traits or to linkage) and environmental variation. Assuming no effect of Q/q on trait variances and correlation, this vector reduces to: $\Theta_{n_1} = (r_1, \mu_{x_1}, \mu_{x_2}, \mu_{y_1}, \mu_{y_2}, \sigma_x, \sigma_y, R)$. For the hypothesis of no segregating QTL in the interval M_1/m_1–M_2/m_2 the parameter vector reduces to: $\Theta = \Theta_{n_0} = (\mu_x, \mu_y, \sigma_x, \sigma_y, R)$. The presence of a segregating QTL was tested by a likelihood-ratio test. If the null hypothesis of no segregating QTL is correct, then the likelihood ratio, LR, is:

$$\mathrm{LR} = 2\ln[\max L(\Theta_{n_1})/\max L(\Theta_{n_0})] \tag{3}$$

$$\Theta_{n_1} \in S_1 \qquad \Theta_{n_0} \in S_0$$

and should be distributed asymptotically with a χ^2 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). However, several QTL detection studies have noted that the empirical distribution of LR is not identical to the theoretical distribution (Jansen 1994). Therefore, for each model the empirical distribution of the test statistic under the null hypothesis was estimated by running 2000 simulations. Significance was then determined based on the empirical 5 and 1% critical values. Note that the likelihood function $L(\Theta_{n_1})$ includes terms calculated over both genotyped individuals (equation 2) and non-genotyped individuals (equation 2′).

For the analysis of two linked QTLs and two unlinked QTLs, the mixture model was modified to take account of the joint segregation of the QTLs as described in Korol et al. (1998). The individuals that were not genotyped were included in the likelihood function as a mixture of four densities, instead of the two that are shown in (2′). The mixture proportions depend on the unknown distance between the putative QTLs. These proportions are 0.25 for unlinked QTLs.

Generation of data and analysis

For each combination of parameters, 200 backcross populations of 2000 individuals were simulated using pseudo-random numbers. Bivariate normal distributions with unit variances were generated for the two QTL genotypes, with substitution effects varied from zero to one. QTL genotypes were then determined by binomial sampling based on Equation (2). With SG, unless noted otherwise, the 200 individuals with the highest and 200 individuals with the lowest phenotypic values were genotyped for seven markers spaced at 24-cM intervals along the chromosome. The primary QTL was assumed to be located at position 60 cM on a chromosome of length 144 cM. Thus, the QTL was at the midpoint of the third marker interval. Marker genotypes were determined based on sampling from a binomial distribution, with an assumed recombination frequency 0.19 between adjacent markers (according to the Haldane mapping function for 24-cM intervals). With random genotyping (RG) 400 individuals were randomly selected for genotyping. For simulations with two linked QTLs the second QTL was simulated at position 84 cM in the middle of the fourth interval. Both coupling and repulsion phases relative to the QTL trait effects were simulated. For simulations with two unlinked QTLs the second QTL was simulated at position 84 cM on a second chromosome also of length 144 cM. Seven markers with equal spacing were also genotyped for the second chromosome. The two QTLs simulations were analyzed with both single- and two-QTL models. The same data sets were analyzed by single- and multiple-trait analysis methods. For the single-QTL model, the power of detection of QTL_1 was estimated as the frequency of rejection of H_0 (no linked segregating QTL) based on the empirical 5 and 1% critical value of the test statistics. Similarly, the power of detection of two QTLs vs one QTL (testing H_2 vs H_1 for the two-QTL model) was determined as the frequency of rejection of H_1 as described in Korol et al. (1998).

Obtaining numerical solutions

Optimization was by the modified-gradient and Newton methods. The possibility of multiple maxima was tested by optimizing a few test cases using various sets of starting values. In all cases only a single maximum was found within the parameter space. Therefore, for all other data sets the simulated parameters were used as initial values.

Results

From the schematic plot presented in Fig. 1, it appears that the parameter estimates for the correlated trait will be biased. In the absence of any effect of QTL Q/q on the trait y ($d_y = 0$), SG for trait x may generate a "ghost" effect on trait y, due to the residual correlation between x and y. A similar effect is noted if marker information from the selected individuals is used to construct the map of the markers; direct application of this design leads to the revealing of spurious linkages (Martinez 1996). Furthermore, it is not known how SG for one trait will affect the power to detect QTLs affecting other correlated traits.

Single-trait analysis with two QTLs

Unlinked QTLs

The simulated data were fitted by two models: (1) a single QTL on the chromosome of interest; thus the second unlinked QTL, which had a larger effect, was ignored; (2) a two-QTL model. For comparison we also simulated and analyzed the situation with a single QTL. The results from SG and randomly genotyped samples are presented in Table 1. It can be concluded that:

(1) despite the presence of an unlinked QTL (QTL_2) not included in the model, the power of detection of QTL_1 and the accuracy of the parameter estimates are increased with SG as compared to RG, even though the effect of the unlinked QTL_2 is several-fold greater than that of the target QTL_1. The bias and accuracy of the estimation of the QTL_1 position are not much affected by the unlinked QTL. Thus, SG is apparently robust with respect to the presence of an additional unlinked QTL. The improvement in the power of detection of the target QTL using SG as compared to random genotyping is the same, or even greater than for a single QTL (Lander and Botstein 1989).

(2) these conclusions also hold when the model includes the co-segregating QTL_2 (section b in Table 1). The power of detection of two QTLs (QTL_1 and QTL_2, H_2 hypothesis) versus one QTL (i.e., QTL_2, H_1 hypothesis) is very close to the power of QTL_1 detection using the single-QTL model. This indicates the robustness of ML estimated with SG. The two-QTL model results in a reduced bias for the estimates of σ and a, as well as a narrower confidence interval in the target QTL location.

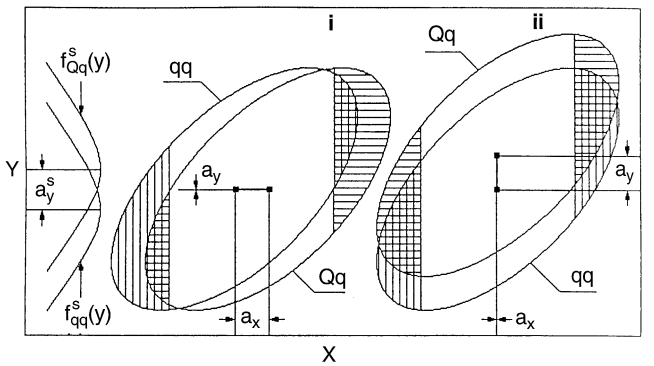

Fig. 1 The effect of selective genotyping in a backcross design when a QTL affects the selected trait x (**i**) or the correlated trait y (**ii**). Projections of the joint (two-dimensional) density function $f(x, y)$ of the traits x and y on the $x - y$ plane are denoted by *elipses* for the alternative QTL groups, qq and Qq. The selected samples at this density are *shaded by vertical lines* for qq individuals and by *horizontal lines* for Qq individuals. In **i**, selection for x induces a "ghost" effect for y. Although the true marginal densities $f_{Qq}(y)$ and $f_{qq}(y)$ have the same mean values, the densities $f^s_{Qq}(y)$ and $f^s_{qq}(y)$ resulting from selection of the tails are shifted relative to each other, with a distance a^s_y

Linked QTLs

The results of two linked QTLs of equal effects, in coupling (CP) and repulsion (RP) phases, residing in adjacent intervals are presented in Table 2. As in the case of non-linked QTLs, the power of detection for linked QTLs is increased relative to random genotyping. For RP, SG increased the power of the 'H_2 vs H_1' test 3–4-fold as compared to RG, and for CP this difference is even greater. In contrast to the results of Lin and Ritland (1996), with ML estimation and including all individuals with trait records, the direction of the bias in the estimated distance between the QTLs does not depend on the linkage phase. The bias of the estimated distance between the two linked QTLs and

Table 1 Increased detection power and robustness to the presence of additional unlinked QTLs of selected genotyping design, using the single-interval mapping model for a single trait affected by two unlinked QTLs. Mean parameter estimates and the power from 200 Monte-Carlo runs are presented; standard errors among runs are given below the means in parentheses. Note that in both models, (1) and (2), the bias and accuracy of the estimation of the target QTL (QTL_1) position in SG design are not much affected by the unlinked QTL, indicating the robustness of ML mapping analysis with SG. However, the two-QTL model results in a reduced bias for the estimates of σ and a, as well as a narrower confidence interval in the target QTL location

a_2^*		L_1	L_2	a_1	a_2	σ	Power%	
							$\alpha \to 5\%$	1%
(1) Fitting a single-QTL model for the 'target' QTL (QTL_1)								
0	SG	60.76		0.257		0.998	92	82
		(1.00)		(0.004)		(0.001)		
	RG	64.33		0.263		0.996	42	24
		(2.15)		(0.010)		(0.001)		
0.25	SG	59.70		0.257		1.008	90	79
		(1.11)		(0.005)		(0.001)		
	RG	65.83		0.278		1.005	45	27
		(1.97)		(0.009)		(0.001)		
1.00	SG	60.74		0.266		1.117	88	72
		(1.39)		(0.005)		(0.001)		
	RG	68.15		0.281		1.113	35	17
		(2.23)		(0.011)		(0.001)		
(2) Fitting a two-QTL model								
0.25	SG	59.25	83.02	0.256	0.252	0.999	91	79
		(0.99)	(1.00)	(0.005)	(0.004)	(0.001)		
	RG	57.46	86.54	0.281	0.266	0.999	44	27
		(1.90)	(1.70)	(0.008)	(0.010)	(0.001)		
1.00	SG	60.14	84.03	0.256	1.004	0.998	88	70
		(1.10)	(0.12)	(0.005)	(0.005)	(0.001)		
	RG	58.07	83.84	0.264	1.004	0.995	41	25
		(1.90)	(0.28)	(0.010)	(0.007)	(0.002)		

Table 2 Increased detection power of a selected genotyping design using a two-interval mapping model for a single trait affected by two linked QTLs. Equal effects of the two QTLs were simulated ($a_1^* = a_2^* = 0.5$) for both repulsion (RP) and coupling (CP) phases. The power of detection is given for the two QTL vs one QTL, that was determined as the frequency of rejection of H_1 (a single segregating QTL) based on the empirical critical value of the test statistics as described in Korol et al. (1998)

Design		L_1	L_2	a_1	a_2	σ	Power %	
							$\alpha \rightarrow 5\%$	1%
RP	SG	54.91 (1.02)	88.77 (0.95)	0.473 (0.022)	− 0.482 (0.022)	1.001 (0.002)	94	83
CP	SG	53.73 (1.37)	88.41 (1.15)	0.492 (0.022)	0.512 (0.022)	0.999 (0.002)	77	52
RP	RG	49.81 (2.23)	91.50 (2.02)	0.448 (0.035)	− 0.446 (0.033)	0.997 (0.003)	32	17
CP	RG	48.04 (2.12)	91.63 (2.00)	0.518 (0.033)	0.500 (0.034)	0.993 (0.003)	19	3

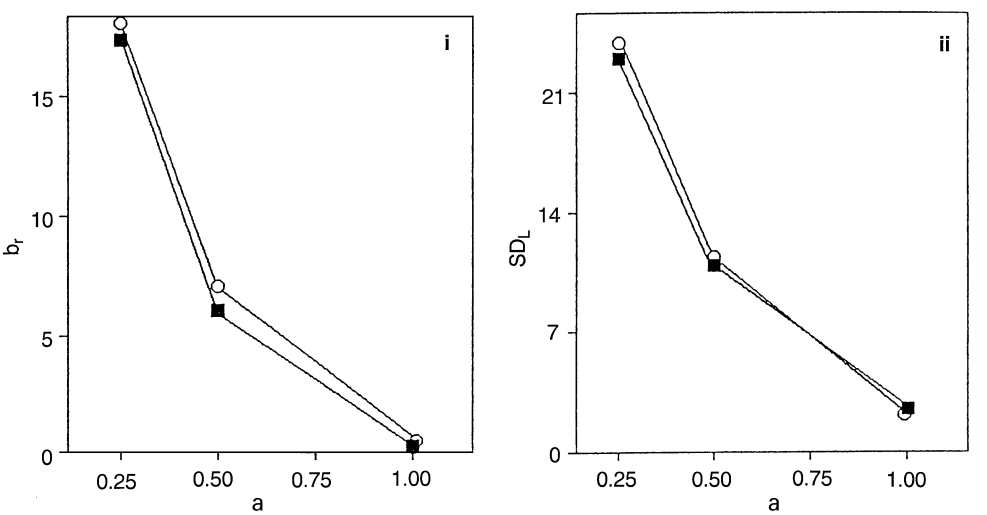

Fig. 2 Parameter estimates resulting from interval analysis with selective genotyping when the trait is affected by two linked QTLs as a function of the QTL effects. **(i)** Bias in the estimated distance between the QTL; **(ii)** standard deviation of the estimated position. *Open circles*, coupling phase; *black squares*, repulsion phase. The size of the mapping population was 4000 individuals

the SD_L of the estimated distance is plotted as a function of the simulated QTL effects in Fig. 2. As in Table 2, both simulated loci had equal effects. Although the distance was always over-estimated, both bias and the SD of the estimated distance decreased with an increase in QTL effects. With SG the confidence intervals for QTL locations were only half of those for RG.

Two-trait analysis

(1) $a_x = 0.25$ and $a_y = 0$. Single-trait ML analysis of populations simulated with a segregating QTL affecting trait x but not y are presented in the upper part of Table 3. Since selection was relative to trait x, ML parameter estimates for this trait should be unbiased in all cases. However, when a non-zero residual correlation between the traits was simulated, biased results were obtained for y. With $R = 0.5$ the estimated effect on y was about 80% of the effect on x. Even relatively "mild" SG resulted in significantly biased estimates for the effect on y (data not shown). The statistical power to detect an effect on y, which should have been equal to the type-I error, was considerably greater with a non-zero residual correlation, and increased as a function of this correlation. Similar to the case described by Martinez and Curnow (1992) a ghost QTL effect is observed when an incorrect model is used for QTL analysis.

The results for the multiple-trait analyses with selective and random genotyping are presented in the remainder of Table 3. SG resulted in a considerably greater power of detection for both the single- and multiple-trait analyses. Even though selection was relative to x, and an effect was simulated only on this trait, the power of detection with correlated traits was slightly greater for the multi-trait analyses than for the single-trait analyses of x. Furthermore, the power increased with an increase in the residual correlation. Similar results with random genotyping were found previously by Korol et al. (1994, 1995, 1998) and Ronin et al. (1995). The SDs of the estimated QTL location as a function of R for both SG and random sampling are

Table 3 Increased detection power, and accuracy of estimation of QTL position and effect using interval two-trait analysis under selective genotyping as compared to single-trait analysis and random genotyping. A single QTL on the genotyped chromosome affects the trait under selection, but not the correlated trait. Backcross populations were simulated for two traits, x and y. The simulated QTL effects were $a_x^* = 0.25$ and $a_y^* = 0$, residual standard deviations were $\sigma_x = \sigma_y = 1$, and the residual correlation between the traits was R^* (0, 0.5, and 0.7)

R^*	a_x	a_y	σ_x	σ_y	R	L	Power %	
							$\alpha \rightarrow 5\%$	1%
Single-trait analysis, SG								
	0.257		0.998			60.76	92	82
	(0.004)		(0.001)			(1.00)		
0.0		0.006		0.995		80.65	8	1
		(0.014)		(0.001)		(3.41)		
0.5		0.206		0.992		68.18	35	18
		(0.006)		(0.001)		(2.27)		
0.7		0.208		0.992		66.14	55	28
		(0.005)		(0.001)		(1.86)		
Two-trait analysis, SG								
0.0	0.254	− 0.001	0.998	0.998	− 0.002	61.20	87	77
	(0.005)	(0.009)	(0.001)	(0.001)	(0.002)	(1.12)		
0.5	0.254	− 0.005	0.998	0.997	0.499	60.87	91	80
	(0.005)	(0.008)	(0.001)	(0.001)	(0.001)	(1.05)		
0.7	0.255	0.006	0.998	0.997	0.699	59.90	93	86
	(0.004)	(0.007)	(0.001)	(0.001)	(0.001)	(0.86)		
Two-trait analysis, RG								
0.0	0.253	− 0.001	0.996	0.997	− 0.001	64.01	33	10
	(0.010)	(0.010)	(0.001)	(0.001)	(0.002)	(2.41)		
0.5	0.273	0.003	0.996	0.996	0.498	64.67	46	25
	(0.010)	(0.010)	(0.001)	(0.001)	(0.001)	(2.11)		
0.7	0.255	0.004	0.996	0.996	0.699	63.67	64	40
	(0.010)	(0.009)	(0.001)	(0.001)	(0.001)	(1.78)		

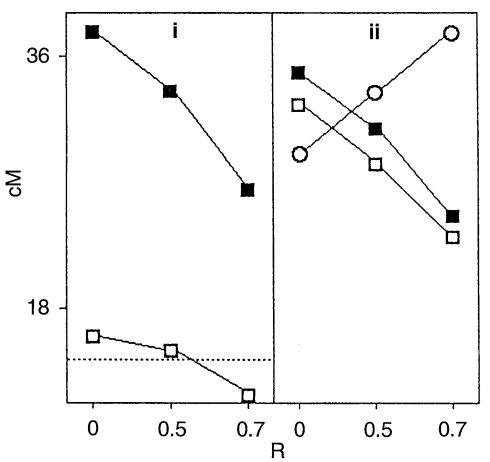

Fig. 3 Standard deviation of the estimated position of the mapped QTLs with selective and random genotyping for single- and two-trait analysis as a function of the residual correlation between the traits. **(i)** The marked chromosome affects the selected trait, and **(ii)** the marked chromosome affects the correlated trait. *Open and black squares* denote two-trait analysis with selective and random genotyping, respectively; the *dotted line* and *open circles* denote results for the single-trait analyses of traits x and y, respectively, under selective genotyping

plotted in Fig. 3. SDs were always less for SG. SDs were less for multi-trait as compared to single-trait analyses, except for an effect simulated on y and a correlation of zero. However, for this case the single- and multi-trait analyses, with or without SG, should be equivalent. The differences among the three alternatives plotted in Fig. 3ii for $R = 0$ are not significant, and apparently represent random variation.

The distribution of the maximum LOD values by chromosomal intervals is given in Fig. 4. The correct interval always has the highest frequency, but SG increased the probability of correct determination. This probability increases with an increase in the residual correlation. As expected, the frequency of correct determination of the QTL location increases with sample size (data not shown).

(2) $a_x = 0$ and $a_y = 0.25$. The results are given in Table 4. In the single-trait analysis the power to reject the null hypothesis of $a_x = 0$ was close to the expected value of α. Figure 1 b indicates that with a non-zero correlation, SG for x should reduce the power of detection, and bias downward the estimate of the QTL effect on y. As the correlation increases from 0 to 0.7, the statistical power of detection for $\alpha = 0.05$ and $N = 400$ decreases from 41 to 24%. Similar trends were also found for other selection levels (data not shown). As the residual correlation increases, the estimate of QTL location tends towards the center of the chromosome, and the SE of the estimate increases.

The null hypothesis for the multivariate analysis is no segregating QTL affecting either trait. With no residual correlation, the power to detect the QTL effect is similar to that obtained for the single-trait analysis for all selection levels. However, the power to detect a QTL increases with an increase in the residual correlation. Estimates of all parameters were close to the

simulated values. The QTL location was generally biased (Hyne et al. 1995) towards the center of the chromosome. As the residual variance and the sample size increased, this bias decreased.

(3) More than one QTL. Similar results were obtained with two unlinked QTLs. In the example presented (Table 5), both QTLs only affect trait x, $a^*_{x_1} = a^*_{x_2} = 0.25$ and $a^*_{y_1} = a^*_{y_2} = 0$, with a residual correlation $R^*_{xy} = 0.5$ and residual SDs $\sigma_x = \sigma_y = 1$. In the single-trait analyses, estimates of the effects of both loci were close to the simulated values for x, but biased for y. The estimates of a_{y_1} and a_{y_2} were more than 70% of the effects simulated for trait x. By contrast, two-trait analysis results in unbiased estimates of the parameters.

(4) SG with a QTL variance effect. The results are presented in Table 6 for a QTL with mean and variance effects on trait x, and no effects on trait y. Residual correlations of 0 and 0.7 were simulated. The power of QTL detection was estimated as the frequency of rejection of H_0 (no segregating QTL on the chromosome genotyped) based on an asymptotic approximation of the test statistics (Wilks 1962). Similarly, the power of detection of the variance effect was determined from the likelihood ratio of unequal and equal variance-parameter estimates. A variance effect of only 15% dramatically increased the power of the QTL detection. Furthermore, the SEs of the estimated QTL chromosomal position were reduced by half. Previous results with random genotyping indicate that analysis with a model which assumes no QTL variance effect, when the QTL does in fact affect the variance, reduces the power of QTL detection (Korol et al. 1996). However, in the present analysis with SG, the power obtained assuming equal genotype variance was nearly the same whether or not a variance effect was simulated. These results demonstrate that if the variance effect is small, the reduction in power may be negligible. Thus the two-trait SG analysis appears to be robust with respect to inaccuracies in the analysis model.

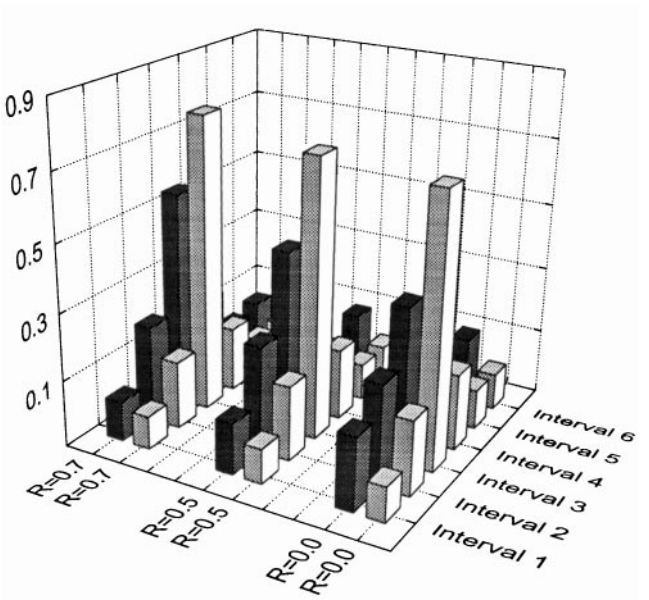

Fig. 4 The frequency of the distribution of the maximum LOD score by chromosomal intervals as a function of the residual correlation based on two-trait interval analysis. The QTL with simulated effect $a^*_x = 0.25$ and $a^*_y = 0$ was located in the middle of interval 3, at $L = 60$ cM. *Light and shaded columns* denote frequencies for selective and random genotyping, respectively

Discussion

Lin and Ritland (1996) found that SG can bias the proportion of recombinant genotypes for linked QTLs, with the bias increasing with the QTL effects. The simulation results presented in Tables 1 and 2 for interval ML analysis demonstrate that, with either

Table 4 Increased detection power, and accuracy of estimation of QTL position and effect using interval two-trait analysis under selective genotyping as compared to single-trait analysis. A single QTL on the marked chromosome affects the correlated trait, but not the trait under selection. The simulated effects of the QTL were $a^*_x = 0$ and $a^*_y = 0.25$

R^*	a_x	a_y	σ_x	σ_y	R	L	Power %	
							$\alpha \to 5\%$	1%
Single-trait analysis								
	0.007		0.997			69.28	5	1
	(0.007)		(0.001)			(3.37)		
0.0		0.273		0.997		61.63	41	18
		(0.008)		(0.001)		(1.99)		
0.5		0.184		1.001		66.06	26	14
		(0.007)		(0.001)		(2.32)		
0.7		0.142		1.003		68.52	24	12
		(0.007)		(0.001)		(2.60)		
Two trait analysis								
0.0	0.006	0.257	0.998	0.997	− 0.002	63.26	32	16
	(0.006)	(0.009)	(0.001)	(0.001)	(0.002)	(2.21)		
0.5	0.007	0.261	0.998	0.997	0.499	63.08	43	21
	(0.005)	(0.007)	(0.001)	(0.001)	(0.001)	(1.94)		
0.7	0.006	0.255	0.998	0.997	0.699	61.14	64	37
	(0.005)	(0.006)	(0.001)	(0.001)	(0.001)	(1.60)		

Table 5 Unbiased parameter estimation in two-trait interval analysis combined with selective genotyping. Two QTLs on different chromosomes affect the trait under selection (x), but not the correlated trait (y). Note, that single-trait analysis results in highly biased estimates of both QTL effects on trait y, whereas two-trait analysis gives unbiased estimates

a_{x_1}	a_{x_2}	σ_x	a_{y_1}	a_{y_2}	σ_y	R_{xy}	L_1	L_2
Single-trait analysis								
0.256	0.252	0.999					59.25	83.02
(0.005)	(0.004)	(0.001)					(0.99)	(1.00)
			0.193	0.182	0.988		50.48	91.84
			(0.010)	(0.012)	(0.002)		(2.96)	(2.45)
Two-trait analysis								
0.252	0.245	1.001	0.001	0.005	0.996	0.502	59.53	83.07
(0.006)	(0.007)	(0.002)	(0.012)	(0.012)	(0.002)	(0.002)	(1.15)	(1.12)

Table 6 Increased detection power and accuracy of estimation of QTL position due to an accounted for variance effect. A single QTL on the genotyped chromosome affects the mean and variance of the selected trait x, but has no effects on the correlated trait y. The simulated QTL effect on mean value and variance where $a_x^* = 0.25$ and $\delta_x = \sigma_{x_1}/\sigma_{x_2} = 1.15$. The power of detection of the variance effect was determined from the likelihood ratio of unequal and equal variance parameter estimates, and is denoted by an asterisk. Note a remarkable reduction in the standard errors of the estimated map position of the QTL

R^*	a_x s_{a_x}	a_y s_{a_y}	δ_x s_{δ_x}	σ_y s_{σ_y}	R s_R	L s_L	Power % $\alpha \to 1\%$	Power % 0.1%
Simulated: $\delta_x = \sigma_{x_1}/\sigma_{x_2} = 1$, model: $\delta_x = 1$								
0	0.254	0.009		0.996	− 0.003	58.26	62	33
	(0.006)	(0.014)		(0.001)	(0.002)	(1.64)		
0.7	0.259	0.016		0.997)	0.699	57.97	72	43
	(0.005)	(0.010)		(0.001)	(0.001)	(1.35)		
Simulated: $\delta_x = \sigma_{x_1}/\sigma_{x_2} = 1.15$, model: $\delta_x = 1$								
0	0.246	0.015		0.996	− 0.003	58.8	68	33
	(0.006)	(0.012)		(0.001)	(0.002)	(1.60)		
0.7	0.243	0.018		0.996	0.697	57.5	74	44
	(0.006)	(0.011)		(0.001)	(0.001)	(1.56)		
Simulated: $\delta_x = \sigma_{x_1}/\sigma_{x_2} = 1.15$, model: $\delta_x \neq 1$								
0	0.251	0.008	1.154	0.997	− 0.003	59.6	97	84
	(0.006)	(0.012)	(0.004)	(0.001)	(0.002)	(0.86)	95*	86*
0.7	0.252	0.007	1.157	0.997	0.700	60.4	99	94
	(0.006)	(0.009)	(0.003)	(0.001)	(0.001)	(0.80)	98*	93*

linked or unlinked QTLs, SG always outperforms RG. SG results in greater detection power, a lower bias and a higher accuracy for the parameter estimates, including both QTL locations and effects.

Henderson (1975) investigated the effect of selection in dairy cattle on sire evaluation using Best Linear Unbiased Methodology, and concluded that sire evaluation will be unbiased if: (1) variances and covariances were known, (2) selection occurred within fixed effects, and (3) all records on which selection were based were included in the analysis. Thus, if cows are selected based on first-parity production records, but later-parity records, which are a selected sample, are included in the analysis, then unbiased genetic evaluations are derived, if all first-parity records are included (Weller 1986 b). Similarly, in the current study, unbiased parameter estimates were derived only if all records for the trait under selection were included in the analysis. Furthermore, the power of QTL detection, and the estimation accuracy of the QTL substitution effects, residual variances, residual correlations, and the chromosomal position of the QTL are increased by SG for a ML multi-trait analysis. With single-trait ML analysis "ghost" QTL effects may be detected, the power of QTL detection is reduced, and estimates of the substitution effects and chromosomal location are biased. Contrary to the single-trait analyses, detection power and the accuracy of parameter estimates increase with an increase in the residual correlation. All parameters were accurately estimated in the multi-variate analyses, except for QTL location with a simulated effect on y. However, even for this parameter, the bias and variance of the estimated location decrease with increased residual correlation and sample size.

Experiments for the detection of QTL with the aid of genetic markers consist of four elements: generation of the population for analysis, scoring individuals for quantitative traits, genotyping for genetic markers, and data analysis. The cost of data analysis will generally be insignificant. Scoring additional traits will usually be independent of both generating the population and genotyping. Unless the objective of the analysis is a single well-defined trait, then it will generally be cost effective to measure additional traits. Even if the objective is to detect QTLs for a single characteristic, it will often be useful to score related traits that may yield additional information about the trait of interest (Korol et al. 1987, 1994, 1995, 1996; Ronin et al. 1995;

Jiang and Zeng 1995). For example, with resistance to a specific disease, there will be more than one symptom that can be scored, and these variables will be only partially correlated. Thus, the question of the effect of SG on correlated traits will nearly always be relevant if SG is employed, unless SG is applied to a set of canonical variables, which are uncorrelated by definition, as suggested by Weller et al. (1996). In view of the results presented, there seems to be little justification for this procedure, unless many traits are considered. An alternative strategy may be an iterative procedure in which each interval is analyzed with a correction for previously fitted QTLs.

Although SG with a multi-trait ML analysis does not decrease the power per individual genotyped in any of the cases considered, the optimum sampling strategy for several correlated traits cannot be derived from the results presented. Further study of this question is suggested. In addition to the covariance matrix among the traits, trait-scoring costs, the relative importance of the traits, and heritability should also be considered. Possibly, methods of linear programming could be applied to optimize the sampling strategy.

Acknowledgements The authors thank Morris Soller and Rohan Fernando for useful discussions. The helpful comments of the reviewers are also acknowledged. This research was supported by the United States–Israel Binational Agricultural Research and Development Fund (BARD), project number IS-2383-94C, and by the Israeli Ministry of Absorption.

References

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

Darvasi A, Soller M (1994) Selective DNA pooling for determination of linkage between a marker locus and a quantitative trait loci. Genetics 138 : 1365–1373

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

Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, Pasquino AT, Sargent LS, Sorensen A, Steele MR, Zhao X, Womack JE, Hoeschele I (1995) Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics 139 : 907–920

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

Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. Biometrics 31 : 423–447

Hyne V, Kearsey MJ, Pike DJ, Snape JW (1995) QTL analysis: unreliability and bias in estimation procedures. Mol Breed 1 : 273–282

Jansen RC (1994) Controlling the type-I and type-II errors in mapping quantitative trait loci. Genetics 138 : 871–881

Jiang C, Zeng ZB (1995) Multiple-trait analysis and genetic mapping for quantitative trait loci. Genetics 140 : 1111–1127

Korol AB, Preygel IA, Bocharnikova NI (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 AB, Preygel IA, Preygel SI (1994) Recombination, Variability and Evolution. Chapman and Hall, London

Korol AB, Ronin YI, Kirzhner VM (1995) Interval mapping of quantitative trait loci employing correlated trait complexes. Genetics 140 : 1137–1147

Korol AB, Ronin YI, Tadmor Y, Bar-Zur A, Kirzhner VM, Nevo E (1996) Estimating the variance effect of QTLs: an important prospect to increase the resolution power of interval mapping. Genet Res 67 : 187–194

Korol AB, Ronin YI, Hayes P, Nevo E (1998) Multi-interval mapping of correlated trait complexes: simulation analysis and evidence from barley. Heredity 80 : 273–284

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

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

Lin JZ, Ritland K (1996) The effect of selective genotyping on estimates of proportion of recombination between linked quantitative trait loci. Theor Appl Genet 93 : 1261–1266

Martinez O (1996) Spurious linkage between markers due to selective genotyping in QTL Mapping. Plant Genome Abst IV, abstract P110

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

Motro U, Soller M (1993) Sequential sampling in determining linkage between marker loci and quantitative trait loci. Theor Appl Genet 85 : 658–664

Payne F (1918) An experiment to test the nature of the variation on which selection acts. Indiana Univ Studies 5 : 1–45

Plotsky Y, Cahaner A, Haberfeld A, Lavi U, Lamont SJ, Hillel J (1993) DNA fingerprint bands applied to linkage analysis with quantitative trait loci in chickens. Anim Genet 24 : 105–110

Preygel IA, Korol AB (1989) Marker analysis of quantitative characters. Uspekhi Sovremennoy Genetiki (Advances in Modern Genetics) USSR (in Russian) 16 : 82–95

Ronin YI, Kirzhner VM, Korol AB (1995) Linkage between loci of quantitative traits and marker loci: multi-trait analysis with a single marker. Theor Appl Genet 90 : 776–786

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

Simpson SP (1989) Detection of linkage between quantitative trait loci and restriction fragment length polymorphisms using inbred lines. Theor Appl Genet 77 : 815–819

Simpson SP (1992) Correction: Detection of linkage between quantitative trait loci and restriction fragment length polymorphisms using inbred lines. Theor Appl Genet 85 : 110–111

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

Soller M (1991) Mapping quantitative trait loci affecting traits of economic importance in animal populations using molecular markers. In: Schook LB, Lewin HA, McLaren DG (eds) Gene mapping. Strategies, techniques and applications. Marcel Dekker, Inc., New York, pp 21–49

Soller M (1994) Marker-assisted selection–an overview. Anim Biotech 5 : 193–207

Soller M, Beckmann JS (1990) Marker-based mapping of quantitative trait loci using replicated progeny. Theor Appl Genet 80 : 205–208

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

Weller JI (1986 b) Comparison of multi-trait andd single-trait multiple parity evaluations by Monte-Carlo simulations. J Dairy Sci 69 : 493–500

Weller JI (1992) Statistical methodologies for mapping and analysis of quantitative trait loci. In: Beckmann JS, Osborn TC. Plant genomes: Methods for genetic and physical mapping. Kluwer Academic Publishers Group, Dordrecht, The Netherlands, pp 181–186

Weller JI (1996) Introduction to QTL detection and marker-assisted selection. In: Beltsville Symp Agric Res XX: Biotechnology's Role in Genetic Improvement of Farm Animals. Beltsville, Maryland, pp 259–275

Weller JI, Wyler A (1992) Power of different sampling strategies to detect quantitative trait loci variance effects. Theor Appl Genet 83 : 582–588

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

Weller JI, Kashi Y, Soller M (1990) Power of "daughter" and "granddaughter" designs for genetic mapping of quantitative traits in dairy cattle using genetic markers. J Dairy Sci 73 : 2525–2537

Weller JI, Wiggans GR, VanRaden PM, 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 92 : 998–1008

Wilks SS (1962) Mathematical Statistics. Wiley, New York