Enhanced Efficiency of Quantitative Trait Loci Mapping Analysis Based on Multivariate Complexes of Quantitative Traits

Abraham B. Korol, Yefim I. Ronin, Alexander M. Itskovich, Junhua Peng and Eviatar Nevo

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

Manuscript received February 20, 2000
Accepted for publication January 10, 2001

ABSTRACT

An approach to increase the efficiency of mapping quantitative trait loci (QTL) was proposed earlier by the authors on the basis of bivariate analysis of correlated traits. The power of QTL detection using the log-likelihood ratio (LOD scores) grows proportionally to the broad sense heritability. We found that this relationship holds also for correlated traits, so that an increased bivariate heritability implicates a higher LOD score, higher detection power, and better mapping resolution. However, the increased number of parameters to be estimated complicates the application of this approach when a large number of traits are considered simultaneously. Here we present a multivariate generalization of our previous two-trait QTL analysis. The proposed multivariate analogue of QTL contribution to the broad-sense heritability based on interval-specific calculation of eigenvalues and eigenvectors of the residual covariance matrix allows prediction of the expected QTL detection power and mapping resolution for any subset of the initial multivariate trait complex. Permutation technique allows chromosome-wise testing of significance for the whole trait complex and the significance of the contribution of individual traits owing to: (a) their correlation with other traits, (b) dependence on the chromosome in question, and (c) both a and b. An example of application of the proposed method on a real data set of 11 traits from an experiment performed on an F_2/F_3 mapping population of tetraploid wheat (*Triticum durum* × *T. dicoccoides*) is provided.

THE detection power and mapping resolution of marker analysis of quantitative traits are the major factors affecting practical applications of quantitative trait loci (QTL) mapping. These characteristics strongly depend on the effect of the QTL in question relative to the phenotypic variance of the trait in the mapping population. The higher the discrepancy between QTL groups (or the contribution of the QTL to the trait heritability H^2, the proportion of genetic variation σ^2_G in total phenotypic variation σ^2_{Ph} of the trait) the better the expected QTL detection power and mapping resolution. As shown by Lander and Botstein (1989), the expected value of the log-likelihood test statistics increases monotonically with H^2:

$$\text{ELOD} = -\tfrac{1}{2}N\log(1 - H^2). \qquad (1)$$

Several strategies have been proposed to improve the precision of QTL mapping. These involve development of (i) new experimental designs to suit specific mapping goals and an organism's breeding system, and (ii) new QTL mapping models and algorithms to extract maximum information about QTL locations and effects. One of the improvements includes multilocus (composite) mapping analysis (Jansen and Stam 1994; Zeng 1994), selective sampling (Lebowitz *et al.* 1987; Darvasi and Soller 1994), replicated progeny testing (Soller and Beckmann 1990), and sequential experimentation (Motro and Soller 1993). For example, in composite mapping the increase in mapping resolution derives from a reduction of the residual variation by taking into account the effects of cosegregating QTL.

In QTL mapping, the experimental design usually includes simultaneous measurements of many related and unrelated quantitative traits and subsequent treatment of the individual traits. Recently, several groups attempted to improve the efficiency of marker analysis of QTL by taking into account possible effects of the putative QTL on several traits simultaneously (Korol *et al.* 1987, 1995, 1998a; Amos *et al.* 1990; Schork *et al.* 1994; Jiang and Zeng 1995; Ronin *et al.* 1995, 1998, 1999; Weller *et al.* 1996; Almasy *et al.* 1997; Boomsma and Dolan 1998; Mangin *et al.* 1998; Henshall and Goddard 1999; Olson *et al.* 1999; Williams *et al.* 1999; Zeng *et al.* 2000). In the simplest case of two noncorrelated traits, the advantage of joint analysis is in the increase of the *multivariate effect* according to $d^2 = (d_x/\sigma_x)^2 + (d_y/\sigma_y)^2$ (Figure 1a), where d_x and d_y are the substitution effects of the QTL for traits x and y, and σ_x and σ_y are the corresponding standard deviations within the QTL groups (residual standard deviations). Consequently, for a population with 1:1 ratio of the alternative QTL groups (like backcross, dihaploid, or recombinant inbreds) the bivariate analogue of H^2 could be represented in the form

Corresponding author: Abraham Korol, Institute of Evolution, University of Haifa, Haifa 31905, Israel. E-mail: korol@esti.haifa.ac.il

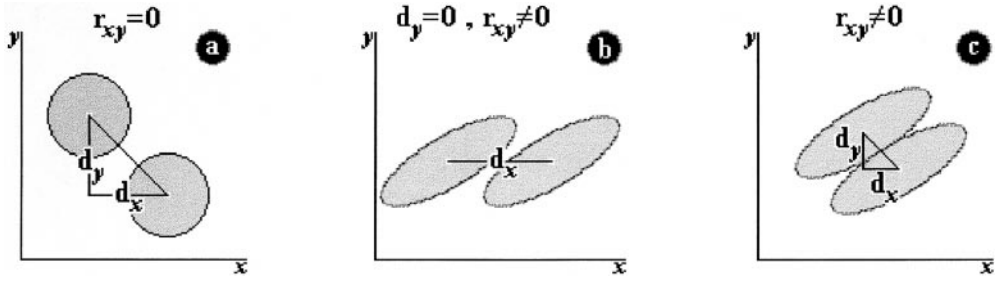

FIGURE 1.—The main sources for improvement of QTL mapping efficiency in multiple-trait analysis: (a) due to the pleiotropic effects; (b) due to correlation between the traits (within the QTL groups) caused by nongenetic effects and segregation of unlinked QTL; or (c) due to *combined effect* of both foregoing factors.

$$H^2_{xy} = \frac{\frac{1}{4}d^2}{1 + \frac{1}{4}d^2}. \qquad (2)$$

The situation becomes more complicated when correlated traits are involved. It can be shown (KOROL *et al.* 1995) that Equation 1 remains valid in bivariate analysis of correlated traits,

$$\mathrm{ELOD}(x, y) = -\tfrac{1}{2}N \log(1 - H^2_{xy}) \qquad (1')$$

with

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

It was shown earlier that either $\mathrm{ELOD}(x, y) \geq \mathrm{ELOD}(x)$ and $\mathrm{ELOD}(x, y) \geq \mathrm{ELOD}(y)$ follow from $H^2_{xy} \geq H^2_x$ and $H^2_{xy} \geq H^2_y$, respectively (KOROL *et al.* 1995; RONIN *et al.* 1999). Given fixed d_x/σ_x or $H^2_x = \frac{1}{4}d^2_x/(\frac{1}{4}d^2_x + \sigma^2_x)$, how will the resolution be affected by other traits being taken into account? Several situations should be considered to explain the expected gain of joint analysis of multiple traits compared to single-trait analysis. For the sake of simplicity, let us consider two traits. As mentioned above, if $R_{xy} = 0$, the effect of an additional trait is simply due to the increased *Euclidean* distance between the (two-dimensional) centers of the QTL groups (see Figure 1a). Consider now the situation when the traits are correlated within each of the QTL groups with residual correlation $R_{xy} \neq 0$. It is easy to see from Equation 3 that if $d_y \neq 0$ and $R_{xy} \neq 0$ and $\mathrm{sign}(R_{xy}d_xd_y) < 0$, then $H^2_{xy} \geq H^2_x$ and one could expect a respective increase in ELOD. Moreover, the inequality $H^2_{xy} > H^2_x$ holds even if $d_y = 0$ but $R_{xy} \neq 0$, independent of the sign of correlation (Figure 1b). Therefore, we can further assume that the increment in H^2_{xy}, compared with H^2_x, will result in an increased resolution of the mapping analysis (in spite of complications due to certain statistical nonequivalence), no matter how this increment in H^2_{xy} was produced, due to (i) the pleiotropic effect of the QTL on x and y, (ii) residual correlation between x and y (within the QTL groups) caused by nongenetic effects or segregation of unlinked QTL, or (iii) the *combined effect* of both factors (i) and (ii) (Figure 1c). In other words, instead of separate analyses of traits x and y, one can conduct joint analysis of these traits that is formally equivalent to transformation of a two-dimensional phenotype into a one-dimensional phenotype. For the new phenotype, a higher ratio of the between-QTL group difference to the residual variation can be achieved owing to the pleiotropic effect of the QTL on both traits, and residual correlation between the traits caused by nongenetic factors and segregation at other QTL. These expectations, illustrated geometrically in Figure 1, are confirmed by both Monte Carlo simulations and analytical approximations, for marker and interval analysis (KOROL *et al.* 1995, 1998a; RONIN *et al.* 1995, 1999). Although the described approach resembles the principal component analysis (PCA), it differs from PCA significantly. Besides technical differences, the main dissimilarity is that our trait transformations are interval dependent (KOROL *et al.* 1995), whereas in PCA the transformation is applied to the initial trait complex (WELLER *et al.* 1996). This difference may be important if the mapping population segregates for more than one QTL (see below).

Clearly, not only statistical reasons are of interest when discussing the advantages of the joint analysis of correlated trait complexes. The multitrait approach allows for an integral evaluation of the effects of genomic segments on a defined group of traits. Because of the internal balance of the organism's systems (SCHMALHAUSEN 1942), such an approach for QTL mapping seems to be much more justified biologically than the usual "trait-by-trait" analysis. It may assist in testing numerous biologically important hypotheses concerning manifold effects of genomic segments on quantitative variation, to distinguish between linkage and pleiotropy as mechanisms of genetic correlation, to address the problem of QTL-by-environment interaction, etc. (KOROL *et al.* 1994, 1998b; JIANG and ZENG 1995; LEBRETON *et al.* 1998; RONIN *et al.* 1999). Such analysis may be of major importance in formulating marker-assisted breeding strategies, dissecting heterosis as a multilocus multitrait phenomenon, developing optimized programs for evaluation and bioconservation of genetic resources, and revealing the genetic architecture of fitness systems in natural populations and of multifactorial diseases in humans.

The increased number of parameters to be estimated complicates the application of this approach when a large number of traits are considered. With n traits to be analyzed simultaneously in the simplest case of a

backcross (as well as a dihaploid or recombinant inbred lines) mapping population using single-interval mapping, the model should include $(n^2 + 5n + 2)/2$ parameters [QTL position, n mean values, n effects, n residual variances, and $n(n - 1)/2$ covariances]. At $n = 10$, this amounts to 76 parameters.

One possible *ad hoc* simplification of the estimation aspects is based on a reduction to two-trait analysis (Korol *et al.* 1995, 1998a; Ronin *et al.* 1995, 1999; Jiang and Zeng 1995) that appeared to be very efficient in allowing for an increase in QTL detection power and mapping resolution. In specific situations, such a reduction to a two-trait analysis may also be justified by the biological nature of the involved traits. However, in real multitrait situations this approach may result in statistical difficulties caused by the large number of trait pairs. Corresponding multiple tests may be interdependent, causing a further complication in defining the critical values of the test statistics. Another possibility is related to the attempt at space transformation, *e.g.*, using the PCA (Weller *et al.* 1996; Mangin *et al.* 1998) applied to the multivariate trait distribution across the entire data set. Although this approach seems to be very attractive, it cannot directly solve the problem when the mapping population segregates for more than one QTL, especially if some of the effects are relatively strong. Indeed, assume that in such a case the PCA transformation was applied to the initial trait complex (without taking out the effects of the target QTL), *i.e.*, for all individuals independent of their genotypes. Then, the independence of the resulting derivative traits over the entire mapping population cannot guarantee their independence within the alternative QTL groups. Moreover, this problem may exist even in the case of one QTL segregating in the mapping population, because the total (across all individuals) variance-covariance matrix of the initial trait complex may differ from the residual one (*i.e.*, for the matrix characterizing the within-QTL group variation). It is noteworthy that the largest principal components may be irrelevant in such an analysis, as can be seen from Figure 1c (see also Olson *et al.* 1999). Nevertheless, in some cases this approach may work (in situations represented by Figure 1a).

Here we present a generalization of our previous two- and three-trait QTL mapping algorithm (Korol *et al.* 1995; Ronin *et al.* 1995), which is free of the mentioned difficulties, to multivariate trait complexes that allow analysis of a large number of traits. It is based on transformation of the initial trait space into a space of a lower dimension. In the simplest case of a single-QTL analysis of a backcross (dihaploid) mapping population, the resulting space is one-dimensional independent of the number of traits, whereas two-QTL analysis for such a population will employ a two-dimensional model (for F_2 these will be three- and eight-dimensional models, correspondingly). The main difference between our previous two-trait models (Korol *et al.* 1995, 1998a; Ronin *et al.* 1995, 1998, 1999; see also Jiang and Zeng 1995) and the foregoing PCA-based models lies in the fact that the residual variance-covariance matrix was considered interval dependent, in the following sense. Its elements are a subset of the vector of unknown parameters to be estimated by the employed procedure *for each interval*, so that for QTL residing in different genomic segments the resulting (transformed) traits could be very different. This interval dependence remains a notable characteristic of our new multivariate algorithm.

THE PROPOSED METHOD

The model: Assume first that only one QTL segregates in the mapping population. Consider the genomic segment carrying this QTL (with alleles Q and q) flanked by markers M_1/m_1 and M_2/m_2, with recombination rates r_1 and r_2 in intervals M_1/m_1–Q/q and Q/q–M_2/m_2. On the basis of the marker scores and the measurements of the trait complex $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, we should test whether or not variation of any trait of $\mathbf{x}$ indeed depends on the interval M_1/m_1–M_2/m_2 and identify the corresponding locus Q/q. The expected joint distributions of the traits $\mathbf{x}$ in each of the marker groups, $U_{m_1m_2} = U_1(\mathbf{x})$, $U_{M_1m_2}(\mathbf{x}) = U_2(\mathbf{x})$, $U_{m_1M_2}(\mathbf{x}) = U_3(\mathbf{x})$, and $U_{M_1M_2}(\mathbf{x}) = U_4(\mathbf{x})$, can be written as

$$U_i(\mathbf{x}) = \pi_i fqq(\mathbf{x}) + (1 - \pi_i) fQq(\mathbf{x}), \quad i = 1, \ldots, 4,$$

where the proportions $\pi_i = \pi_i(r_1, r_2)$ depend on unknown recombination rates r_1 and r_2 and mode of interference. The specification of the n-dimensional densities $fqq(\mathbf{x})$ and $fQq(\mathbf{x})$ depends on the assumptions made about the genetic control of the traits. The simplest case of additive control can be represented by the model

$$\mathbf{x} = \mathbf{m} + \mathbf{0.5d}g_q + \mathbf{e},$$

where $\mathbf{x} = (x_1, \ldots, x_n)$ is the vector of phenotype scores for an arbitrary individual, $\mathbf{e} = (e_2, \ldots, e_n)$ is a vector of random variables that obey multivariate normal distribution with zero expectations for all coordinates and (residual) variance-covariance matrix $\Sigma_{\mathbf{R}} = \{s_{ij}\}$, $\mathbf{m}$ is the vector of trait means, $\mathbf{d}$ is the vector of the effects of substitution at the Q/q locus with respect to mean values of $\mathbf{x}$, *i.e.*, $dx_i = \mu_{x_i}(Qq) - \mu_{x_i}(qq)$, and g_q denotes the genotype at locus Q/q ($g_q = -1$ for qq and 1 for Qq).

Expected improvement owing to multiple-trait analysis: As in the bivariate case, the QTL detection power should depend on the total contribution of the QTL to multivariate phenotypic variation (V_{Ph}) of the correlated trait complex. If V_{R} is the multivariate residual variation (within the QTL groups) and V_{G} is the combined between-QTL-group discrepancy, then

$$H^2_{\mathrm{T}} = V_{\mathrm{G}}/V_{\mathrm{Ph}} = V_{\mathrm{G}}/(V_{\mathrm{G}} + V_{\mathrm{R}}). \qquad (4)$$

In the case of noncorrelated traits, the improvement is

due to the "Euclidean effect," which grows with the number of traits:

$$H^2_{\text{Eu}} = \frac{\frac{1}{4}\sum(d_i/\sigma_i)^2}{1 + \frac{1}{4}\sum(d_i/\sigma_i)^2}. \quad (5)$$

Clearly, in the general case of correlated traits, the pure Euclidean contribution is only a part of the total effect, so that $H^2_{\text{Eu}} < H^2_{\text{T}}$. Note that an analogue of Equation 5 can be obtained by canonical transformation of the initial trait space (with the within-group covariance matrix associated with the QTL under consideration), allowing the evaluation of the total effect as defined by Equation 4. Then, the multivariate effect of the QTL will be manifested as in Equation 5, but with relative effects (d_i/σ_i) in the new coordinate system. Moreover, using scale transformations $x_i' = x_i/\sigma_i$ and corresponding angular transformations, one can map the multivariate space into another multivariate space where the QTL affects only one trait, with $H^2_D = \frac{1}{4}D^2/(1 + \frac{1}{4}D^2)$ being equal to the total contribution of the QTL, as in Equation 4 (where D is the total multivariate effect of the QTL).

The short review in the Introduction indicates that correlation between traits may be no less (if not more) an important factor affecting the detection power of multitrait QTL analysis. Therefore, it is of great interest to evaluate the contribution of correlations between the traits to H^2_{T} in Equation 4. Consequently, in the following illustrations, we present the expected improvement due to the Euclidean effect and the additional contribution due to correlations. Moreover, although no effect is expected from correlations if all effects d_i are 0, situations are possible where for only a small subset of traits $d_i \neq 0$, the remaining traits are still very informative because of their correlations to the foregoing traits (the simplest such example is provided in Figure 1b).

The numerical procedures of interval analysis: The distinctive feature of our analysis is that all the multivariate transformations are *interval-specific* (as can be seen from procedures 1 and 2 described below; see also Korol *et al.* 1987, 1995, 1998a; Jiang and Zeng 1995; Ronin *et al.* 1998), in contrast to the aforementioned attempts based on canonical transformation applied to the entire mixed distribution (Weller *et al.* 1996; Mangin *et al.* 1998). To explain why this is important, let us consider the simplest situation with a mapping population polymorphic for two unlinked QTL, say Q_1/q_1 and Q_2/q_2. A double haploid (or recombinant inbred, backcross, etc.) population will consist of four groups, such as $Q_1Q_1Q_2Q_2$, $Q_1Q_1q_2q_2$, $q_1q_1Q_2Q_2$, and $q_1q_1q_2q_2$. Assume that the residual (nongenetic) multitrait variation is the same in all four groups and can be described by a variance-covariance matrix $\Sigma_{\mathbf{R}}$.

Two possibilities for incorporating the joint variation of the traits exist when single-QTL mapping analysis is conducted on the basis of markers of one chromosome. These are (a) to consider the general variance-covariance matrix of the traits, which differs from $\Sigma_{\mathbf{R}}$ due to the contribution of both QTL (the higher the individual effects of Q_1/q_1 and Q_2/q_2, the higher the difference); and (b) to consider the residual variation for each QTL as a combined result of nongenetic variation and the contribution of all other QTL excluding the one under consideration. The second possibility provides a relevant description of the residual variation for each QTL.

Two different approximated procedures, giving very similar results, were employed to implement this approach. In both, the LOD score serves as the major criterion in interval mapping; the steps of evaluating the QTL effects and QTL position are separated. Both are based on our earlier maximum-likelihood approach (Korol *et al.* 1995). Although the proposed procedures are only approximations of the full procedure, their major advantage is that they allow treatment of a large number of traits simultaneously.

Procedure 1: For each interval, a five-step procedure is conducted.

1. The vector of mean trait values in alternative QTL groups defined by flanking markers M_1M_2 and m_1m_2 is evaluated.
2. The same groups are used to define the elements of the residual (for the current ith interval) covariance matrix, $\Sigma_{\mathbf{R}_i}$. Throughout this article, we assume no variance-covariance effect (but see Korol *et al.* 1995, 1996a), so that $\Sigma_{\mathbf{R}}(QQ) = \Sigma_{\mathbf{R}}(qq)$ and $\Sigma_{\mathbf{R}_i}(QQ) = \Sigma_{\mathbf{R}_i}(qq)$.
3. Transformation of the trait space, as described earlier, reduces the problem to a single-trait analysis. This step includes solving the problem of eigenvalues and eigenvectors of matrix $\Sigma_{\mathbf{R}}$ followed by scale and angular transformations, resulting in a new space with all effects being absorbed by only one variable ("integral" trait; see also Allison *et al.* 1998).
4. For the resulting variable, a single-trait analysis is conducted, with the likelihood function being dependent on four parameters, $\theta = (\mu, D, \sigma, r)$, where μ, D, σ, and r stand for the mean value of the new trait, total substitution effect, residual standard variation, and recombination rate from the left marker, respectively.
5. After getting the estimates, back transformations can be conducted, making it possible to get more precise estimates of mean values of the QTL groups. Consequently, the analysis could be repeated from step (2) until a convergent result is obtained.

Procedure 2: This is a simplified version of procedure one. It includes three steps and gives approximated results compared with those of procedure 1. However, the differences appear to be very small. For each interval, the three-step analysis is conducted.

1. The vector of mean trait values in alternative QTL

groups defined by flanking markers M_1M_2 and m_1m_2 is evaluated.
2. The same groups are used to define the elements of the residual (for the current ith interval) covariance matrix, $\Sigma_{\mathbf{R}i}$.
3. The entire sample is used to calculate the conditional maximum-likelihood estimate of the QTL position within the interval with all other parameters being fixed at the estimates obtained at steps (1) and (2).

Clearly, two factors influence the results obtained by this procedure. First, the estimates of the QTL effects will be biased downward owing to undetectable double recombinants among the parental (for the flanking markers) haplotypes. With interval size of ~10–15 cM this danger is negligible unless high negative interference is characteristic of multiple exchanges in the considered region of the genome. The second factor results in a slight reduction of the sample size: when the QTL effects and the residual covariance matrix are determined according to the foregoing steps (1) and (2), the recombinants for the flanking markers are ignored. Consequently, the sampling error of the estimates is increased by a factor $1/\sqrt{(1-r)}$, where r is the rate of recombination between the flanking markers; for an interval of 10–15 cM the loss of precision is ~5.4–8.5%.

Monte Carlo simulations: For mapping a population of the dihaploid (or recombinant inbred, backcross, etc.) type, 200 individuals were simulated with one, two, and three unlinked QTL and a trait complex including up to 10 traits. For each chromosome six equidistant markers were simulated, with recombination rate $r = 0.1$ between the neighbors and no interference and QTL residing in the middle of the third interval. To get the critical level of the test statistics two approaches were employed: Monte Carlo simulations with parameters corresponding to H_0 (no QTL in the simulated chromosome) and permutation of the data set corresponding to H_1. In both cases, 5000 runs were assayed for each situation. To evaluate the detection power and the precision of the estimated QTL effects and chromosomal position, 500 runs were assayed for each situation. In some isolated examples the numbers of permutation and bootstrap runs were increased to 10,000 and 1000, respectively. The majority of calculations were conducted using the multiple-trait algorithms implemented in the MultiQTL package (http://www.MultiQTL.com). With this program, 1000 permutation runs or 1000 bootstrap runs using a single-QTL model to analyze a 10-variate trait complex for a chromosome with 20 markers and population size 150 genotypes takes a Pentium III 600 MHz ~3.5 min or 2 min, respectively.

RESULTS

QTL detection power and mapping resolution: *Example 1: Improved quality of QTL analysis:* To demonstrate the contribution of different factors to the detection power and mapping resolution of multivariate QTL mapping, a series of variants were simulated that differ with respect to the number of traits (from 1 to 10), the type of the covariance matrix, the number of QTL, the effects of the target QTL(s) on the traits, etc. These were based on four 10×10 covariance matrices $\Sigma_{\mathbf{R}}$ (Table 1), with a common vector of alternating effects $\mathbf{d} = (0.25, -0.25, 0.25, -0.25, \ldots)$ and the same residual standard variation $s_i = 1.0$ for all traits. Table 2 represents a diversity of examples with a single QTL: covariance matrices of the majority of variants were derived as major minors of corresponding dimension of the matrix for the 10-trait problem.

As expected, the increase in H^2_{T} (see Equation 4) owing to higher information content of multivariate complexes of greater dimension than those of lower dimension brought about an appreciable improvement in the quality of the QTL mapping analysis. This is manifested in higher LOD values and, correspondingly, a better detection power and higher precision of QTL mapping (Table 2). As expected, the improvement strongly depends on correspondence between the QTL effects and the signs of correlation coefficients (*e.g.*, compare cases 2 and 6). The same mechanism appeared to work already in the two-trait analysis, as manifested by the inequality $d_x d_y R_{xy} < 0$ being the necessary condition for $\mathrm{ELOD}(x, y) > \mathrm{ELOD}(x)$ and/or $\mathrm{ELOD}(x, y) > \mathrm{ELOD}(y)$ to hold (Korol *et al.* 1995). The remarkable fact is that it makes no difference whether the increase in H^2_{T} is caused by correlation between the traits or by the Euclidean contribution H^2_{Eu} (Figure 2). Indeed, the variants represented in Figure 2 differ qualitatively. These include the number of traits taken from the entire 10-dimensional complex, the values and signs of the effects of the QTL on the selected traits, the values of correlations, and even the covariance matrices in general (*e.g.*, numbers 3, 4, 6, 10, 11, 13–15, 17, and 18). In spite of this diversity, the detection power (P) and mapping precision (σ_L) display a unified pattern across variants reflected in the curves $P(H^2)$ and $\sigma_L(H^2)$ in Figure 2.

The results presented in Table 2 and Figure 2 indicate the high potential for improving the QTL detection power and mapping resolution by employing the information contained in the multivariate trait complex without increasing the sample size. Thus, for the same data set corresponding to the first matrix (case 10) with no d_i/σ_i exceeding 0.25, the detection power grows from 13 to 100% (at significance level 0.01) for single- and 10-trait analyses, respectively. Especially pronounced is the improvement of mapping precision: standard deviation of the estimated QTL position, σ_L, decreases from 14.8 cM in single-trait, to 9.3 cM in 2-trait, to 4.0 cM for the matrix A defined in Table 1 (compare cases 1, 2, and 10), or correspondingly, 14.8, 9.4, and 1.4 cM for the matrix C (compare cases 1, 16, and 18). Clearly,

TABLE 1

Residual covariance matrices and QTL effects used in the simulations

	1	2	3	4	5	6	7	8	9	10	Effect
						A					
1	1.00	0.60	0.50	0.40	0.30	0.20	0.10				0.25
2	0.60	1.00	0.60	0.50	0.40	0.30	0.20	0.10	0.00		−0.25
3	0.50	0.60	1.00	0.60	0.50	0.40	0.30	0.20	0.10		0.25
4	0.40	0.50	0.60	1.00	0.60	0.50	0.40	0.30	0.20	0.10	−0.25
5	0.30	0.40	0.50	0.60	1.00	0.60	0.50	0.40	0.30	0.20	0.25
6	0.20	0.30	0.40	0.50	0.60	1.00	0.60	0.50	0.40	0.30	−0.25
7	0.10	0.20	0.30	0.40	0.50	0.60	1.00	0.60	0.50	0.40	0.25
8		0.10	0.20	0.30	0.40	0.50	0.60	1.00	0.60	0.50	−0.25
9		0.00	0.10	0.20	0.30	0.40	0.50	0.60	1.00	0.60	0.25
10				0.10	0.20	0.30	0.40	0.50	0.60	1.00	−0.25
						B_1 (blocks 2×2)					
1	1.00	0.70									0.25
2	0.70	1.00									−0.25
3			1.00	0.70							0.25
4			0.70	1.00				0.00			−0.25
5					1.00	0.70					0.25
6					0.70	1.00					−0.25
7							1.00	0.70			0.25
8			0.00				0.70	1.00			−0.25
9									1.00	0.70	0.25
10									0.70	1.00	−0.25
						B_2 (blocks 5×5)					
1	1.00	0.60	0.50	0.40	0.30						0.25
2	0.60	1.00	0.60	0.50	0.40						−0.25
3	0.50	0.60	1.00	0.60	0.50			0.00			0.25
4	0.40	0.50	0.60	1.00	0.60						−0.25
5	0.30	0.40	0.50	0.60	1.00						0.25
6						1.00	0.60	0.50	0.40	0.30	−0.25
7						0.60	1.00	0.60	0.50	0.40	0.25
8			0.00			0.50	0.60	1.00	0.60	0.50	−0.25
9						0.40	0.50	0.60	1.00	0.60	0.25
10						0.30	0.40	0.50	0.60	1.00	−0.25
						C					
1	1.00	0.70	0.50	0.32							0.25
2	0.70	1.00	0.70	0.50	0.32						−0.25
3	0.50	0.70	1.00	0.70	0.50	0.32			0.00		0.25
4	0.32	0.50	0.70	1.00	0.70	0.50	0.32				−0.25
5		0.32	0.50	0.70	1.00	0.70	0.50	0.32			0.25
6			0.32	0.50	0.70	1.00	0.70	0.50	0.32		−0.25
7				0.32	0.50	0.70	1.00	0.70	0.50	0.32	0.25
8		0.00			0.32	0.50	0.70	1.00	0.70	0.50	−0.25
9						0.32	0.50	0.70	1.00	0.70	0.25
10							0.32	0.50	0.70	1.00	−0.25

The four multitrait sets (A, B_1, B_2, and C) were used in Monte Carlo experiments presented in Table 2 and Figures 2 and 3. The trait complex B_1 includes five pairs of traits with nonzero correlation (0.7) only within pairs; likewise, trait complex B_2 includes two five-trait blocks with nonzero correlations only within the blocks. Empty cells in the covariance matrices correspond to zero correlation coefficients.

this trend reflects the fact that the increasing H^2 caused by joint multiple-trait analysis results not only in higher LOD values and detection power, but also in increased probability to find the QTL in the true interval (interval 3; see footnote *a* in the right column of Table 2). At the level of an individual experiment, the increased resolution derives from the effect of H^2 on the form of the LOD as a function of chromosomal position (l): at high H^2 values the function LOD(l) is more steep than at small H^2 (Figure 3). Clearly, increased precision of the estimated QTL position should also allow a more accurate estimation of the QTL effect. This is indeed the case, as illustrated by Figure 4. The increase in H^2 accompanied by a more strict slope of the LOD function

TABLE 2

Improvement of the efficiency of QTL mapping owing to joint analysis of multiple-trait complexes

No.	Variant according to Table 1		Scheme of effects										σ_L (cM)	Significance				H^2	Power (%)	Interval distribution of the detected QTL (%)				
			1	2	3	4	5	6	7	8	9	10		$LOD_{0.01}$	LOD_m	LOD_f	LOD_t			1	2	3[a]	4	5
1	A	1 × 1	+										14.8	1.97	2.59	0.94	0.67	0.015	13	16	27	32	17	8
2	A	2 × 2	+	−									9.3	2.62	4.27	3.60	3.27	0.072	71	5	16	55	20	4
3	A	3 × 3	+	−	+								7.9	3.07	5.64	5.09	4.85	0.106	84	2	18	61	16	3
4			+	−	+	−							7.2	3.59	7.03	6.91	6.56	0.140	94	2	17	66	13	2
5	A	4 × 4	+	0	0	0							11.5	3.59	4.33	2.13	1.10	0.025	12	7	18	39	26	10
6			+	−	0	0							8.9	3.59	5.10	4.08	3.35	0.074	57	4	21	55	16	4
7			+	−	+	0							8.2	3.59	5.96	5.40	4.95	0.108	83	3	16	63	15	3
8	A	5 × 5	+	−	+	−	+						6.4	3.94	8.51	8.64	8.27	0.173	98	1	17	68	13	1
9	A	8 × 8	+	−	+	−	+	−	+	−			4.8	5.23	13.14	13.14	12.72	0.254	100	0	8	83	8	1
10	A	10 × 10	+	−	+	−	+	−	+	−	+	−	4.0	5.93	16.41	16.41	15.13	0.294	100	0	8	85	7	0
11	B_1	2 × 2	+	−									8.0	2.55	4.94	4.47	4.30	0.094	84	3	16	59	19	3
12	B_1	4 × 4	+	−	+	−							6.0	3.61	8.42	8.39	8.35	0.175	99	1	11	76	1	1
13	B_1	10 × 10	+	−	+	−	+	−	+	−	+	−	3.5	5.74	18.71	18.71	18.62	0.349	100	0	5	91	4	0
14	B_2	5 × 5	+	−	+	−	+						6.2	3.96	8.73	8.66	8.27	0.173	97	1	13	69	16	1
15	B_2	10 × 10	+	−	+	−	+	−	+	−	+	−	4.1	5.97	15.96	15.96	15.58	0.301	100	0	5	87	8	0
16	C	2 × 2	+	−									9.4	2.64	4.81	4.35	4.30	0.094	85	6	18	57	15	4
17	C	5 × 5	+	−	+	−	+						3.6	3.97	15.36	15.36	15.71	0.304	100	0	5	89	6	0
18	C	10 × 10	+	−	+	−	+	−	+	−	+	−	1.4	5.88	60.81	60.81	65.21	0.777	100	0	0	100	0	0

For each variant, 500 Monte Carlo samples were simulated with residual covariance matrix and effects as denoted in the left columns and in Table 1. σ_L is the standard deviation of the estimated QTL position; $LOD_{0.01}$ is the threshold value of the test statistics (obtained by 5000 simulations under the assumption H_0 that the analyzed chromosome does not affect the trait complex); LOD_m is the mean value of the test statistics averaged over the runs with $LOD > LOD_{0.01}$ whereas LOD_f is the average over all runs; LOD_t is the theoretical value of LOD calculated based on multivariate generalization of the connection between LOD and H^2.

[a] The simulated QTL position was the middle of the third interval.

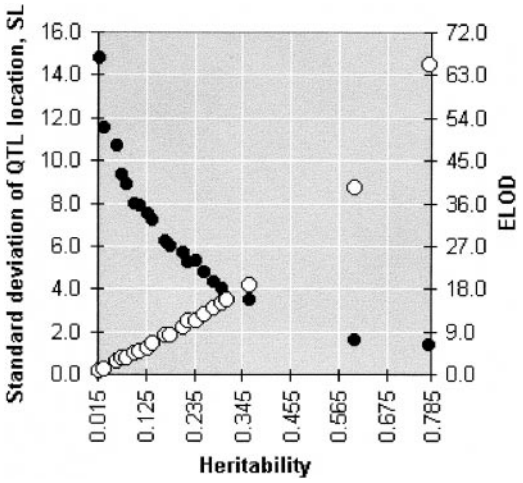

FIGURE 2.—Multivariate heritability as a predictor of the LOD value (affecting QTL detection power) and mapping resolution. Right-hand scale, ELOD (○), left-hand scale, SL (standard deviation of the estimated QTL position; ●). The graphs are based on Monte Carlo simulations described in Table 1 and partially represented in Table 2.

may justify a saturation of the chromosomal region in the detected QTL by additional markers. This may allow a reduction of the chances of incorrect QTL location and finer QTL mapping, as well as an attempt at resolving the pleiotropy-linkage alternative (JIANG and ZENG 1995; ALMASY *et al.* 1997; LEBRETON *et al.* 1998; KOROL *et al.* 1998a; RONIN *et al.* 1999).

It is noteworthy that ELOD calculated on the basis of H^2_T appeared to be a very good predictor of the averaged LOD obtained from Monte Carlo simulations (see the column LOD_m in Table 2). This indicates that Equation 1 obtained by LANDER and BOTSTEIN (1989) for a single trait, and generalized by KOROL *et al.* (1995) for two-trait analysis, also holds in a general multivariate case. The last statement follows from the fact that heritability of a complex of *noncorrelated* traits with a single QTL affecting only one trait can be represented as $H^2_D = \frac{1}{4}D^2/(1 + \frac{1}{4}D^2)$, where D is the multivariate effect and σ^2 the residual variance for the "integral" trait described in *The numerical procedures of interval analysis.*

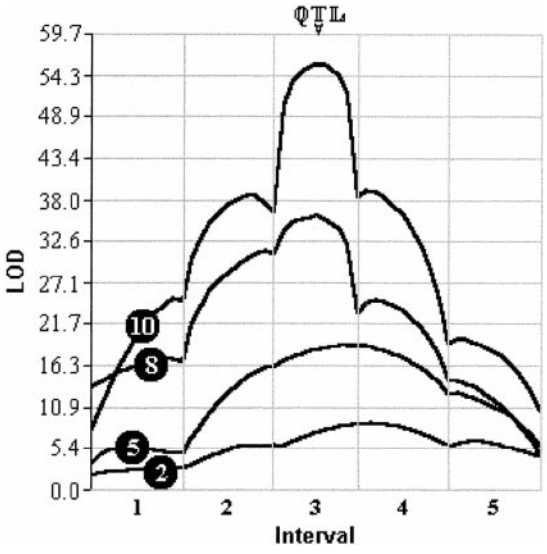

FIGURE 3.—The dependence of the LOD function on the number of traits. The numbers in the solid circles indicate the number of traits; the simulated position of the QTL is marked by an arrowhead (based on the last example of Table 1).

Example 2: Interval-specific estimation of the covariance matrix: Another comment concerns the interval specificity that is a characteristic of our approach to defining the elements of the residual covariance matrix, Σ_R. If, instead of that, one uses the total (interval-independent) covariance matrix defined on the entire sample, the efficiency of mapping may be lowered. The numerical example with a three-trait complex shown in Table 3 illustrates the difference between the two approaches. One can easily see that if the approach based on total covariance matrix is employed, instead of our interval-specific procedure, a reduction in the LOD value (hence lower detection power) and increase in the bias (δ) and standard variation (σ) of the estimated QTL effects (d_i) and, especially, chromosomal position (L), may be obtained. Note that in the foregoing example only a single QTL was simulated in the mapping population. The difference between the methods derives from the noncorrespondence between the residual correlation matrix and the directions of the pleitropic effects. Nevertheless, in some cases, where the total covariance matrix does not differ strongly from Σ_R, the loss will be less pronounced (see MANGIN *et al.* 1998).

Example 3: Multiple QTL: We now illustrate the efficiency of the proposed algorithm in situations with more than one QTL segregating in the mapping population. We simulated two and three identical unlinked QTL with the residual 10 × 10 covariance matrix equal to that of example 10 and the same pleiotropic effects (see Tables 1 and 2). As before, 500 Monte Carlo runs were made. The results (Table 4) confirm the previous conclusion: a dramatic improvement can be achieved by use of joint analysis of the correlated traits. Note that segregation for one or two additional QTL resulted in an increase in the residual variances (as compared with Example 1). Consequently, we obtained a slightly lower detection power and a lower mapping precision. For the 10-trait analysis, the standard deviation of the estimated QTL position (SL) increased from 4.0 to 5.0–5.6 cM in case of two QTL and to 5.8–6.7 cM in the case of three QTL. Clearly, this reduction in mapping precision can be recovered by a composite interval mapping approach (ZENG 1994; JANSEN and STAM 1994) but within the framework of multiple-trait analysis.

Significance of the detected effects: Testing for significance is a difficult problem in QTL mapping analysis, especially when multiple intervals and/or multiple traits are involved (LANDER and BOTSTEIN 1989; LANDER and

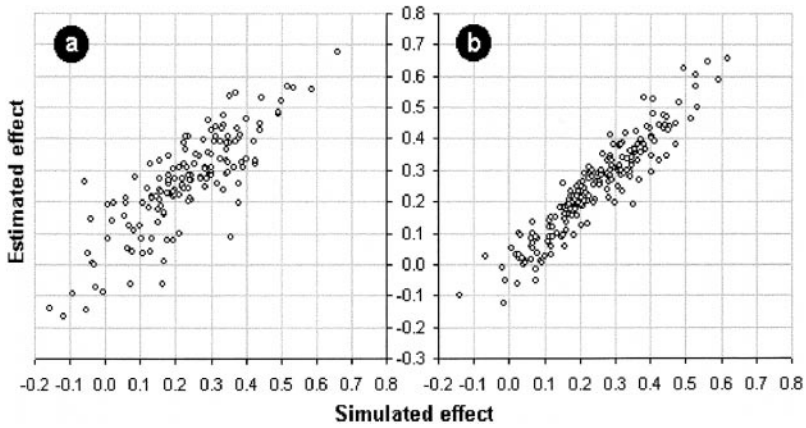

FIGURE 4.—Improved correspondence between the simulated and estimated QTL effects in multiple-trait analysis as compared to single-trait analysis. (a) Single-trait analysis; (b) 10-trait analysis (based on the first example of Table 1).

KRUGLYAK 1995; WELLER *et al.* 1998). To get the critical level of the test statistics in the foregoing analysis we employed Monte Carlo simulations with parameters corresponding to H_0 (no QTL in the chromosome, with 5000 runs per each variant). Clearly, this technique can also be used for real data analysis, but it would be much more preferable to take into account the distribution properties of the real data set. The best way to do this in testing significance is the permutation test (DOERGE and CHURCHILL 1996). A few different, although related, questions about the significance of the results can be recognized in the multiple-trait procedure: (i) What is the significance level of the detected QTL?, (ii) which traits significantly contributed to the criterion (multivariate LOD score)?, and (iii) which traits depend significantly on the detected QTL? The difference between the second question and the third is caused by the fact that the information value of a trait may derive from its correlation to other traits of the complex, from the pleiotropic effect of the QTL on this trait, or from both these factors (see Figure 1).

Example 4: Selecting significant traits and effects: The foregoing aspects are illustrated in a simulated example with seven quantitative traits and a chromosome with five intervals (10 cM each) with a QTL residing in the middle of the third interval. The pleiotropic effects of the simulated QTL, the residual correlation matrix, and residual variances were as shown in Table 5. The results can be outlined as follows:

i. To evaluate the significance of the QTL detected by using seven-dimensional mapping analysis, the entire vector of trait values was reshuffled relative to the marker scores (while retaining the structure within the trait complex). For each such permutated data set, the mapping procedure was applied, resulting in a corresponding value of the test statistics LOD score. This process was repeated many times (10,000 in our experiment). The significance of the H_0 hypothesis (no effect of the considered chromosome on the multivariate trait complex) is calculated as the proportion of permutation runs that resulted in LOD values equal to or exceeding LOD* obtained on the nonpermutated data.

TABLE 3

Comparison of the QTL mapping results obtained by the proposed method (based on interval-specific determination of the residual covariance matrix Σ_R) and by using the total covariance matrix (Σ_{total})

				Effect		
Matrix	LOD	Parameter	L (cM)	d_1	d_2	d_3
Σ_R	48.10	δ	0.02	0.010	−0.003	0.007
		σ	1.44	0.048	0.050	0.050
Σ_{total}	27.58	δ	0.04	0.027	−0.023	0.018
		σ	3.60	0.069	0.068	0.064
Trait	1	2	3		Effect	
1	1.00	0.01	0.70		+0.75	
2	0.01	1.00	0.70		−0.75	
3	0.70	0.70	1.00		+0.50	

Three-trait complex was analyzed, with QTL effects and the residual covariance matrix as presented above. The parameters δ and σ denote the bias and standard deviation of the estimated QTL position (L) and QTL effects (d_1, d_2, and d_3).

TABLE 4

The effect of the number of traits on efficiency of QTL mapping analysis with multiple QTL segregating in the mapping population

	No.		Precision of estimation		Test statistics		Interval distribution of the detected QTL (%)					Power (%)
			δ_L	σ_L	$LOD_{0.01}$	LOD_m	1	2	3[a]	4	5	
					Two QTL							
QTL1	19	1 × 1	1.14	12.72	9.19	12.45	9	29	31	18	13	14
	20	2 × 2	−0.02	9.58	11.74	19.21	5	21	50	20	4	61
	22	10 × 10	−0.01	5.63	28.04	57.16	2	9	79	9	1	100
QTL2	23	1 × 1	−0.82	11.82	9.19	12.46	9	30	34	19	8	13
	24	2 × 2	0.12	9.03	11.74	19.66	3	24	50	18	5	68
	30	10 × 10	0.28	4.98	28.04	56.51	1	6	82	10	1	99
					Three QTL							
QTL1	31	1 × 1	1.09	12.56	9.29	13.51	6	31	35	16	12	13
	32	2 × 2	−0.77	10.09	11.66	19.22	7	19	49	20	5	57
	34	10 × 10	0.09	6.67	27.05	46.52	1	13	71	13	2	97
QTL2	31	1 × 1	1.12	14.09	9.29	12.55	10	18	39	15	18	12
	32	2 × 2	0.15	10.24	11.66	18.44	7	20	50	18	5	60
	34	10 × 10	0.31	5.83	27.04	47.60	1	11	71	16	1	97
QTL3	31	1 × 1	−2.92	12.72	9.29	12.15	16	21	39	19	5	12
	32	2 × 2	0.03	9.94	11.66	18.49	5	19	52	18	6	62
	34	10 × 10	0.29	6.38	27.04	47.67	1	11	72	14	2	95

The parameters δ_L and σ_L denote the bias and standard deviation of the estimated QTL position; $LOD_{0.01}$ is the threshold value of the test statistics (obtained by 5000 simulations under the assumption H_0 that the analyzed chromosome does not affect the trait complex); LOD_m is the mean value of the test statistics averaged over the runs with $LOD > LOD_{0.01}$.

[a] The simulated position of each of the two or three QTL on the corresponding chromosomes was the middle of the third interval.

ii. The second test aimed to evaluate the significance of contributions of each of the traits for the QTL detection power. This test is conducted separately for each trait. For this, the individual values of the trait under consideration are reshuffled relative to the remaining data (the other trait values and marker scores). The resulting data set is treated as before and the proportion of runs with LOD ≥ LOD* is used as the measure of significance of the trait contribution. The permutations are always performed regarding all the traits included in the model independently of the contribution value of the remaining traits. Clearly, some traits may prove to be insignificant because they contribute the same information as one (or a few) of the remaining traits. Thus, one can exclude insignificant traits from consideration by creating a new trait set that does not include the insignificant traits(s). This procedure should be applied by simple steps, excluding only one trait per step and repeating the permutation test for the remainder. The last warning is important because after excluding one of the traits at some step, the significance of contributions of the remaining traits may change.

iii. The same procedure as in (ii) can be used to test the significance of the QTL effect for each of the traits. Namely, we calculate the proportion of permutated cases where the estimated QTL effect for the considered trait x_i fits the condition $abs(d_i) \geq abs(d_i^*)$, where d_i^* is the estimated effect on trait x_i obtained on initial (not reshuffled) data.

In the example of Table 5, trait 7 displayed the lowest contribution and hence was removed after the first step. Reevaluation of the remaining complex revealed the next candidate to remove, trait 3, and then, similarly, trait 4. All the remaining traits (1, 2, 5, and 6) showed significant contribution. This trait complex provides also the narrowest confidence interval for the estimated QTL position (σ_L), as shown by the results of bootstrap analysis. The last result means that maintenance of excessive (noninformative) traits is not neutral, a reduced precision of the estimated QTL position being the penalty. Filtering out of the nonsignificant traits should affect the QTL detection power, but further reduction of the trait complex by removing the significant traits may result in a reduced power and lowered mapping precision (see the characteristics obtained for the last two trait combinations, 1, 2, 5, and, especially, 2, 5, 6).

An example of application to real data: We illustrate the efficiency of the proposed approach using real data on a wheat mapping population characterized for 11

TABLE 5

Permutation test of significance for the contribution of the traits: the multitrait LOD and the pleiotropic effects of the QTL

	1–7		1–6		1, 2, 4–6		1, 2, 5, 6		1, 2, 5		2, 5, 6	
Traits	Trait	Effect	Trait	Effect	Trait	Effect	Trait	Effect	Trait	Effect	Trait	Effect
	Significance (%) based on 10,000 permutations for each tested trait combination											
1	0.00	0.07	0.00	0.07	0.00	0.08	*0.00*	*0.07*	0.00	0.08	—	—
2	0.03	84.16	0.02	84.23	0.02	83.87	*0.02*	*83.01*	1.18	85.10	71.99	83.14
3	21.76	5.10	21.85	5.08	—	—	—	—	—	—	—	—
4	18.46	1.56	19.28	1.56	11.94	1.59	—	—	—	—	—	—
5	0.00	0.00	0.00	0.00	0.00	0.00	*0.00*	*0.00*	0.01	0.00	0.01	0.00
6	0.20	63.17	0.19	63.15	0.16	62.56	*0.15*	*61.52*	—	—	73.06	62.62
7	58.40	41.10	—	—	—	—	—	—	—	—	—	—
	Estimates resulting from bootstrap analysis (1000 runs)											
LOD_m	12.72		12.43		11.90		*11.16*		8.83		4.88	
σLOD_m	2.86		2.87		2.71		*2.71*		2.48		1.86	
Power (%)	99.5		99.6		99.7		*99.7*		95.9		47.4	
H^2	0.346		0.324		0.304		*0.290*		0.204		0.172	
σ_{H^2}	0.086		0.079		0.078		*0.070*		0.062		0.083	
L (cM)	24.12		24.17		24.53		*23.81*		23.58		21.03	
σ_L	5.26		5.09		5.00		*3.88*		5.22		7.04	
Traits	1	2	3	4	5	6	7	Effect	σ			
	Occurred correlation matrix, QTL effects, and residual standard deviations											
1	1.0	0.618	0.062	0.165	0.054	0.155	0.156	0.504	0.955			
2		1.0	0.126	0.020	0.112	0.723	0.156	0.037	1.013			
3			1.0	0.236	0.072	0.092	0.007	0.281	1.002			
4				1.0	−0.069	−0.112	0.032	0.283	1.088			
5					1.0	0.034	−0.050	0.674	0.982			
6						1.0	0.105	0.044	0.968			
7							1.0	−0.149	0.960			

The simulated effects, residual correlation matrix, and standard deviations are as shown in the bottom; note that one out of seven traits, no. 7, was simulated as "noise", trait 2 was independent on the QTL but correlated with traits 1 and 6.

On the basis of the permutation test, the significance of contribution to the LOD score as well as the QTL effect was evaluated for each trait. After the first step, trait 7 that appeared to have the lowest contribution was removed. Reevaluation of the remainder complex revealed the next candidate to remove, trait 3, and then, similarily, trait 4. All the remainder traits, 1, 2, 5, and 6, show significant contribution. This complex (italic) also provides the narrowest confidence interval for the estimated QTL position (σ_L), as shown by the results of bootstrap analysis. This filtering out of the nonsignificant traits did not affect the QTL detection power, whereas further reduction of the trait complex by removing the significant traits may result in a reduced power and lowered mapping precision (see the characteristics obtained for the last two trait combinations).

morphological quantitative traits. The experiment was performed on an F_2/F_3 mapping population derived from a cross between a highly stripe-rust-resistant wild emmer wheat *Triticum dicoccoides* (accession no. H52, from Mt. Hermon, Israel) and a *T. durum* cultivar, Langdon, released in North Dakota. The tetraploid wild emmer, *T. dicoccoides*, is the progenitor of cultivated wheat; hence, the genetic dissection of quantitative trait differences between the wild species and the cultivated crop is of great interest from the viewpoint of domestication evolution. It is also important for the ever-increasing utilization of *T. dicoccoides* as a rich genetic resource for wheat improvement. The molecular markers [microsatellites and amplified fragment length polymorphisms (AFLP)] were scored on 150 F_2 individuals resulting in a rather dense genetic map (Peng *et al.* 2000). The quantitative traits were scored on the selfed progeny in field trials conducted in Neve Yaar Agricultural Experimental Station, Israel, during the 1997–1998 cropping season. Eleven quantitative traits were scored on F_3 progeny (for ~10 individual plants from each family): plant height (HT), plant heading date—the days from sowing to heading (HD); spike number/plant (SNP); spike weight/plant (SWP) including the grains, hulls, and rachis; single spike weight (SSW); kernel number/plant (KNP); kernel number/spike (KNS); kernel number/spikelet (KNL); 100-grain weight (GWH); grain yield/plant (YLD); and spikelet number/spike (SLS).

A detailed QTL description of the obtained QTL mapping results on these traits will be presented elsewhere (J. H. Peng, A. B. Korol, T. Fahima, Y. I. Ronin and E. Nevo, unpublished results). Here we employ the obtained data only to illustrate the efficiency of the multitrait analysis, using as an example markers of chro-

TABLE 6

Interval analysis of a multitrait complex that includes 11 morphological traits scored in F_2/F_3 mapping population of wheat *Triticum durum* × *T. dicoccoides* (Peng *et al.* 2000)

No. of traits	LOD_f σ_{LOD_f}	$L\sigma_L$ (cM)	QTL detection power (%) at three significance levels			Significance, by permutation test
			$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$	
1	4.38 1.63	271 74	91	71	42	0.011
11	19.71 3.43	265 36	100	100	100	<0.0001
5	13.82 2.91	262 30	100	100	99	<0.0001

The example is based on markers of chromosome 7A.

The results of single-trait interval analysis (for trait GWH) are compared with those of the entire 11-trait complex and the "filtered" five-trait complex obtained by excluding nonsignificant traits (as in the example shown in Table 5). The traits remaining in the five-trait complex are: GWH, YLD, HD, HT, and SWP. The significance level for each trait and trait complex was calculated using a permutation test (10,000 runs). In addition to the analysis of the initial data set, 1000 bootstrap runs were conducted, enabling us to evaluate the QTL detection power and precision of the parameter estimates. LOD_f and σ_{LOD_f} are the mean value and standard deviation of the test statistics estimated on the basis of 1000 bootstrap runs; correspondingly, L and σ_L are the QTL map position and its standard deviation.

mosome 7A. With single-trait analysis applied separately to each of the traits, only one significant QTL was found on 7A, for trait GWH, with significance level ~0.01 (Table 6). This level should be corrected for multiple comparisons, taking into account the fact that the analyzed traits are correlated (*e.g.*, by using the method based on factor analysis, as suggested by Spelman *et al.* 1996). Therefore, the corrected significance will be even worse. The mapping precision evaluated by bootstrap analysis is not high (σ_L = 74 cM), as one would expect for the modest population size employed (n = 150). Therefore, it makes sense to attempt improvement of the mapping by utilizing the information contained in the entire trait complex, owing to possible pleiotropic effects of the putative QTL and/or correlations between GWH and the remaining traits. This was done exactly in the same way as described in the foregoing simulated example presented in Table 5. First, the entire complex of 11 traits was analyzed and then the traits that did not contribute significantly to the test statistics were removed. The results presented in Table 6 and Figure 5 show a more than twofold increase in the mapping precision (σ_L decreased from 74 to 30 cM) and an increase in detection power that is especially clear at higher significance level (98.9% *vs.* 42.9%).

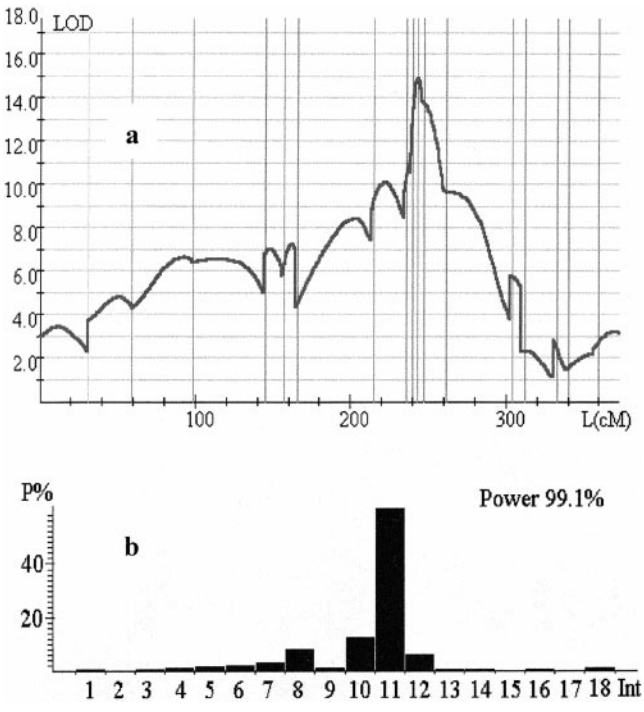

Figure 5.—Joint analysis of 11 traits scored in F_2/F_3 mapping population of wheat *Triticum durum* × *T. dicoccoides* using markers of chromosome 7A. The results of removing nonsignificant traits are presented. (a) LOD score distribution along chromosome 7A for the 5-trait complex (GWH, YLD, HD, HT, and SWP). (b) Interval distribution of the maximum LOD values along chromosome 7A based on 1000 bootstrap runs. Both graphs are outputs of the MultiQTL package (http://www.MultiQTL.com).

DISCUSSION

A multivariate generalization of our previous two-trait QTL mapping analysis (Korol *et al.* 1987, 1995, 1998a; Ronin *et al.* 1995, 1999; see also Jiang and Zeng 1995) is proposed here. It is not difficult to extend this method

to other situations (*e.g.*, analyzing F_3 populations), to deal with linked QTL (similar to the analysis of Korol *et al.* 1998a; Ronin *et al.* 1999), to combine it with selective genotyping design (Ronin *et al.* 1998; Henshall and Goddard 1999), or to adopt composite interval mapping (Jansen and Stam 1994; Zeng 1994). Especially promising may be its application to fine mapping (Y. I. Ronin, E. Britvin, E. Nevo and A. B. Korol, unpublished results). Indeed, the dramatic increase in mapping resolution derived from using the entire multivariate complex, as compared with univariate or even bivariate analysis, effectively increases the score D_n^2 that was found to affect the mapping resolution (Darvasi and Soller 1997). Consequently, it becomes reasonable to saturate the revealed intervals by additional markers even at modest population sizes like 200–500 individuals; usually this is pointless because with small effects no increase in precision is expected by addition of new markers to the map (Darvasi *et al.* 1993). Therefore, the transition from a single- or even two-trait analysis to treatment of genuine multiple-trait complexes significantly improves all aspects of utilizing the mapping information contained in the data.

However, the application of multivariate complexes not only increases the QTL detection power, mapping resolution, and estimation accuracy but it may also increase the power of discriminating various important hypotheses that concern the genetic architecture of complex traits, such as linkage *vs.* pleiotropy (Schork *et al.* 1994; Jiang and Zeng 1995; Almasy *et al.* 1997; Lebreton *et al.* 1998; Ronin *et al.* 1999), genetic interaction within and across QTL (additive *vs.* dominant or overdominant effects, and additive *vs.* epistatic effects or canalization; Ronin *et al.* 1999; Shook and Johnson 1999), and QTL-environment interaction (Fry *et al.* 1998; Korol *et al.* 1998b). Multivariate QTL analysis may be helpful in genetic dissection of such types of complex traits as multifactorial diseases (Mansfield *et al.* 1997), development (Wu *et al.* 1999), longevity and aging (Nuzhdin *et al.* 1997), behavior (Plomin and Craig 1997; Wehner *et al.* 1997), fitness-related trait complexes and species differentiation (Zeng *et al.* 2000), heterosis (Xiao *et al.* 1995), marker-assisted breeding (Lande and Thompson 1990; Visscher *et al.* 1996), characterizing the regulatory networks of structural genes (Damerval *et al.* 1994), bridging between gene-structure-and-function studies (*e.g.*, when looking for functions of massively expressed sequence tags; Lahbib-Mansais *et al.* 1999), analyzing the genetic transmission system (breeding system, recombination, and mutation control; Korol *et al.* 1994; Bernacchi and Tanksley 1997), and so on.

As a not-so-remote analogy, one could compare the situation of multivariate QTL analysis with that characteristic of medical diagnoses: excluding simple situations, a good physician will never rely on one trait (symptom or analysis, etc.) Instead, he/she will try to take into account all available information concerning the patient. However, this does not mean that increasing the number of traits to be analyzed simultaneously will necessarily improve the quality of the QTL mapping results. A technical obstacle with high dimensionality is an increasing probability that many loci may affect the analysis along the chromosome, whereas a small-to-moderate population size could hardly justify fitting more than two or three linked QTL simultaneously. Another problem is the interpretation of the results. Therefore, in choosing the initial set of traits for joint QTL analysis, one may find it reasonable to restrict such sets by functionally related traits. The examples presented in this article, on both simulated and real data, show that maintenance of excessive traits in the model may be penalized. These concerns indicate that in spite of high potential and biological "compatibility" of the multiple-trait analysis to the main targets of QTL analysis, a lot of work remains to be done to fully extract the mapping information hidden in the collected data.

An additional complication that is worth mentioning is the possible effect of the model assumption on the obtained results. It was shown earlier that for testing for linkage, erroneous models may lead to valid tests for linkage (Wright and Kong 1997). For example, QTL mapping analysis may be quite robust to violations of the assumption of normality in single-trait situations (Korol *et al.* 1996b) and more sensitive in multivariate ones. Likewise, the assumption of homoscedastic distributions (*i.e.*, equal residual variances in QTL groups) that is usually applied automatically may be wrong, leading to reduced QTL detection power and biased estimates of parameters. On the contrary, if a correct model is fitted, this may increase the detection power and mapping accuracy compared to situations when no such disturbances exist. We demonstrated these effects earlier for single- and two-trait analysis (Korol *et al.* 1995, 1996a). Especially important is the assumption of a single QTL per chromosome, which being violated may lead to the LOD score peaking in the wrong place (see Knott and Haley 1992; Wright and Kong 1997). For the two-trait case we found that joint analysis of correlated traits increases the power of the test aimed to discriminate between the single QTL and two-linked QTL situations (Ronin *et al.* 1999). All these aspects should be taken into account in multivariate QTL analysis.

The described approach is implemented in the MultiQTL package (http://www.MultiQTL.com) for both single- and two-linked QTL models.

We thank M. Soller, J. Weller, G. Churchill, and three anonymous referees for helpful comments and suggestions on the first version of the manuscript. This study was supported by the Israeli Science foundation (grants 02198 and 9048/99), the United States-Israel Binational Science Foundation (grant 4556), and the German-Israeli Cooperation Project (DIP project funded by the Internationales Büro Deutsch-Israelische des BMBF Projektkooperation).

LITERATURE CITED

Allison, D. B., B. Thiel, P. St. Jean, R. C. Elston, M. C. Infante *et al.*, 1998 Multiple phenotype modeling in gene-mapping studies of quantitative traits: power advantages. Am. J. Hum. Genet. **63:** 1190–1201.

Almasy, L., T. D. Dyer and J. Blangero, 1997 Bivariate quantitative trait linkage analysis: pleiotropy versus co-incident linkages. Genet. Epidemiol. **14:** 953–958.

Amos, C. I., R. C. Elston, G. E. Bonney, B. J. Keats and G. S. Berenson, 1990 A multivariate method for detecting genetic linkage, with application to a pedigree with an adverse lipoprotein phenotype. Am. J. Hum. Genet. **47:** 247–254.

Bernacchi, D., and S. D. Tanksley, 1997 An interspecific backcross of *Lycopersicon esculentum* × *L. hirsutum*: linkage analysis and a QTL study of sexual compatibility factors and floral traits. Genetics **147:** 861–877.

Boomsma, D. I., and C. V. Dolan, 1998 A comparison of power to detect a QTL in sib-pair data using multivariate phenotypes, mean phenotypes, and factor scores. Behav. Genet. **28:** 329–340.

Damerval, C., A. Maurice, J. M. Josse and D. de Vienne, 1994 Quantitative trait loci underlying gene product variation: a novel perspective for analyzing regulation of genome expression. Genetics **137:** 289–301.

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

Darvasi, A., and M. Soller, 1997 A simple method to calculate resolving power and confidence interval of QTL map position. Behav. Genet. **27:** 125–132.

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

Doerge, R. W., and G. A. Churchill, 1996 Permutation tests for multiple loci affecting a quantitative character. Genetics **142:** 285–294.

Fry, J. D., S. V. Nuzhdin, E. G. Pasyukova and T. F. Mackay, 1998 QTL mapping of genotype-environment interaction for fitness in *Drosophila melanogaster*. Genet. Res. **71:** 133–141.

Henshall, J. M., and M. E. Goddard, 1999 Multiple-trait mapping of quantitative trait loci after selective genotyping using logistic regression. Genetics **151:** 885–894.

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

Jiang, C., and Z-B. Zeng, 1995 Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics **140:** 1111–1127.

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

Korol, A. B., I. A. Preygel and N. Bocharnikova, 1987 Linkage between loci of quantitative traits and marker loci. 5. Simultaneous analysis of a set of markers and quantitative traits. Genetika **23:** 1421–1431 (English translation in Sov. Genet. 1988, **23:** 996–1004).

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

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

Korol, A. B., Y. I. Ronin, Y. Tadmor, A. Bar-Zur, V. M. Kirzhner *et al.*, 1996a Estimating variance effect of QTL: an important prospect to increase the resolution power of interval mapping. Genet. Res. **67:** 187–194.

Korol, A. B., Y. I. Ronin and V. M. Kirzhner, 1996b Linkage between loci of quantitative traits and marker loci. Resolution power of three statistical approaches in single marker analysis. Biometrics **52:** 426–441.

Korol, A. B., Y. I. Ronin, P. Hayes and E. Nevo, 1998a Multi-interval mapping of correlated trait complexes: simulation analysis and evidence from barley. Heredity **80:** 273–284.

Korol, A. B., Y. I. Ronin and E. Nevo, 1998b Approximated analysis of QTL-environmental interaction with no limits on the number of environments. Genetics **148:** 2015–2028.

Lahbib-Mansais, Y., G. Dalias, D. Milan, M. Yerle, A. Robic *et al.*, 1999 A successful strategy for comparative mapping with human ESTs: 65 new regional assignments in the pig. Mamm. Genome **10:** 145–153.

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

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

Lander, E., and L. Kruglyak, 1995 Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat. Genet. **11:** 241–247.

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

Lebreton, C. M., P. M. Visscher, C. S. Haley, A. Semikhodskii and S. A. Quarrie, 1998 A nonparametric bootstrap method for testing close linkage vs. pleiotropy of coincident quantitative trait loci. Genetics **150:** 931–943.

Mangin, B., P. Thoquet and N. Grimsley, 1998 Pleiotropic QTL analysis. Biometrics **54:** 88–99.

Mansfield, T. A., D. B. Simon, Z. Farfel, M. Bia, J. R. Tucci *et al.*, 1997 Multilocus linkage of familial hyperkalaemia and hypertension, pseudohypoaldosteronism type II, to chromosomes 1q31-42 and 17p11-q21. Nat. Genet. **16:** 202–205.

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

Nuzhdin, S. V., E. G. Pasyukova, C. L. Dilda, Z-B. Zeng and T. F. Mackay, 1997 Sex-specific quantitative trait loci affecting longevity in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **94:** 9734–9739.

Olson, J. M., S. Rao, K. Jacobs and R. C. Elston, 1999 Linkage of chromosome 1 markers to alcoholism-related phenotypes by sib pair linkage analysis of principal components. Genet. Epidemiol. **17** (Suppl. 1): S271–276.

Peng, J. H., A. B. Korol, T. Fahima, M. S. Röder, Y. I. Ronin *et al.*, 2000 Molecular genetic maps in wild emmer wheat, *Triticum dioccoides*: genome-wide coverage, massive negative interference, and putative quasi-linkage. Genome Res. **10:** 1509–1531.

Plomin, R., and I. Craig, 1997 Human behavioural genetics of cognitive abilities and disabilities. Bioessays **19:** 1117–1124.

Ronin, Y. I., V. M. Kirzhner and A. B. Korol, 1995 Linkage between loci of quantitative traits and marker loci. Multitrait analysis with a single marker. Theor. Appl. Genet. **90:** 776–786.

Ronin, Y. I., A. B. Korol and Y. Weller, 1998 Selective genotyping to detect quantitative trait affecting multiple traits. Theor. Appl. Genet. **97:** 1169–1178.

Ronin, Y. I., A. B. Korol and E. Nevo, 1999 Single- and multiple-trait analysis of linked QTL: some asymptotic analytical approximation. Genetics **151:** 387–396.

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

Schork, N. J., A. B. Weder, M. Trevisan and M. Laurenzi, 1994 The contribution of pleiotropy to blood pressure and body-mass index variation: the Gubbio Study. Am. J. Hum. Genet. **54:** 361–373.

Shook, D. R., and T. E. Johnson, 1999 Quantitative trait loci affecting survival and fertility-related traits in *Caenorhabditis elegans* show genotype-environment interactions, pleiotropy and epistasis. Genetics **153:** 1233–1243.

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

Spelman, R. J., W. Coppieters, L. Karim, J. A. van Arendonk and H. Bovenhuis, 1996 Quantitative trait loci analysis for five milk production traits on chromosome six in the Dutch Holstein-Friesian population. Genetics **144:** 1799–1808.

Visscher, P. M., C. S. Haley and R. Thompson, 1996 Marker-assisted introgression in backcross breeding programs. Genetics **144:** 1923–1932.

Wehner, J. M., R. A. Radcliffe, S. T. Rosmann, S. C. Christensen, D. L. Rasmussen *et al.*, 1997 Quantitative trait locus analysis of contextual fear conditioning in mice. Nat. Genet. **17:** 331–334.

Weller, J. I., G. R. Wiggans, P. M. Van Raden and M. Ron, 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:** 998–1002.

Weller, J. I., J. Z. Song, D. W. Heyen, H. A. Lewin and M. Ron, 1998 A new approach to the problem of multiple comparisons in the genetic dissection of complex traits. Genetics **150:** 1699–1706.

Williams, J. T., P. Van Eerdewegh, L. Almasy and J. Blangero, 1999 Joint multipoint linkage analysis of multivariate qualitative and quantitative traits. I. Likelihood formulation and simulation results. Am. J. Hum. Genet. **65:** 1134–1147.

Wright, F. A., and A. Kong, 1997 Linkage mapping in experimental crosses: the robustness of single-gene models. Genetics **146:** 417–425.

Wu, W. R., W. M. Li, D. Z. Tang, H. R. Lu and A. J. Worland, 1999 Time-related mapping of quantitative trait loci underlying tiller number in rice. Genetics **151:** 297–303.

Xiao, J., J. Li, L. Yuan and S. D. Tanksley, 1995 Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. Genetics **140:** 745–754.

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

Zeng, Z-B., J. Liu, L. F. Stam, C. H. Kao, J. M. Mercer *et al.*, 2000 Genetic architecture of a morphological shape difference between two *Drosophila* species. Genetics **154:** 299–310.

Communicating editor: G. A. Churchill