Approximate Analysis of QTL-Environment Interaction with No Limits on the Number of Environments

Abraham B. Korol, Yefim I. Ronin and Eviatar Nevo

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

Manuscript received December 9, 1996
Accepted for publication December 31, 1997

ABSTRACT

An approach is presented here for quantitative trait loci (QTL) mapping analysis that allows for QTL × environment (E) interaction across multiple environments, without necessarily increasing the number of parameters. The main distinction of the proposed model is in the chosen way of approximation of the dependence of putative QTL effects on environmental states. We hypothesize that environmental dependence of a putative QTL effect can be represented as a function of environmental mean value of the trait. Such a description can be applied to take into account the effects of any cosegregating QTLs from other genomic regions that also may vary across environments. The conducted Monte-Carlo simulations and the example of barley multiple environments experiment demonstrate a high potential of the proposed approach for analyzing QTL × E interaction, although the results are only approximated by definition. However, this drawback is compensated by the possibility to utilize information from a potentially unlimited number of environments with a remarkable reduction in the number of parameters, as compared to previously proposed mapping models with QTL × E interactions.

DIFFERENTIAL expression of a phenotypic trait by genotypes across environments, or genotype × environment (G × E) interaction, is an old problem of primary importance for quantitative genetics and its applications in breeding, conservation biology, theory of evolution, and human genetics (Eberhard and Russel 1966; Falconer 1981; Via and Lande 1987; Tiret *et al.* 1993; Wu and Stettler 1997). Recent successful attempts to dissect quantitative variation into Mendelian genes employing molecular markers (mapping quantitative trait loci, or QTLs) have shifted the focus of G × E interaction analysis from the genotype to gene level (*e.g.*, Paterson *et al.* 1991; Hayes *et al.* 1993; Asins *et al.* 1994; Sari-Gorla *et al.* 1997). For breeding purposes, the primary concern is possible environmental instability in manifestation of mapped QTLs that might become candidates for marker-assisted selection. To evaluate stability of QTL effects in crop species, dozens of immortal mapping populations have been developed for trait scoring under various environmental conditions (Hayes 1994).

Several algorithms and computer packages have been proposed to conduct QTL mapping, allowing for QTL × E interaction effects (Hayes *et al.* 1993; Jansen *et al.* 1995; Tinker and Mather 1995; Romagosa *et al.* 1996; Beavis and Keim 1996; Utz and Melchinger 1996). In addition to testing the hypothesis of QTL × E interaction, simultaneous treatment of data from multiple environments provides a significant increase in statistical power of QTL detection and accuracy of the estimates of QTL position and effect (Jansen *et al.* 1995). However, such an analysis is limited by situations where the environments can be obviously characterized by some parameters, like day length or irrigation-fertilization treatments, *etc.* [like the "fixed effects" model of analysis of variance (ANOVA)]. When these characteristics are not available, the application of the "general" QTL × E mapping model (see below) is accompanied by a tremendous number of parameters involved in the model that increase as a product of the identified QTLs and the number of environments where the traits were measured. In such a case, one could think about a "random effects" model so that the number of parameters for each QTL will include only main effect, the variance of QTL × E interaction, and QTL position. Although this option seems very attractive, it has its own drawbacks, especially if we are going to deal with environmental variation associated with different localities. Indeed, some (or many) localities may manifest quite repeatable differences from each other, justifying the "fixed model" approach (Baker 1996). Moreover, the information of geographically-specific QTL effects may be of practical importance. In such a case, a fixed effects model is, whenever possible, preferable over the random effects model, because the latter hides the biological (geographic) specificity of the QTL effect, compressing all the results to an estimate of variance. The major goal of this paper is to present an approach of QTL mapping analysis, allowing for QTL × E interaction across a large (in fact, unlimited) number of environ-

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

ments, without the necessity for a corresponding increase in the number of parameters. The proposed model is especially relevant in situations of geographic variation of external conditions, where the fixed model approach is desirable but not easy to implement.

THE MODEL

Approximated description of environmental dependence of QTL effect: The main distinction of the proposed model is in the choice of approximation of the dependence of putative QTL effects on environmental states. In reality, each environment is a complex of abiotic (temperature, humidity, ion concentration, *etc.*), biotic (parasites, pathogens, competitors, *etc.*), and agrotechnical features. These could strongly affect the manifestation of quantitative traits and the effects of QTL but are difficult to characterize quantitatively. As first suggested by Eberhard and Russel (1966), we advocate that the measured trait values of the mapping population (*e.g.*, trait means) may serve as objective integral characteristics of the environmental state. Accordingly, a larger number of traits should provide a better "bioindication." In the simplest form, one can approximate the environmental dependence of the effect of allele substitution at a QTL by a polynomial over the mean values of the same trait across the environments. The following example of a QTL mapping in a barley experiment (Hayes *et al.* 1996) with measurements conducted in many environments, illustrates the idea (Figure 1).

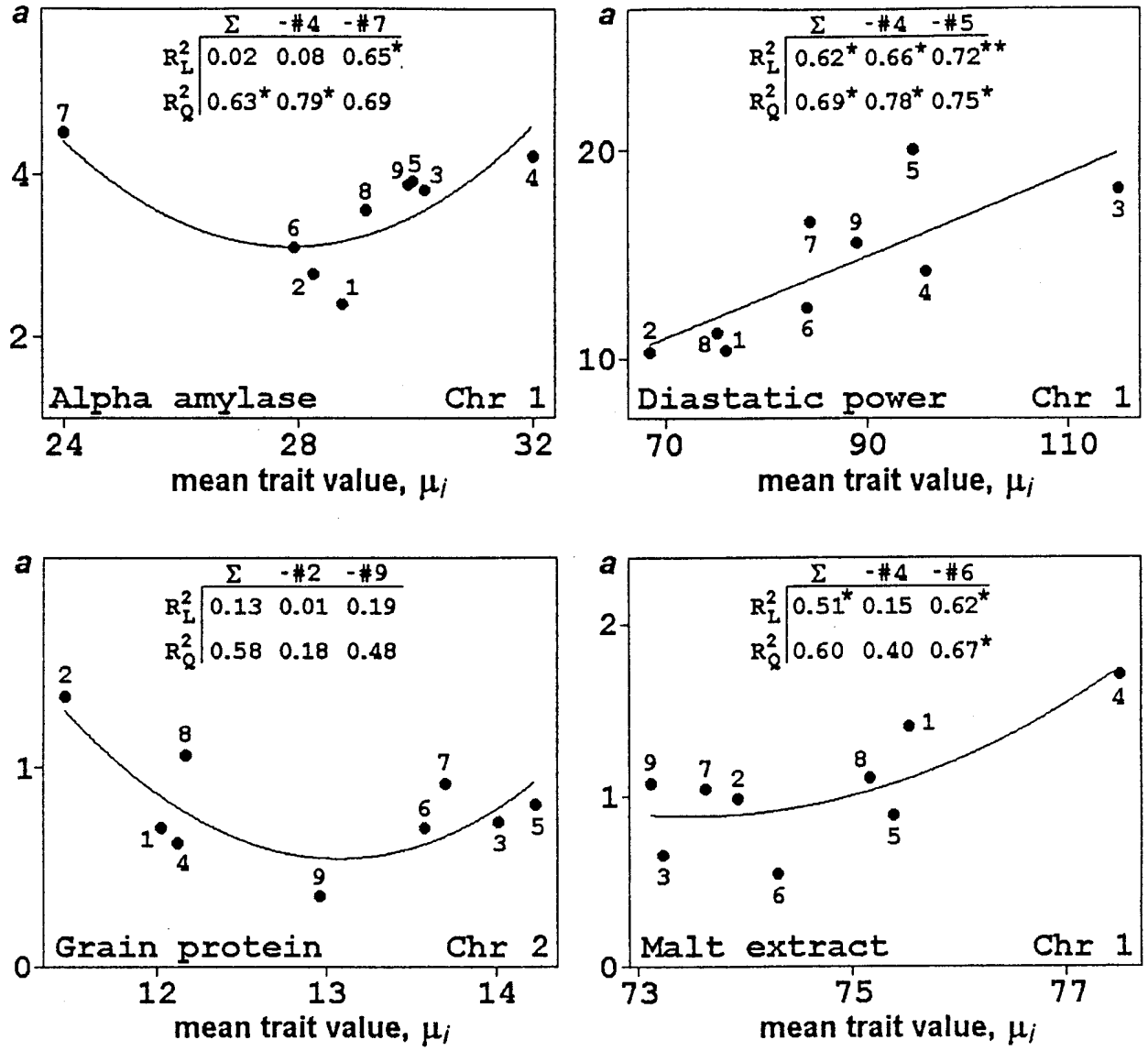

Figure 1.—Regression approximation of QTL substitution effect a as a function of mean trait value μ_i. Data on barley malting quality traits from Hayes *et al.* (1996). Circles represent pairs (a_i,μ_i) where i is the number of environment. RL and RQ denote the explained part of variation of the estimated QTL effect based on linear and quadratic regression, respectively; Σ, all nine environments used; #k, the kth environment is excluded; * and **, regression is significant at $P < 0.05$ and 0.01.

For some putative QTLs, the dependence on mean value of the respective traits explains a large part of the environmental variation of the QTL effect. This suggested approach does not exclude the possibility to take into account any additional information, like temperature, day length, water regime, *etc.*, that might characterize the environments (*e.g.*, Jansen *et al.* 1995). These "physical" characteristics can be introduced into the model parallel to the bioindicatory terms (*e.g.*, polynomial over the mean values) together with terms characterizing the dependence of the putative QTL effect on interaction between the physical and bioindicatory factors. Another approach to analyze QTL × E interaction without direct specification of the physical characteristics of the environments was recently proposed by Romagosa *et al.* (1996). Their algorithm is based on clustering the environments using a few (*e.g.*, two) detected QTL with most variable effects across environments. Actually, this is a different version of the same general idea of bioindicators as a tool for characterizing "anonymous" environments.

Clearly, the results one could obtain by means of the method of QTL × E analysis proposed in this paper will be approximate, allowing, at best, to consider the major part of QTL × E interaction. However, as will be demonstrated below, the possibility to work with an unlimited number of environments without increasing the number of parameters, as compared to usual mapping models with QTL × E interactions, may significantly offset loss-of-accuracy drawback, resulting in increased power to detect QTL × E interactions and in improved accuracy of estimates of QTL genomic location.

Mixture-model of interval QTL mapping: Consider a simplified situation when the trait of interest (x) depends on a single QTL, Q/q. We will confine the analysis to dihaploid mapping populations (which also applies to backcrosses and recombinant inbreds), but it can easily be extended to other population structures. Then, for an arbitrary genotype of the mapping population, the trait measurement in the ith environment can be presented as

$$x_i = \mu_i + 0.5ga_i + e_i, \qquad (1)$$

where μ_i is the mean trait value in the ith environment, g is either +1 (for QQ genotypes) or −1 (for qq genotypes), a_i is the effect of allele substitution at putative QTL on trait in environment i, and e_i is a random variable with zero mean and variance σ_i^2. If we find $a_i \sim a$ for any i, then no G × E interaction is manifested by Q/q.

Assume that Q/q resides in some interval $(k,k+1)$ of a chromosome marked by a series of marker loci, M_j/m_j, with recombination rates r_1 and r_2 in $M_k/m_k - Q/q$ and $Q/q - M_{k+1}/m_{k+1}$, respectively. For simplicity, we confined the analysis to the "no interference" case. For a dihaploid (backcross) mapping population, the expected densities of the trait x in each of the four marker groups $U_{m_km_{k+1}}(x) = U_1(x)$, $U_{M_km_{k+1}}(x) = U_2(x)$, $U_{m_kM_{k+1}}(x) = U_3(x)$, and $U_{M_kM_{k+1}}(x) = U_4(x)$ can be written as

$$U_i(x) = \pi_i f_{qq}(x) + (1 - \pi_i) f_{QQ}(x), \qquad i = \overline{1,4}, \qquad (2)$$

where the proportions $\pi_i = \pi_i(r_1,r_2)$ are dependent on r_1 and r_2. Here, $f_{qq}(x)$ and $f_{QQ}(x)$ are the trait density distributions in the QTL groups qq and QQ, respectively. 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 $\pi_4 = 1 - \pi_1$.

In a single-environment formulation, one could test whether or not the observed variation of x is associated with segregation in interval $M_k/m_k - M_{k+1}/m_{k+1}$ and identify the corresponding locus Q/q. Provided recombination rate between marker loci is known, the vector of n_1 parameters specifying the putative QTL can be presented as $\theta_{n_1} = \{r,\mu,a,\sigma^2\}$. The assumption of no association between segregation in $M_k/m_k - M_{k+1}/m_{k+1}$ interval can formally be presented by another set of parameters, $\theta = \theta_{n_0} = \{\mu,\sigma^2\}$. The null hypothesis $\{H_0\colon \theta = \theta_{n_0}\}$, as contrasted with the alternative $\{H_1\colon \theta = \theta_{n_1}\}$, can be investigated with the likelihood ratio test approach (Wilks 1962). If H_0 is true, the statistic

$$\chi^2(H_1 \textit{ vs. } H_0) = 2\ln\left[\max_{\theta_{n_1} \in S_1} L(\theta_{n_1}) / \max_{\theta_{n_0} \in S_0} L(\theta_{n_0})\right] \qquad (3)$$

is distributed asymptotically as chi-square with $n_1 - n_0$ d.f., where S_0 and S_1 are the parameter spaces corresponding to H_0 and H_1, respectively (Wilks 1962; Zeng 1994). Thus, if χ^2 exceeds some critical value corresponding to a preset level of significance α, the null hypothesis can be rejected. In such a case, numerical values of the parameters that maximize $L(\theta_{n_1})$ are considered maximum likelihood estimates of the parameters characterizing Q/q, its effect and position (Weller 1986; Lander and Botstein 1989; Knott and Haley 1992). As applied to (3), the test statistic will have $n_1 - n_0 = 4 - 2 = 2$ d.f.

In multiple environments, we could use the foregoing to trait measurements obtained under several environmental conditions. Namely, when comparing the foregoing alternatives H_0 and H_1, QTL × E interaction effects could be included in the model and tested against the alternative of no QTL × E interaction. In other words, an additional group of hypotheses $\{H_2\colon \theta = \theta_{n_2}\}$ could be considered that assume a dependence of the target QTL effect and, possibly, of the residual variance, on environment. Vector θ_{n_2} of the full model, corresponding to H_2 with environment-specific parameters a_i, σ_i^2, and μ_i will then contain $3p + 1$ components, where p is the number of environments. Consequently, θ_{n_1} specifying H_1 (constant effect $a_i \sim a$ of the QTL across environments, though allowing for variable σ_i^2 and μ_i) contains $2p + 2$, while θ_{n_0} (no QTL on the tested chromosome) contains $2p$ parameters.

In the simplified case of only one QTL segregating in the mapping population, no correlation between trait measurements across environments are expected. With this assumption, instead of the test statistics (3), one can build its multi-environmental equivalent χ^2 (H_1 *vs.* H_0)

with $df = 2p + 2 - 2p = 2$. If H_0 is rejected ($a_i \neq 0$), then the obvious benefit of the corresponding multi-environmental model is the striking increase in the number of measurements, resulting in higher precision of parameter estimates (*e.g.*, Jansen *et al.* 1995). No less important is the possibility to conduct the following two tests:

$$\chi^2(H_2 \textit{ vs. } H_0) = 2\ln[\max_{\theta_{n_2} \in S_2} L(\theta_{n_2}) / \max_{\theta_{n_0} \in S_0} L(\theta_{n_0})] \quad (3a)$$

with $df = 3p + 1 - 2p = p + 1$, and

$$\chi^2(H_2 \textit{ vs. } H_1) = 2\ln[\max_{\theta_{n_2} \in S_2} L(\theta_{n_2}) / \max_{\theta_{n_1} \in S_1} L(\theta_{n_1})] \quad (3b)$$

with $df = 3p + 1 - (2p + 2) = p - 1$. Note that the same d.f. for the test statistics will be obtained if the entire analysis is conducted for the centered data, so that μ_i are subtracted from the individual measurements at corresponding environments.

The asymptotic distribution of the test statistics (3) in the multi-interval mapping remains unknown (see Zeng 1994), but one could use extensive Monte-Carlo simulations in order to obtain an empirical critical value of the statistics for each considered situation. Our previous simulation studies (Korol *et al.* 1995) have shown that the chi-square distribution is a good approximation for the test statistic (3), and here we will demonstrate that it may also be suitable for the test statistic (3b).

Regression specification of QTL × E interaction: Ignoring possible variation of the QTL effect among environments may lead to erroneous breeding decisions in subsequent applications of the mapping results, an accompanied reduction in the power, and loss of precision in estimated QTL effects and genome location. On the contrary, accounting for QTL × E interaction in the data obtained in multiple environments can strongly increase the resolution of the mapping experiment (Jansen *et al.* 1995; Tinker and Mather 1995). However, this proficiency is seriously attenuated by the necessity to build into the mapping model a large number of parameters specifying the working hypothesis of the QTL effects. For example, an experiment with 10 environments will require a model with 31 parameters when evaluating a single interval.

According to the proposed approach, the unknown effects a_i and, if desirable, the residual variances σ_i^2 are represented by low degree polynomials. For instance, with a cubic approximation for a_i and a quadratic for σ_i^2, we will only need seven parameters instead of 20! Clearly, the main question remains to what extent the bioindicating trait, or a (linear) combination of traits, will indeed be informative with respect to the dependence of the target QTL effect on environmental states. No prior answer is possible, but the foregoing examples on barley (see Figure 1) demonstrate that such an assumption is quite realistic, even with the simplest univariate mode of the approximation, $a_i = f(\mu_i)$. Thus, in the log-likelihood functions $\ln L(\theta_{n_k})$ ($k = 0,1,2$) from Equation 3, a and b, we could replace the corresponding coordinates of the parameter vectors θ_{n_i} by polynomials:

$$a_i = \alpha_0 + \alpha_1\mu_i + \alpha_2\mu_i^2 + \ldots + \alpha_s\mu_i^s,$$
$$\sigma_i^2 = \beta_0 + \beta_1\mu_i + \beta_2\mu_i^2 + \ldots \beta_t\mu_i^t, \qquad i = \overline{1,p}. \quad (4)$$

In other words, instead of estimates of r_1, a_i, and σ_i^2, the procedure will provide ML-estimates of r_1 and regression coefficients α_0, α_1,..., α_s, β_0, β_1, and β_t. The following information about these coefficients will be useful. One can represent the approximation (4) in form of deviations from the mean values of the trait μ averaged over environments, *i.e.*, with terms $\alpha_k(\bar{\mu} - \mu_i)^k$ and $\beta_j(\bar{\mu} - \mu_i)^j$, instead of $\alpha_k\mu_i^k$ and $\beta_j\mu_i^j$. Then, the genetic interpretation of the coefficient α_0 is that it specifies the average substitution effect at the putative QTL, whereas coefficients α_i ($i > 0$) reflect the stability of the QTL effect over environments. This parallels the stability analysis of Eberhard and Russel (1966), although they used only linear regression in their model. There are different reasons why linear approximation for QTL dependence on environment may be not sufficient. For example, one may mention the effect of canalization of gene effects (Gilbert 1961; Rendel 1967; Korol *et al.* 1981, 1994). Likewise, the coefficient β_0 specifies the residual variance under average conditions, whereas β_i ($i > 0$) reflect the stability of the residual variance with deviation from average conditions. The changes in the residual variance may result from either QTL × E interaction at other sections of the genome not accounted by the model or dependence of nongenetic components of residual variation on environment.

The degrees of polynomials in Equation 4 cannot be predetermined before the mapping analysis. By contrast, the analysis includes model adjustment with a series of polynomials $a(\mu) = P_{a_s}(\mu)$ and $\sigma^2(\mu) = P_{\sigma_t^2}(\mu)$, and the final degrees s and t are to be chosen on the basis of maximum statistical significance of the hypothesis QTL × E interaction *vs.* the alternative of no QTL × E interaction (see further analysis).

An important point of concern with the proposed approach is how to proceed in a situation where the employed model allowed us to detect a significant QTL effect, but QTL × E interaction was not detected. Does it mean that no QTL × E interaction is characteristic of the revealed QTL or, alternatively, that this interaction exists, but the chosen parametrization (*e.g.*, regression of QTL effects on mean trait values across environments) poorly approximates the real dependence of the QTL effect on environment. One of the possible ways to overcome this obstacle will be presented below.

Obtaining parameter estimates: Maximum likelihood estimates of all of the parameters, including α and β, are obtained using the procedure of numerical multiparameter optimization of functions $L(\theta)$ from Equation 3, a and b. Optimization was by modified gradient method (Himmelblau 1972). The possibility of multiple maxima was excluded by using various sets of starting values.

RESULTS

The efficiency of the proposed method was tested through Monte-Carlo simulations. Three groups of situations were simulated: a single QTL (situations S_1–S_3), two unlinked QTLs (situations S_4–S_5), and several unlinked QTLs (situation S_6) (Table 1).

Single QTL: In the situation with a single QTL, no "between-environment" correlation is expected for the residual (within QTL groups) variation. Thus, the log-likelihood functions 3a and 3b for the mixture model 3a and 3b could be calculated by summing up over all environments and employing the polynomials of Equation 4. This assumes implicitly that after removing the effects of the QTL under consideration, the residuals are independent across environments. Clearly, such an idealization is correct if all residual genetic variation of the quantitative trait is taken into account by markers of other genomic regions, such as cofactors (Jansen and Stam 1994; Zeng 1994). This may not be the case, calling into question the applicability of the proposed approach to real data analysis. It is indeed a very serious problem, but as shown in the following section, the conclusions may be fairly promising.

As a first step in demonstrating the idea of our method, we here consider the simplest case of a single QTL. The dependence of the simulated QTL effect and the residual variance on environment was modeled as cubic and quadratic functions, respectively (see Table 1). The simulated experiment included 10 environments with mean value of the trait (μ_i) linearly increasing from $\mu_1 = 0$ to $\mu_{10} = 3.6$. The target QTL was positioned in the middle of the third interval of six of a linkage group. Each interval consisted of 24 cM. The size of the mapping population (either dihaploid or backcross) was $n = 200$.

The results obtained with polynomials of different degrees corroborate the expectation that the best resolution is achievable when the adjusted polynomials are of the same degree as those employed in generating the data (not shown). We found such a correspondence to be more important for approximating the substitution effects a_i than the residual variances σ_i^2.

Our intention was to compare the general model (MG), specifying all effects a_i and residual variances σ_i^2 with the proposed model (MA), which utilizes a polynomial approximation of a_i and σ_i^2 (Equation 4) as functions of an environmental bioindicator. Here, we used population mean values of the same trait, but there are other possibilities, *e.g.*, mean values of other traits or some other scores characterizing the performance of the target population or even other species. The main criteria for comparison include the power of detection of QTL effect and QTL × E interaction, and the accuracy and precision of the estimated chromosomal location of the detected QTL. In addition, we employed Akaike's information criterion (AIC), which takes into account the cost of an increased number of parameters in the model (Bozdogan 1987).

The results presented in Table 2 show that adequate approximation of $a(\mu)$ results in an appreciable increase in the power of both tests: H_1 *vs.* H_0 (presence of a QTL, allowing for $a_i = \text{const}$ and $\sigma_i^2 \neq \text{const}$), and H_2 *vs.* H_1 (presence of QTL × E interaction, allowing for $\sigma_i^2 \neq \text{const}$). Utilization of $a(\mu)$ and $\sigma(\mu)$ polynomial approximations resulted in an improved precision in the estimated QTL position (compare the estimated position L and its standard error s_L for MA and MG in Table 2). In situation S_1 with the smallest average simulated QTL effect, the power to detect the QTL at the significance level of 1% was 45 and 34% for MA and MG, respectively, whereas the corresponding figures for detection of QTL × E interaction were 41 and 29%. Note that this increased resolution of the MA model as compared to the MG model was obtained using only eight parameters (less than half of that in MG). Hence, the superiority of the MA model over MG is reflected in the values of AIC. Likewise, the polynomial model provided ~1.5–2-fold reduction in the estimated confidence interval for the a_i estimates across environments (Figure 2).

With simulated data, it is easy to compare "the adequate" and "nonadequate" approximations simply because we know the employed model. The results in Table 3 illustrate this point. As one can see, the adequate model (MA_3) gave the highest power of detection of both the presence of the QTL in question and QTL × E interaction, and the most accurate and precise estimate of QTL location. Note that even the poorest approximate model (MA_1) resulted in a higher power of QTL detection and better estimate of location than the best single-environment model (*i.e.*, for the environment where the QTL effect was the highest). However, the situation will be quite different when real data will be analyzed, *i.e.*, no prior information exists on the form of $a(\mu)$. Thus, the decision about the adequacy should be justified using statistical criteria. This can be done on the basis of the dependence of the evaluated significance level on the degree of the applied polynomials. The corresponding results for the situation S_2 are presented in Table 4.

Table 4 illustrates the possibility to deduce the adequate approximation of the QTL × E interaction based on the analysis of the obtained LOD scores. The columns $\beta_{\epsilon_t} = \beta(\alpha)$ show the power of detection of QTL × E interaction for each of the presented models for three levels of significance (5, 1, and 0.1%). It is noteworthy that the critical values of the test statistics (see Equation 3b) were determined by using: (1) the asymptotic χ^2 distribution, and (2) Monte-Carlo simulations with 5000 runs for each of the models (data in brackets). The obtained results showed a remarkable proximity of these two estimates of the power for all of the models. Clearly, such a correspondence may be disturbed when a QTL not accounted for by the model affects the residual genetic variation, causing correlation between environments (see below). As in Table 3, the highest power of

TABLE 1

Characteristics of the simulated multiple environment experiments with both the QTL substitution and residual standard deviation represented as functions of trait mean value

env.	E_i	1	2	3	4	5	6	7	8	9	10
	μ_i	0.0	0.4	0.8	1.2	1.6	2.0	2.4	2.8	3.2	3.6
						Single-QTL situations					
S_1		$\alpha_i = \mu_i(\mu_i - 2)(\mu_i - 4)\sqrt{3}/18$, $\sigma_i = 0.064\mu_i^2 + 1.2$									
	α_i	0.00	0.22	0.30	0.26	0.15	0.00	−0.15	−0.26	−0.30	−0.22
	σ_i	1.20	1.21	1.24	1.29	1.36	1.46	1.57	1.70	1.86	2.03
	h_i^2%	0.00	0.83	1.40	0.99	0.29	0.00	0.22	0.57	0.63	0.29
S_2		$\alpha_i = \mu_i(\mu_i - 2)(\mu_i - 4)\sqrt{3}/18$, $\sigma_i = 0.032\mu_i^2 + 0.8$									
	α_i	0.00	0.22	0.30	0.26	0.15	0.00	−0.15	−0.26	−0.30	−0.22
	σ_i	0.80	0.81	0.82	0.85	0.88	0.93	0.98	1.05	1.13	1.21
	h_i^2%	0.00	1.86	3.14	2.28	0.70	0.00	0.56	1.49	1.69	0.83
S_3		$\alpha_i = \mu_i(\mu_i - 2)(\mu_i - 4)\sqrt{3}/12$, $\sigma_i = 0.064\mu_i^2 + 0.8$									
	α_i	0.00	0.33	0.44	0.39	0.22	0.00	−0.22	−0.39	−0.44	−0.33
	σ_i	0.80	0.81	0.84	0.89	0.96	1.06	1.17	1.30	1.46	1.63
	h_i^2%	0.00	4.04	6.50	4.51	1.30	0.00	0.89	2.17	2.26	1.03
						Two QTL situations					
S_4		$\alpha_{1i} = \alpha_i(S_3)$, $\alpha_{2i} = (0.064\mu_i^2 + 0.8)2\delta$, $\sigma_i = (0.064\mu_i^2 + 0.8)\sqrt{(1-\delta^2)}$, $\delta = 0.25$									
	α_{1i}	0.00	0.33	0.44	0.39	0.22	0.00	−0.22	−0.39	−0.44	−0.33
	α_{2i}	0.40	0.41	0.42	0.45	0.48	0.53	0.58	0.65	0.73	0.82
	σ_i	0.78	0.79	0.81	0.86	0.93	1.02	1.13	1.26	1.41	1.58
	h_{1i}^2%	0.00	4.04	6.50	4.51	1.30	0.00	0.89	2.17	2.26	1.03
	h_{2i}^2%	6.25	6.00	5.84	5.97	6.17	6.25	6.19	6.11	6.11	6.19
S_5		$\alpha_{1i} = \alpha_i(S_3)$, $\alpha_{2i} = (0.064\mu_i^2 + 0.8)2\delta$, $\sigma_i = (0.064\mu_i^2 + 0.8)\sqrt{(1-\delta^2)}$, $\delta = 0.5$									
	α_{1i}	0.00	0.33	0.44	0.39	0.22	0.00	−0.22	−0.39	−0.44	−0.33
	α_{2i}	0.80	0.81	0.84	0.89	0.96	1.06	1.17	1.30	1.46	1.63
	σ_i	0.69	0.70	0.73	0.77	0.84	0.92	1.01	1.13	1.26	1.41
	h_{1i}^2%	0.00	4.04	6.50	4.51	1.30	0.00	0.89	2.17	2.26	1.03
	h_{2i}^2%	25.0	24.0	23.4	23.9	24.7	25.0	24.7	24.5	24.4	24.7
						Multiple QTL situations					
S_6		$\sigma_{pi} = 0.064\mu_i^2 + 0.8$, $\alpha_{1i} = \alpha_i(S_3)$, $\alpha_{2i} = 2\sigma_{pi}\sqrt{0.1}$, $\alpha_{3i} = 2\sigma_{pi}\sqrt{0.05}$, $\alpha_{4i} = 2\sigma_{pi}\sqrt{0.04}$, $\alpha_{5i} = \alpha_{6i} = 2\sigma_{pi}\sqrt{0.02}$, $\alpha_{7i} = 2\sigma_{pi}\sqrt{0.01}$, $\alpha_{8i} = 2\sigma_{pi}\sqrt{0.005}$, $\alpha_{9i} = \alpha_{10i} = 2\sigma_{pi}\sqrt{0.002}$, $\alpha_{11i} = 2\sigma_{pi}\sqrt{0.001}$.									
	α_{1i}	0.00	0.33	0.44	0.39	0.22	0.00	−0.22	−0.39	−0.44	−0.33
	α_{2i}	0.51	0.51	0.53	0.56	0.61	0.67	0.74	0.82	0.92	1.03
	α_{3i}	0.36	0.36	0.38	0.40	0.43	0.47	0.52	0.58	0.65	0.73
	α_{4i}	0.32	0.32	0.34	0.36	0.39	0.42	0.47	0.52	0.58	0.65
	$\alpha_{5i} = \alpha_{6i}$	0.23	0.23	0.24	0.25	0.27	0.30	0.33	0.37	0.41	0.46
	α_{7i}	0.16	0.16	0.17	0.18	0.19	0.21	0.23	0.26	0.29	0.33
	α_{8i}	0.11	0.11	0.12	0.13	0.14	0.15	0.17	0.18	0.21	0.23
	$\alpha_{9i} = \alpha_{10i}$	0.07	0.07	0.08	0.08	0.09	0.09	0.10	0.12	0.13	0.15
	α_{11i}	0.05	0.05	0.05	0.06	0.06	0.07	0.07	0.08	0.09	0.10
	σ_i	0.69	0.68	0.69	0.75	0.83	0.91	1.01	1.11	1.24	1.40
	σ_{pi}	0.80	0.81	0.84	0.89	0.96	1.06	1.17	1.30	1.46	1.63
	h_{1i}^2%	0.00	4.20	6.95	4.73	1.32	0.00	0.90	2.22	2.32	1.04
		$h_{2i}^2 = 10\%$, $h_{3i}^2 = 5\%$, $h_{4i}^2 = 4\%$, $h_{5i}^2 = h_{6i}^2 = 2\%$, $h_{7i}^2 = 1\%$, $h_{8i}^2 = 0.5\%$, $h_{9i}^2 = h_{10i}^2 = 0.2\%$, $h_{11i}^2 = 0.1\%$.									

The proportion of genetic variance attributed to the QTL is denoted by h^2% (h_{1i}^2, h_{2i}^2, *etc.*, when several QTL were simulated); μ_i is the (expected) mean value of the trait in the environment E_i; the residual standard deviations are denoted as σ_i. In the most complex situation S_6, we show also the phenotypic standard deviations σ_{pi}.

TABLE 2

Estimated location L(cM) and power of detection of QTL effect (β_f for testing $H_1|H_0$) and QTL × E interaction (β_e for testing $H_2|H_1$) employing the general (MG) and approximated (MA) models in single-QTL situations

			β_{et} (%)($H_2|H_1$)			β_{ft} (%)($H_1|H_0$)						
		L(cM)	α%→5	1	0.1	5	1	0.1	n_p	df_e	df_f	*AIC*
S_1	MA	66.0 ± 2.03	67	41	16	67	45	22	8	3	5	10.8
	MG	65.4 ± 2.56	57	29	6	64	34	8	21	9	11	
S_2	MA	61.7 ± 0.89	97	87	70	94	87	67	8	3	5	12.4
	MG	62.5 ± 1.45	88	71	47	91	78	49	21	9	11	
S_3	MA	59.8 ± 0.49	100	99	94	100	99	97	8	3	5	12.2
	MG	61.1 ± 0.68	100	95	81	100	98	90	21	9	11	

The results of 200 Monte-Carlo runs are presented for single-QTL situations (see Table 1). L is the estimated QTL location (the simulated value of L is 60 cM); α is significance level; n_p is the number of parameters specifying the model. To reduce n_p, the vector of mean values μ_i across environments was calculated before starting the optimization procedure for the tests 3, 3a, and 3b (for either MA or MG); df_f and df_e are the degrees of freedom for the tests of QTL presence (H_1 *vs.* H_0) and QTL × E interaction (H_2 *vs.* H_1), respectively.

detection of QTL × E interaction and the most precise estimate of QTL location were obtained with model MA_3. It is not surprising that MA_3 is superior over MG. But less expected is the fact that the nonadequate approximations MA_2 and MA_4 were also superior over MG, whereas the poorest approximation MA_1 gave the closest results to MG, but with fewer parameters. Thus, it is not mandatory to have the adequate approximation to take advantage of the proposed method. It will be sufficient to provide a good approximation. Nevertheless, how can we decide about the adequate model, provided the class of the approximation functions is chosen correctly?

To address the last question, the following procedure was employed. For each run, the data were analyzed using all of the models (MA_1–MA_4, and MG), and models that detected QTL × E interaction at the level of significance α were chosen. Then, the model that: (1) exceeded significantly (at some level α*) all of the more simple models: (2) did not differ significantly (at α*) from more complex models was selected as adequate. The general model also participated in this competition as the most complex one, because of the number of parameters needed. The resulting distribution of the choices of the adequate model is presented in the last three columns of Table 4. It allowed us to conclude that: (1) model MA_3 is an adequate model because it was chosen in more than half of the runs where the QTL × E interaction was detected, and with a frequency that is threefold higher than the next best choice; (2) the models of the polynomial class were chosen 25–30 times more than the exact general model MG. Moreover, even the simplest approximation, MA_1, would be selected 4–6 times more frequently than MG.

Two QTLs: When several QTLs segregate simultaneously in the mapping population, their effects will generate correlations between trait measurements across environments, which should be taken into account. One of the possible ways to account for this correlation is through simultaneous analysis of multiple traits, taking the trait values in different environments as different quantitative traits (Korol *et al.* 1987, 1994, 1995; Jiang and Zeng 1995; Ronin *et al.* 1995). However, the multiple trait analysis limits the number of environments, because it is associated with an increased number of parameters. The approach proposed in this paper does not have this drawback, but introduces other sources of distortions: (1) correlations caused by unaccounted QTLs, and (2) approximated description of QTL dependence on environment based on the bioindication assumption.

Consider the first problem. We should now evaluate to what extent correlations between environments caused by unaccounted QTLs may affect the efficiency of the proposed approach. The second problem will be treated in the next section and in the discussion.

For the simulated cases of two QTLs segregating in the mapping population (S_4 and S_5), we first analyzed the consequences when the proposed approach of accounting QTL × E interaction is applied, ignoring the correlations caused by the effect of Q_2/q_2 (model 1, Figure 3). Then, we re-evaluated the results by applying the proper model (model 2, Figure 3). In these simulations, we considered two situations of relative effects of the "target" QTL (Q_1/q_1) and of the cosegregating QTL (Q_2/q_2): Q_1/q_1 and Q_2/q_2 have comparable effects on the target trait though $a_1 < a_2$ (S_4), and Q_2/q_2 is much stronger than Q_1/q_1 (S_5). The residual variance σ_i^2 of the trait within the QTL groups Q_1/Q_1 and q_1/q_1 was the same in S_3, S_4, and S_5.

First, compare the accuracy of the QTL mapping obtained employing model 1 for situation S_4, with those of S_3 where only the effect of Q_1/q_1 was simulated (the situations S_3 of Table 2 and S_4, model 1, Figure 3). In

both cases, the results clearly demonstrate the superiority of the approximated model MA. Hence, provided that the effect of a cosegregating QTL, Q_2/q_2, does not considerably exceed the effect of the target QTL, Q_1/q_1, the proposed approach provides accurate results even if the effect of Q_2/q_2 is ignored. However, this may not be the case with larger effects of Q_2/q_2, as demonstrated by the results for S_5 (model 1, Figure 3). In general, a correct model should account for the genetic components of the residual variation in the alternative genotypic groups of the target QTL, causing correlation between trait values across environments (*e.g.*, Jiang and Zeng 1995). This is also true for the method proposed here of mapping analysis with data measured in multiple environments.

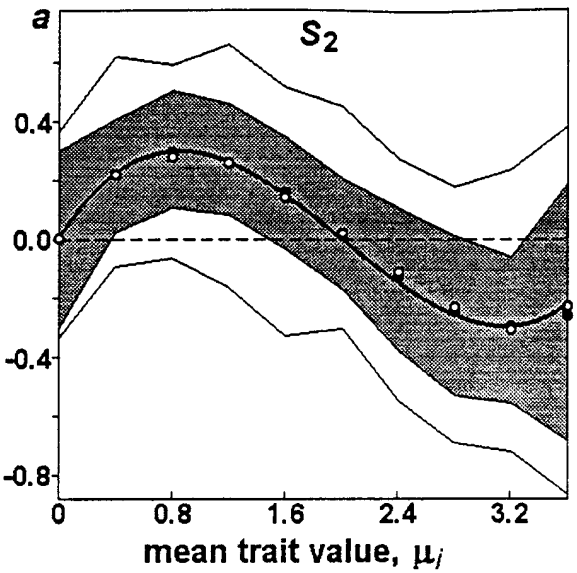

Figure 2.—Accuracy and precision of estimates of QTL effects across environments obtained by the general and approximated single-QTL models applied to single-QTL data. The graph represents estimated average trend and a sampled 95% vicinity of the p-dimensional point $\{a_i/\sigma_i, i = 1,... p\}$ from 200 Monte-Carlo runs for situation S_2 (described in Table 1). Open circles and light region are for the general model MG, black circles and shaded region are for the approximated model MA. The regions were obtained as follows. Let a_i and σ_i be the QTL effect and the residual standard variation at the ith environment (i...., p), and a_{ij} and σ_{ij} be the corresponding estimates at jth run ($j = 1,..., 200$). As a measure of discrepancy of the estimated parameters from the expected, across environments, we used the index

$$\delta_j = \sqrt{\Sigma}(a_i/\sigma_i - a_{ij}/\sigma_{ij})^2.$$

The sampled 95% volume is defined as a p-dimensional sphere, $S_{0.95}$, of minimal radius with center at point $\{a_i/\sigma_i, i = 1,..., p\}$, which includes 95% of estimated results $r_j = (a_{ij}/\sigma_{ij}, i = 1,..., p)$ from the simulations. The sphere provides a 95% region of estimated curves $a_j(\mu(E))$ in the plane of the estimated effects $\{a_{ij}, \mu(E_i)\}$:

$$D_{0.95} = \{a_j(\mu(E)),\ a_{ij} \in [\min_{r_j \in S_{0.95}} a_{ij},\ \max_{r_j \in S_{0.95}} a_{ij}],\ i = 1,... p\}.$$

According to our calculations, the radius of $S_{0.95}$ is smaller for MA than MG (0.672 *vs.* 0.805).

Two possibilities exist for considering the effects of cosegregating QTLs in the mixture mapping model. The first is to represent all QTL groups (four, in our case, of two QTLs cosegregating in a doubled haploid or backcross population) in the likelihood function. Although this procedure is not feasible for mapping multiple QTLs across the genome, it may be very useful in cases of linked QTLs. The second is to include into the mixture model the effects of the cosegregating QTLs as cofactors derived from regression analysis on marker loci (Zeng 1994; Jansen and Stam 1994). The proposed approximated method is equally applicable in both of these approaches. Here, we demonstrate it using the first approach. Although this mixture formulation is more challenging technically, it allows for a proper analysis of potential variance effect of the cosegregating QTL (although we do not deal with this problem here). It is not obvious how to model the effect of a second QTL with regression cofactors. As was shown earlier, variance effect of a QTL may result in increased accuracy of the mapping model if it is included into the model, and may seriously reduce the accuracy with an inadequate model (Korol *et al.* 1996).

With two QTLs, four densitites $f_{q_1q_1q_2q_2}(x)$, $f_{q_1q_1q_2q_2}(x)$, $f_{Q_1Q_1q_2q_2}(x)$, and $f_{Q_1Q_1Q_2Q_2}(x)$ should be considered. Consequently, in calculations of the maximum likelihood function, instead of four marker groups for a current interval, it is necessary to characterize 16 marker groups for any pair of nonadjacent intervals. The application results of the full MG model and the approximated polynomial MA model are presented in Figure 3 (model 2). It is noteworthy, that the proposed approximation of the environmental dependence of QTL effect as a function of the mean value of the trait in a given environment was applied not only to the target QTL Q_1/q_1, but also to the cosegregating Q_2/q_2. This approach may be especially attractive when there are many cosegregating QTLs with environmental dependent effects (*e.g.*, as regression cofactors on respective marker loci). This will result in far fewer parameters. As expected, the full model 2 increased the power of detection of Q_1/q_1 effect on the trait x and $(Q_1/q_1) \times$ E interaction (not shown), as well as increased accuracy of estimates of the chromosome position of Q_1/q_1 and of its effects by the environments (model 2, Figure 3). Again, MA had superior attributes than MG. This conclusion is also supported by the values of AIC.

Several QTL: As we could see before, a strong QTL, if not accounted by the model, may cause correlations between environments resulting in reduced accuracy of estimated parameters. Nevertheless, the distortion caused by a QTL comparable with the target one (*e.g.*, exceeding the target effect no more than two times) is not dramatic (see Figure 3). Including the effects of cosegregating QTLs into the model solves this problem. This can be done by combining the proposed approach with regression cofactors. However, an appreciable pro-

TABLE 3

Comparison of the general model with the polynomial approximations for the power of detection of the QTL (β_{f_t}), QTL × E interaction (β_{e_t}), and the accuracy of QTL location (L)

		β_{f_t} (%)($H_1|H_0$)				β_{e_t} (%)($H_2|H_1$)			
Model	L(cM)	α%→5	1	0.1	df_f	5	1	0.1	df_e
MA_1	60.6 ± 0.74	98	90	74	3	96	85	63	1
MA_2	61.0 ± 0.59	99	98	92	4	98	95	84	2
MA_3	59.8 ± 0.49	100	99	97	5	100	99	94	3
MG	61.1 ± 0.68	99	98	90	11	99	97	83	9
ME_3	60.9 ± 1.44	83	63	40	2				

MA_1, MA_2, and MA_3 are the approximated models based on polynomials of first, second, and third degree, respectively, for the target QTL effect. Situation S_3 (see Table 1) is considered. The results are compared to the best single-environment model (ME_3, for data from the environment 3 where the effect of the target QTL is the largest one).

portion of genetic variation for the analyzed trait may still remain in the residuals, because of combined effect of many small QTLs. This residual genetic variation may be several-fold larger than the effect of the target QTL. Would the resulting correlation between environments preclude the application of the method?

To address this question, let us consider the case S_6 (for detailed specification see Table 1). Here, the genetic variation of the trait depends on the target QTL (Q_1/q_1) (with an average $h^2 \sim 2.5\%$ across environments) and 10 additional unlinked QTLs (Q_2/q_2–Q_{11}/q_{11}). The average (across environments) effect of Q_2/q_2 was $h^2 \sim 10\%$, whereas the combined average effect of Q_3/q_3–Q_{11}/q_{11} was 15%. Thus, the total effect of Q_2/q_2–Q_{11}/q_{11} is 10-fold compared to that of Q_1/q_1, whereas the effect of Q_3/q_3–Q_{11}/q_{11} is sixfold compared to that of Q_1/q_1. One may expect that the power of detection of Q_1/q_1 × E interaction will be very low if the segregation of Q_2/q_2–Q_{11}/q_{11} is not accounted for by the model, hence causing correlation between the environments. This is indeed the case as can be seen from Table 5 (first row for N = 1,3,5). It is noteworthy, that in this case employment of the asymptotic distribution for the critical values of the test statistics gives seriously biased upward estimates $\beta_{e_t} = \beta_{e_t}(\alpha)$ of the power of detection of Q_1/q_1 × E interaction (compared to the estimates $\beta_{e_s} = \beta_{e_s}(\alpha)$ obtained using Monte-Carlo simulations with 5000 runs).

As we see from the foregoing results for the two QTL situations (S_4 and S_5 in Figure 3), an unaccounted QTL will not seriously affect the results for the target QTL if its effect does not exceed the target one by too much (*e.g.*, not more than twofold). In the current situation S_6, each of the simulated effects of Q_3/q_3–Q_{11}/q_{11} fit this condition, whereas this is not true for their combined effect or for the individual effect of Q_2/q_2. It is interesting to explore whether including Q_2/q_2 into the model as a cofactor will improve the situation. This is indeed an important question, because in practice sufficiently strong QTLs can be compensated in such a way (Jansen and Stam 1994; Zeng 1994), but this does not guarantee

TABLE 4

Comparison of the general model with the polynomial approximations for the detection power of QTL × E interaction (β_e), and accuracy of QTL location (L)

		β_e(%)				f_M		
Model	L(cM)	α%→5	1	0.1	df	α%→5	1	0.1
MA_1	62.0 ± 1.46	88 (88)	68 (69)	48 (46)	1	0.18	0.16	0.13
MA_2	61.1 ± 1.31	92 (92)	78 (77)	56 (59)	2	0.18	0.17	0.15
MA_3	61.7 ± 0.89	96 (97)	86 (87)	66 (70)	3	0.54	0.49	0.45
MA_4	61.6 ± 0.92	95 (95)	80 (82)	61 (61)	4	0.05	0.05	0.05
MG	62.5 ± 1.45	88	71	47	9	0.03	0.03	0.03
Total frequency of cases where QTL × E interaction was detected						0.98	0.90	0.78

MA_i (i = 1,...,4) denotes the approximated model based on polynomial of ith degree for the target QTL effect. Situation S_2 (see Table 1) is considered. The power of the test was obtained using Monte-Carlo simulations (see text); corresponding results based on χ^2 asymptotic distribution of the test statistic are given in brackets. The distribution of frequencies of the chosen approximations is presented in the last three columns (f_M) resulted from competition between MA_1–MA_4 and MG, see text.

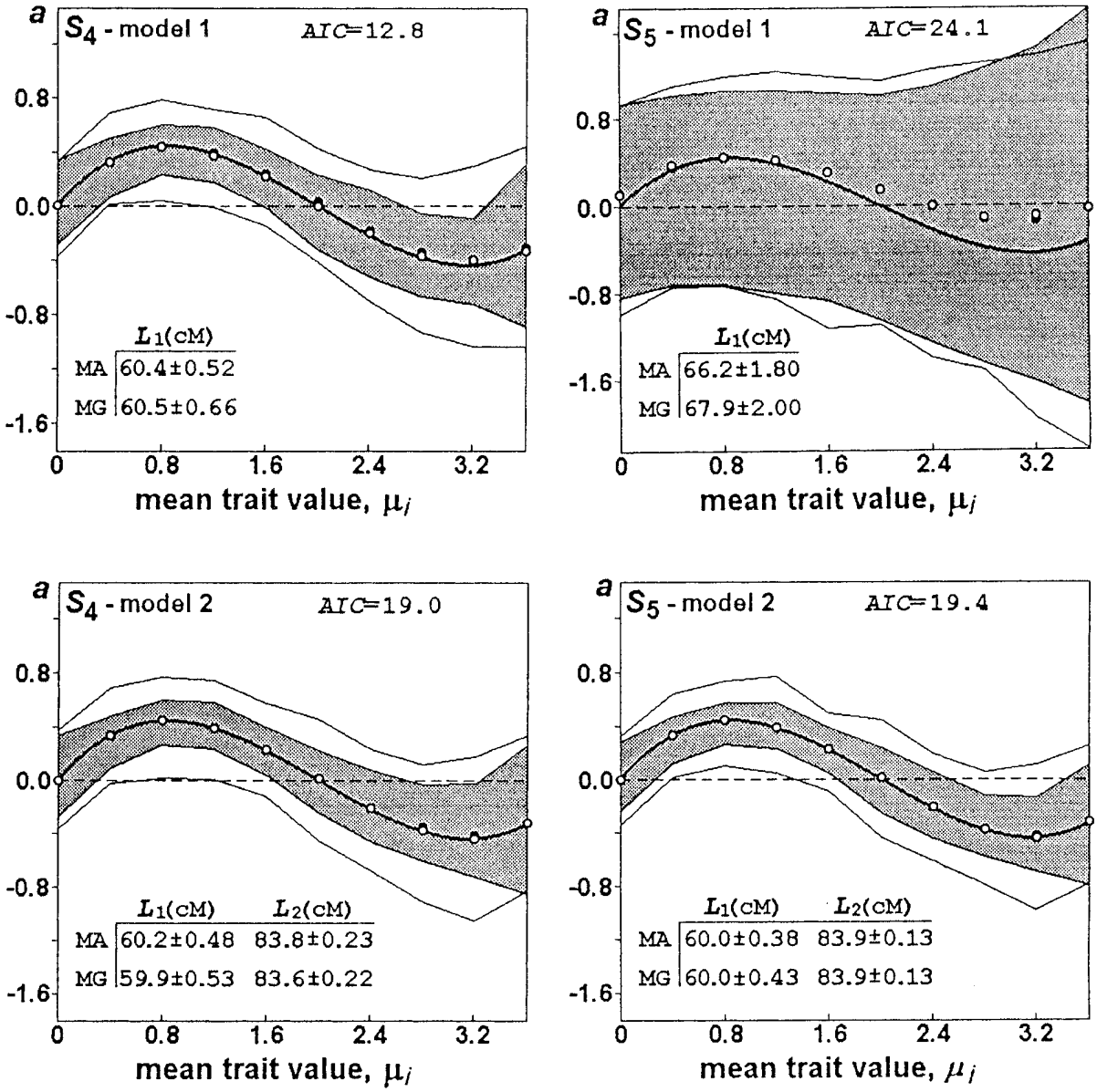

Figure 3.—Accuracy of estimates of QTL effects across environments obtained by the general and approximated models applied to two-QTL data. Here model 1 and model 2 denote a single- and two-QTL mapping models. In situation S_4, the effects of simulated unlinked QTLs were: average $h^2\% \approx 2.5$ (range 0–6.5) for QTL_1 and 6.0 (range 5.84–6.25) for QTL_2, respectively; in situation S_5, the simulated effects were: average $h^2\% \approx 2.5$ (range 0–6.5) for QTL_1 and 24.0 (range 23.4–25.0) for QTL_2, respectively. AIC is the Akaike's information criterion which takes into account the cost of increased number of parameters in the model (Bozdogan 1987).

that the residual variation caused by many small polygenes will not exceed the target effect several times over, thus preventing the application of the proposed method.

The data presented in the second row of Table 5 show that including Q_2/q_2 as a cofactor into the model substantially improved the situation by increasing the detection power of $Q_1/q_1 \times$ E interaction from two- to fivefold (for α, ranging from 0.05 to 0.001) and the precision of Q_1/q_1 estimated location more than twofold. Note that in this case the χ^2 distribution appeared to be a very good approximation for the distribution of the test statistic (compare corresponding values of β_{ε_t} and β_{ε_s}) in spite of the noise caused by Q_3/q_3–Q_{11}/q_{11}. It would be quite desirable to get some idea of the distorting effect of the correlations caused by joint action of the unaccounted QTLs Q_3/q_3–Q_{11}/q_{11}. Therefore, for comparison we provide in the third row the results for the case where all the residual genetic variation caused by Q_3/q_3–Q_{11}/q_{11} is replaced by nongenetic variation. We can conclude that distortion of the basic model assumption of "no correlation between environments" caused by the presence of Q_3/q_3–Q_{11}/q_{11}, which collectively exceed by a factor of six the effect of the target QTL Q_1/q_1, is incomparably smaller than that caused by

TABLE 5

The effect of cofactors on the power of detection of QTL × E interaction (β_e), and accuracy of QTL location (*L*)

N	Model	h^2%	D_{gei}	*L*(cM)	α%→5	β_e (%) 1	0.1
1	1	2.4	0.101	62.9 ± 1.66	45 (70)	27 (58)	15 (42)
	2			59.9 ± 0.74	96 (96)	91 (90)	76 (76)
	3			59.2 ± 0.49	99 (99)	99 (99)	96 (97)
3	1	5	0.015	60.2 ± 0.72	19 (76)	8 (57)	1 (32)
	2			59.3 ± 0.37	42 (56)	21 (37)	6 (13)
	3			59.6 ± 0.22	41 (40)	23 (23)	4 (5)
5	1	2	0.006	62.6 ± 1.52	14 (62)	4 (41)	1 (23)
	2			60.5 ± 0.85	20 (32)	7 (16)	2 (4)
	3			59.9 ± 0.71	18 (19)	8 (8)	1 (1)

The power of the test was obtained using Monte-Carlo simulations (see text); (corresponding results based on χ^2 asymptotic distribution of the test statistic are given in brackets). Data of the situation S_6 (see Table 1) were used with the target QTL on chromosome *N* ($N = 1$, 3 or 5). Three models of the analysis of the residual variation were employed: (1) the cofactors are totally ignored; (2) the effect of the strongest QTL is fitted using two-QTL mixture model; (3) the genetic component of the residual variation is replaced by the equivalent nongenetic variation. D_{gei} is the variance of QTL × E interaction for the target QTL; h^2% is the averaged heritability over environments attributed to the target QTL.

a single QTL, Q_2/q_2, which exceeds the target QTL only by a factor of four. The same analysis was conducted when instead of Q_1/q_1 another QTL was considered as a target one (Q_3/q_3 or Q_5/q_5). The results are presented in the remainder of Table 5 and manifest the same pattern.

Missing data: One can hardly expect that all genotypes will be perfectly represented in all of the environments where the experiment was conducted. Some data will be missed, hence it is of interest to get some idea how it could affect the power of QTL × E detection. Our approximate model allows us to treat this problem easily. It appeared that with a large number of environments, even if a large proportion of genotypes is not represented in each environment, the resulting power of the test of QTL × E interaction and location accuracy of the target QTL are quite high. Monte-Carlo simulations presented in Table 6 illustrate this point. It is noteworthy, that if only 20–50% of the data are available in each of the 50–100 environments, the approximated model is still very satisfactory even when a suboptimal approximation was used (compare the results for MA_1, MA_2, and MA_3 for the two examples with the situation S_4). Clearly, an attempt to apply the general model would mean an unrealistic task of estimation of about 100–200 parameters, in contrast to our model which needs only eight parameters.

Example of application: The trait "alpha amylase activity" from a barley QTL × E study presented in Figure 1 (see Hayes *et al.* 1993, 1996) was used to demonstrate the utility of the proposed procedure. From previous analyses, the largest QTL effect for this trait was associated with segregation on chromosome *1* (Hayes *et al.* 1996). Thus, according to the simulation results of the previous section, even if one ignores the effects of other genomic segments when dealing with markers of chromosome *1*, we did not expect serious reduction in the efficiency of the mapping analysis. As shown in Figure 1, the estimates of a_i for this trait obtained for separate environments can be approximated as a quadratic parabola of the mean value of the trait over the environments. This approximation was used to construct a combined model for testing QTL × E interaction effect and to estimate the QTL location on chromosome *2* (Table 7).

The first step was to decide whether variation in σ_i^2 is significant and should be included into the model. Our prior trial showed that polynomial regression of σ_i^2 on mean trait value is nonsignificant (data not shown). Thus, the hypotheses "constant $\sigma_i^2 = \sigma^2$ across environments" and "variable σ_i^2 across environments" were contrasted employing two models to describe the dependence of a_i on environment: the general model MG and quadratic approximation MA. Both approaches reject the hypothesis of constant σ_i^2 at a highly significant level (with LOD values 21.64 and 20.25 for MG and MA, respectively). Therefore, we should test for the presence of QTL × E interaction given environmental-specific σ_i^2. Here, we can see the advantage of the proposed approach. Indeed, with the polynomial approximation of a_i for H_2 $\{a_i \neq \text{const}\}$, the hypothesis H_1 $\{a_i = a = \text{const}\}$ is rejected at the significance level of $P = 0.010$, while based on the general model MG we can get only $P = 0.095$.

An important question is whether the two models, MA or MG, differ significantly provided H_2 $\{a_i \neq \text{const}\}$ is true. Such a comparison was conducted for both con-

TABLE 6

The effect of missing data on the power of detection of QTL × E interaction and accuracy of QTL location when the number of environments is large

S_i	Model	N_{lin}	N_{env}	$\overline{N}$	L(cM)	β_{e_t} (%) α%→5	1	0.1
S_1	MA_3	200	10	200	66.0 ± 2.03	67	41	16
		200	100	50	59.7 ± 0.94	94	82	60
		200	100	100	60.0 ± 0.45	100	99	98
S_3	MA_3	100	50	40	60.0 ± 0.72	100	98	93
		200	50	40	60.2 ± 0.56	98	96	88
S_4	MA_1	200	50	40	61.4 ± 1.10	81	69	46
	MA_2	200	50	40	61.0 ± 1.10	89	77	56
	MA_3	200	50	40	60.0 ± 0.72	98	93	82
S_4	MA_1	200	50	100	60.1 ± 0.41	98	97	93
	MA_2	200	50	100	60.2 ± 0.39	99	98	97
	MA_3	200	50	100	59.9 ± 0.27	100	100	100

N_{lin} is the total number of genotypes (lines) in the mapping population, N_{env} is the number of environments, and $\overline{N}$ is the mean number of genotypes scored per environments. To generate the data, the initial form of the dependence of the QTL on environment was used as presented in Table 1, but with a correspondingly smaller step in changes of the independent variable.

sidered situations, *i.e.*, with $\sigma_i^2 = \text{const}$ and $\sigma_i^2 \neq \text{const}$, using LOD scores (see the last section of Table 7). In no case was the difference significant, so that MA provides the same solution as MG, but the approximated model is preferable because of a lower number of needed parameters. In other words, MA extracts from the data the same information on variation of the QTL effect a_i across environments as MG, but does it more efficiently.

DISCUSSION

The conducted simulations and the example of barley multiple environments experiment demonstrate the utility of the proposed approximate approach for analyzing QTL × E interaction. Its main benefit is the ability to use data collected from a large number of environments without the necessity of increasing the number of parameters. Earlier, an elegant solution to this problem was proposed by Jansen *et al.* 1995. Their QTL mapping model includes in an obvious way the terms describing the effects of the target QTL and regression cofactors of cosegregation QTL, the effects of multiple environments, and the terms of QTL × E interactions. However, such an analysis is limited by situations where the environments can be obviously characterized by some physical attributes. When such characteristics are not available, the application of the general QTL × E mapping model (our foregoing MG model) is accompanied by a tremendous number of parameters involved in the model. The method proposed in this paper overcomes, though in an approximate form, both these obstacles, allowing us to analyze QTL × E interactions across a large (in fact, unlimited) number of "anonymous" environments. Expressing the dependence of a QTL effect on environmental conditions as a function of environmental mean value of the trait can also be applied to multiple QTLs from independent genomic regions. Therefore, the proposed approach could be very helpful in coping, albeit in an approximate form, with a difficult problem of QTL mapping analysis, *i.e.*, rapid increase in the number of parameters with increasing number of effective QTLs and environments. This improves our ability to efficiently extract more mapping information when more environments are used to evaluate the quantitative trait.

In addition to the large number of parameters to be estimated, the general model MG of QTL × E interaction fails to account for correlation between environments caused by cosegregating QTLs not included into the model. While the first problem is not critical for our method, the second one may be more serious. The foregoing simulations showed that unaccounted QTLs with a strong individual effect may indeed reduce the power of detection of QTL × E interaction and the accuracy of parameter estimation by the proposed approximated method. Therefore, including such QTLs as cofactors into the model is mandatory for applications. However, such a compensation cannot be perfect and a significant genetic component may remain in the residual variation. An important question is whether this residual genetic variation, which can be several-fold larger than the effect of the target QTL, will produce correlation between the environments precluding the application of the method. Our simulations allowed us to conclude that distortion of the basic model assump-

TABLE 7

Detection of QTL × E interaction in a barley dihaploid population scored over nine environments (for a QTL of chromosome 1 affecting 'alpha amylase activity'; see Hayes *et al.* 1996)

	k	$\sigma_j \neq$ const 1	$\sigma_j =$ const 2
MG	$LOD(H_{2k}/H_{1k})$ (df)	2.954 (8)	3.403 (8)
	P	0.095	0.015
MA	$LOD(H_{2k}/H_{1k})$ (df)	1.983 (2)	1.817 (2)
	P	0.010	0.015
MG	$LOD(H_{21}/H_{22})$ (df)	21.64 (8)	
	P	0.000	
MA	$LOD(H_{21}/H_{22})$ (df)	20.25 (8)	
	P	0.000	
	$LOD(H^G_{2k}/H^A_{2k})$ (df)	0.971 (6)	1.586 (6)
	P	0.60	0.29

We used the index k to denote two types of models corresponding to equal ($k = 1$) *vs.* nonequal ($k = 2$) residual variances across environments. Therefore, H_{11} and H_{12} hypotheses here assume the presence of a QTL effect with constant and varying residual variances, respectively. Correspondingly, H_{21} and H_{22} assume the presence of a QTL with varying effect and constant and varying residual variances, respectively. To test whether the two models, MA or MG, differ significantly, provided H_2 $\{a_j \neq \text{const}\}$ is true, both situations, *i.e.*, with $\sigma_i^2 = \text{const}$ and $\sigma_i^2 \neq \text{const}$, were considered using LOD score, $LOD(H^G_{2k}/H^A_{2k})$ (see the last section of the table).

tion of "no correlation between environments" caused by the segregation of several small QTLs, which collectively exceed by a factor of six the effect of the target QTL, is much smaller than that caused by a strong single QTL, which exceeds the target QTL by only a factor of four (see Table 5). Thus, undetectable small QTLs will not attenuate seriously the resolution power of the proposed method, even if their combined effect is severalfold higher than that of the target QTL.

An important question is how to reveal the adequate approximation of the QTL × E interaction. With simulated data, it is easy to compare the adequate and the nonadequate approximations simply because we know the degrees of the polynomials employed in the simulations. However, the situation will be quite different when real data will be analyzed. Thus, the decision about the adequacy of the approximation should be justified statistically, *i.e.*, we should decide about the adequate model, provided the class of the approximation functions is chosen correctly. This allows us to conclude that: (1) the adequate model MA_3 was the best, *i.e.*, it was chosen in more than half of the runs where the QTL × E interaction was detected and with a frequency that was threefold higher than the next best choice.

The last and most difficult problem is how to recognize the situations when the applied approximation is not valid. If the opposite is true, *i.e.*, if the dependence of the QTL effect on environmental conditions can indeed be presented in the form of regression on mean values or any other bioindicators, then the proposed approximated method proved to give a higher detecting power of QTL × E interaction compared to the precise general model (MG). Thus, one can start the procedure using the approximated method, though the general model can also be applied in parallel if the number of environments is not too large, so that the number of parameters for MG is not unrealistically large. However, if the approximated analysis revealed no significant QTL × E interaction, does it really mean an independence of the QTL effect from environmental conditions? Or, alternatively, the interaction may exist, but it cannot be represented as a regression of the target QTL effect on the mean values of the trait or some other bioindicators?

Consider one of the possible ways to cope with this problem. If the general model is applicable, *i.e.*, the number of parameters is not too large, it may be used as a tool to answer the foregoing question. Rejection of the H_0 hypothesis "no QTL × E interaction" by MG will mean that our basic assumption (regression on the bioindicator) does not fit the data. If the number of environments is too large, the general model can be applied for randomly chosen groups of environments. Then, the significance of the interaction may be evaluated from the obtained distribution of the tests using the Bonferroni correction. For example, with $N = 100$ environments, one can produce $k = 20$ samples, each including data of $m = 10$ randomly chosen environments. Let α be the accepted level of significance for the QTL × E interaction test for the whole set of the samples. Then, assuming independence of these samples, one can reject the H_0 hypothesis if at least one of the samples achieved the significance level of α/k. Clearly, due to the postulated independence, which is not the case for $mk > N$, this is a conservative test of QTL × E interaction. Nevertheless, it seems preferable to us than the standard way of multiple-environment data analysis when the data from each environment are treated separately, and the final conclusion is derived from the analysis of the estimated QTL effects across environments (Paterson *et al.* 1991; Stuber *et al.* 1992; Utz and Melchinger 1996).

The foregoing test based on the general model may result in the same conclusion as the approximated model, *i.e.*, "no QTL × E interaction." By contrast, if the general model allowed us to detect QTL × E interaction, but the approximated model did not, it will indicate that the proposed bioindicator(s) is not informative and other explanatory factors could be found. Further studies are needed to develop more optimal algorithms of application of the proposed approach when applied to a large number of environments (and when direct utilization of the general model is impossible). However, even in the current

form, the drawbacks of the proposed method are compensated by the possibility of working with an unlimited number of environments with missing data, and at a remarkable reduction in the number of parameters needed, as compared to the usual way of testing for QTL × E interactions based on ANOVA treatment of QTL estimates obtained on the basis of single-environment analysis.

We are grateful to the North American Barley Genome Mapping Project (Dr. P. M. Hayes) for the data set employed in our illustrative example. We are very thankful to the anonymous referees for their relevant comments and suggestions that assisted in improvement of the manuscript. This work was supported by the Israeli Ministry of Absorption and Ministry of Science and the Ancell-Teicher Research Foundation for Genetics and Molecular Evolution.

LITERATURE CITED

Asins, M. J., P. Mestre, J. E. Garcia, F. Dicenta and E. A. Carbonell, 1994 Genotype × environment interaction in QTL analysis of an intervarietal almond cross by means of genetic markers. Theor. Appl. Genet. **89:** 358–364.

Baker, R. J., 1996 Recent research on genotype-environmental interaction. Proceedings of the 5th International Oat Conference and 7th International Barley Genetics Symposium, Saskatoon, Saskatchewan, Canada, pp. 235–246.

Beavis, W. D., and P. Keim, 1996 Identification of QTL that are affected by environment, in *Genotype by Environment Interactions*, edited by M. S. Kang and H. Gauch, Jr. CRC Press, Inc.

Bozdogan, H., 1987 Model selection and Adaike's information criterion (AIC): the general theory and its analytical extention. Psychometrika **52:** 345–370.

Eberhard, S. A., and W. A. Russel, 1966 Stability parameters for comparing varieties. Crop Sci. **6:** 36–40.

Falconer, D. S., 1981 *Introduction to Quantitative Genetics*. Longman Scientific & Technical, Essex, UK.

Gilbert, N., 1961 Polygene analysis. Genet. Res. **2:** 96–102.

Hayes, P. M., 1994 Genetic stocks available through the North American Barley Genome Mapping Project. Barley Genet. Newsl. **24:** 113–116.

Hayes, P. M., B. H. Liu, S. J. Knapp, F. Chen, B. Jones *et al.*, 1993 Quantitative trait locus effects and environmental interaction in a sample of North American barley germplasm. Theor. Appl. Genet. **87:** 392–401.

Hayes, P. M., G. Briceno and D. Matthews, 1996 The Steptoe × Morex barley mapping population. http://probe.nalusda.gov:8300/edi-bin/browse/graingenes

Himmelblau, D. M., 1972 *Applied Nonlinear Programming*. McGraw-Hill, New York.

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

Jansen, R. C., J. M. Van Ooijen, P. Stam, C. Lister and C. Dean, 1995 Genotype-by-environment interaction in genetic mapping of multiple quantitative trait loci. Theor. Appl. Genet. **91:** 33–37.

Jiang, C., and Z.-B. Zeng, 1995 Multiple trait analysis and 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., A. A. Zhuchenko and A. P. Samovol, 1981 Linkage between loci of quantitative characters and marker loci. III. The bias of estimates upon disturbance of basic assumptions. Genetica (in Russian) **17:** 1234–1247.

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

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.*, 1996 Estimating variance effect of QTL: an important prospect to increase the resolution power of interval mapping. Genet. Res. **67:** 187–194.

Lander, E. S., and D. Botstein, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121:** 185–99.

Paterson, A. H., S. Damon, J. D. Hewitt, D. Zamir, H. D. Rabinowitch *et al.*, 1991 Mendelian factors underlying quantitative traits in tomato: comparison across species, generations and environments. Genetics **127:** 181–197.

Rendel, J. M., 1967 *Canalization and Gene Control*. Academic Press, London.

Romagosa, I., S. E. Ullrich, F. Han and P. M. Hayes, 1996 Use of additive main effects and multiplicative interaction model in QTL mapping for adaptation in barley. Theor. Appl. Genet. **93:** 30–37.

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

Sari-Gorla, M., T. Calinski, Z. Kaczmarek and P. Krajewski, 1997 Detection of QTL-environment interaction in maize by a least squares interval mapping method. Heredity **78:** 146–157.

Stuber, C. W., S. E. Linkoln, S. E. Wolff, T. Helentjaris and E. S. Lander, 1992 Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics **132:** 823–839.

Tinker, N. A., and D. E. Mather, 1995 Methods for QTL analysis with progeny replicated in multiple environments. J. Quant. Trait Loci **1 (2)**. http://probe.nalusda.gov:8000/otherdocs/jqtl/jqtl1995-02/jqtl16r2.html

Tiret, L., L. Abel and R. Rakotovao, 1993 Effect of ignoring genotype-environment interaction on segregation analysis of quantitative traits. Genetic Epidemiology **10:** 581–586.

Utz, H. F., and A. E. Melchinger, 1996 PLABQTL: a program for composite interval mapping of QTL. J. Quant. Trait Loci **2 (1)**. http://probe.nalusda.gov:8000/otherdocs/jqtl/jqtl1996-01/utz.html

Via, S., and R. Lande, 1987 Evolution of genetic variability in a spatially heterogeneous environment: effects of genotype-environment interaction. Genet. Res. **49:** 147–156.

Weller, J. I., 1986 Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. Biometrics **42:** 627–640.

Wilks, S. S., 1962 *Mathematical Statistics*. John Wiley & Sons, New York.

Wu, R., and R. F. Stettler, 1997 Quantitative genetics of growth and development in *Populus*. II. The partitioning of genotype × environment interaction in stem growth. Heredity **78:** 124–134.

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

Communicating editor: B. S. Weir