Theoretical Basis of the Beavis Effect

Shizhong Xu[1]

Department of Botany and Plant Sciences, University of California, Riverside, California 92521

ABSTRACT

The core of statistical inference is based on both hypothesis testing and estimation. The use of inferential statistics for QTL identification thus includes estimation of genetic effects and statistical tests. Typically, QTL are reported only when the test statistics reach a predetermined critical value. Therefore, the estimated effects of detected QTL are actually sampled from a truncated distribution. As a result, the expectations of detected QTL effects are biased upward. In a simulation study, William D. Beavis showed that the average estimates of phenotypic variances associated with correctly identified QTL were greatly overestimated if only 100 progeny were evaluated, slightly overestimated if 500 progeny were evaluated, and fairly close to the actual magnitude when 1000 progeny were evaluated. This phenomenon has subsequently been called the Beavis effect. Understanding the theoretical basis of the Beavis effect will help interpret QTL mapping results and improve success of marker-assisted selection. This study provides a statistical explanation for the Beavis effect. The theoretical prediction agrees well with the observations reported in Beavis's original simulation study. Application of the theory to meta-analysis of QTL mapping is discussed.

THE primary goal of genetic mapping experiments is to identify the locations of genes that affect variable expression of a trait among individuals. Most researchers also use the data from a sampled population to estimate the genetic effects of quantitative trait loci (QTL). Information about the magnitude of the genetic effects of significant QTL is useful in prioritizing subsequent uses of the loci as candidate genes for consideration in genetic engineering and marker-assisted selection (LANDE and THOMPSON 1990). Many statistical methods produce unbiased or asymptotically unbiased estimates of parameters. However, this property does not apply to QTL effects estimated from genome scans. This phenomenon was first discussed by LANDE and THOMPSON (1990).

Almost all QTL mapping procedures can detect QTL with large effects, but not all can detect QTL with intermediate and small effects. Quantitative traits are defined as traits controlled primarily by intermediate and small QTL. BEAVIS (1994, 1998) designed a large-scale simulation experiment to evaluate the efficiency of interval mapping for detecting and estimating polygenes. Hereafter, this simulation experiment is called the Beavis experiment. BEAVIS (1998) simulated either 10 or 40 independent QTL, examining situations where each QTL explained a proportion of the phenotypic variance ranging from 0.75 to 9.5% with sample sizes ranging from 100 to 1000. The Beavis experiment showed that the average estimates of phenotypic variances associated with correctly identified QTL were greatly overestimated if

only 100 progeny were evaluated, slightly overestimated if 500 progeny were evaluated, and fairly close to the actual magnitude when 1000 progeny were evaluated. When the sample size was small, say 100, the statistical power of detecting a small QTL was as low as 3% and the estimated effects were typically inflated 10-fold. This phenomenon has since been referred to as the Beavis effect and has formed the basis of a number of subsequent analyses (*e.g.*, OTTO and JONES 2000; HAYES and GODDARD 2001).

Estimating the number of quantitative trait loci is another goal of QTL mapping experiments (SILLANPAA and ARJAS 1998). OTTO and JONES (2000) acknowledged that the estimated number of QTL using molecular markers (interval mapping) can be more informative than that obtained from analyses using only phenotype (LANDE 1981; ZENG 1992). However, a QTL is reported from the analysis only if it is detected. Therefore, the distribution of the detected QTL is actually inferred from censored data due to missing small-effect QTL. OTTO and JONES (2000) incorporated the Beavis effect into their maximum-likelihood analysis to recover the potential number of QTL using the number of detected QTL, so called "detecting the undetected." It should be emphasized that the purpose of OTTO and JONES (2000) was to infer the number of QTL for a quantitative trait within a single experiment. Some QTL, however, may have large effects and some may have small effects, and estimating the total number depends on the distribution of QTL effects. If the overall distribution of the effects is described by an exponential distribution, the distribution of detected QTL effects becomes a truncated exponential distribution after incorporating the Beavis ef-

[1]*Author e-mail:* xu@genetics.ucr.edu

fect. This should not be confused with the original Beavis experiment where all simulated QTL had an equal genetic effect.

HAYES and GODDARD (2001) conducted a meta-analysis for QTL mapping in livestock to infer the distribution of the effects of genes. They fit the estimated gene effects surveyed from all QTL mapping experiments in pigs and cattle to a gamma distribution. The difference between this meta-analysis and the usual analysis is that estimated genetic effects from multiple experiments were fit in a model rather than estimated from observed measurements. As a consequence, they had to take into account two additional sources of error for the estimated gene effects: the experimental error due to limited sample sizes and bias caused by data censoring. They found that the estimated genetic effects fit a gamma distribution very well. Again, one should not confuse this meta-analysis with the Beavis experiment because HAYES and GODDARD (2001) intended to infer the distribution of genetic effects throughout the genome from multiple experiments. Even if two QTL were identified roughly at the same location by two investigators, they were still treated as different QTL because the estimated positions were not precise enough to be assigned to exactly the same position within the genome. The original Beavis experiment demonstrated that a bias exists in estimating QTL effects. Herein, I present a theoretical basis for the Beavis effect. The theory predicts the amount of bias as a function of the type of progeny used in the experiment, the estimation procedure, the marker density, and the sample size.

STATISTICAL THEORY

Marker analysis: The theory is developed using a backcross (BC) mating design, which provides the simplest genetic model in QTL mapping. The result is then extended to other progeny derived from common mating designs. Let y_j be the phenotypic value of a quantitative trait measured from individual j sampled from a BC population. The linear model describing y_j is

$$y_j = \mu + x_j a + \varepsilon_j, \tag{1}$$

where μ is the population mean, a is the additive genetic effect at the locus of interest, and ε_j is the residual error, which is assumed to follow a normal distribution, $N(0, \sigma^2)$. In the BC progeny, there are only two possible genotypes at the QTL, heterozygote and homozygote for the backcross parent allele. The independent variable x_j is an indicator variable defined as $x_j = 0$ for the homozygote and $x_j = 1$ for the heterozygote.

Let $\hat{a}$ be the estimated genetic effect and $\sigma_{\hat{a}}^2$ be the variance of the estimate for a QTL. For simplicity, let us assume that the QTL is closely linked to a marker so that

$$\sigma_{\hat{a}}^2 = \frac{\sigma^2}{\sum_{j=1}^n (x_j - \bar{x})^2} \approx \frac{\sigma^2}{n\sigma_x^2}, \tag{2}$$

where n is the sample size (the number of BC progeny) and

$$\sigma_x^2 = n^{-1} \sum_{j=1}^n (x_j - \bar{x})^2 \tag{3}$$

is the sampling variance of the genotype. Note that n may be replaced by $n - 1$ for a more precise result.

The variance of x depends on the type of progeny. In a BC population without segregation distortion, half of the individuals will be homozygous and half heterozygous, leading to $\sigma_x^2 = \frac{1}{4}$ in the limit of large n. We have implicitly made the assumption that there is no dominance effect. If there is dominance, it will be confounded with the additive effect; *i.e.*, the genetic effect is actually $a + d$ instead of a. Therefore, all subsequent analyses are based on the additive model.

If the positions of all QTL are known *a priori* (uncensored), the estimate of each QTL effect is approximately unbiased (depending on the method used) so that the distribution of $\hat{a}$ can be assumed to be approximately normal, *i.e.*, $N(a, \sigma_{\hat{a}}^2)$. The normal distribution is mathematically convenient because it allows the standard statistical machinery for censored/truncated data to be used (COHEN 1991). Of course, if the assumption of normal distribution is violated, the theory developed here is not valid. The assumption of a normal distribution for $\hat{a}$ holds better if the residual error is normally, independently, and identically distributed. Some traits may be distributed in a discrete manner but analyzed as a quantitative trait, and some traits may have a skewed distribution with unequal variances. These traits may not have a normal residual error, and thus the theory developed in this study should be used with extra caution for such traits.

In the practice of QTL mapping, estimated genetic effects are reported only for significant QTL. Thus, the reported QTL consists of a censored sample so that the distribution of the estimated QTL effects conditional on statistical significance is a truncated normal distribution with a mean and variance different from those of the original normal distribution.

I now proceed to calculate this truncated distribution. Let us assume that a z-test statistic is used for the significance test so that the critical value for the z-test statistic is defined as $Z_{1-\alpha/2}$, where α is a controlled type I error rate. A QTL is reported only if $|z| > Z_{1-\alpha/2}$, where $z = \hat{a}/\sigma_{\hat{a}}$ is the z-test statistic. In other words, all the QTL reported satisfy $|\hat{a}| > \sigma_{\hat{a}} Z_{1-\alpha/2}$, *i.e.*, $\hat{a} < -\sigma_{\hat{a}} Z_{1-\alpha/2}$ or $\hat{a} > \sigma_{\hat{a}} Z_{1-\alpha/2}$. The two-tailed test leads to the possibility that even if a is positive, it may be detected as a significantly negative effect due to sampling. Denote the truncated $\hat{a}$ by $\hat{a}_T$ so that $\hat{a}_T \sim N_T(a_T, \sigma_{\hat{a}_T}^2)$, where N_T represents a truncated normal distribution, $a_T = E(\hat{a}_T)$ is the expectation, and $\sigma_{\hat{a}_T}^2 = \text{Var}(\hat{a}_T)$ is the variance of the truncated normal distribution. The truncated normal distribution is not symmetric. Let us define the two truncation points in the standardized (z-test statistic) scale as

$$\xi_1 = -Z_{1-\alpha/2} - \frac{a}{\sigma_{\hat{a}}} = -Z_{1-\alpha/2} - \frac{\sqrt{n}\sigma_x a}{\sigma} \quad (4)$$

and

$$\xi_2 = Z_{1-\alpha/2} - \frac{a}{\sigma_{\hat{a}}} = Z_{1-\alpha/2} - \frac{\sqrt{n}\sigma_x a}{\sigma}. \quad (5)$$

Further define

$$\psi_i = \frac{\phi(\xi_i)}{1 + \Phi(\xi_1) - \Phi(\xi_2)} \quad \text{for } i = 1, 2, \quad (6)$$

where $\phi(\xi)$ and $\Phi(\xi)$ are the standardized normal probability density function (pdf) and cumulative distribution function (cdf), respectively.

Often, when using likelihood methods, the LOD score criterion may be used for QTL detection. The critical value of the LOD score can be converted into the critical value of the z-test statistic, using the following approximate relationship,

$$Z_{1-\alpha} \approx \sqrt{\chi^2_{1,1-\alpha}} = \sqrt{4.605\,\text{LOD}} = 2.146\sqrt{\text{LOD}}, \quad (7)$$

where $\chi^2_{1,1-\alpha}$ is the critical value of the chi-square distribution with 1 d.f. and is also the critical value for the likelihood-ratio test statistic. The relationship between the likelihood-ratio test statistic and LOD can be found in LYNCH and WALSH (1998).

According to the standard statistical machinery of truncated normal distributions

$$E(\hat{a}_T) = a + \frac{\sigma}{\sqrt{n}\sigma_x}(\psi_2 - \psi_1) \quad (8)$$

(COHEN 1991), so that the bias is

$$b = \frac{\sigma}{\sqrt{n}\sigma_x}(\psi_2 - \psi_1). \quad (9)$$

The variance of the truncated sample is

$$\text{Var}(\hat{a}_T) = \frac{\sigma^2}{n\sigma_x^2}[(1 + \xi_2\psi_2 - \xi_1\psi_1) - (\psi_2 - \psi_1)^2]. \quad (10)$$

This truncated variance is used later when we discuss the bias in the QTL variance estimate.

The phenotypic variance may be partitioned into the true QTL variance (σ_G^2) and the residual variance (σ^2). The genetic variance of a QTL in a BC population is $\sigma_G^2 = a^2/4$. The corresponding variance in an F_2 population is $\sigma_G^2 = a^2/2$. In general, the genetic variance can be expressed as $\sigma_G^2 = \gamma a^2$, where γ is a constant depending on the filial relationship among segregating progeny; *e.g.*, $\gamma = \frac{1}{4}$ for BC and $\gamma = \frac{1}{2}$ for F_2. Therefore, a commonly used approach to estimating the genetic variance is $\sigma_G^2 = \gamma\hat{a}_T^2$. However, this estimation is also biased because

$$E(\hat{\sigma}_G^2) = E(\gamma\hat{a}_T^2) = \gamma[E^2(\hat{a}_T) + V(\hat{a}_T)]$$

$$= \gamma\left[\left(a + \frac{\sigma}{\sqrt{n}\sigma_x}(\psi_2 - \psi_1)\right)^2\right.$$

$$\left. + \frac{\sigma^2}{n\sigma_x^2}[(1 + \xi_2\psi_2 - \xi_1\psi_1) - (\psi_2 - \psi_1)^2]\right]$$

$$= \gamma\left[a^2 + \frac{\sigma^2}{n\sigma_x^2}\left[1 + \psi_2\left(\xi_2 + 2\frac{\sqrt{n}\sigma_x a}{\sigma}\right)\right.\right.$$

$$\left.\left. - \psi_1\left(\xi_1 + 2\frac{\sqrt{n}\sigma_x a}{\sigma}\right)\right]\right]$$

$$= \sigma_G^2 + \gamma\frac{\sigma^2}{n\sigma_x^2}[1 + (\xi_2'\psi_2 - \xi_1'\psi_1)], \quad (11)$$

where

$$\xi_i' = \xi_i + 2\frac{\sqrt{n}\sigma_x a}{\sigma} \quad \text{for } i = 1, 2. \quad (12)$$

Therefore, the bias in the estimated variance is

$$B = E(\hat{\sigma}_G^2) - \sigma_G^2 = \gamma\frac{\sigma^2}{n\sigma_x^2}[1 + (\xi_2'\psi_2 - \xi_1'\psi_1)]. \quad (13)$$

Note that there are two sources of bias: the contribution of environmental variance to the estimate of a^2 given by $[n\sigma_x^2]^{-1}\gamma\sigma^2$ plus the Beavis effect given by $[n\sigma_x^2]^{-1}\gamma\sigma^2 (\xi_2'\psi_2 - \xi_1'\psi_1)$. Thus, this estimator is biased even in the absence of the Beavis effect. The bias due to the first source of error has been investigated by LUO *et al.* (2003), who also provided a method to correct the bias.

The one-tailed test is a special case of the two-tailed test in that we simply set $\xi_1 = -\infty$ and $\psi_1 = 0$. For the two-tailed test, we can further simplify B into $B = \gamma (\sigma^2/n\sigma_x^2)[1 + (\xi_2\psi_1 - \xi_1\psi_2)]$ due to $\xi_1' = -\xi_2$ and $\xi_2' = -\xi_1$. We did not present this simplified version of B as a general formula because it would not allow us to extend the results simply to the one-tailed test.

The proportion of phenotypic variance explained by the QTL is defined as

$$h_Q^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma^2} = \frac{\gamma a^2}{\gamma a^2 + \sigma^2} = \frac{\gamma a^2}{\sigma_P^2}, \quad (14)$$

where $\sigma_P^2 = \gamma a^2 + \sigma^2$ is the total phenotypic variance among the test progeny. The h_Q^2 notation is adopted from classical quantitative genetics and is known as the heritability. When used as a proportion of variance contributed by an individual QTL, it is no longer called the heritability. The typical estimator for h_Q^2 is $\hat{h}_Q^2 = \gamma\hat{a}_T^2/\hat{\sigma}_P^2$. It is hard to find the expectation of a ratio. However, if we assume that the estimation error of the denominator is negligible, we can take $E(\hat{h}_Q^2) = E(\gamma\hat{a}_T^2)/\sigma_P^2$. This assumption is justified for many quantitative traits because the phenotypic variance is usually accurately estimated. On the basis of this assumption, we get

$$E(\hat{h}_Q^2) \approx \frac{\gamma a^2 + B}{\sigma_P^2} = h_Q^2 + \frac{B}{\sigma_P^2}. \quad (15)$$

Letting $\sigma^2 = (1 - h_Q^2)\sigma_P^2$, the bias of the genetic variance can be rewritten as

$$B = \gamma\frac{(1 - h_Q^2)\sigma_P^2}{n\sigma_x^2}\left[1 + (\xi_2'\psi_2 - \xi_1'\psi_1)\right], \quad (16)$$

which leads to

$$E(\hat{h}_Q^2) \approx h_Q^2 + \frac{\gamma}{n\sigma_x^2}\left[1 + (\xi_2'\psi_2 - \xi_1'\psi_1)\right](1 - h_Q^2). \quad (17)$$

Therefore, the proportion of the phenotypic variance associated with the QTL also is biased. Note that the term $(\xi_2'\psi_2 - \xi_1'\psi_1)$ in Equation 17 involves the genetic effect, which can be expressed using (14) as

$$a = \sigma\sqrt{\frac{h_Q^2}{\gamma(1 - h_Q^2)}}. \quad (18)$$

If we substitute Equation 18 into Equations 4 and 5, σ cancels out from $(\xi_2'\psi_2 - \xi_1'\psi_1)$. Therefore, the bias given in Equation 17 is only a function of h_Q^2 and the parameters associated with the sampled progeny.

We now extend the results to other types of progeny. For simplicity, we assume that dominance is absent. In an F_2 population, there are three possible genotypes whose genotypic values are defined as a for the homozygote with the "high allele," 0 for the heterozygote, and $-a$ for the homozygote with the "low allele." The linear model given in Equation 1 applies here in the F_2 population except that the x variable is now defined as $x_j = 1$, 0, -1 for the three genotypes, respectively. Without segregation distortion, the variance of x in an F_2 population is $\sigma_x^2 = \frac{1}{2}$. Therefore, the results derived above apply here with $\sigma_x^2 = \frac{1}{4}$ in BC progeny replaced by $\sigma_x^2 = \frac{1}{2}$ in F_2 progeny.

The method also can be applied to double-haploid (DH) and recombinant inbred line (RIL) populations. In both DH and RIL, the heterozygous genotype is absent. The x variable is defined as $x_j = 1$ for one homozygote and $x_j = -1$ for the other homozygote. The two types of homozygote have an equal frequency, and thus $\sigma_x^2 = 1$ in both DH and RIL. The results derived above apply by substituting $\sigma_x^2 = \frac{1}{4}$ in the BC by $\sigma_x^2 = 1$ in the DH and RIL.

Note that σ_x^2 and γ are identical for all types of progeny if a marker provides a fully informative genotype for the QTL. Different notation is used because there can be differences when the QTL is not tightly linked to a marker.

Interval mapping: In interval mapping, a QTL may be identified at an intermediate position between markers by inferring the genotype of the QTL from flanking marker information. This will affect both the parameter estimates and the statistical tests of inference. We need to substitute the variance of x by the variance of the estimated x, denoted by $\sigma_{\hat{x}}^2$. This variance depends on the relative position of the QTL within the interval. The

TABLE 1

Predicted genotypic indicator variable and its probability distribution for the four flanking marker genotypes in backcross progeny

Marker class	$\hat{x}_j = p_j(1) = \Pr(x_j = 1\|m_{1j}, m_{2j})$	$\Pr(m_{1j}, m_{2j})$
$m_{1j} = 1, m_{2j} = 1$	$(1 - r_1)(1 - r_2)/(1 - r)$	$(1 - r)/2$
$m_{1j} = 1, m_{2j} = 0$	$(1 - r_1)r_2/r$	$r/2$
$m_{1j} = 0, m_{2j} = 1$	$r_1(1 - r_2)/r$	$r/2$
$m_{1j} = 0, m_{2j} = 0$	$r_1 r_2/(1 - r)$	$(1 - r)/2$

r is the recombination frequency between m_1 and m_2.

explicit expression is tedious, but numerical values may be computed conveniently. For example, in a BC population, there are four possible marker classes, and the estimated x is denoted by $\hat{x}_j = E(x_j|m_{1j}, m_{2j}) = p_j(1)$, where m_{1j} and m_{2j} indicate the flanking marker genotypes and $p_j(1) = \Pr(x_j = 1|m_{1j}, m_{2j})$. Let $\Pr(m_{1j}, m_{2j})$ be the joint probability of the flanking marker genotype. These values are given in Table 1, where r_1 and r_2 are the recombination fractions between the QTL and the two markers, respectively, and r is the recombination between the two markers. From this table, we can calculate $\sigma_{\hat{x}}^2$ using

$$\sigma_{\hat{x}}^2 = E(\hat{x}^2) - E^2(\hat{x})$$
$$= \sum_{j=1}^4 \Pr(m_{1j}, m_{2j})\hat{x}_j^2 - \left[\sum_{j=1}^4 \Pr(m_{1j}, m_{2j})\hat{x}_j\right]^2. \quad (19)$$

Note that when $\sigma_{\hat{x}}^2$ is used in place of σ_x^2, the assumption of a normal distribution for $\hat{a}$ may be less valid.

The method used to derive $\sigma_{\hat{x}}^2$ implicitly assumes that the least-squares method (HALEY and KNOTT 1992) is used because the expectation of the QTL genotype indicator is calculated conditional on markers only. In maximum-likelihood analysis, the expectation is actually calculated conditional on both markers and the phenotypic values. However, it is extremely difficult to derive the variance using expectation conditional on both markers and phenotypes. Therefore, the way we calculate $\sigma_{\hat{x}}^2$ is considered as an approximation.

For a DH population, we define $\hat{x}_j$ in a slightly different way (see Table 2) but still use (19) to calculate $\sigma_{\hat{x}}^2$. For RIL, we use the same table (Table 2), but replace r_1, r_2,

TABLE 2

Predicted genotypic indicator variable and its probability distribution for the four flanking marker genotypes in double-haploid progeny

Marker class	$\hat{x}_j = 2p_j(1) - 1$	$\Pr(m_{1j}, m_{2j})$
$m_{1j} = 1, m_{2j} = 1$	$2(1 - r_1)(1 - r_2)/(1 - r) - 1$	$(1 - r)/2$
$m_{1j} = 1, m_{2j} = -1$	$2(1 - r_1)r_2/r - 1$	$r/2$
$m_{1j} = -1, m_{2j} = 1$	$2r_1(1 - r_2)/r - 1$	$r/2$
$m_{1j} = -1, m_{2j} = -1$	$2r_1 r_2/(1 - r) - 1$	$(1 - r)/2$

$\hat{x}_j = p_j(1) - p_j(-1) = 2p_j(1) - 1$, where $p_j(1) = \Pr(x_j = 1|m_{1j}, m_{2j})$ and $p_j(-1) = 1 - p_j(1)$.

<div align="center">

TABLE 3

Predicted genotypic indicator variable and its probability distribution for the nine flanking marker genotypes in F₂ progeny

</div>

Marker genotype	$\hat{x}_j = p_j(+1) - p_j(-1)$	$\Pr(m_{1j}, m_{2j})$
$m_{1j} = 1, m_{2j} = 1$	$\dfrac{(1 - r_1)^2(1 - r_2)^2 - r_1^2 r_2^2}{(1 - r)^2}$	$\frac{1}{4}(1 - r)^2$
$m_{1j} = 1, m_{2j} = 0$	$\dfrac{r_2(1 - r_1)^2(1 - r_2) - r_1^2 r_2(1 - r_2)}{r(1 - r)}$	$\frac{1}{2}r(1 - r)$
$m_{1j} = 1, m_{2j} = {}^{-}1$	$\dfrac{r_2^2(1 - r_1)^2 - r_1^2(1 - r_2)^2}{r^2}$	$\frac{1}{4}r^2$
$m_{1j} = 0, m_{2j} = 1$	$\dfrac{r_1(1 - r_1)(1 - r_2)^2 - r_1 r_2^2(1 - r_1)}{r(1 - r)}$	$\frac{1}{2}r(1 - r)$
$m_{1j} = 0, m_{2j} = 0$	0	$\frac{1}{2}[r^2 + (1 - r)^2]$
$m_{1j} = 0, m_{2j} = {}^{-}1$	$\dfrac{r_1 r_2^2(1 - r_1) - r_1(1 - r_1)(1 - r_2)^2}{r(1 - r)}$	$\frac{1}{2}r(1 - r)$
$m_{1j} = {}^{-}1, m_{2j} = 1$	$\dfrac{r_1^2(1 - r_2)^2 - r_2^2(1 - r_1)^2}{r^2}$	$\frac{1}{4}r^2$
$m_{1j} = {}^{-}1, m_{2j} = 0$	$\dfrac{r_1^2 r_2(1 - r_2) - r_2(1 - r_1)^2(1 - r_2)}{r(1 - r)}$	$\frac{1}{2}r(1 - r)$
$m_{1j} = {}^{-}1, m_{2j} = 1$	$\dfrac{r_1^2 r_2^2 - (1 - r_1)^2(1 - r_2)^2}{(1 - r)^2}$	$\frac{1}{4}(1 - r)^2$

$p_j(+1) = \Pr(x_j = +1 | m_{1j}, m_{2j})$ and $p_j(-1) = \Pr(x_j = -1 | m_{1j}, m_{2j})$.

and r by c_1, c_2, and c, where $c = 2r/(1 + 2r)$ (LYNCH and WALSH 1998).

In an F₂ population, there are nine flanking marker genotypic classes. The definitions of $\hat{x}_j$'s and their probabilities for the nine classes are given in Table 3. The same formula (Equation 19) is used to calculate $\sigma_{\hat{x}}^2$ except that the summation is taken over nine categories. For example, if the QTL position is in the middle of a 20-cM interval, we have $r_1 = r_2 = 0.0906$ and $r = 0.1648$, leading to $\sigma_{\hat{x}}^2 = 0.4013$. Figure 1 shows the relationship between $\sigma_{\hat{x}}^2$ and the position of the QTL in the 20-cM interval
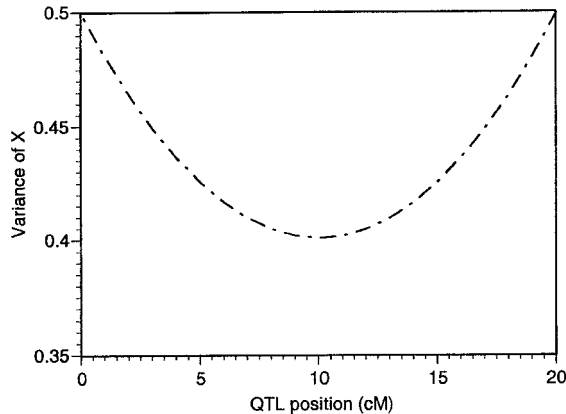

FIGURE 1.—Relationship between $\sigma_{\hat{x}}^2$ and the position of a QTL flanked by two fully informative markers in an interval of 20 cM in an F₂ population in the limit of large n.

flanked by two fully informative markers. Note, as the QTL position approaches a marker, $\sigma_{\hat{x}}^2$ approaches $\frac{1}{2}$.

NUMERICAL EVALUATION

Bias was evaluated numerically by considering the following factors: the sample size, the genetic effect measured by h_Q^2, and the LOD score criterion. In evaluating the bias of QTL effects, the residual variance was set at unity, i.e., $\sigma^2 = 1.0$. If the actual residual variance is not unity, one can always standardize the genetic effect using a/σ in place of a.

The numerical evaluation was conducted only in BC populations because the general trends are similar for all types of progeny (data not shown). In addition, only a one-tailed test was evaluated. The one-tailed test is a special case of the two-tailed test with $\xi_1 = -\infty$ and $\psi_1 = 0$. The functional relationships between the size of the detected QTL and the true sizes are shown in Figure 2. The diagonal lines in the first column of Figure 2 represent the case where $a_T = a$ and those in the second column represent the case where $\hat{h}_Q^2 = h_Q^2$, which holds only when the sample size is infinitely large. With finite sample sizes, the curves deviate from this straight line and the deviation increases as the sample size decreases. The deviation also increases as the LOD criterion increases. The deviation is negligible when $h_Q^2 = 0.5$ (corresponding to $a/\sigma = 0.2$), even if the sample size is as small as $n = 50$. Assume that the com-
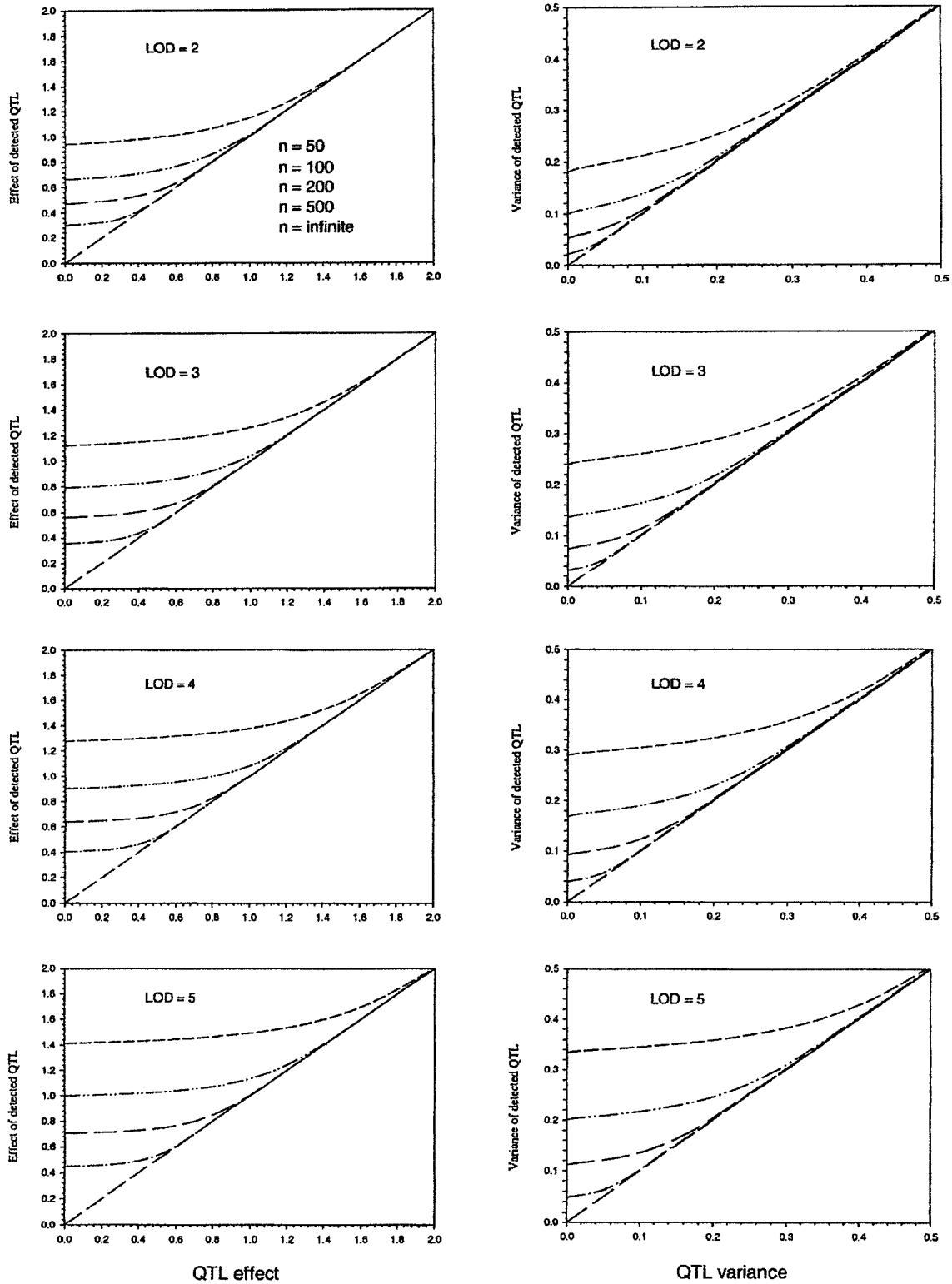
FIGURE 2.—Functional relationships of the detected QTL parameters to the true parametric values under various sample sizes and LOD score critical values in BC mapping. The diagonal lines represent the relationship between the parameter of detected QTL and the true parametric value when the sample size is infinitely large (unbiased relationship). Note that as the LOD criterion increases, the bias associated with detecting a very small QTL ($a \to 0$) becomes worse but the probability of detecting such a QTL becomes much smaller (not shown). Note that in the Beavis experiment, LOD = 2.5 and n = 100, 500, 1000.

monly used LOD criterion is 3 and the sample size is 200; the bias is within 7% of the true value as long as $h_Q^2 \geq 0.10$. The bias becomes more severe, however, for small detectable QTL. This observation is consistent with that observed by Beavis (1998). It is interesting to note that when $h_Q^2 = 0$ and $n = 50$, the bias in $\hat{h}_Q^2$ is as high as 0.18 for LOD = 2 and 0.33 for LOD = 5.

The functional relationships between $\hat{h}_Q^2$ and the sample size under various LOD criteria and h_Q^2 are shown in the right column of Figure 3. The corresponding plots for the effects are given in the left column. When $n \geq 200$ and $h_Q^2 \geq 0.10$, all the biases are negligible (within 10% of the true value), regardless of the LOD criterion. For small h_Q^2, a large sample size, even as large as $n = 500$, is not sufficient to eliminate the bias, again consistent with results of the Beavis experiment (Beavis 1998).

If the average estimated $\hat{h}_Q^2$ is 0.14 among all the experiments surveyed where the average sample size is $\sim$100 and the average LOD criterion is $\sim$3, the true h_Q^2 may actually be zero (found from the second panel from the top of the right column of Figure 2). If $\hat{h}_Q^2$ is 0.15 (just a slight increase), however, the true h_Q^2 is $\sim$0.08. If $\hat{h}_Q^2 = 0.25$, the true h_Q^2 is about the same as $\hat{h}_Q^2$; *i.e.*, very small bias is expected. From these graphs or using Equation 17, one can find the true h_Q^2 retrospectively from $\hat{h}_Q^2$ for all other settings.

We now use the parameter values of the Beavis experiment to compare the biases with those observed by Beavis (1998). Table 4 shows the original data reported by Beavis as well as the predicted biases for both the QTL effects and their variances. The average QTL position is in the middle of a 20-cM interval. However, for an interval this short, the position of a detected QTL rarely coincided with the true position. We observed that the detected position within the interval varied almost uniformly across the interval. Therefore, we choose an average $\sigma_x^2 = (0.5 + 0.4)/2 = 0.45$ as the estimated variance of x, where 0.5 is the variance when the QTL is completely linked with a marker and 0.4 is the variance when the QTL is exactly in the middle of the interval. The agreements between the observed and the predicted biases, judged by the differences between the two, are quite good except when the sample size is small ($n = 100$). However, the percentage differences, (observed-predicted)/predicted, show an opposite trend with >8% discrepancies in the variance observed only when $n = 1000$. The percentage difference may be misleading, however, as the predicted effects become smaller and more sensitive to error with increasing sample size and because, ultimately, one is interested in the absolute error made in inferring the effects of QTL. The large absolute deviations of the predicted biases from the observed values for small sample size may be explained as follows. The critical value of LOD = 2.5 was used by Beavis for a test statistic involving both the additive and dominance effects. Our prediction, however, was based on a test statistic for additive effects only. For small

samples, some QTL were detected due to large estimated dominance effects, even though the dominance was absent in the simulations (see Table 10.5 of Beavis 1998). To verify this, we redid the simulations for all the cases with $n = 100$ and used the pure additive model (dominance was included in neither the simulation nor the model). We used our own interval-mapping program known as QTL-BY-SAS to detect QTL (Xu and Xu 2003). Here, we replicated the experiment 500 times, instead of 200 times. The results are listed in Table 5, which shows very good agreement between the predicted and observed biases. Even though the LOD test statistic of the Beavis experiment involved 2 d.f. (additive and dominance), our model still predicted the biases quite accurately when the sample size was relatively large ($n \geq 500$). This was because for large samples, the LOD score test statistic was determined primarily by the additive effect. The 2-d.f. LOD score and 1-d.f. LOD score were virtually identical.

DISCUSSION

The Beavis effect describes a phenomenon that occurred in the Beavis experiment where all QTL were simulated to have the same effect and distributed independently throughout the genome. The average effect of the detected QTL was biased upward due to censoring. It is more likely that QTL effects vary across the genome and the distribution of the QTL effects may be described by a negative exponential distribution (Xu 2003). In addition, some QTL may be linked within the same chromosomes and thus they do not segregate independently. On the one hand, the Beavis effect will cause the estimated number of QTL to be biased downward (Beavis 1994, 1998; Otto and Jones 2000), because the undetected QTL are not reported. On the other hand, the average effect of the detected QTL will be biased upward.

Using the Beavis effect to interpret results of a meta-analysis of QTL mapping is more straightforward. If a QTL mapping experiment can be repeated many times, the average effect of a chromosome location calculated only from the significant replicates will definitely be biased unless this QTL is detected in all replicates. If one considers incorporating a particular marker into a marker-assisted selection program for an economically important quantitative trait, the Beavis effect will affect the decision. The investigator may decide to search the literature to see how much genetic variance is accounted for by this marker from all published experiments.

The theory developed here helps predict the potential bias in the estimated effect of QTL. The theory may also be used to correct the bias but should be used with caution. Let $\bar{h}_Q^2 = N^{-1}\Sigma_{k=1}^N \hat{h}_{Q(k)}^2$ be the average of the QTL effect estimated from N replicated experiments. To correct the bias, we may simply substitute the expectation given in Equation 17 by the observed average,
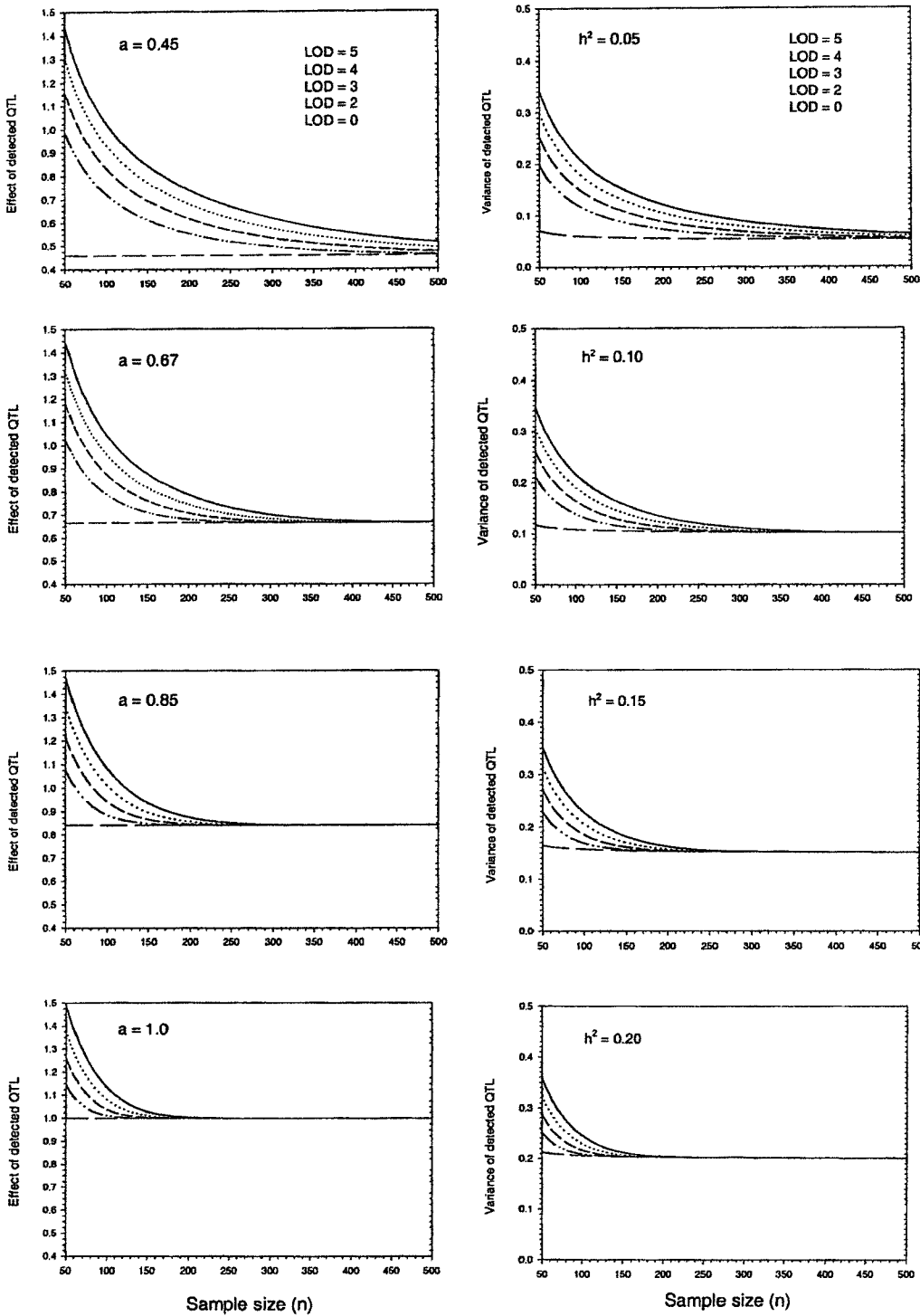
FIGURE 3.—Changes in effect and variance of detected QTL as a function of their true parametric values as sample size increases for various sizes of QTL and LOD score critical values in BC mapping.

$$\overline{h}_Q^2 = h_Q^2 + (1 - h_Q^2)N^{-1}\sum_{k=1}^{N}\frac{\gamma}{n_k \sigma_x^2}\big[1 + (\xi_2'\psi_2 - \xi_1'\psi_1)\big],$$
(20)

and solve for h_Q^2 to obtain the unbiased estimate of h_Q^2. The equation is highly nonlinear, but the solution can be easily solved numerically. Unfortunately, the solution is very sensitive to the sample size (n). It works only when n is sufficiently large, say $n \geq 500$. For a single experiment, we have only one estimate and the above

equation becomes

$$\hat{h}_Q^2 = h_Q^2 + \frac{\gamma}{n\sigma_x^2}\big[1 + (\xi_2'\psi_2 - \xi_1'\psi_1)\big](1 - h_Q^2). \qquad (21)$$

The solution is even more sensitive to n. Therefore, the theoretical prediction of the Beavis effect may not be used retrospectively to correct for the bias when the sample size is small. The correction is necessary when the estimate is obtained by summarizing the results of

TABLE 4

Comparisons of predicted and observed (estimated) biases in estimated QTL effects and variances from Beavis F$_2$ simulation experiments

Simulated conditions[a]	Variance explained			Additive effect			Average estimated location
	Simulated	Observed	Predicted[b]	Simulated	Observed	Predicted[c]	
10-30-100	3.00	16.76	16.0537	2.45	4.96	5.6410	11.3
10-30-500	3.00	4.33	4.1890	2.45	2.89	2.8617	10.53
10-30-1000	3.00	3.02	3.1846	2.45	2.56	2.4868	10.8
10-63-100	6.25	12.65	16.5984	3.55	4.68	5.7328	10.51
10-63-500	6.25	7.08	6.5581	3.55	3.73	3.5829	10.96
10-63-1000	6.25	6.34	6.3566	3.55	3.60	3.5500	11.04
10-95-100	9.50	18.68	17.3883	4.36	5.85	5.8466	10.58
10-95-500	9.50	10.1	9.7082	4.36	4.49	4.3607	11.08
10-95-1000	9.50	9.67	9.6028	4.36	4.44	4.3600	11.19
40-30-100	0.75	15.78	15.6270	1.22	4.40	5.5436	10.83
40-30-500	0.75	3.17	3.3332	1.22	2.35	2.5671	10.17
40-30-1000	0.75	1.46	1.7961	1.22	1.85	1.8790	10.17
40-63-100	1.56	16.31	15.7983	1.77	4.71	5.5999	10.45
40-63-500	1.56	3.54	3.5783	1.77	2.59	2.6582	10.13
40-63-1000	1.56	1.96	2.1435	1.77	2.09	2.0494	10.37
40-95-100	2.40	16.55	15.9694	2.18	5.02	5.6236	10.45
40-95-500	2.40	3.97	3.9190	2.18	2.79	2.7641	10.12
40-95-1000	2.40	2.58	2.6970	2.18	2.36	2.2784	10.29

[a] Numerical values denote the number of QTL-heritability-number of progeny.
[b] Using Equation 17.
[c] Using Equation 8.

many experiments. In that case, we should incorporate not only the mean of the estimates but also the variance of these estimates. The optimal method is the maximum-likelihood method that treats the estimates of the multiple experiments as censored data and infers (or recovers) the parameter of the uncensored data (COHEN 1991).

The theory developed herein applies to segregating populations with two alternative genotypes. For populations with more than two alternative genotypes, *e.g.*, F$_2$, the model is restricted to either the additive or the dominance model but not both. This is because the test statistic utilized is the *z*-test statistic, which is a 1-d.f. test. Further investigation is necessary to predict the biases in both the additive and dominance effects using a 2-d.f. test. Similar extensions can be made for a test with >2 d.f., *e.g.*, four-way crosses (XU 1996) and diallel crosses (REBAI and GOFFINET 1993).

The variance of the genotype indicator (x) determines the estimation error of the QTL effect and thus plays an important role in the Beavis effect. When the QTL under investigation is tightly linked to a fully informative marker, σ_x^2 is a constant depending on the filial

TABLE 5

Comparisons of predicted and estimated biases in estimated QTL effects and variances from our own F$_2$ simulation experiments at sample size 100

Simulated conditions[a]	Variance explained			Additive effect			Average estimated location
	Simulated	Observed	Predicted[b]	Simulated	Observed	Predicted[c]	
10-30-100	3.00	15.29	16.0537	2.45	5.48	5.6410	10.29
10-63-100	6.25	16.93	16.5984	3.55	5.76	5.7328	10.34
10-95-100	9.50	17.75	17.3883	4.36	5.89	5.8466	10.05
40-30-100	0.75	14.62	15.6270	1.22	5.36	5.5436	10.12
40-63-100	1.56	15.01	15.7983	1.77	5.42	5.5999	10.39
40-95-100	2.40	15.50	15.9694	2.18	5.51	5.6236	10.10

[a] Numerical values denote the number of QTL-heritability-number of progeny.
[b] Using Equation 17.
[c] Using Equation 8.

relationships among progeny. If the marker is not fully informative, *e.g.*, a certain percentage of individuals have missing genotypes at the marker, the variance σ_x^2 will decrease and this information loss will have the same ultimate effect as increasing the interval between markers when the QTL is not at the marker. We use σ_x^2 conditional on not only the type of progeny but also the marker information content and the QTL position. This conditional variance is denoted by $\sigma_{\tilde{x}}^2$ and it is used whenever the QTL genotype is not directly observable.

The theory of the Beavis effect is derived on the basis of a single QTL model. However, it applies approximately to multiple QTL experiments, as shown in the Beavis experiment where multiple QTL were actually simulated (Tables 4 and 5). When we predict the bias, the residual variance includes the environmental variance plus the sum of the variances of the remaining QTL excluding the one in question. For example, when we predicted the bias for the QTL explaining 3% of the phenotypic variance, the residual variance was chosen as $100 - 3 = 97$, where 100 is the total phenotypic variance. If the data were analyzed by a multiple-QTL model, *e.g.*, composite interval mapping (Jansen 1993; Zeng 1994), the residual variance should include only the environmental variance given that the remaining QTL have been included in the model. When QTL are linked, the situation will be complicated and further investigation is necessary.

LITERATURE CITED

Beavis, W. D., 1994   The power and deceit of QTL experiments: lessons from comparitive QTL studies, pp. 250–266 in *Proceedings of the Forty-Ninth Annual Corn & Sorghum Industry Research Conference*. American Seed Trade Association, Washington, DC.

Beavis, W. D., 1998   QTL analyses: power, precision, and accuracy, pp. 145–162 in *Molecular Dissection of Complex Traits*, edited by A. H. Paterson. CRC Press, New York.

Cohen, A. C., 1991   *Truncated and Censored Samples*. Marcel Dekker, New York.

Haley, C. S., and S. A. Knott, 1992   A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity **69:** 315–324.

Hayes, B., and M. E. Goddard, 2001   The distribution of the effects of genes affecting quantitative traits in livestock. Genet. Sel. Evol. **33:** 209–229.

Jansen, R. C., 1993   Interval mapping of multiple quantitative trait loci. Genetics **135:** 205–211.

Lande, R., 1981   The minimum number of genes contributing to quantitative variation between and within populations. Genetics **99:** 541–553.

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

Luo, L., C. Mao and S. Xu, 2003   Correcting the bias in estimation of genetic variances contributed by individual QTL. Genetica **119:** 107–113.

Lynch, M., and B. Walsh, 1998   *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.

Otto, S. P., and C. D. Jones, 2000   Detecting the undetected: estimating the total number of loci underlying a quantitative trait. Genetics **156:** 2093–2107.

Rebai, A., and B. Goffinet, 1993   Power of tests for QTL detection using replicated progenies derived from a diallel cross. Theor. Appl. Genet. **86:** 1014–1022.

Sillanpaa, M. J., and E. Arjas, 1998   Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. Genetics **148:** 1373–1388.

Xu, C., and S. Xu, 2003   A SAS/IML program for mapping QTL in line crosses, pp. 1–6 in *Proceedings of the Twenty-Eighth Annual SAS Users Group International Conference (SUGI 28)*. SAS Institute, Cary, NC.

Xu, S., 1996   Mapping quantitative trait loci using four-way crosses. Genet. Res. **68:** 175–181.

Xu, S., 2003   Estimating polygenic effects using markers of the entire genome. Genetics **163:** 789–801.

Zeng, Z-B., 1992   Correcting the bias of Wright estimates of the number of genes affecting a quantitative character—a further improved method. Genetics **131:** 987–1001.

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