Detecting the Undetected: Estimating the Total Number of Loci Underlying a Quantitative Trait

Sarah P. Otto* and Corbin D. Jones†

*Department of Zoology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada and †Department of Biology, University of Rochester, Rochester, New York 14627

Manuscript received December 17, 1999
Accepted for publication July 28, 2000

ABSTRACT

Recent studies have begun to reveal the genes underlying quantitative trait differences between closely related populations. Not all quantitative trait loci (QTL) are, however, equally likely to be detected. QTL studies involve a limited number of crosses, individuals, and genetic markers and, as a result, often have little power to detect genetic factors of small to moderate effects. In this article, we develop an estimator for the total number of fixed genetic differences between two parental lines. Like the Castle-Wright estimator, which is based on the observed segregation variance in classical crossbreeding experiments, our QTL-based estimator requires that a distribution be specified for the expected effect sizes of the underlying loci. We use this expected distribution and the observed mean and minimum effect size of the detected QTL in a likelihood model to estimate the total number of loci underlying the trait difference. We then test the QTL-based estimator and the Castle-Wright estimator in Monte Carlo simulations. When the assumptions of the simulations match those of the model, both estimators perform well on average. The 95% confidence limits of the Castle-Wright estimator, however, often excluded the true number of underlying loci, while the confidence limits for the QTL-based estimator typically included the true value ~95% of the time. Furthermore, we found that the QTL-based estimator was less sensitive to dominance and to allelic effects of opposite sign than the Castle-Wright estimator. We therefore suggest that the QTL-based estimator be used to assess how many loci may have been missed in QTL studies.

GENETIC studies of quantitative trait loci (QTL) are beginning to reveal the genetic basis of phenotypic differences. In several cases, researchers have pinpointed the genetic changes that have occurred during the processes of natural selection, artificial selection, and speciation (for example, Doebley and Stec 1991; Paterson *et al.* 1991; Mackay 1996; Laurie *et al.* 1997; Bradshaw *et al.* 1998). All QTL studies, however, involve a limited number of individuals and markers. Consequently, although QTL studies can detect loci that explain a large fraction of the phenotypic difference between two divergent lines, they have trouble identifying loci of more subtle effect. Thus, QTL studies are known to underestimate the total number of loci involved.

Estimating the total number of underlying loci is valuable for several reasons. First, given that the detected QTL generally represent only a fraction of the total set of QTL, it is worth obtaining accurate estimates of the number and average effect size of the undetected QTL before embarking upon further studies. Although it is tempting to consider simply the fraction of the total parental difference (or segregation variance) accounted for by the detected QTL, this procedure overestimates the importance of what has been detected because of an inherent bias in QTL analyses where the same data are used to detect QTL and to determine their effect sizes [a bias known as the Beavis (Beavis 1994, 1998) effect]. Second, most quantitative genetics theory assumes that many loci contribute to the observed phenotypic variation for a trait. Therefore, to determine the applicability of quantitative genetics theory, it is important to know whether there really are a large number of underlying loci, even if a QTL study detects only a small fraction of them. Third, minor effect QTL may have a much larger effect when different parental lines are involved or under different environmental conditions (Leips and Mackay 2000). Given that the magnitude of a QTL's effect may depend on the experimental crosses and conditions, we may wish to have more general information about how many genes make some contribution to the observed phenotypic difference. For example, we predict that different studies of the same populations would be more likely to detect a different set of QTL when there are many undetected QTL than when there are few.

In this study, we investigate the relationship between the detected number of QTL (n_d) and the total number of genetic differences underlying a trait difference between parental lines (n). Based on the number and magnitude of the detected QTL, we develop a QTL-

Corresponding author: S. P. Otto, Department of Zoology, University of British Columbia, Vancouver, BC V6T 124.
E-mail: otto@zoology.ubc.ca

based estimator (n_{QTL}) for the total number of loci underlying an observed trait difference between parental lines.

To evaluate our estimator, we performed hundreds of simulated QTL experiments. For each "experiment," we generated parental genomes carrying a specified number of QTL (n) with effects drawn at random from a gamma distribution. We then simulated crosses to generate an F_1 and an F_2 population. A QTL analysis was then conducted on the basis of the marker genotypes and phenotypes of these F_2 individuals. We could thus compare our estimated number of underlying loci (n_{QTL}) to the true number (n). We also assessed the performance of the classical Castle-Wright estimator (n_{CW}), which is based on the segregation variance observed among hybrids (reviewed by Lynch and Walsh 1998, pp. 233–243).

For both our QTL-based estimator and the Castle-Wright estimator, the expected distribution of allelic effects must be specified. We begin, therefore, with a discussion of this distribution and its shape. Second, we discuss the Castle-Wright and related estimators. Third, we develop our QTL-based estimator under the assumption of an exponential and a gamma distribution of allelic effect sizes. Finally, we present results from our simulated QTL experiments.

DISTRIBUTION OF EFFECT SIZES

Surprisingly, there is little theory about the suite of genetic differences that are likely to be observed among recently diverged taxa. While we have long known how to calculate the probability of fixation of a single mutation and how this depends on its effect on fitness, few studies have examined the entire set of substitutions likely to result from a period of adaptation.

Using an argument presented by Gillespie (1991, p. 266), we might expect an exponential distribution of selection coefficients among alleles that have fixed within a population. Gillespie noted that if all possible alleles at a locus were ordered in terms of absolute fitness from lowest to highest, then the selection coefficient of the second most fit allele relative to the most fit allele would follow an exponential distribution when measured across all loci. (His proof is based on extreme value theory.) Therefore, if the previously most fit allele at a locus falls only to second place in the fitness ordering when the environment (internal or external) changes, then the selection coefficients of substituted alleles should follow an exponential distribution. That is, we might expect an exponential distribution of selection coefficients among alleles that have fixed within a population if evolutionary change is accomplished primarily by the successive replacement of the two most fit alleles. The proof assumes, however, that genetic interactions are absent and that allelic fitnesses follow the same distribution at each locus, which is unlikely given functional differences among genes.

More recently, Orr (1998) studied the process of adaptation in a geometrical model first proposed by Fisher (1958). In this model, a fitness landscape in multiple dimensions has a single fitness optimum. Initially, the population is displaced from this optimum, perhaps because of a recent shift in the environment. Each new mutation that occurs affects a number of traits but does so to varying degrees. Whether the mutation is beneficial or deleterious depends on the current state of the population and on whether the mutation brings the population closer to the fitness optimum. Orr tracked the series of mutations that were incorporated into the population as it approached the adaptive optimum. In general, adaptation toward a peak in the fitness landscape produced a series of genetic changes whose effects on phenotype followed an approximately exponential distribution, a result that held under a broad variety of plausible distributions for the effects of mutations (Orr 1998, 1999).

Finally, an exponential distribution of fitness effects may also be a reasonable outcome of selection with a moving optimum. To show this, we assume that the fitness effects of new beneficial mutations follow a gamma distribution with mean, μ, and coefficient of variation, C. The shape of the gamma distribution depends on the coefficient of variation (Figure 1). When $C = 1$, the gamma is equivalent to an exponential distribution. For $C > 1$, the distribution becomes more L-shaped, while for $C < 1$, the distribution approaches a bell shape. We also assume that the distribution of newly arising beneficial mutations remains approximately constant throughout the adaptive process, which

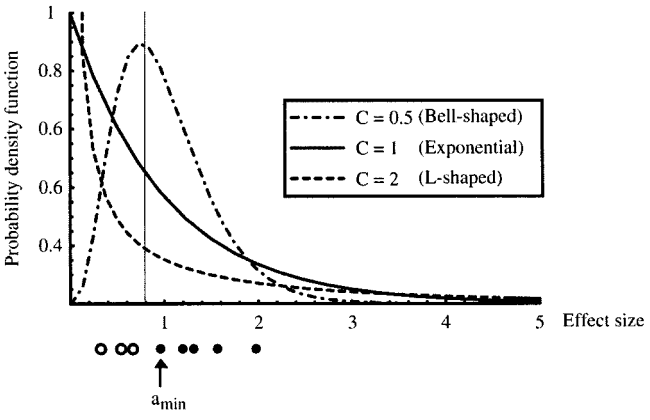

Figure 1.—Potential distributions of allelic effects. Each curve describes a gamma distribution with mean $\mu = 1$ but with different coefficients of variation (C). The QTL underlying a particular phenotypic difference represent draws from the appropriate distribution, as illustrated by the circles under the x-axis. Only those QTL above the threshold of detection ($\theta = 0.8$, thin vertical line) are likely to be detected (solid circles). Those below the threshold are likely to remain undetected (open circles).

is plausible if a population lags steadily behind a moving optimum or fitness threshold. The distribution of fitness effects among those beneficial mutations that fix within a population can be calculated by weighting the original gamma distribution by each allele's probability of fixation, which, in populations of large size, is approximately twice the allele's selective advantage. It can be shown that the distribution of beneficial alleles, conditioned on their fixation, is also gamma, but with mean $\mu(1 + C^2)$ and coefficient of variation $\sqrt{C^2/(1 + C^2)}$. Interestingly, this conditional distribution always has a coefficient of variation <1. Hence, it is either exponential or bell shaped, even if newly arising beneficial mutations have a very L-shaped distribution. Unfortunately, we know little about the distribution of fitness effects of mutations in general and of beneficial mutations in particular. The distribution of spontaneous deleterious mutations is thought to be roughly L-shaped, with estimates of C ranging from 2 to 5 (Keightley 1994; Lynch *et al.* 1999). If newly arising beneficial mutations follow a distribution of similar shape, then those beneficial mutations that fix will follow a gamma distribution that is nearly exponential in shape, with a coefficient of variation between 0.9 and 1.

The above arguments suggest that an exponential distribution might often describe the distribution of effect sizes among alleles that arise and fix during adaptive divergence. There may, however, be circumstances under which alternative distributions are plausible. For example, one might be examining a trait that has diverged as a pleiotropic response to selection on other traits or through neutral drift. In these cases, the distribution should more closely reflect the distribution of effect sizes among new mutations. On the other hand, if one of the parental lines has been subject to rapid and strong selection over a very short period of time, then only large-effect mutations will have had enough time to fix. In this case the distribution of fixed effects may be more normal in shape with a mean offset from zero. Similarly, the fact that researchers choose traits and parental lines that are particularly divergent may bias the distribution of effect sizes such that a greater proportion of allelic effects are large. The gamma distribution is a natural choice to describe these various alternatives, because its shape is so flexible. Here, we focus on the exponential distribution, both because it has theoretical support and because the results are simpler to understand. We also provide a more general derivation that uses a gamma distribution to describe the underlying effects of alleles on the trait of interest and that may allow a more flexible approach to fitting real data.

THE CASTLE-WRIGHT ESTIMATOR

Historically, one of the most widely used methods for estimating the number of loci underlying a trait difference between two lines was developed by Castle (1921) and Wright (1968). Their method is based on the amount of segregation variance observed in controlled crosses. Intuitively, if there is one or very few Mendelian factors underlying a trait, then some F_2 individuals will have the same genotypes as the parents, and the phenotypic range observed among the F_2's should span the entire range of the parental lines, P_1 and P_2. As the number of loci contributing to the trait increases, however, F_2 individuals will more often fall near the average of the parental lines as a result of the central limit theorem. Thus the amount of segregation variance, $\mathrm{Var}(S)$, estimated from hybrid (F_2, F_3, etc.) and backcross generations, contains information regarding the number of alleles underlying the phenotypic difference between two parental lines. If $\Delta\bar{z}$ is the difference between parental lines in the trait of interest, the Castle-Wright estimator for the number of underlying loci is

$$\hat{n}_{\mathrm{CW}} = \frac{(\Delta\bar{z})^2}{8\ \mathrm{Var}(S)}. \qquad (1)$$

Equation 1 assumes that the underlying loci are unlinked and that they have equal effects. Consequently, (1) is said to estimate the "effective" number of underlying loci, equivalent to the number that there would be if all QTL were unlinked and had effects of equal magnitude.

Numerous improvements and extensions have been made to the Castle-Wright estimator (summarized in Lynch and Walsh 1998, chapter 9). In particular, Zeng (1992) analyzed a generalized model that allowed for linkage among loci and variation in their effects. He estimated the number of underlying loci to be

$$\hat{n}_{\mathrm{CWZ}} = \frac{2\bar{c}\,\hat{n}_{\mathrm{CW}} + C^2(\hat{n}_{\mathrm{CW}} - 1)}{1 - \hat{n}_{\mathrm{CW}}(1 - 2\bar{c})}, \qquad (2)$$

where $\bar{c}$ is the average recombination rate between loci, and C is the coefficient of variation for the distribution describing the additive effects. Given that allelic effects are likely to vary from locus to locus, Equation 2 should more closely estimate the true number of underlying loci than does the effective number presented in (1). Equation 2 requires, however, that the distribution of effects be specified. Fortunately, it can be used with a variety of distributions describing the additive effects of the QTL, including an exponential and a gamma distribution.

Sampling variances for $\hat{n}_{\mathrm{CW}}$ and $\hat{n}_{\mathrm{CWZ}}$ have been determined by Lande (1981) and Zeng (1992). Rather than reiterate their general findings, we focus here on the sampling variance relevant to our simulation study. We assume that the parental-line means are known with negligible error and that the trait has been measured under controlled environmental conditions (*i.e.*, the parental difference is attributable to genetic differ-

ences). In this case, the segregation variance reduces to the variance among F_2 individuals, and the variance of the Castle-Wright-Zeng estimator becomes

$$\mathrm{Var}(n_{\mathrm{CWZ}}) = 2\frac{(2\bar{c}(1 + C^2)\hat{n}_{\mathrm{CW}})^2}{(N_{\mathrm{F_2}} + 2)(1 - \hat{n}_{\mathrm{CW}}(1 - 2\bar{c}))^4}, \quad (3)$$

where $N_{\mathrm{F_2}}$ is the number of F_2 individuals measured. Equation 3 can be used to construct confidence limits for $\hat{n}_{\mathrm{CWZ}}$ under the assumption that the error in the estimator is normally distributed.

A QTL-BASED ESTIMATOR

In light of the growing number of QTL studies, we develop an alternate method, based on data generated in a QTL mapping study, to estimate the total number of loci underlying a trait difference. In a typical QTL analysis, a handful of the loci that contribute to the trait difference are detected, each with an estimated additive and dominance effect and position. Our method takes advantage of the fact that the power to detect a QTL depends on the size of its effect. Given an expected distribution of effects, we can estimate the number of loci whose effects were too small to be detected (see Figure 1).

We first need to determine how the probability of detecting a QTL depends on its effect size. This power curve is approximately logistic in shape for a simple QTL study with one marker linked to one QTL (Figure 2). With $N_{\mathrm{F_2}}$ individuals scored in the F_2 generation, the probability of detecting a QTL rises at some point from near 0 to near 1. The more F_2 individuals are examined, the more steeply the curve rises. With multiple markers, however, significance is usually assessed using a permutation test (Churchill and Doerge 1994). Without knowledge of the form of the power curve in a typical QTL study with multiple markers, we approximate the power curve with a threshold function, rising from 0 to 1 at a threshold θ. This approximation should be most accurate when many F_2 individuals are scored. Other power functions can be explored, at least numerically, using the approach developed below, although simulations indicate that the estimator developed using this approximation is reasonably accurate as long as there are enough detected QTL (more than two) to provide a good estimate for the threshold.

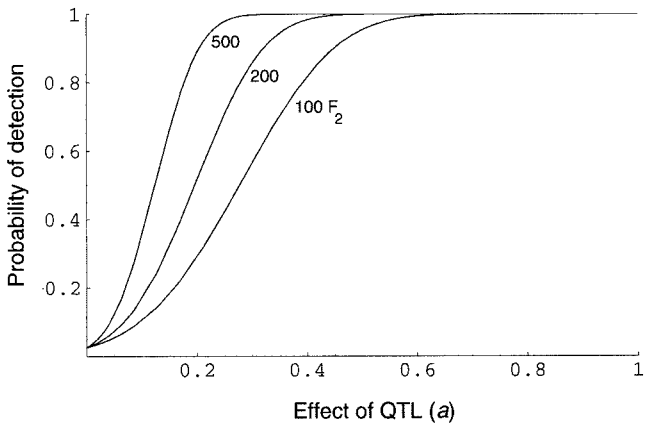

Figure 2.—The power to detect a completely linked QTL with one marker (solid curves), based on equations 15.36b and 15.37 of Lynch and Walsh (1998). The QTL was assumed to be completely additive with effect a, such that the expected difference between homozygotes at this locus was $2a$. The type I error rate, α, was set to 0.05, and the phenotypic variance among F_2's was scaled to 1.

We assume that the phenotypic difference between the parental lines, $\Delta\bar{z}$, is primarily caused by fixed allelic differences at underlying QTL. We define the additive (a) and dominance (d) effects of each allele according to Zeng (1992),

Population:	P_1	F_1	P_2
Genotype:	A_1A_1	A_1A_2	A_2A_2
Average phenotype:	a	d	$-a$,

where we refer to a as the effect size or additive effect of an allele. Here, we assume that each allele contributes to the parental difference in the same direction, that is, a is always positive or always negative. This assumption is reasonable if divergence is primarily a product of selection acting in different directions within the two populations or in a novel direction in one population. (In a later section, we explore the effect of relaxing this assumption through simulations.) We let D denote the sum of the additive effects across all QTL. Under the above assumptions, D also equals half the phenotypic difference between parental lines ($\Delta\bar{z}/2$). We first derive an estimator for the number of loci underlying the trait difference assuming an exponential distribution of allelic effects and given an estimate of θ (the minimum threshold of detection). We then discuss how to obtain confidence limits for this estimator and how to estimate θ from QTL data. Finally, we repeat these steps assuming allelic effects follow the more general gamma distribution.

Exponentially distributed effect sizes: As argued above, we might expect alleles underlying phenotypic divergence to be exponentially distributed. To begin, we proceed under the assumption that the additive effect sizes (a) represent draws from an exponential distribution with mean, μ, and probability density function

$$f[x, \mu] = \frac{\mathrm{Exp}[-x/\mu]}{\mu}. \quad (4)$$

If there are a total of n underlying QTL, then the mean effect size will be $\mu = D/n$. Using (4), we can determine the probability density function for *detectable* QTL, *i.e.*, those that lie above the threshold θ,

$$f_{\mathrm{d}}[x, \mu] = \frac{f[x, \mu]}{\int_{\theta}^{\infty} f[x, \mu]\,dx} = \frac{\mathrm{Exp}[-(x - \theta)/\mu]}{\mu}, \quad (5)$$

which is simply the same exponential distribution with an origin shifted to the right by an amount, θ. The mean additive effect size among detected QTL is therefore expected to equal $\mu + \theta = D/n + \theta$. If the number of QTL actually detected is n_d and their average effect size is M, we can estimate the total number of loci by setting M to its expectation, $D/n + \theta$. Rearranging this equation, the estimated number of underlying loci is

$$\hat{n}_{QTL} = \frac{D}{M - \theta}. \quad (6)$$

Equation 6 can also be derived using a likelihood approach, which will enable us to calculate confidence limits for the number of loci underlying a trait difference. From the probability density function for observable QTL, (5), the likelihood of observing a set of QTL whose additive effect sizes are a_i ($i = 1 \ldots n_d$) is proportional to

$$L[n] = \prod_{i=1}^{n_d} f_d[a_i, \mu] = \frac{\mathrm{Exp}[-n_d(M - \theta)/\mu]}{\mu^{n_d}}. \quad (7)$$

Substituting in $\mu = D/n$ and solving for the maximum, the most likely value for the true number of underlying loci is again given by (6).

Confidence limits: Confidence limits for $\hat{n}_{QTL}$ can be obtained using the likelihood-ratio test, which holds that the values of n whose log likelihoods lie within $\chi^2_{1[\alpha]}/2$ of the maximum log likelihood comprise an approximate $1 - \alpha$ confidence region. This amounts to solving the equation

$$(\ln L[\hat{n}_{QTL}] - \ln L[n]) = \frac{\chi^2_{1[\alpha]}}{2} \quad (8)$$

for n. From (7), the confidence limits for n equal the two roots (found by numerical solution or plotting) to the equation

$$-n_d + \frac{(M - \theta)\, n n_d}{D} - n_d \ln\left[\frac{(M - \theta)\, n}{D}\right] - \frac{\chi^2_{1[\alpha]}}{2} = 0, \quad (9)$$

where $\chi^2_{1[0.05]} = 3.841$ is used to obtain 95% confidence limits. More exact confidence limits can be found using the methods described for exponential distributions in Larsen and Marx (1985, p. 313), but numerical examples suggest that the two limits are similar. Note that this approach does not constrain estimates for n to be greater than the number of detected loci; if desired, a constrained likelihood surface can be examined (with $n \geq n_d$), and the confidence limits can be adjusted accordingly.

Average effect of the undetected QTL: Because of the Beavis effect, the effects of the detected QTL are inflated and may appear to explain more of the trait difference between two parental lines than is the case. Here, we use the exponential distribution to estimate the relative importance of the undetected QTL. Specifically, we estimate the expected effect size of the undetected QTL, $M_{undetected}$, by averaging the exponential distribution within the range, 0 to θ. Recalling that the mean effect of all QTL (μ) is estimated by $M - \theta$, we obtain the estimate

$$M_{undetected} = \frac{\int_0^\theta x f[x, \mu]\, dx}{\int_0^\theta f[x, \mu]\, dx} = M\left(1 - \frac{1 - \tau}{1 - \mathrm{Exp}[-(1 - \tau)/\tau]}\right), \quad (10)$$

where $\tau = (M - \theta)/M$. τ is a measure of the power of a QTL experiment. It ranges from 0 for very weak experiments with a threshold near the mean detected effect to 1 for a very powerful experiment with a threshold near 0. Equation 10 indicates that the average effect among the undetected QTL reaches a maximum of 23% M when $\tau = 0.36$. For experiments with little power (τ near 0), the relative size of the undetected QTL decreases toward zero because any QTL that are detected are likely to have effects well above the average. Similarly, for powerful experiments (τ near 1), most QTL will have been detected, and the remaining ones will have a relatively small effect. For the simulation experiments described below, the average power ranged from 0.27 to 0.61 (see Table 1), where the experiments with 500 F_2's were more powerful. Over this range, the expected effect of the undetected QTL is 17–23% of the average effect of the detected QTL (M).

As an example, in the simulation study described below with 20 underlying QTL and 500 F_2's (see Table 1), the detected QTL appeared to explain just over 100% of the parental difference, on average, even though less than half of the underlying QTL were detected. In fact, the undetected QTL accounted for ~20% of the parental difference. Thus, the effects of undetected QTL can be substantial even when a QTL study indicates that the entire difference between parental lines has been explained.

One can also determine the expected fraction of segregation variance accounted for by the undetected QTL. From Lynch and Walsh (1998) Equation 9.26a, the segregation variance for unlinked loci equals ½ times the sum of the squared average effects of all the QTL. The expected fraction of the segregation variance contributed by loci whose effects lie below the threshold of detection, θ, can be shown to equal

$$1 - \frac{1}{2}\mathrm{Exp}\left[-\frac{1 - \tau}{\tau}\right]\left(1 + \frac{1}{\tau^2}\right).$$

For example, in the simulation experiments, where the average power (τ) ranged from 0.27 to 0.61 (see Table 1), the fraction of the segregation variance that we expect to be accounted for by undetected QTL ranges

from 51 to 3%, decreasing rapidly as the power of the experiment increases. Although one could simply calculate the total segregation variance contributed by the observed QTL, this procedure would overestimate the importance of the detected QTL because of the Beavis effect, which causes an especially large bias in variance calculations where the effect sizes of detected QTL are squared.

Estimating the threshold of detection: We have not yet addressed how to estimate the unknown threshold, θ. Ideally, one would compute the power curve for the probability of detecting a QTL. For example, one could perform Monte Carlo simulations, placing QTL of known effects on simulated genomes using the same number of markers and individuals as in the planned experiment. Alternatively, one can use the observed QTL data to estimate θ, as follows. If there are several detected QTL, then the magnitude of the smallest observed additive effect size ($a_{\min}$) will often be near the threshold and can be used as an approximation for θ. Indeed, $a_{\min}$ is the maximum-likelihood estimator for θ because the shape of the exponential distribution is such that the most probable value for the smallest observed QTL is at the origin of the distribution (*i.e.*, at θ). This is clearly a biased estimate for the threshold, because the smallest observed QTL will not lie below the minimum detectable size and will generally lie above θ. To obtain a less biased estimate for θ, we use a Bayesian approach. We assume that the threshold lies somewhere between 0 and $a_{\min}$ but that every point within this range is equally plausible (θ has a uniform prior distribution). We then weight the prior distribution for θ by the probability that the smallest detected QTL has additive effect size $a_{\min}$. Noting that the minimum of n_d draws from an exponential distribution with mean μ is itself exponential with mean μ/n_d (Feller 1971, p. 18) and recalling that detectable loci follow an exponential distribution shifted to the right by θ, the probability density function for $a_{\min} - \theta$ is exponential with mean μ/n_d. This gives us a posterior distribution for θ, whose average value is our estimated threshold ($\hat{\theta}$), which can be found to solve

$$0 = a_{\min} - \hat{\theta} - \frac{M - \hat{\theta}}{n_d} + \frac{a_{\min}\mathrm{Exp}[-a_{\min}n_d/(M - \hat{\theta})]}{1 - \mathrm{Exp}[-a_{\min}n_d/(M - \hat{\theta})]}. \quad (11)$$

If several QTL are detected, the last term becomes negligible, and the estimate for θ from (11) approaches

$$\hat{\theta} \approx \frac{a_{\min}n_d - M}{n_d - 1}. \quad (12)$$

Equation 12 makes sense: we would expect the smallest observed QTL to lie above θ by μ/n_d (because the distribution of the smallest QTL is an exponential with parameter μ/n_d shifted to the right by θ). Hence, $E[a_{\min}] = \theta + \mu/n_d = \theta + (E[M] - \theta)/n_d$, which rearranges to give (11). With few detected QTL, however, Equation 12 is subject to substantial sampling error and can become negative, which is biologically unrealistic. Equation 11, however, is always bounded between 0 and $a_{\min}$.

One potential problem with using $a_{\min}$ and M in our estimators is that they will be overestimated in QTL studies as a result of the Beavis effect. This occurs because the effect sizes of the QTL are estimated from the same data used to detect the QTL. Those QTL that, by chance, happen to have a larger effect than expected are more likely to be detected and so lead to inflated estimates of a_i. There is currently no analytical method to correct for the Beavis effect, so its impact was tested through simulation (described below). The simulations indicate that the Beavis effect did not cause a large bias in our QTL-based estimators. We suspect that $\hat{n}_{QTL}$ is not extremely sensitive to the Beavis effect because both M and θ tend to be inflated, but the estimator depends primarily on their difference [see (6) and (9)], which may not be as strongly biased.

Gamma-distributed effect sizes: We now describe more general results that apply when allelic effects follow a gamma distribution with mean, μ, and coefficient of variation, C. The probability density function of a gamma distribution is

$$g[x, \mu, C] = \frac{\mathrm{Exp}[-x/(C^2\mu)]x^{(1-C^2)/C^2}(C^2\mu)^{-1/C^2}}{\Gamma[1/C^2]}, \quad (13)$$

where $\Gamma[x]$ is the gamma function (Abromowitz and Stegun 1972). Using the same methods as were used to derive (6), we can show that the estimated number of underlying loci, $\hat{n}_{QTL}$, equals $\Psi D/(M - \theta)$, where Ψ solves

$$\Psi\Gamma\left[\frac{1}{C^2}, \frac{(1-\tau)\Psi}{\tau C^2}\right] = \tau C^2\Gamma\left[1 + \frac{1}{C^2}, \frac{(1-\tau)\Psi}{\tau C^2}\right]. \quad (14)$$

Here, $\tau = (M - \theta)/M$, and $\Gamma[x, y]$ is the digamma function (Abromowitz and Stegun 1972). Ψ is essentially a correction factor for (6) that must be applied when the expected distribution of underlying effect sizes is gamma. If $C = 1$, the gamma distribution reduces to the exponential distribution, and Ψ becomes one. For other values of C, the correction term depends only on C and τ and is illustrated in Figure 3. When $C > 1$, the gamma distribution is L shaped, and more loci have very small effect. These small-effect loci are missed if one mistakenly assumed an exponential distribution and used Equation 6. Therefore, (6) underestimates the number of underlying loci ($\Psi > 1$). The converse is true when $C < 1$, and the gamma distribution is bell shaped. In this case, fewer loci have small effect, and fewer loci fall below the threshold of detection. Consequently, (6) overestimates the number of underlying loci ($\Psi < 1$). The sensitivity of the estimator to the shape of the distribution of allelic effects depends strongly on

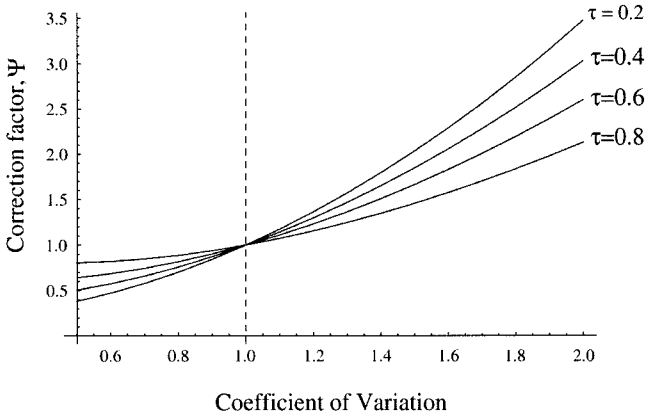

FIGURE 3.—The correction factor, Ψ, as a function of the coefficient of variation of the gamma distribution, *C*. The correction factor when multiplied by (6) gives the estimated number of underlying loci when the distribution of allelic effects follows a gamma distribution. It depends only on *C* and on $\tau = (M - \theta)/M$, which varies from zero for an experiment with little power to detect a QTL to one for an experiment with high power.

the power of the experiment (τ). When the threshold of detection is high and τ is near 0, the estimator is very sensitive to the shape of the distribution. Conversely, when the threshold of detection is low and τ is near 1, the correction term approaches 1, because most alleles are detectable regardless of the shape of the distribution.

Again, $a_{\min}$ can be used as an upwardly biased estimator for θ. Alternatively, one can correct θ using a Bayesian approach that takes into account the probability that $a_{\min}$ is the smallest detectable QTL when effect sizes follow a gamma distribution. This leads to an expression involving θ that must be numerically evaluated simultaneously with (14):

$$\hat{\theta} \approx \frac{\int_0^{a_{\min}} x\Gamma[1/C^2,\ x\Psi/C^2(M - \hat{\theta})]^{-n_d}dx}{\int_0^{a_{\min}} \Gamma[1/C^2,\ x\Psi/C^2(M - \hat{\theta})]^{-n_d}dx}. \quad (15)$$

When $C = 1$, (15) reduces to (11).

SIMULATIONS

To assess the accuracy of our estimator and the Castle-Wright-Zeng estimator, QTL analyses were simulated using the QTLcartographer package (version 1.13a; BASTEN *et al.* 1996). Using the Rmap program, we created a map with 20 chromosomes, each carrying five markers spaced 10 cM apart, with the first and last markers placed at the ends of the chromosome. This map was used for all simulations. On this map, we randomly placed 2, 5, 10, 20, or 100 QTL using Rqtl. For a given number of QTL, 30 different QTL maps were generated (*i.e.*, we constructed 30 different "experimental genomes"). In the basic set of simulations, the following assumptions were made: (1) alleles were additive; (2) all allelic effects, *a*, had the same sign (*i.e.*, each QTL acted in the direction of the observed parental difference); and (3) the QTL effects were drawn from an exponential distribution with a mean of 1/4. These assumptions were then relaxed in turn. To assess the effects of dominance, we used Rqtl to produce QTL with additive effects drawn from an exponential distribution with mean 1/4 and dominance effects drawn from a Beta distribution with shape parameter set to 1 (see BASTEN *et al.* 1996). This routine produced a dominance term that ranged from $-a$ to *a*, with mean equal to 0. That is, alleles from P_1 ranged from recessive to dominant. To assess the effects of having some QTL act in the opposite direction to the observed parental difference, we let the additive effect of a QTL be positive with probability 0.85 and negative with probability 0.15. Finally, to assess the effect of the underlying distribution of effect sizes, we drew effect sizes from a gamma distribution with a coefficient of variation of 0.5 or 2.0 instead of an exponential distribution (see Figure 1). For these extensions, the number of underlying QTL was set to 20, but otherwise the experimental design was the same as for the basic simulations. In all cases, environmental variation was assumed to be negligible.

For each experimental genome generated, Rcross was used to simulate the production of 200 and then 500 F_2 individuals. This F_2 sampling procedure was repeated 10 times. This design allowed us to assess whether the estimators were more sensitive to the realized distribution of QTL within the genome (variation among "experimental genomes") or to the specific set of F_2 individuals analyzed (variation among sampled individuals). For each set of F_2 individuals, the QTL were mapped using Zmapqtl's interval mapping routine (model 3). A single permutation test was used to determine an approximate significance threshold for each number of QTL and F_2 sample size (5% genome-wide significance level). In general, there was not much variation in significance thresholds (data not shown). This threshold was then used in conjunction with Eqtl to estimate the location and effects of the QTL (BASTEN *et al.* 1996). Eqtl identifies putative QTL by scanning through the output of Zmapqtl and noting all intervals with a likelihood peak that exceeds the significance threshold.

Unfortunately, interval mapping methods systematically bias the effects of chromosome regions linked to QTL (LYNCH and WALSH 1998, pp. 457–458). Put simply, an interval not containing a QTL but adjacent to one containing a QTL may exceed the significance threshold because of its linkage to the QTL. This problem is endemic to all QTL analyses and no adequate solution has yet been developed (WHITTAKER *et al.* 1996; GOFFINET and MANGIN 1998). To reduce this problem, we screened our data set for obvious "ghost" QTL. That is, we eliminated intervals containing putative QTL whose position was within another QTL's approximate

support interval (the chromosome interval within which the likelihood of a QTL was within a factor of 10 of the peak likelihood for that QTL; see LYNCH and WALSH 1998, pp. 448–449). In each case, the QTL with the weaker support was eliminated. VISSCHER *et al.* (1996) have shown that a bootstrap approach is much more accurate, but it would have been too time consuming for our analysis.

We chose interval mapping for two reasons. First, interval mapping is an established method that has been repeatedly used in empirical studies and is well explored theoretically. Second, interval mapping is much faster, allowing us to perform more extensive tests. We also suggest that interval mapping provides a conservative test of $\hat{n}_{QTL}$. Interval mapping produces data that are less precise than more advanced methods of QTL analysis, such as composite interval mapping and multiple interval mapping (ZENG 1994; KAO *et al.* 1999); this reduced precision should result in less accurate estimates for $\hat{n}_{QTL}$. Furthermore, interval mapping corresponds less well to our model, which assumes that QTL above the threshold of detection are always detected. Consequently, it seems reasonable that our estimator would perform even better with data from more sophisticated methods for detecting QTL.

Our simulation data were imported into *Mathematica* 3.0 (WOLFRAM 1996; package available at www.zoology.ubc.ca/~otto/Research) for analysis. In our calculations, we used the estimators appropriate for an exponential distribution of underlying QTL effect sizes. Specifically, we used Equation 6 for the QTL-based estimator and Equation 2 for the Castle-Wright-Zeng estimator, with $C = 1$. For $\hat{n}_{CWZ}$, we initially used the average recombination rate between randomly chosen pairs of loci, $\bar{c}$, which equals 0.48 for a genome with 20 chromosomes, each of length 40 cM (LYNCH and WALSH 1998, Equation 9.3). Unfortunately, this often led to nonsensical answers, especially when the number of underlying loci was large (>5). The problem lies in the fact that the denominator in (2) is subject to substantial sampling error; relatively often, the denominator approached zero (causing severe overestimates) or became negative. With 20 underlying QTL, the $\hat{n}_{CWZ}$ estimate for the number of underlying loci *averaged* 49.3 (with 500 F_2's) and −21.8 (with 200 F_2's) over the 300 simulations! To avoid these problems, we set $\bar{c}$ to ½ in all of our analyses, which either made little difference (for small n) or improved the estimates.

Because the QTL estimator relies on the difference between the mean effect of factors found and the minimum effect found, it can only work if at least two QTL have been identified. Therefore, we excluded QTL analyses with only one detected QTL. In experimental genomes with more than two true QTL, very few analyses were excluded. With only two true QTL, however, approximately half of the analyses had to be excluded. Furthermore, in a very small number of cases with two true QTL (6/600), two QTL were detected whose estimated effects were, by chance, equal. To avoid division-by-zero errors, the mean was increased by 5% over the minimum in these cases. (If, instead, these cases were eliminated, the performance of the estimators improved slightly over the results shown in Table 1.)

RESULTS

The results of the simulations are presented in Table 1. As expected, our QTL-based estimator was more accurate when (11) was used to estimate the threshold of detection (θ) than when the smallest detected QTL was used (a_{min}). We therefore focus our discussion on estimates based on (11). Both our estimator and the Castle-Wright-Zeng estimator were fairly accurate on average when there was an intermediate number of underlying QTL ($5 \leq n \leq 20$). With 100 QTL, however, both methods underestimated the true number of QTL but for different reasons. The QTL-based estimator will be biased downward whenever the density of QTL is high, because tightly linked QTL are then rarely separated by recombination. Furthermore, in interval mapping, the number of detected QTL must be less than the number of marker intervals, which was only 80 in our study. This bias could be eliminated by following the lines for more generations (increasing the opportunity for recombination) and by adding more markers to the study (increasing the number of marker intervals). The Castle-Wright-Zeng estimator, on the other hand, becomes less and less accurate as the number of loci increases because the segregation variance approaches zero and becomes harder to estimate precisely. Although increasing the number of F_2 progeny tested and following the lines for more generations may improve the accuracy of the Castle-Wright-Zeng estimator, our simulations indicate no substantial improvement between $N_{F_2} = 200$ and $N_{F_2} = 500$.

With only two true QTL, our estimator performed poorly, but the Castle-Wright-Zeng estimator continued to perform well, on average. Because we often had to exclude cases where only one QTL was detected when there were only two true QTL, it is not surprising that the estimators overestimated the number of underlying QTL. Note that if we also excluded cases where only two QTL were detected, the average of our estimator improved (Table 1, last two rows), which suggests that $\hat{n}_{QTL}$ is biased upward by sampling error when there are few detected QTL. More generally, Table 1 suggests that, unless the number of QTL is very large, the QTL-based estimator tends to overestimate the true number of underlying loci and that this bias is stronger with fewer QTL. We expect an upward bias in our estimator for two reasons: (1) when there are few detected QTL, there will be substantial variation in the denominator ($M - \theta$) of Equation 6, which can approach zero and

TABLE 1

Estimated number of underlying QTL assuming an exponential distribution of effect sizes without dominance

n	N_{F_2}	n_d	$\hat{n}_{QTL}$ ($\theta = a_{min}$)	$\hat{n}_{QTL}$ [θ from (11)]	$\hat{n}_{CWZ}$ ($C = 1$)
100	500	28.84 {21, 38}	46.86 {29.14, 69.87} [1.3%] $\bar{\tau} = 0.37$	45.21 {28.06, 67.63} [0.7%] $\bar{\tau} = 0.39$	38.40 {32.75, 45.37} [0%]
100	200	12.00 {6, 19}	57.72 {25.13, 117.19} [35.7%] $\bar{\tau} = 0.25$	52.51 {23.65, 104.17} [30.7%] $\bar{\tau} = 0.27$	37.43 {30.09, 45.77} [0%]
20	500	9.65 {5, 14}	23.42 {11.85, 42.57} [94.0%] $\bar{\tau} = 0.44$	20.98 {10.39, 39.35} [95.0%] $\bar{\tau} = 0.49$	19.51 {7.86, 43.90} [41.0%]
20	200	5.65 {3, 9}	25.78 {10.98, 58.69} [93.6%] $\bar{\tau} = 0.32$	20.98 {9.02, 48.22} [95.7%] $\bar{\tau} = 0.39$	18.10 {10.78, 53.45} [42.3%]
10	500	6.84 {3, 11}	13.99 {5.80, 25.95} [93.3%] $\bar{\tau} = 0.47$	11.93 {4.80, 21.27} [97.3%] $\bar{\tau} = 0.55$	5.34 {−0.12, 12.61} [17.7%]
10	200	4.94 {2, 8}	17.71 {6.24, 38.67} [87.7%] $\bar{\tau} = 0.34$	13.88 {4.74, 32.32} [94.7%] $\bar{\tau} = 0.43$	9.63 {6.12, 14.14} [64.0%]
5	500	4.19 {2, 8}	8.84 {3.88, 25.89} [90.0%] $\bar{\tau} = 0.49$	6.84 {3.11, 15.62} [96.0%] $\bar{\tau} = 0.61$	5.29 {2.77, 8.81} [48.7%]
5	200	3.85 {2, 7}	12.00 {5.12, 29.66} [84.0%] $\bar{\tau} = 0.33$	8.68 {3.99, 19.32} [91.7%] $\bar{\tau} = 0.44$	5.57 {3.17, 8.05} [72%]
2	500	2.59 {2, 4}	19.38 {3.25, 130} [63.0%] $\bar{\tau} = 0.32$	10.65 {2.59, 65} [81.3%] $\bar{\tau} = 0.45$	2.66 {1.80, 3.25} [24.7%]
2	200	2.74 {2, 4}	25.05 {3.98, 135.2} [61.7%] $\bar{\tau} = 0.30$	13.64 {3.06, 67.58} [81.7%] $\bar{\tau} = 0.43$	2.78 {2.01, 3.87} [47.7%]
2	500	Cases (147) with $n_d > 2$	7.29 {3.25, 13.9} [66.0%] $\bar{\tau} = 0.40$	5.21 {2.84, 9.27} [91.8%] $\bar{\tau} = 0.53$	
2	200	Cases (177) with $n_d > 2$	7.87 {3.97, 22.30} [61.0%] $\bar{\tau} = 0.34$	5.53 {3.07, 14.9} [87.6%] $\bar{\tau} = 0.47$	

For each combination of n (the number of underlying loci) and N_{F_2} (the number of F_2 individuals scored), 30 different QTL maps were generated, where each map was used in 10 replicate QTL experiments. Each row reports the mean estimate of n over the 300 resulting analyses. The 2.5–97.5% range of estimates is given in braces. In brackets is the percentage of cases where the 95% confidence limits for n included the true value of n. The average value of the power of the QTL studies, $\tau = (M - \theta)/M$, is also reported.

generate an overestimate, and (2) the Beavis effect will generate a greater upward bias in θ than in M, which also leads (6) to overestimate the number of underlying loci. We therefore recommend the use of the QTL-based estimator only when three or more QTL have been detected and when the mean detected QTL is substantially above (say >25% above) the minimum threshold of detection.

Interestingly, the largest difference between the two estimators is not in their average performance but in their confidence limits. Appropriate 95% confidence limits should include the true value 95% of the time. The confidence limits based on (9) for our estimator have this property (see numbers in square brackets in Table 1). In those cases where our estimator had little bias ($5 < n < 20$), the confidence limits included the true value 95.1% of the time. On the other hand, the confidence limits for the Castle-Wright-Zeng estimator included the true value only 47.6% of the time. The confidence limits for $\hat{n}_{CWZ}$ often excluded the true number of underlying loci because these limits only account for error caused by sampling a limited number of F_2 individuals. They do not account for the sampling error inherent in having a limited number of QTL, whose effects represent particular draws from an underlying distribution. For example, if the QTL with the largest effect has, by chance, a magnitude that is greater than expected, there will be more segregation variance than expected, and $\hat{n}_{CWZ}$ will underestimate n. Conversely, if the major QTL have, by chance, roughly equal influence, there will be less segregation variance than expected, and $\hat{n}_{CWZ}$ will overestimate n.

In fact, the sampling of different sets of QTL in different experimental genomes accounts for a large fraction of the total variance in estimates of n (Figure 4). This is especially true for $\hat{n}_{CWZ}$, where almost all of the observed variation was among genomes with different sets of QTL (dark gray bars) rather than among the different sets of F_2 individuals sampled from each experimental genome (light gray bars). In other words, $\hat{n}_{CWZ}$ depended little on the exact set of F_2 individuals, but it varied greatly each time a new set of QTL was generated. Figure 4 suggests that, with at least 200 F_2 individuals, the confidence limits for the Castle-Wright-Zeng estimator are based on a minor source of error (F_2 sampling variance) rather than the bulk of the error (QTL sampling vari-

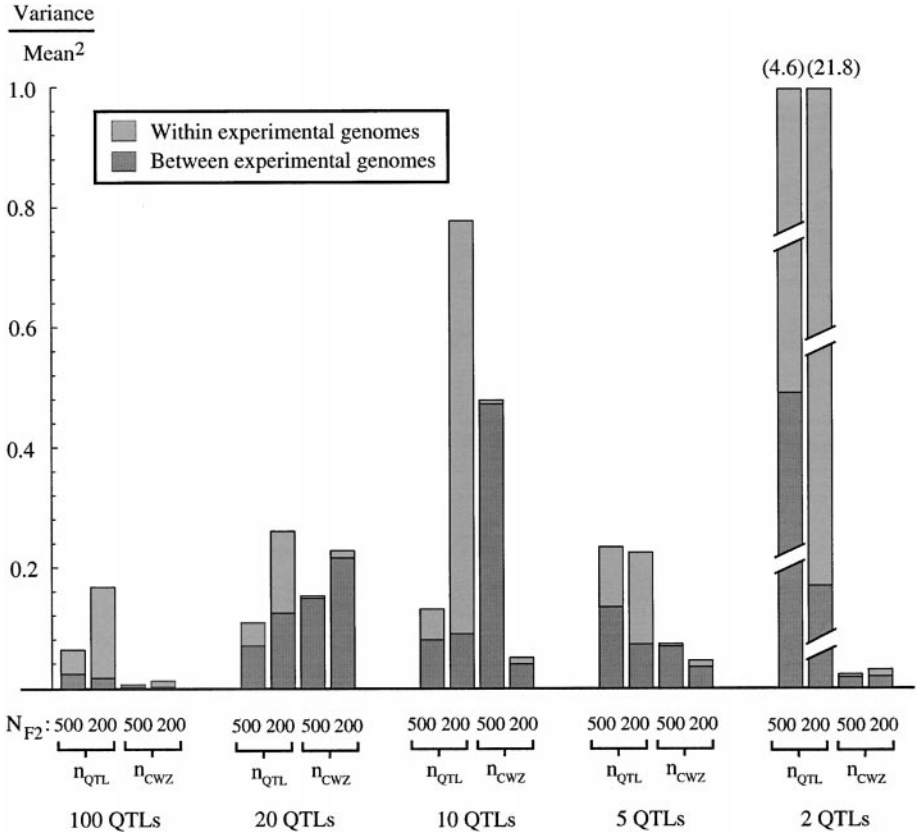

FIGURE 4.—The square of the coefficient of variation for the QTL-based estimator and the Castle-Wright-Zeng estimator. For each combination of n (number of underlying loci) and N_{F_2} (number of F_2 individuals scored), the total variance among the 300 QTL analyses is illustrated, scaled by the square of the mean estimate ($\hat{n}_{QTL}$ or $\hat{n}_{CWZ}$, as appropriate). The square root of the total height of each bar therefore gives the coefficient of variation for each estimator. The dark gray bars (bottom) indicate the proportion of variation observed among experimental genomes (*i.e.*, sets of QTL), while the light gray bars (top) indicate the proportion of variation observed within experimental genomes (*i.e.*, among sets of F_2 progeny). An exponential distribution of effect sizes was assumed in both the simulations and the analyses, and Equation 11 was used to estimate θ. The bars for $\hat{n}_{QTL}$ with 2 QTL were scaled to 1; their true heights are 4.6 (with 500 F_2's) and 21.8 (with 200 F_2's).

ance). In practice, this means that if a researcher is interested in the number of underlying loci that are responsible for a trait difference between two parental lines, the Castle-Wright-Zeng estimator will too often indicate a high degree of confidence in the wrong number and exclude the right number. Furthermore, as indicated by the simulations, reestimating $\hat{n}_{CWZ}$ using a different set of F_2 individuals is unlikely to help because the sampling of F_2's was not the major source of error (Figure 4).

Dominance: In the above results, the simulated QTL had additive effects. To test the impact of dominance on our estimator, we simulated experimental genomes that included 20 QTL with nonadditive effects ranging from fully recessive to fully dominant. Although models have been developed to incorporate dominance into the Castle-Wright estimator, these generally assume that the mean and distribution of dominance coefficients are known. Because most QTL studies lack such information, we continue to use (2) and (6) to estimate the number of underlying loci. That is, we ask, how inaccurate are the two estimators if we assume no dominance when dominance is actually present? The inclusion of dominance did not noticeably affect the performance of the QTL-based estimator, but it caused $\hat{n}_{CWZ}$ to underestimate n by ~20% (Table 2). Dominance tends to inflate the segregation variance inferred from F_2 individuals, because heterozygotes at a locus have genotypic values further from the mean. This effect is even more exaggerated with overdominance or underdominance, in which case the phenotype of the F_2's may lie outside of the range defined by the two parental

TABLE 2

Estimated number of underlying QTL assuming an exponential distribution of effect sizes with dominance

n	N_{F_2}	n_d	$\hat{n}_{QTL}$ ($\theta = a_{min}$)	$\hat{n}_{QTL}$ [θ from (11)]	$\hat{n}_{CWZ}$ ($C = 1$)
20	500	8.87 {3, 14}	21.37 {8.55, 37.35} [94.7%] $\bar{\tau} = 0.46$	18.69 {6.54, 34.24} [91.0%] $\bar{\tau} = 0.52$	15.42 {5.89, 21.75} [21.3%]
20	200	5.78 {3, 9}	27.14 {12.27, 63.41} [93.3%] $\bar{\tau} = 0.32$	22.22 {9.32, 51.23} [96.3%] $\bar{\tau} = 0.39$	16.19 {10.55, 23.19} [42.7%]

Dominance effects for each QTL were drawn at random from purely recessive to purely dominant as described in the text. See Table 1 for more details.

TABLE 3

Estimated number of underlying QTL when there was an 85% chance that the additive effect (a) was positive and a 15% chance that it was negative

n	N_{F_2}	n_d	$\hat{n}_{QTL}$ ($\theta = a_{min}$)	$\hat{n}_{QTL}$ [θ from (11)]	$\hat{n}_{CWZ}$ ($C = 1$)
20	500	4.70 {2, 7}	20.72 {3.70, 47.31} [92.0%] $\bar{\tau} = 0.29$	15.89 {2.63, 38.08} [84.7%] $\bar{\tau} = 0.38$	9.37 {−0.50, 18.17} [3.0%]
20	200	4.41 {2, 7}	21.86 {8.38, 52.81} [93.3%] $\bar{\tau} = 0.29$	16.35 {6.11, 44.01} [81.7%] $\bar{\tau} = 0.38$	8.85 {1.65, 18.09} [6.0%]

See Table 1 for more details.

lines (transgressive segregation). This explains the sensitivity of $\hat{n}_{CWZ}$ to the inclusion of dominance but does not address why $\hat{n}_{QTL}$ was little affected by dominance. We believe that, because the methods used to detect QTL explicitly allow dominance levels to vary among loci, the estimated additive effects of the detected QTL (and hence $\hat{n}_{CWZ}$) are not strongly biased by dominance. Interestingly, the average power of the QTL experiments was also little affected by dominance (compare $\bar{\tau}$ values in Tables 1 and 2). Because $\hat{n}_{QTL}$ is less sensitive to dominance interactions, it is a more appropriate estimator to use than $\hat{n}_{CWZ}$ whenever the nature of dominance is unknown.

QTL of opposite effect: Table 3 presents results for experiments where the additive effect of a QTL had an 85% chance of being positive and a 15% chance of being negative. Although both estimators underestimated the true number of underlying QTL, $\hat{n}_{QTL}$ was much less sensitive than $\hat{n}_{CWZ}$ to the inclusion of loci affecting the trait of interest in the opposite direction. On average, ~16 underlying QTL were estimated with $\hat{n}_{QTL}$, while ~9 were estimated with $\hat{n}_{CWZ}$, whereas the true number was 20. When the effects of QTL oppose one another, the trait values of the F_2's are no longer expected to lie strictly between the parental lines, providing another explanation for transgressive segregation. As with dominance, the segregation variance measured among the F_2's is thus inflated, and the Castle-Wright-Zeng estimator underestimates the number of underlying loci. On the other hand, the inclusion of QTL with opposite effects reduced the power of the QTL experiments (compare $\bar{\tau}$ values in Tables 1 and 3), but the additive effects of the detected QTL (and hence $\hat{n}_{QTL}$) were not strongly biased.

Gamma distribution of effect sizes: Finally, we tested the extent to which the estimators were sensitive to the underlying distribution of effect sizes. Table 4 provides results from simulations where the underlying distribution was gamma (with $C = 0.5$ or 2.0) but where the analyses assumed an exponential distribution ($C = 1.0$). As expected, $\hat{n}_{QTL}$ and $\hat{n}_{CWZ}$ overestimated the true number of QTL when the underlying distribution had a lower coefficient of variation ($C = 0.5$). The extent of the bias was not severe for $\hat{n}_{CWZ}$ in this case; this is consistent with the form of Equation 2, which changes less when C is lowered by a fraction than when it is increased. Conversely, both $\hat{n}_{QTL}$ and $\hat{n}_{CWZ}$ underestimated the true number of QTL by ~50% when the underlying distribution had a higher coefficient of variation ($C = 2.0$), with $\hat{n}_{CWZ}$ performing slightly worse than $\hat{n}_{QTL}$. One can correct these estimators using Equation 2 for $\hat{n}_{CWZ}$, and 14 and 15 for $\hat{n}_{QTL}$. These corrections brought the esti-

TABLE 4

Estimated number of underlying QTL when the distribution of additive effects was not exponential

n	N_{F_2}	n_d	$\hat{n}_{QTL}$ ($\theta = a_{min}$)	$\hat{n}_{QTL}$ [θ from (11)]	$\hat{n}_{CWZ}$ ($C = 1$)
			$C = 0.5$		
20	500	10.32 {7, 14}	34.86 {21.16, 50.67} [70.0%] $\bar{\tau} = 0.38$	31.44 {18.60, 46.53} [85.0%] $\bar{\tau} = 0.42$	26.31 {21.43, 31.28} [9.0%]
20	200	5.80 {3, 8}	49.04 {20.20, 105.53} [66.0%] $\bar{\tau} = 0.24$	39.94 {15.99, 85.23} [79.3%] $\bar{\tau} = 0.3$	26.25 {18.42, 34.11} [37.0%]
			$C = 2.0$		
20	500	5.16 {2, 8}	13.48 {5.50, 29.41} [85.0%] $\bar{\tau} = 0.49$	10.97 {4.31, 23.78} [61.3%] $\bar{\tau} = 0.60$	10.44 {4.63, 17.27} [1.7%]
20	200	5.59 {2, 10}	17.14 {5.47, 41.57} [74.3%] $\bar{\tau} = 0.35$	13.04 {4.55, 30.04} [62.3%] $\bar{\tau} = 0.43$	8.77 {5.41, 14.93} [0.3%]

Simulations were run where QTL effects were drawn from a gamma distribution with a coefficient of variation of either 0.5 or 2.0 (see Figure 1). In this table, the analyses assume (incorrectly) that the underlying distribution of QTL effects was exponential ($C = 1$). See Table 1 for more details.

mates toward the true value of 20 but tended to overcorrect, especially for $\hat{n}_{QTL}$ when $C = 2.0$, perhaps because the observed coefficient of variation of the true effects of the QTL was highly variable in this case. Of course, making these corrections requires external knowledge about the underlying distribution, which will often be lacking. When many QTL have been detected, the number of underlying loci, the threshold of detection, and the shape of the distribution could be simultaneously estimated using a maximum-likelihood approach. One potential problem is that the Beavis effect will bias the shape parameter (lowering C) by making QTL of small effect seem larger.

CONCLUSIONS

Historically, the number of genetic factors, n, underlying an observed difference between two parental lines has been estimated using methods developed by Castle (1921) and Wright (1968). QTL analyses have, to some extent, supplanted the Castle-Wright estimator. QTL analyses localize the genetic factors and estimate their additive and dominance effects without assuming that these effects follow a particular distribution. Unfortunately, QTL analyses are best at finding factors with profound phenotypic effects and often miss factors of moderate to small effect. As a result, the number of observed QTL is a poor indicator of the number of loci contributing to a difference between two parental lines.

This problem is illustrated in Figure 5, which shows the expected number of detectable QTL as a function of the number of underlying QTL. Figure 5 is based on the assumption that there is a threshold below which a QTL is unlikely to be detected and above which it is. Two threshold levels of detection are illustrated, with θ set to 5 or 10% of D, where D is the total additive effect size (*i.e.*, half the parental difference). The first threshold was typical in our simulation studies with 500 F_2's, while the second was typical in simulations with 200 F_2's. The most striking feature of these curves is that they do not increase monotonically with the number of underlying loci. Instead, the expected number of detected loci initially rises, reaches a maximum, and then falls back toward zero as the number of underlying loci increases. These curves suggest two reasons why a certain number of QTL may be observed: (1) there are few underlying QTL, but their average effect is relatively large such that most are above the threshold, or (2) there are several QTL, but their average effect is relatively small such that few are above the threshold of detection. Furthermore, the maxima of these curves is fairly low, indicating that only a handful of QTL will be detected regardless of the true number of loci contributing to the trait difference. These conclusions are the same whether the effects of the underlying loci follow an exponential distribution, an L-shaped gamma distribution, or a bell-shaped gamma distribution. In short,

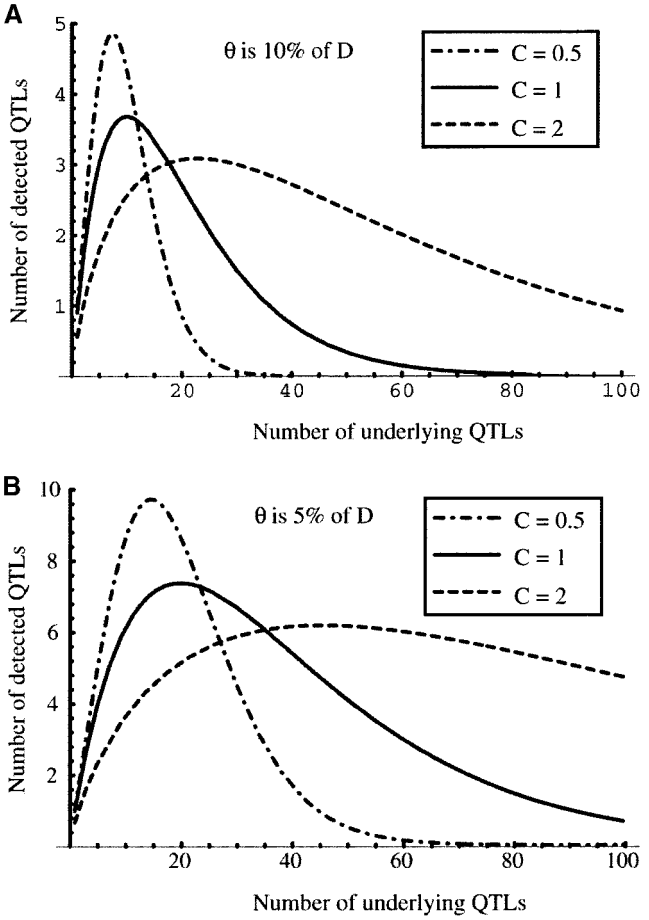

Figure 5.—The expected number of detected loci as a function of the number of underlying loci. The expected number of detected loci is equal to n times the fraction of the probability density function, $g[x, \mu, C]$ given by (13), that lies above θ. It is plotted as a function of the number of underlying loci for a bell-shaped distribution ($C = 0.5$; dot-dashed curve), an exponential distribution ($C = 1$, $\Psi = 1$; solid curve), and an L-shaped distribution ($C = 2$; dashed curve). (A) $\theta = 10\%$ of D, as was typical in our studies with a large number of QTL and 200 F_2's. (B) $\theta = 5\%$ of D, as was typical in our studies with a large number of QTL and 500 F_2's.

because QTL studies predominantly detect loci of large effect, the number of loci detected in a QTL study is not linearly related to the number of underlying loci.

Here, we present a new estimator of gene number, $\hat{n}_{QTL}$, that takes into account the bias of QTL analyses toward detecting loci of large effect. By noting the average size and the minimum size of the detected QTL, we can estimate the number and magnitude of the loci whose effects were too small to be detected. As with the Castle-Wright estimator, this technique requires us to specify the expected distribution of effect sizes. We develop a QTL-based estimator for an exponential distribution and a gamma distribution of effect sizes. Although our method assumes that QTL analyses have a negligible probability of detecting a QTL below a

TABLE 5

Data analysis of stamen length and pistil length from Bradshaw *et al.* (1998)

	Stamen length (mm)	Pistil length (mm)
$D = (P_1 - P_2)/2$	9.0	10.2
n_d, number of detected QTL	5	5
M, mean QTL effect	2.2	2.7
a_{min}, minimum QTL effect	1.6	1.7
Var(S), segregation variance	16.0	21.1
N_{F_2}, number of F_2's	465	465
$\hat{n}_{QTL}$ [θ from (11)]	12.0 {4.3, 25.8}	8.2 {2.9, 17.6}
$\hat{n}_{CWZ}$ ($C = 1$, $\bar{c} = 0.413$)	6.5 {4.8, 8.2}	6.1 {4.5, 7.7}
$\hat{n}_{CWZ}$ ($C = 1$, $\bar{c} = 1/2$)	4.1 {3.4, 4.7}	3.9 {3.3, 4.6}

An exponential distribution of effect sizes was assumed. The average rate of recombination among pairs of loci (0.413) was estimated from Equation 9.3 of Lynch and Walsh (1998) using data from Figure 3 of Bradshaw *et al.* (1998).

threshold (θ) and a 100% probability of detecting QTL above θ, simulations indicate that this simplifying assumption does not generate a substantial bias in the average number of estimated loci ($\hat{n}_{QTL}$).

As Table 1 shows, our QTL-based estimator provides a good approximation for the number of underlying loci unless few QTL were detected ($n_d < 3$) or the genetic map was saturated with QTL (more QTL than marker intervals; $n_d > 80$). Furthermore, in those cases where the average value of the estimator approximately equals the true number of underlying loci (*i.e.*, when $5 \leq n \leq 20$), the 95% confidence limits based on $\hat{n}_{QTL}$ contain $n \sim 95\%$ of the time (Table 1). In contrast, the 95% confidence limits for the Castle-Wright-Zeng estimator often miss the true value for n, despite the fact that the simulations and the estimator both assume an exponential distribution of effect sizes. Essentially, the confidence limits for the Castle-Wright-Zeng estimator do not account for the variance inherent in the sampling of mutations that arise and fix within a population. Because the confidence limits for $\hat{n}_{QTL}$ take into account the sampling error inherent in drawing allelic effects from an underlying distribution, they more often include the true value for the number of underlying loci. An additional benefit of our estimator is that it is less sensitive to dominance (compare Tables 1 and 2) and to violations of the assumption that the additive effects of all alleles have the same sign (compare Tables 1 and 3).

To demonstrate the application of our estimator to real data, we consider two examples. The first, a QTL study by Bradshaw *et al.* (1998), examined floral morphology in monkeyflowers (Table 5). We focus on two of their morphological characters, stamen and pistil length. In both cases, 5 QTL were detected, each acting in the same direction. Assuming an exponential distribution, our QTL-based estimator predicted that 12.0 loci underlie the stamen length differences and that 8.2 loci underlie pistil length differences, suggesting that about half of the QTL have been detected. Using 1/2 as the average rate of recombination among pairs of loci, the Castle-Wright-Zeng estimator inferred 4.1 factors affecting stamen length and 3.9 factors affecting pistil length, both of which fall short of the observed number of QTL. Using the average recombination rate based on the genome map of monkeyflowers (0.413) increased these estimates only slightly (to 6.5 and 6.1, respectively). The Castle-Wright-Zeng estimators are particularly suspect for these traits, however, because both exhibit transgressive variation, where the F_2 distribution is broader than the parental difference. The presence of transgressive variation suggests that there are either strong interactions among alleles (*e.g.*, overdominance) or QTL of opposite effects. In either case, the Castle-Wright-Zeng estimator underestimates the number of underlying loci and is less reliable than the QTL-based estimator (see Tables 2 and 3).

The above example highlights the difference between $\hat{n}_{QTL}$ and $\hat{n}_{CWZ}$. Our next example demonstrates that our estimator could be used to predict the number of loci that may be uncovered in a more powerful QTL study. Liu *et al.* (1996) mapped factors affecting genital arch shape in *Drosophila mauritiana* and *D. simulans* hybrids, employing two backcrosses and <200 individuals. Zeng *et al.* (2000) extended this analysis by increasing the sample size (~500 individuals per backcross) and reanalyzing the old and new data sets using multiple interval mapping. We concentrate on the results presented by Zeng *et al.*, as they were obtained by applying the same mapping methodology and criteria to both data sets.

Liu *et al.*'s backcross analysis suggested that 11–13 QTL are involved in the genitalia difference. [The first backcross to *D. mauritiana* (BM) identified 11 QTL; the backcross to *D. simulans* (BS) identified 13 QTL.] Based on Equation 11, our estimator suggests that the true number is closer to 21.4 for BM and 23.2 for BS. The number of QTL found in the BS analysis is within our 95% confidence intervals (CIs), albeit just barely (95%

CI, 12.7–38.1). The number of QTL found in the BM analysis, however, is not (95% CI, 11.1–36.7). This predicts that a fair number of QTL were probably missed in these analyses, a fact that was confirmed by Zeng *et al.* (2000). After more than doubling the number of individuals and markers genotyped, Zeng *et al.* found 19 QTL involved in genitalia differences in both backcrosses. With this data, $\hat{n}_{QTL} = 16.7$ (95% CI, 10.2–25.3) for the BS lines and 20.0 (95% CI, 12.3–30.5) for the BM lines. The value of 19 QTL found by Zeng *et al.* is well within our 95% confidence intervals and is near the mean of the two estimates. This, plus the fact that the power of the second analysis is quite high ($\tau = 0.97$ for the BS analysis and $\tau = 0.92$ for the BM analysis), indicates that nearly all of the QTL have now been identified. Thus, our estimator accurately estimated the number of QTL that could be found in a powerful experiment given data from a less powerful experiment.

Improving estimates of gene number could have an important impact in both quantitative and evolutionary genetics. First, better estimates would help researchers know how many genes affecting a trait of interest go undetected. They could then make more informed decisions about whether to refine a QTL analysis to uncover missing factors. Second, better estimates of the actual number of genes underlying divergent traits can help us evaluate the applicability of quantitative genetic models, which assume a large number of underlying loci. Third, improved estimates of gene number provide interesting information about the genetic architecture underlying evolutionary change. For example, they can help us identify the sorts of phenotypic changes that are typically accomplished by few allelic substitutions. Although we have taken steps to incorporate data being generated in QTL studies, further work is warranted. In particular, it would be valuable to know how best to use the information contained in *both* the Castle-Wright-Zeng and the QTL-based estimators to obtain an even more powerful estimator of the number of loci underlying phenotypic divergence.

We owe special thanks to H. Allen Orr for encouraging this project along, and for generously sharing his time and ideas, and to John Huelsenbeck for kindly providing access to his computer facilities. We also thank Andrea Betancourt, Thomas Lenormand, J. P. Masly, Allen Orr, Art Poon, Daven Presgraves, Dolph Schluter, Peter Visscher, Bruce Walsh, Michael Whitlock, Zhao-Bang Zeng, and an anonymous reviewer for their helpful comments on the project and manuscript. This work was inspired by a discussion of the Vancouver Evolutionary Group and was sponsored by grants from the Natural Sciences and Engineering Research Council of Canada (S.P.O.), the Peter Wall Institute for Advanced Studies (S.P.O.), the David and Lucile Packard Foundation (H. A. Orr), the National Institutes of Health (GM51932 to H. A. Orr), and a Caspari Fellowship from the University of Rochester (C.D.J.).

LITERATURE CITED

Abromowitz, M., and I. A. Stegun, 1972 *Handbook of Mathematical Functions.* Dover, New York.

Basten, C. J., B. S. Weir and Z.-B. Zeng, 1996 *A Reference Manual and Tutorial for QTL Mapping.* Department of Statistics, North Carolina State University, Raleigh, NC.

Beavis, W. D., 1994 The power and deceit of QTL experiments: lessons from comparative QTL studies. Proceedings of the Corn and Sorghum Industry Research Conference, American Seed Trade Association, Washington DC, pp. 250–266.

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, Boca Raton, FL.

Bradshaw, H. D., Jr., K. G. Otto, B. E. Frewen, J. K. McKay and D. W. Schemske, 1998 Quantitative trait loci affecting differences in floral morphology between two species of monkeyflower (*Mimulus*). Genetics **149:** 367–382.

Castle, W. E., 1921 An improved method of estimating the number of genetic factors concerned in cases of blending inheritance. Proc. Natl. Acad. Sci. USA **81:** 6904–6907.

Churchill, G. A., and R. W. Doerge, 1994 Empirical threshold values for quantitative trait mapping. Genetics **138:** 963–971.

Doebley, J., and A. Stec, 1991 Genetic analysis of the morphological differences between maize and teosinte. Genetics **129:** 285–295.

Feller, W., 1971 *An Introduction to Probability Theory and Its Applications*, Vol. II. John Wiley, New York.

Fisher, R. A., 1958 *The Genetical Theory of Natural Selection*, Ed. 2. Dover, New York.

Gillespie, J. H., 1991 *The Causes of Molecular Evolution.* Oxford University Press, New York.

Goffinet, B., and B. Mangin, 1998 Comparing methods to detect more than one QTL on a chromosome. Theor. Appl. Genet. **96:** 628–633.

Kao, C. H., Z. B. Zeng and R. D. Teasdale, 1999 Multiple interval mapping for quantitative trait loci. Genetics **152:** 1203–1216.

Keightley, P. D., 1994 The distribution of mutation effects on viability in *Drosophila melanogaster.* Genetics **138:** 1315–1322.

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

Larsen, R. J., and M. L. Marx, 1985 *An Introduction to Probability and Its Applications.* Prentice-Hall, Englewood Cliffs, NJ.

Laurie, C., J. R. True, J. Liu and J. M. Mercer, 1997 An introgression analysis of quantitative trait loci that contribute to a morphological difference between *Drosophila simulans* and *D. mauritiana.* Genetics **145:** 339–348.

Leips, J., and T. F. C. Mackay, 2000 Quantitative trait loci for life span in *Drosophila melanogaster*: interactions with genetic background and larval density. Genetics **155:** 1773–1788.

Liu, J., J. M. Mercer, L. F. Stam, G. C. Gibson, Z. B. Zeng and C. C. Laurie, 1996 Genetic analysis of a morphological shape difference in the male genitalia of *Drosophila simulans* and *D. mauritiana.* Genetics **142:** 1129–1145.

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

Lynch, M., J. Blanchard, T. Kibota, S. Schultz, L. Vassilieva *et al.*, 1999 Perspective: spontaneous deleterious mutation. Evolution **53:** 645–663.

Mackay, T. F. C., 1996 The nature of quantitative genetic variation revisited: lessons from Drosophila bristles. Bioessays **18:** 113–121.

Orr, H. A., 1998 The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. Evolution **52:** 935–949.

Orr, H. A., 1999 The evolutionary genetics of adaptation: a simulation study. Genet. Res. **74:** 207–214.

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.

Visscher, P. M., R. Thompson and C. S. Haley, 1996 Confidence intervals in QTL mapping by bootstrapping. Genetics **143:** 1013–1020.

Whittaker, J. C., R. Thompson and P. M. Visscher, 1996 On the mapping of QTL by regression of phenotype on maker-type. Heredity **77:** 23–32.

Wolfram, S., 1996 *The Mathematica Book*, Ed. 3. Wolfram Media/Cambridge University Press, Cambridge, MA.

Wright, S., 1968 *Evolution and the Genetics of Populations.* University of Chicago Press, Chicago.

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.

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

Communicating editor: J. B. Walsh