## ORIGINAL PAPER

Joel Ira Weller · Morris Soller

# An analytical formula to estimate confidence interval of QTL location with a saturated genetic map as a function of experimental design

**Abstract** Analytical formulae are derived for the confidence interval for location of a quantitative trait locus (QTL) using a saturated genetic map, as a function of the experimental design, the QTL allele substitution effect, and the number of individuals genotyped and phenotyped. The formulae are derived assuming evenly spaced recombination events, rather than the actual unevenly spaced distribution. The formulae are useful for determining desired sample size when designing a wide variety of QTL mapping experiments, and for evaluating a priori the potential of a given mapping population for defining the location of a QTL. The formulae do not take into account the finite number of recombination events in a given sample.

## Introduction

Many studies have shown that individual quantitative trait loci (QTL) can be detected and mapped with the aid of genetic markers. With multiple linked genetic markers, the approximate map location of QTL for a given set of experimental data can be determined by single marker or interval mapping. Both analytical (e.g., Lander and Botstein 1989) and empirical methods (Visscher et al. 1996) have been proposed to obtain confidence intervals (CI) for estimated QTL location, based on a given set of experimental results. While it is possible to derive empirical formulas for various specific situations by

J. I. Weller (✉)
Institute of Animal Sciences, Agricultural Research
Organization, The Volcani Center,
P.O. Box 6 Bet Dagan, 50250, Israel
e-mail: weller@agri.huji.ac.il
Tel.: +972-8-9484430
Fax: +972-8-9470587

M. Soller
Department of Genetics, The Hebrew University of Jerusalem,
Jerusalem 91904, Israel

extensive simulation (e.g., Darvasi and Soller 1997; Darvasi et al. 1993; Ronin et al. 2003), it is clearly preferable to have a generally applicable simple formula for the CI of QTL map locations that enables the mapping potential of complex designs to be evaluated without the need for simulations. This is becoming particularly important with the availability of physical maps and complete genome sequences, since the width of the CI of a given QTL map location will determine the potential population of candidate genes for the QTL.

With many tightly linked markers, the limiting factors in locating a QTL are the number of recombination events in the sample and the magnitude of the QTL effect relative to the residual standard deviation (VanRaden and Weller 1994). This enables the map resolution attainable in a given experiment to be estimated by simply estimating the expected number of recombinants in a given interval. The objective of this study is to derive analytical formulas to predict the CI of any QTL map location within a saturated genetic map as a function of sample size, QTL effect, and experimental design.

## Theory

A saturated genetic map consisting of many evenly spaced completely informative markers is assumed. We will further assume that the number of recombination events in a finite sample of individuals is a continuous variable, even though it is in fact discrete. The consequences of these assumptions will be considered in the Discussion. QTL location can then be estimated by single marker mapping, involving a $t$-test at each marker, and the most likely QTL location will be at the marker with the greatest estimated effect. With single markers, Simpson (1989) proposed that linkage of a segregating QTL to a marker could be detected by a likelihood ratio test, with the null hypothesis that the recombination frequency between the QTL and genetic marker is 0.5. Simpson (1992) showed for the backcross (BC) design that the statistical power for this test is equal to that obtained by a $t$-test with the null

hypothesis of equal means for the two marker genotypes. It follows that for single markers, the $(1-\alpha)$ CI for QTL location, $\mathrm{CI}_{(1-\alpha)}$, can be determined from the $\mathrm{CI}_{(1-\alpha)}$ for the QTL effect. Consequently, with single marker mapping, given that the marker with the greatest estimated QTL effect is $M_1$, the CI for QTL location will include marker $M_2$ if the CI for the QTL effect at marker $M_1$ also includes the effect estimated at marker $M_2$. Thus, the CI for QTL location can be derived from the CI for the difference of the expected QTL effect for a marker at the QTL and the expected effect for a marker at some other chromosomal location. Clearly, this difference will be due solely to those individuals that are recombinant in the interval between the two markers. Therefore, given that the marker with the greatest estimated QTL effect is $M_1$, the CI for QTL location will include marker $M_2$ if the CI for QTL effect at marker $M_1$, considering recombinant individuals only, also includes the effect estimated at marker $M_2$.

Assuming a normal distribution of the estimated marker-associated effects and considering recombinant individuals only, the probability that the QTL effect at marker $M_1$ also includes the effect estimated at marker $M_2$ is equal to the probability, $\alpha/2$, of obtaining the value:

$$Z_{\alpha/2} = D/\mathrm{SE}(D) \qquad (1)$$

where $Z_{\alpha/2}$ is the value of the standard normal variable corresponding to a probability of $\alpha/2$. The "contrast", $D= E(M_1)-E(M_2)$, where $E(M_1)$ is the expected QTL effect evaluated at $M_1$, and $E(M_2)$ is the expected QTL effect evaluated at $M_2$; $\mathrm{SE}(D)$ is the standard error of $D$.

$D$ and $\mathrm{SE}(D)$ are functions of the experimental design. Their derivation is now exemplified for a BC design initiated by a cross between two parental lines. The two QTL alleles are denoted Q and q, the QTL is assumed to be located at marker $M_1$, and the parental genotypes are denoted $M_1QM_2/M_1QM_2$ and $m_1qm_2/m_1qm_2$. Relative to the genetic markers, there are two recombinant genotypes in the $BC_1$ generation: $M_1m_2/m_1m_2$ and $m_1M_2/m_1m_2$, with expected mean values denoted $\underline{M_1}/\underline{m_2}$ and $\underline{m_1}\underline{M_2}$. $E(M_1) = \underline{M_1}\underline{m_2} - \underline{m_1}\underline{M_2}$ and $E(M_2) = \underline{m_1}\underline{M_2} - \underline{M_1}\underline{m_2}$, giving:

$$D = E(M_1) - E(M_2) = 2\underline{M_1}\underline{m_2} - 2\underline{m_1}\underline{M_2}$$
$$= 2(\underline{M_1}\underline{m_2} - \underline{m_1}\underline{M_2}) \qquad (2)$$

Letting the phenotypic variance within the marker genotypes equal 1.0, standardized effects at the QTL are: $QQ=+d$, $Qq=h$, and $qq=-d$. Defining $E(M_1)=\delta=d+h$, and $R=$the number of individuals carrying a recombinant chromosome in each marker genotype group, we have:

$$D = 2(d + h) = 2\delta \qquad (3)$$

$$\mathrm{Var}(\underline{M_1}\underline{m_2}) = \mathrm{Var}(\underline{m_1}\underline{M_2}) = 1/R \qquad (4)$$

To derive $\mathrm{SE}(D)$, recall that when $X$ and $Y$ are independent, $\mathrm{Var}[b(X-Y)]=b^2(\mathrm{Var}X+\mathrm{Var}Y)$. Applying this to (2) yields

$$\mathrm{SE}^2(D) = 4[\mathrm{Var}(\underline{M_1}\underline{m_2}) + \mathrm{Var}(\underline{m_1}\underline{M_2})] = 8/R \qquad (5)$$

Substituting (3) and (5) in (1), gives

$$Z_{\alpha/2} = 2\delta/(8/R)^{0.5} \qquad (6)$$

Defining $k$ as the proportion of the mapping population included in each marker genotype group, $r$ as the proportion of recombination between $M_1$ and $M_2$, and $N$ as the population size, we have: $R=rkN$. For the BC design, $k=0.5$. Substituting $rkN$ for $R$ in (6) gives:

$$Z_{\alpha/2} = 2\delta/(8/rkN)^{0.5} = \delta/(2/rkN)^{0.5} \qquad (7)$$

Note that the interval between markers $M_1$ and $M_2$ defines half of the $\mathrm{CI}_{(1-\alpha)}$. Assuming a chromosome of infinite length, the CI will be symmetrical, so that $r= \mathrm{CI}_{(1-\alpha)}/2$, with the CI of QTL map location in units of proportion of recombination. Generally, the CI of map location is given in cM. To convert cM to proportion of recombination, cM are first converted to percent recombination using an appropriate mapping function, such as the Haldane mapping function, and then to proportion of recombination by dividing by 100. Thus, $r=\mathrm{CI}^*_{(1-\alpha)}/200$, where $\mathrm{CI}^*_{(1-\alpha)}$ is the CI expressed as percent recombination. Substituting for $r$ for $2R/N$ gives:

$$R = \mathrm{CI}^*_{(1-\alpha)}N/400 \qquad (8)$$

Substituting (8) in (7), gives $Z_{\alpha/2}=2\delta/(3{,}200/\mathrm{CI}^*_{(1-\alpha)}N)^{0.5}$, and solving for $\mathrm{CI}^*_{(1-\alpha)}$ and $N$, yields:

$$\mathrm{CI}^*_{(1-\alpha)} = 800Z_{\alpha/2}^2/\delta^2 N \qquad (9)$$

and

$$N = 800Z_{\alpha/2}^2/\delta^2\mathrm{CI}^*_{(1-\alpha)} \qquad (10)$$

In a similar, but more complex derivation (see the Appendix) for the $F_2$ design, the contrast between the appropriate marker genotype groups, $D'$, is: $2\delta/(2-r)$, where $E(M_1)=\delta=2d$; and $\mathrm{SE}^2(D')=32/(2-r)^2rN$ (Eqs. 18 and 19). Thus for the $F_2$ design:

$$Z_{\alpha/2} = 2\delta/(32/rN)^{0.5} = \delta/(2/rkN)^{0.5} \qquad (11)$$

For the $F_2$ design only homozygotes for alternative marker alleles are used to construct the contrast (Appendix). Therefore, $k=0.25$ for this design. Letting $r= \mathrm{CI}^*_{(1-\alpha)}/200$, substituting in (1), and solving for $\mathrm{CI}^*_{(1-\alpha)}$ and $N$ yields:

$$CI^*_{(1-\alpha)} = 1,600Z^2_{\alpha/2}/\delta^2 N, \tag{12}$$

and

$$N = 1,600Z^2_{\alpha/2}/\delta^2 CI^*_{(1-\alpha)} \tag{13}$$

Taking $\alpha$=0.05, so that $Z_{\alpha/2}$=1.96 and substituting $\delta$=$d$+$h$ and $\delta$=2$d$ in Eqs. 9 and 12 for $CI^*_{(1-\alpha)}$ yields $CI^*_{(1-\alpha)}$=3,073/$(d+h)^2 N$ for a BC design, and $CI^*_{(1-\alpha)}$=1,537/$d^2 N$ for an $F_2$ design. These equations are virtually identical to those obtained by extensive simulation in Darvasi and Soller (1997).

Equations 7 and 11 can readily be generalized to other mapping designs according to the corresponding values for $\delta$, the expectation of the contrast for the marker $M_1$ located at the QTL, and $k$, the proportion of the mapping population in each marker genotype group making up the contrasts for the markers $M_1$ and $M_2$. More complex mapping designs that accumulate recombination events, such as advanced intercross lines (AIL, Darvasi and Soller 1995), full-sib intercross lines (FSIL, Song et al. 1999), and recombinant inbred lines (RIL, Soller and Beckmann 1990) differ from the BC and $F_2$ designs in the proportion of recombination per cM. To take this into account, Eq. 7 must be modified as follows to convert the proportion of recombination, $r$, which is the proportion of recombination in an $F_2$ or BC generation, into the effective accumulated proportion of recombination obtained in generation $g$:

$$Z_{\alpha/2} = \delta_D/(2/t_D k_D r N)^{0.5} \tag{14}$$

where $\delta_D$ and $k_D$ are the appropriate $\delta$ and $k$ values for the given design; and $t_D$ is a factor that converts the proportion of recombination obtained in generation $g$ into the effective accumulated proportion of recombination ob-

tained in actuality. Substituting $r$=$CI_{(1-\alpha)}/2$ in (14) gives the general expressions:

$$CI^*_{(1-\alpha)} = 400Z^2_{\alpha/2}/\left(\delta^2_D t_D k_D N\right) \tag{15}$$

$$N = 400Z^2_{\alpha/2}/\left(\delta^2_D t_D k_D CI^*_{(1-\alpha)}\right) \tag{16}$$

## Results

The predicted CI with a saturated genetic map for various experimental designs are given in Table 1. BC, $F_2$, and AIL designs are assumed to be initiated from fully inbred parental lines. Half-sib and full-sib designs in outcrossing populations (Soller and Genizi 1978) are the equivalent of BC and $F_2$ designs respectively, assuming that the parents of the families are heterozygous at the QTL and that the family size is sufficiently large so that the marker-QTL phase can be determined virtually without error. The FSIL design is a variant of the AIL design, adapted to outcrossing populations. It is initiated as a large $F_1$ family produced by a mating between two individuals, and is maintained by continued random mating within the families of each generation. In the cumulative AIL and FSIL designs (CAIL and CFSIL), progeny are phenotyped and genotyped in each generation from the $F_2$ generation on, to build a cumulating mapping population consisting of individuals from all of the generations. The mapping resolution of such a population will depend on the accumulated recombinants over all generations.

In the AIL design, from the $F_2$ generation on, only half of the chromosomal regions in any generation will be in the heterozygous state. Thus recombination accumulates at a rate of $t_D$=0.5$g$, relative to the BC and $F_2$ designs, where $g$ is the number of generations. In the FSIL design, three-quarters of the descendants of any one of the four parental chromosomes will be in the heterozygous state. Thus

**Table 1** Confidence interval of QTL location with a saturated genetic map for various experimental designs. $\delta$ represents the contrast value between marker genotype groups for the quantitative trait (codominance is assumed), $t_D$ the effective proportion of recombination per cM, $k_D$ the proportion of the mapping population in each of the marker genotype groups forming the mapping contrast, $CI_{1-\alpha}$ the length of the CI of 1-$\alpha$ in percent recombination, $CI_{(0.95)}$ the length of the 95% CI in percent recombination, and $N$ (10, 0.25) the required total population size for $CI_{(0.95)}$=10 cM with $d$=0.25. The codes representing the design of the population groups are as follows: BC backcross, $AIL(g)$ advanced intercross line

carried to generation $g$, $FSIL(g)$ full-sib intercross line carried to generation $g$, $CAIL(g)$ and $CFSIL(g)$ cumulative AIL and FSIL respectively carried to generation $g$, $RIL(n)$ recombinant inbred lines with $n$ individuals phenotyped per line. The additive effect at the QTL is represented as $d$. $\sigma_f = (2h^2+(1-h^2)/n)^{0.5}$, where $h^2$ equals heritability in the narrow sense and $h^2$=0.25 is assumed. $Z_{\alpha/2}$ gives the value of the standard normal variable with a probability of $\alpha/2$. For the CAIL and CFSIL designs the number of individuals per generation is given in *parenthesis*, and for the RIL designs the total number of individuals phenotyped is given in *parenthesis*

| Design | $\delta$ | $t_D$ | $k_D$ | $CI_{(1-\alpha)}$ | $CI_{(0.95)}$ | $N(10, 0.25)$ |
|---|---|---|---|---|---|---|
| BC | $d$ | 1.0 | 0.5 | $800Z_{\alpha/2}^2/\delta^2 N$ | $3,073/d^2 N$ | 4,917 |
| $F_2$ | $2d$ | 1.0 | 0.25 | $1,600Z_{\alpha/2}^2/\delta^2 N$ | $1,537/d^2 N$ | 2,458 |
| AIL ($g$=6) | $2d$ | 0.5$g$ | 0.25 | $3,200Z_{\alpha/2}^2/g\delta^2 N$ | $512/d^2 N$ | 819 |
| FSIL ($g$=6) | $2d$ | 0.75$g$ | 0.25 | $2,130Z_{\alpha/2}^2/g\delta^2 N$ | $341/d^2 N$ | 547 |
| CAIL ($g$=6) | $2d$ | 0.25$g$ | 0.25 | $6,400Z_{\alpha/2}^2/g\delta^2 N$ | $1,024/d^2 N$ | 1638 (328) |
| CFSIL ($g$=6) | $2d$ | 0.38$g$ | 0.25 | $4,210Z_{\alpha/2}^2/g\delta^2 N$ | $674/d^2 N$ | 1078 (218) |
| RIL ($n$=1) | $2d/\sigma_f$ | 2.0 | 0.5 | $400Z_{\alpha/2}^2/\delta^2 N$ | $480/d^2 N$ | 768 (768) |
| RIL ($n$=20) | $2d/\sigma_f$ | 2.0 | 0.5 | $400Z_{\alpha/2}^2/\delta^2 N$ | $207/d^2 N$ | 330 (6,600) |

recombination accumulates at a rate of $t_D=0.75g$, relative to the BC and $F_2$ designs. In the RIL design the final proportion of recombination over small distances is twice that in the $F_2$ generation, so that $t_D=2.0$ (Soller and Beckmann 1990).

Assuming codominance at the QTL, the contrast values will be $\delta_D=d$ for the BC design and $\delta_D=2d$ for the $F_2$ and AIL designs. The effect for the FSIL designs approaches $\delta_D=2d$, depending on the specific configuration of the marker and the QTL alleles (Song et al. 1999). The contrast value for the RIL design depends on the number of individuals phenotyped in each line, and will be $\delta_D=2d/\sigma_f$, where $\sigma_f^2$ is the variance among means of inbred lines. $\sigma_f^2=2h^2+(1-h^2)/n$ where $h^2$ is the heritability in the narrow sense and $n$ is the number of individuals scored for the quantitative trait from each RIL (Soller and Beckmann 1990).

As in the BC design, $k_D=0.5$ for the RIL design, for which one half of all lines are homozygous for one allele at each marker, and the other half are homozygous for the alternative allele. In the $F_2$ and AIL designs one quarter of the population are in each of the two contrasted marker genotype groups, so that $k_D=0.25$. In the FSIL population, at the optimal configuration of marker and QTL alleles, the proportions will be the equivalent of $k_D=0.25$ in each marker genotype group (Song et al. 1999).

Table 1 also shows the general expressions for $CI_{(1-\alpha)}$ as a function of $Z_{\alpha/2}$, $\delta_D$, $t_D$, $k_D$ and $N$, and specific values for $CI_{(0.95)}$. These expressions demonstrate that the CI is inversely proportional to population size and the square of the contrast value. Thus, methods such as multi-trait analysis that increase the contrast value (Korol et al. 1995) can markedly reduce the CI for given population size. General expressions for $N$ as a function of $Z_{\alpha/2}$, $\delta_D$, $t_D$, $k_D$, and $CI_{(1-\alpha)}$, as derived from Eq. 16, are also presented.

The required number of individuals genotyped to obtain a $CI_{(0.95)}$ of 10 cM with $d=0.25$ are listed in the right-hand column. $g=6$ is assumed for the AIL and FSIL designs, and $h^2=0.25$ for the RIL design. Sample sizes are quite large for the BC and $F_2$ designs, about 5,000 and 2,500 respectively. The numbers of genotyped individuals required by the AIL and FSIL designs are one-third and one-fifth respectively those for the $F_2$. Values for the FSIL assume an optimal configuration of marker and QTL alleles. Cumulative AIL and cumulative FSIL with $g=6$ require just twice the total numbers required for a single generation, but the numbers per generation are only 40% of the total. Thus, these designs are useful when total facilities are limited and not elastic. For the RIL design, the number of individuals genotyped is the total number of lines, while the number of individuals that must be scored for the quantitative trait is the number of lines multiplied by the number of individuals per line. This value is given in Table 1 in parenthesis. When a single individual is phenotyped for each line, 768 RIL have mapping power equivalent to an AIL ($g=6$); but 330 RIL, each evaluated on 20 individuals (6,600 individuals scored for the quantitative trait), has the same mapping power as a BC sample of 5,000 individuals. The formulas given in Table 1 demonstrate that very large samples are required for high resolution mapping. Achieving a 1 cM $CI_{(0.95)}$ would require 50,000 BC individuals, 8,000 AIL ($g=-6$), or 3,300 RIL with 20 individuals scored per line for a total of 66,000 phenotypes.

The equations for the BC and $F_2$ designs are virtually identical to Eqs. 1 and 2 of Darvasi and Soller (1997). The only differences are that the values of $k$ are 3,073 and 1,537, instead of 3,000 and 1,500 respectively, and that the CI is measured in percent recombination, rather than in cM. The value of 3,073 is well within the CI for the empirical estimate of the corresponding parameter derived by Darvasi and Soller (1997) and denoted "$k$". Furthermore, Darvasi and Soller (1997) slightly underestimated "$k$", because they simulated a chromosome of 100 cM. This imposes an artificial upper limit on the CI. This problem can also be noted in Fig. 2 of Darvasi and Soller (1997), where they used only a range of 1–20 cM to estimate "$k$".

The analytical formula derived can be used for any CI up to 50% recombination on either side of the estimated QTL location. The expressions in the sixth column of Table 1 can also be used to derive the minimum value of $d^2N$ for which a valid QTL CI can be derived. A $CI_{(0.95)}\geq100$ will include any point up along the chromosome with up to 50% recombination relative to the point with the maximum test statistic, which means that a $CI_{(0.95)}>100$ is essentially infinite. For the BC design, a valid $CI_{(0.95)}$ is obtain only if $d^2N>30.73$. For example, if $N=1,000$, then a valid $CI_{(0.95)}$ can be obtained only for $d>0.175$. If the estimated QTL position is near one end of the chromosome, a "one-tailed" CI would be more appropriate than a "two-tailed" CI that includes non-existent DNA.

## Discussion

The formulas we have derived here are only asymptotically correct. During analysis of an actual data set, the likelihood profile across the chromosome may be far from symmetrical, because chromosomes are of finite length, information content differs among markers, and marker spacing is not uniform. This would be taken into account when deriving CI from the data. We emphasize, however, that the purpose of this note is to derive a priori expectations for the CI for design purposes, or for general evaluation of the overall mapping resolution of an experiment, and not to derive CIs for specific QTL from an actual data set. In estimating CI for an actual experimental data set, empirical methods, such as bootstrap analysis (Visscher et al. 1996), can be used to obtain the CI of estimated QTL map locations.

Although this study assumed that an infinite number of markers were genotyped, Darvasi and Soller (1994) demonstrated that for most experimental designs if genotyping costs are large compared to phenotyping costs, the power to detect QTL is economically optimized by phenotyping many individuals for fewer markers. For

crosses between inbred lines or half-sib designs the optimum marker spacing is 80 cM, provided that unlimited numbers of individuals are available for phenotyping. Even if phenotyping costs are large relative to genotyping costs, the optimum marker spacing is no less than 20 cM if all markers are completely informative (Darvasi et al. 1993).

Percent recombination is close to cM for small values, but underestimates cM for larger values for most commonly used mapping functions. Measuring the CI in cM, rather than percent recombination, would not affect the relationship between the simulated and the predicted CI if both are given in the same units, as long as the CI $_{(0.95)}$ measured in percent recombination is <100.

The equations derived in this study do not account for the fact that the number of events of recombination in a finite sample is also finite. Although the expectation of the number of recombinants will be equal to $2R$, the number of recombinants in a finite sample will have a binomial distribution, which will increase the variance of the standard error of the contrast. As noted by Kruglyak and Lander (1995) in the BC design, $N$ events of recombination per Morgan are expected in a sample of $N$ individuals. If in the sample of $N$ individuals, there were no events of recombination between point $x_1$ and $x_2$ on the chromosome, then the probability of QTL location will be equal across the chromosomal segment $x_1$–$x_2$. [In this case the likelihood function is completely flat between $x_1$ and $x_2$, and has first and second derivatives of zero. This is one reason why generic software, such as Proc NLIN (SAS 1999), often has difficulty obtaining QTL CI.] Thus even for a gene with complete heritability, the mean length of the "critical interval" for gene location will be $2/N$ in units of Morgans, or $200/N$ in cM. The critical interval, as defined by Kruglyak and Lander (1995), differs from the CI in that with complete heritability there is zero probability that the gene is outside this interval.

The effect of the finite distribution of events of recombination will be negligible, unless the QTL effect is very large relative to the phenotypic standard deviation. For example even if $\delta=0.88$, then for the BC design, CI $_{(0.95)}=20$ for $N=200$. In this case, $2R=20$, and ten recombinants are expected in each marker class. When simulating, taking into account the binomial distribution of the number of recombinants, the standard error of the contrast was increased by 3.7%. As the effect of the QTL decreases, the discrepancy from the analytical formula will also decrease.

We were informed by a reviewer that Visscher and Goddard (2004), using somewhat different methodology, also derived the same formulas to predict CI$_{(0.95)}$ for the BC and F$_2$ designs.

# Appendix

## Derivation of a formula for CI of QTL location for an F2 mapping population using only the recombinant progeny

The contrast for an $F_2$ mapping population is based on individuals homozygous for alternative marker alleles. Thus, to be included in the recombinant $F_2$ mapping population, an individual must be recombinant for at least one chromosome, and homozygous for at least one of the marker alleles. Three of the nine possible $F_2$ marker genotypes do not meet these criteria. These are the two homozygous parental types ($M_1M_2/M_1M_2$ and $m_1m_2/m_1m_2$) and the double recombinant type ($M_1m_2/m_1M_2$). The remaining six $F_2$ marker genotypes that meet these criteria are listed in Table 2. The genotypic value of each genotype, assuming that the QTL is located at marker $M_1$, and the expected number of individuals having that genotype in the mapping population are also listed in the table.

The contrast for an $F_2$ mapping population is composed of the expected mean value of marker genotype groups that are homozygous for the alternative markers $M_1$, $m_1$, $M_2$, and $m_2$. Each of these, however, is composed of two recombinant marker genotype groups. For example, the marker genotype group $M_2M_2$ is composed of the recombinant genotype groups: $m_1M_2/M_1M_2$ (class E in Table 2) with genotypic value $h$ and frequency $2(1-r)r/4$ in the entire $F_2$ population, and marker genotype group $m_1M_2/m_1M_2$ (class D in Table 2) with genotypic value $-d$, and frequency $r^2/4$. The mean genotypic value of the $M_2M_2$ group including recombinants only is the mean of the genotypic values of the classes E and D, weighted by their relative frequency in the $M_2M_2$ recombinant group i.e., $2(1-r)r/[2(1-r)r+r^2]$ and $r^2/[2(1-r)r+r^2]$, which simplifies to $2(1-r)/(2-r)$ and $r/(2-r)$, respectively. These relative frequencies are the "weighting factors" listed in column three of Table 2. The contrast for the $F_2$ mapping population, $D'$, is computed as follows:

$$D' = (\underline{M_1M_1} - \underline{m_1m_1}) - (\underline{M_2M_2} - \underline{m_2m_2}) \qquad (17)$$

Letting A, B, C, D, E, and F represent their respective genotypic values, and letting $k_1=2(1-r)/(2-r)$, and $k_2=r/(2-r)$, so that $k_1+k_2=1$, we have:

**Table 2** Composition of the F$_2$ including only recombinant progeny

| Class | Genotype | Weighting factor | Number in group | Effect |
|-------|----------|------------------|-----------------|--------|
| A | $M_1m_2/M_1M_2$ | $2(1-r)/(2-r)$ | $2(1-r)rN/4$ | $d$ |
| B | $M_1m_2/M_1m_2$ | $r/(2-r)$ | $r^2N/4$ | $d$ |
| C | $m_1M_2/m_1m_2$ | $2(1-r)/(2-r)$ | $2(1-r)rN/4$ | $-d$ |
| D | $m_1M_2/m_1M_2$ | $r/(2-r)$ | $r^2N/4$ | $-d$ |
| E | $m_1M_2/M_1M_2$ | $2(1-r)/(2-r)$ | $2(1-r)rN/4$ | $h$ |
| F | $M_1m_2/m_1m_2$ | $2(1-r)/(2-r)$ | $2(1-r)rN/4$ | $h$ |

$$\underline{M}_1\underline{M}_1 = k_1 A + k_2 B$$
$$m_1\underline{m}_1 = k_1 C + k_2 D$$
$$\underline{M}_2\underline{M}_2 = k_1 E + k_2 D$$
$$m_2\underline{m}_2 = k_1 F + k_2 B$$

Substituting in (17) gives

$$
\begin{aligned}
D' &= [(k_1 A + k_2 B) - (k_1 C + k_2 D)] \\
&\quad - [(k_1 E + k_2 D) - (k_1 F + k_2 B)] \\
&= k_1 A + k_1 F - k_1 C - k_1 E + 2k_2 B - 2k_2 D \\
&= k_1(A + F - C - E) + 2k_2(B - D) \\
&= [1/(2-r)]\left\{ \begin{array}{l}[2(1-r)] \\ {[d+h+d-h] + 2r[d+d]}\end{array}\right\} \\
&= 4d/(2-r)
\end{aligned}
$$

(18)

To calculate SE($D$), note that:

$$\sigma_A^2 = \sigma_F^2 = \sigma_C^2 = \sigma_E^2 = 1/[2(1-r)rN/4] = 4/2(1-r)rN$$
$$\sigma_B^2 = \sigma_D^2 = 4/r^2 N$$

Thus:

$$
\begin{aligned}
SE^2(D') &= 4[2(1-r)/(2-r)]^2 4/[2(1-r)rN] \\
&\quad + 2[2r/(2-r)]^2[4/(r^2 N)] \\
&= 4[2(1-r)]/(2-r)^2[4/rN] \\
&\quad + 2\left[4/(2-r)^2\right][4/N] \\
&= 32(1-r)/(2-r)^2 rN + 32/(2-r)^2 N \\
&= [32(1-r) + 32r]/\left[(2-r)^2 rN\right] \\
&= 32/(2-r)^2 rN
\end{aligned}
$$

(19)

## References

Darvasi A, Soller M (1994) Optimum spacing of genetic markers for determining linkage between marker loci and quantitative trait loci. Theor Appl Genet 89:351–357

Darvasi A, Soller M (1995) Advanced intercross lines, an experimental population for fine genetic mapping. Genetics 141:1199–1207

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

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

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

Kruglyak L, Lander ES (1995) High-resolution genetic mapping of complex traits. Am J Hum Genet 57:1212–1223

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

Ronin Y, Korol A, Shtemberg M, Nevo E, Soller M (2003) High resolution mapping of quantitative trait loci by selective recombinant genotyping. Genetics 174:1757–1777

SAS/STAT User's Guide, Version 8 (1999) SAS Institute, Cary

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

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

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

Soller M, Genizi A (1978) The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. Biometrics 34:47–55

Song JZ, Soller M, Genizi A (1999) The full-sib intercross line (FSIL) design: a QTL mapping design for outcrossing species. Genet Res 73:61–73

VanRaden PM, Weller JI (1994) A simple method to locate and estimate effects of individual genes with a saturated genetic marker map. J Dairy Sci 77[Suppl 1]:249

Visscher PM, Goddard ME (2004) Prediction of the confidence interval of QTL location. Behav Genet 34:477–482

Visscher PM, Thompson R, Haley CS (1996) Confidence intervals in QTL mapping by bootstrapping. Genetics 143:1013–1020