Combining Different Line Crosses for Mapping Quantitative Trait Loci Using the Identical by Descent-Based Variance Component Method

Chongqing Xie, Damian D. G. Gessler and Shizhong Xu

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

Manuscript received October 1, 1997
Accepted for publication February 20, 1998

ABSTRACT

Mapping quantitative trait loci (QTLs) is usually conducted with a single line cross. The power of such QTL mapping depends highly on the two parental lines. If the two lines are fixed for the same allele at a putative QTL, the QTL is undetectable. On the other hand, if a QTL is segregating in the line cross and is detected, the estimated variance of the QTL cannot be extrapolated beyond the statistical inference space of the two parental lines. To reduce the likelihood of missing a QTL and to increase the statistical inference space of the estimated QTL variance, we present a consensus QTL mapping strategy. We adopt the identical by descent (IBD)-based variance component method originally applied to human linkage analysis by combining multiple line crosses as independent families. We explore the properties of consensus QTL mapping and demonstrate the method with F_2, backcross (BC), and full-sib (FS) families. In addition, we examine the effects of the QTL heritability, marker informativeness, QTL position, the number of families, and family size. We show that F_2 families notably outperform BC and FS families in detecting a QTL. There is a substantial reduction in the standard deviation of the estimated QTL position and the separation of the QTL and polygenic variance. Finally, we show that the power to detect a QTL is greater when using a small number of large families than a large number of small families.

LINE crossing is a common experimental design for mapping quantitative trait loci (QTLs) in plants and laboratory animals. Crosses are initiated from at least two inbred lines, such as backcrosses (BC), F_2, and more derived generations. Statistical methods are well developed for QTL mapping using such line crossing data (Lander and Botstein 1989; Haley and Knott 1992; Martínez and Curnow 1992; Jansen 1993; Zeng 1994). These methods are mainly designed to handle a single line cross. The characteristics of line crossing experiments are: (1) a small number of parental lines are involved, (2) the linkage phases of the parental markers are known, and (3) family sizes are usually large. These properties allow the effects of a gene substitution to be tested directly. The methods developed by the above authors all test the effects of a gene substitution (the first moments) and therefore are referred to as the fixed model approach (Xu and Atchley 1995).

Quantitative geneticists are interested not only in detecting QTLs and locating their positions, but also in estimating the contribution of the detected QTLs to a trait. The contribution of a QTL, however, is only meaningful when expressed relative to the total phenotypic variance. Therefore, the effect of a QTL is actually measured by its variance. In a single line cross, the QTL variance is relative to the genetic variance among individuals within that line cross; *i.e.*, the QTL variance is formulated as conditional on the cross. As a result, the variance itself is a variable that differs from one cross to another. Therefore, a QTL variance estimated from a single line cross cannot be extended to a statistical inference space beyond that cross. In addition, the number of founder alleles at any locus is expected to be small in a line cross. For instance, there are at most two alleles at each locus in an F_2 family. With such a single line cross, one's entire effort is invested in this single large family. If the two founder alleles of a QTL are polymorphic, then detection of the QTL is possible with a relatively large family. On the other hand, if the two parental lines are fixed for the same allele at a particular QTL, then this QTL is undetectable, independent of the sample size. To increase the statistical inference space of the estimated QTL variance and ensure that polymorphic alleles are present in the parental gene pool, one needs to sample a sufficient number of parents (Muranty 1996). This can be achieved by combining data from multiple line crosses.

Suppose that there are 10 F_2 families derived from 10 pairs of inbred lines. What is the appropriate statistical method for analyzing the data from these F_2 families? One may simply extend the regression approach to fit 2 additional parameters, 1 mean and 1 gene substitution effect, for each F_2 family added to the data set. This means estimating 20 parameters and testing 10 additive effects. If dominance deviations are considered, 10 additional terms must be estimated and tested. In a single-line, fixed model, one can easily convert the effect of

Corresponding author: Shizhong Xu, Department of Botany and Plant Sciences, University of California, Riverside, CA 92521-0124.
E-mail: xu@genetics.ucr.edu

a gene substitution into the variance via $\sigma_g^2 = \alpha^2/2$, where α is the average effect of a gene substitution (Falconer and Mackay 1996). But as the number of parameters increases, extension of this method becomes complicated.

Data from multiple line crosses, such as diallelic and four-way crosses, can occasionally be analyzed using the methods of Rebai and Goffinet (1993) and Xu (1996a). A survey of the literature shows that the most popular computer software, such as MAPMAKER/QTL (Lincoln *et al.* 1993), QTL Cartographer (Basten *et al.* 1997), MapQTL (van Ooijen and Maliepaard 1996), and MQTL (Tinker and Mather 1995), are designed to handle only a single line cross.

In contrast to the difficulties of the fixed model, the IBD-based variance component method initially developed for human genetic studies can handle multiple families (Haseman and Elston 1972). This method has been referred to as the random model approach because the QTL variance is directly estimated and tested (Xu and Atchley 1995). To separate the QTL variance from the polygenic variance, the IBD-based approach relies on variation in the proportion of genes IBD shared by relatives at the putative QTL. The random model approach is adopted here for combining data from different line crosses because each line cross is effectively a different family.

Before one can apply the random model approach to line crosses, one needs to adjust for the fact that regular full-sib (FS) families and families of line crosses differ in that the latter involves inbreeding (for example, an F_2 individual is equivalent to a progeny resulting from a selfing parent). The traditional IBD-based method must therefore be modified to reflect the inbreeding effect. Our purpose here is to develop such an IBD-based random model methodology for combining data from different line crosses. We examine two types of line crosses: F_2 and BC. We then compare the results with regular noninbred, FS families.

STATISTICAL METHODS

Linear model and likelihood function: We combine line crosses by treating each line cross as a family and using a multipoint QTL mapping methodology. Consider a family with n individuals; the phenotypic value (y_i) of the ith individual is described as $y_i = \mu + g_i + a_i + e_i$ (Goldgar 1990; Xu and Atchley 1995), where μ is the overall mean, g_i is the additive effect of a putative QTL with mean 0 and variance σ_g^2, a_i is the polygenic effect (excluding g_i) with mean 0 and variance σ_a^2, and e_i is the residual error distributed as $N(0, \sigma_e^2)$. Inclusion of a dominance effect is discussed later. In matrix notation, the model is $\mathbf{y} = \mathbf{1}\mu + \mathbf{g} + \mathbf{a} + \mathbf{e}$. The expectation and variance of the model are $E(\mathbf{y}) = \mathbf{1}\mu$ and $\mathrm{Var}(\mathbf{y}) = \mathbf{V} = \Pi\sigma_g^2 + \mathbf{A}\sigma_a^2 + \mathbf{I}\sigma_e^2$, respectively, where $\mathbf{1}$ is a column vector of order n, $\Pi = \{\pi_{ij}\}_{n\times n}$ is a square matrix of order n with the element of the ith row and the jth column being the shared IBD value between sibs i and j at the QTL, $\mathbf{A}$ is an additive relationship matrix of order n with its elements being twice the coancestry coefficients for the polygenic component, and $\mathbf{I}$ is the identity matrix.

Under the assumption that $\mathbf{y}$ is multivariate normal and Π is known, the likelihood function for a particular family is

$$L(\beta|\mathbf{y}\Pi) = \frac{1}{(2\pi)^{n/2}|\mathbf{V}|^{1/2}} \exp[-\tfrac{1}{2}(\mathbf{y} - \mathbf{1}\mu)^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{1}\mu)],$$

where $\beta = [\mu\ \sigma_g^2\ \sigma_a^2\ \sigma_e^2]^{\mathrm{T}}$ are the unknown parameters and the superscript T stands for matrix transposition.

Assume that families are independent so that the overall likelihood function for multiple families is simply the product of these family-specific likelihoods. Therefore, the overall log likelihood for N families is:

$$L = \sum_{k=1}^{N} \log[L_k(\beta|\mathbf{y}_k\Pi_k)],$$

where $L_k(\beta|\mathbf{y}_k\ \Pi_k)$ represents the likelihood of the kth family.

To test the presence of a QTL, a log likelihood ratio test statistic is used, which is $\Lambda = -2(L_0 - L_1)$, where L_1 is the log likelihood value evaluated at the maximum likelihood solution under the alternative model ($\beta_1 = [\mu\ \sigma_g^2\ \sigma_a^2\ \sigma_e^2]^{\mathrm{T}}$) and L_0 is the log likelihood value evaluated at the maximum likelihood solution under the null model ($\beta_0 = [\mu\ \sigma_a^2\ \sigma_e^2]^{\mathrm{T}}$).

The IBD value between two sibs at a QTL: Because of inbreeding, the IBD values among F_2 individuals are different from those among regular full sibs. If the parental lines are fixed for alternative QTLs, then F_2 individuals have three possible genotypes at a QTL: *QQ*, *Qq*, and *qq*. Given the genotypic configuration of individuals i and j, the IBD value is measured as

$$\pi_{ij} = 2\theta_{ij} = \begin{cases} 2 & \text{for } QQ\text{-}QQ \text{ or } qq\text{-}qq \\ 1 & \text{for } QQ\text{-}Qq,\ qq\text{-}Qq, \text{ or } Qq\text{-}Qq \\ 0 & \text{for } QQ\text{-}qq, \end{cases}$$

where π_{ij} are the ijth elements of Π and θ_{ij} is Malécot's (1948) coefficient of coancestry. Note that while π_{ij} between noninbred, full sibs ranges from zero to one, under inbreeding π_{ij} ranges from zero to two. In this usage, π_{ij} is not interpreted as the proportion of alleles IBD, but rather as twice the coefficient of coancestry (Kempthorne 1955; Harris 1964; Cockerham 1983). For example, when the two individuals have a genotypic configuration of *QQ-Qq*, the coefficient of coancestry is

$$\begin{aligned}\theta_{ij} &= \tfrac{1}{4}[\Pr(Q\equiv Q) + \Pr(Q\equiv q) + \Pr(Q\equiv Q) + \Pr(Q\equiv q)] \\ &= \tfrac{1}{4}[1 + 0 + 1 + 0] = \tfrac{1}{2}.\end{aligned}$$

Without inbreeding, the IBD value of an individual with itself (π_{jj}) always takes a value of 1. Under inbreeding, π_{jj} can be greater than 1, depending on whether the individual is homozygous or heterozygous, *i.e.*,

$$\pi_{jj} = 1 + \phi_j = \begin{cases} 2 & \text{if } QQ\text{-}QQ \text{ or } qq\text{-}qq \\ 1 & \text{if } Qq\text{-}Qq \end{cases},$$

where ϕ_j is the inbreeding coefficient of individual j at the QTL. Here, π_{jj} can be interpreted as 1 plus the inbreeding coefficient (Harris 1964; Cockerham 1983).

The elements in the additive relationship matrix $\mathbf{A}$ are IBD values of the polygenic component and can be obtained by taking the unconditional expectation of π_{ij}. In an F_2 family, $\mathbf{A}$ has elements of $A_{ij} = E(\pi_{ij}) = 1$ and $A_{jj} = E(\pi_{jj}) = \frac{3}{2}$, in contrast to $A_{ij} = \frac{1}{2}$ and $A_{jj} = 1$ in a regular FS family.

In BC populations, the π's are derived similarly. They are

$$\pi_{ij} = 2\theta_{ij} = \begin{cases} 2 & \text{for } QQ\text{-}QQ \\ 1 & \text{for } QQ\text{-}Qq \text{ or } Qq\text{-}Qq \end{cases}$$

for between individuals and

$$\pi_{jj} = 1 + \phi_j = \begin{cases} 2 & \text{for } QQ\text{-}QQ \\ 1 & \text{for } Qq\text{-}Qq \end{cases}$$

for the individual with itself. **A** has elements of $A_{ij} = 5/4$ and $A_{jj} = 3/2$.

Inferring the IBD value of a QTL from markers: The IBD value is completely determined by the genotypes of two individuals at the QTL of interest. The actual genotype of an individual, however, is not observable and it must be inferred from its marker information. In F_2 and BC populations, two flanking markers are sufficient if Haldane's mapping function is assumed and the markers are completely informative. The conditional distribution of the QTL genotype given the genotypes of the flanking markers is given by Jiang and Zeng (1996). Denote the conditional probabilities of the three genotypes of the QTL by $p_{j2} = \Pr(QQ|I_M)$, $p_{j1} = \Pr(Qq|I_M)$, and $p_{j0} = \Pr(qq|I_M)$, and let $\mathbf{p}_i = [p_{i2}\ p_{i1}\ p_{i0}]^T$ and $\mathbf{p}_j = [p_{j2}\ p_{j1}\ p_{j0}]^T$. The conditional expectations of the IBD values are $\hat{\pi}_{ij} = E(\pi_{ij}|I_M) = \mathbf{p}_i^T\mathbf{C}\mathbf{p}_j$ for between individuals, and $\hat{\pi}_{ij} = E(\pi_{jj}|I_M) = \mathbf{c}^T\mathbf{p}_j$ for the individual with itself, where

$$\mathbf{C} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{c} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$$

The conditional expectations of the IBD in a BC population are

$$\hat{\pi}_{ij} = E(\pi_{ij}|I_M) = 2p_{i2}p_{j2} + p_{i2}\,p_{j1} + p_{j1}\,p_{i2} + p_{i1}\,p_{j1}$$

for between individuals, and

$$\hat{\pi}_{jj} = E(\pi_{jj}|I_M) = 2p_{j2} + p_{j1}$$

for an individual with itself. When the maximum likelihood method is performed, $\hat{\pi}_{ij} = E(\pi_{ij}|I_M)$ is used in substitution of π_{ij}.

SIMULATION STUDIES

Individuals within an F_2 family are equivalent to full sibs resulting from selfing a single parent. As a consequence, we randomly sampled a single parent from an infinitely large panmictic (or base) population. This single parent was then selfed to produce an F_2 family. Individuals within a BC family were derived by crossing an F_1 hybrid with one of its homozygous parents. The regular (noninbred) FS families were generated from the mating of two unrelated parents sampled from the base population. Families, including those of regular (noninbred) FS families, were analyzed via the maximum likelihood method (Xu and Atchley 1995).

To infer the IBD value of a QTL from markers, we used a multipoint methodology (Fulker *et al.* 1995; Kruglyak and Lander 1995; Olson 1995). In most cases, we simulated one chromosome of length 100 cM with six biallelic markers evenly spaced along the chromosome. The two alleles at each marker were equally frequent. A single QTL with six equally frequent alleles was simulated at position 50. In addition to the QTL of interest, we also simulated 12 independent biallelic loci of equal effects to form the polygenic contribution. A detailed description of the simulation process for random mating populations can be found in Gessler and Xu (1996).

We simulated 50 families each with 10 siblings (a total of 500 individuals). For each run, a single set of phenotypic values was generated with a QTL, polygenic, and residual variance of $\sigma_g^2 = 12.5$, $\sigma_a^2 = 12.5$, and $\sigma_e^2 = 25$, respectively. The joint contribution of the QTL and the polygenic component to the total phenotypic variance was 50%, with 25% for each component. We referred to this as the standard setting.

To examine the effect of different factors on the performance of the methods, we varied each of the following factors successively: (a) the number of families × family size: 20 × 25, 250 × 2, or 500 × 2, (b) QTL heritability, $h_g^2 = 0.10$ ($\sigma_a^2 = 12.5$, $\sigma_g^2 = 5.0$, and $\sigma_e^2 = 32.5$) or $h_g^2 = 0.5$ ($\sigma_a^2 = 12.5$, $\sigma_g^2 = 25.0$, and $\sigma_e^2 = 12.5$), (c) true QTL position at 10 cM or 30 cM, (d) low marker information (two alleles, one with a frequency of 0.9 and the other with 0.1) and high marker information (six equally frequent alleles at each marker), and (e) 11 markers each with two alleles spaced every 10 cM. We report results of 100 repeated simulations for each parametric setting.

To estimate the strength of a false positive signal, we ran an additional 1000 simulations with no QTL segregating. We augmented the polygenic variance such that the total genetic variance remained unchanged. From each simulation we chose the maximum observed likelihood ratio (LR) found across the chromosome and then determined the 95th percentile from the list of 1000 runs as an estimate of the chromosome-wise critical value.

RESULTS

The average likelihood ratio (test statistic) profiles over 100 replications of the three mating designs under the standard setting are depicted in Figure 1. It is evident that the F_2 families notably outperform the two other mating designs. The benefit of QTL mapping using F_2 families is manifest as a signal 70% higher than BC families, with BC families having a slightly higher signal than FS families. Since the critical values of the LR test statistic in the three mating populations are nearly

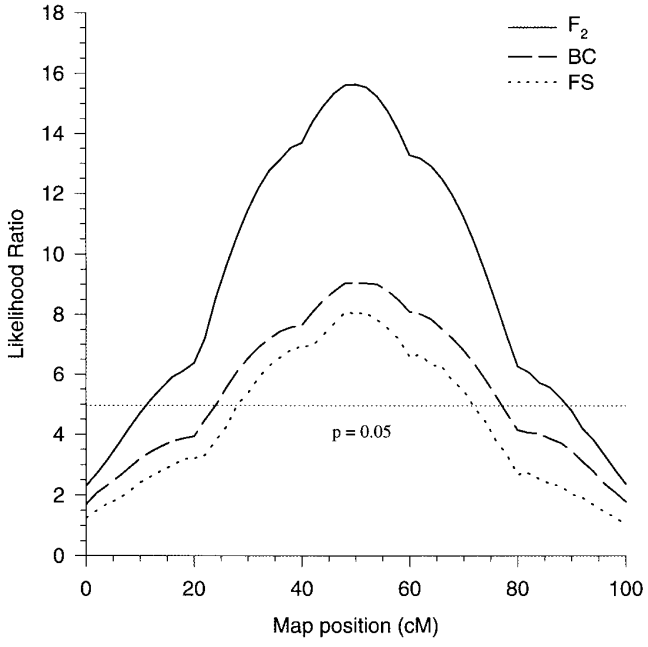

Figure 1.—Comparison of the LR profiles of the standard setting for F_2, backcross (BC), and full-sib (FS) families. The horizontal dotted line indicates the corresponding 5% empirical threshold.

TABLE 1

Estimates of the position, total phenotypic variance (σ_P^2), and QTL (h_g^2) and polygenic (h_a^2) heritabilities in the standard setting

Family	Position	σ_P^2	h_g^2	h_a^2
F_2	50.86 (13.67)	49.06 (3.38)	0.25 (0.076)	0.22 (0.121)
BC	51.46 (18.85)	49.05 (4.17)	0.26 (0.106)	0.22 (0.131)
FS	50.66 (17.32)	49.49 (3.77)	0.24 (0.146)	0.25 (0.177)

The standard setting is a QTL at position 50 cM, a phenotypic variance of 40, $h_g^2 = h_a^2 = 0.25$, medium marker informativeness (defined in the text), a 20-cM marker interval, with 50 families × 10 sibs. Standard deviations among 100 replicates are given in parentheses.

TABLE 2

Estimates of QTL parameters under two levels of heritabilities for three mating populations

Family	True h_g^2	Position	σ_P^2	h_g^2	h_a^2
F_2	0.50	50.58 (5.44)	49.03 (4.64)	0.46 (0.097)	0.28 (0.166)
	0.10	48.58 (22.76)	49.94 (3.27)	0.13 (0.068)	0.23 (0.116)
BC	0.50	49.82 (6.93)	48.43 (4.75)	0.47 (0.138)	0.26 (0.181)
	0.10	49.16 (26.81)	49.97 (3.89)	0.11 (0.096)	0.22 (0.117)
FS	0.50	48.84 (7.17)	49.84 (4.18)	0.50 (0.094)	0.26 (0.174)
	0.10	49.32 (27.76)	49.72 (3.71)	0.14 (0.090)	0.20 (0.122)

See Table 1 for the standard setting. Each additional run differs from the standard setting by parameter change noted in the second column. Standard deviations among 100 replicates are given in parentheses.

TABLE 3

Estimates of QTL parameters with high (or low) marker informativeness and a 10-cM marker interval

Family	Parameter	Position	σ_P^2	h_g^2	h_a^2
F_2	High	49.84 (8.53)	49.69 (3.55)	0.24 (0.071)	0.26 (0.130)
	Low	46.50 (25.92)	49.90 (6.17)	0.21 (0.136)	0.27 (0.156)
	10 cM	50.22 (7.07)	49.83 (3.67)	0.25 (0.076)	0.23 (0.128)
BC	High	51.84 (15.76)	49.76 (4.23)	0.26 (0.094)	0.23 (0.120)
	Low	46.24 (23.52)	50.16 (7.23)	0.27 (0.181)	0.22 (0.186)
	10 cM	48.14 (9.44)	48.67 (4.12)	0.27 (0.099)	0.22 (0.127)
FS	High	50.83 (10.11)	49.47 (4.38)	0.23 (0.087)	0.25 (0.149)
	Low	52.55 (21.70)	50.63 (5.53)	0.26 (0.155)	0.22 (0.189)
	10 cM	49.30 (9.36)	49.41 (4.33)	0.27 (0.094)	0.21 (0.139)

See Table 1 for the standard setting. Each additional run differs from the standard setting by the parameter change noted in the second column. Standard deviations among 100 replicates are given in parentheses.

equivalent, we conclude that QTL mapping using F_2 families has a higher power than BC and FS families under the standard parameter setting.

Under the standard setting, the QTL position and the total phenotypic variance are successfully estimated, while the sum of the heritabilities, $h_g^2 + h_a^2$, is as expected for all three mating designs, implying a fair partitioning of the genetic and residual variances (Table 1). In addition, F_2 families provide more accurate estimates than BC and FS families in the estimated QTL position and various variance components.

The levels of the QTL effect (proportional to the heritability value at the QTL) and marker informativeness produce a strong effect on the precision of the estimated QTL position. As expected, a higher QTL effect or higher marker informativeness decreases the standard deviation of the estimated QTL position (Tables 2 and 3). However, the levels of QTL heritability or marker informativeness have a smaller effect on the precision of the phenotypic variance and estimated heritabilities. Higher marker informativeness tends to decrease the standard deviation of various ML estimates, while a decrease in marker informativeness leads to an increase in the confounding of h_g^2 and h_a^2. In contrast, higher heritability levels tend to be associated with a slightly larger standard deviation in the estimated phenotypic variance and heritabilities.

One clear feature in the simulations is that using a

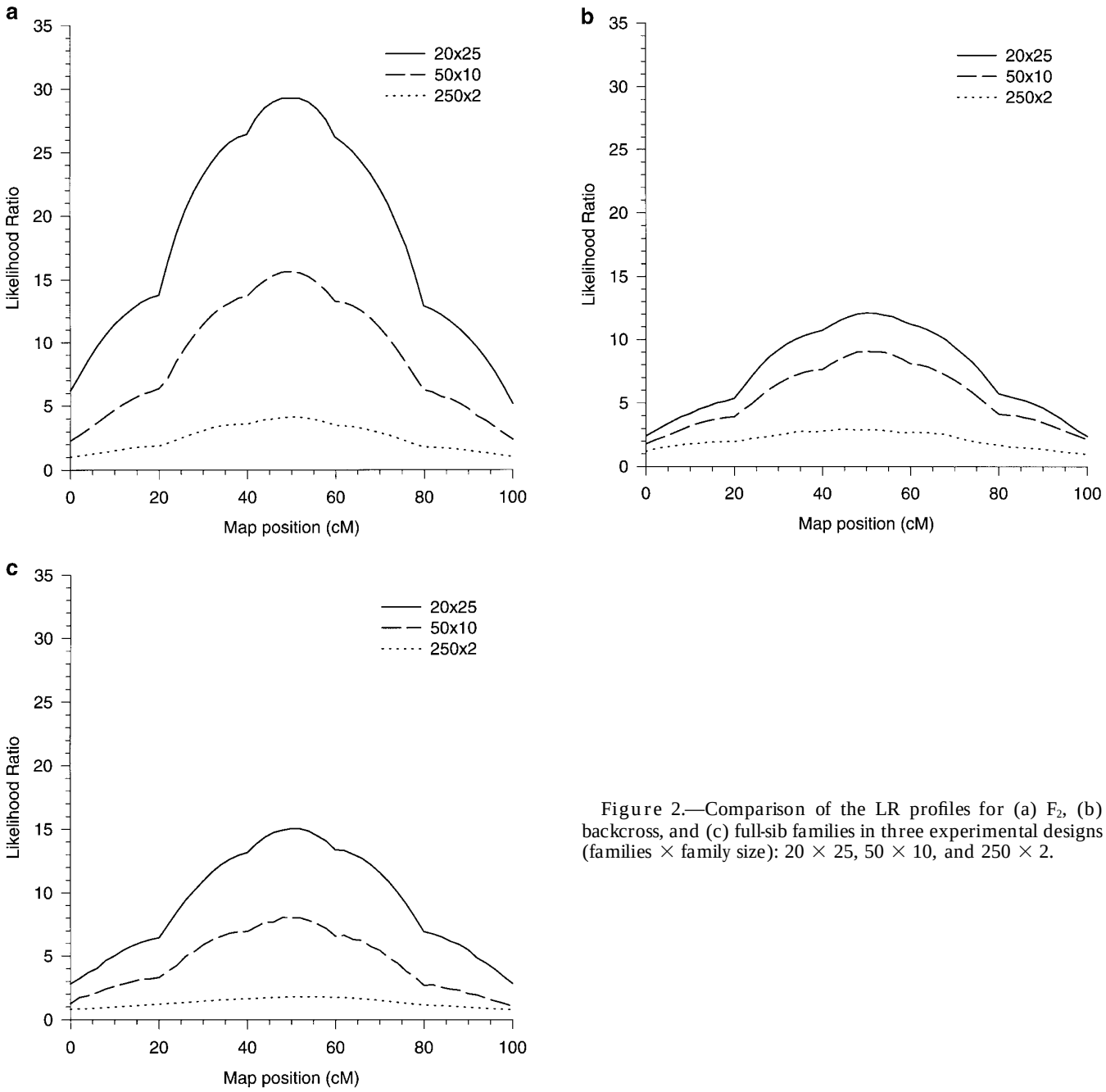

Figure 2.—Comparison of the LR profiles for (a) F_2, (b) backcross, and (c) full-sib families in three experimental designs (families × family size): 20 × 25, 50 × 10, and 250 × 2.

TABLE 4

Estimates of QTL parameters under different experimental designs

Family	Design[a]	Position	σ_P^2	h_g^2	h_a^2
F_2	20 × 25	48.92 (10.24)	48.94 (4.92)	0.23 (0.085)	0.24 (0.190)
	250 × 2	48.57 (22.41)	49.84 (2.95)	0.30 (0.120)	0.20 (0.148)
	500 × 2	49.16 (18.54)	50.15 (2.15)	0.26 (0.087)	0.24 (0.103)
BC	20 × 25	49.72 (13.10)	48.76 (5.59)	0.25 (0.111)	0.22 (0.155)
	250 × 2	49.80 (27.54)	49.70 (3.39)	0.32 (0.152)	0.18 (0.160)
	500 × 2	51.64 (22.78)	50.24 (2.04)	0.30 (0.130)	0.20 (0.133)
DS	20 × 25	50.28 (11.64)	48.88 (5.30)	0.26 (0.093)	0.20 (0.165)
	250 × 2	50.09 (29.93)	49.88 (3.48)	0.30 (0.192)	0.20 (0.212)
	500 × 2	47.47 (28.47)	49.81 (2.18)	0.25 (0.178)	0.25 (0.195)

See Table 1 for the standard setting. Each additional run differs from the standard setting by the parameter change noted in the second column. Standard deviations among 100 replicates are given in parentheses.

[a] Number of families × family size.

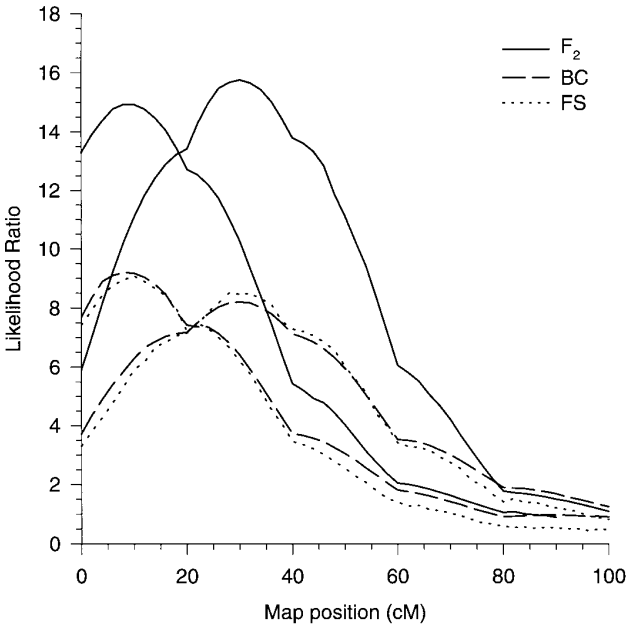

Figure 3.—Comparison of the LR profiles for F_2, backcross (BC), and full-sib (FS) families. All other information is the same as that in Figure 1 except that a single QTL is located at 10 cM for the left set of curves and at 30 cM for the right set of curves.

large sibship per family has a pronounced effect on the ability to detect the QTL. Figure 2 presents the results of sibships for three mating populations. The signal at the QTL with 10 or 25 sibs per family is 250 or 500% higher, respectively, than that for two sibs per family. In addition, with a fixed number of 500 individuals tested, increasing family size from 2 to 25 decreases the standard deviation of the estimated QTL position. It also increases the ability to separate the genetic variance into the polygenic and the QTL components (Tables 1 and 4). The standard deviation of the estimated phenotypic variance increases as the number of families decreases (Table 4).

Figure 3 shows the simulation results with the QTL at position 10 or 30 of a chromosome of length 100 cM. It is generated by taking the average value at each position, and this method shows no bias in predicting the position of the QTL. Alternatively, taking the average maximum value of each run produces a slight bias toward the center of the chromosome, as reported in Table 5. Of the three populations, the BC population has the largest bias in the estimated QTL position. This bias is caused by some runs where the QTL effect is not significant. In these situations, the QTL position, on average, tends to be close to the center.

The empirical threshold values of LR test statistics over 1000 replicated simulations are reported in Table 6. It can be seen that all three mating populations have nearly equivalent critical values. The average LR test statistics and the power estimates (Type I error rate at $\alpha = 0.05$) over 100 replicated simulations are summarized in Table 7. First, the average LR in F_2 families is notably greater than that in BC or FS families, whereas both BC and FS families have similar test statistics and powers. Second, under the condition of low marker informativeness, FS families have a power in QTL detection relatively higher than that of either F_2 or BC families. Recall that an F_2 family is generated from a single parent by selfing. Accordingly, only two alleles at a spe-

TABLE 6

Observed 95th percentile likelihood ratios under the hypothesis of no QTL segregation

Case	F_2	Backcross	Full-sib
Standard	4.97	4.88	5.00
$h_a^2 = 0.75^a$	5.76	5.18	5.89
$h_a^2 = 0.35^a$	5.40	4.83	5.54
Marker informativeness:			
High	5.75	5.22	5.72
Low	5.29	4.88	4.58
10-cM interval	4.19	4.19	4.63
500 families × 2 sibs	5.13	5.88	5.52

See Table 1 for the standard setting. Each additional run differs from the standard setting by the parameter change noted in the first column.

[a] $h_g^2 = 0$.

TABLE 5

Estimates of QTL parameters with the QTL located at position 10 or 30 cM

Family	True position	Estimated position	σ_P^2	h_g^2	h_a^2
F_2	10	13.96 (18.56)	49.12 (3.44)	0.23 (0.084)	0.26 (0.125)
	30	31.78 (11.19)	50.77 (4.16)	0.24 (0.086)	0.24 (0.136)
BC	10	17.90 (18.43)	49.42 (4.12)	0.27 (0.107)	0.22 (0.141)
	30	36.25 (20.08)	49.83 (4.75)	0.26 (0.109)	0.23 (0.135)
FS	10	14.75 (16.17)	50.13 (4.20)	0.28 (0.101)	0.21 (0.159)
	30	31.58 (13.26)	49.87 (3.72)	0.27 (0.107)	0.20 (0.153)

See Table 1 for the standard setting. Each additional run differs from the standard setting by the parameter change noted in the second column. Standard deviations among 100 replicates are given in parentheses.

TABLE 7

Test statistic and power to detect QTLs

Case	F_2		Backcross		Full-sib	
	Test statistic	Power (%)	Test statistic	Power (%)	Test statistic	Power (%)
Standard	17.27 (9.41)	97	9.58 (5.94)	76	9.35 (7.80)	72
$h_g^2 = 0.50$	66.29 (24.16)	100	40.64 (18.70)	99	37.31 (12.32)	100
$h_g^2 = 0.10$	5.50 (4.37)	46	2.59 (3.15)	21	3.32 (3.21)	15
Marker informativeness:						
High	22.63 (11.39)	98	12.33 (7.21)	85	12.54 (6.86)	86
Low	8.07 (8.57)	46	6.60 (7.08)	48	6.39 (5.93)	54
10-cM interval	24.22 (11.61)	100	13.10 (7.51)	90	14.16 (8.42)	89
Number of families × family size:						
500 × 2	7.67 (4.57)	68	6.75 (4.74)	55	3.09 (3.53)	28
250 × 2	5.56 (4.02)	—	4.63 (3.32)	—	2.45 (2.97)	—
20 × 25	30.72 (16.40)	—	15.74 (9.75)	—	17.61 (9.29)	—
QTL position:						
P10 cM	16.01 (9.56)	—	9.74 (5.98)	—	10.13 (5.75)	—
P30 cM	17.50 (9.29)	—	9.95 (6.81)	—	9.82 (6.58)	—

See Table 1 for the standard setting. Each additional run differs from the standard setting by the parameter change noted in the first column.

— Simulations are not performed under the hypothesis of no QTL segregation of these schemes.

cific locus are randomly sampled from the reference population. The two alleles have a large probability being the same state under the condition of low marker informativeness. In contrast, to generate FS families, two parents or four alleles at a specific locus are randomly sampled. This process essentially reduces the chance of a locus being monomorphic and thus increases the marker and QTL informativeness. This explains why the FS design is more powerful than the F_2 and BC when the marker information content is low.

DISCUSSION

What contributes to the variance in the estimate of the total phenotypic variance σ_P^2? For example, in Table 2 the standard deviation is higher for the high heritability—something counter to expectation. This is due to a scaling effect, *i.e.*, the standard deviation being positively correlated with the mean. The phenotypic variance is estimated by $\sigma_P^2 = \sigma_a^2 + \sigma_g^2 + \sigma_e^2$, and thus is the sum of three random variables. Because of the relation $\mathbf{V} = \mathbf{A}\sigma_a^2 + \Pi\sigma_g^2 + \mathbf{I}\sigma_e^2$, the variance in estimates of σ_g^2 is greater than that for σ_e^2. This is due in part to a confounding between σ_g^2 and σ_a^2 greater than that between σ_g^2 and σ_e^2. This means that the variance in σ_P^2 increases with σ_g^2, rendering a standard deviation for $h^2 = 0.5$ higher than that for $h^2 = 0.10$.

Similarly, in Table 4 small families have smaller standard deviations. Note that the phenotypic variance can be partitioned into variance between families and variance within families. When the total number of individuals is fixed, reducing the number of families increases the standard deviation for the between-family variance component (genetic drift) and decreases the standard deviation for the within-family component. When the increase is greater than the decrease, the net effect on the estimated phenotypic variance is an increased standard deviation.

To make the most efficient use of marker data, many QTL mapping experiments are designed to detect a number of economic traits, rather than only one trait (*e.g.*, Edwards *et al.* 1987). How to select parental lines that are fixed for alternative QTLs for multiple traits is a difficult task. The natural choice is to use more than two parental lines in a mating design. Limited investigations have shown that QTL mapping by using multiple line crosses has several advantages. First, it can handle multiple alleles at any locus and thus has a wider statistical inference space than a single line cross. Second, the use of mating designs with an increased number of parents is more efficient than the use of only one F_2-like FS family in outbred populations. This is because the variance attributable to the QTL is better estimated as the number of parents increases (Muranty 1996). However, with a fixed number of individuals, there is an optimal allocation between the number of families and the number of individuals per family where QTL mapping reaches its maximum power and minimum estimation error (Soller and Genizi 1978). Third, a joint test for multiple line crosses is more powerful than a test considering crosses independently (Rebai and Goffinet 1993).

For convenience of presentation, the consensus method of QTL mapping described above assumes that a dominance effect is absent. We now discuss how to relax this assumption using F_2 family data as an example.

The model can be described by $y_j = \mu + g_j + \delta_j + s + e_j$, where δ_j is the dominance deviation of a putative QTL with mean 0 and variance σ^2_δ and s is a family-specific effect distributed as $N(0, \sigma^2_s)$. Note that σ^2_s consists of a portion of the additive polygenic variance and dominance variance (Xu 1996b). The covariance between y_i and y_j is $\text{Cov}(y_i, y_j) = \pi_{ij}\sigma^2_g + \Delta_{ij}\sigma^2_\delta + \sigma^2_s$, where Δ_{ij} is a descent measure indicating whether i and j share both alleles IBD (Harris 1964; Cockerham 1983). The coefficient of the dominance variance (Δ_{ij}) is determined by $\Delta_{jj} = 1$ and

$$\Delta_{ij} = \begin{cases} 1 & \text{for } QQ\text{-}QQ,\ Qq\text{-}Qq\text{-} \text{ or } qq\text{-}qq \\ 0 & \text{otherwise} \end{cases}.$$

The conditional expectation of Δ_{ij} given the marker information is computed using $\hat{\Delta}_{ij} = E(\Delta_{ij}|I_M) = p_{i2}p_{j2} + p_{i1}p_{j1} + p_{i0}p_{j0}$. In our simulation experiments we consider only additive effects at the QTL.

In this study, we have used a random model methodology to detect a QTL. Essentially, the theoretical basis of the random model is based on the variability of the IBD proportion shared by sibs at the putative QTL. For example, the variance of the IBD proportions are, on average, 1/8 for noninbred full sibs and 3/16 for siblings from a BC. In contrast, the variance of the IBD proportion is 1/4 for siblings from F_2. This difference results in QTL mapping using F_2 families being more powerful than BC or FS families, while both BC and FS families have similar test statistics and powers.

The F_2 and BC mating designs require the availability of inbred lines. If no such lines exist in nature, one must develop such lines, and this is costly and time consuming. In this case, the FS mating design is more preferable than F_2 and BC. In self-incompatible organisms, the FS mating design is the only choice.

The consensus QTL mapping proposed here is a general approach for combining or updating data. By setting the relevant Π and **A** matrices on a family-by-family basis, families from all types of mating designs can be sampled at different locations or different laboratories may be combined. Alternatively, data can be combined vertically; that is, data collected in the same laboratory but at different times can be pooled through the consensus mapping strategy.

This research was supported by the National Institutes of Health Grant GM-55321-01 and the United States Department of Agriculture National Research Initiative Competitive Grants Program 97-35205-5075 to S.X.

LITERATURE CITED

Basten, C. J., B. S. Weir and Z. B. Zeng, 1997 *QTL Cartographer.* Version 1.12, North Carolina State University, Raleigh, NC.

Cockerham, C. C., 1983 Covariances of relatives from self-fertilization. Crop Sci. **23:** 1177–1180.

Edwards, M. D., C. W. Struber and J. F. Wendel, 1987 Molecular marker-facilitated investigations of quantitative trait loci in maize. I. Number, distribution, and types of gene action. Genetics **116:** 113–125.

Falconer, D. S., and T. F. C. Mackay, 1996 *Introduction to Quantitative Genetics*, Ed. 4, Longman, NY.

Fulker, D. W., S. S. Cherny and L. R. Cardon, 1995 Multipoint interval mapping of quantitative trait loci using sib pairs. Am. J. Hum. Genet. **56:** 1224–1233.

Gessler, D. D. G., and S. Xu, 1996 Using the expectation or the distribution of the identity by descent for mapping quantitative trait loci under the random model. Am. J. Hum. Genet. **59:** 1382–1390.

Goldgar, D. E., 1990 Multipoint analysis of human quantitative genetic variation. Am. J. Hum. Genet. **47:** 957–967.

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

Harris, D. L., 1964 Genotypic covariances between inbred relatives. Genetics **50:** 1319–1348.

Haseman, J. K., and R. C. Elston, 1972 The investigation of linkage between a quantitative trait and a marker locus. Behav. Genet. **2:** 3–19.

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

Jiang, C., and Z. B. Zeng, 1996 Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics **140:** 1111–1127.

Kempthorne, O., 1955 The correlation between relatives in inbred populations. Genetics **40:** 681–691.

Kruglyak, L., and E. S. Lander, 1995 Complete multipoint sib-pair analysis of qualitative and quantitative traits. Am. J. Hum. Genet. **57:** 439–454.

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

Lincoln, S. E., M. J. Daly and E. S. Lander, 1993 *Mapping Genes Controlling Quantitative Traits Using MAPMAKER/QTL Version 1.1: A Tutorial and Reference Manual.* Whitehead Institute, Cambridge, MA.

Malécot, G., 1948 Les mathématiques de l'hérédité. Masson et Cie, Paris.

Martínez, O., and R. N. Curnow, 1992 Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theor. Appl. Genet. **85:** 480–488.

Muranty, H., 1996 Power of tests for quantitative trait loci detection using full-sib families in different schemes. Heredity **76:** 156–165.

Olson, J. M., 1995 Robust multipoint linkage analysis: an extension of the Haseman-Elston method. Genet. Epidemiol. **12:** 177–193.

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

Soller, M., and A. Genizi, 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.

Tinker, N. A., and D. E. Mather, 1995 *MQTL: Software for Simplified Composite Interval Mapping of QTL in Multiple Environments.* J. Quant. Trait Loci **1** (on line). WWW: http://probe.nalusda.gov:8000/otherdocs/jqtl/jqtl1995-02/jqtl16r2.html.

Van Ooijen, J. W., and C. Maliepaard, 1996 *MapQTL Version 3.0: Software for the Calculation of QTL Positions on Genetic Maps.* Plant Genome IV. San Diego, CA.

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

Xu, S., 1996b Computation of the full likelihood function for estimating variance at a quantitative trait locus. Genetics **144:** 1951–1960.

Xu, S., and W. R. Atchley, 1995 A random model approach to interval mapping of quantitative trait loci. Genetics **141:** 1189–1197.

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

Communicating editor: T. F. C. Mackay