DOI: 10.1534/genetics.104.033746

Improving Quantitative Trait Loci Mapping Resolution in Experimental Crosses by the Use of Genotypically Selected Samples

Zongli Xu,* Fei Zou† and Todd J. Vision*,[1]

**Department of Biology and †Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina 27599*

Manuscript received July 19, 2004
Accepted for publication January 15, 2005

ABSTRACT

One of the key factors contributing to the success of a quantitative trait locus (QTL) mapping experiment is the precision with which QTL positions can be estimated. We show, using simulations, that QTL mapping precision for an experimental cross can be increased by the use of a genotypically selected sample of individuals rather than an unselected sample of the same size. Selection is performed using a previously described method that optimizes the complementarity of the crossover sites within the sample. Although the increase in precision is accompanied by a decrease in QTL detection power at markers distant from QTL, only a modest increase in marker density is needed to obtain equivalent power over the whole map. Selected samples also show a slight reduction in the number of false-positive QTL. We find that two features of selected samples independently contribute to these effects: an increase in the number of crossover sites and increased evenness in crossover spacing. We provide an empirical formula for crossover enrichment in selected samples that is useful in experimental design and data analysis. For QTL studies in which the phenotyping is more of a limiting factor than the generation of individuals and the scoring of genotypes, selective sampling is an attractive strategy for increasing genome-wide QTL map resolution.

THE precision with which a quantitative trait locus (QTL) can be located in a genome-wide survey can be critical to the time, expense, and probability of success of subsequent positional cloning (Remington *et al.* 2001). The precision of QTL position estimates, sometimes referred to as *map resolution*, can be frustratingly low in experimental crosses (Nadeau and Frankel 2000). Map resolution can be affected by the method used for statistical analysis (*e.g.*, Zeng 1994), but there is a limit to the extent to which analysis can compensate for poor experimental design. Factors affecting map resolution that can be controlled during experimental design include the number of individuals in the sample and the nature of the genetic cross (Mackay 2001). These two factors affect resolution, at least in part, by governing the sample of meoitic crossover sites occurring between markers and QTL. This suggests that it might be possible to increase map resolution by directly selecting those individuals to phenotype on the basis of observable crossover events (see also Ronin *et al.* 2003).

A method for choosing mapping samples on the basis of observable crossovers has previously been proposed, although not in the context of QTL mapping. *Selective mapping* is an experimental design strategy for genome-wide, high-density linkage mapping of molecular markers in experimental crosses (Vision *et al.* 2000). In the first step, a limited number of *framework markers* are genotyped in a large *base population*. From the resultant genotype matrix, individuals are selected that collectively provide good coverage (as defined below) of the crossover sites in the larger population. Large numbers of secondary markers can then be genotyped on the selected sample and their positions inferred relative to the previously mapped framework markers. The resolution obtained with selective mapping for a given investment of genotyping effort can considerably exceed that obtained using an equivalently sized random sample of individuals. The gain is most dramatic for small genomes (<1000 cM).

In principle, a similar strategy could also be applied to QTL mapping with the aim of maximizing the resolution obtained when only a limited number of permanent genotypes can be propagated or phenotyped. Although it is generally undesirable to use a small sample for QTL mapping when a larger one is available due to the limited QTL detection power and inaccurate estimates of genetic effect sizes obtained with small samples (Beavis 1998), practical constraints on sample size are commonplace. For instance, the same genotypes may need to be phenotyped at multiple sites in multiple years, and financial constraints may set an upper limit to the number of genotypes that can be used (see also Jin *et al.* 2004).

We refer to the choice of individuals for phenotyping on the basis of their genotypes, namely the inferred positions of crossover sites, as *selective sampling*. Here

[1]*Corresponding author:* Department of Biology, Campus Box 3280, University of North Carolina, Chapel Hill, NC 27599.
E-mail: tjv@bio.unc.edu

we study the statistical consequences of using such a selected sample. In particular, we use simulations to quantify the effects of selective sampling on QTL detection power, sensitivity, and specificity and on the precision of estimated QTL positions.

METHODS

Crossover distribution: *Simulation of genotypes:* Since the principal difference between a selected and a random sample is in the frequency and spatial arrangement of crossover sites, we studied the effect of selective sampling on these features of the sample. It is important to note, however, that there may be other differences between selected and random samples that are not examined here. For instance, we have observed that selective sampling generally reduces the variance in genotypic proportions at each locus (Doganlar *et al.* 2002; T. J. Vision and D. G. Brown, unpublished results).

For most experiments, we used a base population of diploid recombinant inbred lines (RILs), each line derived by recurrent selfing of a unique member of an F_2 population. Backcross recombinant inbred line (BRIL) and doubled haploid (DH) base populations were also studied where indicated. Each individual was assumed to have a single linkage group of length L cM, where L varied according to the experiment. At a given marker density, marker positions were assigned with even spacing or uniformly at random. The expected number of crossovers in each individual was calculated as $z_r = 2 \times (1/100)L$, since the cumulative number of crossovers is approximately twofold higher in a late-generation selfed F_2 RIL than in an F_2 individual. (Haldane and Waddington 1931). The realized number z_i of crossovers in individual i was simulated as a Poisson random variable with expectation z_r. The locations of the crossovers were drawn from a uniform distribution on $(0, L)$ conditional on z_i.

Sampling from a base population: From a base population, n individuals were selected either at random or using selective sampling with the sum of squares of bin lengths (SSBL) objective function (Vision *et al.* 2000). The objective function can be understood as follows. A *bin* is defined on a sample of individuals as an interval along the linkage group within which there are no crossovers in any sampled individual and bounded on either side either by a crossover in at least one individual or by the end of a linkage group (Vision *et al.* 2000). By minimizing the sum of the squares of the bin lengths, we obtain a sample of individuals in which crossovers are more frequent, and the distance between them less variable, than in a random sample. Previous work has shown that SSBL behaves well even when framework markers are widely spaced and the genotyping error rate is high (Vision *et al* 2000).

The proportion of individuals from the base population present in either the random or the selected sample is termed the *sample fraction*, symbolized by f. Selective sampling was performed using the MapPop software package (http:/www.bio.unc.edu/faculty/vision/lab/mappop/).

Crossover enrichment: Use of the SSBL objective function is expected to lead to an enrichment of crossovers in a selected sample. The total number of crossovers in the selected sample relative to that expected in a random sample of the same size is referred to as the *crossover enrichment* (CE). CE was measured for selected samples drawn from simulated base populations in which the sample fraction, marker density, and map length were varied.

Pseudointerference: In addition to crossover enrichment, use of the SSBL objective function is expected to produce bin lengths that are less variable than those in a random sample. This phenomenon, which we call *pseudointerference*, differs from standard crossover interference in that it arises from selection of crossovers present in different individuals rather than from biological interference among crossovers during meiosis.

We use the Karlin map function (Karlin 1984) to quantify the magnitude of pseudointerference. A map function models the relationship between m, the genetic distance in morgans, and r, the recombination rate. In random samples, the positions of crossover sites should be well fit by the Haldane map function, which assumes sites are spaced uniformly at random. When there is positive or negative interference among crossovers, a different map function is needed.

For our purposes, the key property of the Karlin map function,

$$m = 0.5N\left[1 - (1 - 2r)^{1/N}\right], \qquad (1)$$

is that it allows for variable interference by adjustment of the N parameter. When N is large (>5), interference is negligible and the Karlin and Haldane map functions converge. Thus, the value of N allows us to evaluate the consequences of different sampling strategies on the level of pseudointerference and to study the consequences of differing levels of pseudointerference on QTL estimation.

In addition, when QTL analysis is done using interval mapping (Lander and Botstein 1989), it is necessary to specify the map function to accurately estimate the position of a QTL relative to its flanking markers. Another motivation for this aspect of the study is thus to provide some guidance as to what map function would be appropriate to use for interval mapping on selected samples.

We fit the Karlin map function to recombination data from selected samples and estimated the magnitude of N. Map positions were first rescaled by CE. Let the number of crossovers in the interval from 0 to i cM in individual j be denoted x_{ij}. For an interval to be considered recombinant, we required that $\mathrm{mod}(x_{ij}, 2) = 1$ (*i.e.*, there must be an odd number of crossovers in the interval). The recombination rate for the ith interval was

calculated as $\Sigma_j \mathrm{mod}(x_{ij}, 2)/S$, where S is the number of individuals in the sample. Equation 1 was then fit by nonlinear regression using SAS (Cary, NC) to data that included all intervals of integral length starting at 0. While the intervals were not strictly independent, obtaining independent intervals with selective mapping would have been computationally prohibitive, and the large sample size (10,000 individuals) ensured that estimates of N were stable.

CE-adjusted random samples: Differences in QTL estimates obtained with a selected sample *vs.* a random sample of the same size may be due to CE and/or pseudointerference. To separately investigate these two factors, we employed *CE-adjusted random samples* in which the expected number of crossovers was equal to $z_s = z_r \times$ CE. The realized number z_i of crossovers in individual i was simulated as a Poisson random variable with expectation z_s and the locations of the crossovers were drawn from a uniform distribution on $(0, L)$ conditional on z_i. Thus, CE-adjusted samples are free of pseudo-interference. By comparison of CE-adjusted random samples with varying levels of CE, the effects of CE alone can be isolated. Alternatively, by comparison of a selected sample with a CE-adjusted random sample of equivalent CE, the effects of pseudointerference alone can be isolated.

QTL experiments: *Simulation of QTL:* To study the effects of selected sampling on QTL analysis, loci affecting a quantitative trait were added to the base populations simulated above. The *additive effect* was parameterized as $-a$ and a for QTL genotypes qq and QQ, respectively. Environmental deviations were drawn from a standard normal distribution. To calculate the heritability h^2, we used the fact that the additive genetic variance contributed by a QTL is $V_g = a^2$ in an F_2 RIL population. For simulations where additive effects were considered to be random variables, they were sampled from a gamma (1, 2) distribution (Zeng 1992). In simulations involving multiple QTL, the loci were constrained to be spaced at least 100 cM apart.

QTL analysis: QTL analysis was performed via single-marker analysis as implemented in QTL Cartographer v. 1.16 (Basten *et al.* 2002). At each marker, the following model was fit: $y_i = b_0 + b_1 x_i + e$ ($i = 1, 2, \ldots,$ total number of individuals), where y_i is the phenotype of the ith individual, x_i is the indicator variable for the marker genotype, and the error e is assumed to be normally distributed with mean 0 and variance e^2 (Basten *et al.* 2002). A likelihood-ratio (LR) test statistic was computed to test the null hypothesis H_0: $b_1 = 0$ *vs.* the alternative hypothesis H_1: $b_1 \neq 0$. To estimate the genome-wide significance threshold for the LR, data were simulated under the null hypothesis that no QTL was present. The $(1 - \alpha)100$th percentile of the maximum LR was used as the threshold to control the genome-wide type I error at α (after Dupuis and Siegmund 1999).

To estimate the effect of selective sampling on detection power, QTL analysis was performed on samples differing only in crossover enrichment (with uniformly random crossover sites) or on selected *vs.* CE-adjusted random samples. These two comparisons allow us to separate the effects of crossover enrichment and pseudointerference on the power to detect a QTL and on the precision with which it is located. The *detection power* was defined to be the probability that the maximum LR at any marker or position exceeded the significance threshold for a type I error of $\alpha = 0.05$.

To calculate sensitivity and specificity of QTL detection, we adopted the following conventions. A *peak* was defined as a point where the LR value exceeded both the significance threshold and the LR values of adjacent points (or point, if at the end of a linkage group). The *range* of the peak was taken to be the interval on either side of the peak bounded by the end of the linkage group or by that point closest to the peak with an LR value below the threshold, whichever came first. The position of the highest LR peak within the range was taken to be the QTL position. If the range bracketed a true QTL position, then the peak was counted as a true positive (TP); if not, it was counted as a false positive (FP). If a true QTL position was not bracketed by the range of any peak, then it was counted as a false negative (FN). Sensitivity (Sn) and specificity (Sp) were then calculated as follows:

$$\mathrm{Sn} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN})$$

$$\mathrm{Sp} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP}).$$

For comparison of estimated and true genetic effects, we took the estimate of a at the peak to be the estimate of the genetic effect; when no QTL was detected, that replicate was discarded.

QTL mapping resolution was measured using two different methods. The first was to take the difference between the 2.5 and 97.5% quantiles of the estimated QTL position among a set of independent QTL populations sharing a fixed QTL position and effect size (Darvasi and Soller 1997). To approximate an infinite number of markers, we used an even density of 10 markers per centimorgan for QTL analysis and took the marker with the highest LR in each replicate to be the estimated QTL position. The second method was to calculate the *1-LOD drop support interval*, defined as the distance between the two points on either side of the peak where the base-10 logarithm of the LR [the log of odds (LOD) score] declined by 1 unit. In the multiple-QTL simulations, only TP peak ranges were used for this calculation.

RESULTS AND DISCUSSION

Crossover enrichment: We first investigated the nature of CE in selected samples. We found CE to be inversely related to the sample fraction, marker spacing, and map

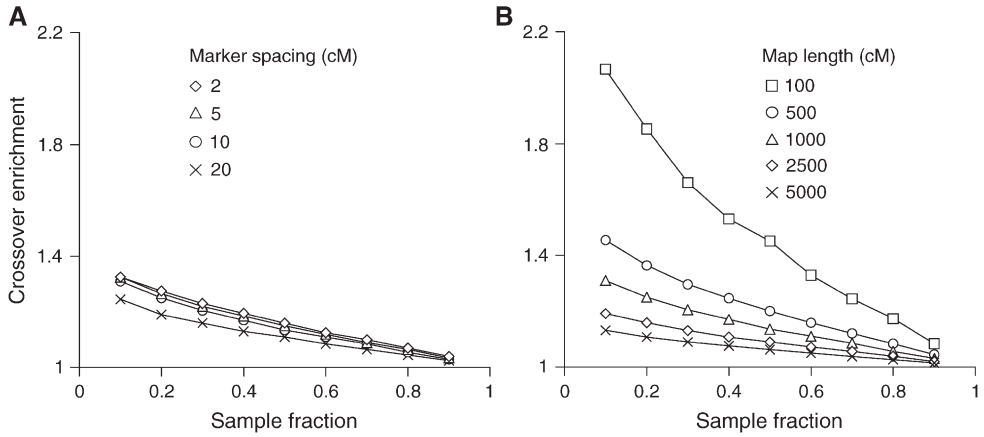

FIGURE 1.—Crossover enrichment after selection of RILs with varied framework marker intervals (A) and chromosome lengths (B). Base population size is 500; $L = 1000$ cM in A. Marker interval is 10 cM in B. Each point is the average of 10,000 individuals.

length (Figure 1). These results make intuitive sense. Since the sample fraction is itself an inverse measure of the strength of selective sampling, a small sample fraction is expected to result in more extreme CE than a larger (and more nearly random) one. Marker spacing affects the precision with which crossovers are detected in the base population as well as the number of framework marker intervals that contain double crossovers. Since double crossovers between adjacent markers are invisible to the selection algorithm, one expects a decline in CE as the framework marker spacing increases. We found that when the marker spacing was already $< \sim$10 cM, CE was fairly insensitive to variability in marker spacing, reflecting the rarity of double-recombinant intervals in such maps. The inverse relationship of CE to map length agrees with previous studies showing that the effectiveness of selective sampling declines with map length (VISION *et al.* 2000; BROWN 2001).

We noted that when the marker spacing was $\leq \sim$10 cM, CE was nearly inversely proportional to the square root of the map length, L, and the sample fraction, f. Remarkably, CE could be very closely predicted by the empirical formula

$$\mathrm{CE} = 1 + 0.5(1 - f)\sqrt{A/L}, \qquad (2)$$

where A is a constant that is determined by the type of base population. For an F_2 RIL population, $A \approx 500$. For BRIL and DH populations, $A = 750$ and 1200, respectively. Within the particular parameter range that we explored ($L = [100, 2500]$, $f = [0.1, 0.9]$, and marker spacing from 1 to 10 cM), nonlinear regression using Equation 2 yielded R^2 values of 0.96, 0.98, and 0.98 for F_2 RIL, BRIL, and DH, respectively. Note that simulations were excluded when the map length was short (100 cM) and the sample fraction was small (0.1), as these gave unusually large deviations. Also, Equation 2 was derived using simulated data in which markers were evenly spaced; CE was found to be smaller when markers were distributed uniformly at random but the difference was slight (<0.1).

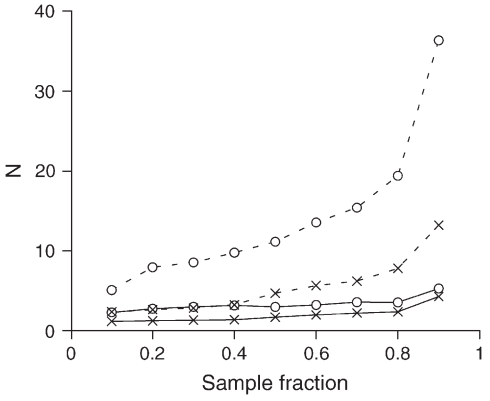

FIGURE 2.—Best-fit Karlin map function parameter N as a function of sample fraction. $L = 100$ cM (solid line) or 500 cM (dashed line). Marker spacing was 5 (○) or 20 cM (×). Each point is obtained from 10,000 individuals.

Pseudointerference: We measured the magnitude of pseudointerference in selected samples by fitting the Karlin map function to simulated recombination data. A large value of N (>5) indicates that pseudointerference is negligible. We examined base populations of 500 individuals with $L = [100, 5000]$, $f = [0.1, 0.9]$, and marker spacings of 5–20 cM. The best-fit parameter of N was found to be sensitive to all three factors: map length, sample fraction, and marker spacing (Figure 2). The map function in Equation 1 was fit to recombination data from 10,000 individuals for each parameter combination (where the individuals came from multiple selected samples). In all cases, the R^2 goodness-of-fit value was >0.94; for most parameter combinations, it was >0.97. Pseudointerference was greater when the sample fraction was small and the map length was short, consistent with the behavior of CE described above. A more surprising result was that pseudointerference was more pronounced when markers were more distantly spaced.

Further analysis offered a potential explanation for the relationship between marker spacing and pseudointerference. We hypothesized that since marker intervals bearing double crossovers are not distinguishable from those bearing no crossovers at all during the selection

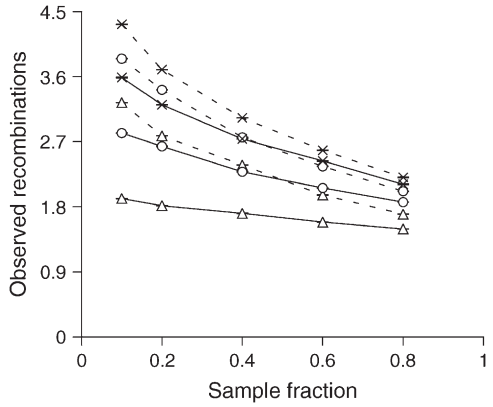

FIGURE 3.—Number of recombinations per individual in selected (dashed lines) and CE-adjusted random (solid lines) samples; $L = 100$ cM; marker spacing was 5 (×), 10 (○), or 20 cM (△). Each value is the average of 10,000 individuals.

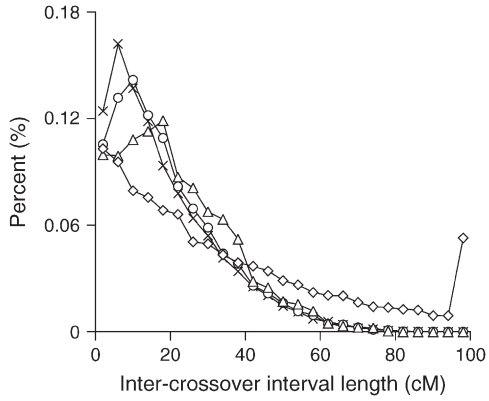

FIGURE 4.—The length distribution of intercrossover intervals. Selective sampling was done using marker spacings of 5 (×), 10 (○), or 20 cM (△); random samples are shown by diamonds (◇); $L = 100$ cM; base population was 500 individuals; $f = 0.1$. Numbers were calculated out of the total number of intercrossover intervals present in 10,000 individuals. The bin shown at length x represents the interval $(x - 4, x]$. The spike in the rightmost bin of the series is due to the occurrence of an appreciable number of chromosomes without crossovers at that marker spacing.

phase, there would be a bias to select individuals that have only one crossover within each marker interval. This would, in turn, lead to crossovers that are spaced relatively evenly. The effect would be most pronounced when markers are sparse because closely spaced crossover sites in the base population would be less likely to result in an observable recombination and thus would be underrepresented in the selected sample.

In support of this hypothesis, we found that there were more observed recombinations in selected samples than in random samples even when random samples were adjusted to have the same expected number of crossovers (Figure 3). Another way to understand the underlying effect of selection on crossovers and recombinations is to note the change in the distribution of intercrossover intervals within an individual in a selected sample. Remarkably, the mode in the intercrossover interval length occurs at the same centimorgan distance as the marker spacing used for selective sampling (Figure 4).

The distribution of crossover sites in our simulated base populations differs from that in a true RIL population in one important respect. In a true RIL population, crossovers accumulate over multiple generations as each recombinant inbred line approaches homozygosity. Only crossovers in heterozygous segments lead to recombinations. This process leads to negative interference: double recombinations will occur within short intervals more often than under our assumption of uniformly distributed crossovers. Nonetheless, we have found that explicitly simulating the process of selfing over multiple generations to produce more realistic RIL genotypes does not have an appreciable affect on the comparison between random and selected samples (results not shown).

QTL analysis: We examined the effect of selective sampling on QTL detection power for simulated base populations in which one QTL was segregating at an equal distance from two flanking markers. The threshold and power for a given experiment were determined empirically, as described in METHODS. We conducted two experiments to quantify the effects of CE and pseudointerference separately from one another.

In the first experiment, we compared CE-adjusted random samples with varying levels of CE under different marker spacings. This comparison allows us to evaluate the effect of CE alone since these samples are free of pseudointerference. We found that power was inversely related to CE but that the relationship was nearly flat when the marker spacing was <5 cM, corresponding to a marker-QTL distance of ≤ 2.5 cM (Figure 5A). Even for CE = 1, the detection power was inversely related to the marker spacing. This indicates that the increasing distance between the flanking markers and the QTL was more important to detection power than the variation in the significance threshold, which was lower for the more widely spaced markers.

In the second experiment, we compared the power between selected samples and CE-adjusted random samples, where the CE did not differ between the two types of samples. This was done to measure the effect of pseudointerference alone. Here, we also found that the QTL-detection power in the selected samples was less than that in CE-adjusted random samples (Figure 5B). Thus, QTL detection power is affected by both CE and pseudointerference.

For experimental design purposes, an investigator would like to know how dense markers need to be, when selecting a predetermined fraction f, to obtain the same QTL detection power as that of an equivalently sized random sample. To study this, we simulated populations in which QTL position was random with respect to the markers and compared random samples to selected

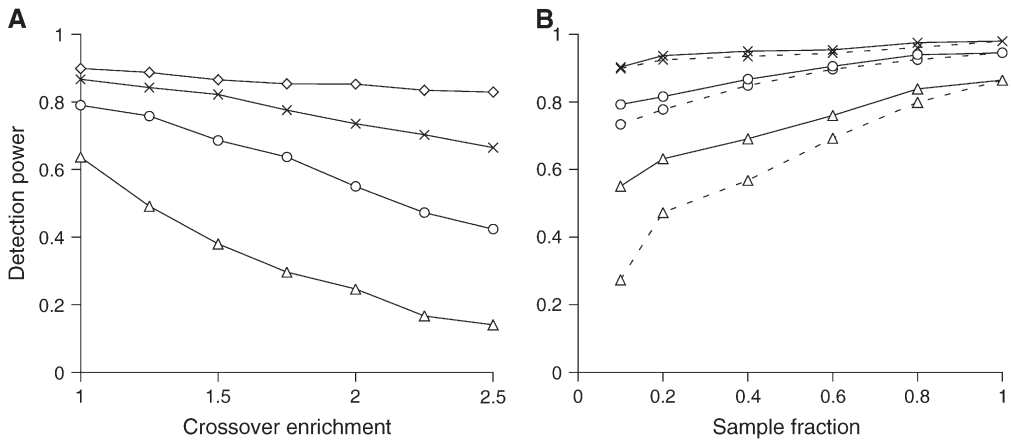

FIGURE 5.—(A) Detection power in samples differing only in crossover enrichment. Lines correspond to marker spacings of 1 (◇), 5 (×), 10 (○), and 20 cM (△). $L = 1000$ cM. (B) Detection power in selected (dotted lines) and CE-adjusted random (solid lines) samples is shown. Lines correspond to marker spacings of 5 (×), 10 (○), and 20 cM (△). $L = 100$ cM. In both A and B, a single QTL (with additive effect of $a = 0.5$ and heritability of $h^2 = 0.2$) was located equidistant from the two centermost markers. Sample size was 100. Each point obtained was from 5000 replicates.

samples differing in both CE and pseudointerference (Figure 6). In the case of a 1000-cM map, 51 markers in a random sample had equivalent power to an $f = 0.5$ selected sample with 59 markers or an $f = 0.1$ selected sample with 72 markers. Thus, a modest increase in marker density can counteract the effects of increased CE and pseudointerference even under extreme selection.

We then measured the effect of selective sampling on the sensitivity and specificity of QTL detection. Both single-QTL and multiple-QTL simulations show that sensitivity is slightly reduced and specificity is slightly increased in selected samples when markers were widely spaced, but that the differences were relatively small when markers were dense (Figure 7). While the reduced sensitivity of selected samples is not surprising in light of the detection power results discussed above, the reason for increased specificity of selected samples is not clear. One thing we can conclude is that it is not due to CE, because CE-adjusted samples do not differ from random samples in their specificity (results not shown). However, this does not necessarily implicate pseudointerference as the cause of the increased specificity.

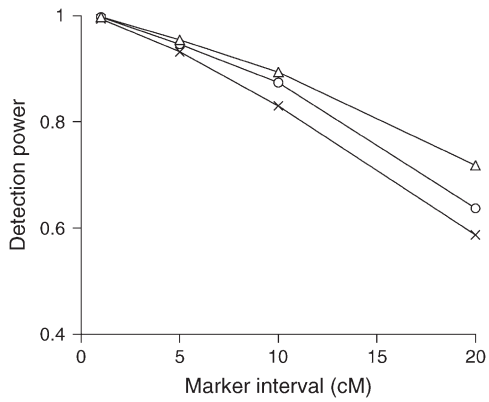

FIGURE 6.—Detection power in random and selected samples for a trait underlain by a single QTL with random position and effect size; $L = 1000$ cM; $h^2 = 0.2$; markers are distributed uniformly at random; $f = 0.2$ (×) or 0.5 (○), corresponding to base population sizes of 500 and 200, respectively. Random samples (of size 100) are denoted by △. Each point was obtained from 5000 replicates.

Estimates of QTL additive-effect size are known to be upwardly biased due to the so-called *Beavis effect* (BEAVIS 1998). The lower the QTL detection power, the greater the bias. Since QTL detection power is reduced by selective sampling, we would expect the effect size estimates to be inflated over that obtained for random samples. We did, in fact, observe this trend, but the added effect of selection was relatively small except in a few cases where selection was unrealistically intense (results not shown).

QTL mapping resolution: Since QTL detection power is reduced considerably for distant markers, but less so for nearby ones, we hypothesized that the confidence or support intervals to which QTL are located in selected samples might be smaller than those in random ones. To test this hypothesis, we examined the effect of selective sampling on two different measures of QTL mapping resolution: the distribution of QTL peak positions among independent simulations (DARVASI and SOLLER 1997) and the 1-LOD drop support interval. For both measures, QTL were located with substantially greater precision in selected samples. The 1-LOD drop support interval results are shown in Figure 8A for simulated populations in which one QTL was segregating. The increase in QTL mapping resolution was similar across a range of additive-effect sizes. Precision was substantially better when selection was done with a shorter map. Precision was also improved, but not as dramatically, by selection using densely spaced markers.

Selective sampling also increased map resolution for simulations in which five QTL were segregating (Figure

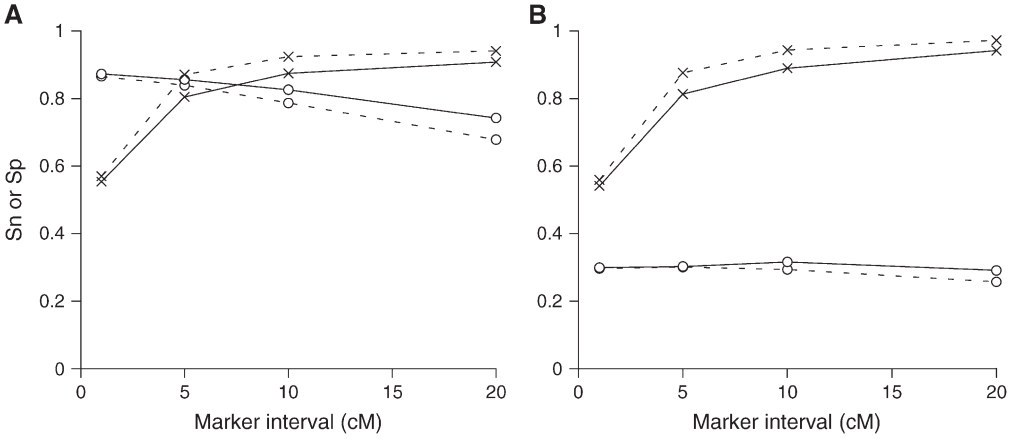

FIGURE 7.—Sensitivity (○) and specificity (×) in random (solid lines) and selected (dashed lines) samples. (A) One QTL with $h^2 = 0.2$. (B) Five QTL with random effect sizes having an overall $h^2 = 0.5$. For both A and B, $L = 1000$ cM, base population was 500 individuals; $f = 0.2$. QTL positions were random although constrained to be >100 cM apart from each other in the five QTL simulations. Each point was obtained from 5000 replicates.

8B). For these simulations, the same markers were used for selective sampling and QTL analysis to increase the realism of the experiment. The absolute difference in the size of the support interval is considerably greater when the markers are sparse, although the proportional difference in the size of the support interval is relatively insensitive to marker spacing. In sum, selective mapping appears to effectively increase map resolution even for a complex trait and even when markers are widely spaced. Furthermore, this conclusion is robust to moderate epistasis (results not shown).

Conclusions: In summary, we have found that the probability of detecting a QTL is somewhat diminished in a selected sample relative to a random one when the QTL is far from a marker. But since this reduction in power disappears when the distance between the marker and the QTL approaches zero, the width of the confidence interval surrounding the QTL is narrowed, resulting in a more precise estimate of QTL location. The increased marker density needed to take advantage of this increased resolution is fairly modest. Additionally, specificity in QTL detection is slightly higher in selected samples.

One reason for the difference between selected and random samples is the increased number of crossover sites or CE. We found that a simple formula can be used to predict CE in a selected sample when the marker spacing is dense (≤10 cM). The value of CE thus obtained can be used to adjust the genetic map prior to statistical analysis of QTL.

A second factor affecting QTL detection power and resolution is the reduced variability in intercrossover interval length within each individual, or what we have termed pseudointerference.

QTL mapping is widely used a first step in the determination of the molecular basis of phenotypic variation relevant to agriculture, medicine, ecology, and evolutionary biology (MACKAY 2001). Since, for most organisms, QTL intervals can encompass tens to hundreds, even thousands, of genes, the major effort in cloning a QTL is the work required to refine the estimated position once a genome-wide survey has been completed (REMINGTON *et al.* 2001). Thus the feasibility of QTL mapping hinges, in part, on the precision with which QTL can be located during this initial scan. A number of strategies are available for increasing QTL map reso-

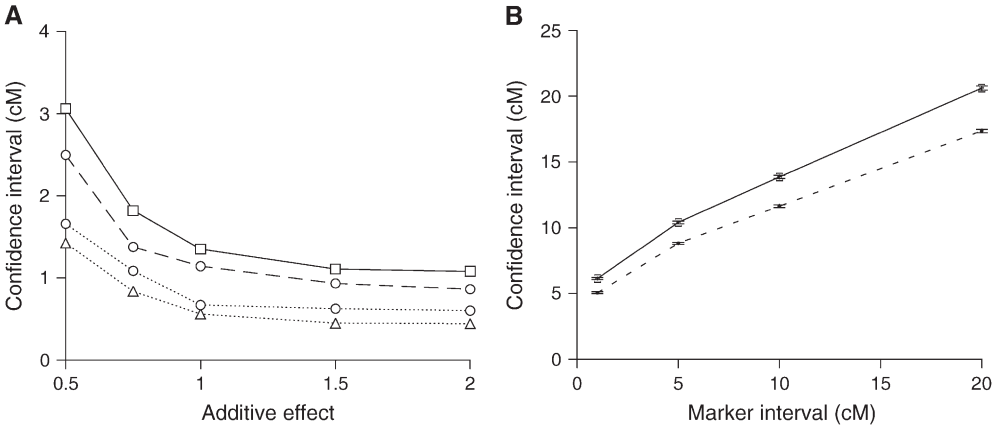

FIGURE 8.—Confidence intervals obtained using the 1-LOD drop method for QTL in random (solid lines) and selected samples ($f = 0.1$, dashed or dotted lines). (A) Samples segregating for one QTL at a fixed position (50 cM from the beginning of the map) with a fixed environmental variance of 1; $L = 100$ (dotted lines) or $L = 1000$ cM (dashed line); marker spacing during selection was 1 (△) or 10 cM (○); QTL analysis was performed only on the first 100 cM of the map. (B) Five QTL with random positions, but constrained to be ≥100 cM from each other; total heritability of $h^2 = 0.5$; $L = 1000$ cM; base population of 100, random sample (solid line) and selected sample ($f = 0.2$, dashed line). Shown are means and standard errors of 5000 replicates.

lution, among them the use of large populations and multiple generations of intercrossing. We have shown here that selective sampling is an additional strategy that can be used to increase QTL map resolution over a random sample of the same size. Selective sampling is suitable for situations in which the number of individuals that can be genotyped is large but the number that can be phenotyped is not. This situation is likely to become increasingly common as large, permanent, genotyped mapping populations are produced for various model organisms and as QTL mapping is applied to ever more subtle and complex traits (*e.g.*, Threadgill *et al.* 2002).

We acknowledge the helpful comments of two anonymous reviewers. This work has been supported by National Science Foundation grant DBI-0110069 to T.J.V.

LITERATURE CITED

Basten, C. J., B. S. Weir and Z-B. Zeng, 2002 *QTL Cartographer: A Reference Manual and Tutorial for QTL Mapping*. (http://statgen.ncsu.edu/qtlcart/manual/).

Beavis, W. D., 1998 QTL analyses: power, precision and accuracy, pp. 145–161 in *Molecular Analysis of Complex Traits*, edited by A. H. Paterson. CRC Press, Boca Raton, FL.

Brown, D. G., 2001 A probabilistic analysis of a greedy algorithm arising from computational biology. Proceedings of the 12th Annual ACM-SIAM Symposium on Discrete Algorithms, Washington, DC, pp. 206–207.

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

Doganlar, S., A. Frary, H.-M. Ku and S. D. Tanksley, 2002 Mapping quantitative trait loci in inbred backcross lines of *Lycopersicon pimpinellifolium* (LA1589). Genomics **45:** 1189–1202.

Dupuis, J., and D. Siegmund, 1999 Statistical methods for mapping quantitative trait loci from a dense set of markers. Genetics **151:** 373–386.

Haldane, J. B. S., and C. H. Waddington, 1931 Inbreeding and linkage. Genetics **16:** 357–374.

Jin, C., L. Lan, A. D. Attie, G. A. Churchill, D. Bulutuglo *et al.*, 2004 Selective phenotyping for increased efficiency in genetic mapping studies. Genetics **168:** 2285–2293.

Karlin, S., 1984 Theoretical aspects of genetic map functions in recombination processes, pp. 209–228 in *Human Population Genetics: The Pittsburgh Symposium*, edited by A. Chakravarti. Van Nostrand Reinhold, New York.

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

Mackay, T. F., 2001 The genetic architecture of quantitative traits. Annu. Rev. Genet. **35:** 303–339.

Nadeau, J. H., and D. Frankel, 2000 The roads from phenotypic variation to gene discovery: mutagenesis versus QTL. Nat. Genet. **25:** 381–384.

Remington, D. L., M. C. Ungerer and M. D. Purugganan, 2001 Map-based cloning of quantitative trait loci: progress and prospects. Genet. Res. **78:** 213–238.

Ronin, Y., A. Korol, M. Shtemberg, E. Nevo and M. Soller, 2003 High-resolution mapping of quantitative trait loci by selective recombinant genotyping. Genetics **164:** 1657–1666.

Threadgill, D. W., K. W. Hunter and R. W. Williams, 2002 Genetic dissection of complex and quantitative traits: from fantasy to reality via a community effort. Mamm. Genome **13:** 175–178.

Vision, T. J., D. G. Brown, D. B. Shmoys, R. T. Durrett and S. D. Tanksley, 2000 Selective mapping: a strategy for optimizing the construction of high-density linkage maps. Genetics **155:** 407–420.

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

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

Communicating editor: A. H. Paterson