

# Poor Performance of Bootstrap Confidence Intervals for the Location of a Quantitative Trait Locus

Ani Manichaikul,\* Josée Dupuis,<sup>†</sup> Saunak Sen<sup>‡</sup> and Karl W. Broman<sup>\*,1</sup>

\*Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland 21205, <sup>†</sup>Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts 02118 and <sup>‡</sup>Department of Epidemiology and Biostatistics, University of California, San Francisco, California 94107

Manuscript received March 24, 2006  
Accepted for publication June 15, 2006

## ABSTRACT

The aim of many genetic studies is to locate the genomic regions (called quantitative trait loci, QTL) that contribute to variation in a quantitative trait (such as body weight). Confidence intervals for the locations of QTL are particularly important for the design of further experiments to identify the gene or genes responsible for the effect. Likelihood support intervals are the most widely used method to obtain confidence intervals for QTL location, but the nonparametric bootstrap has also been recommended. Through extensive computer simulation, we show that bootstrap confidence intervals behave poorly and so should not be used in this context. The profile likelihood (or LOD curve) for QTL location has a tendency to peak at genetic markers, and so the distribution of the maximum-likelihood estimate (MLE) of QTL location has the unusual feature of point masses at genetic markers; this contributes to the poor behavior of the bootstrap. Likelihood support intervals and approximate Bayes credible intervals, on the other hand, are shown to behave appropriately.

**T**HERE is much interest in mapping the genetic loci (called quantitative trait loci, QTL) that contribute to variation in a quantitative trait. Once such a QTL has been identified, interest turns to the calculation of a confidence interval for its location, as such an interval estimate can be a useful guide for the design of further experiments, such as the generation of congenic lines.

LOD support intervals are the most commonly used interval estimates for the location of a QTL. A LOD support interval is defined as the interval in which the LOD score is within some value of its maximum. As an illustration, Figure 1A displays the LOD curve for chromosome 4 for the data of SUGIYAMA *et al.* (2001), concerning salt-induced hypertension in 250 back-cross mice. Assuming that there is a single QTL on this chromosome, the maximum-likelihood estimate (MLE) of the location of the QTL is the position at which the LOD curve achieves its maximum, in this case at marker D4Mit164 (at 30 cM). The 1.5-LOD support interval for the location of the QTL is the region in which the LOD score is within 1.5 of its maximum; here, the interval extends from 19 to 31 cM. (When the relevant region is disconnected, we generally take the conservative approach of forming the longest contiguous interval.)

LANDER and BOTSTEIN (1989) recommended the use of 1- and 2-LOD support intervals. DUPUIS and SIEGMUND (1999) found that 1.5-LOD support intervals

provide ~95% coverage in the case of a dense marker map. However, it has often been observed (see, *e.g.*, MANGIN *et al.* 1994) that the coverage of LOD support intervals depends upon the effect of the QTL, and so they do not behave as true confidence intervals.

VISSCHER *et al.* (1996) recommended the use of a nonparametric bootstrap to derive a confidence interval for the location of a QTL. For experimental cross data on  $n$  individuals, one makes  $n$  draws, with replacement, from the observed individuals to form a new data set in which some individuals are omitted and some appear multiple times. An estimate of QTL location is calculated with these new data, and the process is repeated many times. An ~95% confidence interval for the location of the QTL is obtained as the interval containing 95% of the estimated locations from the bootstrap replicates.

As an illustration, Figure 1B contains a histogram of the results of 10,000 bootstrap replicates using the chromosome 4 data of SUGIYAMA *et al.* (2001). The 95% bootstrap confidence interval extends from 14 to 32 cM. A striking feature of these results is that ~79% of the bootstrap replicates gave an estimated QTL location precisely at one of the 20 genetic markers on the chromosome. (Note that the calculations were performed at the markers and at 1-cM steps along the chromosome.) This is due to an unusual feature of the MLE of QTL location (previously observed by WALLING *et al.* 1998): it has a great tendency to occur precisely at a marker.

WALLING *et al.* (1998, 2002) investigated the performance of bootstrap confidence intervals for QTL

<sup>1</sup>Corresponding author: Department of Biostatistics, Johns Hopkins University, 615 N. Wolfe St., Baltimore, MD 21205-2179.  
E-mail: kbroman@jhsp.edu

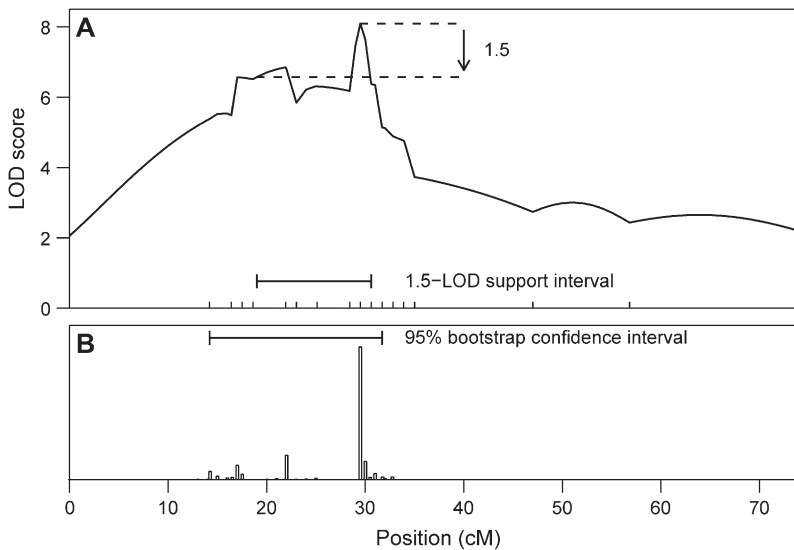


FIGURE 1.—Results for the chromosome 4 data of SUGIYAMA *et al.* (2001). (A) The LOD curve and the 1.5-LOD support interval. Tick marks at the bottom indicate the locations of the genetic markers. (B) A histogram of the estimated QTL locations in 10,000 bootstrap replicates, and the 95% bootstrap confidence interval, calculated by the method of VISSCHER *et al.* (1996).

location and concluded that they provide appropriate coverage. However, the unusual character of the distributions obtained in applications of the bootstrap for this problem, well illustrated in Figure 1B, led us to suspect that the performance of the bootstrap may be less than ideal and that the bootstrap may be inappropriate for the construction of confidence intervals for QTL location. Thus, we conducted a large-scale computer simulation study to investigate the performance of bootstrap confidence intervals for QTL location.

We considered the case of a backcross with a single segregating QTL, normally distributed residual variation, and equally spaced genetic markers exhibiting complete genotype data. While our simulation study is similar to those of WALLING *et al.* (1998, 2002) and differs largely in scale and thus precision, our conclusions are quite different. We find that the coverage of bootstrap confidence intervals for QTL location shows great variation as a function of the location of the QTL relative to the available genetic markers, and so we recommend against the use of the bootstrap for this problem.

One cannot reasonably recommend against the use of a method without providing some alternative, and so we further investigated the performance of LOD support intervals, as well as an approximate Bayes credible interval initially suggested by SEN and CHURCHILL (2001). Both of these types of intervals were found to display relatively stable coverage. On the basis of extensive simulations of backcrosses and intercrosses with varying marker densities and varying sizes of the effect of the QTL, we provide estimates of the appropriate amount to drop for LOD support intervals and the appropriate nominal fraction for the Bayes credible intervals to attain an actual coverage of 95%. The Bayes credible intervals are particularly attractive, as a nominal Bayes fraction of 96.5% in a backcross (and 97% in an intercross) is found to provide consistent coverage,

irrespective of the size of the QTL effect, marker density, and the number of individuals.

## METHODS

We consider the case of a backcross or an intercross with a single segregating QTL. We focus on the single chromosome (taken to have length 100 cM) harboring the QTL and assume equally spaced markers with complete genotype data. The residual variation is assumed to follow a normal distribution, and QTL mapping was performed by standard interval mapping (LANDER and BOTSTEIN 1989), which we briefly describe.

One assumes the presence of a single QTL and considers each position on the chromosome, one at a time, as the putative location of the QTL. (Our analyses were conducted at 1-cM steps along the chromosome.) While the QTL genotype,  $q$ , of an individual is generally not known, its distribution, conditional on the available marker data, may be calculated. Under the assumption of no crossover interference and with complete marker genotype data, the distribution of  $q$  depends only on the genotypes at the flanking markers. Given the QTL genotype, the phenotype is assumed to follow a normal distribution with mean  $\mu_q$  and common standard deviation  $\sigma$ . Given the available marker data, the phenotype follows a mixture of these normal distributions with known mixing proportions (the QTL genotype probabilities, conditional on the marker data). The nuisance parameters (the  $\mu_q$  and  $\sigma$ ) are estimated by maximum likelihood via the EM algorithm (DEMPSTER *et al.* 1977), and a LOD score is calculated, comparing the hypothesis that there is a single QTL precisely at that location to the null hypothesis of no QTL anywhere (in which case the phenotypes are assumed to follow a single normal distribution, independent of genotype).

Let  $\theta$  denote the true location of the QTL. The result of interval mapping is a LOD curve,  $\text{LOD}(\theta)$ , for the position of the QTL along the chromosome,  $\theta$ . This LOD curve is equivalent to a profile log likelihood for the position of the QTL. The MLE of the location of the QTL,  $\hat{\theta}$ , is the position at which the LOD curve achieves its maximum. While analysis at 1-cM steps along the chromosome results in a discrete distribution for  $\hat{\theta}$ , analysis on a finer grid would greatly increase the computational effort and would provide similar results.

LOD support intervals were calculated as the longest contiguous interval in which the LOD score was within some chosen value of its maximum. Bootstrap confidence intervals were constructed via the percentile method, as described by VISSCHER *et al.* (1996). For each of 1000 bootstrap replicates, a sample of the same size as the available data was drawn with replacement from the available individuals, and a new estimate of QTL location ( $\hat{\theta}^*$ ) was obtained by application of standard interval mapping to the resampled data. The endpoints of the 95% bootstrap confidence interval were taken to be the 2.5 and 97.5 percentiles of the  $\hat{\theta}^*$ . Finally, an approximate Bayes credible interval was calculated: we treated the profile likelihood for QTL location as if it were a real likelihood, assigned a uniform prior on the location of the QTL, and so derived an approximate posterior distribution for QTL location,  $f(\theta | \text{data}) = 10^{\text{LOD}(\theta)} / \sum_{\theta=0}^{100} 10^{\text{LOD}(\theta)}$ . From this approximate posterior, a 95% Bayes credible interval was defined to be the interval,  $I$ , for which  $f(\theta | \text{data})$  exceeded some threshold and for which  $\sum_{\theta \in I} f(\theta | \text{data}) \geq 0.95$ .

**Effect of QTL location relative to markers:** In our first simulation study, to investigate the coverage of bootstrap confidence intervals, we considered a backcross of 200 individuals and a single QTL whose position was allowed to vary at 1-cM steps along a chromosome of length 100 cM. Complete genotype data were available at 11 equally spaced markers (thus at a 10-cM spacing). The heritability due to the QTL (the proportion of the phenotypic variance due to the QTL) was taken to be 10%.

For each of the 101 possible QTL positions (at 0, 1, 2, ..., 100 cM), we performed 10,000 simulation replicates. At each replicate, we calculated the LOD curve by standard interval mapping at 1-cM steps along the chromosome and derived the 1-LOD support interval and 95% Bayes credible interval. (We used 1-LOD support intervals here, as they were found to be somewhat conservative in this sparse-map case.) In addition, at each simulation replicate we constructed a 95% bootstrap confidence interval on the basis of 1000 bootstrap replicates, as described above. Great computational effort was expended in this investigation: at each of 1,010,000 simulation replicates (10,000 replicates for each of 101 QTL positions), 1000 bootstrap replicates were performed.

The simulations were performed using the R statistical software (IHAKA and GENTLEMAN 1996) and R/qtl (BROMAN *et al.* 2003), an add-on package to R. For some aspects of our simulation studies, we used C code adapted from the R code in R/qtl to improve computational speed. The total computer time for these simulations was  $\sim 450$  days. The simulations were split across multiple processors on a fast cluster, but still required  $\sim 2$  months of constant computation. Our modifications to R/qtl resulted in a 10- to 15-fold improvement in speed.

**Effect of cross type, sample size, marker density, and QTL effect:** On the basis of the results of the first simulation study, we performed a second simulation study to more completely characterize the coverage of the LOD support and Bayes credible intervals. We varied the type of cross (backcross or intercross), the sample size (200 or 500), the marker density (1-, 2-, 10-, or 20-cM spacing), and the effect of the QTL. We hypothesized that interval coverage might be more clearly expressed as a function of the power to detect the QTL rather than the heritability due to the QTL, and so heritabilities were chosen to give estimated power of 0.3, 0.4, ..., 0.9, where power was defined as the probability of achieving a LOD score of at least 3. These heritabilities were estimated via R/qtlDesign (SEN *et al.* 2005), an add-on package to the R statistical software (IHAKA and GENTLEMAN 1996). Our targeted values for power, calculated with R/qtlDesign, differed somewhat from the power estimated from our simulation results, as we defined power to be the chance of a LOD score  $\geq 3$  somewhere on the chromosome, whereas R/qtlDesign defines it to be the chance of a LOD score  $\geq 3$  at the QTL. The results are presented below using the power seen in our simulations.

The position of the QTL was fixed at a position equidistant between two markers and near the center of the chromosome. For the marker spacings of 1, 2, 10, and 20 cM, the QTL was placed at 49.5, 49, 45, and 50 cM, respectively, as with the QTL equidistant between markers, coverage of the LOD and Bayes intervals was lowest. Thus, the results of our second set of simulations may be viewed as conservative. In the case of 1-cM marker spacing, calculations were performed every 0.5 cM, rather than every 1 cM, as used for the other cases. For each setting (of cross type, sample size, marker density, and QTL effect), 100,000 simulation replicates were performed. The total computation time for these simulations was  $\sim 40$  days.

For each simulation replicate, standard interval mapping was performed to obtain the LOD curve at 1-cM steps along the chromosome. Rather than investigate the coverage of the LOD support and Bayes credible intervals for particular choices of the drop in LOD and the nominal Bayes fraction, we chose to estimate the drop in LOD and the nominal Bayes fraction for which the two types of intervals would attain

95% coverage. These values could be obtained with little additional effort. At each simulation replicate, we kept track of the difference in the LOD score at the MLE and at the true location of the QTL. The 95th percentile of these differences is the value to drop in a LOD support interval to attain 95% coverage. A similar trick applies for the Bayes credible intervals. Note that here we are using a definition for the confidence interval that can lead to a set of disjoint intervals, rather than a single contiguous interval, and so the results are somewhat conservative.

## RESULTS

**Distribution of the MLE of QTL location:** Our initial simulation study, comprising 10,000 replicates with a QTL at each of 0, 1, 2, ..., 100 cM on a chromosome of length 100 cM, allows us to inspect the distribution of the MLE of QTL location and the dependence of this distribution on the location of the QTL relative to the markers. The simulations used a backcross of 200 individuals, 11 equally spaced markers (10-cM spacing), and heritability due to the single QTL at 10%.

Figure 2 displays the distribution of the MLE of QTL position,  $\hat{\theta}$ , as a function of the true location of the QTL,  $\theta$ , for  $\theta = 45, 46, \dots, 50$ . The most striking feature of these distributions is the clear tendency for the  $\hat{\theta}$  to occur exactly at the marker loci. For example, in the case that the QTL is at 49 cM, immediately adjacent to a marker, there is a far greater chance that the QTL is estimated to be at the marker rather than at the true location of the QTL. A similar pattern was seen for other values of  $\theta$ . The standard error (SE) of  $\hat{\theta}$  is smallest when the QTL is at a marker and is  $\sim 25\%$  larger when the QTL is in the center of the interval between markers. When the QTL is near one of the ends of the chromosome,  $\hat{\theta}$  exhibits considerable bias, as we do not examine positions beyond the terminal markers on the chromosome. We calculated the LOD score at 1-cM steps along the chromosome and so estimated QTL position only to within 1 cM. If calculations were performed on a more dense grid, the tendency for the MLE to occur precisely at the markers would be even more striking.

The dependence of the distribution of  $\hat{\theta}$  on the position of the QTL relative to the markers, both with respect to the large mass placed at the markers and with respect to the variation in the SE of  $\hat{\theta}$ , is seen to cause a breakdown in the performance of the bootstrap for this problem.

**Coverage vs. true QTL location:** Performance of a confidence interval is generally assessed by its coverage (the probability that it contains the true parameter value) as a function of the true parameter. Ideally, a 95% confidence interval shows constant 95% coverage, regardless of the true parameter value. In Figure 3, coverage of the 95% bootstrap confidence interval

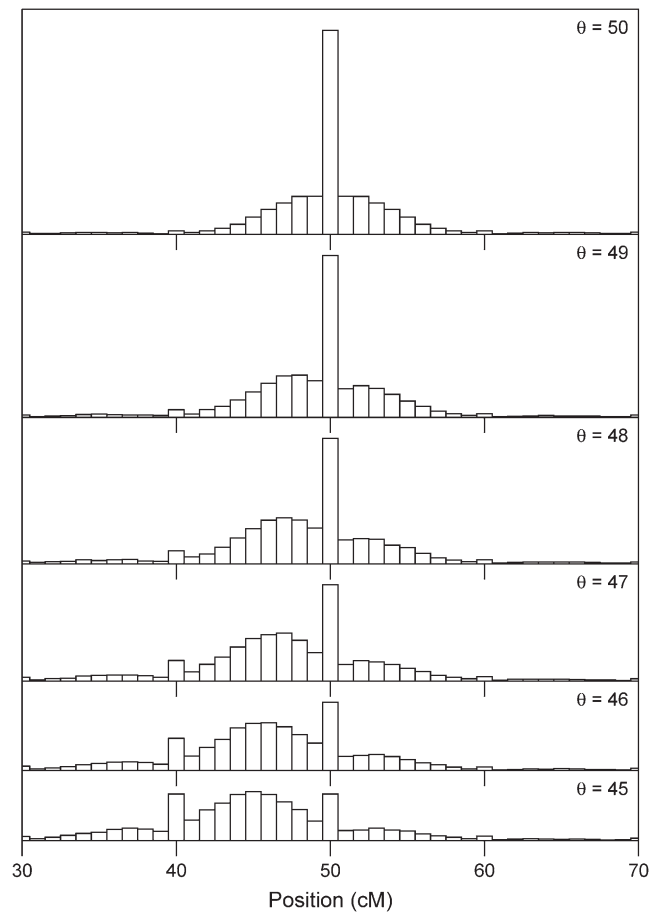


FIGURE 2.—Estimated distribution of the MLE of QTL location,  $\hat{\theta}$ , as a function of the true location of the QTL,  $\theta$ , for  $\theta$  varying from 45 to 50. The results are based on 10,000 simulation replicates of a backcross with 200 individuals for a chromosome of length 100 cM and having 11 equally spaced markers and with the heritability due to the QTL at 10%.

(in black), 1-LOD support interval (in red), and the 95% Bayes credible interval (in blue) is displayed as a function of the true location of the QTL. The bootstrap confidence interval shows extremely high coverage ( $\sim 99\%$ ) when the true QTL is at a marker, low coverage ( $\sim 92.5\%$ ) when the QTL is right next to a marker, and above nominal coverage when the QTL is exactly between markers. Coverage of the 1-LOD support and 95% Bayes credible intervals does not fluctuate as widely, although it is highest when the QTL is at a marker. Note that the SEs of our estimates of coverage are  $\sim 0.3\%$ .

**Coverage vs. estimated QTL location:** It is also of interest to consider coverage as a function of the estimated QTL location,  $\hat{\theta}$ . In our simulations, we performed 10,000 replicates for each of the 101 possible positions of the QTL; here we consider the portion of those 1,010,000 simulation replicates in which the MLE was attained, for example, at 50 cM, and calculate the proportion of those replicates in which each type of confidence interval contained the true parameter value.

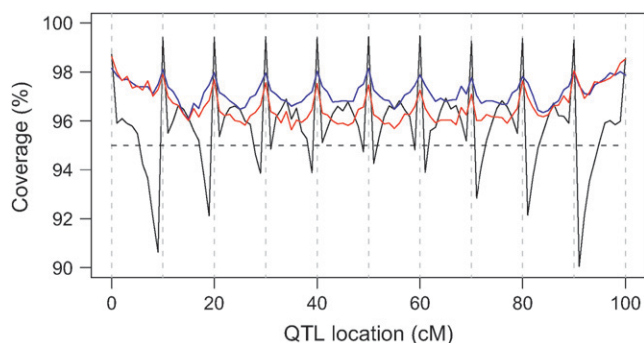


FIGURE 3.—Coverage of 95% bootstrap confidence intervals (black), 1-LOD support intervals (red), and 95% Bayes credible intervals (blue) as a function of the true QTL position,  $\theta$ . The dashed vertical gray lines denote marker positions on the chromosome.

This is an unorthodox mixture of Bayes and frequentist statistics. The coverage of a confidence interval is a quantity of interest only to frequentists; here we are taking the location of the QTL to be uniformly distributed on the positions 0, 1, 2, ..., 100 cM and inspecting the posterior probability, given the observed estimate of the QTL location, that the interval covers its true parameter value. Note that, across the 1,010,000 simulation replicates, each possible value of  $\hat{\theta}$  was observed at least 6700 times, and so the SEs of our estimates of coverage as a function of  $\hat{\theta}$  are  $\sim 0.4\%$ .

Coverage as a function of the estimated location of the QTL is displayed in Figure 4. These results provide a qualitatively different perspective from those of coverage *vs.*  $\theta$ , shown in Figure 3. While coverage of the bootstrap confidence intervals is high when the QTL is at a marker (see Figure 3), coverage is low ( $\sim 92\%$ ) when the QTL is estimated to be at a marker. Coverage of the 1-LOD support interval and 95% Bayes credible interval is less variable as a function of  $\hat{\theta}$  and is entirely above the nominal level, 95%, with the Bayes credible interval exhibiting slightly less variability than the LOD support interval.

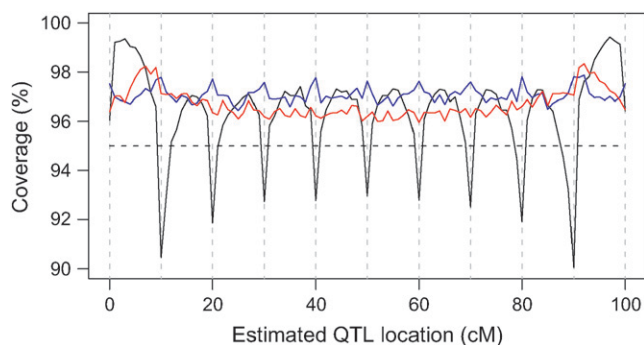


FIGURE 4.—Coverage of 95% bootstrap confidence intervals (black), 1-LOD support intervals (red), and 95% Bayes credible intervals (blue) as a function of the MLE of QTL position,  $\hat{\theta}$ . The dashed vertical gray lines denote marker positions on the chromosome.

We view this perspective (coverage of a confidence interval conditional on the observed estimate,  $\hat{\theta}$ ) as the more relevant one for the user of a confidence interval. One does not know the true location of the QTL, but does know one's estimate of that location, and so coverage as a function of the observed estimate is of greatest interest. But it is from this perspective that coverage of the bootstrap confidence intervals looks worst. While coverage is low only when the estimated location of the QTL is at a marker, it is quite low in that case, and, as we've seen, that is often the case.

**Interval widths:** Another important feature of a confidence interval is its width: one prefers intervals to be as small as possible, while maintaining the appropriate level of coverage. Averaging over all possible values of  $\theta$ , the 95% bootstrap confidence interval, 1-LOD support intervals, and 95% Bayes credible intervals had average widths of 45, 24, and 29 cM, respectively. When the QTL was not close to the end of the chromosome, the 1-LOD support intervals were  $>40\%$  smaller than the bootstrap confidence intervals about half of the time. The 95% Bayes credible intervals and the 1-LOD support intervals were quite similar in width. The 1-LOD support and 95% Bayes credible intervals show not just better coverage properties than the 95% bootstrap intervals (see Figures 3 and 4), but are also generally smaller.

**Coverage with varying cross type, sample size, marker density, and QTL effect:** As described above, coverage of the 95% bootstrap confidence intervals varied greatly according to the position of the QTL relative to the genetic markers; the 1-LOD support and 95% Bayes credible intervals, on the other hand, exhibited relatively stable coverage across the chromosome. We thus omitted the bootstrap confidence intervals from further consideration, but sought a more complete characterization of the performance of the LOD support intervals and the approximate Bayes credible intervals, as a function of sample size, marker density, and QTL effect, and considering both a backcross and an intercross.

Rather than study the coverage of the intervals for a fixed drop in LOD or nominal Bayes fraction, we sought the values that would give 95% coverage at different settings of the parameters of interest. The value to drop in LOD for the coverage of the LOD support interval to have 95% coverage in a backcross is displayed in Figure 5A. The results are displayed as a function of the size of the effect of the QTL, which has been reparameterized as the power to give a LOD score of at least 3. (The displayed values for the power were estimated from 100,000 simulation replicates at each point and so have standard error  $<0.2\%$ .) The black and red curves correspond to sample sizes of 200 and 500, respectively. As seen in Figure 5A, sample size has little effect on the appropriate drop in LOD to give 95% coverage, for a given power to detect the QTL. Of course, the heritability due to the QTL that corresponds to a particular

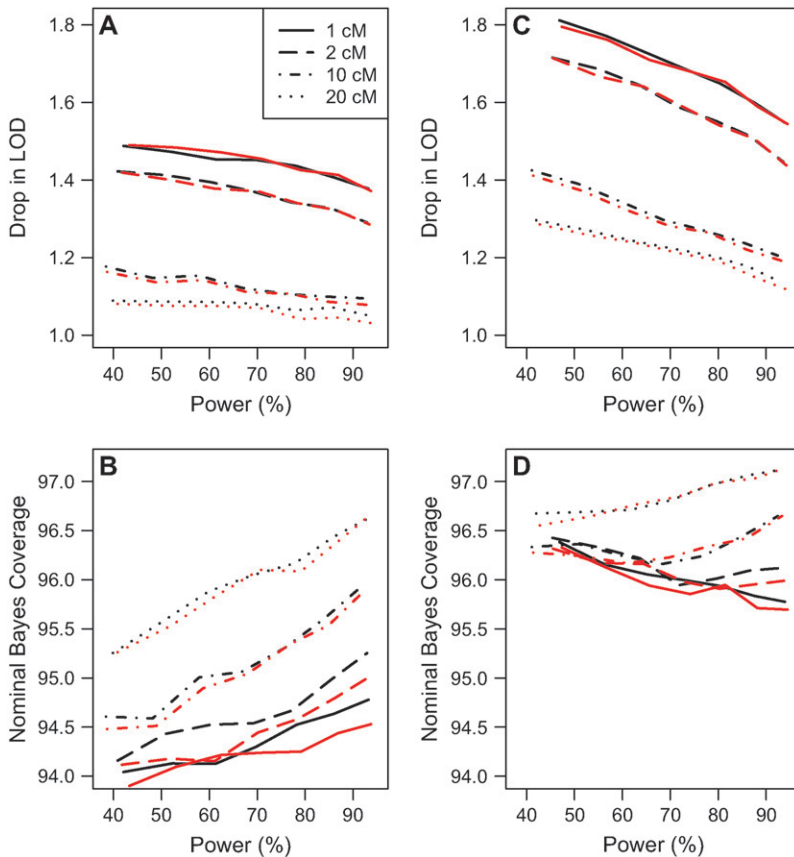


FIGURE 5.—Estimated amount to drop in a LOD support interval (A and C) and the nominal Bayes coverage for the approximate Bayes credible interval (B and D) to give 95% coverage, based on 100,000 simulation replicates. Backcross (A and B) and intercross (C and D) experiments with either 200 (black curves) or 500 (red curves) individuals were considered. The line types indicate different possible marker spacings. Values are plotted as a function of the effect of the QTL, scaled according to the power to detect the QTL.

power is quite different for the two sample sizes. The biggest effect seen concerns the spacing of markers: one must drop  $\sim 1.5$  in LOD to attain 95% coverage when markers are at a 1-cM spacing, but need drop only  $\sim 1.2$  in LOD if the markers are at a 10-cM spacing. A slightly smaller drop is required when the QTL has a larger effect.

The nominal Bayes fractions at which the approximate Bayes credible intervals had 95% coverage in a backcross are displayed in Figure 5B. Again, sample size has little effect, except in the case of very dense markers. The effect of marker spacing and of the size of the QTL effect is seen to be in the opposite direction for the Bayes intervals *vs.* the LOD support intervals. A greater nominal Bayes fraction is needed for sparse markers and for a larger QTL effect.

Figure 5, C and D, shows the corresponding results for an intercross. A greater drop in LOD is required for the LOD support interval to have 95% coverage in the intercross, and the QTL effect appears to have a somewhat greater influence on the appropriate value to drop. Sample size is again seen to have little effect, and the greatest influence comes from the marker spacing, with a greater drop in LOD required in the case of more densely spaced markers. There is remarkably little variation in the appropriate nominal Bayes fraction so that the approximate Bayes credible interval

has 95% coverage in an intercross; for all sample sizes, marker spacings, and QTL effects, the appropriate nominal Bayes fraction was 96–97%.

These results suggest that, for the Bayes intervals, the use of 96.5% for a backcross and 97% for an intercross will provide  $>95\%$  coverage for all possible cases. For the LOD support intervals, if one drops by 1.5 for a backcross and 1.8 for an intercross, coverage will be maintained at  $>95\%$ . The actual coverage obtained with these choices is shown in Figure 6. The Bayes intervals are seen to be particularly attractive, as they exhibit quite stable coverage with sample size, marker density, and QTL effect.

## DISCUSSION

We have shown that coverage of bootstrap confidence intervals for QTL location depends critically upon the location of the QTL relative to the typed genetic markers. Coverage is high when the QTL is at a marker but can be low when the QTL is immediately adjacent to a marker (see Figure 3). Especially interesting results were observed in the consideration of coverage as a function of the estimated location of the QTL, taking the true location of the QTL to be uniformly distributed along the chromosome. This perspective is most relevant for the user of such confidence intervals and

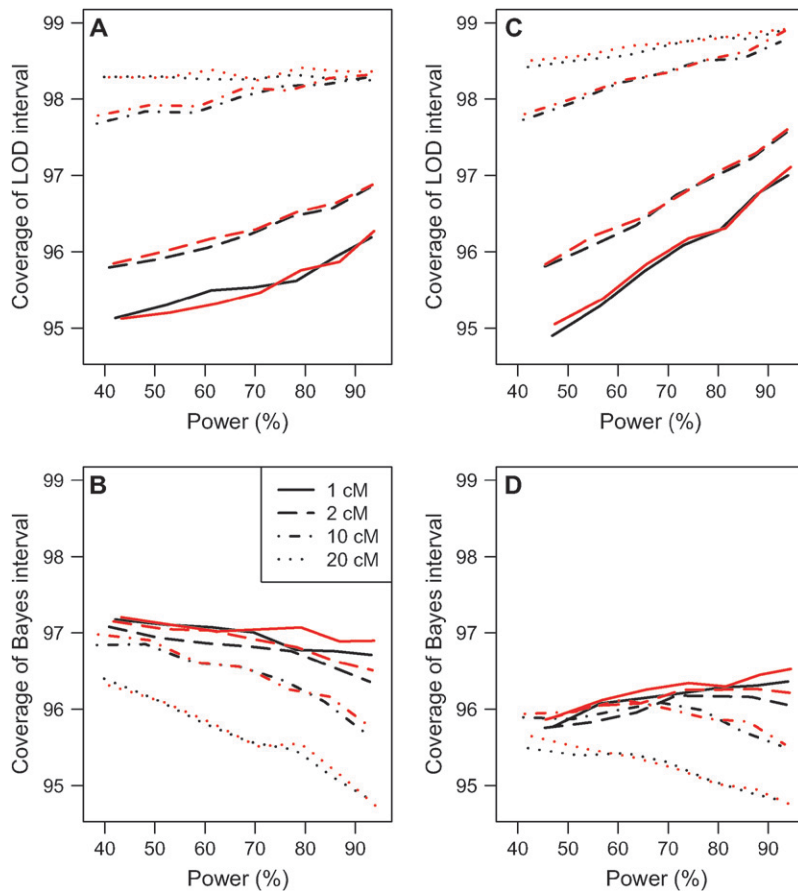


FIGURE 6.—Coverage of the 1.5-LOD support interval in a backcross (A), the 96.5% Bayes interval in a backcross (B), the 1.8-LOD support interval in an intercross (C), and the 97% Bayes interval in an intercross (D), based on 100,000 simulation replicates. The black curves are for 200 individuals; the red curves are for 500 individuals. The line types indicate different marker spacings. Values are plotted as a function of the effect of the QTL, scaled according to the power to detect the QTL.

indicates poor performance of the bootstrap confidence intervals: coverage is quite far below the nominal level when the QTL is estimated to be at a marker (see Figure 4). The bootstrap confidence intervals were also seen to be much wider than the LOD support and approximate Bayes credible intervals.

Our results are similar to those of WALLING *et al.* (1998, 2002), but our conclusions are markedly different. It is important to point out that WALLING *et al.* (1998, 2002) used Haley–Knott regression (HALEY and KNOTT 1992), whereas we have focused on standard interval mapping (LANDER and BOTSTEIN 1989), using maximum likelihood via the EM algorithm. While it was not mentioned above, we did include the use of Haley–Knott regression in our initial simulation study and found similar results by the two methods (data not shown).

Bootstrap methods have desirable properties in a wide variety of statistical problems. However, modifications to the bootstrap are necessary for problems that are not classically regular (BERAN 2003), and the QTL mapping problem is not regular (KONG and WRIGHT 1994; SIEGMUND 2004). Thus, our finding of inadequate bootstrap performance in QTL mapping is consistent with theory.

The poor performance of bootstrap confidence intervals for QTL location derives from the unusual behavior of the MLE of QTL location: the MLE has a

tendency to coincide with a marker position (see Figure 2 and KONG and WRIGHT 1994), and its SE varies greatly according to the location of the QTL relative to the markers. Appropriate performance of the percentile-based nonparametric bootstrap confidence intervals (proposed, for this context, by VISSCHER *et al.* 1996 and studied herein) generally requires the existence of some monotone transformation  $h(\cdot)$  such that  $h(\hat{\theta}) - h(\theta)$  has the same symmetric continuous distribution for all  $\theta$  (SHAO and TU 1995, p. 132). The tendency of the MLE to occur at a marker indicates that no such transformation exists for this problem.

An alternative heuristic for understanding the breakdown of the bootstrap in this problem is as follows: we hope to approximate the sampling distribution,  $f(\hat{\theta} | \theta)$ , by the bootstrap distribution,  $g(\hat{\theta}^* | \hat{\theta})$ . But the bootstrap distribution better reflects the sampling distribution evaluated at the observed estimate,  $f(\cdot | \hat{\theta})$ , than it does the target,  $f(\cdot | \theta)$ . That the MLE is most precise when the QTL is at a marker, and is less precise when the QTL is between markers, indicates that the bootstrap distribution will provide an overly optimistic view of our understanding of the location of the QTL in those cases in which we have estimated the QTL to be at a marker.

WALLING *et al.* (1998) also assessed the performance of the parametric bootstrap for this problem; rather than resampling from the observed data, one simulates

new data taking one's estimate of the QTL location to be the true location. They obtained the surprising result that the parametric bootstrap performed more poorly than the nonparametric bootstrap; the result is surprising because, when one's model is correct (as it was in their simulation study), the parametric bootstrap would be expected to give better performance than the nonparametric bootstrap. This result can now be clearly understood. In the parametric bootstrap, the bootstrap distribution on which the confidence interval is based is simply the sampling distribution of the estimate in the case that the QTL is located at the observed estimate. Thus, when the QTL is estimated to be at a marker, the parametric bootstrap will provide an overly optimistic view of the precision of that estimate.

The tendency of the MLE for QTL location to occur precisely at a genetic marker (see Figure 2) is a major contributor to the failure of the bootstrap in this context. Our explanation of the cause of this behavior is as follows. The profile likelihood exhibits cusps at the markers. (Its first derivative is not continuous at the markers.) This is the result of the fact that, in the case of complete genotype data at the markers, and with the assumption of no crossover interference, the likelihood to the left of the marker incorporates data on the marker to the left but not those for the marker to the right, while the likelihood to the right of a marker incorporates data on the marker to the right but not those for the marker to the left. The abrupt change in the first derivative of the profile likelihood at the markers appears to lead to a greater chance of a change in the direction of the profile likelihood and so to a greater chance that the MLE occurs precisely at a marker.

The unusual features of the MLE distribution do not necessarily imply that it is a poor estimate of QTL location, but just that its use with the bootstrap will not work well. The MLE is approximately unbiased (except for the end-of-chromosome effect), and an improved estimate of QTL location is not so important as an improved interval estimate, which the approximate Bayes interval provides.

It should be emphasized that these results were obtained in a single setting: a backcross of 200 individuals, equally spaced markers at a 10-cM spacing, and heritability due to the QTL at 10%. The behavior of the bootstrap seen here may not hold generally. In fact, for a cross with very dense markers and a QTL of not too strong effect, the bootstrap would likely behave reasonably. However, the setting in which our simulations were conducted is not unreasonable, and that the bootstrap performed so poorly here supports the general conclusion that it should not be used.

It should also be emphasized that we have considered only percentile-based nonparametric bootstrap confidence intervals, as that was the approach recommended by VISSCHER *et al.* (1996). Other forms of bootstrap might be found to work in this context. For example,

one might use a bootstrap to calibrate the LOD support or approximate Bayes credible intervals. However, the good performance of the approximate Bayes credible interval suggests that the computational effort that must be expended in any bootstrap may not be necessary.

We have focused on the simplest possible QTL model: a single QTL with normally distributed residual variation. This simple model is not likely to hold in practice. An especially important departure concerns the presence of multiple linked QTL. A confidence interval for QTL location derived from the results of analysis using single-QTL models has little meaning if multiple QTL exist on the chromosome. The LOD support and Bayes credible intervals have obvious extensions for the case of multiple QTL; their performance, especially in the case of multiple linked QTL, deserves further study.

We conducted a small simulation study to explore the performance of the various intervals in the case that there exist two tightly linked QTL that cannot easily be separated. The setup was similar to our first simulation: 200 backcross individuals and a chromosome of length 100 cM with markers at a 10-cM spacing. Two additive QTL were linked in coupling at positions 15–20 and 40–45, respectively. The total heritability due to the pair was set at 10%, matching the effect of the single QTL in our first simulations. We found that the chance that the Bayes and LOD intervals covered at least one of the QTL was at the nominal level. The bootstrap intervals had quite high coverage, but were considerably wider than the other intervals, as seen in our first simulation: they were typically 40% wider than the LOD and Bayes intervals and were twice as wide ~25% of the time. The greater coverage of the bootstrap intervals is largely attributable to this increased width. All intervals appeared to perform reasonably in the presence of two tightly linked QTL, but the greater width and greater computation of the bootstrap intervals led us to continue to prefer the Bayes intervals.

While we have shown that bootstrap confidence intervals for QTL location perform poorly and so should not be used in this context, the LOD support and approximate Bayes credible intervals were seen to behave appropriately. This is in broad agreement with DUPUIS and SIEGMUND (1999). They studied the performance of LOD support and Bayes credible intervals, focusing on the widths of the intervals. They found that when LOD support and Bayes credible intervals had similar coverage, their widths were generally comparable. For LOD intervals to have the target coverage properties, the LOD drop has to be adjusted, while the Bayes intervals give consistent coverage for a range of marker densities and QTL effects. Thus, the approximate Bayes credible intervals are particularly attractive: a nominal 96.5 or 97% Bayes credible interval was seen to exhibit coverage near 95% for different sample sizes, marker densities, and sizes of QTL effect.



Finally, we emphasize that 95% is not a magic number and investigators may wish to be more conservative (seeking, for example, 99% coverage), so that, for example, the formation of a congenic line does not miss the true location of the QTL.

This work was supported in part by National Institutes of Health grant GM074244 (to K.W.B.) and a National Science Foundation graduate research fellowship (to A.M.).

#### LITERATURE CITED

- BERAN, R., 2003 The impact of the bootstrap on statistical algorithms and theory. *Stat. Sci.* **18**: 175–184.
- BROMAN, K. W., H. WU, S. SEN and G. A. CHURCHILL, 2003 R/qt: QTL mapping in experimental crosses. *Bioinformatics* **19**: 889–890.
- DEMPSTER, A. P., N. M. LAIRD and D. B. RUBIN, 1977 Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. B* **39**: 1–38.
- DUPUIS, J., and D. SIEGMUND, 1999 Statistical methods for mapping quantitative trait loci from a dense set of markers. *Genetics* **151**: 373–386.
- 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.
- IHAKA, R., and R. GENTLEMAN, 1996 R: a language for data analysis and graphics. *J. Comp. Graph. Stat.* **5**: 299–314.
- KONG, A., and F. WRIGHT, 1994 Asymptotic theory for gene mapping. *Proc. Natl. Acad. Sci. USA* **91**: 9705–9709.
- LANDER, E. S., and D. BOTSTEIN, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185–199.
- MANGIN, B., B. GOFFINET and A. REBAI, 1994 Constructing confidence intervals for QTL location. *Genetics* **138**: 1301–1308.
- SEN, S., and G. A. CHURCHILL, 2001 A statistical framework for quantitative trait mapping. *Genetics* **159**: 371–387.
- SEN, S., J. M. SATAGOPAN, K. W. BROMAN and G. A. CHURCHILL, 2005 R/qtDesign: inbred line cross experimental design. Center for Bioinformatics and Molecular Biostatistics, University of California, San Francisco (<http://repositories.cdlib.org/cbmb/qt-powercalculations>).
- SHAO, J., and D. S. TU, 1995 *The Jackknife and the Bootstrap*. Springer, New York.
- SIEGMUND, D., 2004 Model selection in irregular problems: applications to mapping quantitative trait loci. *Biometrika* **91**: 785–800.
- SUGIYAMA, F., G. A. CHURCHILL, D. C. HIGGINS, C. JOHNS, K. P. MAKARITSIS *et al.*, 2001 Concordance of murine quantitative trait loci for salt-induced hypertension with rat and human loci. *Genomics* **71**: 70–77.
- VISSCHER, P. M., R. THOMPSON and C. S. HALEY, 1996 Confidence intervals in QTL mapping by bootstrapping. *Genetics* **143**: 1013–1020.
- WALLING, G. A., P. M. VISSCHER and C. S. HALEY, 1998 A comparison of bootstrap methods to construct confidence intervals in QTL mapping. *Genet. Res.* **71**: 171–180.
- WALLING, G. A., C. S. HALEY, M. PEREZ-ENCISO, R. THOMPSON and P. M. VISSCHER, 2002 On the mapping of quantitative trait loci at marker and non-marker locations. *Genet. Res.* **79**: 97–106.

Communicating editor: A. D. LONG