RESEARCH NOTE

A quick method to calculate QTL confidence interval

HENGDE LI*

The Centre for Applied Aquatic Genomics, Chinese Academy of Fishery Sciences, Beijing 100141, People's Republic of China

[Li H. 2011 A quick method to calculate QTL confidence interval. *J. Genet.* **90**, 355–360]

Introduction

There is much interest in the nature of the genetic variation for quantitative traits, and mapping quantitative trait loci (QTL) is a method to reveal it (Mackay 2001). Once a QTL is identified, interest turns to determining the confidence interval (CI) of QTL location, which is a useful guide for further experimental design and analysis to reveal the real molecular nature of the variation of quantitative traits.

The classic method to determine the CI of QTL location is nonparametric bootstrap suggested by Visscher *et al.* (1996). A series of bootstrap samples are formed by withdrawing observations randomly with replacement from the observed data, and then mapping to detect QTL at each putative position. These positions with largest test statistics of each bootstrap sample form the bootstrap distribution of QTL location. The 2.5 and 97.5 percentiles are lower limit and upper limit of 95% QTL confidence interval. Although bootstrap is time-demanding, especially for large complex populations, it has been used frequently. Some authors' investigation showed that bootstrap CI provide appropriate coverage (Walling *et al.* 1998, 2002). However, some authors' investigation showed that bootstrap CI behave poorly (Manichaikul *et al.* 2006). So the bootstrap CI is not very stable, perhaps because of the unusual character of distributions obtained in applications (Sugiyama *et al.* 2001). Bennewitz *et al.* (2002) presented three methods of permutation bootstrapping, which is a modification of traditional nonparametric bootstrap, to improve the precision of the CI and showed that CI will be short and less biased in a large number of simulated configurations, if the impact of markers was corrected.

A quick method to determine the CI of a QTL location is to use 1-LOD and 2-LOD support intervals as 95% and 99% CI (Lander and Botstein 1989). Supposing the largest LOD score is y, then the largest (smallest) position on the left (right) of QTL where LOD score less than y-x is lower (upper) limit of x-LOD support interval. It was found that 1-LOD rule often gives very small CI, and the CI also depends on the effect of QTL (Mangin *et al.* 1994). Dupuis and Siegmund (1999) found that 1.5-LOD support intervals provide 95% CI if the marker map is dense. An extreme example is the CI will be the whole chromosome if the range of LOD scores of each position is less than 1 or 1.5. So the x-LOD rule does not always give a stable estimate of the CI either.

Darvasi and Soller (1997) derived formulas to calculate the confidence intervals for backcross design and F_2 design through an extensive series of simulations, the formula is:

$$\mathrm{CI} = \frac{3000}{m \cdot n \cdot \delta^2}, \tag{1}$$

where m is the relative number of informative meiosis ($m = 1$ for backcross and $m = 2$ for F_2), N is the sample size and $\delta = (d + h)$, $-d$, h and d are gene effects of QTL genotype qq, Qq and QQ respectively. Later, Weller and Soller (2004) derived an analytical formula of different experimental designs for inbred populations, which incorporate the QTL allele substitution effect and the number of individuals genotyped and phenotyped to estimate CI of QTL location using a saturated genetic map. It is an extension of the results of Darvasi and Soller (1997) where the genetic effects of QTL have to be considered as fixed effects, so it cannot deal with populations with complex pedigree. What is more, the estimated CI is assumed to be symmetric, but in practice, it is not reasonable, because the background effect and random error will lead to asymmetric profile even the QTL is exactly located on the centre of the chromosome.

In theory, the lower and upper limits of CI are continuous variables; however bootstrap method detects only a limit of positions, so the step-size of genome scan will affect the bootstrap CI much. Because of time-consuming nature of bootstrap, inapplicability to outbred populations of formula CI, and instability of x-LOD support interval, a method to calculate QTL CI quickly, that is applicable not only to inbred but also outbred populations is required. In this paper,

*E-mail: hengde_li@yahoo.com.cn.

Keywords. quantitative trait loci; confidence interval.

we present a quick method to determine the CI of QTL, which can be used for different QTL mapping methods and resource population.

Materials and methods

Theory

The popular strategy to map QTL is maximum likelihood (ML) method. The framework is to test whether a QTL is linked to the marker(s) under consideration based on the $-2\times$log-likelihood ratio (LR) statistic (Lynch and Walsh 1998):

$$LR = -2 \cdot \ln \frac{\max l_r(z)}{\max l(z)} \tag{2}$$

where z is the observed data, $\max l_r(z)$ is the maximum of the likelihood function under the null hypothesis of no segregating QTL, $\max l(z)$ is the maximum of the likelihood function under the alternative hypothesis of there being a segregating QTL. The LR for position θ_i can be expressed as:

$$LR_i = -2 \cdot \ln \frac{\max l_r(z)}{\max l(\theta_i|z)} \tag{3}$$

The log-likelihood profile can be obtained if we plot the LR against positions. The estimated QTL location is the position where the peak is. In theory, θ is a continuous variable, if the possibility of each position to harbour a QTL is known, we can get the density curve of θ, then according to the density curve, the CI of QTL location can be estimated easily. It is difficult to get the real frequencies of each θ directly; however, likelihood can be used as a substitute of probability to harbour a QTL, the greater the likelihood is, the greater probability to harbour a QTL; hence likelihoods ratio of two positions reflect the relative frequency to harbour a QTL at these two positions. For example, the likelihoods of positions i and j are $\max l(\theta_i|z)$ and $\max l(\theta_j|z)$ respectively, supposing the probability to harbour a QTL at position θ_j is p, then the probability to harbour a QTL at position θ_i is:

$$\frac{\max l(\theta_i|z)}{\max l(\theta_j|z)} \cdot p,$$

i.e. one LOD difference between positions θ_i and θ_j ($LOD_i - LOD_j = 1$) means the chance to harbour a QTL at position θ_i is 10 times of that at position θ_j; one LR difference between positions θ_i and θ_j ($LR_i - LR_j = 1$) means the chance to harbour a QTL at position θ_i is $e^{1/2}$ times of that at position θ_j.

So let us denote the relative frequency ratio (RFR) with f, the RFR of position θ_i against position θ_j is:

$$\begin{aligned} f_{ij} &= \frac{\max l(\theta_i|z)}{\max l(\theta_j|z)} \\ &= \frac{\max lr(z) \cdot e^{LR_i/2}}{\max lr(z) \cdot e^{LR_j/2}} \\ &= e^{(LR_i - LR_j)/2} \end{aligned} \tag{4}$$

where f_{ij} is the RFR of position θ_i to θ_j. Hence the distribution of QTL location is:

$$\begin{aligned} f(\theta|z) &= \frac{e^{(LR_\theta - LR_j)/2}}{\int_\theta^L e^{(LR_\theta - LR_j)/2} \cdot d\theta} \\ &= \frac{e^{LR_\theta/2}}{\int_\theta^L e^{LR_\theta/2} \cdot d\theta} \end{aligned} \tag{5}$$

where L is chromosome length, according to the distribution we developed the algorithm to calculate the CI:

i) Calculate the $f(\theta_i|z)$ at each scanned position; ii) Calculate the areas (A_i) under the QTL profile from θ_i to θ_{i+1}, when dense positions were scanned, A_i is approximately as: $0.5 \times (f(\theta_i|z) + f(\theta_i|z)) \times (\theta_{i+1} - \theta_i)$ (figure 1). iii) Scale A_i using the sum of A_i. $A_i' = A_i / \sum A_i$. Hence the sum of A_i' is exactly 1. iv) Search from the maximum of LR statistics down to the appropriate threshold with step size of, e.g. 0.01LR to ensure $\int_u^v f(\theta|z)d\theta = 0.95$ or $\int_0^u f(\theta|z)\,d\theta + \int_v^L f(\theta|z)\,d\theta = 0.05$, where u and v are positions at which the LR is the threshold. Then u and v are lower and upper limits of 95% CI of QTL location, respectively.

Simulation

In order to test our approach, we simulate an F_2 population of 50 replicates. The number of individuals is 300; the chromosome length is 95 cM with 20 markers evenly distributed, and Haldane mapping function (Haldane 1919) was used in the simulation. The QTL was located on the centre of the chromosome, the effects are -0.5, 0, $+0.5$ for qq, Qq and QQ genotypes respectively, the random errors follow the standard normal distribution.

To compare our method with other methods, we performed analysis on the simulated data. First we use Haley–Knott regression to do QTL analysis with step size of 1 cM scanning. Then CI was calculated with nonparametric bootstrap method with 1000 replicates, 1 and 1.5-LOD support interval, formula method (Darvasi and Soller 1997) and our approach. Power of every method was measured by the frequency that CI covered the real QTL location among all the simulations.

QTL likelihood profile was influenced by the number of positions scanned; the more positions scanned the QTL likelihood profile was more accurate. In many cases, the QTL likelihood profile did not change much even when a small number of positions were scanned, e.g. profile of 1 cM genome scan is as similar as 2 cM genome scan. In theory nonparametric bootstrap CI and x-LOD support interval will be affected much by the step size of genome scan, because their accuracies are determined by step size of genome scan. However, the approach we presented here may not be affected much so long as the QTL likelihood profile is similar to that of dense genome scan. To test this guess we calculated the CI on 1, 2 and 3 cM genome scan results.

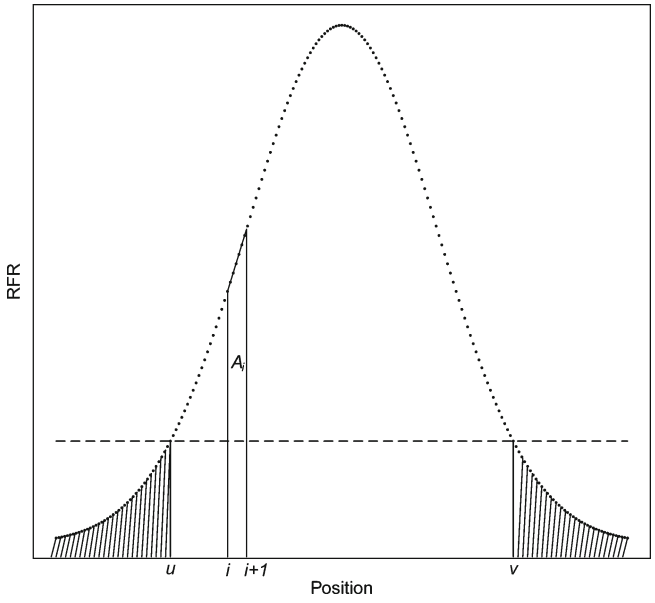

Figure 1. Diagram of calculation of confidence interval for QTL location. The dotted curve is relative frequency ratio (RFR) against position. Area (A_i) under the QTL profile from position i to $i+1$ is appropriately as the area of trapezoid (solid line). When the dashed line dropped until the total area of shaded parts equals significance to level, the position u and v are lower bound and upper bound of confidence interval for QTL location.

Real data

We use the linkage analysis result of two meat quality traits on SSC9 from an outbred swine population (Li *et al.* 2010). These two traits are post mortem pH value in longissmus dorsi (pHLM) and Minolta L* (L*). The details of population description, phenotype measurements and statistical analysis have been described in Li *et al.* (2010). We compared the CI using bootstrap method with 1000 replicates, 1-LOD and 1.5-LOD dropping method and our approach.

Results

Comparison of CI of different methods (table 1.) shows the lower and upper limits of Bootstrap CI and formula CI are not significantly different, the lower and upper limits of our method are between those of 1-LOD and 1.5-LOD support interval, but more close to 1-LOD support interval. Although 1.5-LOD support interval is wider than 1-LOD support interval, we did not find a significant difference between the two. However, among these CI, bootstrap has largest standard deviation. The powers of different methods to calculate CI are quite similar (table 1.), the formula method has slightly greater power than the others, but it also yields wider CI than the other methods.

The QTL CI under 1, 2 and 3 cM step size of genome scan (table 1.) shows that the CI becomes wider with the increase of step size of genome scan, but the changes are very small, indicating that so long as the QTL likelihood profile is determined approximately, the QTL CI can be determined largely with great accuracy.

Table 1. Confidence interval comparison with different methods.

Method	Lower limit*	Upper limit*	Power
Bootstrap	$38.78_a \pm 8.27$	$56.88_a \pm 10.68$	0.96
1-LOD	$42.70_b \pm 3.45$	$51.88_b \pm 2.79$	0.96
1.5-LOD	$41.18_b \pm 4.73$	$54.12_b \pm 6.53$	0.98
Formula**	$37.11_a \pm 4.95$	$57.77_a \pm 3.84$	0.98
1 cM***	$42.68_b \pm 4.29$	$51.91_b \pm 3.03$	0.96
2 cM***	$42.50_b \pm 4.24$	$52.13_b \pm 2.97$	0.96
3 cM***	$42.43_b \pm 4.11$	$52.61_b \pm 2.90$	0.96

*The means with same letter are not significantly different. **Confidence interval calculated using equation 1. ***Confidence interval of different step size of genome scan using our method.

Table 2. A quick method to calculate QTL confidence interval.

```
"qtl.ci" <- function (pos = NULL, lrt = NULL, alpha = 0.05) {
# pos: the vector of scanned positions
# lrt: -2*log-likelihood ratio test statistic
# alpha: the significance level

  tol <- alpha/100
  down <- 0.01       # drop extent each loop to search threshold
  np <- length(pos)

  # Step 1: Calculate RFR at each scanned position

  posq <- which.max(lrt)
  lrtq <- max(lrt)
  f <- exp((lrt - lrtq)/2)

  # Step 2: Calculate the area Ai for position i to i+1

  Area <- (pos[2:np] - pos[1:(np-1)]) * (f[1:(np-1)] + f[2:np])/2

  # Step 3: Scale Ai

  SArea <- sum(Area)
  Area <- Area/SArea

  # Step 4: calculate CI

  thr <- lrtq
  loop <- TRUE
  rp <- rev(pos)
  rf <- rev(f)

  while(loop)
  {
    thr <- thr - down
    fd <- exp((thr - lrtq)/2)

    tArea <- lArea <- rArea <- dl <- dr <- 0

    u <- (lrt <= thr)
    v <- rev(u)
    u <- sum(cumprod(u))
    v <- sum(cumprod(v))

    if (u > 0)
    {
      dl <- (pos[u+1]-pos[u])*(fd - f[u])/(f[u+1]-f[u])
      lArea <- 0.5*dl*(f[u] + fd)/SArea
    }

    if (u > 1) lArea <- lArea + sum(Area[1:(u-1)])
    if (v > 0)
    {
      dr <- (rp[v]-rp[v+1])*(fd - rf[v])/(rf[v+1]-rf[v])
      rArea <- 0.5*dr*(rf[v] + fd)/SArea
    }
```

Table 2 (*contd.*)

```
    if (v > 1) rArea <- rArea + sum(rev(Area)[1:(v-1)])

    tArea <- lArea + rArea

    if ((tArea <= alpha)|(abs(tArea-alpha)<= tol)) loop <- FALSE
  }
  if (u == 0) u <- 1
  if (v == 0) v <- 1

  lower <- pos[u] + dl
  upper <- rev(pos)[v] – dr
  return(c(lower,upper))
}
# End
```

Using the real data, the 95% CI for pHLM are 34–78, 32–82, 28–86, and 34.1–79.9 cM calculated by bootstrap, 1-LOD dropping, 1.5-LOD dropping and our approach respectively; and the 95% CI for Minolta L* are 36–98, 53–59, 33–100 and 35.9–98.7 cM calculated by bootstrap, 1-LOD dropping, 1.5-LOD dropping and our approach respectively. The CI calculated by our approach are similar as bootstrap for both traits. Both 1 and 1.5-LOD support interval are wider than bootstrap and our approach for pHLM; however for Minolta L*, 1-LOD support interval is narrower than those of three other methods.

Discussion

This work was motivated by high computational workload of bootstrapping in practical applications and the inapplicability of the formula method to outbred populations. The method presented here is fast and easy to use, in which the distribution of QTL position was approximated from likelihood. The formula method (Darvasi and Soller 1997; Weller and Soller 2004) is also quick, but loses generality because of symmetry assumption of CI and limitation to use with only inbred populations. From this viewpoint, our method is more general and suitable for any resource population. The x-LOD rule is also quick to get the CI of QTL location, but it is not stable as the results of real data shows sometimes the CI are close to 1-LOD support interval and sometimes close to 1.5-LOD support interval. The shape of QTL profile indicated the relative probability to harbour a true QTL, hence x-LOD rule is not stable, when the shape of QTL profile is sharper, CI may be underestimated; when the shape is flatter, CI may be overestimated. When the range of test statistics of each position is less than x, then x-LOD support interval will cover the entire chromosome. Bootstrap CI may behave poorly (Manichaikul *et al.* 2006), perhaps because bootstrap samples are only a very small part of the total possible bootstrap samples. For a dataset of 100 individuals, the number of total possible samples is $100^{100} = 10^{200}$, however suppose the number of bootstrap sample to determine CI is 10^5, it is only 10^{-195} of the total possible bootstrap samples, it may have a hgih probability to yield biased CI. Our approach shows similar power as the others and relatively smaller standard deviation, indicating it is both powerful and efficient.

In the real data results, CI for both traits calculated by our approach are similar to those by bootstrap. Bootstrap CI is used widely in QTL mapping experiment because its intuitiveness and independence of phenotypic distribution. Compared to the current CI calculation methods for QTL location, our method is quicker, independent of experimental design and the most important, it is directly based on the approximated distribution of QTL location. Simulation results shows that using our method the step size of genome scan does not affect CI much as long as the QTL likelihood profile can stand for the pattern under dense genome scan, it is efficient and powerful to determine the CI after genome scan analysis. The code used in this study was written in R Development Core Team (2010) and is included in table 2., but without guarantee.

Acknowledgement

This work was supported by National Institution General Research and Development Project (2009B001) of Chinese Academy of Fishery Sciences.

References

Bennewitz J., Reinsch N. and Kahn E. 2002 Improved confidence intervals in quantitative trait loci mapping by permutation bootstrapping. *Genetics* **160**, 1673–1686.

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

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

Haldane J. B. S. 1919 The combination of linkage values, and the calculation of distance between the loci of linked factors. *J. Genet.* **8**, 299–309.

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

Li H. D., Lund M. S., Christensen O. F., Gregersen V. R., Henckel P. and Bendixen C. 2010 Quantitative trait loci analysis of swine meat quality traits. *J. Anim. Sci.* **88**, 2904–2912.

Lynch M. and Walsh B. 1998 *Genetics and analysis of quantitative traits*. Sinauer Associates, Sunderland, MA, 01375 U.S.A.

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

Mangin B., Goffinet B. and Rebai A. 1994 Constructing confidence intervals for QTL location. *Genetics* **138**, 1301–1308.

Manichaikul A., Dupuis J., Sen S. and Broman K. W. 2006 Poor performance of bootstrap confidence intervals for the location of a quantitative trait locus. *Genetics* **174**, 481–489.

R Development Core Team 2010 *R: a language and environment for statistical computing*. Reference Index Version 2.12.0. R Foundation for Statistical Computing, Vienna, Austria (http://www.R-project.org).

Sugiyama F., Churchill G. A., Higgins D. C., Johns C., Makaritsis K. P., Gavras H. and Paigen B. 2001 Concordance of murine quantitative trait loci for salt-induced hypertension with rat and human loci. *Genomics* **71**, 70–77.

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

Walling G. A., Visscher P. M. and Haley C. S. 1998 A comparison of bootstrap methods to construct confidence intervals in QTL mapping. *Genet. Res.* **71**, 171–180.

Walling G. A., Haley C. S., Perez-Enciso M., Thompson R. and Visscher P. M. 2002 On the mapping of quantitative trait loci at marker and non-marker locations. *Genet. Res.* **79**, 97–106.

Weller J. I. and Soller M. 2004 An analytical formula to estimate confidence interval of QTL location with a saturated genetic map as a function of experimental design. *Theor. Appl. Genet.* **109**, 1224–1229.

Received 9 July 2010, in final revised form 8 February 2011; accepted 8 March 2011
Published on the Web: 19 August 2011