Selection Bias in Quantitative Trait Loci Mapping

C. Lee

Laboratory of Statistical Genetics, Ilsong Institute of Life Science, Hallym University, Anyang, Kyonggi-do 431-060, Korea.

Address correspondence to Chaeyoung Lee at the address above or e-mail: clee@hallym.ac.kr.

Abstract

A simulation study was performed to see whether selection affected quantitative trait loci (QTL) mapping. Populations under random selection, under selection among full-sib families, and under selection within a full-sib family were simulated each with heritability of 0.3, 0.5, and 0.7. They were analyzed with the marker spacing of 10 cM and 20 cM. The accuracy for QTL detection decreased for the populations under selection within full-sib family. Estimates of QTL effects and positions differed \( (P < .05) \) from their input values. The problems could be ignored when mapping a QTL for the populations under selection among full-sib families. A large heritability helped reduction of such problems. When the animals were selected within a full-sib family, the QTL was detected for the populations with heritability of 0.5 or larger using the marker spacing of 10 cM, and with heritability of 0.7 using the marker spacing of 20 cM. This study implied that when selection was introduced, the accuracy for QTL detection decreased and the estimates of QTL effects were biased. A caution was warranted on the decision of data (including selected animals to be genotyped) for QTL mapping.

The overwhelming number of molecular markers produced by genome projects provides a great opportunity to locate genes controlling the expression of quantitative traits. Various experimental designs have been utilized for quantitative trait loci (QTL) detection (Lynch and Walsh 1998). Inbred lines are, however, often limited in humans and animals, and in such cases QTL analyses are to utilize existing populations as they are. Either natural or artificial selection is usually introduced to the populations. Use of such populations exposes QTL detection to potential biases introduced by genotyping individuals, especially for economically important animals selected through breeding programs. The objectives of this study were to examine if such selection affects QTL mapping and to assess the impact on estimates of QTL effects and positions.

Materials and Methods

Simulation

A population was produced beginning with 80 founder individuals randomly assigned with QTL genotypes (40 males and 40 females). Each male parent was mated to three rotationally selected female parents. That is, the first male was mated to the 1st, 2nd, and 3rd females, the second male to the 2nd, 3rd, and 4th females, through the 40th male to the 40th, 1st, and 2nd females. The 120 full-sib families were produced in the F1 generation. Ten individuals were generated in each full-sib family, and a total of 1200 F1 individuals were produced. Sex was randomly assigned, producing 600 males and 600 females. Phenotypes were simulated, and 180 males and 180 individuals were selected and genotyped based on three criteria of the phenotypes. One was random selection and another was to select 36 of 120 full-sib families based on the large phenotypic means of individuals in the families. The third criterion was to select three individuals with large phenotypic values in each full-sib family. For 120 males and 120 females randomly selected among them, each male was mated to a randomly assigned female, producing 120 full-sib families. For the other animals, 60 males and 60 females were backcrossed to a randomly selected parent, again producing 120 full-sib families. Ten individuals were generated in each full-sib family, and a total of 2400 individuals were produced in the F2 generation. Phenotypes were simulated, and based on the three criteria described above, 720 individuals were selected and genotyped.

One chromosome was simulated with a length of 120 cM. The markers were generated every 10 cM and 20 cM, and a QTL was simulated on the chromosome to control the expression of a quantitative trait. The QTL was located at position 77 cM on the chromosome. The allelic effects of the QTL were simulated based on input values of their variances. Phenotypes were simulated by the following model:

\[ y_{ijk} = \mu + f_i + ax^f_j + dx^f_j + g_k + e_{ijk}, \]
Table 1. Codes of analyses designed in the simulation study

<table>
<thead>
<tr>
<th>Population under random selection</th>
<th>Marker spacing</th>
<th>Heritability</th>
<th>Analysis code</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10 cM</td>
<td>0.3</td>
<td>RS1H3</td>
</tr>
<tr>
<td></td>
<td>20 cM</td>
<td>0.3</td>
<td>RS1H7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>RS1H5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>RS1H7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>RS1H7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>RS1H7</td>
</tr>
<tr>
<td>Population under phenotype-based selection among full-sib families</td>
<td>Marker spacing</td>
<td>Heritability</td>
<td>Analysis code</td>
</tr>
<tr>
<td></td>
<td>10 cM</td>
<td>0.3</td>
<td>AS1H3</td>
</tr>
<tr>
<td></td>
<td>20 cM</td>
<td>0.3</td>
<td>AS1H7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>AS1H5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>AS1H7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>AS1H7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>AS1H7</td>
</tr>
<tr>
<td>Population under phenotype-based selection within full-sib family</td>
<td>Marker spacing</td>
<td>Heritability</td>
<td>Analysis code</td>
</tr>
<tr>
<td></td>
<td>10 cM</td>
<td>0.3</td>
<td>WS1H3</td>
</tr>
<tr>
<td></td>
<td>20 cM</td>
<td>0.3</td>
<td>WS1H7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>WS1H5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>WS1H7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>WS1H7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>WS1H7</td>
</tr>
</tbody>
</table>

where \( y_{ijk} \) is a phenotypic value, \( \mu \) is the overall mean for the trait, and 15.0 was simulated. The sex effect was ignored in this simulation. The variable \( f_j \) is the full-sib litter effect generated from \( N(0, 0.1) \). \( x'_j \) is the dummy variable for the additive QTL effect taking 1, 0, and -1 for genotypes QQ, Qq, and qq, respectively. \( x'_j \) is the dummy variable for the dominance effect taking 0, 1, and 0 for genotypes QQ, Qq, and qq. The \( a \) and \( d \) are regressions for the additive and dominance effects of QTL. The additive and dominance effects were generated, respectively, from \( N(a x'_j, \sigma^2_a) \) and \( N(d x'_j, \sigma^2_d) \). \( g_k \) is polygenic effect and the polygenic effect is defined as the genetic effect remaining after the QTL effect is accounted for. This polygenic effect consists of half of the parents’ polygenic effects (\( g_b \) for boar and \( g_s \) for sow) and Mendelian sampling (\( \phi_k \), that is, \( g_k = 1/2 g_b + 1/2 g_s + \phi_k \)).

Mendelian sampling was simulated using a normal distribution with mean zero and variance equal to the polygenic variance unexplained by parents’ polygenic effects. Since all the boars and sows in this population were not inbred, Mendelian sampling was generated from \( N(0, 0.5 \sigma^2_g) \). \( e_{jk} \) is the random environmental effect and is generated from \( N(0, \sigma^2_e) \). The input values for \( a_0 \) and \( d_0 \) were 2 and 1, respectively. The additive and dominance genetic variances and error variance were determined by the total heritability (\( h^2 \)) which was partitioned into the additive QTL heritability, the dominance QTL heritability, and the polygenic heritability. Hence the total heritability can be expressed with variance components as the following:

\[
h^2 = \frac{\sigma^2_a + \sigma^2_d + \sigma^2_g}{\sigma^2_a + \sigma^2_d + \sigma^2_g + \sigma^2_e}.
\]

The polygenic variance is defined as the fraction of genetic variance remaining after the QTL effect is accounted for.

The additive and dominance variances were calculated with Falconer and Mackay’s (1996) equations:

\[
\sigma^2_a = 2p_Q(1-p_Q)[a + d(1-2p_Q)]^2
\]

\[
\sigma^2_d = [2p_Q(1-p_Q)d]^2,
\]

where \( p_Q \) is the allele frequency of \( Q \) at the QTL.

Given a value of total heritability of the trait, the residual variance is expressed as

\[
\sigma^2_e = \frac{(1-h^2)(\sigma^2_a + \sigma^2_d + \sigma^2_g)}{h^2}.
\]

The input values of the total heritability in the simulation were 0.3, 0.5, and 0.7. A random number generator based on the Box-Muller method was used to generate random Gaussian deviates (Press et al. 1992). A total of nine populations were simulated, and 50 replicates were generated for each population. They were each analyzed with the marker spacing of 10 cM and 20 cM, and each analysis was coded as shown in Table 1.

QTL Mapping

The method for QTL analysis utilized in this study was devised first to estimate the marker genotype means using a mixed model that accounted for full-sib litter effects and then to estimate the QTL effects using a weighted least-squares analysis based on the conditional frequencies of QTL given marker genotypes. For details, see QTLMDSS (http://ilsongls.hallym.ac.kr/hwp/qtlmdss.doc).

Results

QTL Detection

Detection of QTL was performed for 18 analyses designed in the simulation study. The average likelihood ratio was estimated for each marker interval from 50 replicates (Figures 1–4). With the marker spacing of 10 cM, the likelihood ratio estimates were consistently the largest at the interval of markers located at 70 cM and 80 cM where a QTL was generated in the simulation (Figures 1–3). It was also true with the marker spacing of 20 cM (shown in Figure 4), with the maximum likelihood ratios located between 60 cM and 80 cM (data are not shown for populations simulated with heritability of 0.3 and 0.7). Regardless of marker spacing and heritability, the likelihood estimates obtained from the populations under random selection were a little bit larger in
the area of the QTL than those from the population under selection among full-sib families, and they were close in other areas.

On the other hand, the estimates from the populations under selection within the full-sib family were all considerably smaller than those from the other populations. Furthermore, QTL were not detected in the analyses of WS1H3, WS2H3, and WS2H5. Here we also observed that for a population simulated with a heritability of 0.5, the QTL was detected with a marker spacing of 10 cM (Figure 2), but not with a marker spacing of 20 cM (Figure 4). All the likelihood estimates drastically increased with a larger heritability.

QTL Effect and Position

The estimates of additive and dominance effects for the QTL are shown in Table 2. The estimates of its additive and dominance effects for the QTL are shown in Table 2. The estimates of its additive and dominance effects for the QTL are shown in Table 2.
dominance effects all corresponded to the input values for the populations under random selection and under selection among full-sib families \((P > .05)\), but not for the populations under selection within full-sib family \((P < .05)\). Their empirical standard errors were larger for the populations under selection within the full-sib family than for the other populations. Table 2 also shows that the estimates of linkage positions differ from their input values for the populations under selection within the full-sib family \((P < .05)\), but not for the populations under random selection and the populations under selection among full-sib families \((P > .05)\). Their standard errors were larger for the populations under selection within the full-sib family than for the other populations. Within such a population group, marker spacing influenced the linkage position estimates more than heritability. For example, within the population group under random selection, the estimates along with the change of heritability ranged from 76.59 to 76.68 with a marker spacing of 10 cM and from 76.35 to 76.44 with a marker spacing of 20 cM. However, the estimates along with the change in marker spacing ranged from 76.35 to 76.59 with a heritability of 0.3, from 76.43 to 76.66 with a heritability of 0.5, and from 76.44 to 76.68 with a heritability of 0.7.

### Discussion

Selection is often applied to domestic livestock to improve economically important traits. However, various data generated by such selection should be analyzed with caution. For example, genetic groups have been incorporated with the animal model that can correctly explain the first and second moments of animals’ genetic merits (Westell and Van Vleck 1987). Another example can be found in a simulation study where splitting data by gender affected estimates of variance and covariance components because analyzing records from only one sex did not account for selection in the other sex (Lee and Pollak 1997). This problem can be overcome by introducing a multiple trait model treating male and female records as different traits to explain the heterogeneity of the variance and covariance components by gender (Lee et al. 1997). The current study also points out the influence of selective genotyping on QTL mapping. However, the influence of selective genotyping on QTL mapping is not severe and is sometimes negligible (Ayoub and Mather 2002; Darvasi 1997; Ronin et al. 2003). This might be because, with selective genotyping, individuals from the high and low phenotypic extremes are usually genotyped.

The simulation study demonstrated that the accuracy for QTL detection decreased when selection was introduced. This problem was serious for the populations under selection within the full-sib family. Furthermore, estimates of QTL effects and positions were all different from their corresponding input values. However, these problems were negligible when mapping a QTL for the populations under selection among full-sib families. This might be because the full-sib litter effects confounded with selection effects could be largely explained by the analytical model introduced in this study. Although the selection and litter effects were
confounded in the analytical model, they were not the effects of much interest in such QTL detection. I suggested that the analytical model could help to overcome the selection problem in this special situation. Cautions should be used for QTL detection from existing populations when genotyping individuals selected through various breeding schemes.

The heritability level is an important factor in QTL mapping for data under a selection program. An obvious trend of improved QTL detection with a large heritability was observed in the current study. This agreed with the previous findings of Belknap (1998), Kearsey and Farquhar (1998), and Wu (1999). Noor et al. (2001) demonstrated that narrow marker spacing as well as large heritability had salutary effects on QTL detection efficacy. The current study further indicates that when the animals were selected within the full-sib family, the QTL was detected for the populations with a heritability of 0.5 or greater using the marker spacing of 10 cM, and with a heritability of 0.7 using the marker spacing of 20 cM. Further investigation of such a threshold level of heritability associated with marker spacing would be beneficial to determine appropriate experimental designs and optimal sample sizes for QTL analysis.

In conclusion, the simulation study suggested the need for a careful decision on animals to be genotyped for QTL mapping. When analyzing data under selection, the accuracy for QTL detection was decreased and the estimates of additive and dominance effects were biased. This study also implied that genotyping only one sex might create similar problems for a population under selection, because selection bias can be introduced by partitioning data by gender (Lee and Pollak 1997). I suggest the analytical model used in the current study for mapping QTL with data under selection among full-sib families.

Acknowledgments

This research was supported in parts by a grant from the BioGreen 21 Program, Rural Development Administration, Korea, and a research grant from Hallym University, Korea. The author would like to thank two anonymous reviewers for their valuable comments.

References


Received February 8, 2004
Accepted January 23, 2005

Corresponding Editor: Sudhir Kumar