Journal of Korean
Data & Information Science Society
2007, Vol. 18, No. 1, pp. 175 ~ 184

Bootstrapping of Hanwoo Chromosome17 Based on BMS1167 Microsatellite Locus

Jea-Young Lee[1] · Yong-Won Lee[2] · Jung-Sou Yeo[3]

Abstract

LOD scores and a permutation test for detecting and locating quantitative trait loci (QTL) from the Hanwoo economic trait have been described and we selected a considerable major BMS1167 locus for further analysis. K-means clustering analysis, for the major DNA marker mining of BMS1167 microsatellite loci in Hanwoo chromosome17, has been tried and three cluster groups divide four traits. The three cluster groups are classified according to eight DNA marker bps. Finally, we employed the bootstrap test method to calculate confidence intervals using the resampling method to find major DNA markers. We conclude that the major marker of BMS1167 locus in Hanwoo chromosome17 is only DNA marker 100bp.

Keywords : Bootstrap Method, K-means Clustering, LOD Score, Permutation Test, QTL

1. Introduction

Problems detecting and locating quantitative trait loci (QTL) have received considerable attention over the past several years. A variety of methods have been developed to analyze quantitative trait data (Weller 1986, Lander and Bostein 1989, Churchill and Deorge 1994). Many research groups (Hirano et al. 1998; Kim et al. 2003; Yeo et al. 2004)have intensively analyzed the linkage between markers and traits, in order to indentify the chromosomal regions responsible for economically important traits such as meat quality and carcass length. Some traits such as

1) Corresponding Author : (712-749) 경북 경산시 대동 214-1 영남대학교 이과대학 통계학과 교수 E-mail : jlee@yu.ac.kr
2) (712-749) 경북 경산시 대동 214-1 영남대학교 이과대학 통계학과 대학원
3) (712-749) 경북 경산시 대동 214-1 영남대학교 생명공학부 교수

"double muscle" in cattle, and RN in swine were revealed to be the results of particular genes (McPherron and Lee, 1997). Such identification of genes responsible for traits requires a huge amount of research, time and some luck. If gene arrangement along chromosomes is determined completely or nearly completely, one can select gene candidates for traits very efficiently, and speed up identification of the genes responsible for the traits. A common problem to all of these methods is the difficulty in determining appropriate significance thresholds (critical value) against which to compare test statistics (usually LOD scores or likelihood ratios) for the purpose of detecting QTL. Knott and Haley (1992) used simulation study for the distributional properties of likelihood ratio tests for QTL detection. They suggested that the chi-square approximation to the distribution of likelihood ratio test statistic is not reliable in many cases and requires further theoretical work. In 1994, Churchill and Doerge proposed permutation tests to detect the QTL effect in the genome. An introduction to the theory of permutation testing is provided by Good (1994).

In the work previous report(Lee et al.; 2006), we tried a method based on the concept of permutation testing of Good (1994), because major LOD scores don't have theoretical significance levels (critical values or p-values). Ten thousand repetitions of the permutation process were used for critical values. A microsatellite locus BM1167 was selected by permutation testing, which included 8 genes: DNA marker 100bp, 103bp, 105bp, 108bp, 110bp, 113bp, 115bp and 117bp. Next, the relations between DNA markers and the economic trait were identified by K-means clustering analysis. And in here, we applied the bootstrap test (Efron 1987; visscher et al., 1996) to calculate confidence intervals of QTL locations for traits. The number of bootstrap samples for each DNA was 1,000 and 95% confidence intervals were calculated for economically important traits.

2. Materials and Methods

2.1 Animals and Traits

Two hundred and sixty nine steers form 36 paternal half-sib families were used for linkage mapping and QTL from Hanwoo Improvement Center, National Agricultural Cooperation Federation, Korea. Daily weight gain was measured from birth to 720 days of age and marbling scores were measured at slaughter of 720 days of age. Marbling was scored as 19 degrees and classified by 1+, 1, 2 and 3 for marker systems. The grading of the marbling scores, backfat thickness and the M. longissimus dorsi area were measured according to standards of the Korean Animal Products Grading Service.

2.2 Bootstrapping (BCa (bias-corrected and accelerated)) Analysis

Sampling with replacement of n individual observations created bootstrap samples. An observation consists of a marker genotype and a phenotype, so at each bootstrap sample, we drew, with replacement, n observations out of the pool of (n) original observations. Some records can appear more than once in a bootstrap sample, while others are not included at all. After determining the n bootstrap samples, the empirical central 90 and 95% CI of the QTL positions were determined by ordering the n estimates and taking the bottom and top 5th and 2.5th percentile, respectively. The bootstrap idea is simply to replace the unknown population distribution with the known empirical distribution function. The bootstrap distribution for $\hat{\theta}-\theta$ is the distribution determined by generating $\hat{\Theta}$ values which are determined by sampling independently with the replacement form empirical distribution F_n. The bootstrap estimate of the standard error of $\hat{\Theta}$ then becomes the standard deviation of the bootstrap distribution for $\hat{\theta}-\theta$.

It should be noted here that almost any parameter of the bootstrap distribution may serve as a "bootstrap" estimate of the corresponding population parameter. We could consider the skewness, the kurtosis, the median, or the 95th percentile of the bootstrap distribution for $\hat{\theta}$.

The basic idea behind the bootstrap is that the variability of θ^* around $\hat{\theta}$ will be similar to the variability of $\hat{\theta}$ around σ. There is good reason to believe this will be true for large sample sizes, since we see that as n grows larger, F_n becomes comparable to random sampling from F.

We have the following steps to produce BCa (*bias-corrected and accelerated*) bootstrap intervals:

Step 1 : Generate a sample of size n with replacement from the empirical distribution

Step 2 : Compute θ^*, the value of $\hat{\theta}$ obtained by using the bootstrap sample in place of the original sample

Step 3 : Repeat steps 1 and 2 k times.
By replicating steps 1 and $2k$ times, we obtain a Monte Carlo approximation to the distribution of θ^*. Let $\hat{\theta}^*_{(\alpha)}$ indicate the $100 \times \alpha$th percentile of B=1000 bootstrap replications $\hat{\theta}^*_{(1)}, \hat{\theta}^*_{(2)}, \ldots, \hat{\theta}^*_{(B=1000)}$.

Step 4 : The BCa interval end points are also given by percentiles of the bootstrap distribution. The percentiles used however, depend on two numbers, $\hat{\alpha}$(acceleration) and Z_0(bias-correction).

The BCa interval of intended coverage 1-2α is given by

$$BCa;(\hat{\theta}_{lo,}\,\hat{\theta}_{up}) = (\hat{\theta}^{*(\alpha 1)},\,\hat{\theta}^{*(\alpha 2)})$$

Where

$$\alpha_1 = \Phi\hat{Z}_0 + (\hat{Z}_0 + Z^{(\alpha)})/[1-\hat{\alpha}(\hat{X}_0 + Z^{(\alpha)})], \qquad \alpha_2 = \Phi\hat{Z}_0 + (\hat{Z}_0 + Z^{(1-\alpha)}/[1-\hat{\alpha}(\hat{X}_0 + Z^{(1-\alpha)})]$$

Here $\Phi(\bullet)$ is the standard normal cumulative distribution function.
$Z^{(\alpha)}$ is the 100th percentile point of standard normal distribution.

If $\hat{\alpha}$ and $\hat{Z}_0$ equal zero, then the BCa interval is the same as the percentile interval.
If $\hat{\alpha}$ and $\hat{Z}_0$ are not equal to zero, then the BCa interval endpoints change. Bias-correction $\hat{Z}_0$ is obtained from

$$\hat{Z}_0 = \Phi^{-1}[\frac{\sum_{b=1}^{n} I(\hat{\theta}^{*}(b) < \hat{\theta})}{B}]$$

Where F^{-1} is the inverse function of the standard normal cumulative distribution function.

3. Results

3.1 QTL Methodology

LOD scores and the permutation test for detecting and locating quantitative trait loci (QTL) from the Hanwoo economic traits are given in table 1. We selected several loci that had maximum LOD scores exceeding3, which is generally considered significant (Chotai, 1984). However, LOD scores at which significance is declared cannot be obtained theoretically; therefore we applied the genome wise (experiment wise) permutation test (Churchill and Deorge, 1994). An empirical 100(1-P) percentile obtained by 10,000 repetition of permutation process for each locus was referred to as an estimated critical value of the genome wise significance level of P. The critical value of P=0.01 was used to detect the presence of a QTL somewhere in the genome so that the type I error rate may be 0.01 or less (Table 1).
In Table 1, BMS8125, BMS499, BMS941, BMS1167 and HUJ223 are very significant levels of P, but other loci are not significance statistically. In particular, BMS941, BMS499 and BMS1167 were demonstrated best. In this paper, we first

want to try a major DNA marker mining of BMS1167 microsatellite locus only in Hanwoo chromosome17.

3.2 K-means Clustering and Results

Two hundred and sixty nine steers from Hanwoo Improvement Center, National Agricultural Cooperation Federation, Korea were used for the analysis. We analyzed the BMS1167 micro locus in chromosome17. Eight DNA markers were obtained, including 100,

<Table 1> QTL and permutation test of Hanwoo chromosome17 based on economic traits

Loci	Economic Traits							
	Marbling score		Daily gain		Backfat thickness		M. Longissimus dorsi area	
	Lod Score (P-value)	Ratio of QTL variation(%)	Lod Score (P-value)	Ratio of QTL variation(%)	Lod Score (P-value)	Ratio of QTL variation(%)	Lod Score (P-value)	Ratio of QTL variation(%)
BMS8125	3.38 (0.000)	6.22	5.28 (0.000)	9.43	2.01 (0.004)	7.54	9.24 (0.000)	15.33
BMS1825	4.67 (0.000)	8.62	1.43 (0.072)	2.83	2.27 (0.028)	6.57	5.03 (0.000)	9.22
BMS499	3.98 (0.000)	7.39	3.19 (0.000)	6.02	4.6 (0.000)	10.39	9.27 (0.000)	15.65
BMS941	3.75 (0.000)	7.29	5.7 (0.000)	10.72	4.07 (0.000)	4.96	5.11 (0.000)	9.67
BMS1101	3.15 (0.000)	5.89	2.04 (0.119)	3.92	4.28 (0.000)	6.48	6.02 (0.000)	10.68
BMS1879	3.49 (0.000)	6.76	4.98 (0.000)	9.4	2.52 (0.015)	3.02	5.55 (0.000)	10.33
BMS1167	4.71 (0.000)	8.8	6.14 (0.000)	11.22	5.7 (0.000)	9.96	4.98 (0.000)	9.26
HUJ223	3.58 (0.000)	6.46	2.3 (0.002)	4.26	5.04 (0.000)	6.25	2.98 (0.003)	5.44
CSSM033	3.79 (0.000)	7.08	3.39 (0.000)	6.42	1.59 (0.012)	3.41	2.27 (0.011)	4.38

· Test statistic for this significance level is the sum of observations in the first sample(Good, 1994)

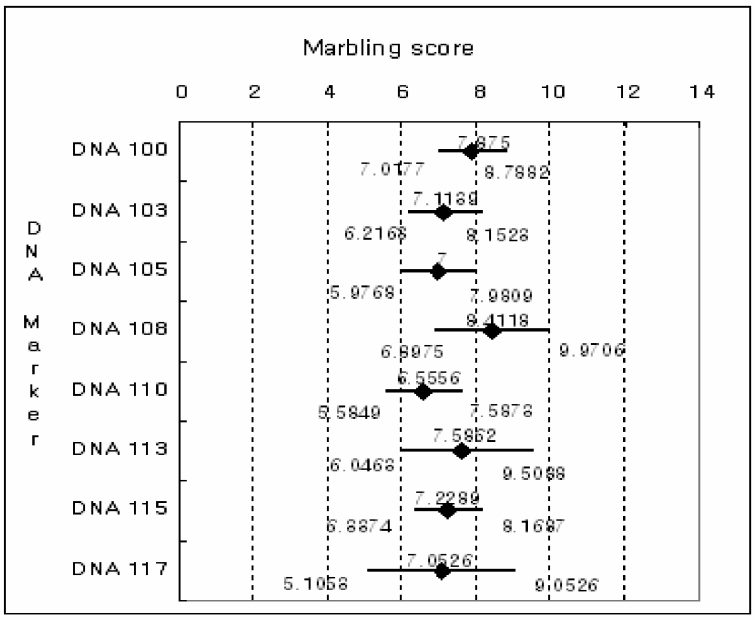
103, 105bp etc., as well as four economic traits data, which were marbling score, daily gain, backfat thickness, and *M. longissimus dorsi* area.

The K-means clustering analysis method applied to the four traits and eight DNA markers resulted in three cluster groups (table 2,3 and figure. 1; see Lee *et al.*,

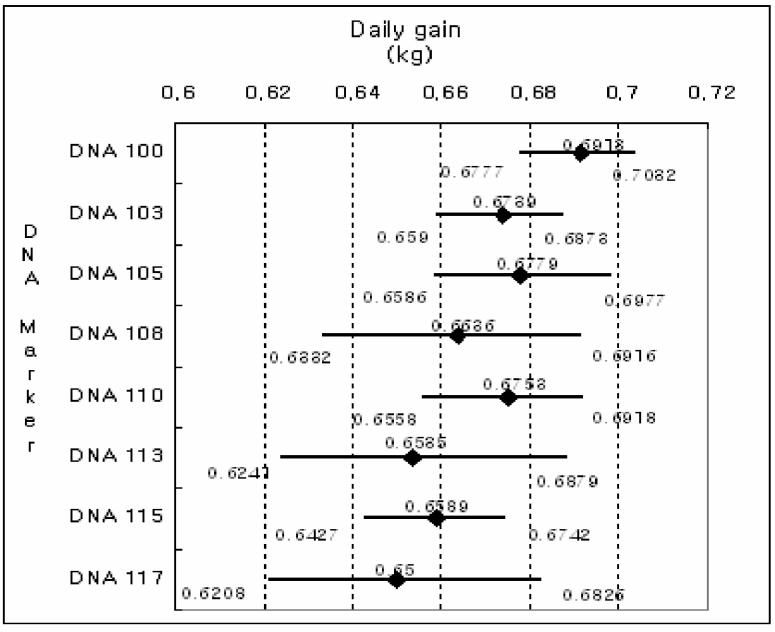
2006). In table2, We can conclude that cluster 1 is a useful group for backfat thickness (high value=0.439), cluster2 is a useful group for marbling score (hight value=1.219), and cluster3 is a useful group for daily gain and the *M. longissimus dorsi* area. Figure 1, which represents clustering ratio comparisons for DNA markers, cluster 1 has the greatest proportion of DNA markers 103, 110 and 117bp, cluster2 has the greatest proportion of 100 and 108bp, and cluster3 has the greatest proportion of 100, 105, 113 and 115bp. Similarly, we recorded standardized mean result of the four economic traits, compared with DNA markers in table3. DNA markers 100 and 108bp present a higher marbling score, markers 113 and 115bp present higher backfat thickness. DNA marker 100bp presents higher *M. longissimus dorsi* area or daily again.
A summary of the results is given in table 4(Lee *et al.*, 2006). That is, DNA markers 100 and 108bp are useful for marbling score; marker 100bp for daily gain and the *M.longissimus dorsi* area. DNA markers 110, 113 and 115bp are important for backfat in the mean result, but they do not match the result of K-means mining, which may be in sufficient for the conclusion.
Therefore, we decided to try bootstrap testing for our final decision.

3.3 Bootstrap (BCa method) Analysis

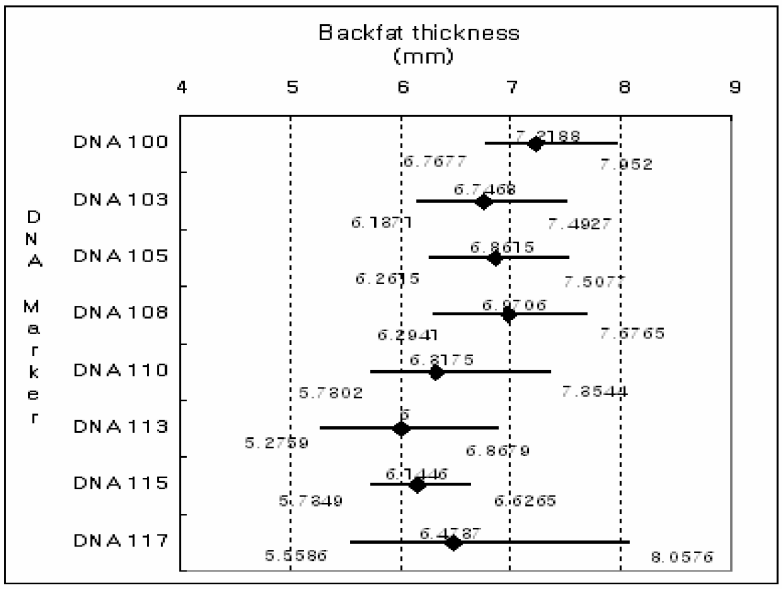
We applied the bootstrap testing method (Visscher et al., 1996) to calculate confidence intervals for finding major DNA markers. Bootstrap samples were created by sampling with replacement each individual DNA marker and trait. The number of bootstrap samples for each DNA was 1,000 and 95% confidence intervals of bootstrap testing were calculated for four traits, i.e. marbling score, daily gain, backfat thickness and *M. longissimus dorsi* area(figures1 through 5).
Figure 1 shows that marker 100 and 108bp have better marbling intervals(7.0177~8.7882) & (6.8975~9.9706) and means 7.8750 & 8.4118 respectively than others. In figure 2, marker 100bp has especially good confidence intervals for daily gain. Figure 3 shows that marker113bp and 115bp have lower backfat thickness confidence intervals (5.2759~6.8679) and (5.7849~6.6265) respectively, which are good but it is only a good influence marker for backfat thickness. In figure 4, we have a good DNA marker 100bp for *M. longissimus dorsi* area. This means that marker 100bp is only good for the K-means clustering method and for bootstrap intervals.



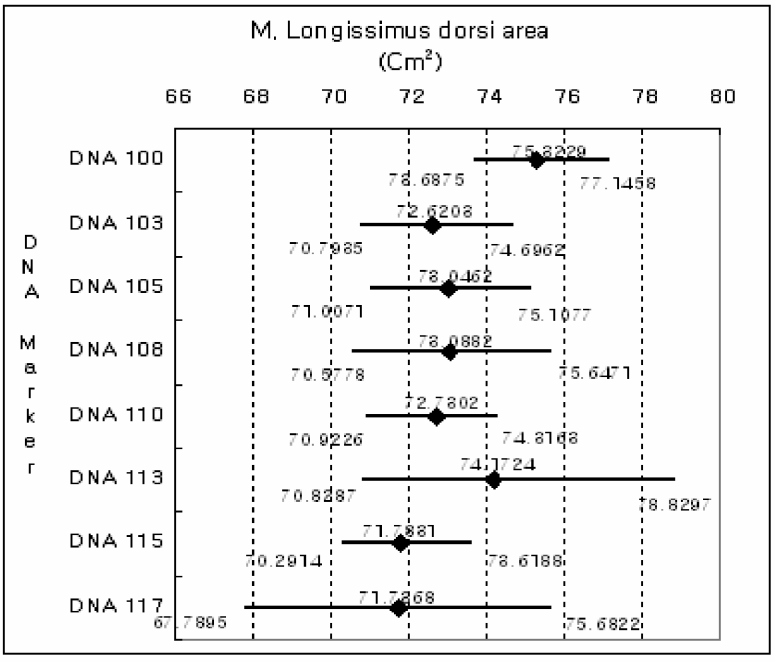
<Figure 1> Bootstrap confidence intervals of BMS1167 for marbling score



<Figure 2> Bootstrap confidence intervals of BMS1167 for daily gain



<Figure 3> Bootstrap confidence intervals of BMS1167 for backfat thickness



<Figure 4> Bootstrap confidence intervals of BMS1167 for M. Longissimus dorsi area

4. Discussion

LOD scores related to economic traits and the permutation test have been applied for the purpose of detecting QTL. QTL for BMS8125, BMS499, BMS941, BMS1167 and HUJ223 are very significant levels of P, but other loci are significant statistically. In particular, BMS941, BMS499, and BMS1167 were demonstrated best. A BMS1167 microsatellite was selected as one of the considerable major loci. Next, K-means clustering analysis for the major DNA marker mining of BMS1167 microsatellite loci in Hanwoo chromosome17, has been described and three cluster groups have been divide into four traits. The three cluster groups are classified according to genes (DNA marker bps). DNA marker 100, 108 and 110bp were selected as being the most useful genes in the BMS1167 locus. We applied the bootstrap test to calculate confidence intervals for traits. Finally, DNA marker 100bp only showed to be a good influence marker for four economic traits. We conclude that the major marker of BMS1167 locus in Hanwoo chromosome17 is a DNA marker 100bp.

References

1. Chotai, j. 1984 On the lod score method in linkage analysis. *Ann. Hum. Genet* 48:359-378.
2. Churchill G.A. and Doerge R. W. 1994 Empirical Threshold Values for Quantitative Trait Mapping, *Genetics* 138:963-971.
3. Efron B. 1987 Better bootstrap confidence intervals. *J. of Amer. Stat. Assoc.* 82:171-199.
4. Good, P. (1994) *Permutation Test : A peactical Guide to resampling for testing Hypothesis*, Spring-Verlag, New York.
5. Hirano T.,Kobayashi N., Nakamaru T., Hara K. and Sugimoto Y. 1998 Linkage analysis of meat quality in Wagyu. *26th International Conference on Animal Genetics*, Auckland, New Zealand:E019.
6. Kim M. J., Lee J. Y., Yeo J. S., Lee Y. W. and Joe Y. J. 2003. A major DNA mining of BM4311 in Hanwoo. *Proceedings of the Spring Conference*, Cheju National Univ. 305-311.
7. Knott S. A. and Haley C. S. 1992 Aspects of maximum likelihood method methods for the mapping of quantitative trait loci in line crosses. *Genet. Res.* 60:139-151
8. Lander E. and Bostein D. 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121; 185-199.
9. Lee J., Lee Y. and Kwon J. 2006 DNA marker mining of BMS1167 microsatellite locus in Hanwoo chromosome17. *J. korean Data and*

Info. 17-2:325-333
10. MacQueen, J. B. 1967 *Some methods for classification and analysis of multivariate observations.* Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, 1, Berkeley, California: University of Press, 281-297.
11. McPherron A. C and Lee S. J. 1997 Double muscling in cattle due to mutations in the myostatin gene. *Proc. Netl. Acad. Sci* USA23, 12457-12461.
12. Visscher Peter, Robin Thompson and Chris S. Haley 1996 Confidence intervals in QTL mapping by bootstrapping. *Genetics* 143; 1013-1020.
13. Weller, J. I. 1986 Maximum likelihood techniques for the mapping and analysis quantitative trait loci with the aid of genetic markers. *Biometrics* 42: 627-640.
14. Yeo, J. S., Lee, J. Y. and Kim, J. W. 2004 DNA marker mining of ILSTS035 microsatellite locus on chromosome6 of Hanwoo cattle. *J. Genetics.* 83; 245-250.

[received date : Oct. 2006, accepted date : Dec. 2006]