

Bootstrap Analysis and Major DNA Markers of BM4311 Microsatellite Locus in Hanwoo Chromosome 6

Jung-Sou Yeo^{1*}, Jae-Woo Kim¹, Hyo-Sub Shin¹ and Jea-Young Lee²

Department of Animal Science, Yeungnam University, Gyeongsan, Gyeongbuk 712-749, Korea

ABSTRACT : LOD scores related to marbling scores and permutation test have been applied for the purpose detecting quantitative trait loci (QTL) and we selected a considerable major locus BM4311. K-means clustering, for the major DNA marker mining of BM4311 microsatellite loci in Hanwoo chromosome 6, has been tried and five traits are divided by three cluster groups. Then, the three cluster groups are classified according to six DNA markers. Finally, bootstrap test method to calculate confidence intervals, using resampling method, has been adapted in order to find major DNA markers. It could be concluded that the major markers of BM4311 locus in Hanwoo chromosome 6 were DNA marker 100 and 95 bp. (*Asian-Aust. J. Anim. Sci.* 2004, Vol 17, No. 8 : 1033-1038)

Key Words : Bootstrap Analysis, LOD Score, QTL, Permutation Test, K-means Clustering

INTRODUCTION

The problems on 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 have intensively analyzed the linkage between markers and traits, in order to identify the chromosomal regions responsible for economically important traits such as meat quality and carcass length. Some traits such as "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 huge research work, time and some luckiness. If gene arrangement along chromosomes is determined completely or nearly completely, one can select genes candidate 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 of 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 is needed further theoretical work. In 1994, Churchill and Deorge proposed permutation

tests to detect QTL effect in the genome. An introduction to the theory of permutation testing is provided by Good (1994).

In this paper, we try a method based on the concept of permutation test (Good, 1994). Because major LOD scores candidate don't have theoretical significant levels (critical value or p-value), 10,000 repetitions of permutation process were used for critical value. BM4311 microsatellite locus was selected by permutation test and includes 6 genes which are DNA marker 95, 100, 103, 105, 107 and 110 bp. Next, the relations between DNA markers and the economic trait are identified by K-means clustering analysis. Finally, we applied the bootstrap test (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 (Figure 2 through 6).

MATERIALS AND METHODS

Animals and trait

One hundred thirty seven steers from 10 paternal half-sib families were used for linkage mapping and QTL from Hanwoo Improvement Center, National Agricultural Cooperation Federation, Korea. Daily gain from birth to 720 days of age and marbling scores at slaughter of 720 days of age were measured. Marbling was scored as 19 degrees and classified by 1+, 1, 2 and 3 for market systems. The grading of marbling score, backfat thickness and *M. longissimus dorsi* area was measured according to standards of the Korean Animal Products Grading Service.

Permutation tests

LOD graphs for detecting and locating quantitative trait loci (QTL) from the Hanwoo marbling scores (exceed 3) have been selected in Table 1. But LOD scores at which

* Corresponding Author: Jung-Sou Yeo. Tel: +82-53-810-2936, Fax: +82-53-813-2936, E-mail: jsyeo@yu.ac.kr

¹ Institute of Biotechnology, Yeungnam University, Gyeongsan, Gyeongbuk 712-749, Korea.

² Department of Statistics, Yeungnam University, Gyeongsan, Gyeongbuk 712-749, Korea.

Received September 9, 2003; Accepted April 13, 2004

Table 1. Permutation test results of Hanwoo chromosome 6 based on marbling score

Trait	Loci	LOD score	P-value*	Ratio of QTL variation (%)
Marbling score	BM3026	3.501999070	p<0.01	9.37
	BMS690	4.602844364	p<0.01	12.53
	ILSTS035	4.991694330	p<0.01	16.54
	BM4311	6.594214985	p<0.01	16.38
	BMS511	4.079206916	p<0.01	11.03
	AFR227	3.150859863	0.07235	19.47
	BMC4203	2.915809464	p<0.01	7.88

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

Table 2. K-means clustering analysis for BM4311

Traits	Cluster 1 (58)	Cluster 2 (36)	Cluster 3 (43)
Marbling score	-0.552124	1.265126	-0.436007
Daily gain	-0.805859	0.253847	0.891291
Backfat thickness	0.34431	0.223596	-0.270384
<i>M. longissimus dorsi</i> area	-0.602908	-0.478173	0.390335
Carcass weight	-0.875109	0.410955	0.83811

(): total number of individuals.

significance is declared cannot be obtained theoretically, we applied the genomewide (experimentwise) permutation test (Churchill and Doerge, 1994). A permutation test in the simplest case is used to detect a location shift in data that are divided into two sets of observations. We will follow the five-step procedures (Good, 1994: p20):

Step 1 Analyze the problem (hypothesis, distribution drawn etc.).

Step 2 Choose the test statistic (sum of observations in first sample) which will distinguish the hypothesis.

Step 3 Compute the test statistic for the original labeling of the observations.

Step 4 Rearrange (permute) the labels and recomputed the test statistic for the arrange labels. Repeat until you obtain the distribution of the test statistic for possible permutations.

Step 5 Calculate labels of significances using this permutation distribution of the statistic.

An empirical 100 (1-P) percentile obtained by 10,000 repetition of permutation process was referred to as an estimated critical value of genomewide 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). Six loci, including BM4311, have been selected. After permutation test, we need to identify the major DNA marker mining of BM4311 based on economically important traits such as meat quality and carcass length.

K-means clustering method

Grouping or clustering can provide an informal means for assessing dimensionality, identifying outliers and suggesting interesting hypotheses concerning relationships. K-means method, which was suggested by MacQueen, is non-hierarchical clustering techniques. The process is composed of the following three steps:

Step 1 Partition the items into *k* initial clusters.

Step 2 Proceed through the list of items, assigning an item to the cluster whose centroid (mean) *m_t* (*t* = 1, ..., *k*) is nearest. Recalculate the centroid for the cluster receiving the new item and for the cluster losing the item.

Step 3 Repeat step 2 until no more reassignments take place.

Distance is usually computed using Euclidean distance with either standardized or unstandardized observations *x_i* (*i*=1, ..., *n*). That is, from data matrix *X* and variance-covariance matrix *S*.

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \dots & x_{pn} \end{bmatrix} = [x_1 \ x_2 \ \dots \ x_n]$$

$$S = \begin{bmatrix} s_{11}^2 & s_{12} & \dots & s_{1p} \\ s_{21} & s_{22}^2 & \dots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \dots & s_{pp}^2 \end{bmatrix}$$

the Euclidean distance depends on

$$d(x_i, m_t) = [\sum_{h=1}^p (x_{ih} - m_{th})^2]^{1/2} \quad (h = 1, A, p)$$

(hand standardized Euclidean distance depends on

$$d(x_i, m_t) = [\sum_{h=1}^p (x_{ih} - m_{th}) / s_{hh}]^2]^{1/2} \quad (h = 1, A, p)$$

Rather than starting with a partition of all items into K preliminary groups in step 1, we could specify K initial centroids and then proceed to step 2.

The K-means for BM4311 results have been obtained in Table 2 through Table 4 with Figure 1.

That is, five traits are divided by three cluster groups and the cluster groups are classified according to genes (DNA marker bps).

Bootstrapping (BCa (bias-corrected and accelerated) analysis

Bootstrap samples were created by sampling with replacement of *n* individual observations. An observation consists of a marker genotype and a phenotype. So, at each bootstrap sample, we draw, 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 *n* bootstrap samples, the empirical central 90 and 95% confidence interval (CI) of the QTL positions was 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 $\hat{\theta}$ is the distribution obtained by generating $\hat{\theta}$ values by sampling independently with replacement from the empirical distribution F_n . The bootstrap estimate of the standard error of $\hat{\theta}$ is then the standard deviation of the bootstrap distribution for $\hat{\theta}$.

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 $\hat{\theta}^*$ 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 as *n* gets larger and larger F_n is almost like 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 $\hat{\theta}^*$, 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 2 *k* times, we obtain a Monte Carlo approximation to the distribution of $\hat{\theta}^*$. Let $\hat{\theta}_{(a)}^*$ indicate the 100×*a*th percentile of *B*=1,000 bootstrap replications $\hat{\theta}_{(1)}^*, \hat{\theta}_{(2)}^*, \dots, \hat{\theta}_{(B)}^*$.

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

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

$$BCa : (\hat{\theta}_{lo}, \hat{\theta}_{up}) = (\hat{\theta}_{(u1)}^*, \hat{\theta}_{(u2)}^*)$$

$$\alpha_1 = \Phi \left(\frac{\hat{z}_0 + (Z^{(\alpha)} - \hat{\alpha}(\hat{\mathbf{x}}_0 + Z^{(\alpha)}))}{1 - \hat{\alpha}(\hat{\mathbf{x}}_0 + Z^{(\alpha)})} \right)$$

$$\alpha_2 = \Phi \left(\frac{\hat{z}_0 + (Z^{(1-\alpha)} - \hat{\alpha}(\hat{\mathbf{x}}_0 + Z^{(1-\alpha)}))}{1 - \hat{\alpha}(\hat{\mathbf{x}}_0 + Z^{(1-\alpha)})} \right)$$

Here $\Phi(\cdot)$ 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-zero then the BCa interval endpoints change. Bias-correction Z_0 is obtained

$$Z_0 = \Phi^{-1} \left(\frac{\sum_{i=1}^n I(\hat{\theta}_i < \hat{\theta})}{n} \right)$$

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

RESULTS

QTL methodology

LOD graph and permutation test for detecting and locating quantitative trait loci (QTL) from the Hanwoo marbling scores have been selected in Table 1. We select several candidate loci that the maximum LOD score exceeds 3 (It is generally considered significant, Chotai, 1984). But LOD scores at which significance is declared cannot be obtained theoretically; we applied the genomewide (experimentwise) permutation test (Churchill and Doerge, 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 genomewide 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, AFR227 is not significant statistically, but others keep very significance level of P. In particular, ILSTS035 and BM4311 were demonstrated best. In this paper, first, we want to try a major DNA marker mining of BM4311 microsatellite locus in Hanwoo chromosome 6.

K-means clustering and results

One hundred thirty seven steers from Hanwoo Improvement Center, National Agricultural Cooperation Federation, Korea have been used for the analysis. We analyzed BM4311 micro locus in chromosome 6. The obtained DNA markers were eleven including 95, 100, 103 bp etc. and five economic traits which were marbling score, daily gain, backfat thickness, *M. longissimus dorsi* area and carcass weight.

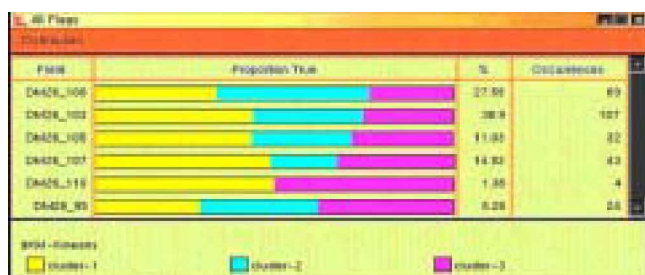
Table 3. Standardized mean results of five traits and DNA markers in BM4311

Traits	DNA markers					
	95 bp (24)	100 bp (80)	103 bp (107)	105 bp (32)	107 bp (43)	110 bp (4)
Marbling score	0.0946	0.2996	0.0633	-0.1014	-0.1239	-0.6893
Daily gain	0.3169	0.0359	-0.0285	-0.0359	0.0241	0.1140
Backfat thickness	0.2263	-0.0182	0.0727	0.0051	-0.0459	0.2866
<i>M. longissimus dorsi</i> area	0.3021	0.0515	-0.0440	-0.1678	0.0985	-0.0674
Carcass weight	0.3497	0.0783	-0.0803	-0.0831	0.0368	0.1860

(): total number of individuals.

Table 4. Clustering comparison between standardized means and K-means mining results

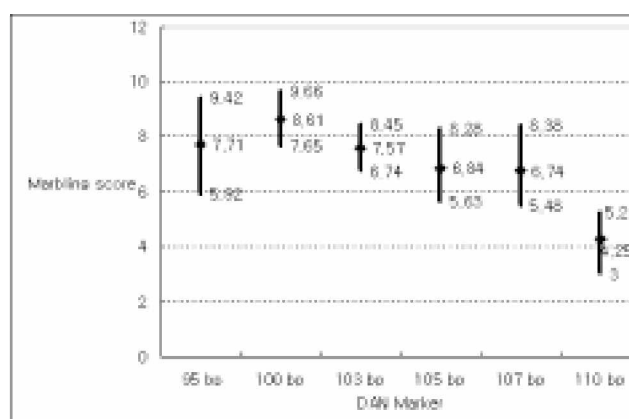
Cluster group	Mean result	K-means mining
Backfat thickness	95 and 110 bp	103, 105, 107 and 110 bp
Marbling score	100 bp	100 and 95 bp
Daily gain, <i>M. longissimus dorsi</i> area and carcass weight	95 and 110 bp	95 and 110 bp

**Figure 1.** Clustering proportional comparisons analysis for DNA markers (DM).

K-means clustering analysis method with those data above has been applied and five traits were divided by three cluster groups in Table 2 and DNA markers were clustered in Figure 1. In Table 2, we conclude that cluster 1 is backfat thickness useful group (high value=0.34431). Cluster 2 is marbling score useful group (high value=1.265126). Cluster 3 is carcass weight, daily gain and *M. longissimus dorsi* area useful group. Next, Figure 1 represents clustering proportional comparisons for DNA markers. Cluster 1 has much proportion for DNA marker 103, 105, 107, and 110 bp, Cluster 2 has much proportion for 100bp, and Cluster 3 has much proportion for 95 and 110 bp. But 110 bp is very a few (n=4) and it may not sufficient to decide conclusions.

Similarly, we have standardized mean results of DNA markers in BM4311 based on traits, in Table 3. DNA marker 100 bp has higher standardized marbling score (0.2996), 95 bp has higher values for all traits except marbling score. A marker 110 bp has higher backfat thickness and carcass weight.

Based on Table 2, 3 and Figure 1, we summarize the results in Table 4. That is, DNA marker 110 bp is useful for backfat thickness but not enough data. 100 bp is useful for marbling score and 95 and 110 bp are useful for daily gain, *M. longissimus dorsi* area and carcass weight. But, although 110bp may be an important for backfat but the individuals are just 4, it is insufficient for the conclusion. Now, we want

**Figure 2.** Bootstrap confidence intervals for BM4311 marbling score.

to try more specific analysis for DNA marker that is bootstrap testing method.

Bootstrap (BCa method) analysis

Bootstrap samples were created by sampling with replacement each individual DNA marker and trait. We applied the bootstrap testing method (Visscher et al., 1996) to calculate confidence intervals for finding major DNA markers. The number of bootstrap samples for each DNA were 1,000 and 95% confidence intervals of bootstrap testing were calculated for five traits, i.e. marbling score, daily gain, backfat thickness, *M. longissimus dorsi* area and carcass weight (Figure 2 through 6).

In Figure 2, DNA marker 100 bp has better marbling interval (7.6536-9.6626) and mean 8.6125 than others. DNA 100 bp has a little better bootstrap confidence interval for backfat thickness in Figure 3. In Figure 4, DNA 95 bp has a high mean value 0.6898, but a little wide confidence interval for daily gain. In Figure 5 and 6, DNA 95 bp is good trait about *M. longissimus dorsi* area and carcass weight. But DNA 110 bp showed that it is bad influence marker for *M. longissimus dorsi* area and carcass weight.

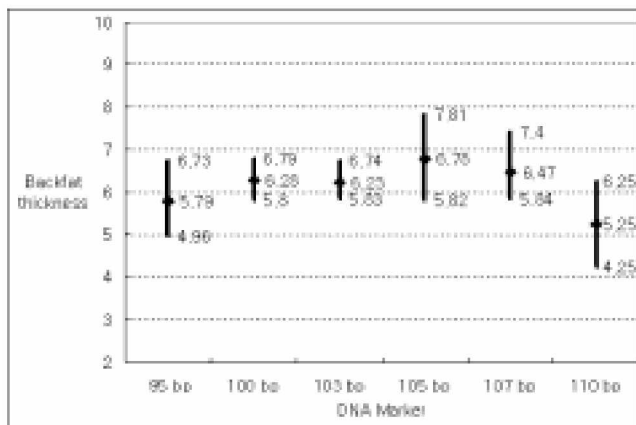


Figure 3. Bootstrap confidence intervals for BM4311 Backfat thickness

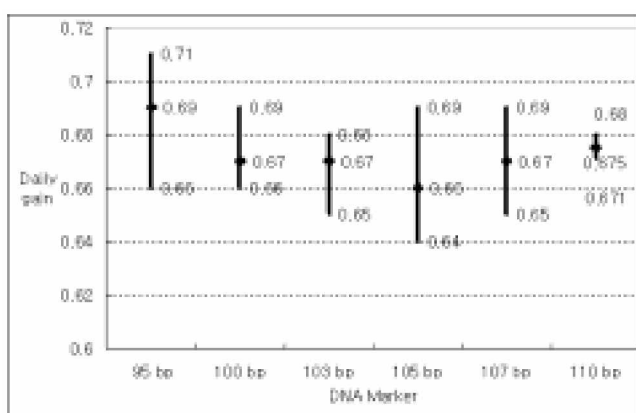


Figure 4. Bootstrap confidence intervals for BM4311 daily gain

Therefore, we conclude that DNA marker 100 bp presents good confidence interval for marbling score and 95 bp for *M. longissimus dorsi* area and carcass weight.

DISCUSSION

LOD scores related to marbling scores and permutation test have been applied for the major DNA marker mining. QTLs for BM3026, BMS690, ILSTS035, BM4311, BMS511 and BMC4203 were demonstrated significances, whereas no significant QTL was detected for AFR227. BM4311 microsatellite was selected as most considerable major locus. Next, K-means clustering, for the major DNA marker mining of BM4311 microsatellite loci in Hanwoo chromosome 6, has been tried and five traits are divided by three cluster groups. Then, the three cluster groups are classified according to genes (DNA marker bps). It was shown DNA maker 95, 100 and 110 bp are most useful genes in BM4311 locus. By the way, DNA marker 110 bp is useful for backfat thickness but not enough data. DNA marker 100 bp is useful for marbling score. DNA marker 95 and 110 bp are useful for daily gain, *M. longissimus dorsi* area and carcass weight. Furthermore, the sample size 4 of

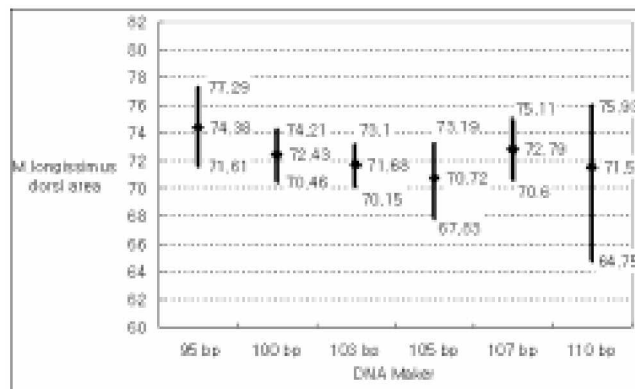


Figure 5. Bootstrap confidence intervals for BM4311 *M. longissimus dorsi* area.

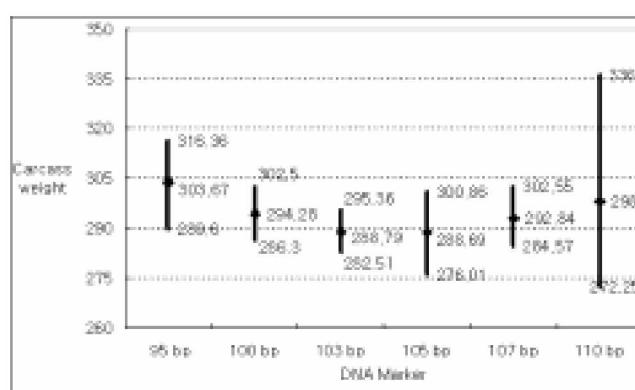


Figure 6. Bootstrap confidence intervals for BM4311 carcass weight.

DNA marker 110 bp is insufficient for the conclusion. Therefore, we applied the bootstrap test to calculate confidence intervals for traits. The 95% confidence intervals were calculated for all traits and then DNA marker 110 bp showed bad influence marker for both *M. longissimus dorsi* area and carcass weight. We concluded that the major markers of BM4311 locus in Hanwoo chromosome 6 are only both DNA 100 and 95 bp. In future research, we have to apply 100 and 95 bp of BM4311 locus not only progeny testing Hanwoo but performance Hanwoo.

ACKNOWLEDGEMENTS

We would like to thank to anonymous reviewers for helpful discussions and comments on the manuscript. This work was supported by Korea Research Foundation Grant (KRF-2001-005-G20008).

REFERENCES

Bishop, M. D., S. M. Kappes, J. W. Keele, R. T. Stone, S. L. F. Sunden, G. A. Hawkins, S. S. Toddo, R. Fries, M. D. Grosz, J. Yoo and C. W. Beattier. 1994. A genetic linkage map for cattle.

- Genetics 136:619-639.
- Casas, E., S. D. Shackelford, J. W. Keele, R. T. Stone, S. M. Kappes and M. Koolmaria. 1999a. Evidence of quantitative trait loci affecting growth and carcass composition traits in cattle segregating the muscle hypertrophy locus. *J. Anim. Sci.* 77 Supplement(1):126-127.
- Casas, E., J. W. Keele, S. M. Kappes and R. T. Stone. 1999b. A quantitative trait locus for birth and yearling weight segregates on chromosome 6 in *Bos Taurus* cattle. Plant & Animal Genome VII Conference, San Diego, CA: p. 150.
- Chotai, J. 1984. On the lod score method in linkage analysis. *Ann. Hum. Genet.* 48:359-378.
- Churchill, G. A. and R. W. Doerge. 1994. Empirical Threshold Values for Quantitative Trait Mapping. *Genetics* 138:965-971.
- Efron, B. and R. Tibshirani. 1993. "An Introduction to bootstrap" CHAMPMAN & HALL
- Georges, M., D. Nielsen, M. Mackinnon, A. Mishra, R. Okimoto, A. T. Pasquino, L. S. Sargeant, A. Sorenson, M. R. Steele, X. Zhao, J. E. Womack and I. Hoeschele. 1995. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* 139:907-920.
- Good, P. 1994. *Permutation Test: A peactical Guide to resampling for testing Hypothesis*, Spring-Verlag, New York.
- Hirano, T., N. Kobayashi, T. Nakamaru, K. Hara and Y. Sugimoto. 1998. Linkage analysis of meat quality in Wagyu. 26th International Conference on Animal Genetics, Auckland, New Zealand:E019.
- Kim, J. W., S. I. Park and J. S. Yeo. 2003. Linkage Mapping and QTL on Chromosome 6 in Hanwoo (Korean Cattle). *Asian-Aust. J. Anim. Sci.* 16(10):1402-1405.
- Kim, J. W., T. K. Jang, Y. A. Park and J. S. Yeo. 2000. Linkage mapping of chromosome 6 in the Korean Cattle (Hanwoo). *Asian-Aust. Anim. Sci.* 13 (Suppl.):235.
- Kim, M. J., J. Y. Lee, J. S. Yeo, Y. W. Lee and Y. J. Joe. 2003. A major DNA mining of BM4311 in Hanwoo. Proceedings of the Spring Conference, Cheju National Univ. 305-311.
- Knott, S. A. and C. S. Haley. 1992. Aspects of maximum likelihood method methods for the mapping of quantitative trait loci in line crosses. *Genet. Res.* 60:139-151.
- Lander, E. and D. Bostein. 1989. Mapping mendelian factors underlying quantative traits using RFLP linkage maps. *Genetics* 121:185-199.
- McPherron, A. C. and S. J. Lee. 1997. Double muscling in cattle due to mutations in the myostatin gene. *Proc. Natl. Acad. Sci. USA* 23:12457-12461.
- Ott, Jurg. 1999. *Analysis of human genetic linkage*, 3rd ed. The Johns Hopkins university press.
- Mannen, H. S., F. Tsuji, F. Mukai, N. Goto and S. Ohtagaki. 1993. Genetic Similarity using DNA fingerprinting in cattle to determine relationship coefficient. *J. Hered.* 84:166-169.
- Peter, M. Visscher, Robin Thompson and Chris S. Haley. 1996. Confidence Intervals in QTL Mapping by Bootstrapping. *Genetics* 143:1013-1020.
- Phillip, G. 1994. "Permutation Test" Springer-Verlag.
- Rassman, K., C. Scholotterer and D. Tautz. 1991. Isolation of simple-sequence loci for use in polymerase chain reaction base DNA fingerprinting. *Electrophoresis* 12:113-118.
- Roslin Institute 2000. Bovine gene map. URL=<http://ri.bbsrc.ac.uk/>
- Stone, R. T., E. Casas, S. M. Cappel and J. W. Keele. 1999. Status of mapping bovine QTL at the U.S. meat animal research center. Plant-Animal Genome VII Conference, San Diego, CA:W34.
- Taylor, J. F., S. K. Davis, J. O. Sanders, J. W. Turner, R. K. Miller and S. B. Smith. 1998. Identification of QTLs for growth and carcass quality in a cross between *Bos indicus* and *Bos taurus*. Plant-Animal Genome VII Conference, San Diego, CA:W19.
- 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.