

DNA Marker Mining of BMS1167 Microsatellite Locus in Hanwoo Chromosome 17

Jea-Young Lee¹⁾ · Yong-Won Lee²⁾ · Jae-Chul Kwon³⁾

Abstract

We describe tests for detecting and locating quantitative traits loci (QTL) for traits in Hanwoo. Lod scores and a permutation test have been described. From results of a permutation test to detect QTL, we select major DNA markers of BMS1167 microsatellite locus in Hanwoo chromosome 17 for further analysis. K-means clustering analysis applied to four traits and eight DNA markers in BMS1167 resulted in three cluster groups. We conclude that the major DNA markers of BMS1167 microsatellite locus in Hanwoo chromosome 17 are markers 100bp, 108bp and 110bp.

Keywords : Bootstrap method, K-means clustering analysis, Permutation test, QTL

1. 서론

21세기는 생명공학의 시대를 맞이하여, 생명공학의 핵심이 될 지놈(genome) 연구가 활발히 진행되고 있다. 지놈 연구란 한 생명체가 생명을 지닌 개체로서 기능을 하도록 하는 정보가 담겨 있는 유전정보의 총 집합체인 지놈을 구성하고 있는 DNA의 서열을 밝히는 프로젝트를 말하는 것이다. 이 지놈 프로젝트로써 21세기 생명체로서 활동하는 개체에 대한 근본정보를 모두 밝히게 될 예정이다. 선진국에서는 이미 인간지놈 프로젝트(Human Genome Project) 뿐만 아니라 경제성이 높은 동물의 지놈 프로젝트도 추진되고 있어, 우리나라에서도 선진국에 종속되지 않는 고유한 재래유전 자원을 이용한 지놈 프로젝트의 필요성이 끊임없이 제기되어져 왔으며, 우리나라 재래가축 유전 자원중 가장 경쟁력이 확보되고 경제성이 높은 것으로 확인된 한우를 대상으로 유전자지도 작성(gene mapping)이 시도되고 있다(Kim 등, 2000; Yeo 등,

1) 제 1 저자 : (712-749) 경북 경산시 대동 214-1 영남대학교 이과대학 수확통계학부 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 이과대학 통계학과 대학원

2004). 한우를 비롯한 축우의 유전자지도 작성에서는 주로 microsatellite를 이용하고 있다. microsatellite란 생물체에 존재하는 단순염기서열(simple repetitive sequence)들의 수많은 반복단위 수의 변이에 따른 DNA다형성(polymorphism)을 나타내는 것으로 1~6개의 nucleotide가 순차적 반복염기(tandem repeat sequence)로 구성되어 있는 것으로 보고되었으며(Rassman 등, 1991), 이들 microsatellite들은 주변의 특이 염기서열로부터 프라이머(primer)를 이용한 PCR 방법은 사용이 용이하고 증폭된 산물들은 단순염기서열 반복단위의 수를 변이 시킴으로서 높은 다형성을 나타내게 된다. 그리고 이들이 나타내는 marker들은 개체에서 특이적인 다형성과 공우성(codominance)을 나타내게 되어 동일 종의 혈연관계가 가까운 개체나 계통간에도 높은 수준의 이형 접합성을 보인다. 그러므로 개체 내 연관분석에 매우 효과적으로 이용할 수 있어 동물의 유전자지도 작성에 유용한 도구로 사용할 수 있는 것으로 밝혀져(Bishop 등, 1994; Georges 등, 1995), 개체식별 및 친자감별 등에 microsatellite가 유용한 도구로 사용되어졌다(Mannen 등, 1993).

특히, 소에서는 Bishop 등(1994)이 468개의 microsatellite primer를 사용하여 170여 개의 유전자 좌위(loci)로 평균 15cM 간격의 유전자 연관 지도를 작성하였으며, Taylor 등(1998)은 Cri-Map 프로그램을 이용하여 bovine 염색체 1번 (BTA1)에서 26개의 microsatellite 좌위들의 유전자지도를 작성하여 각각의 유전자 좌위간의 유전적 거리가 평균적으로 2cM 이하인 암, 수를 선별, 그들 평균(sex-average)이 163.6cM (Female 159.5cM, Male 165.3cM)으로 나타났다. 또한 소에서 일당증체량, 근내지방도, 등지방 두께 및 산자수 등의 중요한 경제 형질에 관여하는 유전자 연관 표식인자가 몇 번 염색체의 어느 위치에 존재하는가를 밝히기 위한 QTL(Quantitative Trait Loci)에 대한 연구가 진행되고 있다. Hirano 등(1998), Taylor 등(1998)과 Stone 등(1999)은 근내지방(marbling)과 고기연도(meat tenderness)에 관련되는 유전자가 27번 염색체에 있음을 보고하였고, Stone 등(1999)은 bovine 염색체 6번에 생시 및 1세령 체중에 연관된 유전자가 존재함을 그리고 Casas 등(1999 a,b)은 bovine 염색체 6번에 위치한 7개의 microsatellite loci(INRA133, BM1329, BMS2508, BMS382, BM306, BMS483 및 BM415)가 생시 체중과 증체에 연관되어 있음을 보고하였다. 또한 Yeo 등(2004)은 염색체 6번에 위치한 ILSTS035 microsatellite locus에서 DNA Marker 235bp를 우수 DNA Marker로 선별하였다.

본 연구는 한우의 주요 경제형질인 육질과 육량에 연관된 유전자들이 다수 위치하는 17번 염색체의 microsatellite loci를 대상으로 우선 육질에 연관된 loci의 QTL 분석을 하여 선별된 loci값을 계산하고(2장), 그 이후 데이터 마이닝에 의해 우수 DNA를 Mining하려고 시도하고(3장), 이들 loci들이 가지는 DNA marker를 이용한 한우의 형질의 능력 개량을 위하여 본 연구를 시행하였다.

2. QTL (quantitative trait loci) 모델링과 Permutation Test

한우에서 경제형질연관 DNA marker를 이용한 육질형질의 능력 개량을 위해 QTL 분석의 활용은 한우의 육질능력 개량에서 보다 신속하고, 정확한 육질 연관 DNA marker의 활용을 가능하게 할 수 있다. 따라서 한우가 국제경쟁력을 가지기 위해서는 육질에 연관된 한우의 고유한 유전자를 찾고, 이를 위한 linkage map을 기초로 한 한

우 집단에 적합한 새로운 QTL 분석 model 및 program의 개발이 반드시 필요하다. 이러한 측면에서 볼 때 (Kim 등, 2000)이 보고 한 한우의 6번 염색체의 유전자 지도 작성을 이용한 일당증체량 연관 QTL의 분석에서 적용되어진 QTL 분석 model은 반형매 가계를 이용한 linkage mapping과 QTL 분석의 가능성을 충분히 제시하고 있어 본 연구의 수행에 적극 활용하였다. 본 연구는 한우의 17번 염색체의 유전자 지도 작성을 이용한 경제형질 연관 QTL의 LOD값을 계산하여 경제형질의 기록이 있는 한우 농가 개체들의 QTL 검정하였다. LOD 값은 LOD (logarithm of the odds ratio, 1984; Chotai)공식에 적용해서 얻어진다. 실제로 LOD 값은 관측값으로부터 우수 유전형질의 추정에 이용되어왔으며, “LOD 값>3.0”이면 통계적으로 유의한 유전형질로 여긴다(Ott, 1999). LOD의 구체적 설명은 Chotai(1984), Ott(1999)등에 따른다. 그리고, 기존의 개량모형의 예상결과와 비교하여 한우의 선발체계를 구축할 수가 있을 것으로 판단했다. 하지만, “LOD 값>3.0”을 판정기준에 대한 문제점이 발생하여 LOD 값의 통계적 유의성을 검정해주기 위해 Permutation Test를 실시하였다(Churchill and Doerge, 1994). 각 loci별로 10,000번의 simulation을 통해 p-value를 계산한다.

예를 들어 BM3026의 근내지방에 대한 Permutation Test의 단계에 대하여 살펴보면(Phillip, 1994)

단계 1 : 근내지방에 대하여 Anova 분석과 사후검정을 통하여 평균이 높은 그룹과 높지 않은 그룹으로 이분화 한다.

단계 2 : 나누어진 그룹 중 평균이 높은 그룹에 속해있는 DNA Marker를 파악한 후, 분석하고자 하는 그룹의 검정통계량을 선별한다.

단계 3 : 선별된 검정통계량을 계산한다.

단계 4 : 배열을 바꾸어가며(permute) 검정통계량을 재계산한다. 이 과정을 시행횟수(10,000번)만큼 반복한다.

단계 5 : 최초로 선별된 이분화된 그룹의 통계량과 비교하여 p-value를 계산한다.

그 결과는 <표 1>과 같다.

〈표 1〉 BMS941의 우수 Loci에 대한 Permutation Test 결과

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(%)
BMS 8125	3.38 (0.000)	6.22	5.28 (0.000)	9.43	2.01 (0.004)	7.54	9.24 (0.000)	15.33
BMS 1123	4.67 (0.000)	8.62	1.43 (0.072)	2.83	2.27 (0.028)	6.57	5.03 (0.000)	9.22
BMS 499	3.98 (0.000)	7.39	3.19 (0.000)	6.02	4.6 (0.000)	10.39	9.27 (0.000)	15.65
BMS 941	3.75 (0.000)	7.29	5.7 (0.000)	10.72	4.07 (0.000)	4.96	5.11 (0.000)	9.67
BMS 1101	3.15 (0.000)	5.89	2.04 (0.119)	3.92	4.28 (0.000)	6.48	6.02 (0.000)	10.68
BMS 1879	3.49 (0.000)	6.76	4.98 (0.000)	9.4	2.52 (0.015)	3.02	5.55 (0.000)	10.33
BMS 1167	4.71 (0.000)	8.8	6.14 (0.000)	11.22	5.7 (0.000)	9.96	4.98 (0.000)	9.26
HUJ 223	3.58 (0.000)	6.46	2.3 (0.002)	4.26	5.04 (0.000)	6.25	2.98 (0.003)	5.44
CSSM 033	3.79 (0.000)	7.08	3.39 (0.000)	6.42	1.59 (0.012)	3.41	2.27 (0.011)	4.38

〈표 1〉에 따르면 BMS499, BMS941, BMS1167의 loci가 모든 경제형질에서 우수한 값으로 나타났다. 우선 본 논문에서는 BMS1167에 대해서 우수 DNA를 선별하고자 한다.

3. K-평균 군집화 기법을 이용한 우수 DNA Mining

본 실험은 농림기획과제로 수행중인 한우고유의 우수 유전인자의 관찰을 위해 실험 중인 후대검정한우 269두를 사용하였다. 이중 한우 17번 염색체중 선별된 BMS1167 microsatellite locus의 DNA marker를 분석하였다. 변수의 개수는 12개이다. 12개의 변수 중 각 DNA marker의 유무에 대하여 나누어진 변수가 8개 DNA Marker 100, …, DNA Marker 117로 정의되고 이들은 각각 염기서열로부터 primer를 이용한 PCR 방법에서 얻어진 정량적(metric)자료이며, 한우 특성에 관한 변수가 4개이다. 한우 특성에 관한 변수에는 근내지방, 일당증체, 등지방, 및 등심이 있다. 이런 한우의 특성에 관한 변수들에 대하여 군집을 나누어서 어떤 DNA marker가 어느 군집에 많이 속해 있는지를 발견하고자 하였다. 변수들 간의 단위가 달라서 군집 형성시 문제가 발생할 가능성이 있기 때문에 처음 변수들을 표준화 시켰다. 그리고, 등지방은 값

이 적을수록 좋은 것이 되므로, 군집 분석 전에 역 등지방이란 새로운 변수를 만든다. 역 등지방을 계산하는 방식은 (1/등지방)×10으로 하였다.

한우의 특성에 관한 4개의 변수들의 특성을 DNA marker와의 모형구축을 위해 비계층적 군집방법인 K-평균 군집분석을 시도하였다. K-평균 군집화의 알고리즘은 총 5개의 단계로 이루어진다.

- 단계 0 : 사전적 군집수 3개 지정,
- 단계 1 : 각 군집에 임의로 군집중심 정함,
- 단계 2 : 모든 개체 각각 가장 가까운 군집중심에 배속,
- 단계 3 : 각 군집 중심 산출,
- 단계 4 : 단계 2,3 변화 없을 때까지 반복한다.

각 개체 $x_i = (x_{1i}, x_{2i}, \dots, x_{pi})'$ ($i = 1, \dots, n$)와 군집중심 m_t ($t = 1, \dots, k$)간의 거리는 유클리드 방식과 표준화된 유클리드 방식으로 계산되어진다. 즉, 자료행렬 X 와 표본 공분산행렬 S 는 다음과 같이 표시된다.

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

$$S = \begin{bmatrix} s_1^2 & s_{12} & \dots & s_{1p} \\ s_{21} & s_2^2 & \dots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \dots & s_p^2 \end{bmatrix}$$

이때, 유클리드 방식은 변수의 척도

$$d(x_i, m_t) = \left[\sum_{h=1}^p (x_{hi} - m_{hi})^2 \right]^{1/2} \quad (h = 1, \dots, p)$$

에 의존하고, 표준화된 유클리드 방식은 변수의 척도에 무관하게 군집을 이루게 된다.

$$d(x_i, m_t) = \left[\sum_{h=1}^p \left(\frac{x_{hi} - m_{hi}}{s_h} \right)^2 \right]^{1/2} \quad (h = 1, \dots, p)$$

이런 단계를 거쳐 여기서는 Clementine의 K-Means node를 이용한 군집화를 시도한 결과 3개의 군집이 잠정 결정되어, 그 상세 분석을 하고자 한다. Clementine의 군집화 결과를 다음 <표 2>로 나타내어 군집의 특성을 파악한 후 각 군집별 특성을 정의하였다. 즉, 군집1은 (역)등지방에 양호한 영향을 주는 DNA marker군이고, 군집2는 근내지방에 양호한 영향을 주는 DNA marker군이며, 군집3은 일당중체, 등심, 도체중에 양호한 영향을 주는 DNA marker군이다.

<표 2> K-Means 군집화 결과 () : 두수

	군집 1(98)	군집 2 (87)	군집 3 (84)
근내지방	-0.677	1.219	-0.473
일당증체	-0.822	0.159	0.797
(역)등지방	0.439	-0.173	-0.333
등심	-0.665	0.241	0.526



<그림 1> 각 DNA Marker에 대한 군집들의 비율

<그림 1>은 각 DNA marker에 대한 군집별들의 비율을 나타낸다. DNA marker 103bp, 110bp, 117bp는 군집1(역 등지방에 양호한 영향을 주는 군집)에 많다. DNA marker 100bp, 108bp, 115bp는 군집 2(근내지방에 양호한 영향을 주는 군집)에 많다. DNA marker 100bp, 105bp, 113bp, 115bp는 군집 3(일당증체, 등심에 양호한 영향을 주는 군집)에 많다.

한편, 표준화된 데이터에 대한 각 DNA marker의 경제형질별 표준화 평균을 그 특징을 분석하고자 하였다. 표준화된 데이터를 사용하였기 때문에 0을 기준으로 하여 평균차이를 비교하였다. <표 3>에서 근내지방의 평균이 높게 나타난 DNA marker는 100bp, 108bp였고, (역) 등지방의 평균이 높게 나타난 DNA marker는 110bp, 113bp, 115bp였으며, 일당증체, 등심의 평균이 높게 나타난 DNA marker는 100bp였다.

<표 3> BMS941 내 DNA marker들 경제형질별 평균 () : 두수

경제형질	DNA marker							
	100bp (96)	103bp (79)	105bp (65)	108bp (34)	110bp (63)	113bp (29)	115bp (103)	117bp (19)
근내지방	0.1406	-0.0388	-0.0657	0.2671	-0.1705	0.0725	-0.0497	-0.0533
일당증체	0.2371	0.0062	0.0507	-0.1338	0.0276	-0.2818	-0.1382	-0.3373
등지방	-0.1592	0.0333	-0.0204	-0.2194	0.1838	0.2412	0.1987	0.079
등심	0.2148	-0.1099	-0.0587	-0.0537	-0.0967	0.0765	-0.1599	-0.216

<표 4> 표준화된 데이터 평균값에 따른 DNA marker와
 Mining된 DNA marker

	데이터의 평균결과 < 표 3 >	K-means Mining < 그림 2 >
(역) 등지방 (군집 1)	110, 113 and 115bp	103, 110 and 117bp
근내지방 (군집 2)	100 and 108bp	100, 108 and 115bp
일당증체, 등심 (군집 3)	100bp	100, 105, 113 and 115bp

<표 2>, <표 3>, <그림 1>로부터 한우의 주요형질인 육질과 육량에 관련한 결과들을 <표 4>로 요약할 수 있다. (역)등지방에 우수 유전인자로 DNA marker 110bp로 나타났으며, 근내지방에 우수 유전인자로 DNA marker 100bp, 108bp로 나타났으며, 일당증체, 등심에 우수한 인자로 DNA marker 100bp로 나타났다.

4. 결론

한우 17번 염색체 유전자 지도에서 QTL분석을 통하여 먼저 잠정적 우수 loci를 선별하였다. 잠정적 우수 loci를 대상으로 Lod score 값의 통계적 유의성을 검정해주기 위해 Permutation Test를 실시하였다. 그 결과 BMS499, BMS941, BMS1167 microsatellite loci가 모든 경제형질에서 우수 loci로 선별되었다. 본 논문에서는 우선적으로 BMS1167 microsatellite locus에 대해서 우수 DNA를 선별하고자 하였다. 우수 DNA를 선별하기 위하여 BMS1167 microsatellite locus의 각 경제형질에 대한 표준화 평균값과 Data Mining 기법 중 K-평균 군집화 방법을 적용하여 한우의 특성부위에 따른 유전적 요인의 DNA marker의 영향을 파악하였다. 그 결과, 경제형질 중 등지방 우수 DNA Marker로 DNA Marker 110bp를 선택할 수 있었으며, 등지방을 제외한 나머지 경제형질인 근내지방, 일당증체 그리고 등심에서의 우수 DNA Marker로 DNA Marker 100bp, 108bp를 선택할 수 있었다. 우리는 본 분석의 보다 정확한 값을 도출하기 위해, 향후 BMS499 와 BMS941 microsatellite loci의 연구결과를 종합하여 DNA marker를 더 정확하게 발굴할 때 양질의 한우 DNA marker연구에 기초적인 토대를 만들 수 있을 것이다.

참고문헌

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

- 136:619-639.
2. Casas, E., Shackelford, S. D., Keele, J. W., Stone, R. T., Kappes, S. M. and Koohmaraie, M., (1999a). Evidence of quantitative trait loci affecting growth and carcass composition traits in cattle segregating the muscle hypertrophy locus. *Journal of Animal Science* 77 Supplement(1) :126-127.
 3. Casas, E., Keele, J. W., Kappes, S. M. and Stone, R. T., (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:P150.
 4. Chotai, J.,(1984). On the lod score method in linkage analysis. *Ann. Hum. Genet* 48:359-378.
 5. Churchill, G. A. and Doerge, R.W., (1994). Empirical Threshold Values for Quantitative Trait Mapping, *Genetics* 138:963-971.
 6. Georges, M., Nielsen, D., Mackinnon, M., Mishra, A., Okimoto, R., Pasquino, A. T., Sargeant, L. S., Sorenson, A., Steele, M. R., Zhao, X., Womack, J. E. and Hoeschele, I., (1995). Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* 139:907-920.
 7. 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.
 8. Kim, J. W., Jang, T. K. Park, Y. A. and Yeo, J. S. (2000). Linkage mapping of chromosome 6 in the Korean Cattle(Hanwoo). *Asian-Aust. Anim. Sci.*13(Suppl.):235
 9. Mannen, H., Tsuji, S., Mukai, F., Goto, N. and Ohtagaki, S., (1993). Genetic Similarity using DNA fingerprinting in cattle to determine relationship coefficient. *J. Hered.* 84:166-169.
 10. Ott, Jurg. (1999). *Analysis of human genetic linkage, 3rd ed.* The Johns Hopkins university press.
 11. Rassman, K., Schlotterer, C. and Tautz, D., (1991). Isolation of simple-sequence loci for use in polymerase chain reaction base DNA fingerprinting. *Electrophoresis* 12:113-118.
 12. Stone, R. T., Casas, E., Cappel, S. M. and Keele, J. W., (1999). Status of mapping bovine QTL at the U.S. meat animal research center. *Plant-Animal Genome VII Conference*, San Diego, CA:W34.
 13. Taylor, J. F., Davis, S. K., Sanders, J. O., Turner, J. W., Miller R. K.,

- and Smith, S. B., (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.
14. Phillip, G., (1994). *Permutation Test*, Springer-Verlag
 15. Yeo, J. S., Lee, J. Y. and Kim, J. W., (2004). DNA Marker Mining of ILSTS035 Microsatellite Locus on Chromosome 6 of Hanwoo Cattle. *J. Genetics*, 83, 245-250bn

[2006년 2월 접수, 2006년 4월 채택]