

## 한우의 Chromosome 6에서 BM4311의 중요 DNA Mining

김문정<sup>1)</sup>, 이재영<sup>2)</sup>, 여정수<sup>3)</sup>,  
이용원<sup>4)</sup>, 조용주<sup>5)</sup>

### 요 약

K-Means 모델을 이용하여 한우 유전자 6번의 BM4311의 중요 DNA marker를 찾기위해 여러 가지로 시도해 왔다. 이번 논문에선 QTL(Quantitative Trait Loci)과 data mining modeling를 이용하여 BM4311에서 중요 DNA marker를 찾아 보도록 하겠다.

주요용어 : K-Means, QTL, Gene mapping, Microsatellite, Linkage mapping

### 1. 서론

21세기는 생명공학의 시대를 맞이하여, 그 핵심이 될 지놈(genome) 연구가 활발히 진행되고 있으며 얼마전 인간의 지놈 지도도 완성되었다. 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. 선진국에서는 경제성이 높은 동물의 Genome Project가 추진되고 있어, 우리나라에서도 재래가축 유전자원 중 가장 경제성이 높아 경쟁력이 있는 것으로 확인된 한우를 대상으로 유전자지도 작성(gene mapping)이 시도되고 있다(Kim 등, 2000, 2001). 한우를 비롯한 축우의 유전자지도 작성에서는 주로 microsatellite를 이용하고 있다.

특히, 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 진행되어지고 있으며, 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)은 bovine 염색체 6번에 위치한 7개의 microsatellite loci(INRA133, BM1329, BMS2508, BMS382, BM306, BMS483 및 BM415)가 생시 체중과 증체에 연관되어 있음을 보고하였다.

따라서 본 연구는 한우의 주요 경제형질인 육질과 육량에 연관된 유전자들이 다수 위치하

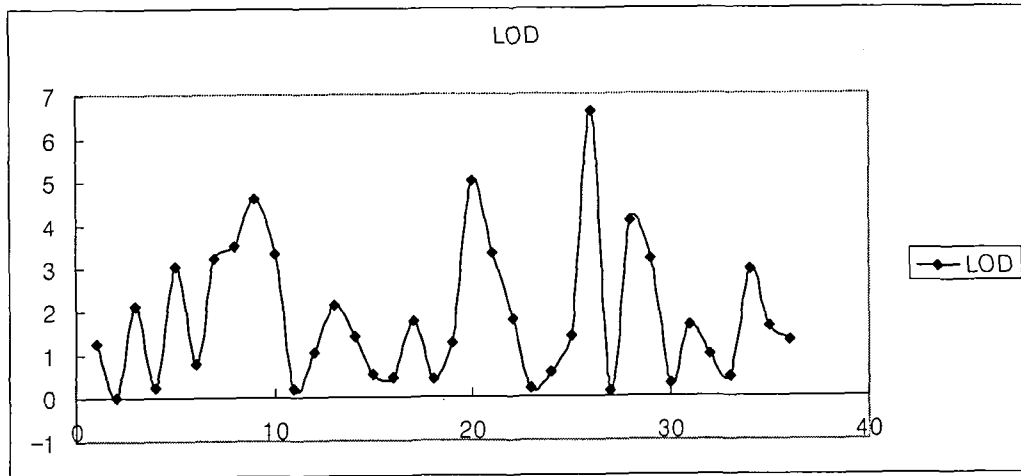
1. Graduate, Department of Statistics, Yeungnam University, 214-1 Daedong, Kyungsan, Kyungbuk, 712-746, South Korea
2. Associate Professor, Department of Statistics, Yeungnam University, 214-1 Daedong, Kyungsan, Kyungbuk, 712-746, South Korea  
E-mail : jlee@yu.ac.kr
3. Professor, Department of Animal resources and Biotechnology, Yeungnam University 214-1 Daedong, Kyungsan, Kyungbuk, 712-746, South Korea
4. Graduate, Department of Statistics, Yeungnam University, 214-1 Daedong, Kyungsan, Kyungbuk, 712-746, South Korea
5. Graduate, Department of Statistics, Yeungnam University, 214-1 Daedong, Kyungsan, Kyungbuk, 712-746, South Korea

는 6번 염색체의 microsatellite loci를 대상으로 우선 육질에 연관된 loci를 QTL 분석에 의해 한우 고유의 잠정우수 경제형질을 선정하고(2장), 그 이후 데이터 마이닝에 의해 우수 DNA를 Mining하려고 시도함으로써(3장), 이들 육질연관 loci들이 가지는 DNA marker를 이용한 한우의 육질형질의 능력 개량을 위하여 본 연구를 시행하였다.

## 2. QTL(quantitative trait loci) 모델링

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 가지도록 하기 위해서는 육질에 연관된 한우의 고유한 유전자를 찾고, 이를 위한 linkage map을 기초로 한 한우 집단에 적합한 새로운 QTL 분석 model 및 program의 개발이 반드시 필요하다. 이러한 측면에서 볼 때 Kim 등(2000)이 보고한 한우의 6번 염색체의 유전자 지도 작성을 이용한 일당중체량 연관 QTL의 분석에서 적용되어진 QTL 분석 model은 반형매 가계를 이용한 linkage mapping과 QTL 분석의 가능성을 충분히 제시하고있어 본 연구의 수행에 적극 활용하였다.

본 연구는 한우의 6번 염색체의 유전자 지도 작성을 이용한 근내지방량 연관 QTL의 LOD값을 계산하여 경제형질의 기록이 있는 한우농가 개체들의 QTL에 대해 검정하였다. 그리고, 기존의 개량모형의 예상결과와 비교하여 한우의 선발체계를 구축할 수가 있을 것으로 판단했다. 그림1은 후대검정우 6번 염색체의 근내지방량을 활용한 microsatellite loci 추정값을 대상으로 얻어진 그래프포, BM4311 microsatellite loci(loci #26)를 가장 의미 있는 잠정 우수형질로 선정하였다. 다음장에서 BM4311 microsatellite loci를 중심으로 DNA 분석을 데이터마이닝 기법으로 시행하였다.



< 그림 1 > 6번 염색체의 근내지방량 연관 QTL의 LOD값 그래프

## 3. 데이터마이닝 기법을 이용한 DNA Mining

연구에 활용된 자료는 농림기획과제로 수행중인 한우의 경제형질 우수 유전자 선별을 위해 실험되고 있는 후대검정우 165두를 사용하였다. 이중 한우 6번 염색체 BM4311

microsatellite loci의 DNA marker를 사용하였다. 변수의 개수는 11개이다. 11개의 변수 중 각 DNA marker의 유무에 대하여 나누어진 변수가 6개 DM95,……,DM110으로 정의되고, 한우 특성에 관한 변수가 5개이다. 한우 특성에 관한 변수에는 근내지방, 일당증체, 등지방, 등심 및 도체중이 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 값이 적을수록 좋은 것이 되므로, 군집 분석 전에 역 등지방이란 새로운 변수를 만든다. 역 등지방을 계산하는 방식은 (1/등지방)×10으로 하였다.

한우의 특성에 관한 5개의 변수들의 특성을 DNA marker과의 모형구축을 위해 비관리학습 신경망분석 방법인 K-Means 군집분석을 시도하였다. 즉 K-Means 군집화의 알고리즘은 총 5개의 단계로 이루어진다.

- 단계 0 : 사전적 군집수 K개 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.

< 표 1 > BM4311 내 DNA marker들의 표준화 평균 ( ) : 두수

	DNA marker					
	# 95 (24)	# 100 (80)	# 103 (107)	# 105 (32)	# 107 (43)	# 110 (4)
근내지방	0.0946	0.2996	0.0633	-0.1014	-0.1239	-0.6893
일당증체	0.3169	0.0359	-0.0285	-0.0359	0.0241	0.1140
역등지방	0.2263	-0.0182	0.0727	0.0051	-0.0458	0.2866
등심	0.3021	0.0515	-0.0440	-0.1678	0.0985	-0.0674
도체중	0.3497	0.0783	-0.0803	-0.0831	0.0368	0.1860

<표 1>에서 근내지방의 평균이 DNA marker 100에서 가장 높으며 일당증체, 역등지방, 등심, 도체중의 평균은 DNA marker 95에서 높게 나타나고 있다. 위의 표준화 평균과 신경망 분석을 통해 얻어진 결과와 그 특징을 비교분석 하기 위해서 클레멘타인 패키지의 K-Means node를 이용한 군집화를 시행하였다.

< 표 2 > K-Means 군집화 결과

	군집 1 (69)	군집 2 (47)	군집 3 (49)
근내지방	-0.552124	1.265126	-0.436007
일당증체	-0.805859	0.253847	0.891291
(역)등지방	0.34431	0.223596	-0.270384
등심	-0.602908	-0.478173	0.390335
도체중	-0.875109	0.410955	0.83811

한우의 Chromosome 6에서 BM4311의 중요 DNA Mining

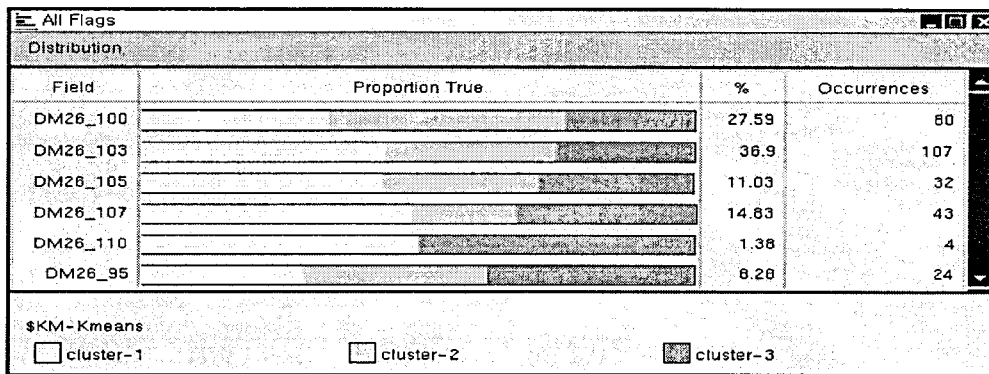
<표 2>에서는 각 군집에 대한 수치적인 특징을 나타내고 있으며, <표 1>와 비교하여 군집의 특성을 파악해 보도록하겠다.

군집 1은 (역)등지방에 양호한 영향을 주는 반면 다른 형질에 있어서는 좋지 못한 영향을 주는 군집이라고 할 수 있으며 군집2에서는 근내지방에 상당히 양호한 영향을 주는 군집이라고 할 수 있다. 마지막으로 군집3은 일당증체, 도체중, 등심에 양호한 영향을 주는 군집이라고 정의 할 수 있겠다.

이렇게 정의된 각 군집에 대해서 DNA marker별 특징을 조사해 보았다.

<그림 2>은 각 DNA marker에 대한 군집들의 비율을 나타낸 것이다.

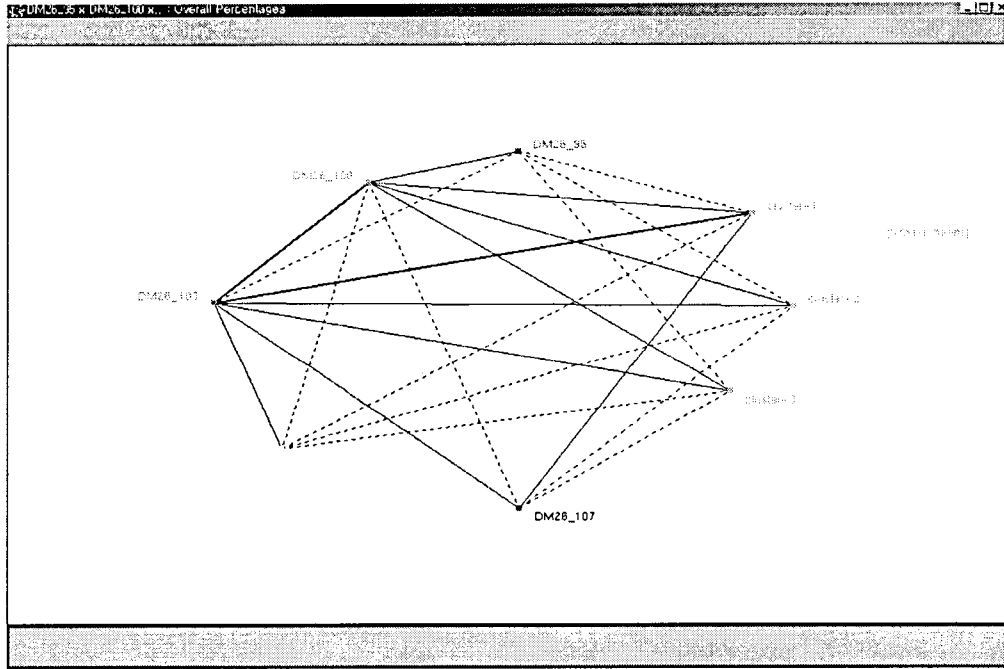
< 그림 2 > 각 DNA marker에 대한 군집들의 비율



DNA marker별로 군집을 살펴보면 DNA marker 100은 각 군집이 골고루 나타나지만 군집2의 성향이 조금 더 있는 것으로 보이며 DNA marker 103, DNA marker 105, DNA marker 107, DNA marker 110은 군집 1의 성향이 강하며 DNA marker 95, DNA marker 110, DNA marker 107은 군집 3의 성향이 있는 것으로 보여진다.

군집결과를 바탕으로 Web node를 그려서 시각적으로 각 군집별로 어느 DNA marker와 관계가 있는지 살펴보도록 하겠다.

< 그림 3 > 각 군집별 DNA marker와의 관계



<그림 3>에서 군집 1은 DNA marker 103이 높은 관계를 가지고 있으며 DNA marker 100, DNA marker 107과도 관계가 있는 것으로 나타나고 있다. 군집 2에서는 DNA marker 100, DNA marker 103이 관계성이 있으며 군집 3 역시 DNA marker 100, DNA marker 103과 관계성이 있는 것으로 나타나고 있다. 따라서 <표 1>, <표 2>, <그림 2>, <그림 3>을 종합하여 보도록하겠다.

<표 3> 표준화된 데이터 평균값에 따른 DNA marker와 Mining된 DNA marker와 Web

	데이터의 평균결과	K-means Mining	Web
<군집1> 역등지방에 좋은 영향을 주는 군집	DNA marker 95, DNA marker 110	DNA marker 103, DNA marker 105, DNA marker 107, DNA marker 110	DNA marker 103, DNA marker 100, DNA marker (107), DNA marker (95)
<군집2> 근내지방에 좋은 영향을 주는 군집	DNA marker 100	DNA marker 100, DNA marker 95	DNA marker 100, DNA marker 103, DNA marker (95)
<군집3> 도체중, 일당중체, 등심에 좋은 영향을 주는 군집	DNA marker 95, DNA marker 110, DNA marker 107	DNA marker 95, DNA marker 110, DNA marker 107	DNA marker 100, DNA marker 103, DNA marker (107)

데이터의 평균결과와 K-means Minig 결과에서 DNA maeker 100이 근내지방에 좋은 영향을 주는 것으로 나타났으며 DNA marker 95과 DNA marker 110이 영향을 주는 것으로 밝혀졌다. 하지만 역등지방에 영향을 주는 DNA marker는 군집2와 군집3과 같이 명백하게 밝혀지는 않는다. 또한 Web의 그림상으로는 보았을 때 군집1에서 평균결과가 높았던 DNA marker 95와 DNA marker 110이 K-menas Mining에서는 나타나고 있으나 Web그림상으로는 약한 관계이거나 관계성이 나타나지 않는 것으로 보여지고 있으며 군집2와 군집 3에서도 Web 그림상으로는 차이가 없는 것으로 보여지고 있다. 이는 Web의 그림은 데이터의 개수에 많은 영향을 받는 것이기 때문에 DNA marker 103의 경우는 데이터의 수가 107개로 가장 많은 DNA marker이므로 web상으로는 가장 강한 관계가 있는 것으로 나타나고 있으며 또한 데이터의 개수가 4개로 가장 작은 DNA marker 110의 경우는 관계이 없는 것으로 나타나고 있다. 따라서 Web의 그림으로 평균값과 군집 분석의 결과와 비교를 하기 위해서는 우선 데이터의 개수를 맞추는 것이 필요하다고 생각되어진다.

#### 4. 결론

한우 6번 염색체 유전자 지도에서 BM4311 microsatellite loci #26의 우수 DNA marker Mining을 해 보았다. BM4311 microsatellite loci #26에서 DNA marker 100은 근내지방에 영향을 주며 DNA marker 95 DNA marker 110이 영향을 주는 것을 알아 낼 수있게 되었다. 하지만 Web에서의 시각적으로 보다 뚜렷한 결과를 얻기 위해서는 데이터의 수를 동일하게 한 후 분석을 해보아야 할 필요성이 있다고 생각되어 진다. 따라서 우리는 데이터의 수를 일정하게 만들기 위한 resampling 작업이 필요하며 Bootstrap method 이나 Permutation Test와 같은 통계적 기법을 이용할 필요가 있다고 보여진다.

#### 참고자료 및 문헌

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., Yoo, J., 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. (1999). 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. Glowatzki-Mullis, Gaillard, C., Wigger, G., and Fries, R. (1995). Microsatellite based on parentage control in cattle. *Animal Genetics*, 26, 7-12.
4. 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
5. Kim, J. W., S. I. Park, Y. A. Park, T. S. Kang, N. H. Park and J. S. Yeo. 2001. Linkage mapping of hanwoo chromosome 6 using half-sib family. *The 10th Annual Congress on Korean Society of Anim. Sci. and Tech*, PA21037
6. Lander, E.S. and Botstein, D. (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, 121, 185-199.
7. Ma, R. Z., Beever, J. E., Da, Y., Green, C. A., Russ, I., Park, C., Heyen, D. W.,

- Everts, R. E., Fisher, S. R., Overton, K. M., Teale, A. J. et al. (1996). A male linkage map of the cattle(*Bos taurus*)genome. *J. Hered.*, 87, 261-271.
8. Rassman, K., Scholotterer, C., and Tautz, D. (1991). Isolation of simple-sequence loci for use in polymerase chain reaction base DNA fingerprinting. *Electrophoresis*, 12, 113-118.
9. Roslin Institute, 2000. Bovine gene map. URL=<http://ri.bbsrc.ac.uk/>
10. Stone, R. T., Casas, E., Cappes, 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.
11. 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.