

The Effects of Breeding Environment Adjustment in FABP4 Gene Identification of Korean Cattle

Hyun-Ji Kim^a · Jea-Young Lee^{a,1}

^aDepartment of Statistics, Yeungnam University

(Received April 16, 2015; Revised May 7, 2015; Accepted August 21, 2015)

Abstract

Economic-traits of livestock are affected by environmental and genetic factors. We are interested in genetic factors that influence the economic-traits of Korean cattle. It is necessary to adjust environmental factors in order to enhance the accuracy of the genetic effect analysis. In this paper, we propose a statistical model of Korean cattle that exclude environmental breeding farm and age factors. We formulated an adjusted economic-trait value, and applied multifactor dimensionality reduction (MDR) method to data of before-and-after adjustment to identify major FABP4 genes. We were able to increase the accuracy of the analysis after adjustment and identify superior FABP4 genes that influence grade and fatty acid.

Keywords: adjusting, environmental factors, genetic factors, MDR, SNP

1. 서론

소고기의 지방성분은 인체에 필요한 열량과 영양성분을 제공할 뿐만 아니라 고기의 맛에 크게 관여하는 것으로 알려져 있다 (Blumer, 1963). 단일불포화지방산(monounsaturated fatty acid; MUFA)은 소고기의 맛과 부드러움에 영향을 미치며, 포화지방산(saturated fatty acid; SFA) 함량에 대한 MUFA의 비율은 소고기의 맛을 결정하는 간접적인 지표가 될 수 있다고 보고되었다 (Sturdivant 등, 1991). 또한 올레인산(oleic acid; C18:1)은 MUFA의 중심 역할로써 요리된 소고기의 풍미 및 다즙성에 있어서 중요한 위치에 있다고 보고된 바 있다 (Melton 등, 1982; Mandell 등, 1998; Oh 등, 2011). 그리고 축산물품질평가원의 한우 등급판정의 도체형질 항목인 등지방두께(backfat thickness; BFT), 등심단면적, 도체중량(carass weight; CWT)과 근내지방도(marbling score; MS)는 한우고기 품질의 주요한 지표이다.

최근 이러한 한우의 맛과 품질에 관련된 경제형질을 향상시키기 위해서 한우의 경제형질에 영향을 미치는 우수 유전자를 찾는 연구가 활발히 이루어지고 있으며, 특히 단일 유전자의 효과에서 더 나아가 여러 유전자의 상호작용에 대한 연구가 활발하다. 이러한 소고기 유전자의 상호작용을 계산하기 위해서 선형 모형과 같은 통계적 방법을 사용해 왔으나, 유전자의 수가 늘어남에 따라 그 모형의 복잡도가 증가되어 사실상 해석이 어려워지는 경우가 발생한다. 그래서 제안된 방법이 다중인자차원축소(multifactor dimensionality reduction; MDR; Ritchie 등, 2001) 방법이다. 이러한 MDR 방법은 이분형 자료에만 활용가능하다는 한계가 있어서 연속형 데이터에도 적용 가능하도록 CART 알고리즘을 이용한 Expanded

¹Corresponding author: Department of Statistics, Yeungnam University, Kyongsan 712-749, Korea.

E-mail: jlee@yu.ac.kr

MDR(E-MDR; Lee 등, 2008) 방법과 더미변수를 이용한 Dummy MDR(D-MDR; Lee와 Lee, 2009) 방법이 있고, 연속형 자료를 이분형 자료로 분류하는 방법으로 Support vector machin 알고리즘을 이용한 Support vector muchin MDR(SVM-MDR; Lee와 Lee, 2010) 방법이 개발되었다.

한우의 경제형질과 같은 개체의 표현형에는 유전적인 요인뿐만 아니라 환경적인 요인도 영향을 미치기 때문에, 순수한 유전적인 요인을 찾기 위해서는 환경요인이 보정되어야 분석의 정확도를 높일 수 있다. 이러한 환경요인, 유전요인 그리고 그 상호작용에 대한 연구에는 이제까지 선형모형이 사용되어왔으나 (Casas 등, 2005; Matsuhashi 등, 2011), 우리는 선형모형을 이용하여 환경요인을 사전에 보정하고, 보정된 새로운 경제형질에 MDR 기법을 적용하여 우수 유전자를 선별한다.

본 연구에서는 한우의 품질에 영향을 미치는 주요 경제형질 C18:1, SFA, MUFA, CWT, BTF, MS와 연관 있는 유전적인 요인을 찾고자 하며, Damcotta 등 (2004)과 Chmurzynska (2006)에 의해 한우의 등급 및 지방산과 깊은 연관이 있다고 알려진 FABP4(adipocyte fatty acid binding protein 4) 유전자의 13가지 SNP 중에서, 6가지 SNP(g.2634+1018 A>T, c.280 A>G, c.388 G>A, c.408 G>C, g.3977-325 T>C, c.456 A>G)에 초점을 맞추었다 (Oh 등, 2012). 이때 일반선형모형에 유전요인 뿐만 아니라 환경요인인 연속형 변수 도축일령(age)의 공변량 효과와 사육환경(farm) 효과를 추가하여 통계모형을 구축하고, 공변량 효과 추정값과 사육환경 효과의 추정값을 지시변수를 이용하여 제거하여 새로운 보정된 경제형질 값을 구한다. 그리고 보정 전·후 데이터에 각각 MDR 방법을 적용하여, 원 데이터에 비해서 보정 된 데이터를 활용하여 정확도를 높일 수 있었음을 보이고, 경제형질에 대한 우수 FABP4 유전자를 선별한다.

본 연구는 다음과 같이 구성되었다. 2절에서는 한우의 경제형질에 대한 통계적 모형을 구축하고, 환경요인이 보정된 경제형질에 대한 모형을 공식화 한다. 그리고 우수 유전자를 규명하기 위해 사용할 MDR 기법을 설명한다. 3절에서는 환경적인 요인에 대한 보정 전·후의 한우자료를 비교하고, 각각에 MDR을 적용하여 유전자를 선별하고 그 결과를 비교해 본다. 마지막으로 4절에서는 연구의 결과를 요약한다.

2. 통계적 방법

2.1절에서는 유전연구에서 개체의 표현형에 영향을 미치는 환경적인 요인과 유전적인 요인에 대한 통계 모형을 소개하고, 2.2절에서는 이 모형을 한우 데이터에 적용하여 한우의 통계적 모형을 구축하고, 환경적인 요인이 보정된 경제형질값에 대한 모형을 공식화 한다. 그리고 2.3절에서는 환경요인 보정 전·후 경제형질 값을 이용하여 우수 유전자를 선별하기위해 적용한 MDR 기법을 설명한다.

2.1. 유전연구에서 개체의 표현형에 대한 통계모형

일반적으로 유전적 관련성 연구에서 어느 개체의 표현형에 영향을 미치는 요인은 크게 유전적인 요인과 환경적인 요인으로 나눌 수 있다. 그러므로 표현형에 영향을 미치는 유전적인 요인에 대한 연구를 위해서는 환경적인 요인을 찾아내어 이를 보정하는 작업이 필요하다.

우선, 개체의 표현형을 나타내면 다음과 같다.

$$P = E + G. \quad (2.1)$$

식 (2.1)에서 P 는 개체의 표현형(phenotype)이고 E 는 환경적인 요인의 효과(environmental effect), G 는 유전적인 요인의 효과(genetic effect)이다. 이러한 요인들의 영향력을 평가하는 방법으로 가장 널

리 쓰이는 방법은 다중선형회귀모형(multiple linear regression)이며 (Casas 등, 2005; Matsuhashi 등, 2011), 다음과 같이 나타낼 수 있다.

$$Y = X\beta + \epsilon, \quad (2.2)$$

여기서 Y 는 개체의 표현형 행렬이며, X 는 표현형에 영향을 미치는 유전적·환경적 요인을 포함하는 설명변수 행렬, β 는 각 요인들의 효과를 나타내는 회귀계수벡터이다. ϵ 은 추정된 회귀식으로 설명할 수 없는 $N(0, \sigma^2 I)$ 인 확률변수항이다. 식 (2.2)를 식 (2.1)의 형태로 바꾸면

$$Y = \mu + E\alpha + G\beta + \epsilon \quad (2.3)$$

으로 나타낼 수 있다. 여기서 μ 는 총 평균 벡터, E 는 표현형에 영향을 미치는 환경적 요인을 포함하는 설명변수 행렬, α 는 각 환경적인 요인들의 효과를 나타내는 회귀계수벡터이다. 그리고 G 는 표현형에 영향을 미치는 유전적인 요인을 포함하는 설명변수 행렬, β 는 유전적인 요인들의 효과를 나타내는 회귀계수벡터이다.

2.2. 한우의 경제형질에 대한 모형 및 환경요인 보정

2.2.1. 한우의 경제형질에 대한 통계모형 본 연구에서는 한우의 경제형질에 유의한 영향을 미치는 변수인 사육환경, 도축일령 그리고 SNP를 이용한다. SNP는 유전적인 요인이며, 사육농가인 farm과 도축일령인 age는 환경적인 요인이다. 우리는 경제형질에 영향을 미치는 유전적인 요인을 규명하는 것에 목적이 있으므로, 유전적인 요인 외에 경제형질에 영향을 미치는 환경적인 요인은 보정한 새로운 경제형질값을 구하여 유전자 효과를 분석할 것이다. 그러므로 다음과 같은 모형을 사용한다. 여기서 사육환경(사육농가)은 l 개, SNP는 m 개, 한우의 개체수는 n 개 이다.

$$y_k = \mu + \alpha_0(\text{age}_k - \overline{\text{age}}) + \alpha_1 \text{farm}_{1k} + \dots + \alpha_{l-1} \text{farm}_{l-1,k} + \beta_1 \text{SNP}_{1k} + \dots + \beta_m \text{SNP}_{mk} + \epsilon_k, \\ i = 1, 2, \dots, l; j = 1, 2, \dots, m; k = 1, 2, \dots, n, \quad (2.4)$$

여기서 y_k 는 k 번째 한우의 경제형질, μ 는 경제형질의 전체 평균을 나타내는 모수, age_k 는 k 번째 한우의 도축일령, $\overline{\text{age}}$ 는 도축일령의 전체 평균, α_0 는 도축일령의 효과, farm_{ik} 는 k 번째 한우의 i 번째 사육농가의 지시변수, α_i 는 i 번째 사육농가의 효과, SNP_{jk} 는 k 번째 한우의 j 번째 유전자, β_j 는 j 번째 유전자의 효과, ϵ_k 는 $N(0, \sigma^2)$ 인 확률변수이다.

또한 양적변수인 도축일령(age)의 효과 α_0 는 모든 처리(유전자)에 공통으로 작용하는 공변량 효과이고 질적변수인 사육농가를 나타내는 변수 farm_{ik} 는 다음과 같이 지시변수로 정의한다.

$$\text{farm}_{ik} = \begin{cases} 1, & k\text{번째 한우가 } i\text{번째 농장 출신일 때 } (i = 1, 2, \dots, l-1), \\ 0, & \text{아닐 때.} \end{cases}$$

따라서 각 사육농가의 효과인 α_i , ($i = 1, 2, \dots, l$)는 한 개체에, 해당하는 하나의 효과만 갖게 된다.

2.2.2. 환경요인을 보정한 통계모형 2.2.1절에서 소개한 한우의 경제형질에 대한 모형식 (2.4)를 식 (2.3)의 형태로 나타내면 다음과 같다.

$$Y = \mu + E\alpha + G\beta + \epsilon, \quad (2.5)$$

여기서 한우 사육환경과 도축일령은 식 (2.4)의 환경적인 효과 $E\alpha$ 에 해당하며, SNP는 식 (2.4)의 유전

적인 효과 $G\beta$ 에 해당한다.

$$E\alpha = \begin{pmatrix} A_1 & \text{farm}_{11} & \text{farm}_{21} & \cdots & \text{farm}_{l-1,1} \\ A_2 & \text{farm}_{12} & \text{farm}_{22} & \cdots & \text{farm}_{l-1,2} \\ A_3 & \text{farm}_{13} & \text{farm}_{23} & \cdots & \text{farm}_{l-1,3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_n & \text{farm}_{1,n} & \text{farm}_{2,n} & \cdots & \text{farm}_{l-1,n} \end{pmatrix} (\alpha_0 \ \alpha_1 \ \alpha_2 \ \cdots \ \alpha_{l-1})',$$

$$G\beta = \begin{pmatrix} \text{SNP}_{11} & \text{SNP}_{21} & \cdots & \text{SNP}_{m1} \\ \text{SNP}_{12} & \text{SNP}_{22} & \cdots & \text{SNP}_{m2} \\ \text{SNP}_{13} & \text{SNP}_{23} & \cdots & \text{SNP}_{m3} \\ \vdots & \vdots & \ddots & \vdots \\ \text{SNP}_{1n} & \text{SNP}_{2n} & \cdots & \text{SNP}_{mn} \end{pmatrix} (\beta_1 \ \beta_2 \ \beta_3 \ \cdots \ \beta_m)',$$

$$A_k = \text{age}_k - \overline{\text{age}}; \quad i = 1, 2, \dots, l; \quad j = 1, 2, \dots, m; \quad k = 1, 2, \dots, n.$$

이와같은 모형식을 이용하여, 우리는 환경적인 요인인 도축일령과 사육농가의 효과 $\hat{\alpha}$ 를 추정하여, 경제형질 Y 에서 환경요인효과 추정치 $E\hat{\alpha}$ 를 제거한 순수 유전요인에 대한 경제형질 Z 를 구한다. 이를 나타내면 다음과 같다.

$$Z = Y - E\hat{\alpha},$$

여기서 $Z, Y, \hat{\alpha}$ 는

$$Z = \begin{pmatrix} \text{adj}(y_1) \\ \text{adj}(y_2) \\ \text{adj}(y_3) \\ \vdots \\ \text{adj}(y_n) \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}, \quad \hat{\alpha} = \begin{pmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_{l-1} \end{pmatrix}$$

이다.

본 연구에서는 순수 유전요인에 대한 경제형질인 Z 를 다음 절에서 소개하는 MDR 방법에 적용하여 우수 유전자를 선별한다.

2.3. 다중인자차원축소방법

MDR(multifactor dimensionality reduction; Ritchie 등, 2001)기법은 인간의 질병 유무와 같은 이분형 자료에 대한 요인의 상호작용 효과를 찾기 위해 제시되었으며, 일반화된 선형 모형의 전통적 통계기법과 달리 모수에 대한 추정과 모형에 대한 가정이 필요하지 않은 비모수적 방법이다. 이 방법은 실험-대조군의 비율을 통해 독립변수를 고 위험군과 저 위험군으로 분류한 뒤 목표변수에 대한 오분류율을 비교하는 방법으로써 검정력 평가를 통해서 높은 검정력이 입증되었다.

본 연구에서는 한우 데이터에 MDR 방법을 적용하여 경제형질에 영향을 미치는 FABP4 유전자를 선별한다. 여기서 경제형질($Y = 1$)은 연속형 데이터이므로 K -평균 군집분석을 이용하여 상위($Y' = 1$) 그룹과 하위($Y' = 0$) 그룹으로 이분화하여 사용한다. 이분화된 경제형질을 MDR 기법에 적용하여 FABP4의 각 유전자형을 고 위험군과 저 위험군으로 분류하고 그 오분류율을 계산하여 오분류율이 가장 낮은 단일SNP와 SNP조합을 찾는다. MDR 방법을 적용한 절차는 다음과 같다 (Figure 2.1).

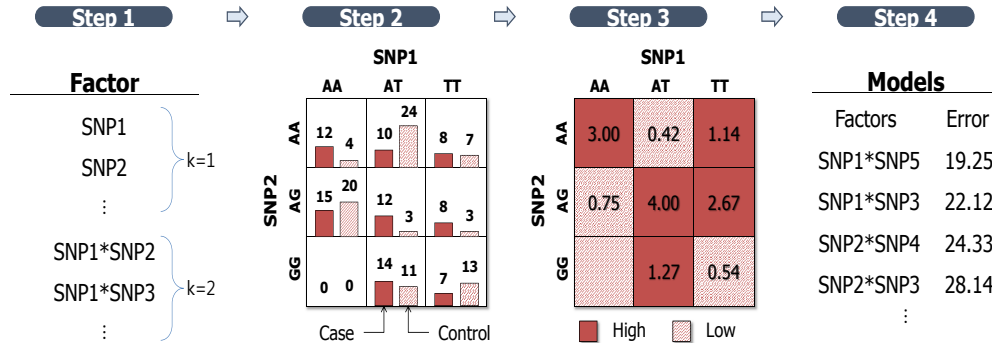


Figure 2.1. MDR algorithm.

- Step 1. 데이터셋에 수준수가 3인 SNP가 p 개 있다고 가정하고, p 개 중 k 개의 SNP를 선택한다.
- Step 2. 선택된 k 개 SNP의 각 수준을 기초로한 3^k 개 셀을 가진 테이블을 구성하고, 각 셀에 실험군($Y' = 1$)과 대조군($Y' = 0$)의 비율을 기술한다. 이때 각 셀은 유전자형을 의미한다.
- Step 3. 전체 데이터셋의 실험-대조군 비율을 계산하여 한계점(threshold; T)으로 설정한다. 그리고 각 유전자형의 실험-대조군 비율(R)을 구하여, $R \geq T$ 인 유전자형을 고 위험군(high risk)으로 정하고, $R < T$ 인 유전자형을 저 위험군(low risk)으로 정한다.
- Step 4. 각 셀을 고위험-저위험군으로 분류한 새로운 이분형 변수에 대한 오분류율을 계산한다.

$$\text{Error} = \frac{N_{(Y'=0|High)} + N_{(Y'=1|Low)}}{N_{total}}$$

이와같은 프로세스를 통해 가능한 모든 k -조합(k 개 SNP들의 조합)을 평가한다. 여기서 전반적으로 가장 좋은 모델을 선택하기 위해서 MDR 기법은 교차검증(tenfold cross-validation)을 사용한다. 그 절차는 다음과 같다.

- Step 1. 전체 데이터셋을 크기가 동일한 10개의 셋으로 나눈다. 그 중 9개를 학습용(train) 자료, 나머지 1개를 테스트(test) 자료로 설정한다.
- Step 2. 학습용 자료, 테스트 자료에서 각 k -조합의 오분류율을 계산한다.
- Step 3. 10개의 데이터셋이 모두 한번씩 테스트 자료로 사용될 때 까지 프로세스를 반복한다.
- Step 4. k -조합 각각에서 학습용 자료의 오분류율의 평균과 검증용 자료의 오분류율의 평균을 구하고, 각 k -조합이 10번의 반복 중에서 가장 좋은 조합으로 선택된 횟수(cross-validation consistency; CVC)를 기록한다.
- Step 5. 테스트 자료의 오분류율이 가장 작고, CVC 값이 가장 큰 k -조합을 최종적으로 선택한다.

3. 적용 및 결과

3.1절에서는 한우 원자료에 대해서 살펴보고, 3.2절에서는 환경적인 요인을 보정한 한우 자료에 대해서 살펴본다. 마지막으로 3.3절에서는 원자료와 환경요인을 보정한 자료에 각각 MDR 기법을 적용하여 우수 유전자를 선별하고, 선별된 우수 유전자가 경제형질을 분류하는 정확도를 비교한다.

Table 3.1. Mean and standard deviation of economic trait (Y)

Economic trait	N	Mean	SD	Min	Max
C18:1	513	44.30	2.66	36.85	53.69
SFA	513	40.60	2.86	30.36	48.44
MUFA	513	53.50	2.97	45.82	63.46
MS	513	5.43	1.94	2.00	9.00
CWT	513	427.25	43.28	321.00	573.00
BFT	513	13.22	5.13	3.00	33.00

Y is an economic traits before correcting for environmental factors.

C18:1 = oleic acid; SFA = saturated fatty acid; MUFA = monounsaturated fatty acid;

MS = marbling score; CWT = carcass weight; BFT = backfat thickness.

Table 3.2. Mean and standard deviation of economic trait (Y) which are divided into two classes

Economic trait	Group (Y')	N	Mean	SD
C18:1	1	230	46.64	1.71
	0	283	42.40	1.55
SFA	1	255	38.33	1.77
	0	258	42.85	1.73
MUFA	1	250	55.94	1.90
	0	263	51.19	1.66
MS	1	275	6.99	0.98
	0	238	3.62	0.94
CWT	1	210	469.68	27.29
	0	303	397.84	23.26
BFT	1	366	10.63	2.82
	0	147	19.67	3.72

Y is an economic traits before correcting for environmental factors.

C18:1 = oleic acid; SFA = saturated fatty acid; MUFA = monounsaturated fatty acid;

MS = marbling score; CWT = carcass weight; BFT = backfat thickness.

3.1. 한우의 경제형질

본 연구는 경북지역의 17개의 사육농장에서 얻은 513두의 한우데이터를 사용하였으며 (Oh 등, 2011), 경제형질(Y)은 한우의 맛과 향에 영향을 준다고 알려진 올레인산(oleic acid; C18:1)과 단일불포화 지방산(monounsaturated fatty acid; MUFA), 포화지방산(saturated fatty acid; SFA), 한우의 품질에 영향을 주는 등지방두께(backfat thickness; BFT), 도체중량(carass weight; CWT)과 근내지방도(marbling score; MS)를 분석에 사용하였다 (Oh, 2014). Table 3.1은 각 경제형질(Y)의 평균과 표준편차를 나타낸 것이다.

우리는 우성인 집단(Y' = 1)과 열성인 집단(Y' = 0)으로 이분화된 경제형질을 분석에 활용한다. Table 3.2는 집단(Y')에 따른 각 경제형질(Y)의 평균과 표준편차를 나타낸 것이다. 여기서 C18:1, MUFA, CWT 그리고 MS는 한우의 품질과 양의 상관관계를 갖고 있으므로 우성 집단의 평균이 더 높고, SFA와 BFT는 음의 상관관계를 가지므로 우성 집단의 평균이 더 낮다.

3.2. 환경적인 요인을 보정한 한우의 경제형질

한우의 경제형질은 유전적인 효과와 환경적인 효과의 영향을 복합적으로 받는다. 우리는 이 중 유전적

Table 3.3. Mean and standard deviation of adjusted economic trait (Z)

Economic trait	N	Mean	SD	Min	Max
C18:1	513	43.65	1.34	41.28	48.31
SFA	513	41.91	1.11	38.07	44.14
MUFA	513	52.53	1.44	49.89	57.37
MS	513	5.58	0.60	3.84	7.38
CWT	513	426.66	11.80	379.36	468.21
BFT	513	13.16	1.59	1.64	24.25

Z is an economic traits after correcting for environmental factors.

C18:1 = oleic acid; SFA = saturated fatty acid; MUFA = monounsaturated fatty acid;

MS = marbling score; CWT = carcass weight; BFT = backfat thickness.

Table 3.4. Mean and standard deviation of adjusted economic trait (Z) which are divided into two classes

Economic trait	Group (Z')	N	Mean	SD
C18:1	1	105	46.01	0.75
	0	408	43.05	0.50
SFA	1	97	39.90	0.65
	0	416	42.38	0.51
MUFA	1	97	55.15	0.85
	0	416	51.92	0.65
MS	1	171	6.26	0.39
	0	342	5.24	0.34
CWT	1	176	439.67	7.89
	0	337	419.86	6.68
BFT	1	151	11.78	1.64
	0	362	13.74	1.15

Z is an economic traits after correcting for environmental factors.

C18:1 = oleic acid; SFA = saturated fatty acid; MUFA = monounsaturated fatty acid;

MS = marbling score; CWT = carcass weight; BFT = backfat thickness.

인 효과(유전자의 효과)에 관심이 있으므로 한우의 경제형질에 영향을 미치는 우수 유전자를 선별하기에 앞서, 한우의 환경적인 요인을 보정하여 유전자 선별의 정확도를 높이려 한다. 따라서 경제형질에 영향을 미치는 환경요인인 사육농장의 효과와 도축일령의 효과를 보정하였고, Table 3.3에 보정된 경제형질 Z 의 평균과 표준편차를 나타내었다.

환경요인이 보정된 경제형질 Z 도 마찬가지로 각 경제형질에서 우성인 집단($Z' = 1$)과 열성인 집단($Z' = 0$)으로 이분화하여 우수유전자형을 선별한다. Table 3.4는 집단(Z')에 따른 보정된 경제형질(Z)의 평균과 표준편차를 나타낸 것이다.

두 경제형질 Y 와 Z 를 살펴보면, Y 에 비해서 Z 의 편차가 줄어든 것을 확인할 수 있다. 이는 환경적인 요인의 영향이 경제형질에서 제거되었기 때문이며, 그만큼 환경적인 요인이 경제형질에 영향을 많이 미친다는 것을 알 수 있다. 또한 Y' 보다 Z' 의 우성 그룹의 N 이 더 작다. 이는 Y 는 환경적인 요인과 유전적인 요인이 함께 고려된 상태에서 경제형질을 우성으로 판단하였고, Z 는 유전자의 효과만을 고려한 상태에서 경제형질을 우성으로 판단하였기 때문이다. 따라서 Z' 가 우성 그룹을 더욱 정밀하게 표현한다는 것을 알 수 있다.

Table 3.5. Accuracy of each economic trait (Y)

Economic trait (Y)	SNP single	Train data accuracy	Test data accuracy	CVC
	SNP combination			
C18:1	g.3977-325 T>C	0.6148	0.6148	10/10
	c.408 G>C*g.3977-325 T>C	0.6362	0.6206	8/10
SFA	g.3977-325 T>C	0.5945	0.5945	10/10
	c.408 G>C*g.3977-325 T>C	0.6098	0.5968	6/10
MUFA	g.3977-325 T>C	0.6054	0.6054	10/10
	c.408 G>C*g.3977-325 T>C	0.6301	0.6010	10/10
MS	g.3977-325 T>C	0.5749	0.5749	10/10
	c.408 G>C*g.3977-325 T>C	0.5871	0.5491	10/10
CWT	c.388 G>A	0.5605	0.5345	7/10
	c.388 G>A*g.3977-325 T>C	0.5732	0.5231	8/10
BFT	c.388 G>A	0.5491	0.5224	7/10
	c.388 G>A*g.3977-325 T>C	0.5945	0.5895	10/10

Y is an economic traits before correcting for environmental factors.

SNP = single nucleotide polymorphism; CVC = cross-validation consistency; C18:1 = oleic acid;

SFA = saturated fatty acid; MUFA = monounsaturated fatty acid; MS = marbling score;

CWT = carcass weight; BFT = backfat thickness.

3.3. MDR 기법 적용 및 보정 전·후 정확도 비교

각 경제 형질별로 우수 단일SNP와 SNP조합을 선별하기 위해서 다중인자차원축소(multifactor dimensionality reduction; MDR) 기법을 사용하였다. 유전자는 한우의 등급 및 지방산과 깊은 연관이 있다고 알려진 FABP4(adipocyte fatty acid binding protein 4) 유전자의 6가지 SNP(g.2634+1018 A>T, c.280 A>G, c.388 G>A, c.408 G>C, g.3977-325 T>C, c.456 A>G)이다 (Damcotta, 2004; Chmurzynska, 2006; Oh 등, 2012). Table 3.5와 Table 3.6은 각각 보정 전(Y)-후(Z) 경제 형질별로 선별된 우수 단일SNP와 SNP조합을 나타낸 것이다. 그리고 MDR 기법으로 선별된 SNP의 훈련용 데이터와 테스트 데이터의 분류 정확도를 계산하여 나타내었다.

C18:1, SFA, MUFA, MS의 경우 우수 단일SNP로 g.3977-325 T>C가 선별되었고, C18:1은 보정 전의 테스트 정확도가 61.48%, 보정 후의 테스트 정확도가 88.10%로 정확도가 26.62%p 높아졌다. SFA는 보정 전의 테스트 정확도가 59.45%, 보정 후의 테스트 정확도가 91.24%로 정확도가 31.79%p 높아졌으며 MUFA는 보정 전의 테스트 정확도가 60.54%, 보정 후의 테스트 정확도가 91.24%로 정확도가 30.07%p 높아졌다. 마지막으로 MS는 보정 전의 테스트 정확도가 57.49%, 보정 후의 테스트 정확도가 72.95%로 정확도가 15.46%p 높아졌다. CWT의 경우는 우수 단일SNP로 c.388 G>A가 선별되었고 보정 전의 테스트 정확도가 53.45%, 보정 후의 테스트 정확도가 59.89%로 정확도가 6.44%p 높아졌다. CWT는 다른 경제형질에 비해 정확도가 소폭 증가하였는데 이는 CWT가 유전적인 요인의 영향($p > 0.05$) 보다 환경적인 요인의 영향($p < 0.001$)을 강하게 받기 때문에 나타난 결과이다. BFT는 보정 전에는 CWT와 동일하게 c.388 G>A가 정확도 52.24%로 선별되었으나, 보정 후에는 다른 경제형질(C18:1, SFA, MUFA, MS)과 동일하게 g.3977-325 T>C가 정확도 75.55%로 선별되었고, 정확도는 보정 후에 23.31%p 높아졌다.

다음으로 우수 SNP조합은 C18:1, SFA, MUFA, MS의 경우 c.408 G>C * g.3977-325 T>C가 선별되었고, C18:1은 보정 전의 테스트 정확도가 62.06%, 보정 후의 테스트 정확도가 98.53%로 정확도가 34.91%p 높아졌다. SFA는 보정 전의 테스트 정확도가 59.68%, 보정 후의 테스트 정확도가 95.05%로

Table 3.6. Accuracy of each adjusted economic trait (*Z*)

Economic trait (<i>Z</i>)	SNP single SNP combination	Train data accuracy	Test data accuracy	CVC
C18:1	g.3977-325 T>C	0.8810	0.8810	10/10
	c.408 G>C*g.3977-325 T>C	0.9853	0.9853	10/10
SFA	g.3977-325 T>C	0.9124	0.9124	10/10
	c.408 G>C*g.3977-325 T>C	0.9505	0.9505	10/10
MUFA	g.3977-325 T>C	0.9124	0.9124	10/10
	c.408 G>C*g.3977-325 T>C	0.9573	0.9462	10/10
MS	g.3977-325 T>C	0.7295	0.7295	10/10
	c.408 G>C*g.3977-325 T>C	0.8174	0.8129	10/10
CWT	c.388 G>A	0.5989	0.5989	10/10
	g.2634+1018 A>T*c.388 G>A	0.6229	0.5782	6/10
BFT	g.3977-325 T>C	0.7555	0.7555	10/10
	c.280 A>G*g.3977-325 T>C	0.8979	0.8979	10/10

Z is an economic traits after correcting for environmental factors.

SNP = single nucleotide polymorphism; CVC = cross-validation consistency; C18:1 = oleic acid;

SFA = saturated fatty acid; MUFA = monounsaturated fatty acid; MS = marbling score;

CWT = carcass weight; BFT = backfat thickness.

정확도가 34.07%p 높아졌으며 MUFA는 보정 전의 테스트 정확도가 60.10%, 보정 후의 테스트 정확도가 94.62%로 정확도가 32.72%p 높아졌다. 마지막으로 MS는 보정 전의 테스트 정확도가 54.91%, 보정 후의 테스트 정확도가 81.29%로 정확도가 23.03%p 높아졌다. CWT와 BFT의 경우에는 보정 전·후의 우수 유전자조합에 차이가 있었다. CWT는 보정 전에 우수 SNP조합으로 c.388 G>A*g.3977-325 T>C가 테스트 정확도 52.31%로 선별되었으나, 보정 후 우수 SNP조합으로 g.2634+1018 A>T*c.388 G>A가 테스트 정확도 57.82%로 선별되었다. 정확도의 증가폭은 5.51%p로 다른 형질에 비해 소폭 증가하였으며, 이는 단일SNP에서와 마찬가지로 CWT가 환경적인 요인의 영향을 강하게 받기 때문이다. BFT도 보정 전에는 CWT와 동일하게 우수 SNP조합으로 c.388 G>A*g.3977-325 T>C가 테스트 정확도 52.24%로 선별되었으나, 보정 후 우수 유전자조합으로는 c.280 A>G*g.3977-325 T>C가 테스트 정확도 89.97%로 선별되었고, 정확도는 37.73%p 높아졌다.

이와같은 결과는 유전적인 요인의 영향보다 환경적인 요인의 영향을 많이 받는 CWT 외의 다른 경제형질들은 두 요인의 영향을 복합적으로 받기 때문에 경제형질에서 환경적인 요인의 영향을 보정하게 되면 순수하게 유전적인 요인의 영향만 남게 되어 유전자 선별의 정확도를 높일 수 있다는 것을 보여준다.

또한 Table 3.6에서 단일SNP의 정확도와 SNP조합의 정확도를 비교해보면, SNP조합의 정확도가 단일SNP 보다 C18:1에서 10.43%p 더 높고, SFA에서 3.81%p, MUFA에서 3.38%p, MS에서 8.34%p, BFT에서 14.24%p 더 높다. 따라서 단일SNP보다 SNP의 상호작용을 고려했을 때 분류의 정확도를 더 높일 수 있었다.

4. 결론

가축의 경제형질에는 유전적인 요인과 환경적인 요인이 복합적으로 영향을 미친다. 우리는 우수 유전자를 찾는 것에 목적이 있으므로, 환경적인 요인을 배제하고 순수하게 유전적인 요인만을 고려하고자 했다. 따라서 한우의 경제형질에 대한 통계 모형을 구축하고, 순수한 유전요인을 찾기 위해서 환경적인 요

인을 보정하였고, 결론적으로 유전요인에 대한 분석에서 우수 유전자 선별의 정확도를 높일 수 있었다.

본 연구에서는 경북지역 17개의 사육농장에서 얻은 513두의 한우 데이터를 이용하여, 한우의 사육환경과 유전자로 한우의 품질에 영향을 미치는 주 경제형질들에 대한 통계모형을 구축했다. 여기서 한우의 품질에 영향을 미치는 경제형질 Y 는 올레인산(C18:1), 포화지방산(SFA), 단일불포화지방산(MUFA), 근내지방도(MS), 도체중량(CWT), 등지방두께(BFT) 각각에 초점을 맞추었다. 그리고 유전적 요인으로써 사용된 유전자는 한우의 등급과 지방산과 깊은 연관이 있다고 알려진 FABP4 유전자의 6가지 SNP(g.2634+1018 A>T, c.280 A>G, c.388 G>A, c.408 G>C, g.3977-325 T>C, c.456 A>G)이며, 환경적 요인은 도축일령과 사육농장효과를 사용하였다. 따라서 이 통계모형에서 경제형질인 Y 값에 영향을 미치는 환경요인인 도축일령(age)의 공변량 효과와 사육환경요인에 해당하는 17개의 사육농장(farm)의 각 효과를 추정하고, 경제형질 Y 값에서 이 두가지 환경요인 효과를 제거하여 순수한 유전자의 효과만 남은 보정된 경제형질 Z 값을 구하였다. 그리고 보정 전의 경제형질 Y 값과 보정 후의 경제형질 Z 값 각각을 MDR 기법에 적용하여 우수 단일SNP와 SNP조합을 선별하였다.

보정 전의 경제형질 Y 값과 보정 후의 경제형질 Z 값에 영향을 미치는 우수 유전자 선별의 정확도를 비교해 보면 올레인산, 포화지방산, 단일불포화지방산, 근내지방도의 경우 경제형질 Y 값과 Z 값 모두에서 우수 단일SNP로 g.3977-325 T>C가 선별되었고, 우수 SNP조합으로 c.408 G>C * g.3977-325 T>C가 선별되었다. 이때 경제형질 Y 값과 Z 값에 대한 단일SNP와 SNP조합의 분류 정확도를 비교해 보면, 올레인산의 경우 각각 26.62%p, 34.91%p 높아졌고 포화지방산은 각각 31.79%p, 34.07%p 높아졌으며 단일불포화지방산은 각각 30.07%p, 32.72%p 높아졌다. 그리고 근내지방도는 각각 15.46%p, 23.03%p 높아졌다. 등지방두께의 경우에는 Y 값과 Z 값의 선별결과에 차이가 있지만 우수 단일SNP는 동일하게 g.3977-325 T>C가 선별되었으며 정확도는 23.31%p 높아졌다. 따라서 한우의 경제형질에 대한 유전요인 연구에서 유전적 요인과 환경적 요인 모두의 영향을 받는 경제형질의 경우에는 환경적 요인의 보정을 거친 경제형질 Z 값을 활용하는 것이 Y 값을 사용하는 것보다 유전자 선별의 정확도를 높일 수 있었다. 또한 도체중량의 경우에는 유전적 요인보다 환경적 요인의 영향을 많이 받는 경제형질이기 때문에 환경요인을 보정하더라도 유전요인의 선별에 영향을 미치지 않으므로 정확도가 크게 높아지지 않는다는 것을 확인하였다.

아울러 보정 후의 경제형질 Z 값을 이용한 분석에서 단일SNP의 정확도와 SNP조합의 정확도를 비교해 보면, SNP 조합이 단일 SNP 보다 정확도가 더 높았다. 이로 인해 단일 유전자의 효과보다 유전자의 상호작용효과가 경제형질을 더 잘 설명한다는 것을 확인하였다.

우리는 한우의 경제형질에서 환경적 요인인 사육농장과 도축일령을 보정하여 순수한 FABP4 유전자에 대한 새로운 경제형질 값을 구하였고, 이 값을 분석에 활용하여 우수 유전자 선별의 정확도를 높였으며, 가장 우수한 유전자로 g.3977-325 T>C, 유전자조합으로 c.408 G>C * g.3977-325 T>C를 선별하였다.

References

- Blumer, T. N. (1963). Relationship of marbling to the palatability of beef, *Journal of Animal Science*, **22**, 771-778.
- Casas, E., White, S. N., Riley, D. G., Smith, T. P. L., Brenneman, R. A., Olson, T. A., Johnson, D. D., Coleman, S. W., Bennett, G. L. and Chase, C. C. (2005). Assessment of single nucleotide polymorphisms in genes residing on chromosomes 14 and 29 for association with carcass composition traits in *Bos indicus* cattle, *Journal of Animal Science*, **83**, 13-19.
- Chmurzynska, A. (2006). The multigene family of fatty acid-binding proteins (FABPs): function, structure and polymorphism, *Journal of Applied Genetics*, **47**, 39-48.

- Damcotta, C. M., Moffetta, S. P., Feingolda, E., Barmadaa, M. M., Marshallb, J. A., Hammanb, R. F. and Ferrella, R. E. (2004). Genetic variation in fatty acid-binding protein-4 and peroxisome proliferator-activated receptor γ interactively influence insulin sensitivity and body composition in males, *Metabolism*, **53**, 303–309.
- Lee, J. Y., Kwon, J. C. and Kim, J. J. (2008). Multifactor dimensionality reduction (MDR) analysis to detect single nucleotide polymorphisms associated with a carcass trait in a Hanwoo population, *Asian Australasian Journal of Animal Sciences*, **21**, 784–788.
- Lee, J. Y. and Lee, H. G. (2009). Multifactor Dimensionality Reduction (MDR) Analysis by Dummy Variables, *The Korean Journal of Applied Statistics*, **22**, 435–442.
- Lee, J. Y. and Lee, J. H. (2010). Support vector machine and multifactor dimensionality reduction for detecting major gene interactions of continuous data, *Journal of the Korean Data & Information Science Society*, **21**, 1271–1280.
- Mandell, I. B., Buchanan-Smith, J. G. and Campbell, C. P. (1998). Effects of forage vs grain feeding on carcass characteristics, fatty acid composition, and beef quality in Limousin-cross steers when time on feed is controlled, *Journal of Animal Science*, **76**, 2619–2630.
- Matsuhashi, T., Maruyama, S., Uemoto, Y., Kobayashi, N., Mannen, H., Abe, T., Sakaguchi, S. and Kobayashi, E. (2011). Effects of bovine fatty acid synthase, stearoyl-coenzyme A desaturase, sterol regulatory element-binding protein 1, and growth hormone gene polymorphisms on fatty acid composition and carcass traits in Japanese Black cattle, *Journal of Animal Science*, **89**, 12–22.
- Melton, S. L., Amiri, M., Davis, G. W. and Backus, W. R. (1982). Flavor and chemical characteristics of ground beef from grass-, forage-grain-and grain-finished steers, *Journal of Animal Science*, **55**, 77–87.
- Oh, D. Y. (2014). Identification of the SNP (single nucleotide polymorphism) within candidate gene associated with fatty composition in Hanwoo, Ph. D. Thesis, Yeungnam University.
- Oh, D. Y., Lee, Y. S., La, B. M., Yeo, J. S., Chung, E. Y., Kim, Y. Y. and Lee, C. Y. (2011). Fatty acid composition of beef is associated with exonic nucleotide variants of the gene encoding FASN, *Molecular Biology Reports*, **39**, 4083–4090.
- Oh, D. Y., Lee, Y. S. and Yeo, J. S. (2012). Identification of the SNP (single nucleotide polymorphism) for fatty acid composition associated with beef flavor-related FABP4 (fatty acid binding protein 4) in Korean cattle, *Asian Australasian Journal of Animal Sciences*, **24**, 757–765.
- Ritchie, M. D., Hahn, L. W., Roodi, N., Bailey, L. R., Dupont, W. D., Parl, F. F. and Moore, J. H. (2001). Multifactor dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer, *The American Society of Human Genetics*, **69**, 138–147.
- Sturdivant, C. A., Lunt, D. K., Smith, G. C. and Smith, S. B. (1991). Fatty acid composition of subcutaneous and intramuscular adipose tissues and M. longissimus dorsi of Wagyu cattle, *Meat Science*, **32**, 449–458.

한우의 FABP4 유전자 선별에서 사육환경 보정 효과

김현지^a · 이제영^{a,1}

^a영남대학교 통계학과

(2015년 4월 16일 접수, 2015년 5월 7일 수정, 2015년 8월 21일 채택)

요약

가축의 경제적인 특성은 환경적인 요인과 유전적인 요인의 영향을 복합적으로 받는다. 우리는 한우의 경제적 특성에 영향을 미치는 유전적인 요인에 관심이 있으며, 우리의 목적은 경제형질에서 환경적인 요인을 보정하여 유전효과를 더욱 정확하게 검증하는 것이다. 본 연구에서는 환경적인 요인과 유전적인 요인으로 구성된 통계모형을 구축하고, 이 모형에서 환경적인 요인인 사육농가의 효과와 도축일령의 효과를 제거하여 보정된 경제형질값을 구한다. 그리고 보정 전·후 경제형질값을 다중인자차원축소 방법에 적용하여 각각의 우수 유전자와 유전자조합을 선별하고, 정확도를 비교한다. 그 결과, 우리는 환경요인을 보정한 경제형질값을 활용하여 우수 유전자 선별의 정확도를 높였고, 한우의 등급과 지방산과 깊은 연관이 있는 우수한 FABP4 유전자를 선별하였다.

주요용어: 다중인자차원축소, 단일염기다형성, 보정, 유전요인, 환경요인

¹교신저자: (712-749) 경북 경산시 대동 214-1, 영남대학교 통계학과. E-mail: jlee@yu.ac.kr