

고품질 한우 생산 유전자 연구에서 환경 요인을 보정한 통계적 모형 제안

장지은¹ · 이제영² · 오동엽³

¹²영남대학교 통계학과 · ³경상북도축산기술연구소

접수 2015년 10월 5일, 수정 2015년 11월 27일, 게재확정 2015년 11월 28일

요약

개체의 표현형은 대부분 유전적인 요인의 영향과 환경적인 요인의 영향을 모두 받는다. 따라서 한우의 경제적인 특성과 연관이 있는 유전자 마커 선별 연구에서도 관심이 있는 유전적인 요인의 효과를 좀 더 정확히 보기 위해서는 환경적인 요인의 보정이 필요하다. 본 연구의 목적은 고품질 한우 생산을 위한 우수 유전자 마커 선별 연구에서 환경적인 요인이 보정된 새로운 통계 모형을 제안하고 그 효과를 규명하는 데 있다. 먼저 환경적인 요인과 유전적인 요인을 모두 포함한 통계모형을 구축한 뒤, 환경적인 요인인 도축일령과 사육농가의 효과를 제거하여 보정된 경제형질의 값을 구한다. 그리고 다중인자차원축소 방법을 보정 전·후 데이터에 각각 적용하여 우수 유전자 마커 조합을 선별하고 정확도를 비교한다. 사용된 경제형질은 C18:1, SFA, MUFA, MS, CWT, BFT이며 사용된 유전자 마커는 49개 LPL 유전자 마커 중 지방산 조성 및 경제 형질 능력 검정을 통해 나머지에 비해 더 뛰어난 유전자 마커로 선별된 6개 (g.6960 A>T, g.6974 G>A, g.21604 G>A, g.22488 G>T, g.22649 G>A, g.25670 C>T)이다.

주요용어: 다중인자차원축소 방법, 보정 효과, 유전적인 요인, 환경적인 요인.

1. 서론

일반적으로 소고기에 대한 소비자들의 선호도는 맛과 육질에 따라 달라진다. 따라서 소비자들의 입맛을 사로잡기 위한 고급육 생산 연구가 활발히 진행되고 있다. 소고기의 육질은 근내지방도 (marbling score; MS), 육색 및 조직감 등에 의해 결정되며 이런 요인들은 연도, 풍미, 다즙성과 밀접한 관련이 있다 (Lee 등, 1994; Monson 등, 2005; Robins 등, 2003; Vander Wal 등, 1997). 그 중 풍미는 불포화지방산의 함량에 영향을 받으며, 한우가 불포화지방산 중 특히 oleic acid (C18:1) 함량이 많은 것으로 알려져 있다. 또한 oleic acid (C18:1)은 MUFA (mono-unsaturated fatty acid)의 80%이상을 차지하고 있으며, 소고기의 풍미를 좌우하는 요소라고 보고된 바 있다 (Dryden 과 Marchello, 1970; Tsuji, 2008; Yoshimura 와 Namikawa, 1983). 등지방두께 (backfat thickness; BFT)와 도체중량 (carcass weight; CWT)도 축산물품질평가원의 한우 등급판정에 있어 중요한 도체형질 항목이다. 또한 소고기의 맛은 지방산 조성에 영향을 많이 받는데 이러한 지방산 합성 과정에는 SCD (stearoyl-CoA desaturase), FASN (fatty acid synthase), LPL (lipoprotein lipase), SREBPs (sterol regulatory element binding proteins) 등의 여러 유전자들이 관여 한다 (Oh 등, 2014).

¹ (712-749) 경북 경산시 대동 214-1, 영남대학교 통계학과, 석사과정.

² 교신저자: (712-749) 경북 경산시 대동 214-1, 영남대학교 통계학과, 교수. E-mail: jlee@yu.ac.kr

³ (750-871) 경북 영주시 안정면 대룡산로 186, 경상북도축산기술연구소, 연구원.

최근 여러 나라들과의 FTA에 따른 쇠고기 시장 개방으로 한우의 입지가 좁아지고 있어 한우의 맛, 품질 등과 연관이 있는 경제형질들과 이 경제형질들에 긍정적 영향을 미치는 유전자에 대한 연구가 점점 더 중요해지고 있다. 그동안 유전자 마커 선별 연구에서는 유전자 마커들의 상호작용을 계산하기 위해 선형모형과 같은 통계적 모형을 사용해왔으나 이 경우 유전자 마커의 수가 늘어남에 따라 모형의 복잡도가 증가되고 해석이 어려워지는 등의 문제가 발생한다. 그래서 제안된 방법이 다중인자차원축소 (multifactor dimensionality reduction, MDR; Ritchie 등, 2001) 방법이다. MDR 방법은 이분형 데이터에만 적용이 가능하기 때문에 연속형 데이터에도 MDR 방법을 적용하기 위한 CART (classification and regression tree) 방법을 활용한 확장된 다중인자 차원 축소 (expanded multifactor dimensionality reduction; E-MDR; Lee 등, 2008) 방법, SVM (support vector machine)을 이용한 다중인자 차원 축소 (support vector machine multifactor dimensionality reduction; SVM-MDR; Lee 와 Lee, 2010) 방법, 유전자 행렬 맵핑을 활용한 우수 유전자 조합 선별 (Lee 등, 2010) 등도 개발되었으며, 유전자 관련 연구에서 다양한 형태로 쓰이고 있다 (Yeo 등, 2013; Oh 와 Lee, 2014). 본 연구에서도 한우의 경제형질에 긍정적 영향을 주는 우수 유전자 마커를 선별하기 위해 MDR 방법을 적용하고자 한다.

한우의 경제형질을 비롯한 개체의 표현형은 유전적인 요인과 환경적인 요인의 영향을 받는다. 하지만 한우의 유전자 관련 연구에서는 한우의 맛과 품질에 영향을 주는 유전적 요인에만 관심이 있다. 따라서 순수한 유전적 요인의 효과를 보기 위해 환경적 요인의 보정이 필요하다. 이전 연구에서는 선형모형이 사용되어 왔으나 (Casas 등, 2005; Matsuhashi 등, 2011), 우리는 선형모형을 이용하여 환경적 요인을 사전에 보정하는 새로운 통계 모형을 구축한 뒤, 그 모형으로부터 얻어진 보정된 경제형질에 MDR 방법을 적용하여 유전자 마커를 선별할 것이다.

본 연구에서는 한우의 맛과 질에 영향을 주는 경제형질 6개 (C18:1, SFA, MUFA, MS, CWT, BFT)에 긍정적으로 작용하는 유전자 마커를 선별하는데 있어 환경적 요인의 보정 효과를 규명하고자 한다. 사용된 유전자 마커는 지방산 합성 과정에서 소장과 간에서 생산된 hylomicron과 VLDL (Very-Low-Density Lipoprotein)를 FA (Free fatty acid) 형태로 바꾸어 주는 역할을 하는 (Gotoda 등, 1991; Havel 등, 1970; Larosa 등, 1970) LPL 유전자의 49개 마커 중 지방산 조성 및 경제형질 능력 검정을 통해 나머지에 비해 더 뛰어난 유전자 마커로 선별된 6개 (g.6960 A>T, g.6974 G>A, g.21604 G>A, g.22488 G>T, g.22649 G>A, g.25670 C>T)이다 (Oh, 2014). 유전적 요인과 환경적 요인을 포함한 일반선형모형에서 환경 요인인 도축일령 (age)과 사육환경 (farm) 효과를 제거한 보정된 경제형질 값을 계산한 뒤, 보정 전·후 데이터에 각각 MDR을 적용하여 우수 유전자 마커 조합을 선별하고 정확도를 비교한다.

2. 통계적 방법

먼저 2.1절에서는 한우의 경제형질 연구에서 사용되는 유전적인 요인과 환경적인 요인을 모두 포함한 일반선형모형을 살펴보고, 2.2절에서는 2.1절에 소개된 일반선형모형에서 환경적인 요인의 효과를 제거한 보정된 일반선형모형인 새로운 통계적 모형을 제시한다. 마지막으로 2.3절에서는 본 연구에 사용된 유전자 마커 조합 선별 방법인 MDR 방법에 대해 소개한다.

2.1. 한우의 경제형질에 대한 일반선형모형

일반적으로 한우의 경제형질과 같은 개체의 표현형은 유전적인 요인과 환경적인 요인의 영향을 모두 받는다. 따라서 특정 요인의 영향에 대한 연구를 하는 경우 다른 요인의 영향을 보정할 필요가 있다. 우

선, 개체의 표현형을 식으로 나타내면 다음과 같다.

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

여기서 P 는 개체의 표현형 (phenotype)이고 E 는 환경적인 요인의 효과 (environmental effect), G 는 유전적인 요인의 효과 (genetic effect)이다. 각 요인들의 영향력을 평가하는 방법으로는 주로 다중선형 회귀모형 (multiple linear regression)이 쓰이는데 (Casas 등, 2005; Matsuhashi 등, 2011) 다음과 같은 식으로 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. 위의 식에서 유전적인 요인과 환경적인 요인을 나누어 나타내면 다음과 같다.

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

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.

본 연구에서는 개체의 표현형을 나타내는 위 모형식을 바탕으로 한우의 경제형질에 유의한 영향을 미치는 변수들을 포함하여 모형을 구축한다. 유전적인 요인으로는 SNP, 환경적인 요인으로는 사육농가인 farm과 도축일령인 age를 이용한다. 식은 다음과 같다.

$$\begin{aligned} y_k = & \mu + \alpha_0(\text{age}_k - \overline{\text{age}}) + \alpha_1 \text{farm}_{1k} + \cdots + \alpha_{l-1} \text{farm}_{l-1,k} \\ & + \beta_1 \text{SNP}_{1k} + \cdots + \beta_m \text{SNP}_{mk} + \epsilon_k \end{aligned} \quad (2.4)$$

$$i = 1, 2, \dots, l; j = 1, 2, \dots, m; k = 1, 2, \dots, n$$

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 마커, β_j 는 j 번째 유전자 마커의 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. 환경적인 요인이 보정된 통계적 모형

본 연구는 한우의 경제형질에 영향을 미치는 유전적인 요인을 규명하는 것에 목적이 있으므로 2.1절에서 소개한 유전적인 요인과 환경적인 요인이 모두 포함된 모형식에서 경제형질에 영향을 미치는 환경적인 요인을 보정한 새로운 통계적 모형을 구축하여 그 모형으로부터 얻어진 경제형질 값으로 유전적 효

과를 분석할 것이다. 한우의 경제형질에 대한 일반선형모형인 식 2.4를 2.3의 모형식 형태로 나타내면 다음과 같다.

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{E}\boldsymbol{\alpha} + \mathbf{G}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2.5)$$

여기서 한우 사육환경과 도축일령은 식 2.3의 환경적인 효과 $\mathbf{E}\boldsymbol{\alpha}$ 에 해당하며, SNP는 식 2.3의 유전적인 효과 $\mathbf{G}\boldsymbol{\beta}$ 에 해당한다.

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

$$\mathbf{G}\boldsymbol{\beta} = \begin{pmatrix} SNP_{11} & SNP_{21} & \cdots & SNP_{m1} \\ SNP_{12} & SNP_{22} & \cdots & SNP_{m2} \\ SNP_{13} & SNP_{23} & \cdots & SNP_{m3} \\ \vdots & \vdots & \ddots & \vdots \\ SNP_{1n} & SNP_{2n} & \cdots & SNP_{mn} \end{pmatrix} \begin{pmatrix} \beta_1 & \beta_2 & \beta_3 & \cdots & \beta_m \end{pmatrix}'$$

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

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

$$\mathbf{Z} = \mathbf{Y} - \mathbf{E}\hat{\boldsymbol{\alpha}}$$

여기서 \mathbf{Z} , \mathbf{Y} , $\hat{\boldsymbol{\alpha}}$ 는 다음과 같다.

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

본 연구에서는 순수 유전요인에 대한 경제형질인 \mathbf{Z} 를 다음 절에서 소개하는 MDR 방법에 적용하여 우수 유전자 마커를 선별한다.

2.3. 다중인자차원축소 방법

인간의 질병 유무와 같은 이분형 자료에 대한 유전자의 상호작용 효과를 찾기 위해 다중인자 차원 축소 (MDR; Richie 등, 2001) 방법이 제시되었으며, 이 방법은 일반화된 선형 모형의 전통적 통계기법과는 달리 모수에 대한 추정과 모형에 대한 가정이 필요하지 않은 비모수적 방법이다. 실험군-대조군의 비율을 통해 독립변수를 고위험군과 저위험군으로 분류한 뒤, 목표변수에 대한 오분류율을 비교하는 방법으로 검정력 평가를 통해 높은 검정력이 입증되었다. MDR 방법의 절차는 아래와 같다.

- Step 1. 데이터를 크기가 동일한 10개의 셋으로 나눈다. 그 중 9개를 학습용 자료, 나머지 하나를 검증용 자료로 설정한다.
- Step 2. 선택된 위험 요인 조합에서 요인의 각 수준을 기초로 한 개체들을 각 셀에 기술한다. 예를 들어 $k=2$ 인 경우, 위험 요인은 2개의 수준으로 되어있으므로 $3^2=9$ 개의 셀을 가진다. 9개의 셀에 각각 실험-대조군의 도수를 기술한다.
- Step 3. 전체 데이터셋의 실험-대조군 비율을 계산하여 한계점으로 설정한 뒤, 각 셀의 실험-대조군의 비율을 구하여 한계점보다 크거나 같으면 고 위험군 (high risk), 한계점보다 작으면 저 위험군 (low risk)으로 정한다.
- Step 4. 학습용 자료의 모든 셀에서 잘못 분류된 비율인 학습용 자료의 오분류율과 검증용 자료를 이용하여 잘못 예측된 분류 비율인 검증용 자료의 오분류율을 구한다.
- Step 5. 단계 1에서 정의한 10쌍의 모든 데이터 셋에 대해서 위의 과정을 반복하여 학습용 자료의 오분류율의 평균과 검증용 자료의 오분류율의 평균을 구한다.
- Step 6. 구해진 학습용 자료의 오분류율의 평균과 검증용 자료의 오분류율의 평균값을 비교하여 가장 낮은 조합을 우수 위험 요인 조합으로 선정한다.

본 연구에서는 위 절차에 따라 MDR 방법을 보정 전·후 데이터에 적용하여 각각의 데이터에서 우수 유전자 마커 조합을 선별한 뒤, 선별된 우수 유전자 마커 조합들의 정확도를 비교한다.

3. MDR 적용 및 결과

3.1. 실험자료

본 연구에 사용된 데이터는 경북지역에서 18개의 씨수소를 통해 얻어진 513두의 한우로부터 수집되었다 (Oh 등, 2011). 경제 형질은 한우의 맛과 향에 영향을 주는 올레인산 (Melton 등, 1982; Mandell 등, 1998; Matsuhashi 등, 2011)과 육질의 부드러움에 긍정적으로 작용하는 불포화지방산 (Melton 등, 1982), 그리고 한우의 품질에 중요한 지표가 되는 근내지방도, 도체중량, 등지방두께, 포화지방산을 사용하였다. SNP는 한우의 맛과 육질에 영향을 미치는 지방산 합성 과정에서 중요한 역할을 하는 LPL 유전자의 49개 마커 중 지방산 조성 및 경제 형질 능력 검정을 통해 나머지에 비해 더 뛰어난 유전자 마커로 선별된 6개 (g.6960 A>T, g.6974 G>A, g.21604 G>A, g.22488 G>T, g.22649 G>A, g.25670 C>T)를 이용하였다 (Oh, 2014).

Table 3.1 Mean and standard deviation of economic trait

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

Table 3.1은 보정 전 513두 데이터에 대한 각 경제형질의 평균과 표준편차를 나타낸 것이다. MDR 방법은 실험-대조군으로 이루어진 이분형 데이터에만 적용이 가능하기 때문에 본 연구에 사용된 6개의 경제형질을 이용하여 K-평균 군집분석 방법을 통해 데이터를 이분화 하였다. Table 3.2는 이분화 된 데이터의 각 경제형질의 평균과 표준편차를 나타낸 것이다. C18:1, MUFA, MS, CWT는 한우의 품질과 양의 상관관계를 가지므로 우성 집단의 평균이 더 높고, SFA와 BFT는 음의 상관관계를 가지므로 우

성 집단의 평균이 더 낮다. Table 3.3은 환경적인 요인의 효과를 보정한 513두 데이터에 대한 각 경제형질의 평균과 표준편차를 나타낸 것이다. 보정된 데이터도 6개의 경제형질을 이용하여 K-평균 군집분석 방법을 통해 데이터를 이분화 하였다. Table 3.4는 이분화 된 데이터의 각 경제형질의 평균과 표준편차를 나타낸 것이다.

Table 3.2와 Table 3.4를 비교해보면, 보정 전에 비해 보정 후에 우성 집단과 열성 집단 간의 경제형질 차이 폭이 좁아졌음을 알 수 있다. 이는 경제형질에서 환경적인 요인의 영향을 제거했기 때문이며, 따라서 경제형질이 환경적인 요인의 영향을 받는다는 것을 알 수 있다.

Table 3.2 Mean and standard deviation of economic trait which are divided into two classes

Economic trait	Group	N	Mean	SD
C18:1	high	180	47.12	1.63
	low	333	42.78	1.69
SFA	high	176	37.58	1.63
	low	337	42.18	1.94
MUFA	high	174	56.72	1.76
	low	339	51.85	1.92
MS	high	275	6.99	0.98
	low	238	3.62	0.94
CWT	high	163	477.76	25.74
	low	350	403.72	26.34
BFT	high	418	11.36	3.28
	low	95	21.40	3.58

Table 3.3 Mean and standard deviation of adjusted economic trait

Economic trait	N	Mean	SD	Min	Max
C18:1	513	43.29	1.32	36.98	46.92
SFA	513	42.23	1.88	39.07	47.76
MUFA	513	52.13	1.42	46.02	56.13
MS	513	5.46	0.57	3.22	7.22
CWT	513	427.79	7.01	403.66	444.85
BFT	513	13.27	1.32	9.39	17.37

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

Economic trait	Group	N	Mean	SD
C18:1	high	441	43.66	0.98
	low	72	41.06	0.85
SFA	high	434	41.87	0.81
	low	79	44.23	0.80
MUFA	high	408	52.66	0.97
	low	105	50.07	0.89
MS	high	389	5.66	0.44
	low	124	4.82	0.42
CWT	high	411	430.44	4.52
	low	102	417.13	4.83
BFT	high	211	12.00	0.75
	low	302	14.15	0.82

3.2. 보정 전·후 데이터 적용 및 결과

본 연구에서는 보정 전·후 데이터를 이용하여 우수 유전자 마커 조합을 선별하기 위해 MDR 방법을 활용하였다. 보정 전·후의 각 데이터를 6개의 경제형질을 이용하여 K-평균 군집분석 방법으로 이분화한 뒤, MDR을 적용한 결과는 다음과 같다. 먼저 Table 3.5는 보정 전 513두 데이터를 MDR에 적용시킨 결과이다. 각 경제형질별로 학습용 자료의 정확도가 높은 상위 5개의 조합을 선별하였다. Table 3.6은 환경적인 요인의 효과를 보정한 513두 데이터를 MDR에 적용하여 각 경제형질별로 학습용 자료의 정확도가 높은 상위 5개의 조합을 선별한 결과이다.

Table 3.5 Accuracy of top 5 two-way gene interactions from 6 traits before adjusting environmental factor

Rank	SNP combination	C18:1		SNP combination	SFA		SNP combination	MUFA	
		Train-data accuracy	Test-data accuracy		Train-data accuracy	Test-data accuracy		Train-data accuracy	Test-data accuracy
		1	(g.6960 A>T, g.21604 G>A)		0.6214	0.6099		(g.6974 G>A, g.22649 G>A)	0.5754
2	(g.21604 G>A, g.22488 G>T)	0.5887	0.5673	(g.21604 G>A, g.22488 G>T)	0.5717	0.5231	(g.22488 G>T, g.25670 C>T)	0.5957	0.4973
3	(g.6960 A>T, g.22488 G>T)	0.5860	0.5246	(g.6960 A>T, g.21604 G>A)	0.5672	0.5551	(g.6974 G>A, g.22649 G>A)	0.5918	0.5445
4	(g.21604 G>A, g.25670 C>T)	0.5834	0.5477	(g.22488 G>T, g.25670 C>T)	0.5625	0.5442	(g.21604 G>A, g.22488 G>T)	0.5838	0.5480
5	(g.21604 G>A, g.22649 G>A)	0.5821	0.5532	(g.22488 G>T, g.22649 G>A)	0.5623	0.5290	(g.22649 G>A, g.25670 C>T)	0.5819	0.5658

Rank	SNP combination	MS		SNP combination	CWT		SNP combination	BFT	
		Train-data accuracy	Test-data accuracy		Train-data accuracy	Test-data accuracy		Train-data accuracy	Test-data accuracy
		1	(g.21604 G>A, g.22488 G>T)		0.5641	0.4915		(g.6960 A>T, g.22488 G>T)	0.5796
2	(g.6974 G>A, g.25670 C>T)	0.5631	0.4969	(g.22488 G>T, g.25670 C>T)	0.5690	0.5432	(g.22488 G>T, g.25670 C>T)	0.6132	0.5914
3	(g.6974 G>A, g.22488 G>T)	0.5580	0.5580	(g.6974 G>A, g.22488 G>T)	0.5665	0.5380	(g.21604 G>A, g.22488 G>T)	0.6086	0.6086
4	(g.21604 G>A, g.25670 C>T)	0.5579	0.5014	(g.22488 G>T, g.22649 G>A)	0.5661	0.5414	(g.6974 G>A, g.22649 G>A)	0.6081	0.5939
5	(g.6974 G>A, g.22649 G>A)	0.5518	0.5367	(g.21604 G>A, g.22488 G>T)	0.5639	0.5365	(g.22649 G>A, g.25670 C>T)	0.6067	0.5986

보정 전·후 데이터의 MDR 적용 결과에서 각 경제형질별로 가장 우수한 유전자 마커 조합의 정확도를 비교해보면, C18:1은 보정 전 (g.6960 A>T, g.21604 G>A)조합이 62.14%의 정확도로 가장 우수한 유전자 마커 조합으로 선별되었으나, 보정 후에는 80.89%의 정확도로 (g.21604 G>A, g.22488 G>T)조합이 선별되었으며 정확도는 18.75%p 높아졌다. SFA에서는 보정 전 57.54%의 정확도로 (g.6974 G>A, g.22649 G>A)조합이 보정 후에는 94.07%의 정확도로 (g.21604 G>A, g.22488 G>T)조합이 가장 우수한 유전자 마커 조합으로 선별되었고, 정확도는 36.53%p 높아졌다. MUFA는 보정 전 60.50%의 정확도로 (g.6960 A>T, g.21604 G>A)조합이 보정 후 83.29% 정확도로 (g.21604 G>A, g.22488 G>T)조합이 선별되었으며 정확도가 22.79%p 높아졌다. 또한 MS에서는 보정 전 (g.21604 G>A, g.22488 G>T) 조합이 56.41% 정확도로 가장 우수한 조합으로 선별되었고, 보정 후에는 87.66%의 정확도로 (g.6960 A>T, g.21604 G>A)조합이 선별되었으며 정확도는 31.25%p 높아졌다. CWT는 보정 전 57.96% 정확도로 (g.6960 A>T, g.22488 G>T)조합이 보정 후 90.32% 정확도로 (g.6974 G>A, g.22649 G>A)조합이 선별되었으며, 정확도는 32.36%p 높아졌다. 마지막으로 BFT에서는 보정 전 62.02% 정확도로 (g.6974 G>A, g.22488 G>T)조합이 가장 우수한 유전자 마커 조합

로 선별되었고, 보정 후 85.10%의 정확도로 (g.6960 A>T, g.25670 C>T)조합이 선별되었다. 정확도는 23.08%p 높아졌다.

Table 3.6 Accuracy of top 5 two-way gene interactions from 6 traits after adjusting environmental factor

Rank	SNP combination	C18:1		SNP combination	SFA		SNP combination	MUFA	
		Train-data accuracy	Test-data accuracy		Train-data accuracy	Test-data accuracy		Train-data accuracy	Test-data accuracy
1	(g.21604 G>A g.22488 G>T)	0.8089	0.7636	(g.21604 G>A g.22488 G>T)	0.9407	0.9407	(g.21604 G>A g.22488 G>T)	0.8329	0.8329
2	(g.6960 A>T g.21604 G>A)	0.7998	0.7628	(g.21604 G>A g.22649 G>A)	0.8561	0.8561	(g.21604 G>A g.22649 G>A)	0.7745	0.7745
3	(g.22488 G>T g.25670 C>T)	0.7452	0.7251	(g.6960 A>T g.22488 G>T)	0.8285	0.8285	(g.6960 A>T g.22488 G>T)	0.7731	0.7731
4	(g.21604 G>A g.25670 C>T)	0.7380	0.7380	(g.22488 G>T g.22649 G>A)	0.8031	0.7963	(g.6960 A>T g.21604 G>A)	0.7694	0.7360
5	(g.22488 G>T g.22649 G>A)	0.7290	0.7290	(g.6960 A>T g.21604 G>A)	0.8008	0.8008	(g.22488 G>T g.22649 G>A)	0.7580	0.7526
Rank	SNP combination	MS		SNP combination	CWT		SNP combination	BFT	
		Train-data accuracy	Test-data accuracy		Train-data accuracy	Test-data accuracy		Train-data accuracy	Test-data accuracy
1	(g.6960 A>T g.21604 G>A)	0.8766	0.8766	(g.6974 G>A g.22649 G>A)	0.9032	0.9032	(g.6960 A>T g.25670 C>T)	0.8510	0.8510
2	(g.6960 A>T g.25670 C>T)	0.8096	0.7912	(g.6974 G>A g.22488 G>T)	0.8935	0.8935	(g.21604 G>A g.25670 C>T)	0.8241	0.8173
3	(g.6960 A>T g.22488 G>T)	0.8034	0.8034	(g.6974 G>A g.25670 C>T)	0.7934	0.7934	(g.6960 A>T g.22649 G>A)	0.8013	0.8013
4	(g.6960 A>T g.22649 G>A)	0.7916	0.7916	(g.6960 A>T g.6974 G>A)	0.7745	0.7745	(g.6960 A>T g.6974 G>A)	0.7974	0.7974
5	(g.21604 G>A g.25670 C>T)	0.7651	0.7651	(g.6974 G>A g.21604 G>A)	0.7745	0.7745	(g.6960 A>T g.22488 G>T)	0.7827	0.7750

각 경제형질별로 정확도가 높은 상위 5개의 조합을 전체적으로 살펴보면, 보정 전의 경우에는 (g.21604 G>A, g.22488 G>T)조합이 6개의 경제형질 모두에서 상위 5개의 조합 안에 포함되어 종합적으로 봤을 때 우수한 유전자 마커 조합으로 볼 수 있다. 보정 후의 결과에서는 6개의 경제형질 모두에서 상위 5개의 조합 안에 포함된 조합은 없으나 (g.6960 A>T, g.21604 G>A)조합은 C18:1, SFA, MUFA, MS 4개의 경제형질에서 상위 5개의 조합 안에 포함되었고, (g.6960 A>T, g.22488 G>T)조합은 SFA, MUFA, MS, BFT 4개의 경제형질에서 상위 5개의 조합 안에 포함되어 종합적으로 봤을 때 (g.6960 A>T, g.21604 G>A), (g.6960 A>T, g.22488 G>T) 조합이 우수한 유전자 마커 조합임을 알 수 있다.

보정 전·후 데이터의 적용 결과에서 각 경제형질 별로 가장 우수한 유전자 마커 조합을 비교한 결과와 종합적으로 비교한 결과 모두 보정 전·후 다른 유전자 마커 조합이 선별됨을 알 수 있었다. 또한 모든 조합에서 보정 전보다 보정 후에 정확도가 훨씬 증가함을 확인할 수 있었다. 이는 한우의 경제형질에 긍정적인 영향을 주는 유전자 마커 조합 선별 연구에서 환경적인 요인을 보정하는 경우, 더 높은 정확도로 순수하게 유전적인 요인의 영향만을 고려하여 유전자 마커를 선별할 수 있다는 것을 보여준다. 그리고 6개의 경제형질 중 보정 전·후의 정확도 변화가 36.53%p로 가장 높아진 SFA는 다른 경제형질들에 비해 유전적인 요인의 영향을 더 많이 받는다고 볼 수 있고, 정확도의 변화가 18.75%p로 가장 적게 높아진 C18:1의 경우는 다른 경제형질들에 비해 환경적인 요인의 영향을 더 많이 받는다고 볼 수 있다.

4. 결론 및 토의

본 연구는 유전적인 요인과 환경적인 요인의 영향을 복합적으로 받는 한우의 경제형질 관련 유전자 마커 선별 연구에서 환경적인 요인의 효과는 제거하고 순수한 유전적 요인의 효과만을 이용하여 유전자 마커를 선별하기 위해 환경적인 요인을 보정하여 유전자 마커 선별의 정확도를 높이는 데 목적을 두었다. 사용된 SNP는 한우의 지방산 합성 과정에서 중요한 역할을 하는 LPL 유전자의 49개 마커 중 지방산 조성 및 경제 형질 능력 검정을 통해 나머지에 비해 더 뛰어난 유전자 마커로 선별된 6개 (g.6960 A>T, g.6974 G>A, g.21604 G>A, g.22488 G>T, g.22649 G>A, g.25670 C>T)를 이용하였으며, 경제형질은 한우의 맛과 향, 품질 등에 영향을 미치는 6개 (C18:1, SFA, MUFA, MS, CWT, BFT)를 이용하였다. 먼저 유전적인 요인과 환경적인 요인이 모두 포함된 일반선형모형을 구축한 뒤, 환경적인 요인을 보정하여 경제형질 값을 구하였다. 그리고 보정 전·후의 데이터에 우수 유전자 조합 선별 기법인 MDR을 적용하여 보정 전·후 선별된 우수 유전자 마커 조합과 선별 정확도를 비교하였다. 그 결과, C18:1에서는 보정 전·후 각각 (g.6960 A>T, g.21604 G>A), (g.21604 G>A, g.22488 G>T) 조합이 가장 우수한 조합으로 선별되었고 정확도는 18.75%p 높아졌다. SFA에서는 보정 전·후 각각 (g.6974 G>A, g.22649 G>A), (g.21604 G>A, g.22488 G>T) 조합이 선별되었고 정확도는 36.53%p 높아졌으며, MUFA에서는 보정 전·후 각각 (g.6960 A>T, g.21604 G>A), (g.21604 G>A, g.22488 G>T) 조합이 선별되었고 정확도는 22.79%p 높아졌다. MS는 보정 전·후 각각 (g.21604 G>A, g.22488 G>T), (g.6960 A>T, g.21604 G>A) 조합이 선별되었고 정확도는 31.25%p 높아졌으며, CWT에서는 보정 전·후 각각 (g.6960 A>T, g.22488 G>T), (g.6974 G>A, g.22649 G>A) 조합이 선별되었고 정확도는 32.36%p 높아졌다. 마지막으로 BFT에서는 보정 전·후 각각 (g.6974 G>A, g.22488 G>T), (g.6960 A>T, g.25670 C>T) 조합이 선별되었고 정확도는 23.08%p 높아졌다. 이 중 정확도의 변화폭이 가장 큰 SFA는 다른 경제형질에 비해 상대적으로 유전적인 요인의 영향을 많이 받는다고 볼 수 있으며, 변화폭이 가장 작은 C18:1은 다른 경제형질에 비해 상대적으로 환경적인 요인의 영향을 많이 받는다고 볼 수 있다. 하지만 종합적으로는 환경적인 요인의 영향을 보정하여 순수한 유전적인 요인의 영향으로만 유전자 마커를 선별하는 경우, 모든 경제형질에서 상위 5개의 조합으로 선별된 조합들의 정확도가 높아짐을 확인하였다.

References

- 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.
- Dryden, F. D. and Marchello, J. A. (1970). Influence of total lipid and fatty acid composition upon the palatability of three bovine muscles. *Journal of Animal Sciences*, **31**, 36-41.
- Gotoda, T., Yamada, N., Kawamura, M., Kozaki, K., Mori, N., Ishibashi, S., Shimano, H., Takaku, F., Yazaki, Y. and Furuichi, Y. (1991). Heterogeneous mutations in the human lipoprotein lipase gene in patients with familial lipoprotein lipase deficiency. *Journal of Clinical Investigation*, **88**, 1856-1864.
- Havel, R., Shore, V. and Bier, D. (1970). Role of specific peptides of serum lipoproteins in the action of lipoprotein lipase. *Circulation*, **41**, 111-116.
- Larosa, J. C., Levy, R. I., Herbert, P., Lux, S. E. and Fredrickson, D. S. (1970). A specific apoprotein activator for lipoprotein lipase. *Biochem. Biochemical and Biophysical Research Communications*, **41**, 57.
- Lee, H., Lee, M. H. and Chung, M. S. (1994). Comparison of flavor characteristics and palatability of beef obtained from various breeds. *Korean Journal for Food Science of Animal Resources*, **26**, 500-506.
- 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.

- 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., Lee, J. H. and Lee, Y. W. (2010). Detection of major genotype combination by genotype matrix. *Journal of the Korean Data & Information Science Society*, **21**, 387-395.
- Mandell, I., Buchanan-Smith, G. and C. P. Campbell (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*, **67**, 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, stearyl-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.
- Monson, F., Sanudo, C. and Sierra, I. (2005). Influence of breed and ageing time on sensory meat quality and consumer acceptability in intensively reared beef. *Meat Science*, **71**, 471-479.
- Oh, D. Y. (2014). *Identification of the SNP (single nucleotide polymorphism) within candidate gene associated with fatty composition in Hanwoo*, Yeungnam University, Kyungsan.
- Oh, J. S. and Lee, S. Y. (2014). An extension of multifactor dimensionality reduction method for detecting gene-gene interactions with the survival time. *Journal of the Korean Data & Information Science Society*, **25**, 1057-1067.
- Oh, D. Y., Lee, T. 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., Yeo, J. S. and Lee, J. Y. (2014). Major SNP identification for oleic acid and marbling score which are associated with Korean cattle. *Journal of the Korean Data & Information Science Society*, **25**, 1011-1024.
- 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. *American Journal of Human Genetics*, **69**, 138-147.
- Robins, K., Jense, J., Ryan, K. J., Homco-Ryan, C., McKeith, F. K. and Brewer, M. S. (2003). Consumer attitude towards beef and acceptability of enhanced beef. *Meat Science*, **65**, 721-729.
- Tsuji, S. (2008). Kobe beef as international brand. *Symposium of Gyeongbuk Hanwoo Cluster*
- Vander Wal, P. G., Engel, B. and Hulsegge, B. (1997). Causes for variation in pork quality. *Meat Science*, **46**, 319-322.
- Yeo, J. S., La, B. M., Lee, H. G., Lee, S. W. and Lee, J. Y. (2013). Power and major gene-gene identification of dummy multifactor dimensionality reduction algorithm. *Journal of the Korean Data & Information Science Society*, **24**, 277-287.
- Yoshimura, T. and Namikawa, K. (1983). Influence of breed, sex and anatomical location on lipid and fatty acid composition of bovine subcutaneous fat. *Japanese Journal of Zootechnical Science*, **54**, 97-100.

Proposal of statistical model adjusted environmental factor in genetic research for high quality Hanwoo production

Ji-Eun Jang¹ · Jae-Young Lee² · Dong-Yep Oh³

¹Department of Statistics, Yeungnam University

³Gyeongsangbuk-Do Livestock Research Institute

Received 5 October 2015, revised 27 November 2015, accepted 28 November 2015

Abstract

Individual phenotype is mostly influenced by genetic factors as well as the effects of environmental factors. Therefore, adjustment of environmental factors are needed in order to see more clearly the effects of genetic factors that we are interested in gene screening study related to Hanwoo's economic trait. The purpose of this study is to propose new statistical model that was adjusted environmental factor and identify adjustment effect in a superior gene marker screening study for producing high quality Hanwoo. First, statistical model including both genetic factor and environmental factor establishes and adjusted value of economic trait find by removing effect of environmental factor such as age, breeding farm. Finally, we identify superior gene marker combination and compare accuracy by applying MDR to data of before and after adjustment. Economic trait is used C18:1, SFA, MUFA, MS, CWT, BFT and SNP marker is used 6 markers of LPL that were identified as more excellent SNP marker than the others among 49 markers through fatty acid composition and economic trait performance test.

Keywords: Adjustment, environmental factors, genetic factors, LPL, MDR.

¹ Graduate student, Department of statistics, Yeungnam University, Kyungsan 712-749, Korea.

² Corresponding author: Professor, Department of statistics, Yeungnam University, Kyungsan 712-749, Korea. E-mail: jlee@yu.ac.kr

³ Researcher, Gyeongsangbuk-do Livestock Research Institute, Yeungju 750-871, Korea.