

Research Article

Bayesian Method를 이용한 한우 도체형질의 GEBV 및 Reliability 추정

김성진¹, 손지현¹, 이종재², 최태정³, 이득민⁴, 이정규^{5,6}, 임현태^{5,6}, 구양모^{1*}

¹한국종축개량협회 유전정보사업부, ²중앙대학교 동물생명공학과, ³국립축산과학원 가축개량평가과, ⁴한경대학교 동물생명환경과학과, ⁵경상국립대학교 축산학과, ⁶경상국립대학교 농업생명과학연구원

Estimation of GEBV and Reliability of Hanwoo Carcass Traits using Bayesian Method

Sungjin Kim¹, Ji-Hyun Son¹, Jung-Jae Lee², Tae-Jeong Choi³, Deuk-Min Lee⁴, Jung-Gyu Lee^{5,6}, Hyun-Tae Lim^{5,6}, Yang-Mo Koo^{1*}

¹Genetic & IT Solutions Department, KAIA, Seocho, 06668, Korea

²Chungang University, CAU, Anseong, 17546, Korea

³Animal Genetics & Breeding Division, NIAS, Cheonan, 31000, Korea

⁴Department of Animal Life & Environment Science, Hankyong National University, Anseong, 17579, Korea

⁵Department of Animal Science, GNU, Jinju, 52828, Korea

⁶Institute of Agriculture & Life Science, GNU, Jinju, 52828, Korea

*Corresponding author: Yang-Mo Koo, Genetic & IT Solutions Department, KAIA, Seocho, 06668, Korea, Tel : +82-2-588-9301, E-mail : greatman009@gmail.com

ABSTRACT

This study was conducted to use the Bayes approach, which can be applied in various pre-distribution for genomic selection application in the evaluation of Hanwoo carcass traits, as basic data for genomic selection through genetic accuracy and reliability estimation. As for the data used in the analysis, 8,413 DNA collected by various methods from Hanwoo farms nationwide were extracted. The Axiom Bovine 60k version 3(Affymetrix Inc., 2006) SNP panel was used to generate genomic information. Quality control removed SNPs whose SNPs were on sex chromosomes or whose position on chromosomes was not identified, with a total value of more than 64 markers removed, with an SNP call rate of 95% or less, a Minor Allele Frequency(MAF) of 0.01 or less, and a Hardy-Weinberg Equilibrium(HWE). Among the 8,413 with genetic information, duplicate animals, pedigree errors, and animals that did not match dependent variables were removed, and 6,616 reference groups with genotype information were used for analysis. The KPN used in the analysis was extracted using semen, and the DNA was extracted using tissue samples provided by the Korea Institute for Animal Products Quality Evaluation, and in the case of cows, tail hairs were collected, and DNA was extracted and used for analysis. MBV accuracy was estimated to be 0.25 to 0.55, and accuracy is affected by various factors. In the case of clustering, it is considered that a change in MBV accuracy occurs because the composition of the training group and the verification group varies depending on the number of groups and the clustering method. In addition, MBV accuracy changed according to Bayes B and Bayes C methods assuming different prior distributions. Among the Bayesian methods, the accuracy of MBV in Bayes C was higher than that of Bayes B. This is believed to be because the assumption of prior distribution to estimate the SNP effect is different, and considering that there is no SNP with a great effect, Hanwoo slaughter data is quantitative, so it is considered that the genes involved are related to multiple genes, not a specific one. As genotype information was added to the existing phenotype information, the reliability of the Hanwoo carcass traits GEBV increased by 0.243 ± 0.060 . In particular, in the group with low reliability, the contribution to estimating the GEBV was low, but the increase in reliability was the highest. It was found that the ranking of genetic evaluation on phenotype data changed due to mixing with MBV. Compared to the GEBV, the Two Step method was more correlated than the Single Step method. Therefore, when using genotype

information, the difference between the Single Step and the Two Step is not much different, but when a lot of genotype data is collected in the future, it is considered that there will be a difference in reliability and GEBV.

Key words: GEBV, Reliability, Bayesian Method, MBV, DEBV, Genomic Enhanced-EBV

서론

유전능력평가를 위한 전통적인 방법은 표현형 자료(Phenotype), 혈통정보(Pedigree) 및 BLUP을 활용하여 육종가를 추정하였으며, 이를 기반으로 선발(Selection) 및 도태(Culling)를 실시하여 유전적 개량을 진행해왔다. 최근 다양한 SNP 패널(Affymetrix, Illumina, Neogen-GeneSeek 등)플랫폼의 개발 및 판매로 개량을 위한 유전체 정보가 적용될 수 있었다. 유전체 선발(Genomic selection)은 유전체 정보를 활용하여 유전체육종가를 추정 후 선발에 적용하는 방법으로 전통적인 BLUP방법보다 세대간격 단축, 신뢰도 향상 등의 장점이 있다. 유전체 선발을 위한 유전체 육종가 추정에서 가장 큰 문제는 제한된 개체 정보를 이용하여 수많은 SNP 마커의 효과를 추정하는데 있으며, 이를 해결하기 위해 Meuwissen et al (2001)은 SNP 마커 효과를 균일하게 낮추는 Ridge Regression BLUP (RR-BLUP)과 Bayes A, Bayes B 등 Bayesian 방법이 유전체 선발에 적용이 되었다. 이후 효율적인 유전체 선발을 위한 Least Absolute Angle and Selection Operator (LASSO) (Usai et al., 2009), Elastic Net (Zou and Hastie 2005), Support Vector Regression with a linear kernel (Moser et al. 2009; Xu et al. 2018) 및 Bayesian LASSO, BayesC π 및 BayesD π (de los Campos et al. 2009; Habier et al. 2011) 등과 같은 다양한 방법들이 제안되었다. 또한, Bayes 접근법과 같은 SNP 마커 효과 추정 방법은 BLUP을 이용해 추정된 육종가(또는 유전체육종가)와 SNP 정보를 이용하여 추정된 MBV (Molecular Breeding Value)를 혼합 공식(Kachman, 2008; VanRaden, 2008)을 이용하여 유전체 정보가 더해진 육종가(GE-EBV)를 추정하는 방법과 혈연관계인 NRM (Numerator Relationship Matrix)과 GRM (Genomic Relationship Matrix)를 이용한 H-matrix (IBS, Identity by status)를 이용하는 single-step GBLUP (Misztal et al. 2009) 등이 유전체 선발을 위해 이용되고 있다.

따라서, 본 연구에서는 한우 도체형질의 유전능력평가 시 유전체 선발 적용을 위해 다양한 사전 분포 적용이 가능한 Bayes 접근법을 활용하여 유전체 정확도 및 유전체 육종가 신뢰도 추정을 통하여 한우 도체형질의 유전체 선발 기초자료로 활용하고자 수행하였다.

재료 및 방법

1. 공시자료

1) 종속 변수(Response Variable)

유전체 분석을 위한 종속변수는 Kim et al (2022)에서 추정된 유전모수와 육종가를 활용하였다. 종속변수는 추정 육종가 및 신뢰도를 이용하여 부모의 효과가 제외된 DEBVexcPA, 부모의 효과가 포함된 DEBVincPA 및 육종가(Estimated Breeding Value)를 추정하여 사용하였다. 또한, Deregressed EBV는 개체마다 다른 신뢰도를 가지기 때문에, 이러한 이질적 분산(Heterogeneous variance)을 설명하기 위해 아래와 같은 가중치(Weighting Factor) 공식(Garrick et al., 2009)을 이용하여 모형에 적용하였다.

$$w_i = \frac{1 - h^2}{c + [(1 - r_i^2)/r_i^2]h^2}$$

여기서, r_i^2 은 종속 변수의 신뢰도, h^2 는 유전력, c 는 SNP 정보에 의해 설명되지 않는 유전적 분산의 비율($c = 0.4$, Saatchi et al., 2012), 역회귀육종가(DEBVincPA, DEBVexcPA)로 전환한 후 신뢰도가 0.01 이하인 개체를 제거하였다.

2) 유전체 자료

본 연구를 위하여 전국의 한우농가에서 다양한 방법으로 채취된 8,413두의 DNA를 추출하였다. 유전체 정보를 생성하기 위해 Axiom Bovine 60k version3 (Affymetrix Inc., 2006) SNP패널을 이용하였다. 유전체 분석의 정확성을 높이기 위한 Quality Control은 SNP가 염색체에 상에 있거나 염색체상의 위치(Position)가 확인되지 않은 SNP를 제거하였으며, SNP Call rate가 95% 이하, Minor Allele Frequency (MAF)가 0.01 이하, Hardy-Weinberg 불평형(HWE)을 측정한 Chi-square (χ^2) 값이 95%를 초과한 마커를 제거한 총 64,973개의 SNP 마커를 이용하였다. 유전체 정보를 가진 8,413두 중 중복개체, 혈통오류, 종속변수와 매칭이 되지 않는 개체는 제거하여 유전체 정보가 있는 참조 축군 6,616두를 분석에 이용하였다. 분석에 사용된 보증씨수소는 정액을 이용하여 DNA추출을 진행하였으며, 거세우는 축산물품질평가원에서 제공받은 조직 샘플을 이용하여 DNA추출을 실시하였고, 암소의 경우, 모근을 채취하여 모근에 있는 DNA를 추출하여 분석에 이용하였다.

3) 혈통 및 도체자료

혈통자료의 경우 6,616두와 관련된 5계대의 자료를 한국종축개량협회에서 수집하였다. 분석에 이용된 자료는 총 5,153,168두에 대한 혈통자료를 이용하였다. 도체자료는 5,153,168두 중 도체자료를 가지고 있는 개체 중 결측치 및 이상치를 사전제거 한 후 남은 2,376,865두에 대해서 축산물품질평가원에서 자료를 수집하여 분석에 이용하였다.

이용된 형질은 도체중(Carcass Weight, CW), 등심단면적(Eye Muscle Area, EMA), 등지방두께(Backfat Thickness, BF) 및 근내지방도(Marbling Score, MS)의 4개 형질을 고려하였다.

2. 통계 분석

1) Bayesian Method

유의미한 SNP 마커의 효과를 추정하기 위하여 Bayes 접근법 중 BayesB, BayesC 방법을 이용하였으며, 다음의 통계모형이 SNP 마커 효과를 추정하기 위해 이용되었다.

$$y = \mu + \sum_{j=1}^k Z_{ij} u_j \delta_j + \epsilon_i$$

여기서, y_i 는 종속 변수(Deregressed EBV, EBV), μ 는 전체 집단의 평균, k 는 SNP 마커의 수, Z_{ij} 는 i 번째 개체의 j 번째 마커의 대립유전자 상태(0, 1, 2), u_j 는 SNP 마커 효과이며, δ_j 는 SNP 마커의 모델 상 존재 유무(0 또는 1)이다.

BayesB 방법은 각 SNP 마커마다 다른 분산을 가지게 되며, π 값의 설정에 따라 SNP 마커 효과를 0으로 가정하는 혼합 모델(Mixture model)을 이용하며, SNP 마커 효과에 대한 사전 분포(Prior distribution)는 t-분포(t-distribution)를 사용한다. 또한, BayesC 방법 역시 BayesB 방법과 같이 π 값의 설정에 따라 SNP 마커 효과를 0으로 가정하는 혼합 모델을 이용하며 SNP 마커 효과에 대한 사전 분포는 정규분포(Normal-distribution)를 사용하며, SNP 마커는 등분산(Common variance)을 가지는 가정으로 샘플링 되어진다(Haiber et al., 2011).

모든 절차는 Gensel4R(Garrick 및 Fernando., 2013) 프로그램을 이용하여 추정하였다.

2) 유전체 정확도 (Accuracy of Molecular Breeding Value)

유전체 정확도를 추정하기 위하여 5개의 집단(5-fold) 교차 검증을 이용하였으며, 교차 검증은 K-means, 임의의 군집 방법을 이용하였다(Figure 1).

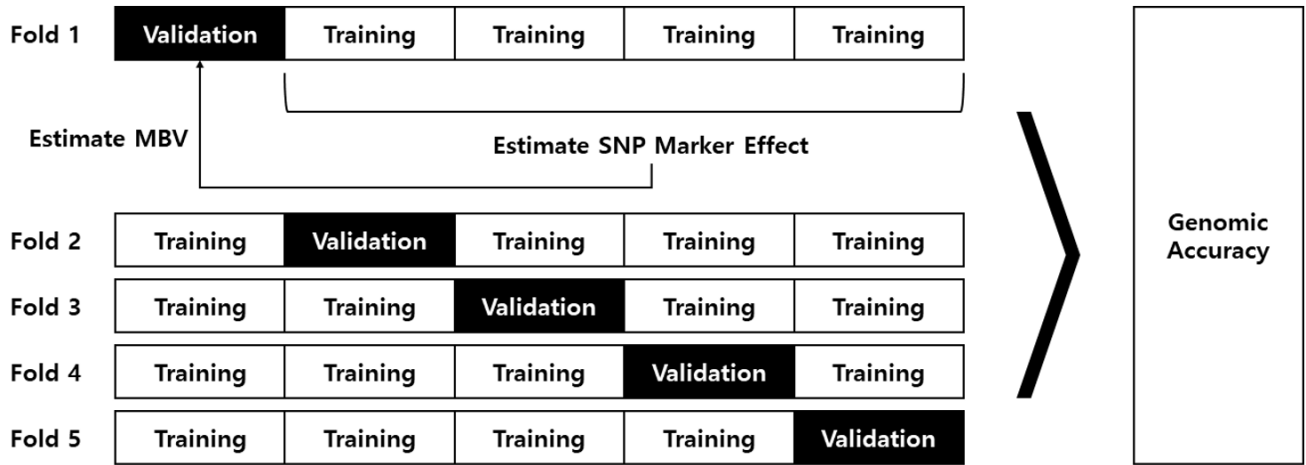


Figure 1. Diagram for cross validation

여기서, K-means 군집 방법은 혈통 정보를 이용하여 혈연 행렬을 추정한 후 이를 거리 행렬(Distance Matrix, D matrix)로 변환하여 집단 내 혈연관계를 최대화, 집단 간 혈연관계를 최소화하였다. D matrix에 대한 공식은 다음과 같다.

$$d_{ij} = 1 - \frac{a_{ij}}{\sqrt{a_{ii} \cdot a_{jj}}}$$

여기서, d_{ij} 는 i 번째 개체와 j 번째 개체 사이의 거리, a_{ij} 는 i 번째 개체와 j 번째 개체의 상가적 유전 관계(Additive genetic relationship), $a_{ii}(a_{jj})$ 는 개체의 NRM의 대각 성분(개체의 혈연계수)이며, 이는 근친 효과를 제거하고 D matrix의 대각 성분을 0으로 만든다.

군집화를 통해 참조 집단(Reference population)은 훈련 집단(Training set) 및 검증 집단 (Validation set)으로 나뉘며, 검증 집단 내 개체의 MBV는 훈련 집단에서 추정된 모든 SNP 효과의 사후 평균(posterior mean) 합으로 계산되어지며, 공식은 다음과 같다.

$$MBV_i = \sum_{j=1}^k z_{ij} \hat{u}_j$$

여기서, MBV_i 는 검증 집단에서 i 번째 개체의 MBV, z_{ij} 는 i 번째 개체의 j 번째 SNP의 유전자형(Genotype, i.e., AA=0, AB=1, BB=2), $\hat{u}_j$ 는 훈련 집단의 McMC 샘플링에서 5,000번의 번인구간 이후의 SNP 마커 효과의 사후 평균(Posterior mean)이다.

각 검증 집단에서 추정된 유전체 육종가(MBV)와 종속변수를 이변량 개체 모형(Bivariate animal model)을 이용한 유전상관을 유전체 정확도(Genomic accuracy)로 정의하였으며, 종속변수(T)와 MBV의 유전상관($r_{g(T, MBV)}$)은 다음과 모형과 같다.

$$\begin{bmatrix} DEBV \\ MBV \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

여기서, β_1 및 β_2 는 고정효과의 벡터(β_1 은 형질의 평균, β_2 는 군집화 그룹 수에 대한 효과), α_1 은 형질, α_2 는 MBV의 임의 상가적 유전 효과의 벡터이며, $\text{Var}(\alpha_1) = G\sigma_{a1}^2$, $\text{Var}(\alpha_2) = G\sigma_{a2}^2$ 및 $\text{Cov}(\alpha_1, \alpha_2) = G\sigma_{a1a2}$ 이며, σ_{a1}^2 는 형질의 상가적 유전 분산, σ_{a2}^2 는 MBV의 상가적 유전분산($\sigma_{a2}^2/\sigma_{a1}^2$ 는 마커에 의해 설명되는 형질의 상가적 유전분산)이며, σ_{a1a2} 는 형질과 MBV의 유전상관이다. G 는 같은 그룹에 있는 개체사이의 혈통 기반의 NRM에 기초한 A의 0이 아닌 요소로 이루어져 있으나, 다른 그룹에 있는 개체들 사이의 공분산은 다음과 같이 0이 된다.

$$G = \begin{bmatrix} A_{11} & 0 & 0 & \cdots & \cdots & 0 \\ 0 & A_{22} & 0 & \cdots & \cdots & 0 \\ 0 & 0 & A_{33} & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & & 0 \\ \vdots & \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & A_{fold} \end{bmatrix}$$

이 접근법은 집단 개수 모두에 걸쳐 정보를 효과적으로 모아 유전모수를 추정한다. ε_1 및 ε_2 는 두 형질에 대해 상호 관련이 없는 임의의 잔차 효과의 벡터이며, $\text{Var}(\varepsilon_1) = I\sigma_{\varepsilon_1}^2$ 및 $\text{Var}(\varepsilon_2) = W\sigma_{\varepsilon_2}^2$, $\text{cov}(\varepsilon_1, \varepsilon_2) = 0$ 이며, 여기서 I 는 항등 행렬(identity matrix)이며, W 는 종속 변수의 신뢰도에 기초한 가중치를 포함하는 대각행렬이다(SNP 마커 효과의 추정에 사용된 것과 동일한 가중치). 이변량 개체 모형을 활용하기 위해 ASREML 4.1(Gilmour et al., 2015)을 이용하였다.

3) Genomic Enhanced-EBV 추정

Kachman(2008)에 의해 제시된 “Correlated traits” 방법을 이용하여 BLUP 방법에 의해 추정된 육종가 및 SNP 마커 효과의 합으로 추정된 MBV간 혼합 공식을 통하여 GE-EBV를 추정하였으며, 공식은 다음과 같다.

$$GE-EBV = \frac{(1-R_i^2)}{(1-r_g^2 R_i^2)} MBV_i + \frac{(1-r_g^2)}{(1-r_g^2 R_i^2)} EBV_i$$

$$Reliability = \frac{(1-R_i^2)r_i^2 + (1-r_g^2)R_i^2}{(1-r_g^2 R_i^2)}$$

여기서, R_i^2 는 i 번째 개체의 신뢰도, r_g^2 는 유전체 정확도, MBV_i 는 i 번째 개체의 유전체 육종가이며, EBV_i 는 BLUP에 의해 추정된 i 번째 개체의 육종가이다.

결과 및 고찰

1. 군집화 결과

군집화 방법으로 K-means, 임의의 방법 등 여러 가지 방법들이 있는데 이전 연구에서 K-means와 임의의 군집화를 이용한 연구와 일치한 결과를 나타냈다는 자료가(Campos et al (2018); Boddhireddy et al (2014)) 있어 K-means 방법을 이용한 5-군집 방법을 이용하여 MBV의 정확도와 유전체 정보가 더해진 육종가(GE-EBV)를 추정하였다. 결과는 각각 Table 1에 나타났다. K-means clustering을 이용한 방법은 Random clustering을 이용한 그룹에 비하여 그룹 내 혈연관계가 높게 나타났으며, 이는 K-means 클러스터링은 NRM(Numeric Relationship Matrix)을 이용함으로써 혈연관계가 높은 개체가 한 그룹에 편성되기 때문인 것으로 사료된다.(Campos et al (2018); Boddhireddy et al (2014)). K-means 클러스터링은 훈련 집단과 검증 집단 사이의 비율(그룹 내/그룹 간)이 타 논문보다 높게 나타났는데(Lee et al, 2017), 이는 한 그룹에 편성되는 개체들이 더 높은 혈연관계를 가지는 개체로 편성되기 때문인 것으로 사료되며, 개체 수와 비교한다면 정액의 치우침이 발생했다고 사료된다. 또한, 3그룹에서 낮은 개체 수를 보였는데 이는 기존 집단과 혈연관계가 적은 외부 유전자원이 한 그룹에 편성되었기 때문인 것으로 사료된다.

2.가중치

Figure 2는 6,619두의 유전체 자료에 대한 가중치 분포를 나타냈다. 0.4부터 2.3까지 다양한 가중치 분포를 나타내고 있다. 신뢰도가 높을수록 가중치 값이 상승하는데 이를 증명하듯, 1.0부터 2.3까지의 가중치를 받은 개체는 전부 한우 씨수소였으며, 0.6에서 0.8 사이를 받은 씨수소가 있는 것도 확인을 했지만 이들은 최근에 보증 종모우 정액으로 사용된 씨수소로 후대가 많지 않아 신뢰도가 낮은 것으로 나타났다. 0.4에서 0.6으로 많이 나타나는 개체들은 주로 암소와 거세우들 이었으며, 거세우들은 0.5~0.6의 값을 나타냈고, 암소는 0.4~1.5로 거세우보다 다양한 경향을 나타냈다. 따라서, 정확도를 이용한 가중치는 암소 및 거세우 자료가 후대 검정을 진행한 씨수소들 보다 정확도가 떨어지므로 이를 보정하기 위하여 가중치를 이용하여 역회귀육중가를 구하였다.

Table 1. Comparison of relationships among animals within and across cluster in 5-fold cross validation

Group name	N	inbreC		Within Group $\alpha_{\max}$		Between group $\alpha_{\max}$		Within Group AverageR		Between Group AverageR		ratio
		Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	
1	2,165	0.04	0.016	0.393	0.117	0.380	0.126	0.018	0.007	0.017	0.007	1.06
2	1,258	0.011	0.015	0.439	0.110	0.383	0.127	0.044	0.010	0.024	0.005	1.83
3	591	0.010	0.018	0.435	0.111	0.377	0.155	0.086	0.018	0.023	0.006	3.74
4	1,395	0.008	0.019	0.482	0.081	0.365	0.141	0.067	0.017	0.022	0.006	3.05
5	1,207	0.012	0.022	0.483	0.088	0.392	0.141	0.078	0.019	0.024	0.006	3.25
Avg.		0.016		0.446		0.379		0.059		0.022		2.586

* inbreC = Inbreeding coefficients within clusters; within Group $\alpha_{\max}$ = the average of $\alpha_{\max}$ values (the maximum value of relationships for each individual) within clusters; between Group $\alpha_{\max}$ = the average of $\alpha_{\max}$ values between clusters (training and validation); within Group AverageR= the average of α_{ij} values (relationships) within clusters; between Group AverageR= the average of α_{ij} values between clusters (training and validation); ratio = the ratio between “Within Group AverageR” and “Between Group AverageR”

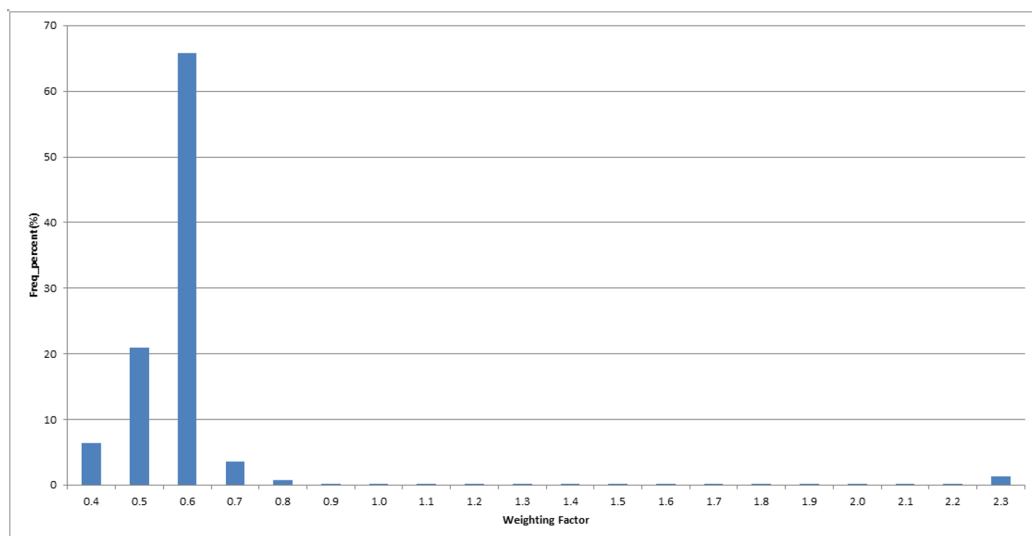


Figure 2. Distribution of genomic data according to weighting factor

3. MBV 정확도

Table 2에서는 Axiom 플랫폼을 이용하여 종속변수인 역회귀육종가와 육종가들을 가지고 Bayesian B와 C 방법으로 분포가 0.99일 경우에 대하여 정확도를 나타냈다. 역회귀육종가와 육종가 사이의 정확도 차이를 보면 Bayes B 방법은 근내지방도에서 12%의 차이를 보였으며 도체중에서 26%의 큰 차이를 보였다. Bayes C방법에서도 근내지방도에서 12%의 차이를 보였으며 도체중과 등심단면적에서 26%의 정확도 차이를 보였다. Lee et al. (2019)을 보면 일반적으로 Bayesian방법에서 역회귀육종가와 육종가 사이의 정확도 차이는 약 6%이하정도의 차이를 보였는데 결과가 많이 차이나는 이유는 첫 번째로 유전체 자료의 참조집단이 부족했으며, 씨수소의 유전체 자료가 암소와 거세우보다 월등히 적었고 무엇보다도 Lee et al. (2019)의 논문에서 사용한 자료는 씨수소의 자료만 이용하였지만 이번 연구에서는 암소와 씨수소의 후대측인 거세우의 자료가 더 많이 사용됨으로 인해서 정확도가 기존 연구보다 많이 떨어진다는 것이라고 사료된다. 추후 유전체 자료가 늘어나고, 씨수소 유전체 자료가 추가된다면 정확도가 상승하고, 농가 또는 지역단위 암소평가에 더 정확한 자료가 될 것이라고 사료된다. Table 3은 앞서 육종가가 유전체 육종가보다 정확도가 떨어진다는 결과가 나왔으므로 역회귀육종가와 육종가인 종속변수와 Bayesian 방법에서 편의가 어느 정도 나왔는지 확인하기 위하여 회귀값을 산출하였다. 1의 값으로부터 어느 정도 떨어져 있는지를 나타낸 것으로 Bayes B방법의 역회귀육종가값이 편의가 가장 많이 발생한 것으로 나타났으며, Bayes C방법의 육종가 값이 편의가 가장 적게 발생하였다. Song et al. (2020)에서는 정확도가 더 높게 추정이 되었지만 편의가 많이 발생한 반면, 이번 연구에서는 정확도가 낮게 추정되었지만 편의가 상대적으로 적게 발생하였다. 반면, Bayes B방법보다 Bayes C방법이, 육종가값이 역회귀육종가보다 편의가 적게 발생하였다는 점은 동일한 것으로 나타났다. 앞선 결과와 마찬가지로 유전체 자료의 참조집단의 크기가 작으며, 암소 및 거세우와 혈통으로 연결되어 있는 씨수소의 자료가 추가된다면 농가 또는 지역단위 암소평가에 이용할 수 있는 자료로 이용할 수 있을 것이라 사료된다.

Table 2. Comparisons of the genomic prediction accuracy with standard error among Axiom platform according to the Bayesian methods ($\pi=0.99$) in each response variable

Traits	Response variables	Bayes B	Bayes C
BF	DEBVincPA	0.28(0.03)	0.28(0.03)
	EBVs	0.45(0.03)	0.46(0.03)
CW	DEBVincPA	0.29(0.06)	0.30(0.05)
	EBVs	0.55(0.08)	0.56(0.08)
EMA	DEBVincPA	0.25(0.08)	0.25(0.09)
	EBVs	0.50(0.04)	0.51(0.04)
MS	DEBVincPA	0.35(0.02)	0.36(0.02)
	EBVs	0.47(0.03)	0.48(0.03)

* CW: Carcass Weight; EMA: eyes muscle area; BF: back fat thickness; MS: marbling score. DEBVincPA: DEBV-including parent average; EBVs: Estimate Breeding Values; Bayes : Bayesian Method average(standard deviation)

Table 3. Regression of the genomic prediction accuracy with standard error among Axiom platform according to the Bayesian methods ($\pi=0.99$) in each response variable

Traits	Response variables	Bayes B	Bayes C
BF	DEBVincPA	1.15(0.13)	1.11(0.12)
	EBVs	0.83(0.05)	0.91(0.06)
CW	DEBVincPA	1.19(0.25)	1.27(0.24)
	EBVs	0.97(0.27)	1.02(0.26)
EMA	DEBVincPA	1.25(0.3)	1.32(0.36)
	EBVs	0.91(0.03)	0.98(0.03)
MS	DEBVincPA	1.28(0.08)	1.25(0.07)
	EBVs	0.88(0.05)	0.94(0.05)

* CW: Carcass Weight; EMA: eyes muscle area; BF: back fat thickness; MS: marbling score. DEBVincPA: DEBV-including parent average; EBVs: Estimate Breeding Values; Bayes : Bayesian Method average(standard deviation)

4. 육종가 순위변화

Table 4는 ssGBLUP 방법으로 구한 유전체육종가와 유전체 정보가 포함된 육종가에 대하여 육종가의 순위를 나타냈다. 개체의 상태를 나타내는 지표(Identity by status, IBS)인 MBV와 종속변수의 Blending에 의해 GE-EBV를 구하였다. GE-EBV의 순위상관은 Bayes C ($\pi=0.99$, Response-variable=EBV, Cross Validation=5-fold Clustering)와 Single-Step 방법으로 추정된 유전체육종가의 상관을 나타냈다. 육종가 순위의 변화를 통해 더 높은 신뢰도를 바탕으로 우수한 능력의 한우 선발이 가능할 것으로 사료된다. EBV와 GEBV간의 상관관계를 형질별로 분석하여 Figure 3-6에 나타내었다. Single Step 방법에서는 도체중, 등심단면적, 등지방두께, 근내지방도순으로 0.89, 0.92, 0.91, 0.94의 상관을 보였으며, Two Step 방법에서는 0.93, 0.94, 0.93, 0.93으로 나타났다. 도체중, 등심단면적, 등지방두께에서는 Two Step 방식이 EBV와 비슷하게 나타났으나 근내지방도에서는 Single Step 방식이 상관이 더 높았다. 하지만 이 차이는 극히 미미하므로 큰 차이를 보이지 않는다고 사료된다.

Table 4. Rank correlation between corrected EBV(Estimated Breeding Value) and Single-Step GEBV(Genomic EBV) and Bayes GE-EBV(Bayesian C($\pi = 0.99$), Response variables = EBVs)

Traits	EBV*Single-Step ^(a)	EBV*Bayes C ^(b)	Difference ^(b-a)
CW	0.89	0.93	0.03
EMA	0.92	0.94	0.02
BF	0.91	0.93	0.02
MS	0.94	0.93	-0.01

CW: Carcass weight (kg), EMA: Eye muscle area (cm²), BF: Backfat thickness (mm), MS: Marbling score.

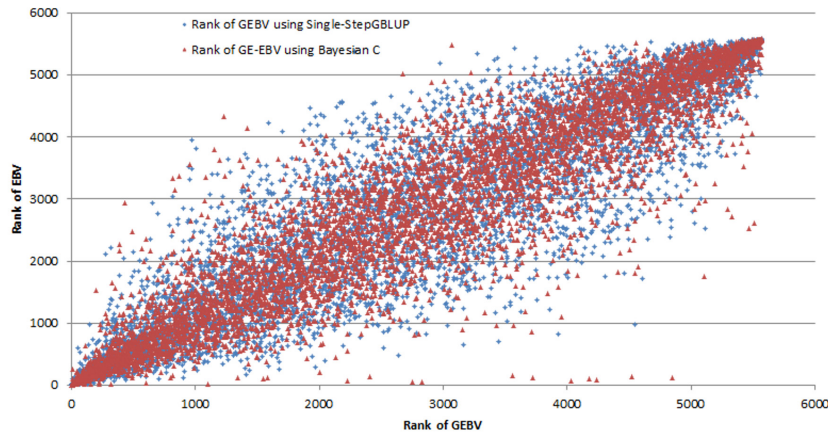


Figure 3. Compare with EBV rank and GEBV rank of carcass weight

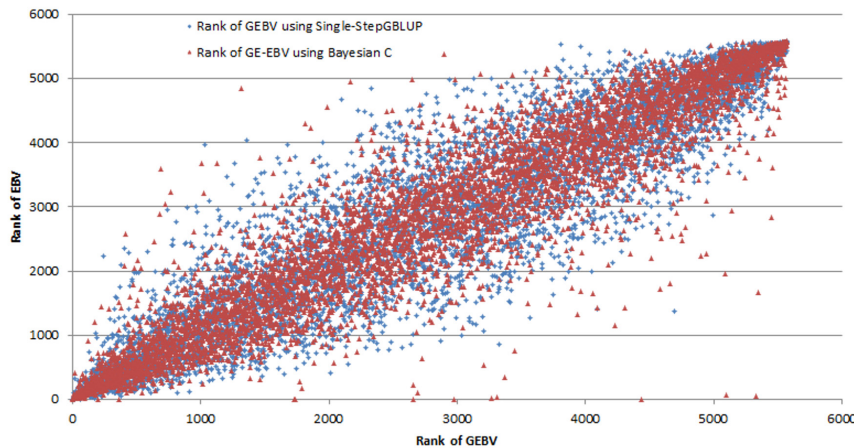


Figure 4. Compare with corrected phenotype rank and EBV rank of eye muscle area

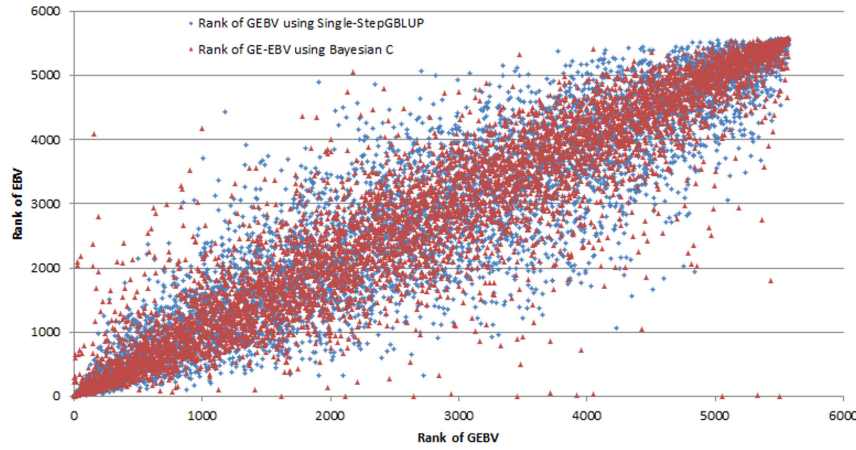


Figure 5. Compare with corrected phenotype rank and EBV rank of backfat thickness

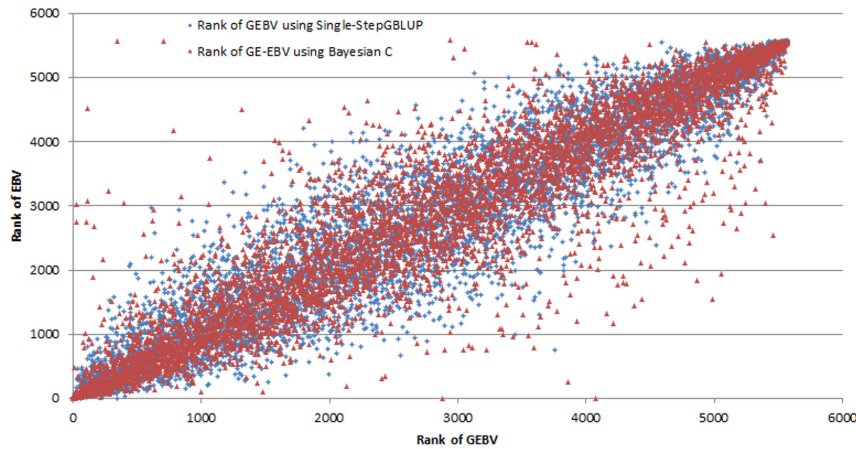


Figure 6. Compare with corrected phenotype rank and EBV rank of marbling score

요약

본 연구에서는 한우 도체형질의 유전능력평가 시 유전체 선발 적용을 위해 다양한 사전 분포 적용이 가능한 Bayes 접근법을 활용하여 유전체 정확도 및 유전체 육종가 신뢰도 추정을 통하여 한우 도체형질의 유전체 선발 기초자료로 활용하고자 수행하였다. 분석에 이용된 자료는 전국의 한우농가에서 다양한 방법으로 채취된 8,413두의 DNA를 추출하였다. 유전체 정보를 생성하기 위해 Axiom Bovine 60k version3(Affymetrix Inc., 2006) SNP패널을 이용하였다. 유전체 분석의 정확성을 높이기 위한 Quality Control은 SNP가 성염색체에 상에 있거나 염색체상의 위치(Position)가 확인되지 않은 SNP를 제거하였으며, SNP Call rate가 95% 이하, Minor Allele Frequency(MAF)가 0.01 이하, Hardy-Weinberg 불평형(HWE)을 측정한 Chi-square(χ^2) 값이 95%를 초과한 마커를 제거한 총 64,973개의 SNP 마커를 이용하였다. 유전체 정보를 가진 8,413두 중 중복개체, 혈통오류, 종속변수와 매칭이 되지 않는 개체는 제거하여 유전체 정보가 있는 참조 축군 6,616두를 분석에 이용하였다. 분석에 사용된 보증씨수소는 정액을 이용하여 DNA추출을 진행하였으며, 거세우는 축산물품질평가원에서 제공받은 조직 샘플을 이용하여 DNA추출을 실시하였고, 암소의 경우, 모근을 채취하여 모근에 있는 DNA를 추출하여 분석에 이용하였다. MBV 정확도는 0.25~0.55로 추정되었으며, 정확도는 다양한 요인에 의해 영향을 받는다. 군집화의 경우 집단의 수, 군집화 방법에 의해 훈련 집단 및 검증 집단의 구성이 달라지기 때문에 MBV 정확도의 변화가 발생하는 것으로 사료된다. 또한, 서로 다른 사전 분포를 가정하는 Bayes B, Bayes C 방법에 따라 MBV 정확도가 변하였다. Bayesian 방법 중 Bayes C에서 MBV의 정확도가 Bayes B 보다 높은 결과를 나타냈다. 이는 SNP 효과를 추정하기 위한 사전 정보(Prior distribution)의 가정이 다르기 때문이라고 사료되며, 큰 효과를 가지는 SNP가 없는 것으로 보아 한우 도축자료는 양적형질이기 때문에 관여하는 유전자가 특정 하나의 유전자가 아닌 다수의 유전자가 관련이 있는 것으로 사료된다. 기존 표현형 정보에 유전체 정보를 추가함에 따라 한우 도체형질 육종가의 신뢰도는 0.243 ± 0.060 상승하였다. 특히, 신뢰도가 낮은 그룹에서는 유전체 육종가 추정을 위한 기여도는 낮게 나타났으나 신뢰도 상승 폭은 가장 높았다. 또한, 표현형 자료에 대한 유전능력평가의 순위는 MBV와의 혼합으로 인해 유전체

육종가의 순위가 변화하는 것으로 나타났다. 육종가와 비교하였을 때 Single Step 방법보다 Two Step 방법이 더 상관이 높게 나왔다. 따라서, 유전체 정보를 활용하는 경우 Single Step과 Two Step 간의 차이는 크게 다르지 않지만 추후 유전체 자료가 많이 모이는 경우, 유전체육종가 뿐만 아니라 신뢰도도 차이가 날 것으로 사료된다.

검색어 : GEBV, Reliability, Bayesian Method, MBV, DEBV, Genomic Enhanced-EBV

REFERENCES

- Bodhiredy P, Kelly MJ, Northcutt S, Prayaga KC, Rumph J, DeNise S. 2014. Genomic predictions in Angus cattle: comparisons of sample size, response variables, and clustering methods for cross-validation. *J. Anim Sci.* 92(2):485-497.
- Campos GS, Reimann FA, Cardoso LL, Ferreira CER, Junqueira VS, Schmidt PI, Neto JB, Yokoo MJ, Sollero BP, Boligon AA, Cardoso FF. 2018. Genomic prediction using different estimation methodology, blending and cross-validation techniques for growth traits and visual scores in Hereford and Braford cattle. *J. Anim Sci.* 96(7):2579-2595.
- de los Campos G, Naya H, Gianola D, Crossa J, Legarra A, Manfredi E, Weigel K, Cotes JM. 2009. Predicting Quantitative Traits With Regression Models for Dense Molecular Markers and Pedigree. *Genetics.* 182(1):375-385.
- Garrick DJ, Fernando RL. 2013. Implementing a QTL detection study (GWAS) using genomic prediction methodology. *Methods in molecular biology (Clifton, NJ).* 1019:275-298.
- Garrick DJ, Taylor JF, Fernando RL. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet Sel Evol.* 41:55.
- Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J., & Thompson, R. 2015. ASReml User Guide Release 4.1 Functional Specification, VSN International Ltd, Hemel Hempstead, HP1 1ES, UK www.vsn.co.uk
- Habier D, Fernando RL, Kizilkaya K, Garrick DJ. 2011. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics.* 12(1):186.
- Kachman S. 2008. Incorporation of marker scores into national cattle evaluations. *Proceedings of the 9th Genetic Prediction Workshop.* December 10 to 11, 2008. Kansas City (MO). Beef Improvement Federation. 92-98.
- Kim S.J., Choi T.J., Son J.H., Lee D.M., Lee J.J., Lee J.G., Lim H.T., and Koo Y.M. 2022. Estimation of Genomic Estimated Breeding Value(GEBV) and Reliability for Hanwoo Carcass Traits using ssGBLUP. *Journal of Animal Breeding and Genomics.* 6(3):57-72.
- Lee J., Kachman SD, Spangler ML. 2017. The impact of training strategies on the accuracy of genomic predictors in United States Red Angus cattle. *J. Anim Sci.* 95(8):3406-14.
- Lee J., Lee S., Park J.E., Moon S.H., Choi S.W., Go G.W., Lim D., and Kim J.M. 2019. Genome-wide association study and genomic predictions for exterior traits in Yorkshire pigs. *J. Anim Sci.* 97(7):2793-2802.
- Meuwissen TH, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 157(4):1819-1829.
- Misztal I, Legarra A, Aguilar I. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J. Dairy Sci.* 92(9):4648-4655.
- Moser G, Tier B, Crump RE, Khatkar MS, Raadsma HW. 2009. A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. *Genet Sel Evol.* 41:56.
- Saatchi M, Schnabel RD, Rolf MM, Taylor JF, Garrick DJ. 2012. Accuracy of direct genomic breeding values for nationally evaluated traits in US Limousin and Simmental beef cattle. *Genet Sel Evol.* 44:38.
- Song HJ. 2020. Study for variation of birth weight on piglets within litter. Chungnam National University. Daejeon. Korea.
- Usai MG, Goddard ME, Hayes BJ. 2009. LASSO with cross-validation for genomic selection. *Genet Res (Camb).* 91(6):427-436.
- VanRaden PM. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91(11):4414-4423.
- Xu Y, Wang X, Ding X, Zheng X, Yang Z, Xu C. 2018. Genomic selection of agronomic traits in hybrid rice using an NCII population. *Rice (New York, NY).* 11(1):32.
- Zou H, and Hastie T. 2005. Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society Series B (Statistical Methodology).* 67(2):301-320.