

유전체 관계행렬 구성에 따른 Landrace 순종돈의 육종가 비교

이준호¹ · 조광현² · 조충일² · 박경도¹ · 이득환^{1*}

¹국립한경대학교 동물생명환경과, ²농촌진흥청 국립축산과학원

Comparison of Breeding Value by Establishment of Genomic Relationship Matrix in Pure Landrace Population

Joon-Ho Lee¹, Kwang-Hyun Cho², Chung-Il Cho², Kyung-Do Park¹ and Deuk Hwan Lee^{1*}

¹Department of Animal Life Resources, Hankyong National University, Anseong-si, Gyeonggi-do, 456-749, Korea,

²National Institute of Animal Science, R.D.A., Cheonan-si, Chungnam, 330-801, Korea

ABSTRACT

Genomic relationship matrix (GRM) was constructed using whole genome SNP markers of swine and genomic breeding value was estimated by substitution of the numerator relationship matrix (NRM) based on pedigree information to GRM. Genotypes of 40,706 SNP markers from 448 pure Landrace pigs were used in this study and five kinds of GRM construction methods, G05, GMF, GOF, GOF* and GN, were compared with each other and with NRM. Coefficients of GOF considering each of observed allele frequencies showed the lowest deviation with coefficients of NRM and as coefficients of GMF considering the average minor allele frequency showed huge deviation from coefficients of NRM, movement of mean was expected by methods of allele frequency consideration. All GRM construction methods, except for GOF*, showed normally distributed Mendelian sampling. As the result of breeding value (BV) estimation for days to 90 kg (D90KG) and average back-fat thickness (ABF) using NRM and GRM, correlation between BV of NRM and GRM was the highest by GOF and as genetic variance was overestimated by GOF*, it was confirmed that scale of GRM is closely related with estimation of genetic variance. With the same amount of phenotype information, accuracy of BV based on genomic information was higher than BV based on pedigree information and these symptoms were more obvious for ABF than D90KG. Genetic evaluation of animal using relationship matrix by genomic information could be useful when there is lack of phenotype or relationship and prediction of BV for young animals without phenotype.

(Key words : Genomic selection, Genomic relationship matrix, Genomic breeding value, SNP markers, Landrace)

서 론

고밀도 유전체 정보를 활용하는 유전체 선발 방법이 제시된 이후 (Meuwissen 등, 2001) SNPchip이 개발되어 단기간에 다량의 유전체 정보를 수집하는 것이 가능해지면서 세계 각국에서 유전체 선발기술을 종축개량에 활용하기 위한 연구가 활발하게 진행되고 있다 (Loberg 등, 2009; Hayes 등, 2009). 기존의 유전체 연구가 기능탐색연구로 진행되어 산업 적용이 어려웠던 반면, 유전체 선발 기술은 형질발현에 관련된 모든 유전자의 기능을 일일이 탐색하는 것이 아니라 유전체 전장에 골고루 분포하는 유전변이들이 형질 발현에 관계된 유전자와 연관불평형이 발생하는 것을 활용하므로 무한소모형 (Infinitesimal model)에 가까운 양적형질좌위에 대한 유전변이를 잘 설명할 수 있어 실제적인 종축 선발에 잘 적용될 수 있다 (Hayes 등, 2008; Meuwissen 등, 2001).

유전체 선발 기술은 다량의 SNP 표지인자들의 효과를 추정하고 이를 이용하여 육종가를 계산하는 방법 (Meuwissen 등, 2001; Gianola, 2009; Habier 등, 2011; Legarra 등, 2008)과 유전체 관계행렬을 구성하여 기존의 혈연계수행렬에 대치시키는 방법 (VanRaden, 2008; Misztal 등, 2009)으로 구분된다. 표지인자의 효과를 추정하는 방법은 양적형질좌위 탐색에 활용될 수 있으며 (Sun, 2011), 유전체 관계행렬을 구성하는 방법은 다형질 모형 등의 기존의 혈통정보를 이용한 육종가 추정모형 적용이 용이하다는 각각의 장점이 존재한다 (Zhang 등, 2007). 그리고 유전체 정보를 이용한 혈연계수는 다량의 유전변이 구성의 평균 유사도를 이용하기 때문에 보다 정밀한 유전관계를 측정할 수 있으며, 추정된 육종가의 정확도를 향상시킨다 (Habier 등, 2007; Forni 등, 2011).

본 연구의 목적은 유전체 혈연계수행렬 구성의 방법론 차이가 혈연계수, 유전모수 및 육종가 추정에 미치는 영향을 확인하고 추정

* Corresponding author : Deuk Hwan Lee, Department of Animal Life Resources, Hankyong National University, Anseong-si, Gyeonggi-do, 456-749, Korea, Tel: +82-31- 670-5091, E-mail: dhlee@hknu.ac.kr

된 육종가의 정확도를 비교하여 유전체 선발 효과를 구명하는데 있다.

재료 및 방법

1. 공시재료

랜드레이스 GGP 순종 집단의 웅돈과 모든 490두의 혈액을 채취하여 DNA를 추출하고 Porcine SNP60 chip을 이용하여 62,163개 SNP 표지인자의 유전자형을 수집하였다. 분석형질은 90 kg 도달일령 (Days to 90 kg; D90KG)과 3개 부위에서 초음파로 측정된 평균등지방두께 (Average Back-fat Thickness; ABF)로 25,325개의 혈통자료와 24,411개의 표현형 자료를 이용하였다. GGP 랜드레이스 순종돈 집단의 평균등지방두께는 1.26 cm으로, 생후 90 kg까지 도달하기까지 약 146일이 소요되는 것으로 나타났다 (Table 1).

사전 분석단계에서 성별 (S), 주간배치 (B) 효과는 모든 형질에서 유의적으로 영향을 미치는 것으로 나타났으며, 모든 산차 (MP) 효과의 경우 90 kg 도달일령에 대하여 높은 통계적 유의성이 나타난 반면 평균등지방두께에서는 낮은 수준의 유의성을 나타내었다.

2. 유전체 자료 처리

Porcine SNP60 chip 분석결과로 얻어진 각 표지인자별 유전자형의 유효성 관리 (Quality Control)를 위하여 결측률이 10% 이상인 표지인자, 소수대립유전자빈도 (Minor Allele Frequency)가 0.01 이하인 표지인자, 하디-바인버그 불평형 검사 (Hardy-Weinberg Disequilibrium test)의 χ^2 값이 23.93 ($p < 0.000001$, $df = 1$) 이상인 표지인자와 모두 이형으로 나타난 표지인자는 이상치로 간주하여 제거하였으며, 성염색체 및 염색체 위치정보가 없는 표지인자 역시 제거하였다. 따라서 이용된 표지인자는 총 40,706개였으며, 잔여 결측 유전자형은 fastPHASE 프로그램을 이용하여 대체하였다 (Stephens 등, 2003).

3. 통계적 모형

분석형질들에 대한 유전모수 및 육종가를 추정하기 위한 모형은 다음과 같다.

$$ABF = S_i + MP_j + B_k + \beta BW/age_{ijkl} + a_{ijkl} + e_{ijkl}$$

$$D90KG = S_i + MP_j + B_k + a_{ijkl} + e_{ijkl}$$

위에서, ABF는 3개부위에서 초음파로 측정된 등지방 두께의 평균, D90KG는 90 kg 도달일령, S는 성별효과, MP는 모든의 산차효과, B는 주간배치효과, BW/age는 측정일령 대비 체중의 공변량, β 는 공변량에 대한 선형회귀계수이며, a는 개체의 상가적 유전효과, e는 임의오차이다. 유전체 육종가 및 유전모수는 혼합모형방정식의 혈연계수 역행렬 부분을 각 유전체 관계행렬의 역행렬로 대체시켜 추정하였으며, EM-REML 방법을 이용하였다. 추정된 육종가의 정확도(r)는 $r = \sqrt{1 - d_1 a}$ (d_1 는 혼합모형방정식 역행렬 중 i번째 개체 부분의 대각원소, a는 σ_e^2 / σ_a^2 로 개체의 상가적 유전분산 대비 오차분산의 비율) 공식으로 추정하였으며, REMLF90 (Misztal, 2011)을 참조하여 GFortran 언어로 직접 개발한 분석프로그램을 이용하였다.

4. 유전체 관계행렬의 구성

유전체 전장에 분포하는 다량의 단일염기다형 유전자형자료를 이용하여 혈연계수를 추정하는 유전체 관계행렬은 Forni 등 (2011)이 제시한 G05, GMF, GOF, GOF* 및 GN의 5가지 방법을 이용하였다. 개체의 수를 n, 표지인자의 수를 m이라고 하면, 행의 크기는 n이고 열의 크기는 m인 M의 행렬을 구성한 뒤, M 행렬의 각 원소는 소수 대립유전자 (Minor allele)의 동형접합체는 -1, 다수 대립유전자 (Major allele)의 동형접합체는 1 그리고 이형접합체는 0으로 유전자형을 수량화 하였다 (VanRaden, 2008). 일반적인 유전체 관계행렬 (Gr)은 MM' 행렬에서 아래의 수식과 같이 유전자빈도 및 분산으로 표준화하는 과정을 거치는데 Forni 등은 아래와 같이 5가지 방법을 제시하고 있다.

$$Gr = \frac{(M-P)(M-P)'}{2 \sum_{j=1}^m p_j (1-p_j)} \quad (1)$$

(1) G05

G05 방법은 위의 수식(1)과 같이 M 행렬에서 P 행렬을 빼는 방식이 적용된다. P 행렬은 아래와 같이 j번째 표지인자의 유전자 빈도와 기준 빈도 0.5의 편차값에 2를 곱해준 것이며, 이러한 경우 대립유전자의 평균값 (mean value of the allele effect)이 0으로 표준화된다 (VanRaden, 2008). G05 방법으로 생성되는 유전체 관

Table 1. Basic statistics of average backfat thickness and days to 90 kg

Trait	Record	Mean	STD	Min	Max
ABF ¹⁾	24,411	1.26	0.23	0.50	2.82
D90KG ²⁾	24,411	145.72	10.17	114.60	188.80

¹⁾ Average Back-fat Thickness, ²⁾ Days to 90 kg.

계행렬(Gr)은 $(M-P)(M-P)'$ 행렬을 유전체 관계행렬의 척도(scale)인 $2\sum_{j=1}^m p_j(1-p_j)$ 로 나누게 된다.

$$P = 2(p_j - 0.5)$$

(2) GMF

GMF 방법은 G05 방법과 매우 유사하나 수식(1)의 P 행렬을 구성할 때 기준빈도를 0.5 대신 모든 표지인자의 소수대립유전자빈도평균(MAF)을 이용하는 것으로 P 행렬은 다음과 같다.

$$P = 2(p_j - \overline{MAF})$$

(3) GOF

GOF 방법은 G05 방법 및 GMF 방법과는 다르게 유전체 관계행렬의 척도(scale)인 $2\sum_{j=1}^m p_j(1-p_j)$ 을 표지인자별로 세분화하여 관측 유전자 빈도(Observed Gene frequency)를 고려한 방법이며 다음과 같다.

$$GOF = (M-P) D (M-P)', P = 2(p_j - 0.5),$$

여기서 D 행렬은 대각행렬이며 이 행렬의 j번째 원소 D_{jj} 는 아래와 같다.

$$D_{jj} = \frac{1}{m[2p_j(1-p_j)]}$$

(4) GOF*

GOF* 방법은 원리적으로 GOF 방법과 유사하나 유전자 빈도의 분산을 위하여 베타함수를 적용한 것이며 다음과 같다.

$$GOF^* = \frac{(M-P)(M-P)'}{\left[(p_0 - q_0)^2 + \left(\frac{\sum_{i=1}^m p_i(1-p_i)}{m} \right) \left(\frac{\alpha + \beta + 2}{\alpha + \beta} \right) \right] m}, P = 2(p_j - 0.5)$$

여기서 α 와 β 는 Beta process $B(\alpha, \beta)$ 의 두 모수이며 아래와 같이 추정하였다.

$$\hat{\alpha} = \bar{p} \left(\frac{\bar{p}(1-\bar{p})}{m} - 1 \right), \hat{\beta} = (1-\bar{p}) \left(\frac{\bar{p}(1-\bar{p})}{m} - 1 \right)$$

여기서 \bar{p} 는 평균 대립유전자 빈도이며, p_0, q_0 는 다음과 같다.

$$E(p) = \frac{\alpha}{(\alpha + \beta)} = p_0, E(q) = \frac{\beta}{(\alpha + \beta)} = q_0$$

(5) GN

GN 방법은 근교되지 않은 혈통 혈연계수행렬과 유전체 관계행렬을 유사하게 만들기 위하여 대각행렬값을 1에 가깝게 표준화한 것이며, 다음과 같다.

$$GN = \frac{(M-P)(M-P)'}{\text{trace}[(M-P)(M-P)'] / n}, P = 2(p_j - 0.5)$$

여기서 n은 유전체 정보를 소유한 개체수이며 trace는 대각합이다.

혈통정보를 이용한 혈연관계행렬은 Henderson (1976)이 제안한 순환참조방법(recursive method)를 적용한 작표 방법(tabular method)으로, 각 유전체 관계행렬은 Aguilar (2010b)의 ORIG 방법을 이용하였으며, 각 혈연계수 행렬의 역행렬을 구성하여 유전모수 및 육종가 추정에 이용하였다.

결과 및 고찰

1. 혈통 혈연계수와 유전체 혈연계수 비교

25,325개의 혈통자료를 이용하여 혈연관계행렬(numerical relationship matrix; NRM)을 구성한 뒤 유전체 정보를 소유한 490두의 조합별 혈연계수를 삼삼각행렬 형태로 추출하였다. 또한 이상 표지인자 및 성염색체상의 표지인자를 제거한 490두의 40,706개 SNP 표지인자를 이용하여 5가지 유전체 관계행렬(genomic relationship matrix; GRM)을 구성하고 이를 혈통을 이용한 혈연계수와 비교하였다.

개체별 유전체 정보의 유효성을 판정하기 위하여 혈통혈연계수 대비 유전체 혈연계수(G05 방법을 이용)를 도화한(plotting) 결과 Fig. 1의 왼쪽 그림에서 표기한 바와 같이 혈통정보와 유전체 정보가 어긋나는 개체들이 발견되었다. 이러한 개체들의 유전체 정보는 채혈, DNA 추출, chip 분석, 유전자형 스캔 등의 일련의 과정에서 발생한 표기오류(Labelling Error)에 의한 것으로 사료된다. 혈통-유전체의 부합(mismatch) 검증 결과 총 490두의 유전체 형성 개체 중 42두가 혈통정보와 유전체 정보가 서로 부합되는 것으로 나타나 제거하였으며, 제거 이후 혈연계수 선형회귀분석의 결정계수 값이 0.33에서 0.70으로 크게 향상되었다(Fig. 1).

Fig. 1의 오른쪽 그림에서와 같이 혈통혈연계수에 대한 유전체 혈연계수의 회귀직선 기울기는 1로 나타났으나 절편이 -0.17로 혈통혈연계수와 유전체 혈연계수가 잘 일치하지만 혈통정보를 이용한 혈연계수는 근교계수만큼 중심이 이동한 것을 확인할 수 있다. 이러한 차이의 원인은 혈통혈연계수 계산 시 기초집단부터 현세대까지의 모든 혈통정보를 이용하였으나 유전체 혈연계수는 생존하고 있는 가장 최근 세대만의 유전체 정보만을 이용하였기 때문인 것으로 사료된다.

Fig. 2는 혈통 혈연계수와 5가지 유전체 혈연계수의 편차값을 계산하여 빈도분포를 확인한 결과로 GOF*를 제외한 모든 방법에서 정규분포형태의 멘델리안샘플링을 나타내었다. GMF 방법을 제

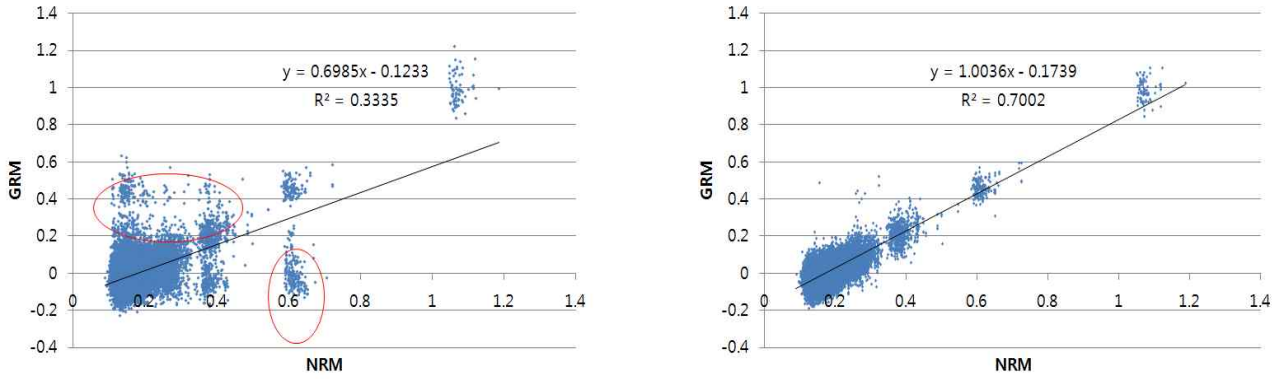


Fig. 1. Plot of relationship coefficients using pedigree information (NRM) and genomic information (GRM, using G05) before(left) and after (right) the elimination of unmatched animals.

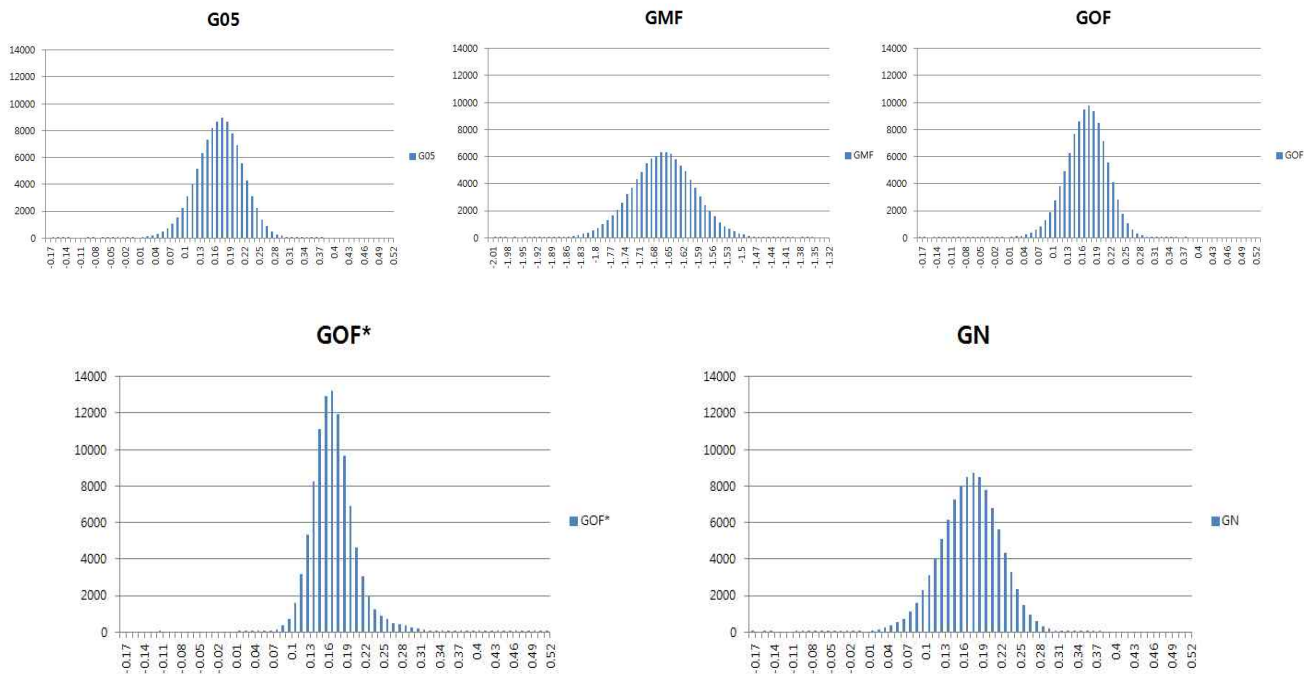


Fig. 2. Distributions of deviation between relationship coefficient using pedigree and genomic information by methods.

외한 나머지 방법에서 혈통혈연계수와 유전체 혈연계수의 편차인 0.179를 중심으로 정규분포 형태를 나타내고 있으며, GMF의 경우 기준 대립유전자빈도를 0.5 보다 작은 평균소수대립유전자빈도로 표준화하여 -1.652로 분포의 중심이 이동한 것을 확인할 수 있다. 혈통 혈연계수와 각 방법론별 유전체 혈연계수 편차값의 표준편차 비교 결과 GMF의 편차가 가장 크고, GOF의 혈연계수가 혈통혈

연계수와 가장 유사한 것으로 나타났다(Table 2).

2. 유전모수 추정

Table 3에서와 같이 평균등지방두께의 유전율은 전체 표현형 정보와 혈통을 이용할 때보다(Ntot; 집단 전체 개체 24,411두의 표

Table 2. Deviation of relationship coefficients estimated using pedigree and genomic information (NRM-GRM)

	G05	GMF	GOF	GOF*	GN
Mean	0.179	-1.652	0.179	0.180	0.179
Standard deviation	0.046	0.065	0.043	0.045	0.047

Table 3. Estimated variance components and heritability for average backfat thickness and days to 90 kg using pedigree and genomic information

ABF							
Parameter	Ntot ¹⁾	N448 ²⁾	G05	GMF	GOF	GOF*	GN
σ_a^2	0.024	0.013	0.015	0.015	0.018	0.029	0.015
σ_e^2	0.022	0.009	0.007	0.007	0.007	0.007	0.007
h^2	0.518	0.610	0.674	0.674	0.691	0.798	0.668
D90KG							
Parameter	Ntot	N448	G05	GMF	GOF	GOF*	GN
σ_a^2	38.27	29.85	23.74	23.74	24.20	44.21	23.20
σ_e^2	41.92	35.82	40.69	40.69	40.34	40.27	40.70
h^2	0.477	0.455	0.369	0.369	0.375	0.523	0.363

¹⁾ Ntot; using 24,411 phenotype records (all animals in GGP) and pedigree

²⁾ N448; using 448 phenotype records (genotyped animals) and pedigree.

현형 및 혈통이용) 유전체 정보를 소유한 448개의 표현형 자료만을 이용한 경우에서 높게 나타났으며, 유전모수가 감소한 것을 고려할 때 잔차분산이 상대적으로 낮게 추정된 이유로 사료된다. 유전체 정보를 형성한 개체만을 고려해보면 혈통정보를 이용할 때보다 (N448; 유전체정보를 형성한 448두의 표현형 및 혈통이용)에 비하여 유전체 정보를 이용한 방법들에서 유전분산이 더 크게 나타나 상대적으로 높은 유전율을 나타내었다. 90 kg 도달일령의 경우 평균등지방두께와는 다르게 전체 자료를 이용한 것보다 448두의 표현형 정보만을 이용하였을 경우 유전율이 낮게 나타났다. 특히 Ntot와 N448의 유전율은 서로 비슷하게 나타났으나 유전체 관계 행렬을 이용한 나머지 방법론에서는 큰 차이를 보였다.

GOF*의 경우 모든 형질에서 유전분산이 크게 추정되었는데 이는 유전체 관계행렬의 척도(scale)에 베타함수를 이용하는 차이에서 의한 것으로 사료되며, 같은 형태의 척도를 이용하는 G05와 GMF는 유전분산이 유사하게 나타나, 유전체 관계행렬 구성 시 대립유전자빈도를 표준화하는 P 행렬의 차이는 혈연계수의 평균이동을 유발하며, 분모부분을 구성하는 척도 부분은 유전모수 추정치에 영향을 주는 것을 확인할 수 있다.

3. 유전체 육종가 추정 및 정확도 비교

Table 4은 평균등지방두께 및 90 kg 도달일령에 대한 방법론별 추정 육종가의 상관분석 결과이다. 혈통자료를 이용한 Ntot와 N448의 육종가의 상관은 ABF 형질과 D90KG 형질에서 각각 0.614, 0.615로 형질에 상관없이 비슷하게 나타났으나 Ntot와 나머지 유전체 관계행렬을 이용한 육종가 상관은 형질별로 서로 다르게 나타났다. 특히 ABF 형질에서 집단전체의 정보를 이용한 Ntot의 육종가는 같은 혈통정보를 이용한 N448의 육종가보다도 유전체 정보를 육종가와 상관성이 더 높게 나타났으며, D90KG 형질에서는 Ntot와 N448의 육종가 상관성이 Ntot와 나머지 유전체 육종가와 상관보다 높게 나타났다. 혈통정보를 이용한 육종가와 유전체 정보를 이용한 육종가의 상관은 GOF가 가장 높게 나타났으며, 이는 Table 2에서와 같이 혈통정보를 이용한 혈연계수와 GOF 방법을 이용한 유전체 혈연계수와 상관성이 가장 높게 나타난 것이 반영된 결과로 사료된다. GOF를 제외한 G05, GMF, GOF*, GN 방법을 이용한 유전체 육종가들의 상관은 서로 1로 나타나 큰 차이가 없는 것으로 나타났으며, Forni (2011)의 연구결과와는 유사하나 GOF*, GN 방법에서 차이를 보였다.

Table 4. Correlations between estimated breeding values using different relationship matrices

Method	Ntot	N448	G05	GMF	GOF	GOF*	GN
Ntot		0.615	0.594	0.594	0.597	0.597	0.594
N448	0.614		0.908	0.908	0.919	0.914	0.907
G05	0.624	0.920		1.000	0.997	1.000	1.000
GMF	0.624	0.920	1.000		0.997	1.000	1.000
GOF	0.627	0.923	0.997	0.997		0.997	0.997
GOF*	0.626	0.925	1.000	1.000	0.998		1.000
GN	0.624	0.920	1.000	1.000	0.997	1.000	

Lower triangle : ABF, Upper triangle : D90 kg.

육종가 분석모형이 적용된 MME 역행렬의 대각행렬 중 개체의 상가적 유전효과 부분을 추출하여 Table 5에 혈통정보와 유전체 정보를 이용한 육종가 추정 정확도를 계산하여 비교하였다. GMF, GOF* 방법을 이용한 유전체 관계행렬은 대각행렬원소의 평균치가 달라 정확도가 과대 또는 과소 추정되기 때문에 계산하지 않았다. 모든 형질에서 Ntot의 육종가 추정 정확도가 가장 높았으며, 성별에 의한 차이가 많이 발생하지 않았다. 반면 소규모 표현형 자료와 혈통정보를 이용한 N448의 경우 정확도가 가장 낮게 나타났으며, 옹돈과 모돈에서의 차이가 크게 나타났다. 혈통상의 대부분의 개체들이 자기 자신의 기록을 가지고 있고 폐쇄집단에서의 장기간 교배에 의하여 개체별 혈연이 잘 연결되어 있기 때문에 전반적으로 육종가 추정의 정확도가 매우 높게 나타났다. 그러나 N448의 경우 제한된 표현형자료를 사용하기 때문에 가까운 혈연 관계이면서 기록을 가지고 있는 방계친척의 정보를 충분히 활용할 수 없어 정확도가 낮게 나타났다. 이와 같은 이유로 집단 내 혈연으로 연결된 개체가 많은 옹돈에서 정확도가 높게 나타나며, 상대적으로 자기 자신 이외에 표현형을 소유하면서 혈연으로 연결된 개체수가 적은 모돈에서 정확도가 낮게 나타나는 것을 확인할 수 있다. 유전체 정보를 이용할 경우 동일한 표현형 정보량을 이용하되 혈통정보를 이용한 N448 보다 모든 방법론에서 높은 육종가 추정 정확도를 보이고 있으며, ABF 형질에서 그 차이가 더 크게 나타났다. 유전체 정보를 이용하여 추정된 육종가의 정확도는 방법론별로 서로 비슷하나 GOF, G05의 정확도가 GN 방법에 비하여 높게 나타났으며 N448과는 다르게 옹돈과 모돈에서 정확도가 서로 비슷하게 나타났다. Forni (2011)의 연구에서는 GMF, GOF*

방법의 육종가 추정 정확도를 계산하였으나 본 연구와는 달리 전체 혈통정보와 유전체 정보를 결합시킨 행렬 (combined pedigree-genomic relationship matrix)을 구성하여 육종가를 추정하였으며, 유전율이 낮은 산자수 형질을 이용하여 본 연구 결과와는 차이를 나타내었다. 그러나 본 연구에서와 같이 G05 및 GOF 방법을 이용하였을 경우 GN 방법보다 육종가 추정의 정확도가 높게 나타났으며, 혈통을 이용한 육종가보다 유전체정보를 이용한 육종가에서 더 높은 정확도를 나타내었다.

본 연구에 이용된 평균등지방두께 및 90 kg 도달일령 형질은 개체 자신 및 혈연으로 연결된 다수의 개체들이 표현형 기록을 가지고 있기 때문에 혈통정보를 이용하여도 높은 정확도의 육종가를 추정할 수 있음을 예상할 수 있다. 오히려 규모가 작은 참조집단을 이용할 경우 집단의 누적된 표현형 자료를 활용할 수 없어 Table 5에서의 결과와 같이 유전체 육종가의 정확도가 전체 혈통과 표현형 자료를 이용한 육종가보다 정확도가 낮게 나타날 수 있다. 그러나 집단 내 누적 표현형자료가 부족한 경우, 외래 유전자원이 도입되어 집단 내 혈연관계가 부족할 경우 또는 멘델리안 분포가 전혀 고려되지 않는 어린 동복자손의 육종가를 예측해야 하는 경우에는 동일 표현형 정보량에서 혈통이용 육종가 대비 유전체 육종가 추정의 정확도가 높으므로 이를 활용할 수 있을 것으로 사료된다.

본 연구에서 사용된 유전체 정보량이 많지 않음에도 불구하고 동일 표현형 정보량에서 혈통정보 대비 유전체 육종가 추정의 정확도가 상당히 높게 나타난 것을 고려하면 추가적인 유전체 정보량의 누적을 통하여 지속적인 육종가 정확도 향상이 예상된다. Van Ranen 등 (2009)은 표지인자 밀도를 늘리는 것 보다 참조집단의 규모를 늘렸을 때 신뢰도 향상측면에서 2배의 효과가 나타났다고 보고하였으며, 또한 유전체 정보를 형성한 효과는 11두의 딸소를 형성한 것과 같은 정보량의 효과를 가진다고 언급하였다. 따라서 지속적인 참조집단 규모를 증가시켜야 유전체 선발을 효과적으로 적용할 수 있다고 사료되며, 집단의 기존 표현형 자료를 활용하기 위하여 새롭게 제시되고 있는 혈통-유전체 정보 결합 행렬 (Aguilar 등, 2010a; Legarra 등, 2009; Mistztal 등, 2009)에 관련된 추가 연구를 진행해야 할 것으로 사료된다.

사 사

본 연구는 농림수산식품부 첨단생산기술개발사업 (과제번호: 309022-5)에 의해 이루어진 것임

요 약

돼지 유전체 전장의 고밀도 단일염기다형 유전자형을 이용하여 혈연관계행렬을 구성하고 이를 이용하여 유전체 육종가를 추정하였다. 이상치를 제거한 랜드레이스 순종돈 448두의 40,706개 단일염기다형 유전자형 정보를 이용하였으며, G05, GMF, GOF, GOF* 및 GN의 5가지 방법을 이용하여 유전체 관계행렬을 구성하고 이

Table 5. Average accuracy estimates for breeding values of genotyped animals using pedigree and genomic relationship coefficients

	ABF		
	*Total	Genotyped Male	Genotyped Female
	n=448	n=31	n=417
Ntot	0.653	0.639	0.654
N448	0.280	0.335	0.276
G05	0.544	0.601	0.540
GN	0.534	0.592	0.530
GOF	0.546	0.607	0.541
	D90KG		
	*Total	Genotyped sire	Genotyped Female
	n=448	n=31	n=417
Ntot	0.631	0.618	0.632
N448	0.194	0.240	0.190
G05	0.358	0.376	0.357
GN	0.343	0.362	0.342
GOF	0.353	0.376	0.352

*: Total = (Genotyped Male + Genotyped Female)

를 이용하여 유전체 육종가를 추정하였다. GOF 방법에 의하여 계산된 혈연계수가 기존의 혈통정보를 이용한 혈연계수와 가장 작은 편차를 나타내고 평균소수대립유전자빈도를 이용하는 GMF 방법에서는 큰 차이가 나타나 대립유전자빈도 기준이 혈연계수의 평균이동을 유발함을 확인하였으며, GOF*를 제외한 모든 방법에서 정규분포형태의 멘델리안샘플링이 나타나는 것을 확인하였다. 등지방두께 평균과 90 kg 도달일령에 대한 육종가 추정 모형을 설정하고 유전체 관계행렬을 이용하여 유전모수와 육종가를 추정한 결과 혈통정보를 이용한 육종가와외의 상관은 GOF 방법에서 가장 높게 나타났으며, 유전체 관계행렬의 척도(scale)에 베타함수를 이용한 GOF*의 경우 모든 형질에서 유전분산이 크게 추정되어 분모부분을 구성하는 척도는 유전모수 추정치에 영향을 미치는 것을 확인하였다. 동일한 표현형 정보량을 이용할 경우 유전체관계행렬을 이용한 육종가 추정의 정확도가 혈통정보를 이용한 육종가보다 높게 나타났으며, 90 kg 도달일령보다는 등지방두께 평균에서 그 차이가 더 크게 나타났다. 집단 내 누적 표현형자료가 부족한 경우, 외래 유전자원이 도입되어 집단 내 혈연관계가 부족할 경우 또는 멘델리안 분포가 전혀 고려되지 않는 어린 동복자손의 육종가를 예측해야 하는 경우에 유전체 정보를 활용하면 유전능력 평가의 정확성을 크게 향상시킬 수 있을 것으로 사료된다.
(주제어: 유전체 선발, 유전체 관계행렬, 랜드레이스, 단일염기다형, 육종가)

인 용 문 헌

Aguilar, I., Misztal, I., Johnson, D., Legarra, A., Tsuruta, S. and Lawlor, T. 2010a. A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93:743-752.

Aguilar, I., Misztal, I., Legarra, A. and Tsuruta, S. 2010b. Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation. *J. Anim. Breed. Genet.* 128(6):422-428.

Forni, S., Aguilar, I. and Misztal, I. 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genetics Selection Evolution* 43:1-7.

Gianola, D., de los Campos, G., Hill, W. G., Manfredi E. and Fernando R. 2009. Additive genetic variability and the Bayesian alphabet. *Genetics* 183:347-363.

Habier, D., Fernando, R. L., Kizilkaya, K. and Garrick, D. J. 2011. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* 12:186-197.

Habier, D., Fernando, R. and Dekkers, J. C. M. 2007. The Impact of genetic relationship information on genome-assisted breeding

values. *genetics* 177: 2389-2397.

Hayes, B. J., Bowman, P. J., Chamberlain, A. J. and Goddard, M. E. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* 92(2):433-443.

Henderson, C. R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* 32:69-83.

Legarra, A. and Misztal, I. 2008. Technical Note: Computing Strategies in Genome-Wide Selection. *J. Dairy Sci.* 91:360-366.

Legarra, A., Aguilar, I. and Misztal, I. 2009. A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* 92:4656-4663.

Loberg, A. and Durr, J. W. 2009. Interbull survey on the use of genomic information. *Proceedings of the interbull technical workshop* 3-13.

Meuwissen, T. H., Hayes, B. J. and Goddard, M. E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819-1829.

Misztal, I., Legarra, A. and Aguilar, I. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J. Dairy Sci.* 92:4648-4655.

Stephens, M. and Donnelly, P. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics*, 73:1162-1169.

Sun, X., Habier, D., Fernando, R. L., Garrick, D. J. and Dekkers, J. C. M. 2010. Genomic breeding value prediction and QTL mapping of QTLMAS2010 data using Bayesian Methods. *BMC Proceedings* 5(3):S13.

VanRaden P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414-4423.

VanRaden P. M., Van Tassell, C. P., Wiggans, G. R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F. and Schenkel, F. S. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* 92:16-24.

Zhang, Z., Todhunter, R. J. Buckler, E. S. and Van Vleck, L. D. 2007. Technical note: Use of marker-based relationships with multiple-trait derivative-free restricted maximal likelihood. *J. Anim. Sci.* 85:881-885.

Mrode, R. A. 2005. *Linear Models for the Prediction of Animal Breeding Values*, 2nd ed. CABI, Scottish Agricultural College, Edinburgh, UK pp10.

Misztal, I. 2011. *Computational techniques in animal breeding*. University of Georgia, Athens, GA.

(Received Apr. 5, 2013; Revised Jun. 24, 2013; Accepted Jun. 25, 2013)