

한우의 유전체 표지인자 활용 개체 혈연관계 추정

이득환¹ 조충일¹ 김내수²¹한경대학교, ²충북대학교

Prediction of Genomic Relationship Matrices using Single Nucleotide Polymorphisms in Hanwoo

Deukhwan Lee¹, Chungil Cho¹ and NaeSoo Kim²¹Hankyong National University, ²Chungbuk National University

ABSTRACT

The emergence of next-generation sequencing technologies has led to application of new computational and statistical methodologies that allow incorporating genetic information from entire genomes of many individuals composing the population. For example, using single-nucleotide polymorphisms (SNP) obtained from whole genome amplification platforms such as the Illumina BovineSNP50 chip, many researchers are actively engaged in the genetic evaluation of cattle livestock using whole genome relationship analyses. In this study, we estimated the genomic relationship matrix (GRM) and compared it with one computed using a pedigree relationship matrix (PRM) using a population of Hanwoo. This project is a preliminary study that will eventually include future work on genomic selection and prediction. Data used in this study were obtained from 187 blood samples consisting of the progeny of 20 young bulls collected after parentage testing from the Hanwoo improvement center, National Agriculture Cooperative Federation as well as 103 blood samples from the progeny of 12 proven bulls collected from farms around the Kyong-buk area in South Korea. The data set was divided into two cases for analysis. In the first case missing genotypes were included. In the second case missing genotypes were excluded. The effect of missing genotypes on the accuracy of genomic relationship estimation was investigated. Estimation of relationships using genomic information was also carried out chromosome by chromosome for whole genomic SNP markers based on the regression method using allele frequencies across loci. The average correlation coefficient and standard deviation between relationships using pedigree information and chromosomal genomic information using data which was verified using a parentage test and eliminated missing genotypes was 0.81 ± 0.04 and their correlation coefficient when using whole genomic information was 0.98, which was higher. Variation in relationships between non-inbred half sibs was 0.22 ± 0.17 on chromosomal and 0.22 ± 0.04 on whole genomic SNP markers. The variations were larger and unusual values were observed when non-parentage test data were included. So, relationship matrix by genomic information can be useful for genetic evaluation of animal breeding.

(**Key words** : Relationship matrix, Genomic and pedigree information, Single nucleotide polymorphism, Hanwoo)

서 론

통계유전학적 방법에 의한 가축개량의 기본 원리는 부모와 자손 간의 유사도에 기초하여 발전되어 왔다. 즉 혈연관계가 있는 개체 간의 닮음의 크기는 Mendel의 유전법칙에 기초하여 이들 개체 상호간의 공변이로 표현되며, 이는 해당 형질의 유전변이에 대한 합수로 정의될 수 있다. 여기서 혈연관계가 있는 다수의 개체들 간의 유사도의 크기를 나타내는 계수를 Wright (1922)는 혈연계수로 정의하였으며, Henderson (1976)의 개체들간 혈통정보를 이용한 혈연계수행렬 및 이에 대한 역행렬을 손쉽게 계산하는 방법을 응용하

여 통계적 혼합모형식을 통한 육종가를 추정하는 방법 (Henderson, 1984)으로 발전하였다. 따라서 통계육종학적 방법에 의한 가축의 유전능력 추정은 개체간 혈연계수행렬의 추정이 필수적이라 할 수 있다.

한편 DNA 분자구조가 밝혀진 이후 분자유전학의 비약적인 발전과 더불어 분자수준의 유전적 다양성에 대한 많은 자료가 생성되고 있다 (Schumm et al., 1988). 최근 단일염기서열 분석기술의 발전에 따라 유전체 전장에 있어서 개체간 염기서열의 차이 (SNP)를 이용한 유사도의 크기를 추정하고 이를 이용하여 가축의 유전능력을 예측하기 위한 연구도 급속도로 이루어지고 있다.

* Corresponding author : Deukhwan Lee, Department of Animal Life and Resources, Hankyong National Univ. Seokjeong-dong Ansong-si Gyeonggi-do, Rep. of Korea 456-749. Tel: 031-670-5091, Fax: 031-676-5091, E-mail: dhlee@hknu.ac.kr

DNA 정보를 이용한 개체간 유사도에 대한 연구는 인체의 유전병을 중심으로 많이 이루어졌는데 특히 이는 가계도 분석을 통하여 두 개체가 동일 유전자(Identity by decent: IBD)를 가질 확률을 예측하는데서 시작되었다. 하나 또는 수개의 유전자(또는 표지인자)를 이용한 개체간 IBD 행렬의 추정에 대한 수많은 연구가 있었는데, Lynch(1988)는 인간의 DNA 지문에 있어서 VNTR(variable number of tandem repeat) 좌위들을 이용한 혈연관계의 크기를 추정하는 방법을 제시하였고, Guo(1996)는 다수 좌위의 정보를 갖는 유전자에 대한 개체간 IBD 추정에 있어서 배우자형성 과정을 고려한 분석방법을 제시하였다. 또한 Nolte와 Meerman(2002)은 초위성체 또는 SNP haplotype이 유사할 때 IBD일 확률을 계산하는 방법을 제시하였으며, Cannings(2003)은 염색체 단위에서 Markov Chain을 이용한 IBD를 계산하는 방법을 제시하였다. Chapman과 Thompson(2003)은 제한된 유효집단 크기를 갖는 임의교배집단에서 IBD 추적을 위한 분석모형을 제시하였으며, Leutenegger 등(2003)은 인간의 유전체 전장의 정보를 이용한 근교계수 추정방법을 제시하였다. Hernandez-Sanchez 등(2004)은 동물 유전자형 정보를 활용하여 3개 또는 4개 좌위에서 동시에 IBD일 확률(근교계수)을 추정하는 방법을 제시하였으며 Libiger와 Schork(2007)은 염색체의 크기를 고려한 임의의 개체간에 염색체 단편이 IBD일 확률을 계산하는 방법을 제시하였다. 또한 Hill과 Hernandez-Sanchez(2007)는 임의교배집단에서 다수 좌위의 정보를 고려한 IBD 계산방법을 제시하였고 Browning(2008)은 고밀도 SNP Haplotypes 표본으로부터 Hidden Markov model을 활용한 IBD 추정방법을 제시하였다.

Blouin(2003)은 임의집단에서 분자차원의 유전정보를 활용하여 혈통을 재구성하고 혈연관계를 분석하는 기술에 대한 발전과정을 기술하였다.

집단유전학적 개념에서 DNA 정보를 이용한 개체간의 유사도를 추정하는 또 다른 접근방법으로써 유전자(또는 표지인자)를 선조와 자손간에 공유할 확률을 계산하는 Coalescence process(Hudson, 1985)을 들 수 있다. Meuwissen과 Goddard(2000)은 Coalescence process 방법으로써 혈통기록이 없이 DNA Marker Haplotype 정보를 활용하여 LD 값을 계산하고 이에 의한 IBD 행렬을 추정하는 방법을 제시하였으며, Meuwissen와 Goddard(2001)는 이러한 방법에 혈통정보를 포함하여 IBD 행렬을 추정하는 방법을 제시하였다.

가축에 있어서 Marker Assisted Selection(MAS)를 목적으로 QTL과 연관된 표지인자 정보를 활용한 개체간 혈연관계 행렬 추정에 대한 연구는 Fernando와 Grossman(1989)에 의해 수행되었다. Arendonk 등(1994)은 Fernando와 Grossman(1989)이 제시한 IBD 행렬 즉 배우자혈연계수행렬의 역행렬을 계산하는 방법을 제시하였으며 Wang 등(1995)은 QTL에 대한 상가적 유전효과를

추정하기 위한 공분산행렬 및 이에 대한 역행렬을 계산하는 방법을 일반화하였다. 이후 혈통기록을 활용하여 하나 또는 수개의 표지인자에 대한 개체간 혈연관계행렬 및 이에 대한 역행렬을 계산하는 방법에 대한 많은 연구가 있었다(Perez-Enciso et al., 2000; Pong-Wong et al., 2001; Abdel-Azim and Freeman, 2001; Liu et al., 2002; Marsuda and Iwaiszki, 2002; Totir et al., 2004; Tuchscherer et al., 2004; Sargolzaei et al., 2006; Hernandez-Sanchez et al., 2006; Abraham et al., 2007; Meuwissen and Goddard, 2007).

주어진 표지인자에 의한 개체간 IBD 추정은 이와 연관된 QTL의 임의효과에 대한 개체간 상관추정의 함수로 나타낸다. 이러한 상관추정은 다수의 표지인자로 확대될 수 있다. 다수의 표지인자를 활용할 경우 염색체 단편의 개체간 유사도 즉 AIS(alike-in-state)를 활용하여 IBD를 추정할 수 있다(Eding and Meuwissen, 2001). 이러한 방법은 혈통기록에 의한 정보를 활용하여 혈연계수행렬을 추정할 때 발생될 수 있는 많은 문제점을 해결할 수 있는 장점이 있다. 즉 만일 혈통기록 정보를 활용하면 자료의 수집상에 많은 어려움과 신조의 기록을 모두 확보해야 하는 어려움, 일부 그릇된 혈통기록이 있을 경우에 추정의 신뢰도가 저하되는 문제점이 있다(Hayes and Goddard, 2008).

본 연구는 이러한 측면에서 한우의 유전체 정보를 활용하여 염색체 단위의 개체간 혈연계수행렬을 추정하고, 기존의 혈통정보를 활용한 혈연계수행렬과 비교함으로써 종축의 유전능력 평가에 있어서 유전체 정보를 활용한 혈연계수행렬의 유용성을 검증하기 위하여 수행하였다.

재료 및 방법

1. 분석자료 수집

유전체 및 표현형 정보를 추출하기 위하여 공시된 검정축군은 농협중앙회 한우개량사업소에서 수행한 20두의 후보종모우 자손 187두의 혈액시료와 경북 지역농가에서 수집한 보중종모우 12두의 자손 103두의 혈액시료 등, 총 290두의 혈액을 분석에 공시하였다. 특히 농협중앙회 공시 검정우는 혈액으로부터 초위성체 DNA 검사를 통하여 친자가 확인된 개체들이었으나 경북에서 조사된 개체는 친자가 확인되지 않은 개체들로 구성되었다. 290두 검정우의 혈액으로부터 추출된 DNA는 단일염기다형(SNP)으로부터 최대한의 다형정보를 얻을 수 있는(Carlson et al., 2004) Illumina BeadArray™ BovineSNP50 assay를 이용하여 SNP 표지인자의 다형성 조사에 이용되었다. 조사된 유전체 전장의 SNP 자료로부터 Minor Allele Frequency(MAF)가 5% 이상인 표지인자를 추출하여 각 표지인자 좌위별 다형성 조사 및 유전체 혈연관계 추정에 이

용하였다. SNP 정보를 활용한 혈연관계 추정은 친자가 확인 된 187두의 기록만을 이용한 경우(data set 1)와 공시된 총 290두의 기록을 모두 포함한 경우(data set 2)로 구분하였다. 혈연관계 추정에 일부 좌위의 유전자형에 대한 결손(missing genotype-MISS)이 어떠한 영향을 미치는지 확인하기 위하여 분석에 고려된 개체 중 단 한 개체라도 특정 좌위가 MISS 되었다면 해당 좌위는 분석에서 제외한 경우와 MISS된 좌위를 모두 포함하는 경우로 다시 세분류하였다. 즉 Data set 1에서 MISS 좌위를 포함한 경우(DS1)와 이들 좌위를 제외한 경우(DM1)로 분류하여 자료를 정리하였고, 또한 Data set 2에서도 동일한 방법으로 MISS 좌위를 제외한 경우(DS2) 및 이를 포함한 경우(DM2)로 분류하여 유전체 혈연관계 추정을 위한 자료로 정리하였다. 개체별 혈통기록정보는 조사된 개체의 생산이력관리를 위하여 부여한 개체식별번호를 이용하여 한국종축개량협회 인터넷 홈페이지로부터 개체별 혈통정보를 선조 4대까지 추적하여 구성하였다.

2. 혈연관계 추정

유전체 SNP 정보를 활용한 혈연관계 추정은 각 염색체별로 추정하였으며 혈연관계 추정을 위한 함수식 (VanRaden, 2008)은 집단 내의 각 좌위에 있어서 대립유전자의 평균효과를 “0”으로 할당시키는 방법으로 다음과 같이 추정하였다.

만일 n 개체가 각각 m개의 표지인자를 갖는다면 이를 M_{mm} 행렬로 표기하였다. M 행렬의 각 원소를 minor allele에 의한 동형접합체 이면 0, 이형접합체이면 1, Major allele만으로 구성된 동형접합체이면 2로 정의하면 MM' 행렬은 n×n 크기의 행렬이고 이 행렬의 대각선 값은 해당 개체의 동형접합체 수이고 대각선 밖의 값은 두 개체간에 공동으로 갖는 allele의 수가 된다.

i번째 좌위의 2번째 allele의 빈도를 p_i로 정의하고 집단내 각 좌위별 유전자 빈도에 대한 행 벡터를 P_{ixm}으로 정의하면 각 좌위별 유전자 빈도를 고려한 행렬 Z=[M- ×1×P]로 정의하였다. 여기서 1_{mx1}은 행의 수가 n인 열벡터이고 M은 상기의 개체별 표지인자 정보를 갖는 행렬이다. 여기서 유전체정보를 이용한 혈연계수행렬 (G)은 다음 Eq (1)과 같이 계산하였다 (VanRaden, 2007).

$$G = \frac{ZZ'}{2\sum_i p_i (1-p_i)} \dots\dots\dots \exists q (1)$$

만일 실험의 오차 등으로 유전자형의 오류가 있을 경우에 선조의 유전체 정보와 혈통정보가 있을 경우에 Gengler 등 (2007)이 제시한 방법을 활용할 수 있으나 본 연구는 선조의 유전체 정보가 없었기 때문에 VanRaden (2008)이 제시한 방법으로 다음과 같이 수정하여 추정하였다.

상기의 Z행렬에서 j번째 개체가 i번째 좌위에서 MISS라면 z_{ji}=0으로 설정하고 기타의 경우에는 상기의 Z 행렬 설정과 동일하게 구성한다. 또한 MISS 좌위들에 대한 가중치 값을 각 개체들에 대하여 보정 후 혈연계수행렬을 추정하였다. 여기서 가중치에 대한 행렬은 대각선 행렬이고 이를 W로 정의하면 j번째 개체에 대한 가중치 벡터는 다음과 같이 계산하였다.

$$W_{jj} = \sqrt{\frac{\sum_i^{mk} p_i (1-p_i) \text{ for all loci}}{\sum_i^{mk} p_i (1-p_i) \text{—only nonmissing loci}}}$$

여기서 mk는 총 표지인자 수이다.

MISS 좌위들에 대한 가중치를 고려하여 상기의 행렬을 활용한 유전체 혈연계수행렬은 다음 Eq (2)과 같이 추정하였다.

$$G = \frac{WZZ' W}{2\sum_i p_i (1-p_i)} \dots\dots\dots \exists q (2)$$

혈통정보를 이용한 혈연계수행렬은 Wright (1922)의 혈연계수에 기초하여 Aguilar와 Misztal (2008)이 제시한 recursive algorithm 으로부터 얻어진 근교계수를 활용하여 개체간 혈연계수를 추정하였다.

결과 및 고찰

1. 한우의 유전체 자료분석

본 연구에 이용된 검정우 290두는 32두의 종모우 자손으로 구성 되어 평균 종모우 두당 9두의 반형매 집단이었으며 이들 집단으로부터 SNP 자료 및 혈통자료를 조사하였다. BovineSNP50 assay 를 통하여 조사된 총 44,819 SNP 표지인자들 중 약 12.4% (5,554 markers)가 다형성이 없는 것으로 분석되었고 약 14.9% (6,698)의 표지인자들에 있어서 Minor Allele Frequency (MAF)의 빈도가 5% 이하이었으며 여기서 MAF가 5% 이하인 표지인자는 제거하고 남은 총 32,567 SNP 표지인자가 개체별 다형성 분석에 이용되었다 (Fig. 1).

다형성이 없거나 MAF가 5% 이하인 표지인자를 제외한 후, 본 연구에 이용된 자료의 일반통계량은 Table 1에 제시하였는데, 통계 분석에 이용된 개체 290두에 대하여 실험상의 오류 등으로 유전자형을 알 수 없는 좌위를 포함한 분석자료(DS2)에서 유전자형의 결손이 있는 좌위수는 개체별 평균 109개 좌위로서 0.335%이었으며 이들 자료는 본 연구의 목적상 fastPHASE (Sheet and Stephens, 2006) 또는 NPUTE (Roberts et al., 2007) 등의 방법은 적용하지 않았으며 표지인자간 거리의 평균(±SD)은 0.078 Mb

Table 1. Total number of markers, average distance between flanking markers, and fractional missing genotypes on whole genomic scan by different data sets (including or excluding missing genotypes and parentage test in the number of 290 Hanwoos

	Animals with parentage test (N=187)	Animals regardless of parentage test (N=290)
Including missing genotypes (Abbr.)	DS1	DS2
No. Markers	32,432	32,567
Ave. distance (Mb)	0.078	0.078
Std. distance (Mb)	0.076	0.076
Prop. missing genotypes (%)	0.363	0.335
Excluding missing genotypes (Abbr.)	DM1	DM2
No. Markers	22,342	19,931
Ave. distance (Mb)	0.114	0.127
Std. distance (Mb)	0.113	0.129

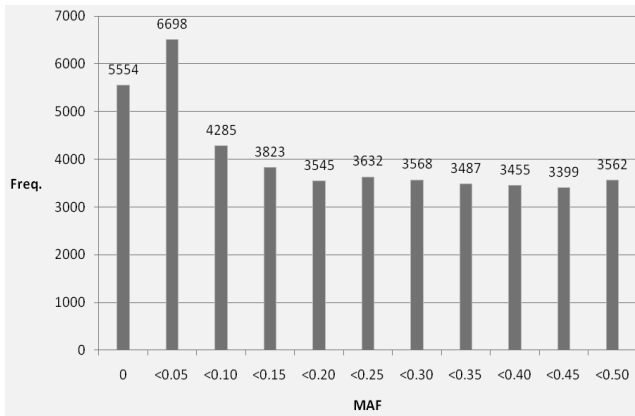


Fig. 1. Numbers of markers by the minor allele frequencies on 44,819 SNP markers in the number of 290 Hanwoos.

(±0.076 Mb)이었다. 이들 290두의 개체의 SNP 자료에서 MISS 좌위를 모두 제거한 후에 이용된 자료(DM2)의 구성을 살펴보면 (Table 1) 총 분석에 이용된 표지인자 수는 19,931 좌위이었으며 표지인자간 거리는 평균 0.127 Mb이고 표준편차는 0.129 Mb 이었다. 반면에 본 연구에 이용된 자료들 중 한우 후대검정 대상우는 초위성체를 활용하여 친자가 확인된 개체 (농협중앙회, Personnel communication) 187두를 포함하고 있었는데 친자확인인 된 한우 후대검정자료만을 추출하여 MISS 좌위를 포함한 자료(DS1)는 총 표지인자 수가 32,432 좌위이었고 표지인자간 거리의 평균 및 표준편차와 MISS 좌위 수는 DS2와 유사한 결과를 갖는 자료로 구성되었다. 반면에 이들 자료들 중 MISS 좌위를 제거하고 구성된 자료(DM1)는 유전자형의 오류가 경북지역에서 조사된 개체의 유전자형 자료보다 다소 양호하여 총 표지인자 수는 22,342 좌위에

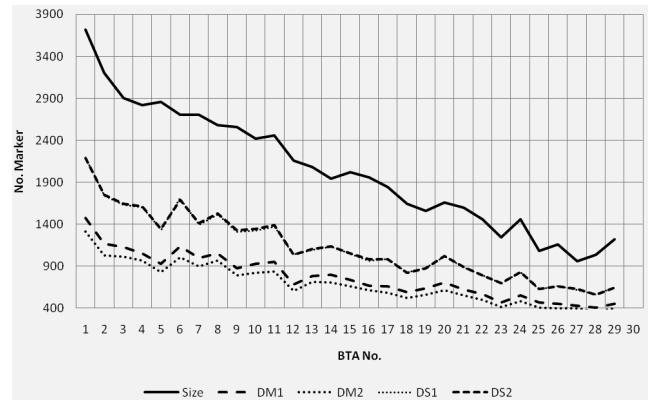


Fig. 2. Plots of numbers of SNP markers by each bovine chromosome (BTA) on each data set from the number of 290 Hanwoos.

Where DM1: data having paternity test and excluding miss genotypes, DM2: data including without paternity test and excluding miss genotypes, DS1: data having paternity test and including miss genotypes, DS2: data including without paternity test and including miss genotypes, Size: chromosomal size (2×10^5 bp, Snelling et al., 2007).

서 평균 좌위 간 거리는 0.224Mb이었다 (Table 1). 유전체 정보를 이용한 혈연관계 추정에 이용된 각 염색체별 표지인자 수에 대한 정보를 Fig. 2에 제시하였다. Fig. 2에 제시된 바와 같이 DS1 자료와 DS2 자료간에 염색체별 표지인자수는 커다란 차이가 없었으나 DS1 대비 DM1 자료는 염색체별 평균 표지인자 수가 69%, 그리고 DS2 대비 DM2의 자료는 62%의 표지인자 수가 개체간 혈연관계 추정에 이용되었다. 즉 Fig. 2에서는 개체별 특정 marker에 있어서 genotyping error로 SNP 자료를 얻지 못한 것을

missing으로 간주하고 해당 marker를 본 분석에서 이용하였을 때와 그렇지 않았을 때 염색체별 이용된 marker수를 보여주었다.

2. 유전체 혈연계수 추정

(1) 혈통기록이 옳은 경우

친자가 확인된 개체에서 MISS 좌위를 모두 제거한 후에 구성된 유전체 전장의 SNP 자료(DM1)를 활용하여 추정된 개체간 혈연관계와 혈통정보에 의하여 추정된 혈연관계를 비교하였다. 유전체 자료를 이용한 혈연관계는 1) 각 염색체별 유전자형 자료를 이용하여 추정된 혈연관계와 2) 유전체 전장의 유전자형을 이용하여 추정하였으며 여기서 얻은 혈연관계는 혈통정보에 의한 혈연관계 추정치와 비교하였다. Fig. 3에는 DM1 자료의 1번 염색체 1,476 좌위의 유전자형을 이용한 개체간 혈연계수 추정치와 혈통자료를 이용한 혈연계수 추정치들의 자료를 상호 산포도 및 회귀식으로 제시하였다. Fig. 3에 제시된 바와 같이 단일연관그룹으로 정의되는 단일 염색체 상의 표지인자들을 이용한 개체간의 혈연관계 추정은 혈통정보를 이용한 혈연관계 추정치와 87%의 상관관계를 갖는 것으로 추정되었으며 회귀계수에 있어서도 1.001으로 혈통정보와 표지인자 정보간에 추정치의 편차가 없는 것을 알 수 있었다. 혈통정보에 의한 개체간 혈연관계는 Mendelian sampling의 효과를 무시한 개체간 혈연관계의 기대치로써 예를 들면 반형매간 혈연관계는 0.25이지만 유전체정보를 활용할 때에는 Mendelian sampling 효과를 고려할 수 있기 때문에 다양한 변이를 갖는데, 만일 비교 대상의 두 개체가 근친이 되지 않고 표지인자 간에 완전 연관상이라면 그 범위는 0.0~0.5의 범위를 갖는다. 즉 만일 표지인자를 이용한 혈연관계 추정에 있어서 이용된 표지인자간에 상호 독립적이라면, 즉 연관 관계가 없다면 그 변이의 폭은 독립인자의 수와 함수관계가

존재하는데 (VanRaden, 2007) 반형매의 경우 하나의 좌위를 고려할 때, 하나의 대립유전자를 공통으로 가질 확률 (혈연관계=0.5)이 0.50이고 하나의 대립유전자도 공통으로 갖지 않을 확률 (혈연관계=0.0)이 0.50이다. 이를 두 개의 독립된 좌위로 확대하여 공통으로 갖는 대립유전자의 수가 0, 1, 2이면 혈연관계는 각각 0.00, 0.25 및 0.50이고 이때의 확률은 각각 0.25, 0.50 및 0.25이며 이를 일반화하여 n개의 독립된 좌위를 고려한다면 k개의 대립유전자를 공통으로 가질 확률은 $(\frac{1}{2})^n \binom{n}{k}$ 이고 좌위수가 무한히 많아지면 공통의 대립유전자수에 대한 확률분포는 정규분포에 근사하여 $0.25 \pm (\frac{1}{2})^n \frac{1}{\sqrt{2m}}$ 의 통계량을 갖는다 (VanRaden, 2007). 이러한 현상을 알아보기 위하여 연관그룹으로 정의될 수 있는 염색체별 SNP 정보를 이용한 반형매간 혈연관계의 크기를 산포도 (Fig. 4) 및 기초 통계량으로 제시하였으며 (Table 2) 유전체 전장의 SNP 정보에 의한 반형매 간 혈연관계의 크기에 대한 산포도 및 기초통계량도 함께 제시하였다. Table 2에 제시된 바와 같이 단일 연관그룹별로

Table 2. Means, standard deviations (SD), minimum (Min) and maximum (Max) of relationship for half sibs (548 pairs) by chromosomal SNP markers (CG) and whole SNP markers (WG) on data including with paternity test in the number of 187 Hanwoos

		Mean	STD	Min	Max
CG	DM1	0.22	0.17	- 15	0.71
	DS1	0.22	0.17	- 14	0.71
WG	DM1	0.22	0.04	0.11	0.37
	DS1	0.22	0.04	0.10	0.37

DM1 = data with excluding miss genotypes,
DS1 = data with including miss genotypes.

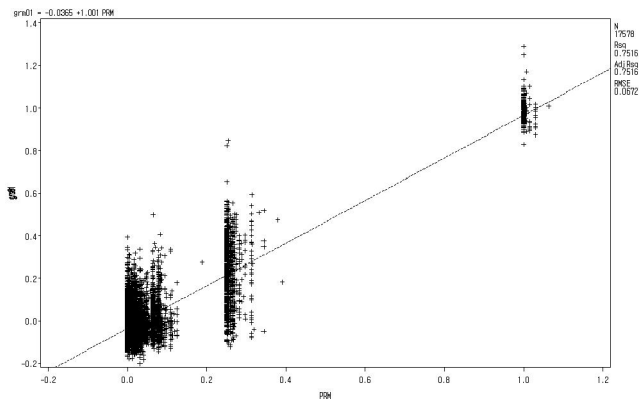


Fig. 3. Plots of genomic relationship matrix using 1,476 SNP markers on BTA1 by pedigree relationship matrix and regression coefficient on data having paternity test and excluding miss genotypes (DM1) in the number of 187 Hanwoos.

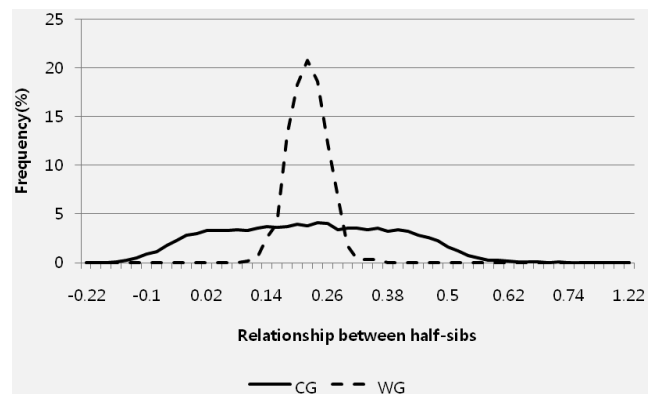


Fig. 4. Frequencies of relationships between non-inbred half sibs using chromosomal SNP markers (CG) and whole genomic SNP markers (WG) in the number of 187 Hanwoos.

추정한 반형매 간 혈연관계에 대한 변이의 크기는 표준편차에서 0.17으로 추정되었는데 이는 상기의 VanRaden (2007)이 제시한 공식을 이용할 때, 단일연관 그룹으로 정의되는 염색체별 독립 표지인자수가 1.08개로 추정될 수 있다. 이러한 추정치는 염색체의 크기가 각각 다르기 때문에 (Fig. 2) 염색체별 독립 표지인자 수는 다소 차이가 있을 것으로 판단되며 이러한 독립 표지인자 수는 표지인자 간의 연관의 크기에 영향을 것으로 판단된다. 유전체 전장을 이용한 반형매 간 혈연관계의 크기에 대한 변이는 표준편차에서 0.04으로 추정되어 독립 표지인자 수는 모두 31.5개로 추정되었다. 이러한 결과는 동일염색체상의 표지인자 간에 상호 강한 연관관계를 갖기 때문에 유전체 정보를 이용한 육종가 추정에 크게 영향을 것으로 사료되었다.

개체 자신에 대한 혈연관계의 크기는 1+F (근교계수)로 정의될 수 있는데 단일염색체상의 특정 표지인자에 대한 근교계수는 IBD로 정의될 수 있으며 개체간 변이의 크기는 개체의 근친도에 영향을 받는다고 할 수 있다. 만일 i번째 좌위의 IBD일 확률이 F_i 라면 i번째 좌위에서 동일한 혈연정보를 갖는 개체간 표준편차는 $\sqrt{F_i(1-F_i)}$ 으로 표현되며 (Weir and Cockerham, 1974) 만일 n개의 독립된 표지인자를 고려한다면 개체간의 표준편차는 $\sqrt{nF_n(1-F_n)}$ 으로 정의될 수 있다 (Hernandez-Sanchez et al., 2004). 여기서 F_n 은 n개의 서로 다른 표지인자가 모두 IBD일 확률로 정의된다. Fig. 5는 단일 연관그룹 및 유전체 전장의 경우에 개체 자신에 있어서 혈연관계의 변이를 분포도로 제시하였다. 단일 연관그룹에서 개체자신에 대한 혈연관계의 기대치는 최대 2인 반면에 본 연구는 0.08의 표준편차 범위를 갖는 것으로 추정되었다 (Table 3). 또한 유전체 전장의 정보를 이용하여 추정된 개체간 변이의 폭은 0.02으로 (Table 3) 매우 높은 첨도를 갖는 분포양상을 보였다 (Fig. 5). 이러한 결과로 볼 때 유전체 정보를 이용한 혈연관계를 이용하여 육종가를 추정하면 혈통정보에 의한 혈연관계를 이용한 육종가 추정에서 보다 개체들 간에 발생하는 염색체의 Mendelian sampling 현상을 설명할 수 있기 때문에 추정의 정확도를 향상시킬 수 있을 것으로 사료되었다 (Visscher et al., 2006).

단일 연관그룹의 SNP 정보를 이용하여 추정된 혈연관계와 혈통정보에 의한 혈연관계의 상관 추정치는 Fig. 2에서 제시한 바와 같이 1번 염색체의 SNP 정보를 이용하였을 때 상관 추정치가 0.87인 반면에 29개 염색체 모두에서 평균 0.81으로 추정되었으며 (Table 4) 이는 단일 연관그룹내 SNP 정보가 적을수록 추정치가 낮게 추정되었다. 또한 유전체 전장의 정보를 이용하였을 때에는 0.97으로 매우 높게 추정되었으며 회귀계수는 모두 0.99의 추정치를 얻을 수 있었기 때문에 (Table 4) 이러한 결과로 볼 때 SNP 정보를 이용한 개체간 혈연관계 추정은 개체별 육종가 추정에 크게 이용할 수 있을 것으로 판단되었다.

표지인자를 이용한 유전적 혈연관계를 추정하는데 있어서 표지인

Table 3. Means, standard deviations (SD), minimum (Min) and maximum (Max) of relationship for animal itself (187 pairs) by each bovine chromosomal SNP markers (CG) and whole SNP markers (WG) on data including with paternity test in the number of 187 Hanwoos

		Mean	STD	Min	Max
CG	DM1	0.99	0.08	0.81	1.39
	DS1	0.99	0.07	0.83	1.39
WG	DM1	0.99	0.02	0.93	1.09
	DS1	0.99	0.02	0.94	1.10

DM1 = data with excluding miss genotypes,
DS1 = data with including miss genotypes.

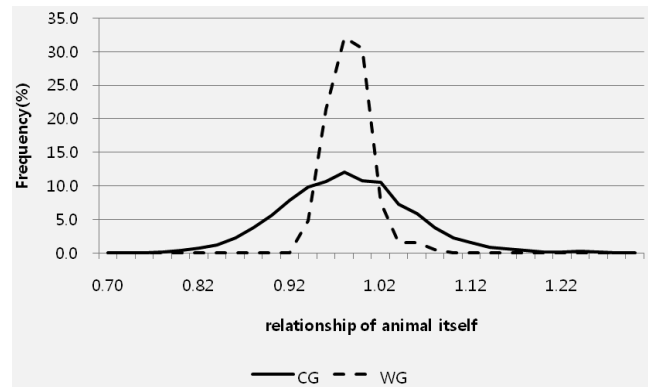


Fig. 5. Frequencies of relationships of animal itself (1 + inbreeding) using chromosomal SNP markers (CG) and whole genomic SNP markers (WG) in the number of 187 Hanwoos.

자 수 및 표지인자간 거리가 혈연관계행렬에 어느 정도 영향을 미치는지 알아보기 위하여 DS1 자료를 이용한 단일그룹별 혈연관계추정치와 유전체 전장의 정보를 이용한 혈연관계 추정치에 대한 기초통계량을 알아보았는데 (Table 3, Table 4) 반형매 간 혈연관계 및 개체 자신에 대한 혈연관계의 크기가 DM1 자료를 이용하였을 때 보다 변이의 폭이 소량 감소하는 경향치를 보였으나 유의적인 차이는 없었다. 표지인자가 결손된 좌위를 포함하여 혈연관계를 추정시 분석에 고려된 좌위수가 45.2% 증가하였으며 (22,342 vs 32,432, Table 1 참조) 이때 단일 연관그룹에 있어서 혈통정보에 의한 혈연관계와의 상관 추정치에 대한 평균이 DM1의 자료에서 보다 1% 향상된 0.82으로 추정되었고 유전체 전장의 자료를 이용하였을 때에는 0.98으로 추정되어 DM1의 자료를 이용하였을 때 보다 조금 높게 추정되었으나 커다란 차이는 없었다. 또한 유전체 정보를 통하여 얻은 혈연관계의 크기에 대한 편의를 나타내는 회귀계수는 단

Table 4. Regression coefficients (b) and correlations (r) for relationships by SNP information to relationship by pedigree information on data including with paternity test in the number of 187 Hanwoos

	Single linkage group ¹⁾		Whole genomic information	
	b	r	b	r
DMI	0.99	0.81	0.99	0.97
DS1	1.00	0.82	1.00	0.98

DMI = data with excluding miss genotypes,

DS1 = data with including miss genotypes

¹⁾ Means of estimates on each chromosome.

일 연관그룹인 염색체 단위로 추정된 혈연관계와 유전체 전장에서 얻은 혈연관계 모두에서 혈통기록을 통하여 얻은 혈연관계의 추정치와의 회귀계수에서 0.99~1.00으로 추정되어 어느 정도 불편추정임을 알 수 있었다. 이러한 결과로 살펴볼 때, Eq (2)를 이용하면 실험상 오류로 인한 SNP 표지인자의 결손된 좌위를 포함한 유전체 정보를 포함한 혈연관계 추정에 있어서 결손자료의 기중치를 고려한 추정 함수식의 타당성이 있을 것으로 사료되었다.

각 염색체별 유전체를 이용한 혈연관계 행렬의 유사성을 알아보기 위하여 DMI 자료를 이용한 염색체별 개체간 혈연관계 추정치를 변량으로 간주하고 염색체 간의 상관추정치와 해당 염색체에 이용된 표지인자수의 곱을 로그화값에 대하여 산포도를 알아보았으며 두 변량간의 회귀식을 추정하여 보았다(Fig. 6). 추정된 회귀식은 $y_{ij} = b_0 + b_1 \log_e(x_i \times x_j) + e_{ij}$ 이며 여기서 y_{ij} 는 i번째 염색체와 j번째

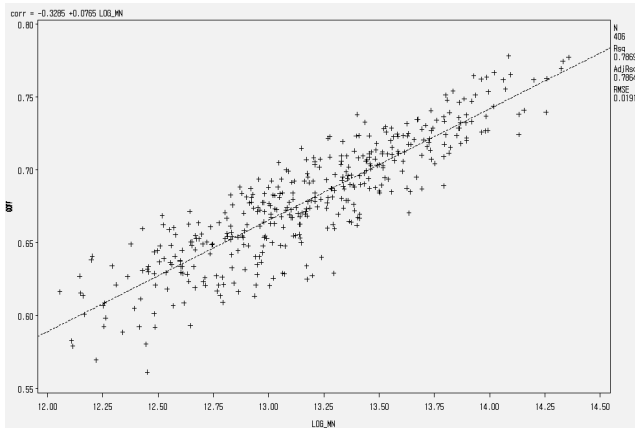


Fig. 6. Plot of correlations between genomic relationship matrices estimated from each chromosome by logarithms of combinations of number of markers on their chromosomes.

Where LOG_MN = $\ln(M \times N)$, M = number of markers on first chromosome, N = number of markers on second chromosome.

염색체의 187두 상호간의 혈연관계에 대한 상관추정치이고 $x_i(x_j)$ 는 i(j)번째 염색체의 표지인자 수이다. Fig. 6에 제시한 바와 같이 두 변량간에는 0.89 이상의 상당히 강한 상관관계가 있는 것으로 추정되었는데, 이는 단일 연관그룹 즉 염색체당 표지인자 수가 많을수록 염색체별 혈연관계 추정치 간에 유사성을 보이는 것을 알 수 있었으며 이는 또한 표지인자 수에 영향받고 있음을 알 수 있었다. 이러한 경향은 DS1 자료를 이용하여 추정된 결과에서도 유사한 경향을 보였는데 이러한 현상을 유전학적 측면에서 해석한다면 비 연관그룹 간에 상호 관련성이 있을 것으로 추측되지만 이에 대한 더 많은 연구가 필요할 것으로 사료되었다.

(2) 혈통기록이 그릇된 경우를 포함한 자료

친자확인이 안된 개체 103두의 유전체 정보를 포함한 총 290두의 유전체 정보를 이용하여 추정된 혈연관계와 혈통정보에 의한 혈연관계를 비교하여 보았다. Fig. 7은 유전자형 결손 자료를 포함한 290두의 유전체 전장의 정보를 이용한 개체간 혈연관계와 혈통정보에 의한 혈연관계를 산포도로 나타낸 것이다. 특이한 사실은 Fig. 7의 A에 표시한 바와 같이 혈통정보에 의한 혈연관계가 0.25인 두 개체간에 유전체 정보를 이용하였을 경우 혈연관계가 약 1.0으로 추정되었다. 이는 실제 개체의 기록내역을 조사해 본 결과 상기의 두 개체는 동일한 개체이거나 일란성 쌍생일 가능성이 있었으나 일란성 쌍생일 가능성 보다는 혈통정보의 오류로 기인한 것으로 판단되었다. 또한 Fig. 7의 B에 표시된 혈통기록에 의한 반형매 집단에 있어서 개체간 혈연관계는 유전체 정보를 이용하였을 경우, 혈연관계가 전혀 없거나 미약한 것으로 판정되어 혈통정보의 오류로 사료되었다. 이러한 현상은 Fig. 7의 C에 해당하는 개체간 혈연관계에 있어서 혈통정보에 의한 혈연관계가 없는 개체간에 유전체를 이용한 혈연관계는 반형매의 관계 또는 그와 유사한 혈연관계를

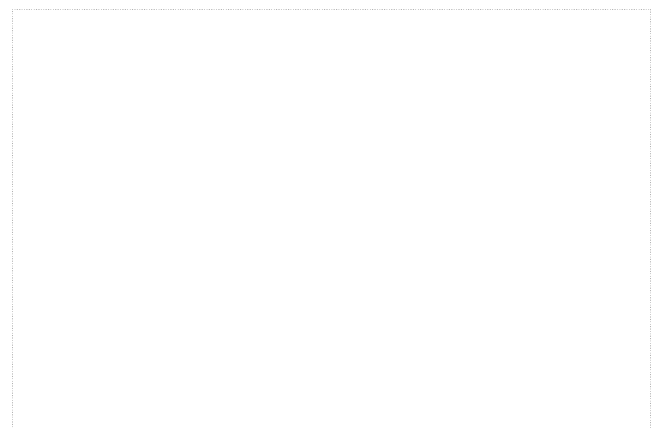


Fig. 7. Plots of genomic relationship matrix by pedigree relationship matrix on data including without paternity test and including miss genotypes (DS2) in the number of 290 Hanwoos.

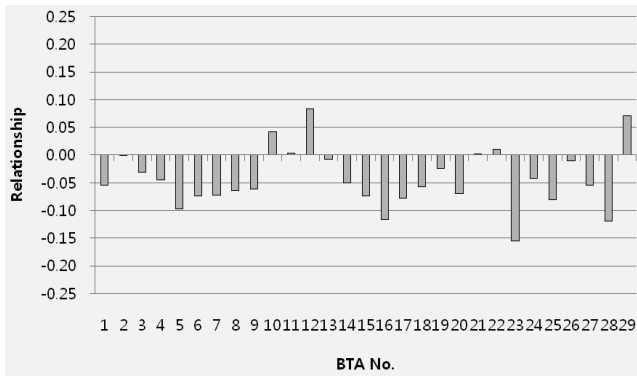


Fig. 8. Plots of genomic relationships estimated by each bovine chromosome (BTA) between two individuals on which relationship has 0.281 by pedigree information.

갖는 것으로 사료되었다. 이러한 현상을 좀더 자세히 알아보기 위하여 혈통정보에 의한 혈연관계가 0.281인 두 개체간에 각 염색체별 유전체 정보에 의한 혈연관계를 추정하여 보았는데(Fig. 8) 각 염색체별 두 개체간 혈연관계는 - 0.16~0.07의 변이로써 평균 혈연관계는 없는 것으로 분석되었다. 여기서 음수의 혈연관계 추정치는 집단내 해당 좌위의 이형접합성의 평균을 '0'으로 수치 변환하여 추정되기 때문에 추정치의 noise로 나타날 수 있는 값이다. 이러한 결과에 의하면 여기에 해당하는 두 개체간의 혈연관계는 없으나 혈통정보의 오류로 반형매의 혈연관계가 있는 것처럼 추정되었다고 사료되었다.

이러한 결과로 살펴볼 때, 그릇된 혈통정보에 의해 잘못 추정되는 개체간 혈연관계를 유전체 정보를 활용하면 이를 보완할 수 있으므로 이러한 결과를 유전능력 평가에 활용하면 유전능력 평가치의 정확도를 크게 향상시킬 수 있을 것으로 사료되었다.

요 약

한우의 유전체 전장의 정보를 Illumina BeadArray™ Bovine SNP50 assay를 이용하여 단일염기다형 현상을 조사한 결과, 유전적 다양성을 보이는 좌위가 약 32,567 좌위 이상에서 다양성을 보이고 있었으며 약 5,554 좌위에서 다양성이 조사되지 않았다. 이는 조사된 자료의 가계집단의 수가 크게 제한되었기 때문에 기인될 수 있으며 또 다른 원인으로는 한우 종축집단의 크기가 작을 수 있다는 현상을 반증한다고 사료된다. 유전분석의 기초가 되는 혈통기록에 의한 개체간 혈연관계를 유전체 정보에 의한 혈연관계와 비교하여 본 결과, 유전체 정보에 의한 혈연관계의 크기가 혈통기록에 의한 혈연관계보다 좀 더 정확하게 추정될 수 있다는 장점이 있으며 혈통기록상의 오류로 그릇된 혈연관계의 크기를 유전체 정보를 통

하여 보완할 수 있다는 장점이 있다. 이러한 장점을 활용하면 유전체정보를 이용한 유전능력 평가의 정확성을 크게 향상시킬 수 있을 것으로 사료되었다.

사 사

본 연구는 농림기술수행과제의 일환으로 수행하였음. 본 연구에 정액 및 DNA 시료를 제공한 농협중앙회 한우개량사업소 및 경북 한우크리스터 사업단에 감사드립니다.

인 용 문 헌

Abdel-Azim, G. and Freeman, A. 2001. A rapid method for computing the inverse of the gametic covariance matrix between relatives for a marked quantitative trait locus. *Genet. Sel. Evol.* 33:153-173.

Abraham, K. J., Totir, L. R. and Fernando, R. L. 2007. Improved techniques for sampling complex pedigrees with the Gibbs sampler. *Genet. Sel. Evol.* 39:27-38.

Aguilar, I. and Misztal, I. 2008. Technical note: Recursive algorithm for inbreeding coefficients assuming nonzero inbreeding of unknown parents. *J. Dairy Sci.* 91:1669-1672.

Arendonk, J. A. M., Tier, B. and Kinghorn, B. P. 1994. Use of multiple genetic markers in prediction of breeding values. *Genetics* 137:319-329.

Blouin, M. S. 2003. DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *TRENDS in Ecology and Evolution* 18:503-511.

Browning, S. R. 2008. Estimation of pairwise identity by descent from dense genetic marker data in a population sample of haplotypes. *Genetics* 178:2123-2132.

Cannings, C. 2003. The identity by descent process along the chromosome. *Human Heredity* 56:126-130.

Carlson, C. S., Eberle, M. A., Rieder, M. J. and Yi, Q. et al. 2004. Selecting a maximally informative set of single nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am. J. Hum. Genet.* 74:106-120.

Chapman, N. H. and Thompson, E. A. 2003. A model for the length of tracts of identity by descent in finite random mating populations. *Theor. Popul. Biology* 64:141-150.

Eding, H. and Meuwissen, T. H. E. 2001. Marker-based estimates of between and within population kinships for the conservation of genetic diversity. *J. Anim. Breed. Genet.* 118:141-159.

Fernando, R. L. and Grossman, M. 1989. Marker assisted selection using best linear unbiased prediction. *Genet. Sel. Evol.* 21:467-

477.

- Gengler, N., Mayeres, P. and Szydlowski, M. 2007. A simple method to approximate gene content in large pedigree populations: application to the myostatin gene in dual-purpose Belgian Blue cattle. *Animal* 1:21-28.
- Guo, S. 1996. Gametogenesis processes and multilocus gene identity by descent. *Am. J. Hum. Genet.* 58:408-419.
- Hayes, B. J. and Goddard, M. E. 2008. Technical note: Prediction of breeding values using marker-derived relationship matrices. *J. Anim. Sci.* 86:2089-2091.
- Herdenson, 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.
- Henderson, C. R. 1984. Applications of linear models in animal breeding. *Can. Catal. Publ. Data. Univ. Guelph, Guelph, Ontario, Canada.*
- Hernandez-Sanchez, J., Haley, C. S. and Woolliams, J. A. 2004. On the prediction of simultaneous inbreeding coefficients at multiple loci. *Genet. Res.* 83:113-120.
- Hernandez-Sanchez, J., Haley, C. S. and Woolliams, J. A. 2006. Prediction of IBD based on population history for fine gene mapping. *Genet. Sel. Evol.* 38:231-252.
- Hill, W. G. and Hernandez-Sanchez, J. 2007. Prediction of multilocus identity-by-descent. *Genetics* 176:2307-2315.
- Hudson, R. 1985. The sampling distribution of linkage disequilibrium under an infinite allele model without selection. *Genetics* 109: 611-631.
- Leutenegger, A. L., Prum, B., Genin, E., Verny, C., Lemainque, A., Clerget-Darpoux, F. and Thompson, E. A. 2003. Estimation of the inbreeding coefficient through use of genomic data. *Am. J. Hum. Genet.* 73:516-523.
- Libiger, O. and Schork, N. J. 2007. A simulation-based analysis of chromosome segment sharing among a group of arbitrarily related individuals. *European J. of Human Genetics* 15:1260-1268.
- Liu, Y., Jansen, G. B. and Lin, C. Y. 2002. The covariance between relatives conditional on genetic markers. *Genet. Sel. Evol.* 34: 657-678.
- Lynch, M. 1988. Estimation of relatedness by DNA fingerprinting. *Mol. Biol. Evol.* 5:584-599.
- Matsuda, H. and Iwaisaki, H. 2002. A recursive procedure to compute the gametic relationship matrix and its inverse for marked QTL clusters. *Genet. Sel. Evol.* 77:123-130.
- Meuwissen, T. H. E. and Goddard, M. E. 2000. Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. *Genetics* 155:421-430.
- Meuwissen, T. H. E. and Goddard, M. E. 2001. Prediction of identity by descent probabilities from marker-haplotypes. *Genet. Sel. Evol.* 33:605-634.
- Meuwissen, T. H. E. and Goddard, M. E. 2007. Multipoint identity-by-descent prediction using dense markers to map quantitative trait loci and estimate effective population size. *Genetics* 176:2551-2560.
- Nolte, I. M. and Meerman, G. J. 2002. The probability that similar haplotypes are identical by descent. *Ann. Hum. Genet.* 66:195-209.
- Perez-Enciso, M., Varona, L. and Rothschild, M. F. 2000. Computation of identity by descent probabilities conditional on DNA markers via a Monte Carlo Markov Chain method. *Genet. Sel. Evol.* 32:467-482.
- Pong-Wong, R., George, A. W., Woolliams, J. A. and Haley, C. S. 2001. A simple and rapid method for calculating identity-by-descent matrices using multiple markers. *Genet. Sel. Evol.* 33:453-471.
- Roberts, A., McMillan L., Wang, W., Parker, J., Rusyn, I. and Threadgill, D. 2007. Inferring missing genotypes in large SNP panels using fast nearest-neighbor searches over sliding windows. *Bioinformatics.* 23:i401-i407.
- Sargolzaei, M., Iwaisaki, H. and Colleau, J. J. 2006. Efficient computation of the inverse of gametic relationship matrix for a marked QTL. *Genet. Sel. Evol.* 38:253-264.
- Scheet, P. and Stephens, M. 2006. A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *Am. J. Human Genetics.* 78:629-644.
- Schumm, J. W., Knowlton, R. C., Braman, J. C., Baarker, D. F., Botsrein, D., Akots, G., Brown, V. A., Gravius, T. C., Helms, C., Hsiao, K., Rediker, K., Thurston, J. G. and Donis-Keller, H. 1988. Identification of more than 500 RFLPs by screening random genomic clones. *Am. J. Hum. Genet.* 42:143-159.
- Snelling, W. M., Chiu, R., Schein, J. E., Hobbs, M., Abbey, C. A., Adelson, D. L., Aerts, J., Bennett, G. L., Bosdet, I. E., Boussaha, M., Brauning R., Caetano, A. R., Costa, M. M., Crawford, A. M., Dalrymple, B. P., Eggen, A., Wind, A. E., Floriot, S., Gautier, M., Gill, C. A., Green, R. D., Holt, R., Jann, O., Jones, S., Kappes, S. M., Keele, J. W., Jong, P. J., Larkin, D. M., Lewin, H. A., McEwan, J. C., McKay, S., Marra2, M. A., Mathewson, C. A., Matukumalli, L. K., Moore, S. S., Murdoch, B., Nicholas, F. W., Osoegawa, K., Roy, A., Salih, H., Schibler, L., Schnabel, R. D., Silveri, L., Skow, L. C., Smith, T., Sonstegard, T. S., Taylor, J., Tellam, R., VanTassel, C., Williams, J. L., Womack, J. E., Wye, N. H., Yang, G. and

- Zhao, S. 2007. A physical map of the bovine genome. the International Bovine BAC Mapping Consortium. *Genome Biology*. 8:R165-1.
- Totir, L. R., Fernando, R. L., Dekkers, J. C. M., Fernandez, S. A. and Gulbrandsen, B. 2004. The effect of using approximate gametic variance covariance matrices on marker assisted selection by BLUP. *Genet. Sel. Evol.* 36:29-48.
- Tuchscherer, A., Mayer, M. and Reinsch, N. 2004. Identification of gametes and treatment of linear dependencies in the gametic QTL-relationship matrix and its inverse. *Genet. Sel. Evol.* 36: 621-642
- VanRaden, P. M. 2007. Genomic measures of relationship and inbreeding. *Interbull Bull.* 37:33-36.
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414-4423.
- Visscher, P. M., Medland, S. E., Ferreira, M. A., Morley, K. I., Zhu, G., Combs, B. K., Montgomery, G. W. and Martin, N. G. 2006. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet.* 2:e41.
- Wang, T., Fernando, R. L., Van Der Beek, S., Grossman, M. and Arendonk, J. A. M. 1995. Covariance between relatives for a marked quantitative trait locus. *Genet. Sel. Evol.* 27:251-274.
- Weir, B. S. and Cockerham, C. C. 1974. Behavior of pairs of loci in finite monoecious populations. *Theoretical Population Biology* 18:396-429.
- Wright, S. 1922. Coefficients of inbreeding and relationship. *The American Naturalist* 56:330-338 (available online at <http://aipl.arsusda.gov/publish/other/wright1922.pdf>, accessed at Aug. 20, 2009)
- (접수일자 : 2010. 4. 21 / 수정일자 : 2010. 10. 11 /
채택일자 : 2010. 10. 14)