Revista Colombiana de Ciencias Pecuarias
ISSN: 0120-0690
rccpecuarias@rccp.udea.edu.co
Universidad de Antioquia
Colombia

Echeverri, Julián; Zambrano, Juan C; López Herrera, Albeiro
Genomic evaluation of Holstein cattle in Antioquia (Colombia): a case study
Revista Colombiana de Ciencias Pecuarias, vol. 27, núm. 4, octubre-diciembre, 2014, pp. 306-314
Universidad de Antioquia
Medellín, Colombia

Case Study

rccp Revista Colombiana de Ciencias Pecuarias

Genomic evaluation of Holstein cattle in Antioquia (Colombia): a case study¤

Evaluación genómica de ganado Holstein en Antioquia (Colombia): estudio de caso

Avaliação genômica do gado Holandês de Antioquia (Colômbia): estudo de caso

Julián Echeverri[1]*, Zoot, MSc, PhD; Juan C Zambrano[2], Qco, MSc, PhD(c); Albeiro López Herrera[1], Zoot, MV, MSc, DrSci.

[1]*Grupo de investigación BIOGEM, Departamento de producción Animal, Facultad de Ciencias Agropecuarias, Universidad Nacional de Colombia, Medellín, Colombia.*

[2]*Grupo de investigación BIOGEM, Posgrado en Biotecnología, Facultad de Ciencias, Universidad Nacional de Colombia, Medellín, Colombia.*

(Received: March 15, 2013; accepted: December 10, 2013)

Summary

Background: DNA markers have been widely used in genetic evaluation throughout the last decade due to the increased reliability of breeding values (BV) they allow, mainly in young animals. **Objective:** to compare breeding values estimated through the conventional method (best linear unbiased predictor, BLUP) with methods that include molecular markers for milk traits in Holstein cattle in Antioquia (Colombia). **Methods:** predictions of breeding values were performed using three methods: BLUP, molecular best linear unbiased predictor (MBLUP), and Bayes C. The breeding values were compared using Spearman's correlation coefficient and linear regression coefficient. **Results:** all Spearman correlation coefficients between breeding values obtained by different methods were greater than 0.5, while linear regression coefficients ranged between -2.10 and 1.58. **Conclusions:** prediction of breeding values through BLUP, MBLUP and Bayes C showed different results in terms of magnitude from the estimated values. However, animal ranking according to breeding values was not significantly different.

Keywords: *genetic markers, genomic selection, breeding value, milk quality, milk traits.*

Resumen

Antecedentes: en la última década, los marcadores de DNA han sido ampliamente usados en evaluaciones genéticas porque incrementan la confiabilidad de valores genéticos principalmente en animales jóvenes. **Objetivo:** comparar valores genéticos (BV) estimados por el método convencional (mejor estimador lineal insesgado, BLUP) y métodos que incluyen marcadores moleculares para algunas características lecheras en

¤ To cite this article: Echeverri J, Zambrano JC, López-Herrera A. Genomic Evaluation of Holstein Cattle in Antioquia (Colombia): a case study. Rev Colomb Cienc Pecu 2014; 27:306-314.

* Corresponding author: Julián Echeverri. Departamento de producción Animal, Facultad de Ciencias Agropecuarias, Universidad Nacional de Colombia. Calle 59A No 63-20 –Núcleo El Volador, Bloque 50, Oficina 306. Medellín, Colombia. Email: jjecheve@unal.edu.co

ganado Holstein de Antioquia (Colombia). **Métodos:** la predicción de valores genéticos se realizó mediante tres métodos: BLUP, mejor predictor lineal insesgado molecular (MBLUP) y Bayes C. Los valores genéticos fueron comparados usando el coeficiente de correlación de Spearman y el coeficiente de regresión lineal. **Resultados:** todos los coeficientes de correlación de Spearman entre los valores genéticos obtenidos por los diferentes métodos fueron mayores de 0,5. Mientras que los coeficientes de regresión lineal oscilaron entre -2,10 y 1,96. **Conclusiones:** la predicción de valores genéticos empleando los métodos BLUP, MBLUP y Bayes C fue diferente en términos de la magnitud de los valores estimados. Sin embargo el ranking o clasificación de los animales por sus valores genéticos no fue alterado significativamente.

Palabras clave: *calidad de leche, características de la leche, marcadores genéticos, selección genómica, valor de cría.*

Resumo

Antecedentes: na última década, os marcadores moleculares que identificam polimorfismos no DNA têm sido utilizados amplamente nas avaliações genéticas porque aumentam a fiabilidade dos valores genéticos (BV) estimados principalmente em animais jovens. **Objetivo:** comparar valores genéticos estimados pelo método convencional (melhor preditor linear não-viesado, BLUP) e métodos que incluem marcadores moleculares para algumas características leiteiras no gado holandês de Antioquia (Colômbia). **Métodos:** as predições dos valores genéticos foram realizadas por meio de três métodos: BLUP, melhor preditor linear não-viesado molecular (MBLUP) e Bayes C. Os valores genéticos foram comparados por meio de coeficientes de correlação de Spearman e de coeficientes de regressão linear. **Resultados:** os coeficientes de correlação de Spearman entre os valores genéticos obtidos pelos diferentes métodos foram maiores que 0,5. Enquanto os coeficientes de regressão linear variaram entre -2,10 e 1,96. **Conclusões:** a predição dos valores genéticos usando os métodos BLUP, MBLUP e Bayes C foi diferente em quanto à magnitude dos valores estimados. No entanto, o ranking ou classificação de animais por seus valores genéticos não foi alterada significativamente.

Palavras chave: *características do leite, marcadores genéticos, qualidade do leite, seleção genômica, valor genético.*

Introduction

Breeding values (BV) are commonly calculated by using the best linear unbiased prediction (BLUP; Henderson, 1984). BV have been very useful to select animals with high genetic merit. However, an important limitation of this method lies in obtaining continuous phenotypic records because of the high costs implied. Accordingly, since the presence of alleles carrying the fundamental causative mutations affecting quantitative traits can determine genetic merit, genetic evaluations in recent years have included information on DNA markers (Haley, 1995).

The use of DNA markers in selection schemes is very useful due to the increased reliability of the estimated breeding values (EBVs), mainly for young animals (Meuwissen and Goddard, 1996). In spite of considerable efforts for implementing marker-assisted selection (MAS), the low density of DNA markers makes it difficult to find markers in linkage disequilibrium (LD) with quantitative trait loci (QTL). Many genes affect quantitative traits. Consequently, the benefit from MAS is limited by the proportion of the genetic variance explained by QTL (Meuwissen *et al.,* 2001).

Meuwissen *et al.* (2001) devised genomic selection, an excellent method to solve the limitations of MAS. This process allows the estimation of BVs using high-density SNP markers. The Single Nucleotide Polymorphism (SNP) markers are uniformly distributed across the entire genome; therefore, each QTL is in LD with some of these markers across the entire population. However, the practical applications of genomic selection became feasible only a few years later with the recent development of DNA chip technologies, which have led to a rapid adoption of this method in selection schemes in order to improve dairy cattle (Schaeffer *et al.,* 2006).

Milk production in Colombia has taken increasing importance. However, few genetic evaluations have been conducted in dairy cattle in Colombia (Quijano

et al., 2011) and few studies including molecular markers have been associated with milk production traits (Rincon *et al.,* 2012). Thus, the current situation of livestock in Colombia requires initiating new research to improve the genetic composition of domestic dairy herds.

The objective of this study was to compare BV estimated through the conventional method BLUP and methods that include molecular markers for milk production traits in Holstein cattle in Antioquia, Colombia.

Materials and methods

This study was approved by the Ethics Committee for Animal Research of the National University from Colombia (Approval letter number: CEMED-015 May, 2012).

Population

The traditional estimated breeding value (TEBV), estimated breeding value (EBV), and molecular estimated breeding value (MEBV) were obtained from 231 Holstein animals (cows and bulls); the genomic estimated breeding value (GEBV) was obtained from 13 Holstein bulls. Phenotypic information was taken from 59 dairy herds located in the high tropics of the Antioquia province, Colombia. The number of lactations used for the analyses were: 1,494, 1,295, 1,645, and 1,140 for milk yield (MY), milk fat percentage (FP), milk protein percentage (PP) and somatic cell count (SCC), respectively. SCC was transformed to somatic cell score (SCS) using the following equation: SCS = [((SCC-100/12) *0.5015) + 0.0434] in order to achieve normality of data distribution (Roman, 2012). All the phenotypic information was managed and analyzed using the Control 1 software, version 1.0 (Echeverri *et al.*, 2010).

Animal genotyping

A total of 231 animals (cows and bulls) were genotyped for bovine growth hormone (bGH), kappa-casein (KC), and prolactin (PRL) genes through PCR-RFLP methodology as described by, Medrano *et al.* (1990), Dybus (2002) and Rincon *et al.* (2012), respectively. Furthermore, 13 Holstein bulls were genotyped using the Illumina BovineSNP50 Beadchip (Illumina, San Diego, CA, USA). The Beadchip provides information of 54,001 SNPs distributed throughout the entire bovine genome (Matukumalli *et al.,* 2009). Upon editing the database of the SNPs, 13 bulls with 40,753 SNPs were available. The database was edited using the SAS/STAT® software, version 9.1 (SAS Institute Inc., Cary, NC, USA) and PLINK programs (Purcell *et al.,* 2007).

Statistical analysis

Analysis of Association for Biallelic Markers. An analysis of association between each marker (bGH, PRL, and KC) and each trait (MY, FP, PP, and SCS) was conducted through a generalized linear model in which the markers were included as fixed effects. The model used for this analysis was:

$$y_{ijklmno} = \mu + PN_i + H_j + \beta_k DL_k + GH_l + KC_m + PRL_n + e_{ijklmno}$$

Where:

$y_{ijklmno}$= dependent variable (MY, FP, PP, and SCS).

μ= overall mean.

PN_i= fixed effect of the *i*th parity.

H_j= fixed effect of the *j*th herd.

β_k= linear regression coefficient of lactation length.

DL_k= lactation length covariate.

GH_l= fixed effect of the *l*th genotype (+/+, +/- and -/-) for the bGH marker.

KC_m= fixed effect of the *m*th genotype (AA, AB and BB) for the KC marker.

PRL_n= fixed effect of the *n*th genotype (AA, AB and BB) for the PRL marker.

$e_{ijklmno}$ = residual.

The statistical analysis was conducted using the General Linear Model (GLM) procedure of SAS/STAT® software, version 9.1 (SAS Institute Inc., Cary,

NC, USA). Differences between treatment means were determined by least squares and analyzed by ANOVA. The Tukey's multiple comparison test was used to compare treatment means ($p<0.05$).

Calculation of traditional estimated breeding value (TEBV). A univariate animal model was used for each trait to estimate TEBV, which was defined as the breeding value obtained using the conventional method (BLUP). The statistical model used for this analysis was:

$$y = Xb + Za + e$$

Where:

y= vector of observations (MY, FP, PP, and SCS).

X= design matrix relating records and fixed effects.

b= vector of the following fixed effects: calving year, calving month, region, contemporary group (Herd-parity number), linear regression coefficients for lactation length covariate (for all traits) and milk production covariate (only for PP, FP, and SCS) respectively.

a= vector of random genetic additive effect.

Z= incidence matrix relating records and random genetic additive effect.

e= residual.

The estimate breeding values (EBV) were predicted in the same way as the TEBV, but included the molecular markers (bGH, PRL and KC) as fixed effects. TEBVs and EBVs were estimated via a derivative-free algorithm by using the MTDFREML program (Boldman *et al*., 1995).

Calculation of molecular estimated breeding value (MEBV). The method used to estimate the molecular marker effects (bGH, KC and PRL) and polygenic effect was the MBLUP (Hayes *et al*., 2009). The model used for this analysis was:

$$y = 1_n'\mu + \sum_{j=1}^{p} X_j \beta_j \delta_j + Zu + e$$

Where:

y= vector of n traditional estimated breeding values (TEBV) corrected for fixed effects as described above (TEBV for MY, FP, PP, and SCS).

m= overall mean.

1_n= vector of 1s.

X= is (n x p) design matrix allocating records to the p markers (KC, bGH and KC), with element X_{ij} = 0, 1 or 2 if the genotype of animal *i* at marker *j* is AA, AB, or BB for KC and PRL genes and +/+, +/- or -/- for the bGH gene, respectively.

g= (p x 1) vector of molecular marker effects (g represents the sum of the linear regression coefficients of TEBV on genotype (0, 1 and 2) of three molecular marker (bGH, KC and PRL).

Z= design matrix allocating records to TEBVs.

u= vector of polygenic effects of the *i*th animal, with variance $A\sigma_u^2$; where *A:* is the average relationship matrix of the animals genotyped with p molecular markers.

e= residual error also assumed to be normally distributed, $e \sim N(0, I\sigma_e^2)$;

where:

I= the n x n identity matrix.

Molecular estimated breeding values (MEBV) were determined through the following equation:

$$y = \hat{u} + X\hat{g}$$

MEBVs were estimated via a derivative-free algorithm by using the MTDFREML software (Boldman *et al*., 1995).

Accuracy of estimated breeding values. The reliabilities of the estimated breeding values (TEBV, EBV and MEBV) were obtained through the following equation: $R^2 = 1\text{-}d_i\alpha$,

where:

d_i= *i*th diagonal element of C^{22} of the generalized inverse of the mixed model equations, $\alpha = \sigma_e^2 / \sigma_a^2$ and accuracy (R) is the square root of reliability (Mrode and Thompson, 2005).

Calculation of genomic estimated breeding values (GEBV). The estimation of GEBVs was carried out in two steps through the Bayes C method. 1) Estimation of the effects of each SNP marker and, 2) Prediction of the genomic estimated breeding values (GEBV). Bayes C method assumes a mixture of distributions for the SNP effects reflecting the assumption that there is a large number of SNPs with zero or near zero effect and a second smaller set of SNPs with larger effect (Kizilkaya *et al.,* 2010, Verbyla *et al.,* 2010). The general statistical model may be written as:

$$y = 1_n'\mu + \sum_{j=1}^{p} X_j \beta_j \delta_j + Zu + e$$

Where:

y= is the vector of traditional estimated breeding values (TEBV) corrected for fixed effects as described above (TEBV for MY, FP, PP, and SCS) for n individuals (n = 13 bulls).

μ= overall mean.

1_n= vector of ones of length n.

X_j= vector of indicator variables representing the genotypes of the *j*th marker for all individuals, at each *j*th marker there are three possible combinations of two alleles (A or B), the homozygote of one allele (AA), the heterozygote (AB) and homozygote of the other allele (BB); these are then quantitatively represented by 0, 1 and 2 respectively (i.e., X_{ij} = 0, 1 or 2).

β_j= is the random substitution effect for locus *j*, which is conditional on σ_β^2 and is assumed normally distributed $N(0, \sigma_\beta^2)$ when δ_j = 1, but β_j = 0 when δ_j = 0.

δ_j= is a random 0/1 variable indicating the absence (with probability π) or presence (with probability 1-π) of locus j in the model.

u= vector of random polygenic effects of length n (Z is the associated design matrix) and can be thought of as fitting the genes no accounted for by the markers-locus effects in β, additionallv u is assumed to be normally distributed, $u \sim N(0, A\sigma_u^2)$ where A is the pedigree derived additive genetic relationship matrix of the genotyped animals.

e= residual error, also assumed to be normally distributed, $e \sim N(0, I\sigma_e^2)$ here I = the nxn identity matrix.

GEBVs of the animals (whose genotype was known) were predicted through the following equation:

$$y = X\hat{\beta} + \hat{u}$$

The SNP effects and GEBVs were obtained by using the GS3 program (Legarra *et al.,* 2011a).

Methods for comparing breeding values

The Spearman's rank correlation coefficients between the breeding values obtained by BLUP, MBLUP and Bayes C methods ($r_{TEBV;MEBV}$, $r_{EBV;MEBV}$ and $r_{EBV;GEBV}$) were calculated and used as a measure of the degree of similarity between the ranking or classification of the animals by their breeding values. The linear regression coefficients ($b_{TEBV;MEBV}$, $b_{EBV;MEBV}$ and $b_{EBV;GEBV}$) were also calculated and used as a measure of the change in magnitude between the breeding values. A regression coefficient of one indicates no bias between the methods of prediction and that the breeding values are equal in magnitude.

Results

Descriptive analysis of milk traits

The mean and standard deviations for MY, PP, FP and SCS were: 5324 ± 1437 L/lactation, 3.03 ± 0.24%, 3.67 ± 0.43%, and 17.7 ± 39.37, respectively (Table 1). MY and SCS were the traits with greatest coefficients of variation (26.9 and 222%, respectively).

Table 1. Descriptive analysis of milk traits for Holstein cattle in Antioquia.

Trait	Mean±SD	CV	Min	Max
MY (L/lactation)	5324 ± 1437	26.9	3000	9000
PP (%)	3.03 ± 0.24	8.0	2.50	3.92
FP (%)	3.67 ± 0.43	11.8	2.05	4.50
SCS	17.7 ± 39.37	222	-37	54

MY: milk yield; PP: protein percentage; FP: fat percentage; SCS: somatic cell score; SD: standard deviation; Min: minimum; Max: maximum; CV: coefficient of variation (%).

Association analysis for biallelic markers and milk traits

Table 2 shows the genotype frequencies of the PRL, bGH and KC genes and the means of each trait per genotype. Through the use of Tukey's multiple comparison test it was possible to determine that genotypes AA and AB of PRL gene ($p<0.01$) and genotype BB of KC gene ($p<0.05$) were the most favorable for MY. On the other hand, BB genotype of PRL gene ($p<0.05$), the genotype (+/-) of bGH gene ($p<0.05$) and genotype BB of KC gene were the most favorable for PP ($p<0.05$). In the case of FP, only genotype (-/-) of bGH gene showed significant association with greater fat content in milk ($p<0.01$).

Estimated breeding values (including molecular markers)

The TEBV, EBV and MEBV means were close to zero in all cases, but the coefficient of variation (CV) and accuracy (R) differed among them. Accuracies (R)

Table 2. Association between genotypic frequencies for PRL, bGH and KC gene and milk traits in Holstein cattle of Antioquia.

Trait	PRL			bGH			KC		
	Genotype	GF	Mean	Genotype	GF	Mean	Genotype	GF	Mean
MY (L/lactation) (n = 1024)	AA	74.7	5549^{A}	+/+	77.1	5558^{a}	AA	59.1	5600^{ab}
	AB	23.6	5520^{A}	+/-	21.7	5392^{a}	AB	35.9	5363^{b}
	BB	1.7	4773^{B}	-/-	1.2	6152^{a}	BB	5.0	5875^{a}
PP (%) (n = 957)	AA	74.4	3.05^{a}	+/+	77.0	3.05^{a}	AA	59.0	3.04^{a}
	AB	23.9	3.06^{a}	+/-	21.8	3.08^{a}	AB	35.7	3.06^{a}
	BB	1.7	3.19^{b}	-/-	1.2	2.94^{b}	BB	5.3	3.13^{b}
FP (%) (n = 972)	AA	74.3	3.75^{a}	+/+	76.9	3.73^{A}	AA	58.9	3.77^{a}
	AB	24.1	3.75^{a}	+/-	21.9	3.82^{A}	AB	35.9	3.75^{a}
	BB	1.6	3.97^{a}	-/-	1.2	4.38^{B}	BB	5.2	3.67^{a}

MY: milk yield; PP: protein percentage; FP: fat percentage; n: number of animals; GF: genotypic frequencies; PRL: prolactin; bGH: bovine growth hormone, KC: kappa casein; columns with different superscripts differ significantly: capital ($p<0.01$); small letter ($p<0.05$).

Table 3. Descriptive analysis of breeding values obtained by using BLUP and MBLUP methods for milk traits in Holstein cattle of Antioquia.

Trait	EBV		MEBV			TEBV		
	Mean±SD	CV	Mean±SD	CV	R	Mean±SD	CV	R
MY (L/lactation)	-3.18± 134	42	-2.84±17	58	0.31	3.47±345	99	0.40
FP (%)	0.00±0.13	32	0.08±0.17	2	0.50	0.01±0.1	9	0.37
PP (%)	-0.01±0.13	20	0.00±0.13	1300	0.69	0.00±0.07	165	0.40
SCS	-1.54±24	16	-1.40±24	17	0.86	0.09±2	28	0.30

MY: milk yield; PP: protein percentage; FP: fat percentage; SCS: somatic cell score; SD: standard deviation; CV: coefficients of variation (%); R: accuracy of the estimated breeding values.

were greater for MEBVs compared to TEBVs in all traits except for MY (Table 3).

SNP effects and genomic estimated breeding values (GEBVS)

The effects of 40,753 SNPs were determined for MY, PP, FP and SCS, and their means were: -0.03520 L/lactation, -0.000034, -0.00019 and 0.000048%, respectively (Table 4).

On the other hand, the GEBVs for MY, PP, FP and SCS were estimated and means and standard deviations were: 359 ± 311 L/lactation, 0.123 ± 0.19%, 0.276 ± 0.20%, 0.501 ± 0.75, respectively (Table 4).

Table 4. Descriptive analysis of the SNP effects and GEBVs obtained by using Bayes C method for milk traits in Holstein cattle of Antioquia.

Trait	SNP effects			GEBV		
	Mean±SD	Min	Max	Mean±SD	Min	Max
MY (L/lactation)	-0.0352 ± 11.423	-64.2744	62.0965	359±311	-186	802
PP (%)	-0.000034±0.029	-0.25295	0.22176	0.12±0.19	-0.30	0.32
FP (%)	-0.00019±0.024	-0.16961	0.17709	0.18±0.20	0.00	0.88
SCS	0.00005±0.024	-0.20082	0.18344	0.50±0.75	-1.33	1.08

MY: milk yield; PP: protein percentage; FP: fat percentage; SCS: somatic cell score; SD: standard deviation; Min: minimum; Max: maximum.

Correlation and regression coefficients between breeding values

The Spearman correlation coefficients between EBV and MEBV for MY, FP, PP and SCS were: 0.796, 0.763, 0.936 and 0.999, respectively; and between TEBV and MEBV were: 0.823, 0.783, 0.962 and 0.620, respectively. These results indicate a high and favorable degree of association between breeding values. Finally, the correlations between EBV and GEBV were medium: 0.780, 0.500, 0.500 and 0.580, since the number of phenotypic records for EBVs was greater than for GEBVs (Table 5).

The comparison of breeding values obtained by different methods (BLUP, MBLUP and Bayes C) shows that regression coefficients were highly variable. For example, the regression coefficients of EBV on MEBV for MY, FP, PP, and SCS were: -2.140, 0.205, -0.015 and 0.999, respectively; for TEBV on MEBV were: 1.227, 1.163, 1.958, and 0.003, respectively; and finally, the regression coefficients of EBV on GEBV were: 0.784, 0.077, 0.380, and 1.110, respectively (Table 5).

Table 5. Correlation and regression between breeding values obtained using BLUP, MBLUP and Bayes C methods for milk traits in Holstein cattle of Antioquia.

Trait	Correlation Coefficient (SE)			Regression Coefficient (SE)		
	$r_{EBV;MEBV}$	$r_{TEBV;MEBV}$	$r_{EBV;GEBV}$	$b_{EBV;MEBV}$	$b_{TEBV;MEBV}$	$b_{EBV;GEBV}$
MY	0.796 (0.040)	0.823 (0.038)	0.780 (0.04)	-2.140 (0.290)	1.227 (0.150)	0.784 (0.194)
FP	0.763 (0.043)	0.783 (0.041)	0.500 (0.057)	0.205 (0.190)	1.163 (0.250)	0.077 (0.018)
PP	0.936 (0.023)	0.962 (0.018)	0.500 (0.057)	-0.015 (0.007)	1.958 (0.150)	0.380 (0.150)
SCS	0.999 (0.008)	0.620 (0.052)	0.580 (0.054)	0.999 (0.000)	0.003 (0.000)	1.110 (0.360)

MY: milk yield; PP: protein percentage; FP: fat percentage; SCS: somatic cell score; SE: standard error.

Discussion

Traditionally, breeding values are obtained by using the best linear unbiased predictor (BLUP) (Henderson, 1984), which assumes that phenotypic traits are determined by an infinite number of unlinked additive loci, each one having an infinitesimal small effect (infinitesimal model) (Fisher, 1918). However, the finite loci model has been proposed to explain the genetic variation observed in quantitative traits. This model assumes a finite number of loci that explains the genetic variation of quantitative traits (Thompson and Skolnick, 1977). In this perspective, several methods that include molecular markers have been evaluated to estimate breeding values.

Legarra *et al*. (2011b) evaluated five methods that include molecular markers (Bayesian Lasso with one variance (BL1Var), Bayesian Lasso with two variances (BL2Var), GBLUP, MCMC-GBLUP and Het-Var-GBLUP). The genomic estimated breeding values (GEBV) obtained through those methods were compared with the double daughter yield deviation (2DYD) by the correlation coefficient ($r_{2DYD;GEBV}$). The correlations between 2DYD and GEBV (obtained through the methods mentioned previously) ($r_{2DYD;GEBV}$) for fat percentage (FP) were: 0.53, 0.73, 0.59, 0.61, and 0.71, respectively; and for protein percentage (PP) were: 0.36, 0.48, 0.44, 0.46, and 0.47, respectively. We found similar results for PP and FP using MBLUP and Bayes C methods (Table 5).

On the other hand, Moser *et al.* (2009) evaluated the following methods: fixed regression-least squares (FR-LS), random regression BLUP (RR-BLUP), Bayes A, support vector regression (SVR), and partial least squares regression (PLSR). They estimated the molecular breeding value (MBV) of young Holstein bulls using only genomic information and the GEBV obtained from the same bulls (combining the MBV with the pedigree). The MBVs and GEBVs obtained through the previously mentioned methods were compared with the Australian estimated breeding value (EBV) by using the correlation coefficient. Correlations between EBV and MBV ($r_{EBV;MBV}$) were: 0.43, 0.56, 0.56, 0.58 and 0.55, respectively; and between EBV and GEBV ($r_{EBV;GEBV}$) were: 0.49, 0.57, 0.60, 0.62, 0.60, and 0.62, respectively. The correlations obtained by Moser *et al.* (2009) were medium, and the authors attributed these results to the low amount of data. Legarra *et al.* (2011b) suggests that if correlations are high (equal or close to 1), prediction methods have the same accuracy and the prediction errors of breeding values are very similar.

We calculated correlations between breeding values ($r_{EBV;MEBV}$, $r_{TEBV;MEBV}$, and $r_{TEBV;GEBV}$) for milk traits (MY, PP, FP, and SCS), which ranged from 0.500 to 0.999. Furthermore, considering the correlations between EBV and MEBV ($r_{EBV;MEBV}$) for PP (0.936) and SCS (0.999), and between TEBV and MEBV ($r_{TEBV;MEBV}$) for PP (0.962), the ranking was not affected

The regression coefficients of TEBV on MEBV and EBV on GEBV ($b_{TEBV;MEBV}$ and $b_{EBV;GEBV}$) obtained in this study were different from 1 (ranged from -2.10 to 1.58). These regression coefficients should ideally be 1. However, the regression coefficients were less than 1 for MY ($b_{EBV;GEBV}$ = 0.784), FP ($b_{EBV;GEBV}$ = 0.077) and PP ($b_{EBV;GEBV}$ = 0.380) and were greater than 1 for MY ($b_{TEBV;MEBV}$ = 1.227), FP ($b_{TEBV;MEBV}$ = 1.163), and PP ($b_{TEBV;MEBV}$ = 1.958).

Bennewitz *et al.* (2009) determined GEBVs using Bayes-BLUP method and two nonparametric kernel regressions methods (ELM, ULM). The GEBVs were compared with true estimate breeding value (TEBV) (obtained by simulation) and determined the regression coefficients of TEBVs on GEBVs ($b_{TEBV;GEBV}$), which were: 1.376, 0.722, and 0.626, respectively. On the other hand, Legarra *et al.* (2011b) obtained regression coefficients of 2DYD on GEBV ($b_{2DYD;GEBV}$) (using the previously mentioned methods). The regression coefficients ($b_{2DYD;GEBV}$) for PP were 0.35, 1.10, 0.83, 1.10, and 0.99, respectively; and for MY were 0.25, 0.67, 0.59, 0.66, and 0.67, respectively. Legarra *et al.* (2011b), suggest that most of the methods frequently inflate the variances of the genomic estimated breeding values (GEBVs) for some production traits, thus obtaining regression values below 1. Contrary to this, they also suggest that genetic variance is captured by QTL with large effect on some compositional traits, what leads to regression values greater than 1.

The prediction of breeding values (TEBV, EBV, MEBV and GEBV) by using BLUP, MBLUP and Bayes C methods showed different results in terms of magnitude from the estimated values according to

the regressions obtained. However, the correlations between breeding values obtained by using methods that include molecular markers were similar, despite the different assumptions underlying the models. Finally, the results suggest that it is necessary to increase the number of records and genotyped animals to improve the prediction of GEBVs.

Acknowledgements

The authors wish to thank the Universidad Nacional de Colombia (Medellin), COLANTA Cooperative, and the Colombian Ministry of Agriculture and Rural Development for their valuable support in conducting this work.

Conflicts of interest

The authors declare they have no conflicts of interest with regard to the work presented in this report.

References

Bennewitz J, Solberg T, Meuwissen THE. Genomic breeding value estimation using nonparametric additive regression models. Genet Sel Evol 2009; 41:20.

Boldman K, Kriese L, Van Vleck L, Van Tassell C, Kachman S. A Manual for Use of MTDFREML. A Set of programs to obtain estimates of variances and covariances. USDA, Agricultural Research Service; 1995.

Dybus A. Associations between Leu/Val polymorphism of growth hormone gene and milk production traits in Black and White cattle. Arch Tiertz 2002; 45:421-428.

Echeverri J, López A, Parra J. Software control 1. Manejo y control de producción para hatos lecheros. Universidad Nacional de Colombia sede Medellín; 2010.

Fisher R. The correlation between relatives on the supposition of mendelian inheritance. T Roy Soc Edin 1918; 52:399-433.

Haley S. Livestock QTLs-bringing home the bacon? Trends Genet 1995; 11:488-492.

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. Invited review: Genomic selection in dairy cattle: Progress and challenges. J Dairy Sci 2009; 92:433-443.

Henderson CR. Applications of linear models in animal breeding. 3ra ed. Guelph (ON): CGIL Publications; 1984

Kizilkaya K, Fernando RL, Garrick DJ. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand nucleotide polymorphism genotypes. J Anim Sci 2010; 88:544-551.

Legarra A, Ricard A, Filangi O. GS3, Genomic selection, Gibbs sampling, Gauss Seidel and Bayes Cpi. 2011a. [Access date: Feb 10, 2011] URL: http://snp.toulouse.inra.fr/~alegarra/

Legarra A, Robert-Granié C, Croiseau P, Guillaume F, Fritz S. Improved Lasso for genomic selection. Genet Res Camb 2011b; 93:77-87.

Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, Jeff O'Connell, Moore SS, Smith TPL, Sonstegard TS, Van Tassell CP. Development and Characterization of a High Density SNP Genotyping Assay for Cattle. PloS ONE 2009; 4:e5350.

Medrano J, Aguilar E. Genotyping of bovine kappa-casein loci following DNA sequence amplification. Biotechnology 1990; 8:144-146.

Meuwissen THE, Hayes B, Goddard M. Prediction of total genetic value using genome-wide dense marker maps. Genetics 2001; 157:1819-1829.

Meuwissen THE, Goddard ME. The use of marker haplotypes in animal breeding schemes. Genet Sel Evol 1996; 28:161-176.

Moser G, Tier B, Crump RE, Khatkar MS, Raadsma HW. A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. Genet Sel Evol 2009; 41:56.

Mrode RA, Thompson R. Linear models for the prediction of animal breeding values. 2nd ed. Cambridge (MA): CABI Publishing; 2005.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D, Maller J, Sklar P, Paul IW de Bakker, Daly M, Sham P. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am J Hum Genet 2007; 81:559-575.

Quijano JH, Echeverri J, López A. Evaluación genética de toros Holstein y Jersey en Condiciones tropicales. Medellín (Colombia): Centro de Publicaciones Universidad Nacional de Colombia Sede Medellín; 2011.

Rincon JC, López A, Echeverri J. Genetic variability of the bovine prolactin-Rsal loci in Holstein cattle in Antioquia province (Colombia). Rev Col Cienc Pec 2012; 25:191-201.

Roman SI. Genomic aspects of genetic improvement for mastitis resistance in dairy cattle. (Thesis PhD). Universitá degli Studi di Milano, Italy; 2012.

SAS/STAT. User's Guide. Version 9.1. Cary (NC, USA): SAS Institute Inc; 2006.

Schaeffer LR. Strategy for applying genome-wide selection in dairy cattle. J Anim Breed Genet 2006; 123:218-223.

Thompson EA, Skolnick MH. Likelihoods on complex pedigrees for quantitative traits. In: Pollack E, Kempthorne O, Bailey TB, editors. Proc Int Conf Quant Genet. Ames, Iowa: Iowa State University Press; 1977. p.815-818.

Verbyla KL, Bowman PJ, Hayes BJ, Raadsma H, Goddard ME. Sensitivity of genomic selection to using different prior distributions. BMC Proceeding 2010, 4(1):S5.